

EAI/Springer Innovations in Communication and Computing

Chirag Paunwala · Mita Paunwala ·
Rahul Kher · Falgun Thakkar ·
Heena Kher · Mohammed Atiquzzaman ·
Norliza Mohd. Noor *Editors*

Biomedical Signal and Image Processing with Artificial Intelligence

 **EAI**
RESEARCH MEETS INNOVATION

 Springer

EAI/Springer Innovations in Communication and Computing

Series Editor

Inrich Chlamtac, European Alliance for Innovation, Ghent, Belgium

The impact of information technologies is creating a new world yet not fully understood. The extent and speed of economic, life style and social changes already perceived in everyday life is hard to estimate without understanding the technological driving forces behind it. This series presents contributed volumes featuring the latest research and development in the various information engineering technologies that play a key role in this process. The range of topics, focusing primarily on communications and computing engineering include, but are not limited to, wireless networks; mobile communication; design and learning; gaming; interaction; e-health and pervasive healthcare; energy management; smart grids; internet of things; cognitive radio networks; computation; cloud computing; ubiquitous connectivity, and in mode general smart living, smart cities, Internet of Things and more. The series publishes a combination of expanded papers selected from hosted and sponsored European Alliance for Innovation (EAI) conferences that present cutting edge, global research as well as provide new perspectives on traditional related engineering fields. This content, complemented with open calls for contribution of book titles and individual chapters, together maintain Springer's and EAI's high standards of academic excellence. The audience for the books consists of researchers, industry professionals, advanced level students as well as practitioners in related fields of activity include information and communication specialists, security experts, economists, urban planners, doctors, and in general representatives in all those walks of life affected ad contributing to the information revolution.

Indexing: This series is indexed in Scopus, Ei Compendex, and zbMATH.

About EAI - EAI is a grassroots member organization initiated through cooperation between businesses, public, private and government organizations to address the global challenges of Europe's future competitiveness and link the European Research community with its counterparts around the globe. EAI reaches out to hundreds of thousands of individual subscribers on all continents and collaborates with an institutional member base including Fortune 500 companies, government organizations, and educational institutions, provide a free research and innovation platform. Through its open free membership model EAI promotes a new research and innovation culture based on collaboration, connectivity and recognition of excellence by community.

Chirag Paunwala • Mita Paunwala • Rahul Kher •
Falgun Thakkar • Heena Kher •
Mohammed Atiquzzaman • Norliza Mohd. Noor
Editors

Biomedical Signal and Image Processing with Artificial Intelligence

 Springer

 **EAI**
RESEARCH MEETS INNOVATION

Editors

Chirag Paunwala
Electronics & Communication Engineering
Sarvajanik College of Engineering and
Technology
Surat, India

Mita Paunwala
Electronics & Communication Engineering
C. K. Pithawala College of Engineering and
Technology
Surat, India

Rahul Kher
Electronics & Communication Engineering
G. H. Patel College of Engineering &
Technology
Vallabh Vidyanagar, Gujarat, India

Falgun Thakkar
Electronics & Communication Engineering
G. H. Patel College of Engineering &
Technology
Vallabh Vidyanagar, Gujarat, India

Heena Kher
A. D. Patel Institute of Technology
New Vallabh Vidyanagar, India

Mohammed Atiquzzaman
School of Computer Science
University of Oklahoma
Norman, OK, USA

Norliza Mohd. Noor
UTM Razak School, Menara Razak
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia

ISSN 2522-8595

ISSN 2522-8609 (electronic)

EAI/Springer Innovations in Communication and Computing

ISBN 978-3-031-15815-5

ISBN 978-3-031-15816-2 (eBook)

<https://doi.org/10.1007/978-3-031-15816-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book mainly focuses on advanced techniques used for feature extraction, analysis, recognition, and classification in the area of biomedical signal and image processing. This is the era of the Internet of Things, cloud computing, and artificial intelligence for biomedical signals and images. This book will provide a great platform to the researchers who are working in the area of artificial intelligence for biomedical applications. Biomedical data analysis plays a very important role in research as well as in clinical purpose for different types of diagnosis. Moreover, processing a huge amount of data is a challenging task that requires parallel processing. For advanced-level research, deep learning–based approaches have been adopted by researchers since the last few years. The chapters in this book will cover all aspects of artificial intelligence, machine learning, and deep learning in the field of biomedical signal and image processing using novel and unexplored techniques and methodologies.

Chapter “[Voice Privacy in Biometrics](#)” summarizes voice privacy in biometrics. In this chapter, the design of second-order resonator and the linear prediction modeling of speech production is exploited to design voice privacy system. The performance of the proposed system is compared with the secondary baseline system of the INTERSPEECH 2020 voice privacy challenge. Improved performance-wise EER and WER are achieved for various subsets of the corpora. Furthermore, anonymization is achieved by cryptography.

Chapter “[Histopathology Whole-Slide Image Analysis for Breast Cancer Detection](#)” lists a novel method for weakly supervised histopathology whole-slide image (WSI) classification for addressing the breast cancer detection task. Some of the salient aspects of our approach include extracting embeddings from a pre-trained CNN network, using a cosine-loss and training schedule for the classification network, and suggesting an overall decision-making criteria for the WSI based on intermediate decisions on local random selection. We also provide an extensive review of closely related methods with an elaborate compression analysis of the embeddings used in these.

Chapter “[Lung Classification for Covid-19](#)” presents a cloud-based lung disease classification system where medical practitioner can upload their patients’ chest X-

ray onto the cloud, and the system will classify either disease absent (normal) and disease present (abnormal). For disease present, the system will then classify into lung infected with Covid-19 and non-Covid.

Chapter “[GRU-Based Parameter-Efficient Epileptic Seizure Detection](#)” presents a gated recurrent unit-based deep learning architecture for accurate epileptic seizure detection has been proposed to reduce the burden on the medical and the paramedical fraternity. The developed model automates the entire process and circumvents the requirement of deploying any manual feature-extraction steps.

Chapter “[An Object Aware Hybrid U-Net for Breast Tumour Annotation](#)” describes digital examination of histopathological slides, and the pathologist annotates the slides by marking the rough polygonal boundary around the suspected tumor region. The polygonal boundary covers the extent of the tumor in the slide.

Chapter “[VLSI Implementation of sEMG Based Classification for Muscle Activity Control](#)” presents the real-time classification of EMG-based pattern recognition using linear discriminant analysis (LDA) and quadratic discriminant analysis.

Chapter “[Content Based Image Retrieval Techniques and Their Applications in Medical Science](#)” addresses the CBIR techniques which are classified into multiple categories based on the feature extraction and retrieval mechanism. These categories are feature-based, machine learning-based, and deep learning-based methods. The pioneer techniques for each category are explained in detail in this chapter.

Chapter “[Data Analytics on Medical Images with Deep Learning Approach](#)” discusses several techniques and decision policies to dynamically decide computation offloading for smart devices. It adopts a binary offloading policy so that each task of the smart device is executed onboard or completely offloaded to the Edge Server. The algorithms jointly optimize the dense network of devices and reduce the overall latency and increase the battery lifetime.

Chapter “[Analysis and Classification of Dysarthric Speech](#)” attempts to understand how dysarthric speech is different from normal speech through various analyses, such as time-domain representation, linear prediction residual, Teager energy profile, and time-frequency-domain representation. In addition, this chapter also explores the deep learning method for the classification of normal vs dysarthric speech.

Chapter “[Skin Cancer Detection and Classification Using DWT-GLCM with Probabilistic Neural Networks](#)” presents skin cancer detection and classification using DWTGLCM with probabilistic neural networks. Authors used the maximum efficiency of the system by using, PNN for classification of skin cancer with the gray level co-occurrence matrix(GLCM); discrete wavelet transform (DWT) and statistical color features, respectively.

Chapter “[Manufacturing of Medical Devices Using Artificial Intelligence Based Troubleshooter](#)” discusses a process in which an artificial intelligent (AI) agent, independent of human skills, would learn the tricks of trade in exactly the same fashion as a human would. This work showcases an AI agent that gains knowledge of the manufacturing process exactly the same way as an operator learns on the production floor.

Chapter “[Enhanced Hierarchical Prediction for Lossless Medical Image Compression in the Field of Telemedicine Application](#)” addresses two algorithms

of MHPCA for high frequency regions which improves coding efficiency and temporal scalability for Enhanced Hierarchical Prediction for Lossless Medical Image Compression in the Field of Telemedicine Application.

Chapter “[LBP Based CAD System Designs for Breast Tumor Characterization](#)” proposes an efficient CAD system for characterization of breast ultrasound images based on LBP texture features and morphological features. The results illustrate that CAD system based on ANFC-LH algorithm yields optimal performance for breast tumor characterization.

Chapter “[Detection of Fetal Abnormality Using ANN Techniques](#)” proposes an approach of neural modeling for the diagnosis of fetus abnormality using ultrasound (US) images. The proposed method is a hybrid approach to image processing methods and artificial neural network as a classifier to extract fetus abnormality.

Chapter “[Machine Learning and Deep Learning-Based Framework for Detection and Classification of Diabetic Retinopathy](#)” is a review of examining the prior and recent new algorithms designed to automatically detect and classify diabetic retinopathy.

Chapter “[Applications of Artificial Intelligence in Medical Images Analysis](#)” discusses how the use of AI has shown promising results in the field of radiology, where the disease can be diagnosed and assessed accurately for efficient decision-making and planning of the treatment procedures.

Chapter “[Intelligent Image Segmentation Methods Using Deep Convolutional Neural Network](#)” presents the underlying general mathematical operations combined with the currently used handy performance metrics for Intelligent Image Segmentation Methods using Deep Convolutional Neural Network.

Chapter “[Artificial Intelligence Assisted Cardiac Signal Analysis for Heart Disease Prediction](#)” discusses a detailed survey of various mathematical and artificial intelligence (AI)-based cardiac signal analysis models for coronary disease prediction.

Chapter “[Early Lung Cancer Detection by Using Artificial Intelligence System](#)” is about computer-aided diagnosis (CAD) system used for the prediction of lung cancer, which helps to attain a high detection rate and reduces the time consumed for analyzing the sample.

Chapter “[An Optimal Model Selection for COVID 19 Disease Classification](#)” introduces a study for understanding which deep learning models give the best result when classifying COVID-19 patients using chest CT images.

Surat, India
 Surat, India
 Vallabh Vidyanagar, Gujarat, India
 Vallabh Vidyanagar, Gujarat, India
 Vallabh Vidyanagar, Gujarat, India
 Norman, OK, USA
 Kuala Lumpur, Malaysia

Chirag Paunwala
 Mita Paunwala
 Rahul Kher
 Falgun Thakkar
 Heena Kher
 Mohammed Atiquzzaman
 Norliza Mohd. Noor

Contents

Voice Privacy in Biometrics	1
Priyanka Gupta, Shrishti Singh, Gauri P. Prajapati, and Hemant A. Patil	
Histopathology Whole Slide Image Analysis for Breast Cancer Detection	31
Pushap Deep Singh, Arnav Bhavsar, and K. K. Harinarayanan	
Lung Classification for COVID-19	57
Norliza Mohd. Noor and Muhammad Samer Sallam	
GRU-Based Parameter-Efficient Epileptic Seizure Detection	73
Ojas A. Ramwala, Chirag N. Paunwala, and Mita C. Paunwala	
An Object Aware Hybrid U-Net for Breast Tumour Annotation	87
Suvidha Tripathi and Satish Kumar Singh	
VLSI Implementation of sEMG Based Classification for Muscle Activity Control	107
Amit M. Joshi, Natasha Singh, and Sri Teja	
Content-Based Image Retrieval Techniques and Their Applications in Medical Science	123
Mayank R. Kapadia and Chirag N. Paunwala	
Data Analytics on Medical Images with Deep Learning Approach	153
S. Saravanan, K. Surendheran, and K. Krishnakumar	
Analysis and Classification Dysarthric Speech	167
Siddhant Gupta and Hemant A. Patil	
Skin Cancer Detection and Classification Using DWT-GLCM with Probabilistic Neural Networks	183
J. Pandu, Umadevi Kudtala, and B. Prabhakar	
Manufacturing of Medical Devices Using Artificial Intelligence-Based Troubleshooters	195
Akbar Doctor	

Enhanced Hierarchical Prediction for Lossless Medical Image Compression in the Field of Telemedicine Application	207
Ketki C. Pathak, Jignesh N. Sarvaiya, and Anand D. Darji	
LBP-Based CAD System Designs for Breast Tumor Characterization	231
Kriti, Jitendra Virmani, and Ravinder Agarwal	
Detection of Fetal Abnormality Using ANN Techniques	259
Vidhi Rawat, Vibhakar Shrimali, Alok Jain, and Abhishek Rawat	
Machine Learning and Deep Learning-Based Framework for Detection and Classification of Diabetic Retinopathy	271
V. Purna Chandra Reddy and Kiran Kumar Gurralla	
Applications of Artificial Intelligence in Medical Images Analysis	287
Pushpanjali Gupta and Prasan Kumar Sahoo	
Intelligent Image Segmentation Methods Using Deep Convolutional Neural Network	309
Mekhla Sarkar and Prasan Kumar Sahoo	
Artificial Intelligence Assisted Cardiac Signal Analysis for Heart Disease Prediction	337
Prasan Kumar Sahoo, Sulagna Mohapatra, and Hiren Kumar Thakkar	
Early Lung Cancer Detection by Using Artificial Intelligence System	373
Fatma Taher	
An Optimal Model Selection for COVID 19 Disease Classification	399
Prמוד Gaur, Vatsal Malaviya, Abhay Gupta, Gautam Bhatia, Bharavi Mishra, Ram Bilas Pachori, and Divyesh Sharma	
Index	417

About the Authors

Chirag Paunwala is working as a professor, EC Department, and dean of R&D at Sarvajani College of Engineering and Technology, Surat. His research interests include image processing, pattern recognition, deep learning, and medical signal processing. He has published more than 60 research publications in reputed conferences and journals as well as book chapters. He is the first recipient of the Regional Meritorious Service Award by the IEEE Signal Processing Society, USA, in 2017. Chirag has also served as chairman of SPS Chapter, Gujarat Section, and won the “Best Chapter Award” consecutively three times during his tenure. He was Chapter Chair Coordinator for IEEE, SPS, USA, for the year 2019. Currently, he is volunteering as a vice-chair for IEEE Gujarat Section. Chirag is a reviewer for many reputed journals by IEEE, Elsevier, and Springer. He has served as technical program chair for signal and image processing track for reputed conferences like INDICON, TENSYP, and TENCON.

Mita Paunwala received her BE (Electronics) from Sarvajani College of Engineering and Technology, Surat, in 1999; MTech (Communication System) from Sardar Vallabhbhai National Institute of Technology, Surat, in 2008; and PhD degree from NIT, Surat, India, in 2014. She is an associate professor in the Electronics and Communication Engineering Department, CKPCET, Surat, India. She has teaching and research experience of over 20 years. Her area of interest is image, video and signal processing, pattern recognition, machine learning, deep learning, and healthcare systems. Mita has published more than 25 research papers in various renowned conferences, journals, and books. She was vice chair of IEEE Signal Processing Society, Gujarat chapter, for the period 2019–2021. Mita has reviewed many papers for renowned journals by IET, Springer, Elsevier, and IEEE Access, among others.

Rahul Kher received his BE (Electronics) from Sardar Patel University in 1997; MTech (Electrical Engineering) from the Indian Institute of Technology, Roorkee, in 2006; and PhD (Electronics and Communication Engineering) from Sardar Patel University in 2014. He has teaching and research experience of over 21 years. His research interests include biomedical signal and image processing, medical

image analysis, and healthcare monitoring systems. Rahul has published 4 books and more than 70 research papers in various international journals and conferences. He is a senior member of IEEE and was the founder secretary of Signal Processing Society (SPS) Chapter of IEEE Gujarat Section during 2013–2015. He has been on the reviewer panel/ TPC member of many international journals and conferences including the *IEEE Communication Society Magazine*; *Journal of Biomedical Signal Processing and Control* (Elsevier); *Journal of Computer Networks* (Elsevier); *International Journal of Advanced Intelligence Paradigms* (Inderscience); *Biomedical Engineering: Application, Basis and Communication* (World Scientific); 1st Global IoT Summit (GIoTS 2017), Geneva, Switzerland; 3rd Global IoT Innovation Forum, Barcelona, Spain; 3rd Annual Int. Conf. on Wireless Comm. and Sensor Networks (WCSN 2016); 2016 IEEE World Forum on Internet of Things (WF-IoT), Virginia, USA, and many more. He has visited Japan, the USA, and the UK for various academic purposes.

Falgun Thakkar obtained his PhD from the National Institute of Technology Allahabad in February 2018. He graduated from Birla Vishvakarma Mahavidyalaya (BVM) in 2004 and completed his Master of Engineering in Communication from GCET, S P University, V V Nagar, in 2010. Dr. Falgun has published more than 25 research articles in various international and national journals and conferences. He has served as reviewer of many international journals and conferences of repute. Falgun has published a book in the domain of compressed sensing–based ECG signal compression. His areas of interest include antenna design, hf transmission line, microwave engineering, wavelet-based image and signal processing, medical image security, compressive sensing, and optimization techniques like PSO and GA. He has guided more than five ME students in their dissertation as well as more than 10 projects of BE students. At present, he is guiding six PhD students in the domain of microwave antenna design and medical image processing.

Heena Kher received her BE in instrumentation and control from Sarvajani College of Engineering and Technology, Surat, in 2001; ME in microprocessors systems and applications from M. S. University, Baroda, in 2006; and PhD from Sardar Patel University, Vallabh Vidyanagar, in 2014. She is Assistant Professor of A. D. at Patel Institute of Technology, New Vallabh Vidyanagar. She has 18 years of teaching experience. Her areas of interest are digital image processing, machine learning, deep learning, biomedical signal processing, and optimization techniques. She has guided ten dissertations of ME students and is DPC member of 3 PhD students. Heena has presented/published 30 research papers in various international/national conferences and international/national journals. She is a reviewer of many reputed journals.

Mohammed Atiquzzaman's research interests and publications are in next-generation computer networks, wireless and mobile networks, satellite networks, switching and routing, optical communications, and multimedia over networks. Many of the current research activities are supported by the National Science Foundation (NSF), National Aeronautics and Space Administration (NASA), and

the U.S. Air Force. He has served as the editor-in-chief of the *Journal of Network and Computer Applications* and the Vehicular Communications journal, and as associate editor of *IEEE Communications Magazine*, *Journal of Wireless and Optical Communications*, *International Journal of Communication Systems*, *International Journal of Sensor Networks*, *International Journal of Communication Networks and Distributed Systems*, and *Journal of Real-Time Image Processing*.

Norliza Mohd. Noor is currently attached as an associate professor in UTM Razak School of Engineering and Advanced Technology, University Technology Malaysia (UTM), Kuala Lumpur Campus. She received her BSc in electrical engineering from Texas Tech University in Lubbock, Texas, and her Master of Electrical Engineering (by research) and PhD (Electrical Engineering) from UTM. Her research area is in image processing and image analysis. Her current work concentrates on medical image analysis for lung diseases. She has published many papers in journals and in indexed conference proceedings, and has published one academic book and two book chapters. Currently, she is the head of the Electrophysiology Research Group., UTM Razak School.

Voice Privacy in Biometrics



Priyanka Gupta, Shrishti Singh, Gauri P. Prajapati, and Hemant A. Patil

1 Introduction

The definition of privacy was given first time by Warren and Brandeis in 1890 [1], as “the right to be left alone”. However, apart from the individuals, some sensitive information can also need protection in such a way that only a certain set of people are allowed (authorized) to access it. This authorization to access a specific information is given by a biometric system. Biometric systems are used for security purposes in a way that they prevent unauthorized access to an important information or data (*information privacy*). The access granted by the biometric is done by capturing traits of humans, which make all human beings unique w.r.t. that particular trait. This means that no two traits are the same. For example, fingerprints and iris are the most common physical traits that are captured by a biometric system. However, forging of such traits is quite prevalent, which poses a great security threat to the biometric system. We can say a biometric system that acts like a guard to protect sensitive and confidential information itself suffers from forgeries. Therefore, the need for privacy preservation in biometrics is all the more important [2].

Apart from the physical traits, such as fingerprints, hand geometry, and iris, there are behavioural traits, such as the speech, keystroke, and gait of an individual. One can recognize a person just by listening to his/her voice. Therefore, speech carries a lot more information than what it just sounds to be. This means, apart from the linguistic content of the speech, there are traits of the speaker, such as accent,

P. Gupta · S. Singh · G. P. Prajapati (✉) · H. A. Patil
Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT),
Gandhinagar, India
e-mail: priyanka_gupta@daiict.ac.in; shrishti_singh@daiict.ac.in; gauri_prajapati@daiict.ac.in;
hemant_patil@daiict.ac.in

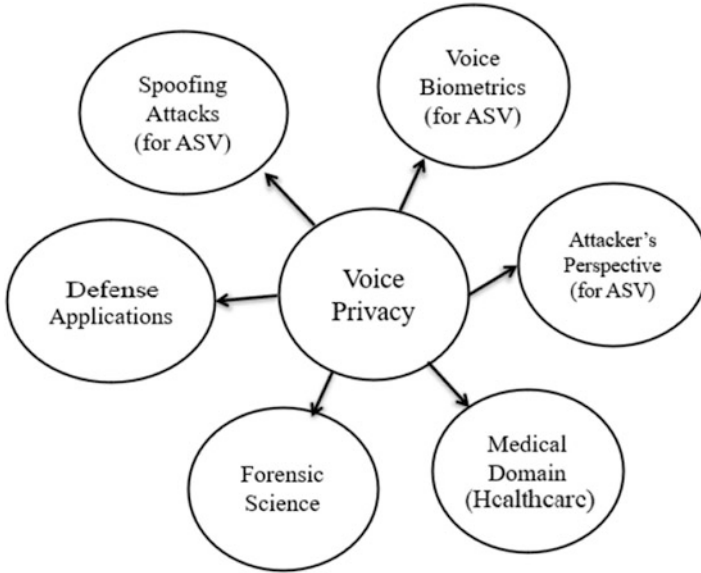


Fig. 1 Applications of voice privacy

pitch (fundamental frequency, F_0), tone, rhythm, idiosyncrasies, etc. Therefore, speaking of an individual's identity, it is not just his/her name, but it is also other the traits captured by the speech signal, such as gender, age, health status, personality, emotional state, and accent. So far as the practical deployment of Automatic Speaker Verification (ASV) technology is concerned, for example, banking transactions, access to restricted buildings, designing a privacy-preserving voice privacy system is crucial. A voice privacy system can be used for real-world applications, such as in forensics, voice biometric systems, medical domain, and to study and analyse attacker's perspectives to build more secure ASV systems as shown in Fig. 1.

1.1 Motivation for Voice Privacy

The notion of privacy in the field of healthcare is very old. With the advancement in technology comes the easy data collection and processing technologies [3]. At the same time, the detail and diversity of information collected in the context of biomedical research are increasing at an unprecedented rate. The easy availability of such large amount of data has also raise the concerns of privacy invasions [4]. It is important to understand the scope and frequency of these invasions. There are cases where medical records of people are illegally accessed for the purpose of identity fraud. Due to privacy concerns, people change the behavioural activities, such as visiting another doctor for check-up, not seeking care when needed in order to avoid

disclosure of information, self-treating or medicating themselves, not disclosing full information about their medical history, paying out of pocket despite of being insured, and hesitate to participate in the surveys that require data from people in the fear of data getting misused, etc. This privacy-protective behaviour shows the trust issues of people. Therefore, focus on the privacy preservation technologies should be given utmost importance to reduce the vulnerability of the data. It becomes all the more important in the case of patients suffering from speech disorders, and diseases such as dysarthria, which affect the speech characteristics. In such cases, the medical practitioner may have to record and save the patient’s speech data (with patients’ consent). However, the risk of availability of patients’ unprotected speech data will exist. Moreover, this risk will turn severely damaging if the patient is enrolled as a genuine speaker on a voice biometric (ASV) system. This risk can be mitigated to a large extent if voice privacy measures are applied to the speech data.

Furthermore, apart from the risk of unprotected speech data, ASV systems are prone to attacks. With the availability of patient’s unprotected speech data, the spoofing attacks become easier to mount and can even be specifically targeted at a particular patient and thus causing more damage. In the ASV literature, there are various spoofing attacks, such as Voice Conversion (VC) [5, 6], Speech Synthesis (SS) [7, 8], replay [9–11], twins [12], and impersonation [13] (Fig. 2).

Corresponding to each spoofing attack, there are Spoofed Speech Detection (SSD) systems. Recently, there have been efforts to develop countermeasures for replay attack detection on voice assistants or Intelligent Personal Assistants (IPA) [14, 15]. Furthermore, SSD systems based on the state-of-the-art Constant Q Cepstral Coefficients (CQCC), and Cochlear Filter Cepstral Coefficients and Instantaneous Frequency (CFCCIF) features were able to detect attacks by Speech Synthesis and Voice Conversion [16, 17]. However, the same features performed poorly in the case of a replay attack [18]. This means that we are far away from designing a generalized for these attacks on ASV and that there is no *versatile* SSD system to prevent all kinds of spoofing attacks. Since SSD systems have been known to prevent only a specific type of attack, their deployment in the real-world applications is not suitable. This is because the attacker has the liberty to perform any kind of attack on the ASV system, irrespective of whether the SSD system

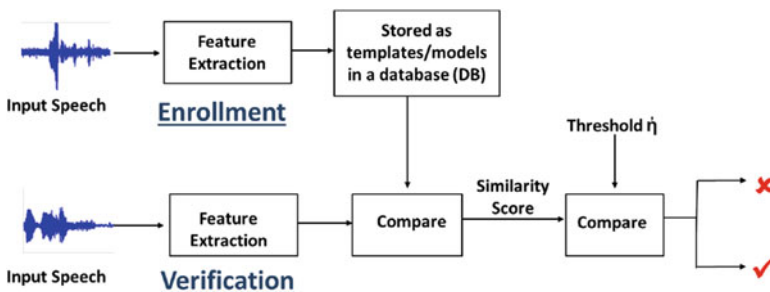


Fig. 2 A conventional voice biometric (ASV) system

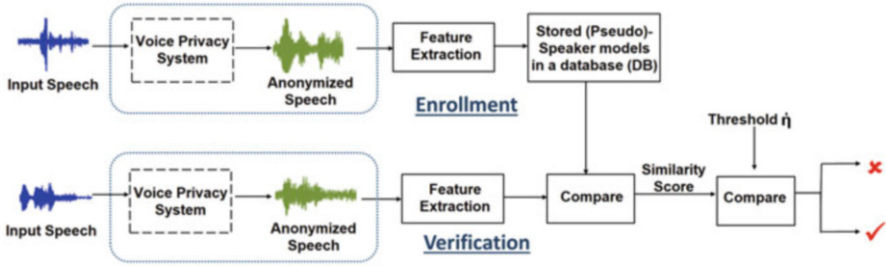


Fig. 3 Significance of proposed voice privacy for ASV system

is prepared for it or not. Therefore, voice privacy brings in the *universality* in protection from all the (at least most of the) types of attacks. Hence, the authors believe that voice privacy system could alleviate in principle, the need of SSD systems as shown in Fig. 3.

1.2 De-identification vs. Anonymization

The terms de-identification and anonymization are used interchangeably since both of them aim to protect or hide a person's identity. However, there is a subtle difference between them. In particular, de-identification is bi-directional, whereas anonymization is not. This means that de-identification procedures are reversible in nature, and hence, the original identity can be recovered from the pseudo-identity. This usually requires the knowledge of some extra (additional) information, such as a *key*. On the other hand, anonymization is irreversible. This means that the identity transformation is an irreversible function. Another notable difference is that in case of de-identification, if re-/de-identification (i.e., identification) usually requires a key. Therefore, de-identification methods are *generally* based on cryptographic methods.

Moreover, cryptographic methods fail to retain the intelligibility and the naturalness of the speech, whereas anonymization *generally* uses Voice Transformation (VT) techniques, which retain the quality of speech to a certain extent [6]. The VT approaches usually include anonymization by Voice Conversion, Speech Synthesis, and the other techniques of speech processing. One such speech processing technique is using linear prediction of speech, which is discussed in detail in this chapter.

2 Voice Privacy and Attacker's Perspective

In previous years, research has been centred around the issue of security of ASV systems. This has been done by developing respective countermeasures (CMs), for each kind of spoofing attack. Speech detection system (SSD) is an anti-spoofing system that detects the presence of an attack and allows only genuine speech into the ASV system. SSD was used preliminary to the ASV system, thus making it a two-class problem, as shown in Fig. 4. The first initiative w.r.t. the defence against spoofing attacks was in the form of a challenge organized by INTERSPEECH in 2015 [19]. Subsequent challenges were organized in 2017 and 2019, as shown in Table 1. It is worth noting that in ASVspoof 2015 and 2017, the assessment of CMs was done using equal error rate (EER), independent of ASV systems [19, 22–29]. However, for real-world applications, this type of assessment is not very useful.

2.1 Target Selection by Attacker and Voice Privacy System

In [30], speakers are classified as *lambs*, *wolves*, and *goats* on the basis of their effect on the equal error rate (EER). Since each of the type of speakers (i.e., lambs, wolves, and goats) has a different effect on the EER, they can also be classified in terms of their vulnerability levels to attacks, as shown in Fig. 5. From an attacker's perspective, the *lamb* type speakers should be the target. The attacker can perform target selection using an ASV system, to choose the most vulnerable speaker from the pool of speakers, as the target [31]. However, if voice privacy is used, the target

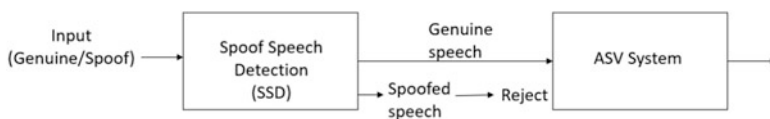


Fig. 4 Spoofed-speech detection (SSD) for ASV system

Table 1 ASVspoof challenge campaigns, after [19, 21]

INTERSPEECH ASVspoof 2015	INTERSPEECH ASVspoof 2017	INTERSPEECH ASVspoof 2019
Countermeasures were proposed using various kinds of feature extraction techniques	Countermeasures for real-replay attacks were proposed	Countermeasures for simulated replay attacks were proposed [20]
Signal processing techniques and Gaussian mixture model (GMM) were used to classify a speech as genuine or spoof	Paradigm shift from signal processing to deep-learning-based algorithms	Real physical access (PA) dataset was released. Tandem-DCF was used as the objective metric for joint evaluation of SSD and ASV

Fig. 5 Types of speakers on the basis of their vulnerability levels and their effect on EER scores

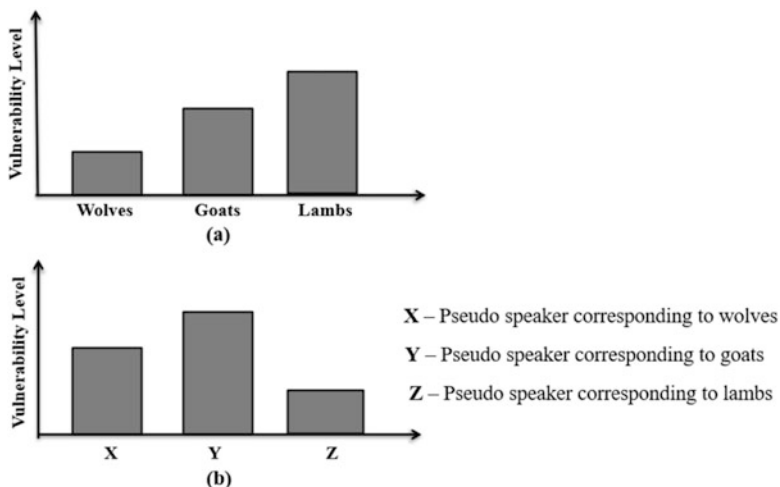
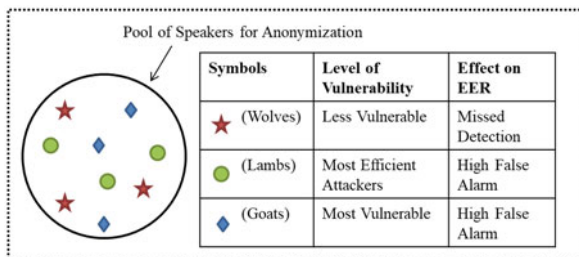


Fig. 6 Schematic representing effect of voice privacy on target selection. (a) Without voice privacy, target selection is successful. (b) With voice privacy, target selection gives misleading results

selection procedure by the attacker will yield incorrect results. Thus, the attacker will be fooled into selecting a not-so-vulnerable speaker as the target, as shown in Fig. 6.

2.2 Enrolled Users with Malicious Intent

In principle, an enrolled user has more power to attack than an attacker (usually a non-enrolled outside entity). An enrolled user with a malicious intent may attempt to spoof the system, which is all the more, a greater security concern/threat. A real-world example of this type of attack is the twin fraud in HSBC bank, where the bank's voice authentication system was spoofed by a BBC journalist, and his non-identical co-twin speaker [32, 33].

Another interesting point to note here is that if an SSD system as a countermeasure for twins attack is used, it will prevent malicious twins from impersonating (which is based on physiological characteristics, in particular, size and shape of the vocal tract system [34]). However, it will also prevent genuine and zero effort imposters from verification and hence increasing the False Rejection Rate (FRR). With the deployment of a voice privacy system instead, this kind of attack will not be possible. Moreover, the issue of preventing genuine and zero effort imposters will also be alleviated, and hence, there will be no increase in FRR.

3 Voice Privacy Using Linear Prediction Model

3.1 Speech Production Model

Depending on the signal shape and structure in time domain, speech signal can be divided into *voiced* and *unvoiced* speech. Voiced sounds are produced due to quasi-periodic vibrations of the vocal folds. These vibrations occur because of the sudden closing of the vocal folds (causing quasi-periodic vibration). In particular, when the air rushes from the lungs, it hits the vocal folds making them vibrate, because of the decrease in air pressure and tension on the vocal folds (i.e., by invoking Bernoulli's principle from fluid dynamics). One can actually touch and feel the vibrations of the vocal folds by placing a thumb near the throat while uttering a voiced sound such as a vowel (e.g., /a/). However, in the case of unvoiced speech, such as /h/, one does not feel any vibrations of the vocal folds (also called as *aspiration*). This is because, for unvoiced sounds, the vocal folds are just lightly open, and therefore, the air rushing from the lungs produces turbulence at the vocal folds. This turbulence is modeled as a noisy signal as shown in Fig. 7, which shows discrete-time speech production model for voiced and unvoiced sounds. For voiced sound, the gain is A_v , which corresponds to the loudness. Similarly, A_N corresponds to the loudness of the unvoiced sound. Considering the voiced case, the overall transfer function of the speech production model is $H(z) = G(z)V(z)R(z)$, where $G(z)$ is the transfer function of the glottal system, $V(z)$ is the transfer function of the vocal tract system, and $R(z)$ is the lip radiation effect. $G(z)$, $V(z)$, and $R(z)$ represent z-domain system functions. Mathematically, $G(z)$ is given by [35]

$$G(z) = \frac{1}{(1 - e^{-cT}z^{-1})^2}, \quad (1)$$

where c and T denote the velocity of sound and the time period of glottal pulse, respectively. Furthermore, the vocal tract system $V(z)$ and lip radiation $R(z)$ are given by [35]

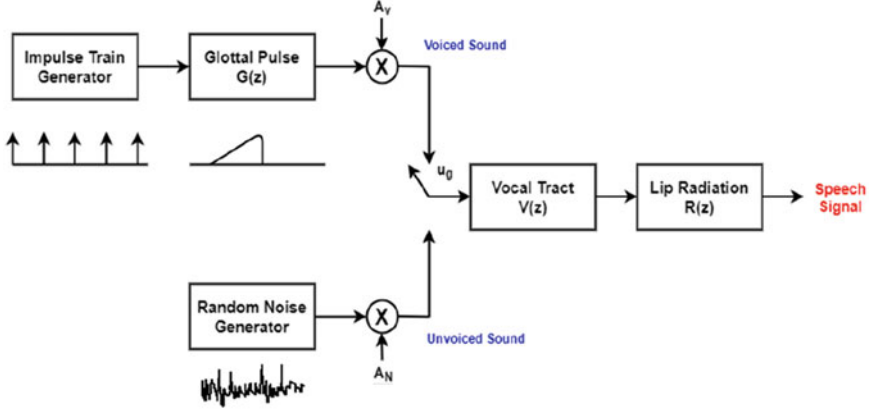


Fig. 7 Discrete-time speech production model, after [35]

$$V(z) = \frac{G}{\prod_{k=1}^{N/2} (1 - 2r_k \cos \theta_k z^{-1} + r_k^2 z^{-2})}, \quad (2)$$

$$R(z) = R_o(1 - z^{-1}), \quad (3)$$

where G is the gain of $V(z)$; r_k and θ_k are the pole radius and pole angle, respectively, of k th complex pole pair. If $e^{-cT} \approx 1$, then $H(z)$ will be

$$H(z) = \frac{\sigma}{1 - \sum_{k=1}^p a_k z^{-k}}, \quad (4)$$

where σ is the gain of $H(z)$. Vocal tract system, $V(z)$, is modeled as a linear time-invariant (LTI) all-pole system by cascading the 2nd-order digital resonators corresponding to each formant (Fig. 8). As per L. G. Kersta, who reported one of the first studies in speaker recognition, resonance is defined as *reinforcement* of spectral energy at or around a particular frequency [36]. The resonance frequencies of the vocal tract system are called formant frequencies. The formant frequencies specify the shape of the vocal tract system, thus forming the spectrum. The peaks in the spectrum are referred to as formant peaks. The formants change with different sizes and shapes of vocal tract configurations [37]. Therefore, the vocal tract system by cascading the four 2nd-order digital resonators (for first four formants) is given by

$$V(z) = \prod_{i=1}^4 H_i(z), \quad (5)$$

where each $H_i(z)$ is a 2nd-order digital resonator. Transfer function for 2nd-order digital resonator for i th formant is given by

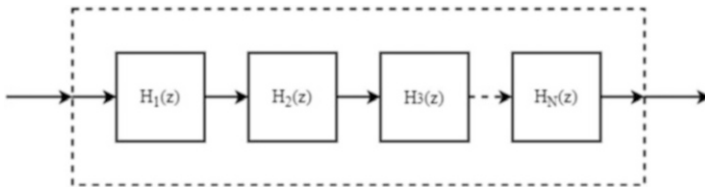


Fig. 8 Vocal tract system, $V(z)$, modeled by cascading 2nd-order digital resonators

$$H_i(z) = \frac{1}{(1 - p_{1i}z^{-1})(1 - p_{2i}z^{-1})}, \quad (6)$$

where p_1 and p_2 are the complex conjugate pole pairs of 2nd-order resonator transfer function. For i th formant, i.e., $p_{1i} = p_{2i}^* = r_i e^{\pm j\omega_{o_i}}$. Taking discrete-time Fourier transform (DTFT) of $H_i(z)$ frequency response of i th formant is given by

$$H_i(z)|_{z=e^{j\omega}} = H_i(e^{j\omega}) = \frac{1}{(1 - r_i e^{j\omega_{o_i}} e^{-j\omega})(1 - r_i e^{-j\omega_{o_i}} e^{-j\omega})}, \quad (7)$$

where ω_o is the pole angle and r_i is the pole radius. Now, taking magnitude of $H_i(e^{j\omega})$, we get

$$|H_i(e^{j\omega})| = \frac{1}{|(1 - r_i e^{j\omega_{o_i}} e^{-j\omega})||1 - r_i e^{-j\omega_{o_i}} e^{-j\omega}|}. \quad (8)$$

For resonance, $|H_i(e^{j\omega})| \rightarrow \max$, therefore,

$$\frac{d|H_i(e^{j\omega})|}{d\omega} = 0, \quad (9)$$

solving Eq (9) will give resonant frequency, ω_{r_i} ,

$$\omega_{r_i} = \cos^{-1} \left[\frac{1 + r_i^2}{2r_i} \cos \omega_{o_i} \right]. \quad (10)$$

Considering pole radius, $r_i \rightarrow 1$, then we get

$$\omega_{r_i} \approx \omega_{o_i}. \quad (11)$$

Impulse response of 2nd-order digital resonator is given by

$$h_i[n] = Kr_i^n \sin \omega_{o_i} (n + 1)u[n], \quad (12)$$

where r_i is the radius of poles and K is the overall gain. The pole radius is *inversely* proportional to the -3 dB bandwidth. When radius = 1 (i.e., bandwidth = 0), sharp peaks in the spectrum are observed with highest possible ($\sim\infty$) quality (Q)-factor. The change in pole radius will correspond to various energy losses, which is discussed in Sect. 3.2. Using physics of speech production, due to various sources of energy losses, dissipation of energy occurs in the system that causes the decrease in resonant frequencies and leads to the broadening of the -3 dB formant bandwidths. Thus, the effect of the damping factor r_i^n is observed. Using impulse-invariant transformation (IIT) to map Laplace domain (s-domain) pole to z-domain pole, relationship between -3 dB bandwidth and pole radius r is given by [38],

$$r_i = e^{-\pi B_i T}, \quad (13)$$

where B is the -3 dB bandwidth (in Hz) and T is the sampling interval (in seconds). Therefore, for larger radius, sharp high peaks will be observed at the resonance frequencies. Hence, to achieve speaker anonymization, radius of the pole should be decreased, which will eventually lead to the broadening of the bandwidth around the resonant frequency, causing the energy to get spread. Thus, there will be no presence of sharp and distinct peaks around formant frequencies, which will make identification of the speaker difficult.

3.2 Energy Losses

Ideally, the oral cavity is assumed to be a uniform tube with no losses due to the fact that the poles of corresponding transfer function Eq. 17 (which is the ratio of DTFT of volume velocities at lips and glottis, respectively) are strictly on $j\omega$ axis in s-plane. This oral cavity has roughly constant cross-section area with one end connected to the glottis and another at the lips [35]. However, in reality, this oral cavity can be modeled by time-varying and non-uniform cross-sectional area. Due to these variations, various energy losses occur. These losses affect the formant frequencies and their -3 dB formant bandwidths, which can be analysed from frequency response via suitable numerical simulations [39]:

- **Viscosity and Thermal Loss:**

The air particles' effects in flowing from glottis to the lips have some friction with vocal tract walls that resist the air flow from the glottis. This friction can be introduced as a resistor in an electrical equivalent circuit of the cavity. This friction represents viscous energy loss. Another loss in the form of heat loss (also called as thermal loss) is incurred due to the vibrations of the vocal tract walls. A small decrease in formant frequencies and increase in formant bandwidth can be observed while considering these losses along with the wall vibration loss. The increase in bandwidth is more at higher frequencies [40].

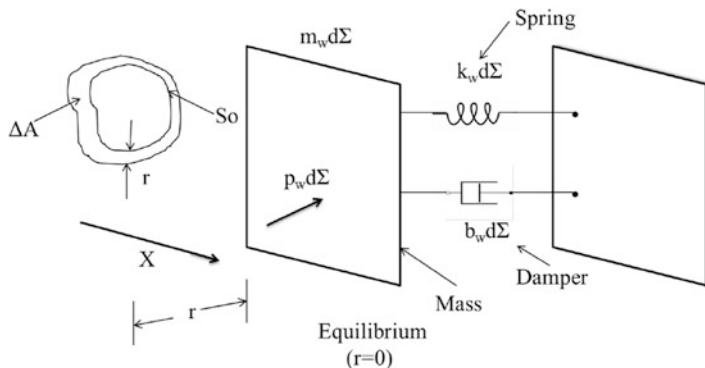


Fig. 9 Mechanical model of differential surface element $d\Sigma$ of vibrating wall, after [35, 40]

- **Wall Vibrations:**

Consider a tube whose cross-section is non-uniform. Furthermore, assume that the cross-sectional area changes slowly with time and space. The small differential sections of surface of the wall ($d\Sigma$) are assumed to be independent by Portnoff [40]. Each of these small sections can be then mechanically modeled as shown in Fig. 9, where m_ω = mass, k_ω = spring constant, and b_ω = damping constant per unit surface area.

Considering the two boundary conditions of the volume velocity sources $u(0, t)$ (known), and the output pressure $p(l, t)$ (where l = length of the vocal tract modeled as uniform tube), three coupling equations—two for sound wave propagation and one for 2nd-order differential equation from Fig. 9—can be approximated as [35].

$$-\frac{\partial p}{\partial x} = \frac{\rho}{A_0} \frac{\partial u}{\partial t}, \quad (14)$$

$$-\frac{\partial u}{\partial x} = \frac{A_0}{\rho c^2} \frac{\partial p}{\partial t} + \frac{\partial \Delta A}{\partial t}, \quad (15)$$

$$p = m_\omega \frac{d^2 \Delta A}{dt^2} + b_\omega \frac{d \Delta A}{dt} + k_w \Delta A, \quad (16)$$

where A_0 = average cross-section (constant), ΔA = linear perturbation about the average cross-section, $S_0(x, t)$ is the average vocal tract perimeter at equilibrium, r is the perpendicular displacement of the wall, and ρ = density of air particles. For the steady-state condition of the system described above, assume the system to be an LTI system. An input $u_g(t) = u(0, t) = U(\Omega)e^{j\Omega t}$

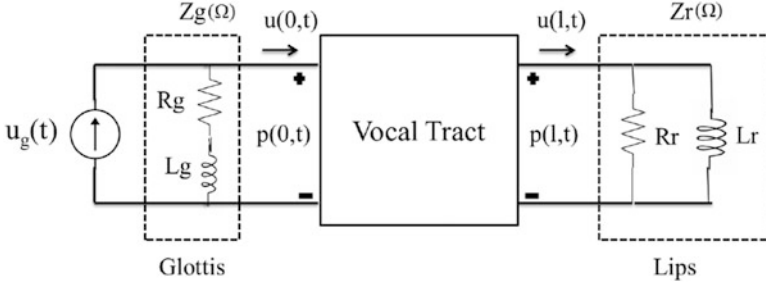


Fig. 10 Glottal and lip boundary conditions as impedance loads, after [35]

gives solutions, $p(x, t) = P(x, \Omega)e^{j\Omega t}$, $u(x, t) = U(\Omega)e^{j\Omega t}$, and $\Delta A(x, t) = \Delta \hat{A}(x, \Omega)e^{j\Omega t}$. Portnoff has used standard numerical simulation techniques to solve these coupled equations, which results in frequency response as shown in Eq. 17 [40].

$$V_a(\Omega) = \frac{U(l, \Omega)}{U_g(\Omega)}. \quad (17)$$

While producing voiced speech, due to air pressure from the lungs, glottis will vibrate (by invoking Bernoulli's principle in fluid dynamics). Since vocal tract walls are pliant, they will move under pressure induced by sound propagation in the vocal tract system. These vibrations lead to energy losses in the cavity, and hence, the poles of Eq. 17 are moved from $j\Omega$, becoming complex from only imaginary (ideally). Hence, the -3 dB bandwidth is non-zero, and formant frequency is increased. At low frequencies, inertial mass of vocal tract walls results in more motion, making it more dominant at lower frequencies compared to the higher frequencies.

- **Lip Radiation Loss:**

The effect of radiation at lips can be analysed by finding the acoustic impedance seen by the vocal tract from the lip end. This leads to the consideration of glottal and radiation load (at the lips) in the cavity model, as shown in Fig. 10. R_r is the radiation resistance due to sound propagation through lips, and L_r is the radiation inductance that is the inertial mass sent out at lips. Parallel combination of them contributes to the acoustic impedance (Eq.18) [35].

$$Z_r(\Omega) = \frac{P(l, \Omega)}{U(l, \Omega)} = \frac{1}{\frac{1}{R_r} + \frac{1}{j\Omega L_r}} = \frac{j\Omega L_r R_r}{R_r + j\Omega L_r}. \quad (18)$$

For very small $\Omega \approx 0$, $Z_r \approx 0$, so the radiation load acts as a short circuit with pressure at the lips equal to zero, i.e., $p(l,t)=0$. For very large Ω with condition $\Omega L_r \gg R_r$, $Z_r \approx R_r$, making it resistive at higher frequencies. The radiation energy loss is happened due to real part of the complex impedance Z_r ,

Table 2 Frequency response of uniform tube with various losses with $p(l, 0) = 0$, after [35, 40]

Formants	Vibrating walls		Vibrating walls, viscus, and thermal loss		Vibrating walls, viscus, thermal, and radiation loss	
	Frequency (Hz)	Bandwidth (Hz)	Frequency (Hz)	Bandwidth (Hz)	Frequency (Hz)	Bandwidth (Hz)
1st	504.6	53.3	502.5	59.3	473.5	62.3
2nd	1512.3	40.8	1508.9	51.1	1423.6	80.5
3rd	2515.7	28.0	2511.2	41.1	2372.3	114.5
4th	3518.8	19.0	3513.5	34.5	3322.1	158.7

which is proportional to R_r from Eq. 18. Thus, more radiation loss will occur at higher frequencies with monotonic increase in R_r . From this discussion, it can be observed that this radiation impedance behaves as a high-pass filter (HPF). Hence, to approximate the lip radiation, we can model the impedance as a HPF before we apply any algorithm on a speech signal.

Considering all the three losses together shows a slight decrease in formant frequencies, however a very high increase in -3 dB bandwidth as shown in Table 2, particularly for higher frequencies [40]. Here, comparison is made to lossless system's formant values (i.e., odd multiples of 500 Hz) for a particular case when tube length is 17.5 cm with a cross-sectional area of 5 cm².

The most important thing to note here is that every human being has different configurations of vocal tract system. In addition, the size and shape of lips during speaking vary differently for everyone. These facts connect lip radiation loss to speaker-specific characteristics of a speech signal. As the speaker-specific characteristics lie in the higher formants (i.e., F_3 and F_4), the energy losses become more important when we deal with the de-identification. In this chapter, authors tried to validate this conclusion using various experiments that changes -3 dB bandwidth to change the speaker's identity.

3.3 Linear Prediction (LP) Model

LP is one of the most powerful methods to analyse speech signals especially in speech coding for wireless communication services. LP coefficients for speech implicitly represent time-varying vocal tract area function. It is an iterative method to find current sample of speech $s[n]$, using past p speech samples because linear prediction coefficients capture implicitly time-varying area function of vocal tract during speech production, where p represents predictor memory [41]. With respect to source-filter model of speech production, LP method decomposes speech signal into two components: LP coefficients (representing vocal tract system using LP filter) and LP residuals (representing speech excitation source) [38]. By minimizing the squared differences between the actual speech samples and the linear predicted

speech samples, a unique set of predictor coefficients can be obtained. LP model is based on all-pole model. Generally, the all-pole model is preferred because it is computationally more efficient and of its acoustic tube model for speech production. It can model sounds, such as vowels well enough and the other consonants (except nasal consonants that require zeros in the transfer function). The zeros arise only in the nasals and in the unvoiced sounds.

In LP analysis, sample at n th instant is represented as a linear combination of past p samples, i.e.,

$$\tilde{s}[n] = a_1s[n-1] + a_2s[n-2] + \dots + a_p s[n-p], \quad (19)$$

where a_1, a_2, \dots, a_p are called as LP coefficients.

The z-domain system function for p th-order predictor is given as

$$P(z) = \sum_{k=1}^p \alpha_k z^{-k}. \quad (20)$$

The error signal or the LP residual signal, $e(n)$, is the difference between the actual (true) speech signal and the estimated speech signal. LP residual is given by

$$e[n] = s[n] - \tilde{s}[n] = s[n] - \sum_{k=1}^p \alpha_k s[n-k]. \quad (21)$$

In z-domain, error signal or LP residual $e(n)$ can be seen as the output of the prediction error filter $A(z)$ to the input speech signal $s(n)$ and is given by

$$E(z) = A(z)S(z), \quad (22)$$

where prediction error filter $A(z)$ is defined as

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k} = 1 - P(z). \quad (23)$$

The whole LP model can be viewed in two parts, the analysis and the synthesis. The LP analysis filter suppresses the formant structure of the speech signal and leaves a lower energy output prediction error that is often called the LP residual. LP residuals are used as an excitation source for the production of speech. The synthesis part takes the error signal as an input that gets filtered by the *inverse* filter, which is the inverse of the prediction error filter, and gives the speech signal as the output. When the vocal tract system is modeled as an LTI all-pole system, then a pole at $r_{o_i} e^{j\omega_{o_i}}$ and $r_{o_i} e^{-j\omega_{o_i}}$ corresponds to i th formant of vocal tract system. Vocal tract length has *inverse* relationship with the formant frequencies. Thus, we can observe the difference in the formant frequencies between the male and the female speaker

[42]. In particular, a male speaker (due to larger vocal tract length) tends to have lower formants than female speakers [35].

In LP model, LP coefficients govern the pole locations that in turn determine formant frequency and formant bandwidth [43]. Mathematically, formant frequency is given by $\frac{F_s \theta}{2\pi}$, where θ is the angle of the pole in radians, given F_s is the sampling frequency in Hz. The formant bandwidth is given by $\frac{F_s}{\pi}(-\log(r))$, where r is the radius of the pole [35]. As per M.R. Schroeder, human beings emit and perceive sounds by emitting spectral peaks more dominantly than the spectral valleys [44]. Therefore, we can achieve speaker de-identification by modifying the formant frequencies leading to the change in the formant spectrum with *naturalness* and *intelligibility* retained in the anonymized speech. Hence, by performing controlled shift in the pole angle and the pole radius, speaker de-identification can be achieved without the loss of intelligibility in the anonymized speech signal.

3.4 Experimental Setup

3.4.1 Baseline System

Recently, efforts are made to develop privacy preservation solutions for speech technology. In the light of moving forward towards this development, the first voice privacy challenge is being organized in INTERSPEECH 2020 to motivate researchers in this direction [45, 46]. In the baseline system, provided by the organizers of the Voice Privacy Challenge 2020, anonymization is achieved by only modifying the pole angle. It is based on employing McAdam’s coefficient [47] to the pole angle, which are extracted using linear prediction (LP) method [48]. The performance of the anonymization is evaluated using these parameters, namely, equal error rate (EER), calibration cost (Cllr), and word error rate (WER). Former two objective metrics are based on ASV systems whose higher values are desired. Whereas, the intelligibility score is determined through WER, which should be as low as possible. However, for better anonymization of the speaker’s identity, improved version of the baseline system has also been discussed in sub-section “Proposed Voice Privacy System”, in which not only pole angle but radius of the poles is also modified in order to incorporate effect of various energy losses during natural speech production.

3.4.2 Corpora Used

For development data, subsets from two corpora, namely, LibriSpeech-dev-clean and VCTK, are provided [49, 50]. These subsets are further divided into trial and enrolment subsets. There are 40 speakers in LibriSpeech-dev-clean. There are 29 speakers in enrolment utterances and 40 speakers in trial utterances. From these 40 speakers of trial subset, 29 speakers are also included in the enrolment subset.

Table 3 Statistics of the development datasets, after [45]

Subsets of corpus	Particulars	Female	Male	Total
LibriSpeech: Dev-clean	Speakers in enrolment	15	14	29
	Speakers in trials	20	20	40
	Enrolment utterances	167	176	343
	Trial utterances	1018	960	1978
VCTK-dev	Speakers (same in enrolment and trials)	15	15	30
	Enrolment utterances	300	300	600
	Trial utterances (common part)	344	351	695
	Trial utterances (different part)	5422	5255	10677

Table 4 Statistics of the evaluation datasets, after [45]

Subsets of corpus	Particulars	Female	Male	Total
LibriSpeech: test-clean	Speakers in enrolment	16	13	29
	Speakers in trials	20	20	40
	Enrolment utterances	254	184	438
	Trial utterances	734	762	1496
VCTK-test	Speakers (same in enrolment and trials)	15	15	30
	Enrolment utterances	300	300	600
	Trial utterances (common part)	346	354	700
	Trial utterances (different part)	5328	5420	10748

In VCTK-dev dataset, there are a total of 30 speakers that are the same for both trial and enrolment utterances. Furthermore, for trial utterances, there are two parts, denoted as *common part* and *different part*. Both the parts are disjoint in terms of utterances; however, they have the same set of speakers. The *common part* of the trials has utterances from #1 to #24 in the VCTK corpus, which are the same for all the speakers. The *common part* of the trials is meant for subjective evaluation of speaker verifiability/linkability in a text-dependent manner. #25 onward utterances are distinct and hence are included in the *different part* of the VCTK-dev dataset. For evaluation, the structure is the same as that of development set, except for the number of utterances (Tables 3 and 4).

3.4.3 Proposed Voice Privacy System

At first, speech signal is divided into smaller frames that are fed to LP source-filter analysis in order to obtain to LP coefficients and residual.

Only LP coefficients are taken into account for further processing, while the residual is left unchanged. LP coefficients are then employed to obtain pole positions of the LP model. Poles whose imaginary value is not zero are considered, and their pole angle " ϕ " is calculated. Since every complex conjugate pole pair corresponds to one formant frequency, only one pole out of complex conjugate pole pair is

Algorithm 1 Voice privacy by LP modeling of speech production

Ensure: Speech signal is divided into frames.

LP coefficients and residuals are extracted.

LP coefficients are converted to poles.

Radius of the complex poles is shifted to 0.975 of the original value of radius.

Poles ϕ of the complex poles are shifted to $\phi^{0.8}$.

New LP coefficients are formed.

The new anonymized speech signal is re-synthesized.

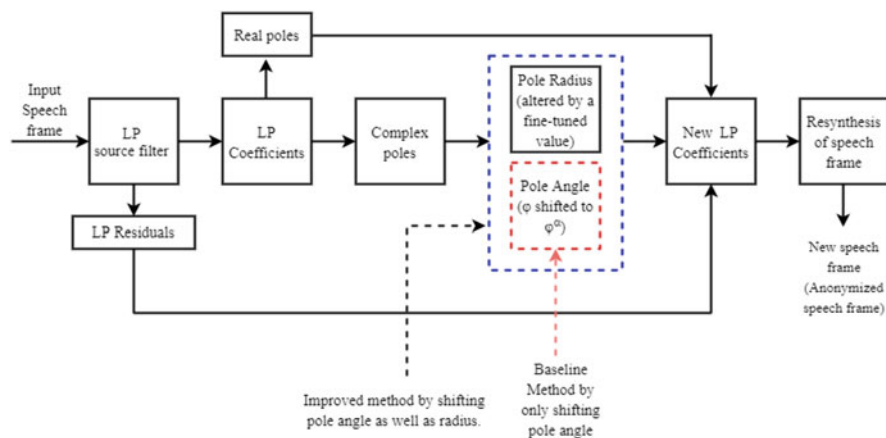


Fig. 11 Proposed LP-based anonymization system, after [14, 45, 46]

considered for achieving speaker anonymization [51]. For further improving the baseline system, pole radius is changed along with the pole angles. The pole angle is shifted by raising it to the power of McAdams coefficient [47] “ $\alpha = 0.8$ ”, i.e., ϕ^α . Values of α and ϕ determine the positive or negative shift in the pole locations. The pole radius is decreased by 15%, 5%, and 2.5% of original pole radius [14]. With these new sets of pole radii and angles, a new set of poles are fabricated therefore forming new LPC coefficients. These new coefficients along with original LP residuals are used to synthesize new speech signal, hence achieving the anonymization of speech. Motivated by original studies in the speech coding literature [52–54], residuals are kept intact because they are used to retain the naturalness and intelligibility of the speech signal (Fig. 11).

3.4.4 Experimental Results

In the experiments, decreasing the radius of the poles results in the expansion of the formant bandwidth. On studying the experimental results, it is observed that 2.5% decrease in the radius along with the phase changed to $\phi^{\alpha=0.8}$ gave higher values of %EER and lower values of %WER, which is desired. As discussed

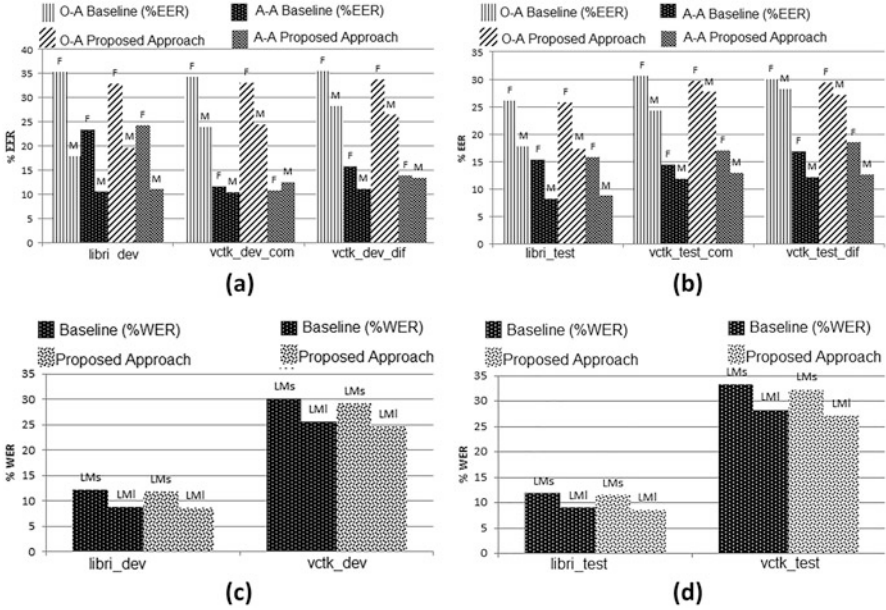


Fig. 12 %EER (o—original, a—anonimized) for (a) development data, (b) test data, and %WER (for two trigram LMs: LM_s —small LM, and LM_l —large LM) for (c) development data, (d) test data (for radius = 0.975 to its value and $\alpha = 0.8$), after [14]

earlier, by decreasing the pole radius, the corresponding formant bandwidth will increase. According to the resonator (theory discussed in Sect. 3.1), increase in the bandwidth will decrease the Q-factor of the resonator. Spectrum peaks will no longer be distinctly present causing the loss of speaker-specific information. Hence, the quality of original speech signal degrades, which contributes to the higher EER scores. The results of the experiment in terms of %EER and %WER for test data and development data are shown in Fig. 12a–d [14, 55].

3.4.5 Gender-Based Analysis

From the experimental results obtained for voice privacy, noticeable information came out to be the *higher* values of the %EER for the female speakers than the male speakers under the condition that the anonymization technique on the utterances is the same for both the female and male speakers as shown in Fig. 12a, b. This result can be supported by the fact that spectral resolution for female speech is poor as compared to the male speech [56]. The mass of the vocal folds in female speakers is less than the male speakers due to which movement of vocal folds becomes sluggish in male speakers, and hence, the glottal vibrations are more rapid (fast) in female speakers, and therefore, high pitch frequency is observed for female voice. Hence,

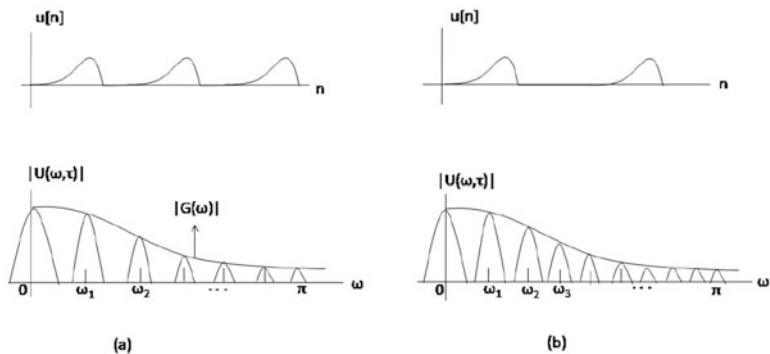


Fig. 13 Illustration of periodic glottal flow and its spectrum. (a) Higher pitch (female speaker), (b) lower pitch (male speaker), after [35, 56]

in spectral domain, the pitch-source harmonics are observed to be in larger distance with each other, which results in the poor spectral resolution of the female speaker. This can be beneficial for our aim to achieve speaker de-identification because the perception of the female speech through ASV systems can become difficult. In addition, in the glottal cycle waveform, glottal closure instant and period during closure provide characteristics for discriminating speaker's voice from one another. Provided the same pitch period and impulse response of the vocal tract system, even a slight variation in the glottal waveform can result in a considerable amount of change in the voice characteristics. Therefore, due to the larger pitch duration in male speakers, they get sufficient time for the closure of the glottis and to perform activity near the glottal closure. However, in the case of female speakers, the pitch duration is almost half the pitch duration of the male speakers (near about $10ms$ in male speakers and $5ms$ in female speakers), due to which female speakers do not have enough time for the closure of the glottis and to perform activity near the glottis before the glottal opening. This large variation in the glottal waveform changes the speaker's characteristics drastically. The speaker recognition techniques use information based on the $1 - 2ms$ glottal closure period. Hence, tracking this large variation in $1 - 2ms$ of glottal closure period becomes difficult for the ASV systems, which can lead to the higher $\%EER$ values. The illustration of the spectral resolution problem is presented in Fig. 13. In particular, $u[n]$ represents the glottal flow waveform model that can be given by

$$u[n] = g[n] * p[n], \quad (24)$$

where $g[n]$ is the glottal flow waveform over a single glottal cycle and $p[n]$ is an impulse train with spacing, P [35]. $U(\omega)$ and $G(\omega)$ are the Fourier transforms of $u[n]$ and $g[n]$. ω_k is the harmonics of the glottal flow waveform. The magnitude of the spectral shaping function, $G(\omega)$, is referred as the spectral envelope of the harmonics.

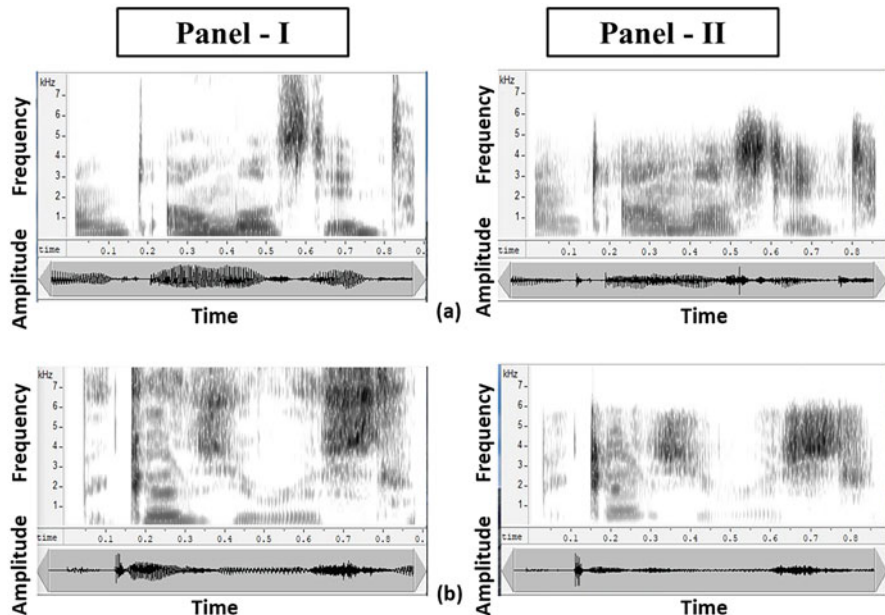


Fig. 14 Panel I: Analysis for original speech signal. Panel II: Analysis for anonymized speech signal. (a) Spectrogram and speech signal for a female speaker, (b) spectrogram and speech signal for a male speaker

Moreover, the analysis of spectrogram of original and anonymized speech of both the female and male speakers is done. The original speech of both the male and female speakers has undergone same anonymization method that was discussed in sub-section “Proposed Voice Privacy System”. According to the change in the pole angle ϕ , corresponding formants will be shifted, i.e., for $\phi < 1$, the formants will be shifted to a higher value and vice versa. Due to this reason, lower formants in male speech will shift to a higher value, and high pitch-source harmonics for female speech will be observed. In addition, more uniform energy distribution in the spectrum is observed in the male speech as compared to the female speech (Fig. 14).

4 Technological Challenges in Voice Privacy

We have seen how a voice privacy system can alleviate so many issues, and hence, it can have applications in a large number of domains. However, it should be noted that there is no algorithm in the world that can extract a speaker’s identity *explicitly*, because in a speech signal, the speaker’s identity is embedded *implicitly*, and thus, this will be the case for speaker de-identification and voice privacy system as well. Therefore, there are some technical challenges that if proven to be solvable in the

near future, then the voice privacy system can actually prove to be a real boon for the security of ASV systems. In this context, some of the technical challenges are discussed in this section.

4.1 Evaluation Metrics for Speech Quality

Evaluation metrics for assessment of speech quality are usually categorized as objective and subjective metrics. Objective metrics result from assessment via machines, while subjective metrics result from assessment via listening tests done by different listeners. Though objective metrics, such as EER and WER, can give us information about the extent of anonymization and intelligibility of the speech, respectively, these measures are insufficient to measure the naturalness and related factors. Thus, subjective evaluations are done by performing manual listening tests. However, subjective tests suffer from inaccuracies because of various following factors:

- **Cognitive State of the Listeners:** Cognitive factors of listeners, such as their attention span, mood, and environmental noise. The output's accuracy (for example, Mean Opinion Score (MOS) and Perceptual Evaluation of Speech Quality (PESQ)) is also affected because the listeners have not understood the test properly. Hence, they evaluate the speech wrongly. In addition, if naturalness is not be evaluated, the effect of intelligibility also creeps in due to the way humans perceive sound. During perception, naturalness and intelligibility cannot be *distinctly* separated from each other.
- **Correlation Between Objective and Subjective Results:** Correlation between objective and subjective measures is done to check the effectiveness of the used methodology for voice privacy. Pearson's Correlation Coefficient (PCC) is used to find this correlation that is given by

$$PCC = \frac{\sum_i^n (X_i - \mu_x)(Y_i - \mu_y)}{\sqrt{\sum_i^n (X_i - \mu_x)^2} \sqrt{\sum_i^n (Y_i - \mu_y)^2}}, \quad (25)$$

where X_i are the subjective scores for i th system, μ_x is the mean of all the subjective score vectors, Y_i are the features values for i th system, and μ_y is the mean of all objective scores. It has been observed that PCC between MOS and PESQ is negative [57].

4.2 Machines vs. Human Perception of Speech

Signals that are perceived as natural and intelligible by the human ear might not be accepted by the machines (i.e., Automatic Speech Recognition (ASR) system).

In particular, the factors that are acoustically significant may not be perceptually detected by our perceptual system. For example, due to non-linear source-filter interaction of glottal airflow with formants of the vocal tract system, a sinusoid-like ripple (fine structure) is observed during opening phase of the derivative of glottal flow waveform. However, perceptual test cannot detect its significance as it is acoustically significant for speaker identification.

4.3 Robustness vs. Vulnerability

With regards to the usage of a voice privacy system in an ASV, one should also consider certain feature characteristics that are listed below:

1. Stability over time
2. Robustness under noisy environments
3. Robustness over emotional and health status
4. Robustness over mic distance variability

However, if most of the characteristic requirements are met, it would also pave a way for the attacker to attempt a successful attack. For example, if robustness under MIC variability and noisy environment is considered, then the chances of a replay attack are high. Thus, our good intention of designing robust ASV system makes it more vulnerable for various spoofing attacks. In particular, distinguishing between natural vs. replay spoof becomes more difficult due to these robust features.

5 Voice Privacy and Cryptography

Cryptography aims to prevent any malicious usage of data. Two primary types of cryptographic algorithms are symmetric key (private-key) encryption and asymmetric key (public-key) encryption. Symmetric key encryption itself requires security for protection of the key. Moreover, the total number of keys required for p parties should be $p(p - 1)/2$. Hence, due to these key-management issues, public-key encryption has taken over most of the security applications. In the following subsections, we will discuss the most widely used RSA algorithm and its time complexity.

5.1 Public-Key Encryption

Public-key encryption uses two kinds of keys—public and private. Public keys are accessible to everyone including the attacker. However, private keys are known only to a single user. Each user has his own private key, which is not to be shared with

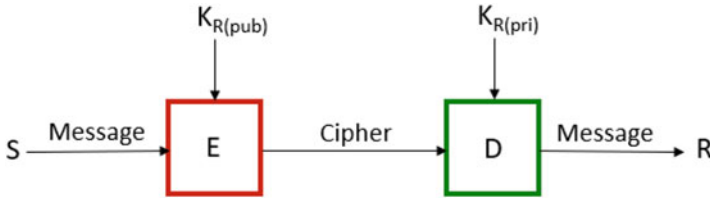


Fig. 15 Public-key encryption and decryption

Algorithm 2 Key Generation

Require: Two prime numbers, p and q , of roughly equal lengths are generated randomly.

$$n = p * q$$

Ensure: Euler’s totient function, $\Phi(n) = (p - 1) * (q - 1)$ Choose an integer e such that it satisfies the following two conditions:

$$1 < e < \Phi(n)$$

$$GCD(e, \Phi(n)) = 1$$

Calculate d such that it is the multiplicative inverse of e , i.e., $d \equiv e^{-1} \pmod{\Phi(n)}$. This means $e.d \equiv 1 \pmod{\Phi(n)}$.

return (e, n) and (d, n) (public key and private key (of the receiver, respectively)).

Algorithm 3 Encryption

Require: The message to be encrypted is represented as an integer m such that $m > 0$ and m lies the interval $(0, n - 1]$.

Ensure: The sender has receiver’s public key (e, n) .

Sender computes cipher $c = m^e \pmod{n}$.

return The cipher c is then sent to the receiver for decryption.

Algorithm 4 Decryption

Ensure: The receiver has received the cipher from the sender.

The cipher will be decrypted as $m = c^d \pmod{n}$, by the receiver.

return Decrypted message m .

anyone. As shown in Fig. 15, the message is encrypted by the sender S , with the help of receiver’s public key, $K_R(pub)$. Examples of public-key algorithms are RSA, Diffie–Hellman, and El-Gamal encryption [58].

Though key management is not a major issue with such types of algorithms, however, these algorithms are slower than the symmetric key encryption algorithms and are mathematically more complex and intensive. As an example, the famous Rivest–Shamir–Adleman (RSA) algorithm is shown below as [59–61]

Time Complexity

The complexity of the RSA algorithm is majorly contributed by three operations, which are exponentiation, inversion, and modular operation. Modular operations, such as modular addition operations, exist, whose complexity is of the order of $O(\log n)$, where n is the size of the input.

Algorithm 5 Modular multiplication using square and multiply technique

Require: Inputs are 2 numbers A and B , in k -bit binary representation.

Initialise output $P = 0$

for $i = 0$ to $k - 1$ **do**

$P = 2P + A.B_{k-1-i}$

$P = P \bmod n$

end for

return P

Modular multiplication is done using squaring and multiply technique as shown in the following algorithm: To get $m^e \bmod n$, modular multiplication is used. Considering the complexity of multiplication $O(\log n^2)$, i.e., repeated addition of two numbers of $\log n$ bits each, the complexity of the modular exponentiation is about $O(\log n^3)$.

Using Euclidean extended GCD from (extended Euclidean algorithm), inverse of a number can be calculated in $O(\log n^2)$ [62]. Thus, for N -digit number space, the overall time complexity of key generation will be $O(N^2)$, and the overall time complexity of encryption and decryption will be of the order of $O(N^3)$.

These modular operations are used repeatedly and intensively for the other cryptographic approaches also, such as homomorphic encryption (HE) [63]. The size of the key used should be 2048-bits, and therefore, the inputs to the modular operations are also nearing the same order, which makes the overall computational overhead high.

5.2 Limitations of Cryptographic Approaches for Voice Privacy

Though cryptographic approaches are meant to be used for security purposes, there are practical issues with their implementations in already complex systems, such as ASVs. The limitations are discussed in this sub-section:

- The security of cryptography lies under the concept of *computational* difficulty of solving the discrete logarithmic problem. However, the same reason is responsible for the limitation of cryptographic techniques in deployment to real-world applications. Therefore, cryptography is costly in terms of both time and money:
 - Addition of cryptographic techniques in the information processing leads to delay.
 - The set-up and maintenance of cryptographic implementations, such as public-key infrastructure and HE, require a large computing power, varying overhead of communications and rounds of interactions, and, hence, a big monetary budget.

- Most cryptographic techniques use modulo arithmetic operations on *integers*. However, given the nature of a speech signal, representation of signals and computations on them requires modulo arithmetic on *floating-point* operations [2].
- Variable speech signal quality should also be reflected in the encrypted output. However, this requires computations of matrix inversions and log determinants, which are expensive computations.
- Vulnerabilities and threats can come up because of the poor (hardware) implementation of systems, protocols, and procedures. A poor hardware implementation can open the way to many hardware-based attacks, such as side-channel attack [64].
- Cryptographic implementations can become vulnerable to attacks if they are not maintained and updated regularly. Since the security lies in the computational difficulty, regular breakthroughs that solve those computationally difficult problems keep coming up [58]. Hence, the current implementations should regularly update their *difficulty* levels.
- With the advent of quantum cryptography, the existing system whose security is based on the computational difficulty of solving a mathematical problem will completely collapse. Therefore, *post-quantum* solutions in cryptography are desirable, but they too come with cost in terms of both time and money.

6 Summary and Conclusions

Recent trends in speech technology are making biomedical systems much more effective than they were previously. However, they also have some challenges of privacy of patient's medical data that should be handled carefully. Hence, in this chapter, we have discussed the importance of a voice privacy system that can be used to protect patient's voice data. Voice privacy is secure against attacking approaches, such as target selection and enrolled users with malicious intent. LP modeling of speech and the fact that a speaker's identity is embedded in the energy losses while speech is produced are used to design a robust voice privacy system. Moreover, a few technological challenges related to the designing of a voice privacy system are also listed. Issues arising with the application of cryptographic techniques are also discussed. In the future, extracting speaker-specific features from the glottal flow waveform and the associated ripple in it can be done. Moreover, the other LP models, such as Residual Excited Linear Prediction (RELP) and Mixed Excited Linear Prediction (MELP), can be used. Apart from signal processing techniques to design a voice privacy system, deep-neural network techniques can also be used, in the future. The speech community is developing privacy systems that will ensure voice data privacy along with data usability.

References

1. WARREN, S.D. and BRANDEIS, L.D. (1890) The Right to Privacy. *Harvard Law Review* : 193–220.
2. NAUTSCH, A., JIMÉNEZ, A., TREIBER, A., KOLBERG, J., JASSERAND, C., KINDT, E., DELGADO, H. *et al.* (2019) Preserving Privacy in Speaker and Speech Characterisation. *Computer Speech & Language* **58**: 441–480.
3. MALIN, B.A., EMAM, K.E. and O'KEEFE, C.M. (2013), Biomedical data privacy: problems, perspectives, and recent advances.
4. BOYER, B.B. (1975) Computerized medical records and the right to privacy: the emerging federal response. *BuFF. L. REv.* **25**: 37.
5. STYLIANOU, Y., CAPPÉ, O. and MOULINES, E. (1998) Continuous probabilistic transform for voice conversion. *IEEE Transactions on Speech and Audio Processing* **6**(2): 131–142.
6. STYLIANOU, Y. (2009) Voice transformation: A survey. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (Taipei, Taiwan): 3585–3588.
7. ZEN, H., TOKUDA, K. and BLACK, A.W. (2009) Statistical parametric speech synthesis. *Speech Communication* **51**(11): 1039–1064.
8. DE LEON, P.L., PUCHER, M., YAMAGISHI, J., HERNAEZ, I. and SARATXAGA, I. (October, 2012) Evaluation of speaker verification security and detection of HMM-based synthetic speech. *IEEE Transactions on Audio, Speech, and Language Processing* **20**(8): 2280–2290.
9. ALEGRE, F., JANICKI, A. and EVANS, N. (2014) Re-assessing the threat of replay spoofing attacks against automatic speaker verification. In *International Conference of the Biometrics Special Interest Group (BIOSIG)* (Darmstadt, Germany): 1–6.
10. PAUL, A., DAS, R.K., SINHA, R. and PRASANNA, S.M. (2016) Countermeasure to handle replay attacks in practical speaker verification systems. In *2016 International Conference on Signal Processing and Communications (SPCOM)* (IISc, Bengaluru, India): 1–5.
11. PRAJAPATI, G.P., KAMBLE, M.R. and PATIL, H.A. (18-21 January, 2020) Energy separation based features for replay spoof detection for voice assistant. *28th European Signal Processing Conference (EUSIPCO)* : pp. 386–390.
12. WU, Z., EVANS, N., KINNUNEN, T., YAMAGISHI, J., ALEGRE, F. and LI, H. (2015) Spoofing and countermeasures for speaker verification: A survey. *Speech Communication* **66**: 130–153.
13. LAU, Y.W., WAGNER, M. and TRAN, D. (2004) Vulnerability of speaker verification to voice mimicking. In *International Symposium on Intelligent Multimedia, Video, and Speech Processing* (Hong Kong): 145–148.
14. GUPTA, P., PRAJAPATI, G.P., SINGH, S., KAMBLE, M.R. and PATIL, H.A. (7-10 December, 2020) Design of voice privacy system using linear prediction. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* (Auckland, New Zealand: IEEE): 543–549.
15. GONG, Y., YANG, J. and POELLABAUER, C. (2020) Detecting Replay Attacks Using Multi-Channel Audio: A Neural Network-Based Method. *IEEE Signal Processing Letters*.
16. PATEL, T.B. and PATIL, H.A. (2016) Cochlear Filter and Instantaneous Frequency based Features for Spoofed Speech Detection. *IEEE Journal of Selected Topics in Signal Processing* **11**(4): 618–631.
17. PATEL, T.B. and PATIL, H.A. (6-10 September, 2015) Combining Evidences from Mel Cepstral, Cochlear Filter Cepstral and Instantaneous Frequency Features for Detection of Natural vs. Spoofed Speech. In *INTERSPEECH* (Dresden, Germany).
18. KAMBLE, M.R., PULIKONDA, A.K.S., KRISHNA, M.V.S. and PATIL, H.A. (1-5 November, 2020) Analysis of Teager Energy Profiles for Spoof Speech Detection. In *Odyssey The Speaker and Language Recognition Workshop, Tokyo, Japan*.

19. ZHIZHENG, W., KINNUNEN, T., EVANS, N., YAMAGISHI, J., HANILÇI, C., SAHIDULLAH, M. and SIZOV, A. (6-10 September, 2015) ASVspooF 2015: The First Automatic Speaker Verification Spoofing and Countermeasures Challenge. In *INTERSPEECH* (Dresden, Germany): 2037–2041.
20. TODISCO, M., WANG, X., VESTMAN, V., SAHIDULLAH, M., DELGADO, H., NAUTSCH, A., YAMAGISHI, J. *et al.* (2019) AsvspooF 2019: Future Horizons in Spoofed and Fake Audio Detection. *arXiv preprint arXiv:1904.05441* .
21. Automatic Speaker Verification-Spoofing and Countermeasures Challenge <https://www.asvspooF.org/>. [Last Accessed: 2021-03-15].
22. NOVOSELOV, S., KOZLOV, A., LAVRENTYEVA, G., SIMONCHIK, K. and SHCHEMELININ, V. (20-25 March, 2016) STC Anti-spoofing systems for the ASVspooF 2015 Challenge. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Shanghai, China: IEEE): 5475–5479.
23. WESTER, M., WU, Z. and YAMAGISHI, J. (6-10 September, 2015) Human vs Machine Spoofing Detection on Wideband and Narrowband Data. In *INTERSPEECH* (Dresden, Germany): 2047–2051.
24. WANG, L., YOSHIDA, Y., KAWAKAMI, Y. and NAKAGAWA, S. (6-10 September, 2015) Relative Phase Information for Detecting Human Speech and Spoofed Speech. In *INTERSPEECH* (Dresden, Germany): 2092–2096.
25. LIU, Y., TIAN, Y., HE, L., LIU, J. and JOHNSON, M.T. (6-10 September, 2015) Simultaneous Utilization of Spectral Magnitude and Phase Information to Extract Supervectors for Speaker Verification Anti-spoofing. In *INTERSPEECH* (Dresden, Germany): 2082–2086.
26. XIAO, X., TIAN, X., DU, S., XU, H., CHNG, E.S. and LI, H. (6-10 September, 2015) Spoofing Speech Detection using High-Dimensional Magnitude and Phase Features: The NTU Approach for ASVspooF 2015 Challenge. In *INTERSPEECH* (Dresden, Germany): 2052–2056.
27. FONT, R., ESPÍN, J.M. and CANO, M.J. (20-24 August, 2017) Experimental Analysis of Features for Replay Attack Detection-Results on the ASVspooF 2017 Challenge. In *INTERSPEECH* (Stockholm, Sweden): 7–11.
28. WITKOWSKI, M., KACPRZAK, S., ZELASKO, P., KOWALCZYK, K. and GALKA, J. (20-24 August, 2017) Audio Replay Attack Detection Using High-Frequency Features. In *INTERSPEECH* (Stockholm, Sweden): 27–31.
29. WANG, X., XIAO, Y. and ZHU, X. (20-24 August, 2017) Feature selection based on CQCCs for automatic speaker verification spoofing. In *INTERSPEECH* (Stockholm, Sweden): 32–36.
30. DODDINGTON, G., LIGGETT, W., MARTIN, A., PRZYBOCKI, M. and REYNOLDS, D. (1998) *Sheep, Goats, Lambs and Wolves: A Statistical Analysis of Speaker Performance*. Tech. rep., National Institute of Standards and Technology (NIST), Gaithersburg Md.
31. GUPTA, P. and PATIL, H.A. (2021, Brno, Czechia) A Survey of Attacker’s Perspective on Automatic Speaker Verification (ASV) Systems. *Submitted to INTERSPEECH 2021* .
32. (2017) HSBC reports high trust levels in biometric tech as twins spoof its voice id system. *Biometric Technology Today* **2017**(6): 12. <http://www.sciencedirect.com/science/article/pii/S0969476517301194>. [Last Accessed: 2021-03-15].
33. TEAM, E. (2017), Twins fool HSBC voice biometrics - BBC. <https://www.finextra.com/newsarticle/30594/twins-fool-hsbc-voice-biometrics--bbc>. [last accessed: 2021-03-15].
34. ROSENBERG, A.E. (1976) Automatic speaker verification: A review. *Proceedings of the IEEE* **64**(4): 475–487.
35. QUATIERI, T.F. (2004) *Discrete-Time Speech Signal Processing: Principles and Practice* (2nd Edition, Pearson Education India).
36. KERSTA, L.G. (1962) Voiceprint identification. *Nature* **196**(4861): 1253–1257.
37. FANT, G. (1970) *Acoustic Theory of Speech Production* (2nd Edition, Walter de Gruyter).
38. ATAL, B.S. and HANAUER, S.L. (1971) Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. *The Journal of the Acoustical Society of America (JASA)* **50**(2B): 637–655.
39. FLANAGAN, J.L. (2013) *Speech Analysis Synthesis and Perception*, **3** (Springer Science & Business Media).

40. PORTNOFF, M.R. (1973) *A Quasi-One-Dimensional Digital Simulation for the Time-Varying Vocal Tract*. Ph.D. thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, USA.
41. MARKEL, J.D. and GRAY, A.J. (2013) *Linear Prediction of Speech*, **12** (Springer Science & Business Media).
42. EIDE, E. and GISH, H. (1996) A Parametric Approach to Vocal Tract Length Normalization. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Atlanta, Georgia, USA: IEEE), **1**: 346–348.
43. MIZUNO, H. and ABE, M. (1996) A Formant Frequency Modification Algorithm Dealing with the Pole Interaction. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)* **79**(1): 46–55.
44. SCHROEDER, M.R. (May 1966) Vocoders: Analysis and Synthesis of Speech. *Proceedings of the IEEE* **54**(5): 720–734.
45. The Voice Privacy 2020 Challenge Evaluation Plan. <https://www.voiceprivacychallenge.org>.
46. TOMASHENKO, N., SRIVASTAVA, B.M.L., WANG, X., VINCENT, E., NAUTSCH, A., YAMAGISHI, J., EVANS, N. *et al.* (24–28 October, 2020) Introducing the voice privacy initiative. In *INTERSPEECH* (Shanghai, China). {Last Accessed: 2021-03-15}.
47. MCADAMS, S. (May, 1984) Spectral fusion, spectral parsing, and the formation of auditory image. *Ph.D. Thesis, Department of Hearing and Speech, Stanford University, California, USA*.
48. PATINO, J., TODISCO, M., NAUTSCH, A. and EVANS, N. (2020) *Speaker Anonymisation using the McAdam's Coefficient*. Tech. rep., EURECOM. <http://www.eurecom.fr/publication/6190> Last Accessed: 2021-03-15.
49. PANAYOTOV, V., CHEN, G., POVEY, D. and KHUDANPUR, S. (19–24 April, 2015) LibriSpeech: an ASR corpus based on public domain audio books. In *2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (Brisbane, Australia: IEEE): 5206–5210.
50. YAMAGISHI, J., VEAUX, C., MACDONALD, K. *et al.* (2019) CSTR VCTK Corpus: English Multi-Speaker Corpus for CSTR Voice Cloning Toolkit (Version 0.92).
51. SLIFKA, J. and ANDERSON, T.R. (1995) Speaker Modification with LPC Pole Analysis. In *1995 International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (Detroit, Michigan, USA: IEEE), **1**: 644–647.
52. UN, C. and MAGILL, D. (1975) The residual-excited linear prediction vocoder with transmission rate below 9.6 kbits/s. *IEEE Transactions on Communications* **23**(12): 1466–1474.
53. SCHROEDER, M. and ATAL, B. (1985) Code-excited linear prediction (CELP): High-quality speech at very low bit rates. In *ICASSP'85. IEEE International Conference on Acoustics, Speech, and Signal Processing (IEEE)*, **10**: 937–940.
54. MCCREE, A.V. and BARNWELL, T.P. (1995) A mixed excitation LPC vocoder model for low bit rate speech coding. *IEEE Transactions on Speech and Audio Processing* **3**(4): 242–250.
55. GUPTA, P., PRAJAPATI, G., SINGH, S., KAMBLE, M.R. and PATIL, H.A. (2020) System description : Design of voice privacy system using linear prediction <https://www.voiceprivacychallenge.org/docs/DA-IICT-Speech-Group.pdf>. {Last Accessed: 15-01-2021}.
56. PATIL, H.A., DUTTA, P. and BASU, T. (2006) On the Investigation of Spectral Resolution Problem for Identification of Female Speakers in Bengali. In *2006 IEEE International Conference on Industrial Technology (ICIT)* (Mumbai, India: IEEE): 375–380.
57. SAILOR, H.B. (2013) *Objective Evaluation of Speech Quality of Text-to-Speech (TTS) Synthesis Systems*. Master's thesis, DA-IICT, Gandhinagar, India.
58. STINSON, D.R. and PATERSON, M. (2018) *Cryptography: Theory and Practice* (CRC press).
59. STALLINGS, W. (2006) *Cryptography and Network Security: Principles and Practices* (Pearson Education India).
60. RIVEST, R.L., SHAMIR, A. and ADLEMAN, L. (1978) A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications of the ACM* **21**(2): 120–126.

61. BAI, X., JIANG, L., LIU, X. and TAN, J. (2014) RSA Encryption/Decryption Implementation Based on ZedBoard. In *International Conference on Trustworthy Computing and Services* (Springer): 114–121.
62. DIXON, J.D. (1970) The Number of Steps in the Euclidean Algorithm. *Journal of Number Theory* **2**(4): 414–422.
63. GENTRY, C. and BONEH, D. (2009) *A Fully Homomorphic Encryption Scheme*, **20** (Stanford University).
64. NARA, R., SATOH, K., YANAGISAWA, M., OHTSUKI, T. and TOGAWA, N. (2010) Scan-based Side-Channel Attack Against RSA Cryptosystems Using Scan Signatures. *IEICE transactions on Fundamentals of Electronics, Communications and Computer Sciences* **93**(12): 2481–2489.

Histopathology Whole Slide Image Analysis for Breast Cancer Detection



Pushap Deep Singh, Arnav Bhavsar, and K. K. Harinarayanan

1 Introduction

Cancer is a group of diseases, distinguished by uncontrollable cell growth, invasion and spread of cells from primary location, to other locations in the body. In India according to the fact sheet presented by GLOBOCAN in 2018 [1], total number of new mammary gland cancer cases were 162,468 and deaths due to this type of cancer were 87,090 and this disease is having highest number of cases in a year and most deaths according to the statistics shared in [1].

Adenocarcinoma is a type of cancer which occurs in glandular tissue (e.g., mammary gland). Mammary gland cancer is categorized based on the cell structure which is observed under a microscope. Major portion of mammary gland cancers are carcinomas, a cancer type that originates in the organ lining. Some of the important carcinoma types are listed below:

- Ductal carcinoma in situ (DCIS): DCIS is distinguished by degenerated cells that are restricted to the milk ducts. If DCIS is left unattended for a long time, the cancer cells may spread to nearby tissue. It is the most common type of noninvasive cancer. DCIS can be categorized by the appearance of the tumor, which can be solid, cribriform, micropapillary, papillary, and comedo.
- Invasive ductal carcinoma (IDC): IDC originates from the milk ducts and expands outside the duct. IDC accounts for 80 percent of the invasive mammary gland cancer.

P. D. Singh · A. Bhavsar (✉)
School of Computing and Electrical Engineering, IIT Mandi, Kamand, India
e-mail: arnav@iitmandi.ac.in

K. K. Harinarayanan
Aindra Pvt. Ltd., Bengaluru, India

- Lobular carcinoma: Lobular carcinoma originates from the lobes (glands responsible for making milk). The probability of this type of cancer to become invasive is very low (0.1).

There exists a variety of tests which can be done to screen and diagnose the mammary gland cancer. These include physical examination of the gland, mammography, ultrasound of the mammary gland, magnetic resonance imaging (MRI) of the gland, lymph node biopsy. Among all of these methods, lymph node biopsy is considered to be the gold standard to detect presence of cancer in the mammary gland.

1.1 Lymph Node Biopsy and the Need for Computer Aided Diagnosis

In humans lymphatic system is an important part of the immune and circulatory system. Lymphatic system is a network of the nodes, vessels, and a special organ known as spleen. Figure 1 shows the lymphatic system in the human body. The green colored dots denote the nodes, pink colored region depicts the spleen and remaining denote the vessels[2]. For detecting cancer one of the methods is the biopsy of the lymph nodes present near tumor. In detecting mammary gland cancer, samples of tissue are collected from the auxiliary lymph nodes.

Figure 2 shows the bean shape structured lymph node which holds cells responsible for the fighting with germs and eliminate harmful cells. The lymph nodes present near mammary gland, auxiliary lymph nodes, will have cancer cells if cancer is present in the mammary gland.

The collected tissue samples are carefully observed by a pathologist carefully under microscope at different resolutions for detecting abnormalities. This is a time-consuming process, requiring a lot of expertise, and is quite subjective, resulting in inter-observer variability. Thus, to aid the pathologists in terms of efficient training to reduce the workload, or to facilitate an objective assessment, computer aided diagnosis systems (CAD) are gaining importance as an important part of the digital pathology domain. Such CAD systems can involve machine learning based methods to assess the histopathology images of the slides containing the tissue samples.

1.2 Whole Slide Imaging, and Weakly Supervised Image Classification

One of the important challenges in machine learning based methods for classification of histopathology images is that the images corresponding to a whole slide of tissue sample are very large, often consisting of pixels numbering in the order of many millions or even billion. The sizes of such whole slide image (WSI)—

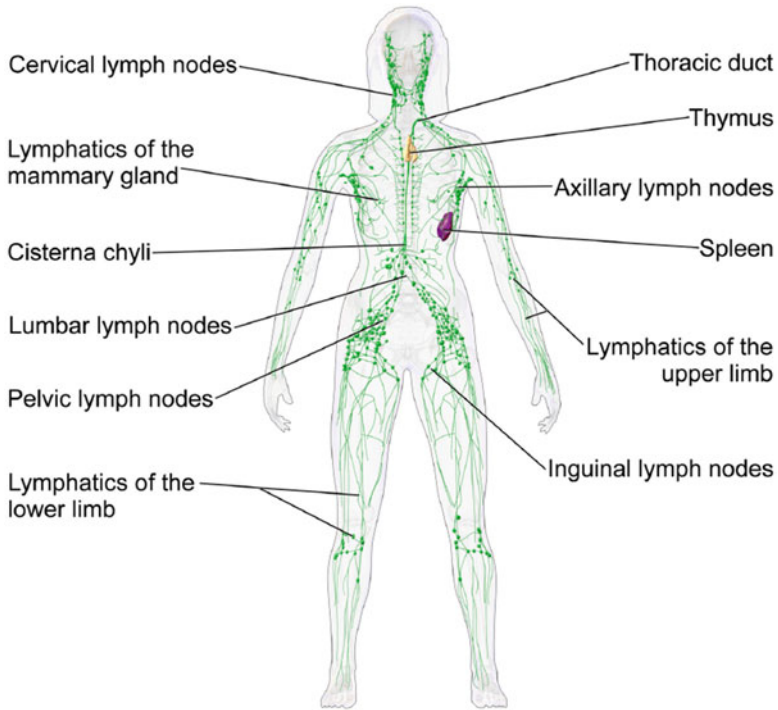


Fig. 1 Depiction of the lymphatic system in humans. Green color shows the lymphatic system in human body. Figure reproduced from [2]

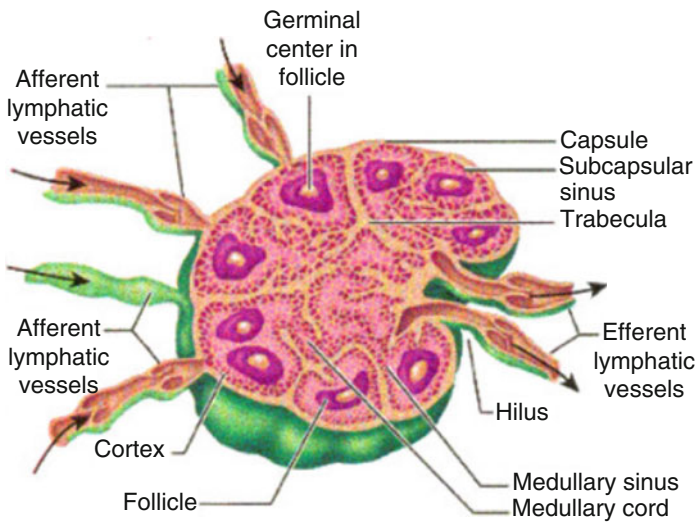


Fig. 2 Lymph node, a part of the lymphatic system. Figure reproduced from [2]

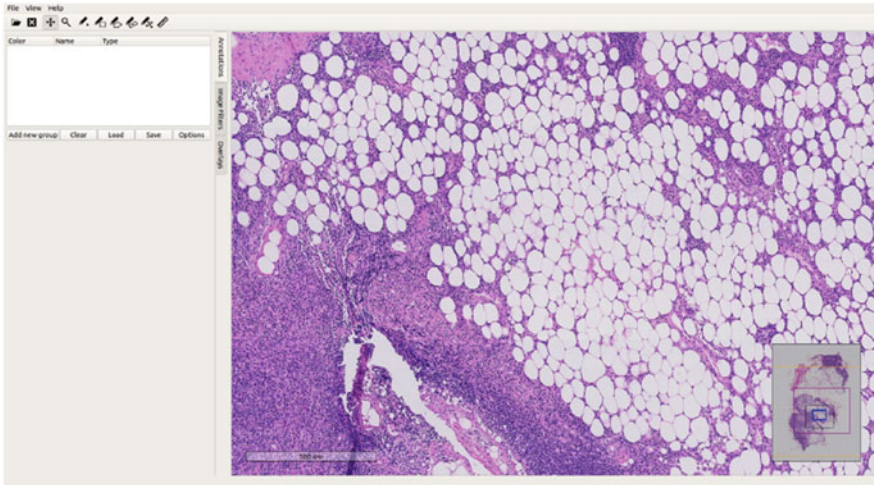


Fig. 3 Automated slide analysis platform (ASAP) by computational pathology group at RMUC

also known as gigapixel images—are typically far larger than what are typically encountered in standard computer vision based applications.

Pioneered by Renato et al. [3], whole slide imaging [4] is an approach to digitize the slides. In this method of digitizing the slides, authors of [3] used a combination of robot, microscope and a computer to generate composite image of the slide under observation. The composite image of the slide was a result of the mosaic pattern of the image tiles. After this initial work, WS imaging methods have significantly improved in efficiency and hardware, and the current WSI scanners which take less than a minute or a few minutes to generate a digitized version of the slide. This can in turn be viewed on computer with the aid of specialized software. Typical image viewers in the digital computer can only show the images which are small in size and can be easily decompressed into the available RAM of the system. These viewers cannot be used to view the whole slide images as they contain billion of pixels. Hence, there are special software which are used to view these images. One such software is associated with the Camelyon17 dataset and is termed as Automated Slide Analysis Platform ASAP [5]. Figure 3 shows an example of a WSI using ASAP [5].

To train machine learning classification algorithms, one often requires manual ground-truth labels about the abnormalities in such large images. However, for such large WSIs, it is tedious and time-consuming for a pathologist to demarcate and provide labels for the local abnormal regions in a WSI.

A much easier labeling process for the pathologist is to provide the overall label for the image, which indicates whether the WSI contains abnormalities or if it corresponds to a normal sample. This approach of labeling does not involve provide exact locations of the abnormalities, thus reducing the labeling load significantly.

Thus, as indicated in [6], image level labels and pixel level labels, and the underlying machine learning frameworks can be defined as follows:

- Image level labels indicate that the no annotation is provided for pixels. The region of interest specifying abnormalities is not given but an overall label is given to the image. Images labeled in this way are termed as weakly labeled or having a global label. The corresponding machine learning methods using such a data are also known as weakly supervised.
- Pixel level labels indicate that abnormal region is labeled by an expert. As the region of interest is known in the given image, all the information required at each pixel, train a machine learning methods is available. This type of learning is known as fully supervised learning.

Considering the difficulty of the weakly supervised scenario for the WSI classification task, it is quite natural that there are various fully supervised methods reported in literature, but few methods are available for address the weakly supervised classification task. However, noticing the practical importance of the weakly supervised classification, in this work we propose a deep learning based method for the same.

The rest of the chapter is organized as follows. In Sect. 2, we discuss some existing approaches for WSI classification and highlight our contribution. Section 3 consists of the description of the dataset that we use in this work. In the Sect. 4, we discuss our methodology in detail, and the results are provided and discussed in the Sect. 5. In the Sect. 6, we conclude the chapter.

2 Literature Survey

In this section we discuss some related work on WSI image classification which includes both weakly supervised and fully supervised approaches.

2.1 *Weakly Supervised Learning*

The proposed work is more closely related to weakly supervised methods. Hence, below we discuss two contemporary works on weakly supervised methods in relatively more detail.

2.1.1 Tellez et al. [7]

The work in paper [7] involves unsupervised representation learning algorithms, to express patches (or tiles) from the gigapixel WSI in a more compact manner, in an embedded feature space. These embeddings are spatially arranged so as to

maintain the same ordering of the corresponding tiles in the WSI image. This representation is then used for classification. The datasets used are CAMELYON16 [8] and TUPAC16 [9]. The embeddings for tiles are computed in different ways, as summarized below:

- **Reconstruction error minimization:** Here, an autoencoder(AE) is used to learn the representation of the data. AE utilizes an encoder-decoder architecture.
- **Contrastive training:** In this case, a model is trained to learn a feature space wherein the difference between similar and different images is increased. Such encodings are then used to represent the WSI image.
- **Adversarial feature learning:** In this method, a bidirectional generative adversarial network is used, which consists three sub-networks: a discriminator, a generator, and an encoder. The encoder maps actual image $x \in \mathbf{R}^{Z \times Z \times 3}$ to an embedding $e \in \mathbf{R}^C$. A generator maps $z \sim \mathcal{N}(0, 1)$ to generate an image $x' \in \mathbf{R}^{Z \times Z \times 3}$. A discriminator then tries to discriminate between generated and actual embedding-image pairs, i.e., $\{z, x'\}$ and $\{e, x\}$.

2.1.2 Courtiol et al. [10]

The authors in [10] have used a pretrained model to extract the representation of the WSI tiles at highest resolution. They have extracted tiles of the size $224 \times 224 \times 3$ from the dataset provided by the Camelyon16 which results into the total number of tiles 200,000, which is still quite a large number, leading to total of 200,000 embeddings.

First, the authors have used Otsu algorithm to remove the background from the whole slide image. Then color normalization of slide is done so as to reduce the effect of the variation in the H&E staining of the slides.

In the network, a minmax layer is used to get the top R negative and positive instances from the features extracted. These 2R values are then fed to multilayer perceptron classifier with 200 and 100 neurons with sigmoid activation. The method does not involve the spatial arrangement of the tiles, but a number of tiles are sampled using a sampling formula as is described in the paper [10]. This results in little compression for the gigapixel image as the volume remains the same. The authors have mentioned that Resnet-50 can be used for getting representation for the high resolution images as they have found it gives the best results among other pretrained models without need for fine-tuning.

2.1.3 Compression Analysis of the Above Weakly Supervised Methods

In this section we provide the compression analysis of the weakly supervised methods, which we believe is important, both these approaches (and also ours) involve the feature embeddings of the WSI to better represent the same.

Let a whole slide image or a gigapixel image c be represented as $c \in \mathbf{R}^{A \times B \times 3}$ where A is number of rows, B is the number of columns and 3 represents color channels (RGB). Let c' be the representation of the gigapixel image after passing through encoder, each high resolution patch of the gigapixel image, and placing them in spatial agreement with the patches (tiles) extracted from the gigapixel image. A two-step process to find the c' is described below:

1. From whole slide image c high resolution patches are extracted. Let $Z = z_{i,j}$ denotes the set of the high resolution patches extracted from c such that $z_{i,j} \in \mathbf{R}^{P \times P \times 3}$ is sampled from c 's i^{th} row and j^{th} column of a uniform square grid with patch size P and stride S .
2. After getting the set of patches at the highest resolution(40x) each patch is independently encoded using an encoder E generating a set of encoding vectors of a particular length(here it is C) which are placed in agreement with spatial location of the patches.

$$c \in \mathbf{R}^{A \times B \times 3} \xrightarrow{E} c' \in \mathbf{R}^{X \times Y \times C} \quad (1)$$

In the equation 1, $X = \frac{A}{S}$, $Y = \frac{B}{S}$ and C is a constant.

We now discuss below how compression occurs. Let $F = A \times B \times 3$ and let $F' = X \times Y \times C$. Dividing F by F' gives the following:

$$\frac{F}{F'} = 3 \frac{S^2}{C} \implies F = 3 \frac{S^2}{C} \times F' \quad (2)$$

Now, using the above, for weakly supervised method [7] for $S = C = 128$ then $F = 384 \times F'$, we compute that the volume reduces by a factor of 384. For weakly supervised method [10], using $S = 224$, $C=2048$ then $F = 73.5 \times F'$, the volume reduces by a factor of 73.5.

2.2 Fully Supervised Learning

While there are many fully supervised learning methods, in this part we discuss the three best performing algorithms in the Camelyon17 challenge; a challenge that was associated with the dataset that we are using in this work. For discussion related to Camelyon16 challenge, refer to [11].

In the paper [12] the authors have used modified Deeplab v3+ network which is trained using patches, where the patches are annotated. The patches for pixel level supervision are extracted using the annotations available from Camelyon16 and Camelyon17. At each epoch, the model generates inference for all patches of whole slide image and accept patch as a part of training whose intersection over union (IOU) with original mask is less than 0.95.

In [13] the authors have used Camelyon16 and Camelyon17 dataset which is annotated for training the network. This work has two steps, involving segmentation to get the segmentation map of the cancerous region and secondly, extraction of handcrafted features are extracted from the segmentation map for classification purpose. The authors have proposed an ensemble approach where many segmentation models learned on different pixel resolutions are combined in a directed acyclic graph structure. Since the segmentation models perform the pixelwise classification, this allows the concatenation of different models. The authors in [13] have used DeepLab models for different pixel resolution to train the pixel wise segmentation. The segmentation map then is used for prediction of the tumor using random forest.

In [14] authors have used a semi-supervised learning, utilizing whole slide images from 10 organs. The training of network is divided into 3 parts. In the first part, cells are independently annotated for fully supervised training. In second part, new regions are proposed for possible detection by the network and are indeed very much probable when shown to a pathologist and these new proposed regions are also added for training. In third part, cooperative training is employed. To mitigate problem of getting stuck at local minima while self-training, SRCDetectors with different backbones have been used and are trained on newly generated labels from other to minimize the problem of self-amplification of error.

In conclusion, while the fully supervised methods show a relatively high performance, they require detailed manual annotations, which are very tedious to generate, because of requirement of hours of analysis of whole slide image at different resolution. Thus, we focus on the weakly supervised methods where only image level labels are required, are in more preferable by pathologists in practice.

2.3 Contribution of This Work

The overall philosophy of our work closely follows the weakly supervised methods discussed earlier, of computing an embedded representation, followed by classification. However, some specific contributions of our work are as follows:

- First, we have worked out and provided an elaborate compression analysis of both the weakly supervised methods, which was discussed above. While providing the comparison with these methods, in the section discussing the results, we also provide a comparison of the compression achieved in the proposed approach.
- As a part of the method, we present a simplistic histogram-based algorithm for removing the patches that do not contribute to information regarding cells while compressing the whole slide image.
- We show that a pretrained network can be used effectively to get the compressed map of whole slide images, and such a representation yields encouraging performance.

- We also demonstrate that a cosine loss [15] combined with a learning rate schedule (cosine annealing) can be used for classification task on very small dataset where total number of samples for training are less than 400.
- We suggest a novel decision making method for the WSI, based on random crops from the embedded representation, and have also analyzed the impact of random crops on the accuracy of the model.
- Finally, to our knowledge, we believe that ours is the first work considering the weakly supervised scenario for the Camelyon17 dataset, and thus, sets the benchmark for the same.

3 Dataset

We have used the Camelyon16 and Camelyon17 datasets which are publicly available [16]. These are collected from five medical hospitals in Netherlands,

- Radboud university medical center (RUMC)
- Utrecht university medical center (UMCU)
- Rijnstate hospital (RST)
- Canisius-Wilhelmina hospital (CWZ)
- LabPON (LPON)

The most common way to detect cancer in mammary gland is to analyze the regional lymph node using sentinel lymph node procedure [8]. In this procedure a blue dye or radioactive material is injected close to tumor and then the lymph node which receives the dye or material first called sentinel lymph node is operated for the sample. Now, the sample collected by the procedure described above is sent out for analysis by a pathologist. The tissue samples are stained with hematoxylin and eosin solution.

The WSI scanners used by RUMC, CWZ, and RST are 3dhistech panoramic flash II 250, whereas UMCU and LPON used Hamamatsu NanoZoomer-XR C12000-01 scanner and Philips ultrafast scanner, respectively [8]. In these datasets, the cancerous WSIs can be categorized into three types:

- Isolated tumor cells (ITC): Number of cells in the cluster formed by metastasized tumor cells is less than 200 or cluster is not greater than 0.2mm.
- Macro-metastasis: If cluster of the tumor cells is greater than 2mm.
- Micro-metastasis: Cluster of cells larger than 0.2mm but smaller than 2mm and containing cells more than 200.

The distribution of data from the five medical center for the Camelyon17 dataset is shown in Table 1. Normal column in the table means that the cells in whole slide image are not cancerous. The images at various resolution are combined and converted to single file which is TIFF (tagged image file format). Each pixel in image is composed of three channels which are red, green, and blue. Number of

Table 1 Distribution of data from different centers for Camelyon17

Center	Total WSI	Normal	Macro	Micro	ITC
CWZ	100	64	15	10	11
LPON	100	64	25	4	7
RST	100	60	11	22	7
RUMC	100	60	19	13	8
UMCU	100	75	15	8	2
Total	500	323	85	57	35

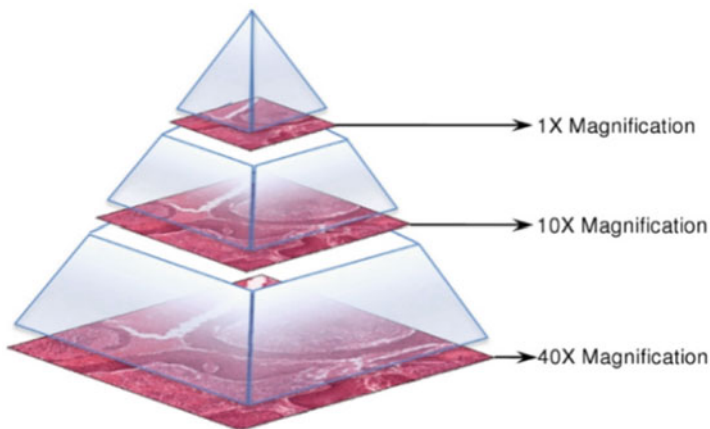


Fig. 4 Pyramid visualization of WSI. Figure reproduced from [17]

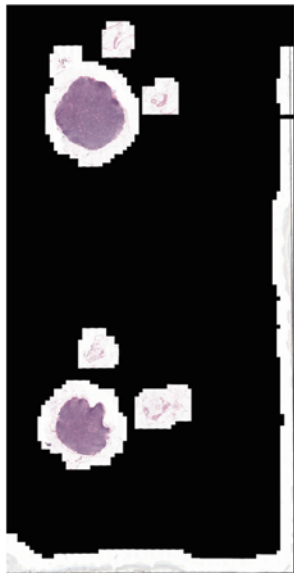
bits per channel are 8. The WSI can be visualized as a multi-resolution pyramid as shown in Fig. 4.

Labelling of the WSI For this work, as in many earlier works, the ITC, macro-metastasis, and micro-metastasis are commonly labelled as cancerous/abnormal samples, and the normal are labelled as non-cancerous/normal. Thus, the classification considered here is a 2-class classification. For Camelyon17 we have total 179 cancerous and 301 non-cancerous whole slide images. For Camelyon16 we have total 131 cancerous and 192 non-cancerous whole slide images from training dataset. We also have 49 cancerous and 80 non-cancerous whole slide images from the official test dataset of the Camelyon16. This dataset can be found on the Camelyon17 official website, link is given in [16].

4 Methodology

Our approach involves dividing the complete WSI image into tiles (patches), processing these patches for classification, and then making a decision for the WSI based on considering the classification outputs of all the patches.

Fig. 5 Patient 1. Image size at 40x: $196,000 \times 97,000 \times 3$



Thus, the primary modules in our overall approach, which we describe below in detail, are as follows: (1) Histogram-based thresholding for removing patches (tiles) of WSI which do not contain useful cellular information, (2) Extraction of feature embeddings from the tiles from a convolutional neural network (CNN), (3) Constructing a compressed version of the input WSI image using the embeddings, (4) A CNN for classification of the embedded representation, (5) Final decision making for deciding the WSI class label.

4.1 Histogram-Based Selection of Patches from WSI

We start with dividing the WSI into patches (tiles) of size $960 \times 960 \times 3$ (3 being the color channels). To remove the patches (tiles) which do not contain cellular structures, we use an approach involving the histogram analysis of the patches. Examples of WSI images, in Figs. 5 and 6, depict that some regions on a slide contains the tissue and many other regions are empty at 6mm resolution. Assuming that a mask of the regions containing useful information is not available, we suggest a simplistic approach to select the useful patches.

We show examples of some patches with their histograms in Figs. 7, 8, 9, 10, 11, 12. Our histogram-based selection method considers the histogram bins between a lower limit (a_2) and an upper limit (a_1). If the bins are not empty in this range, we consider the patch for further processing. In our case, we are able to remove most of the patches which do not contain the cells by setting $a_1 = 180$ and $a_2 = 30$. The overall algorithm is provided below as Algorithm 1.

Fig. 6 Patient 45. Image size at 40x: 83,400×53,000×3

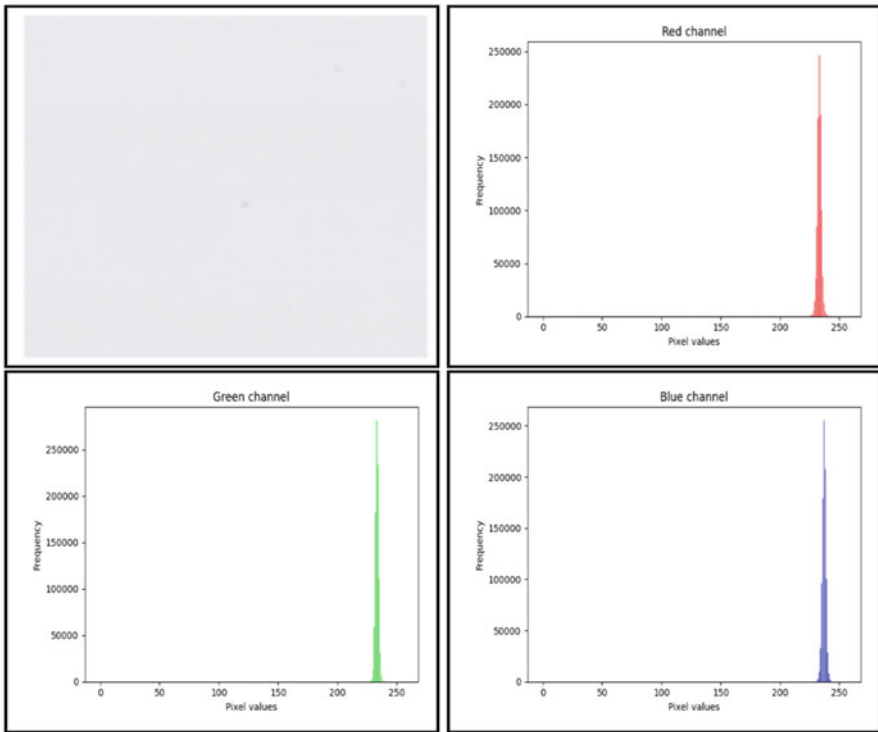
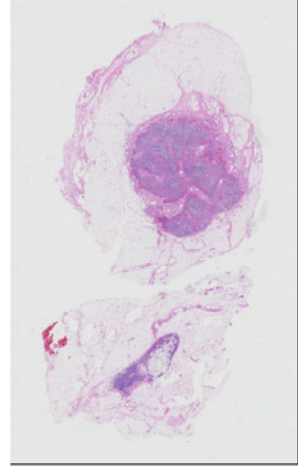


Fig. 7 Grey Patch. Top left is the actual patch, top right is the histogram of red channel, bottom left is histogram of the green channel, and bottom right is the histogram of blue channel of the patch

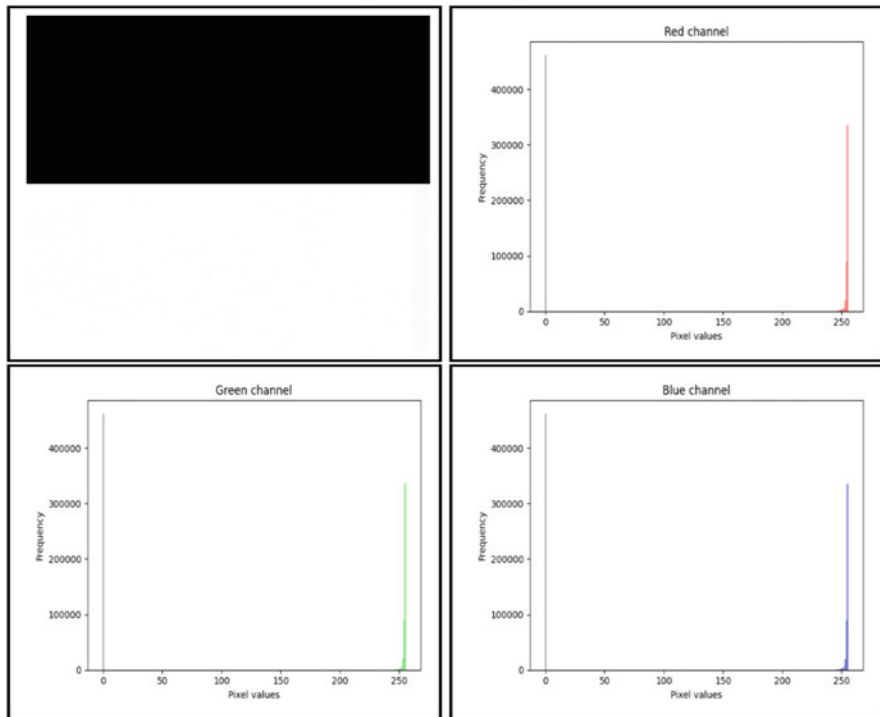


Fig. 8 Black and white patch. Top left is the actual patch, top right is the histogram of red channel, bottom left is histogram of the green channel, and bottom right is the histogram of blue channel of the patch

4.2 Extraction of Feature Embeddings Using a Pretrained Network

As indicated earlier, in [7], the authors have explored an unsupervised learning methods for encoding the patches and use the encoded feature embeddings. Unlike in [7], in this work we have used a pretrained network (Resnet50) to get the patch embeddings of the high resolution patches. We choose to employ a pretrained model for extracting the embeddings, as it is shown to be well suited for the histopathological image analysis as studied by authors of [10].

We extract the patches at highest magnification level which is 40x as shown in Fig. 4. Before extracting the features from high resolution patch of slide it was downsampled by a factor of 3 and the size of patch after downsampling is $320 \times 320 \times 3$. We extract the embedding of the patch from the pool5 layer of the Resnet-50 architecture which gives a vector of length 2048.

We divide the Whole slide image into grid of the patches where each grid cell has shape $960 \times 960 \times 3$. A patch is taken from i th grid cell is passed through the

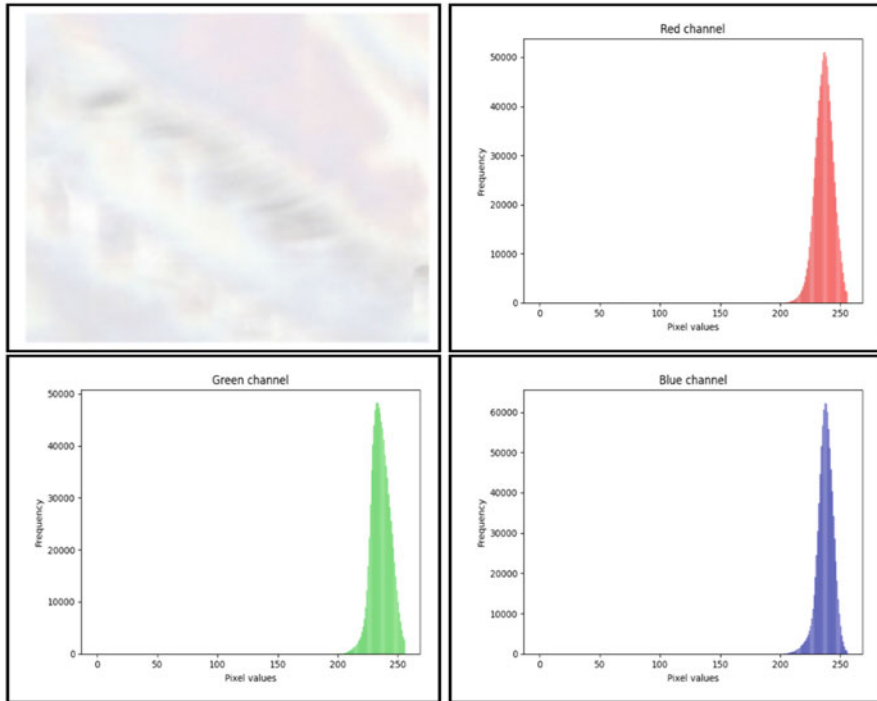


Fig. 9 No cells. Top left is the actual patch, top right is the histogram of red channel, bottom left is histogram of the green channel, and bottom right is the histogram of blue channel of the patch

feature extractor and an embedding of the size 2048 is placed at the i th location in the array of the size $A \times B \times 2048$, where 2048 represents the number of the channels in the new array, A is the total number of the rows and B is the total number of columns in grid structure of the whole slide image.

4.2.1 Compression of the WSI Images Using the Embeddings

As indicated above, an overall WSI image is now represented by an array of size $A \times B \times 2048$. Note that for encoding the same procedure is adopted as in [7]. However, instead of using $P = S = C = 128$ as used by authors in [7] we have set $P = S = 960$ and $C = 2048$.

Using the compression analysis shown in Sect. 2, we can deduce that the compression in our case is increased and is 3.51 more than the compression in [7]. Moreover, since we are using tile size of $960 \times 960 \times 3$, this also leads to reduction in encoding time. Increasing the patch size also reduces storage space on disk in comparison to when the tiles of the size $224 \times 224 \times 3$ (resulting in more number of patches) are encoded.

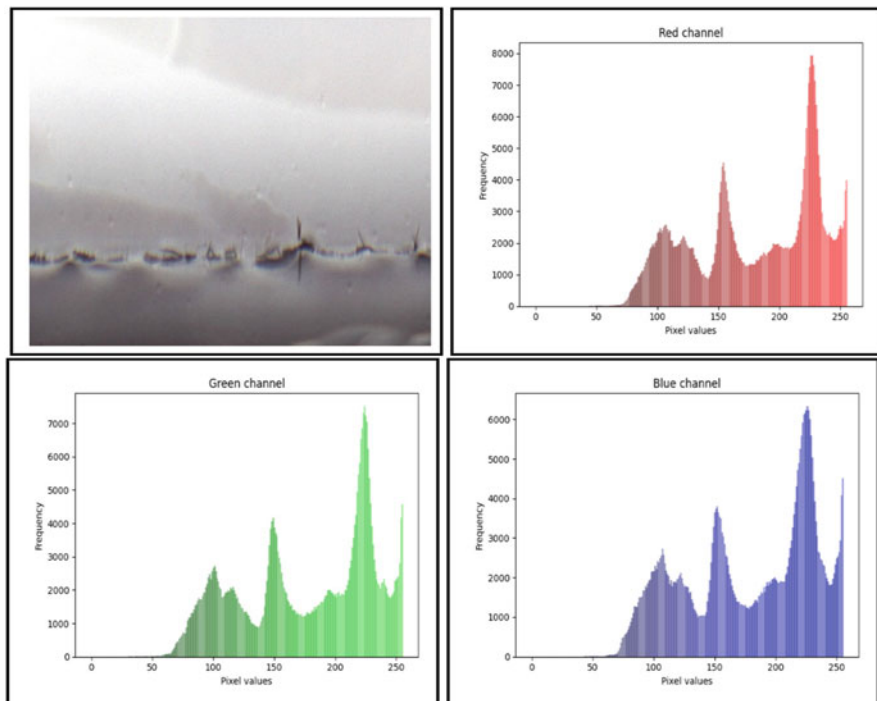


Fig. 10 No cells. Top left is the actual patch, top right is the histogram of red channel, bottom left is histogram of the green channel, and bottom right is the histogram of blue channel of the patch

Figures 13b and 14b show some example visualizations of the features extracted. Each image shows some randomly selected channels out of 2048 channels from the pool5 layer of the Resnet50. Here, a general observation is that, due to the compressed representation, even though the size of the representation is reduced, one can still visualize enough textural variations in the feature embeddings, thus indicating a good quality representability.

4.3 Classification Using the CNN

For classification of the WSIs, we have used the compressed map of the WSI, as computed above, and use it as an input to a convolution neural network (CNN), the architecture of which is described next. Specifically, the input to the convolution neural network is random crops of the compressed map of the size $60 \times 60 \times 2048$. These random crops help in preventing the overfitting of the network [7]. Note that the label associated with each random crop, to train the network, is still the image level label corresponding the original WSI, thus maintaining the weakly supervised

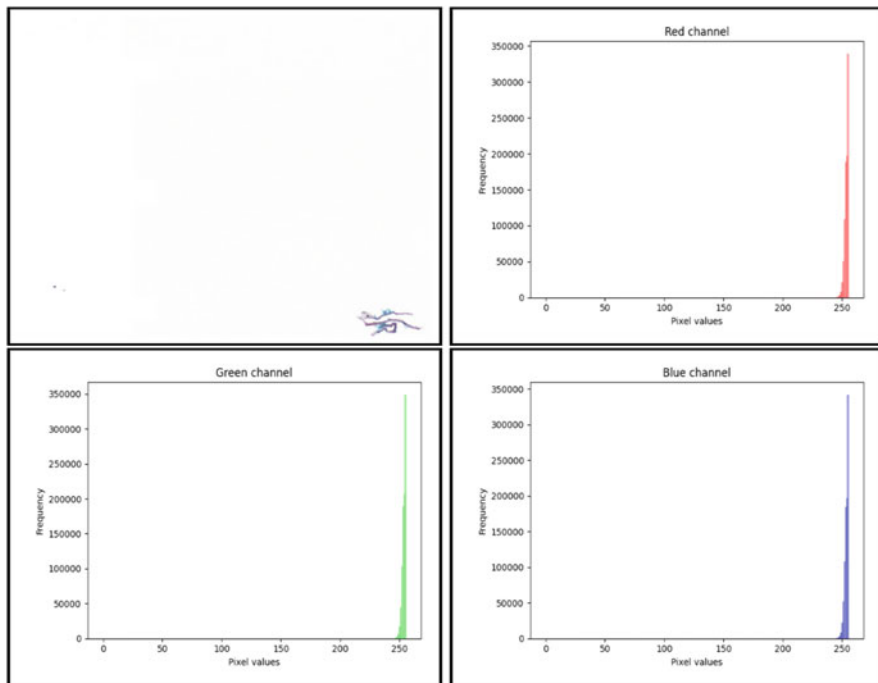


Fig. 11 Stain misplacement. Top left is the actual patch, top right is the histogram of red channel, bottom left is histogram of the green channel, and bottom right is the histogram of blue channel of the patch

paradigm. The final decision for a WSI image during test time is based on the decisions from a multitude of such individual random crops, as discussed a little later.

4.3.1 Network Architecture

We use the classification network as show in Fig. 15. Here, convolution layer is followed by batch normalization and ReLu activation. The last fully connected layer is followed by l2 normalization layer.

4.3.2 Cosine Loss

In this study we have used the network shown in Fig. 15. We have used the Adam optimizer to optimize the model with learning rate schedule (cosine annealing) adapted from [18] and l2 norm to convert the prediction space to the unit vector norm space. The final model used for testing was the one which provided the best

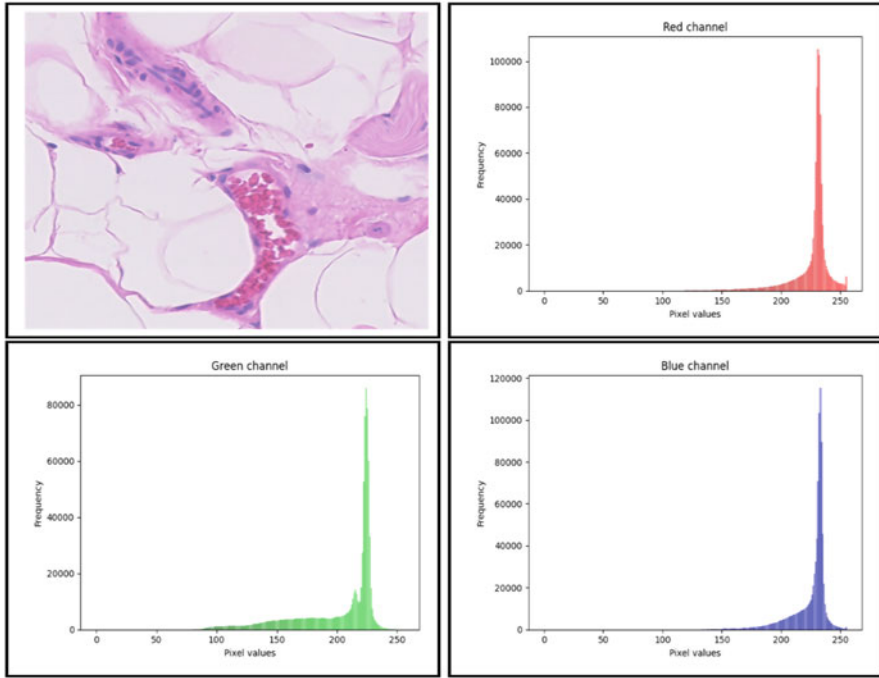


Fig. 12 Cell containing patch. Top left is the actual patch, top right is the histogram of red channel, bottom left is histogram of the green channel, and bottom right is the histogram of blue channel of the patch

Algorithm 1 Algorithm to remove patches not containing tissue

```

y = list(all patches in directory)
len_y = length(y)
for i = 0 to len_y do
    ▷ %comment: calculate the histogram of the file under analysis%
    hist = histogram(y[i])
    ▷ %comment: now take out the individual histograms of the red, blue and green channels%
    hist_red = hist[0 : 256]
    hist_green = hist[256 : 512]
    hist_blue = hist[512 : 768]
    ▷ %comment: set upper limit and lower limit to get the patch as containing cell%
    a1 = limit_upper
    a2 = limit_lower
    ▷ %comment: calculate if bin is empty in the range lower limit
    and upper limit of individual histogram%
    sum_total = sum(hist_red[a2 : a1]) + sum(hist_green[a2 : a1]) + sum(hist_blue[a2 : a1])
    if (sum > 0) then
        encode_image = encode(y[i])
    end if
end for
    
```

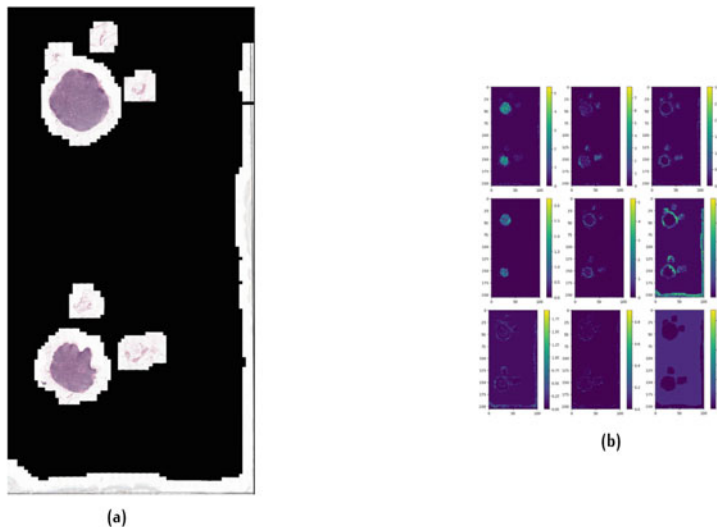


Fig. 13 Feature space visualization. (a) Actual WSI at 6mm resolution. Image size at 40x: $196,000 \times 97,000 \times 3$. (b) Channels visualization. Size of each channel is $206 \times 101 \times 3$

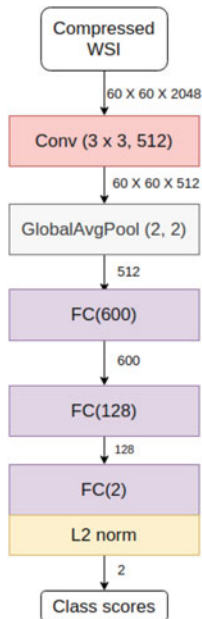


Fig. 14 Feature space visualization. (a) Actual WSI at 6mm resolution. Image size: $123,000 \times 85,000 \times 3$. (b) Channels visualization. Size of each channel is $92 \times 131 \times 3$

training and validation accuracy. Cosine loss is given in Eq. 3. While calculating the loss the label given to the random sample from the compressed map is same as the label of the whole slide image. Since, this is weakly supervised classification, annotations were not used for giving the label to the sample passed through the classification network.

$$loss = 1 - \frac{\langle a, b \rangle}{\|a\|_2 \cdot \|b\|_2} \quad (3)$$

where a and b can represent the output of the network and the corresponding ground-truth. We note that in comparison to the standard cross-entropy loss, the cosine loss exhibits two different properties: First, cosine loss takes value in a

Fig. 15 Network architecture

bound $[0, 2]$ whereas cross-entropy can take high values. Secondly, cosine loss only depends upon the direction of the feature vectors not on their magnitude, it is invariant against scaling of the feature space. These two properties have shown to improve the accuracies in the small datasets [15].

4.3.3 Learning Rate Schedule

Instead of using a fixed learning rate we have used learning rate schedule. Earlier works have suggested starting from a high learning rate and decaying it by a constant factor. Such an approach is vulnerable and is more likely to get stuck at a local minima. Instead of decaying the learning rate [19] has proposed to overcome the issue by increasing and decreasing the learning rate periodically.

$$learningrate(\epsilon) = lr_{minimum} + \frac{1}{2} \times (lr_{maximum} - lr_{minimum}) \times (1 + \cos(\frac{e_{passed}}{l_{cycle}} \pi)) \quad (4)$$

In Eq. 4, ϵ denotes the epoch number, $lr_{minimum}$ is minimum learning rate, $lr_{maximum}$ is maximum learning rate, e_{passed} is the number of epochs passed after the cycle started, and l_{cycle} is the length of the cycle. The length of the cycle is increased by a factor of 2 after each cycle. We have used learning rate schedule given in Eq. 4 in network architecture given in Fig. 15.

Table 2 Analysis of the random crops

Mean	Standard deviation	Minimum	Maximum
61.45%	24.046%	5.82%	100%

4.3.4 Classification Decision for the Whole Slide Image

As indicated above, the training of the network uses random crops from the embedded representation (compressed map) of the WSI. The testing process also follows a similar approach, where the network takes intermediate decisions on random crops (of size $60 \times 60 \times 2048$) extracted from the embedded representation of a test WSI image, and then the final decision is based on such intermediate decisions.

More specifically, for the test time classification at the whole slide image level, we have taken 100 random crops from one compressed map of the whole slide image. We label the overall WSI sample as abnormal, if the number of random crops out of 100, predicted by the network as abnormal, is greater than a threshold T where T lies in the range 0 to 100. In our section discussing the results, we provide the results for various threshold values T .

4.3.5 Study: Analysis of Random Crops

Indeed to analyze the chances of encountering abnormal regions among such random crops, we have used the available masks (annotations) for cancer regions to compute such a percentage. Note that the annotations are used only for this study, and not in the approach (thus, does not violate the weakly supervised nature of the method). A total 45 WSI were used in this study. We analyzed 10^4 random crops of the size $60 \times 60 \times 2048$ from the compressed map of each of the 45 WSI and computed the mean, min, max over such 45 WSI of the fraction of that we encounter a cancerous region. Table 2 shows the results of the analysis. Based on such an analysis we can appropriately choose/justify a range of threshold T , discussed above, to decide on the label of the WSI.

5 Results

In this section we present a variety of results from our proposed approach. These include accuracy at various threshold values discussed in Sect. 4.3.4, ROC curves, and comparisons with other weakly supervised methods discussed earlier.

5.1 *Experimental Details*

We have trained and validated our network shown in Fig. 15 separately for Camelyon16 and Camelyon17 dataset, the samples from two datasets were not mixed. For Camelyon16, considering both training and test WSI samples given in the dataset, there are a total of 452 samples. In one of our experiments we have used a 3-fold cross-validation on this total 452 samples (180 samples are cancerous and 272 samples are non-cancerous). For each fold test samples are non-overlapping with the training data. For Camelyon16, we have also carried out another experiment, where we have used the test data provided officially, separately for testing. For testing our algorithm with the official test data we have trained our network only on the training data (131 cancerous samples and 192 non-cancerous samples). Finally, for Camelyon17 we have done 5-fold cross-validation on the Camelyon17 dataset (having a total number of 179 cancerous samples and 301 non-cancerous samples) where each test partition is non-overlapping with the training partition.

5.2 *Accuracy with Varying Threshold T*

In this section, we present the accuracy on the Camelyon16 and Camelyon17 dataset when the threshold T (Sect. 4.3.4) is varied. We consider the complete range of the threshold T from 0 to 100 with stride of 1. For each fold in cross-validation, we calculate the accuracy for a particular threshold T and then average the accuracy over all the folds for that threshold T .

Figure 16a and b shows the average accuracy vs threshold for the Camelyon16 and Camelyon17. Interestingly, in Fig. 16a, which involves the same data as was used to analyze the threshold in Sect. 4.3.5, it can be seen that our results are consistent with our analysis given in Sect. 4.3.5 as accuracy is maximum at $T=62$ for both datasets.

In general, considering the weakly supervised nature of the task, the accuracy in all cases is quite encouraging. Note that, to our knowledge, this is the first work on Camelyon17 dataset for the case of weakly supervised learning. Figure 17 shows the accuracy vs threshold curve for the test dataset of the Camelyon16.

5.3 *ROC-AUC Results*

The ROC curves are computed using the 100 random samples of the size $60 \times 60 \times 2048$ from a single WSI sample representation, during the test time. ROC curve is plot of the false positive rate vs true positive rate as the threshold T for sample to be in a positive class (cancerous) is varied in the range 0 to 100. If the number of

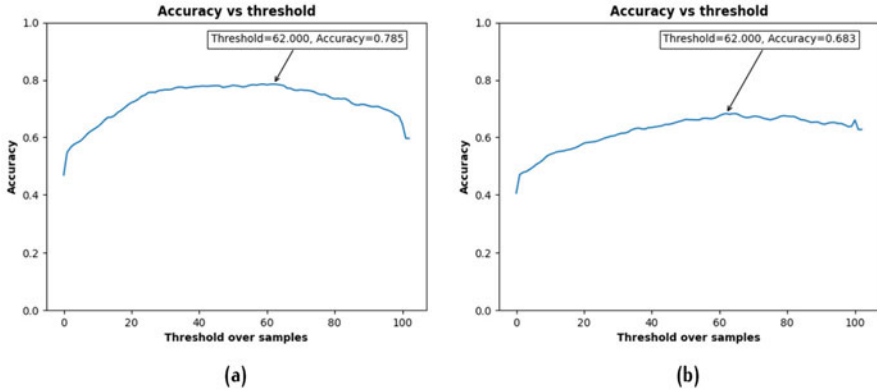


Fig. 16 Average accuracy vs Threshold curves for cross-validation of Camelyon16 and Camelyon17. (a) Camelyon16. (b) Camelyon17

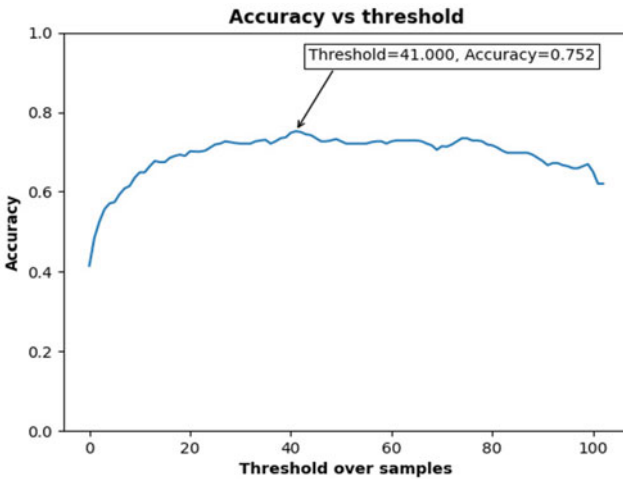


Fig. 17 Camelyon16 test data accuracy vs threshold curve

abnormal random samples predicted for a particular sample is greater than threshold T then whole slide image is labelled as cancerous (positive class) otherwise it is labelled as non-cancerous. For each threshold we calculate the true positive, false positive, true negative and false negative, based upon these values we calculate the true positive rate and false positive rate for a particular threshold T .

The ROC curves AUC for Camelyon16 can be seen in Fig. 18. From the curves for the three folds it can be inferred that the change in the true positive rate is relatively less as compared to the change in the false positive rate in the interval of 0.2 to 1 of the false positive. Hence, we can say that our algorithm has learned well to achieve a good true positive of about 0.8, even at a low false positive value.

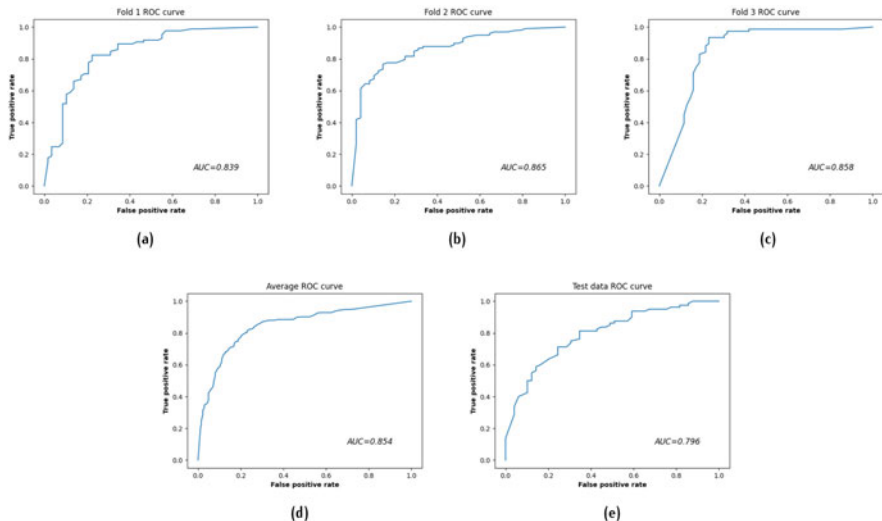


Fig. 18 ROC curves for 3 Fold cross-validation and the test dataset of Camelyon16. (a) Fold 1. (b) Fold 2. (c) Fold 3. (d) Average ROC curve for 3 folds. (e) Test data ROC curve

For blind test data (official test data of Camelyon16), it can be inferred that our network performs quite well achieving ROC-AUC of 0.79.

The ROC curves of Camelyon17 are shown in Fig. 19. Here, we note that the Camelyon17 dataset is more difficult to classify as compared to the Camelyon16 dataset. Considering that this is an early weakly supervised work on the Camelyon17 dataset, we believe that a ROC-AUC of 0.71 sets an encouraging benchmark for the future approaches.

As indicated earlier, to best of our knowledge, the weakly supervised prediction of the metastasis at WSI level is not performed for Camelyon17 dataset earlier. Thus, in Table 3, we only show our results for the ROC-AUC, over the 5-fold cross-validation, and their average.

5.4 Comparison

Table 4 presents the results for our model as compared to results presented in [7] and [10], which are the two state-of-the-art weakly supervised methods reported on the Camelyon16 dataset. The Test column represents the results for the test dataset provided in Camelyon16. Note that our method outperforms the neural image compression work (BiGAN method) in terms of AUC on cross-validation as well as on test dataset.

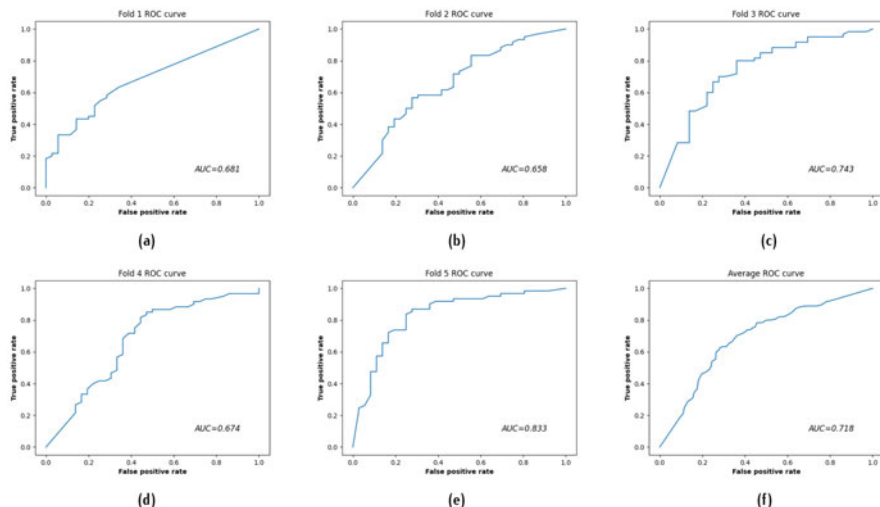


Fig. 19 ROC curves for the 5 fold cross-validation of Camelyon17. (a) Fold 1. (b) Fold 2. (c) Fold 3. (d) Fold 4. (e) Fold 5. (f) Average

Table 3 Camelyon17 metastasis presence prediction at WSI level (ROC-AUC)

Metric	1st fold	2nd fold	3rd fold	4th fold	5th fold	Average
ROC-AUC	0.681	0.657	0.743	0.674	0.833	0.718

Table 4 Camelyon16 metastasis presence prediction at WSI level (ROC-AUC)

Network	Cross-validation	Test
[7]	0.725	0.704
[10]	0.903	0.858
Ours	0.854	0.795

While the approach in [10] performs better in terms of the evaluation metrics on the Camelyon16 dataset, it significantly lags in the compression ability of the representability, as given below, thus suggesting a practical trade-off.

From the compression analysis given in Sect. 2, [7] achieves a compression of 384 and [10] achieves a compression of 73.5 with respect to whole slide image. We have achieved the compression of 1350, which is 3.51 times more than [7] and 18.36 times more than [10]. Hence our method achieves higher compression rate than both [7] and [10], outperforms [7] and achieves the ROC-AUC comparable to the method given in [10]. Thus, overall, we believe that our approach can be considered as a competitive method, especially given that such weakly supervised methods are quite sparse.

Finally, a limited set of experiments suggested that the result from a trained network in our approach can be obtained within 10 min after a slide is digitized, which suggests a reasonably good efficiency, considering the purpose of a CAD system for efficient training.

6 Conclusion

The focus of this work was two-fold: To highlight the importance and challenges in the weakly supervised histopathology WSI image classification for breast cancer detection, and to propose an approach in a direction to address the task. We provided a detailed overview of a small number of works for weakly supervised classification in this domain, especially stressing on the compressed representation of WSI images, which is an important consideration in addition to the classification performance, from a system perspective. Maintaining the same important philosophy of achieving classification via a compressed representation, we suggested some new directions for the classification in a deep learning framework including a patch-based decision criteria, cosine loss based network training, and using a non-conventional learning rate schedule. We have also shown that instead of training neural networks from scratch, pretrained networks can also yield useful embeddings. Our results on two publicly available datasets are encouraging, and quite competitive in terms of performance and compressed representation. Considering that there are few weakly supervised methods in this domain, we believe that the proposed work also sets a good benchmark in this area.

References

1. Globocan 2018, "Fact sheet for cancer in India", <https://gco.iarc.fr/today/data/factsheets/populations/356-india-fact-sheets.pdf>
2. Dr. Brian Koffman, "The chronic lymphocytic leukemia" <https://cllsociety.org/quarter-4-2015-volume-1-issue-2/>
3. Renato Ferreirat et al, "The virtual microscope" *AMIA, Inc.*, 1997.
4. Navid Farahani and Anil V Parwani and Iiron pantanowitz, "Whole slide imaging in pathology: advantages, limitation and emerging perspectives" *Pathology and Laboratory Medicine International*, 2015.
5. ASAP Source: <https://computationalpathologygroup.github.io/ASAP/#home>
6. Pedro O. Pinheiro, Ronan Collobert "From Image-level to Pixel-level Labeling with Convolutional Networks" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1713–1721
7. David Tellez and Geert J. S. Litjens and Jeroen van der Laak and Francesco Ciompi, "Neural Image Compression for Gigapixel Histopathology Image Analysis" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
8. Geert et al, "1399 HE-stained sentinel lymph node sections of breast cancer patients, the CAMELYON dataset" *Oxford GigaScience*, 2018.
9. M. Veta et al., "Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge", *Medical image analysis*, pp. 111–121, 2019.
10. Pierre Courtiol, Eric W. Tramel, Marc Sanselme, and Gilles Wainrib, "Classification and disease localization in histopathology using only global labels: a weakly supervised approach", <http://arxiv.org/abs/1802.02212>, 2020.
11. B. Ehteshami Bejnordi et al, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer", *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017.

12. Sanghun Lee, Joonyoung Cho, Sun Woo Kim, "AUTOMATIC CLASSIFICATION ON PATIENT-LEVEL BREAST CANCER METASTASES", *CAMELYON 2017*, 2019.
13. Nicolas Pinchaud, "Camelyon17 challenge" *CAMELYON 2017*, 2019.
14. Jiahui Li et al, "signet ring cell detection with a semi-supervised learning framework" *CAMELYON 2017*, 2019.
15. Björn Barz, Joachim Denzler, "Deep Learning on Small Datasets without Pre-Training using Cosine Loss" *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020.
16. Camelyon dataset link : <https://camelyon17.grand-challenge.org/Data/>
17. Pyramid structure of WSI, Source: CAMELYON 2016 website, <https://camelyon16.grand-challenge.org/Data/>
18. Björn Barz, Joachim Denzler, "Deep Learning is not a Matter of Depth but of Good Training", *International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI)*, pages 683–687, 2018.
19. Ilya Loshchilov, Frank Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts" *International Conference on Learning Representations (ICLR)*, 2017.

Lung Classification for COVID-19



Norliza Mohd. Noor and Muhammad Samer Sallam

1 Introduction

In December 2019, in the city of Wuhan, China, Corona virus Disease 2019 (COVID-19) launched a pandemic, triggering a Public Health Emergency of International Significance (PHEIC) [1]. The World Health Organization (WHO) called COVID-19 as a novel infectious disease linked to coronaviruses (CoV) and hazardous viruses [2, 3]. In some cases, this results in critical respiratory problems, such as Extreme Acute Respiratory Syndrome (SARS-CoV) and Middle East Respiratory Syndrome (MERS-CoV), leading inevitably to breathing failure and death. Currently, the total number of confirmed cases has reached 367,166 cases worldwide with 6 million deaths [4]. Every year, widespread lung infections such as viral and bacterial pneumonia also result in thousands of deaths [5]. Pneumonia is another form of lung disease that is comparable to COVID-19. Via the accumulation of pus and other liquids in air sacs, this pneumonia disease causes fungal infection of one or both sides of the lungs. Viral pneumonia symptoms develop progressively and they are mild. But bacterial pneumonia, especially among children, is more severe [6]. Many lobes of the lung may be affected by this form of pneumonia. The real-time polymerase chain reaction (RT-PCR) assay of sputum [7] is the gold standard of diagnosis for common pneumonia diseases and coronaviruses. To confirm positive COVID-19 cases, however, these RT-PCR tests showed high false negative levels. Alternatively, radiological tests are also

N. M. Noor (✉)

Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur Campus, Malaysia

e-mail: Norliza@utm.my

M. S. Sallam

Quaking Aspen, Dublin, Ireland

being used to determine the health status of infected patients, using chest X-rays and computed tomography (CT) scans [8, 9]. CT scans are an efficient tool for screening, diagnosis, and evaluation of improvement in patients with COVID-19 [10]. Nonetheless, the modality of CT imaging involves high dose exposure to patients scanned and also a high hospital bill for CT screening [11]. Conventional X-ray devices, on the other hand, are still available in hospitals and clinical centers to provide two-dimensional (2D) images of a rapid scan of the patient's lungs. Chest X-ray scans are, therefore, the first method for clinicians to diagnose pneumonia or confirm cases of COVID-19 [8, 12]. To diagnose most of the patients' COVID-19 infected cases, standard Chest X-ray and computed tomography scan images were used. The limited number of COVID-19 test kits and the lack of accuracy often lead to specialized physician lakes, especially in remote areas, saving medical professionals' valuable time [13]. Due to the rapid rise in COVID-19 cases and in order to rapidly treat affected patients, the importance of creating an automated diagnostic assistance system has become an urgent need. These challenges can be solved and assisted with accurate disease detection by the advent of deep learning as an artificial intelligence technique. In helping to interpret the computed tomography scan images, deep learning presented promising results [14–16]. Wang S et al. [17] provided an automated diagnostic and prognostic analysis of COVID-19 based on DenseNet121-FPN for raw chest computed tomography scan image segmentation. Ozturk et al. [18] suggested real-time identification and classification of COVID-19 using X-ray images for COVID-19. They suggested real-time detection and classification for COVID-19 from X-ray images based on 17 convolutionary layers with different filtering on each layer using the DarkNet model. Seung et al. [19] suggested CNN and the PyTorch frame-dependent deep learning model using three binary decision-trees as classifiers of infected or uninfected chest X-ray images. Elaziz et al. [20] indicated that new fractional multi-channel exponent moments were used to extract the function from chest X-ray images. They proposed that the multi-core parallel calculation framework was used to speed up the computational process. Then, the essential characteristics were chosen using a modified manta ray foraging optimization strategy using differential evolution.

2 Materials and Methods

2.1 Data Collection

In this section, some publicly available COVID 19 datasets used in the development of the cloud-based lung classification system for COVID–non-COVID lungs are described in detail. Mainly, the section is divided into two parts. The first section describes each dataset individually wherein total 17 datasets have been covered. The second section provides a high-level overview for the available datasets, which is helpful for the researchers to combine multiple datasets together.

2.1.1 Description of the COVID-19 Dataset

A total of 17 datasets have been discovered and described on the Internet. The following list shows all the 17 datasets in which the used name for each dataset matches exactly the name from the original source:

1. Twitter COVID-19 CXR dataset (twitter.com/ChestImaging)
2. NIH Chest X-rays (www.kaggle.com/nih-chest-xrays/data) [21]
3. Labeled Optical Coherence Tomography (OCT) (data.mendeley.com/datasets/rscbjbr9sj/2) [22]
4. Kaggle (X-ray images of Pneumonia) (www.kaggle.com/paultimothymooney/chest-xray-pneumonia) [22]
5. Figure1-COVID-chest X-ray dataset (github.com/agchung/Figure1-COVID-chestxray-dataset)
6. Dropbox dataset (www.dropbox.com/s/09b5nutjxotmftm/data_upload_v2.zip?dl=0)
7. COVID19_Pneumonia_Normal_Chest_X-ray_PA_Dataset (www.kaggle.com/asraf047/covid19-pneumonia-normal-chest-xray-pa-dataset)
8. COVID19_classifier_dataset (www.kaggle.com/rgaltro/newdataset)
9. COVID19 High quality images (www.kaggle.com/theroyakash/covid1)
10. COVID-Net dataset (<https://github.com/ieee8023/covid-chestxray-dataset>) [23]
<https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>)
11. COVID-CT dataset (github.com/UCSD-AI4H/COVID-CT) [24]
12. COVID-19 X-rays (www.kaggle.com/andrewmvd/convid19-X-rays)
13. COVID-19 Radiography Database (<https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>) [25]
14. Covid-19 Image Dataset (www.kaggle.com/pranavrakokte/covid19-image-dataset)
15. CoronaHack Chest X-Ray-Dataset (www.kaggle.com/praveengovi/coronahack-chest-xraydataset)
16. Chest X-ray (COVID-19 and Pneumonia) (www.kaggle.com/prashant268/chest-xray-covid19-pneumonia)
17. Actualmed-C OVID-chest X-ray-dataset (github.com/agchung/Actualmed-COVID-chestx-ray-dataset)

2.2 Datasets Summary

This section provides a high-level overview for all the available dataset allowing researchers to compare the available datasets easily. The following details over all the datasets will be discussed in this section; datasets sizes, average datasets images size, number of classes per dataset, number of images per dataset, from classes perspective, from extensions perspective, from data types perspective, from average datasets images height perspective, and from average datasets images width

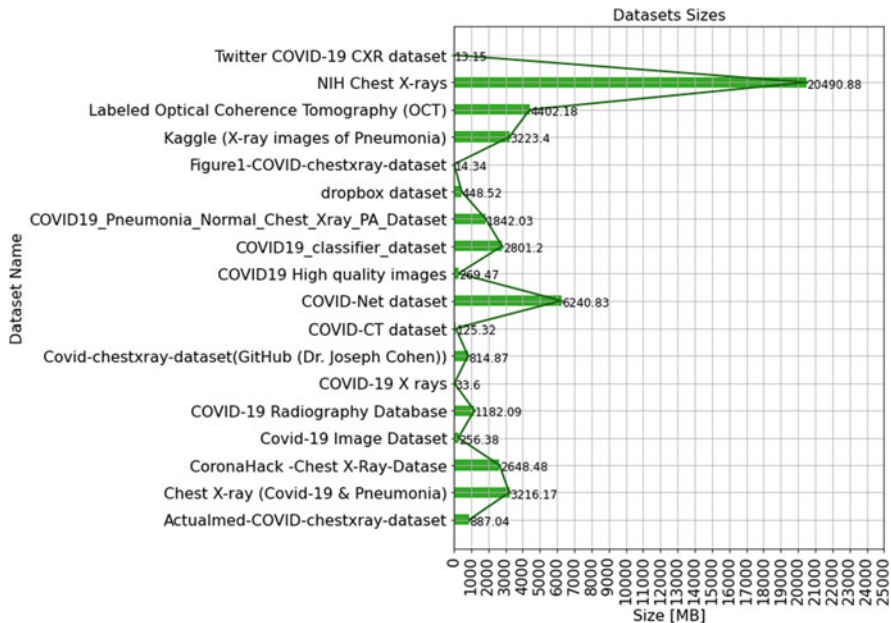


Fig. 1 Datasets sizes

perspective. Figure 1 shows that nine datasets have size less than 1000 MB, one dataset has size more than 20000 MB, and eight datasets are in the range of 1000–7000 MB. Figure 2 shows that the majority of the dataset have image size less than 1 MB. Figure 3 shows the number of classes per dataset, where one dataset has one class, one dataset has four classes, seven datasets have two classes, and nine datasets have three classes. It can be seen from Fig. 4 that one dataset has about 53000 images, ten datasets have 1000 to 18000 images, and seven datasets have less than 1000 images. Figure 5 shows the class distribution over all the datasets where pneumonia images and normal images are in majority in these datasets. Figure 6 shows the percentage of images per class, and Fig. 7 shows the number of images per class. It is clear that the majority of images represent other class, pneumonia class, and normal class. So far, there are only 4801 COVID images. Figure 8 shows the number of images per dataset from image type perspective, which includes X-ray and CT. Figure 9 shows the percentage of images per image type, and Fig. 10 represents the number of images per image type. It is clear that the majority of the images represent X-ray (99 %). Figure 11 shows that all of the images in the datasets are JPG and PNG, where 96.8% of the images or 117238 images are JPG images.

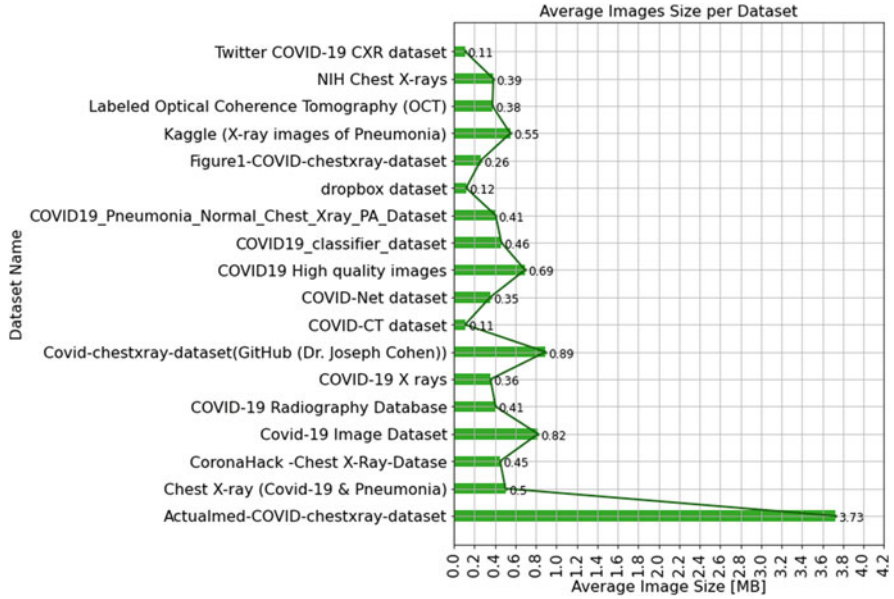


Fig. 2 Average images size per dataset

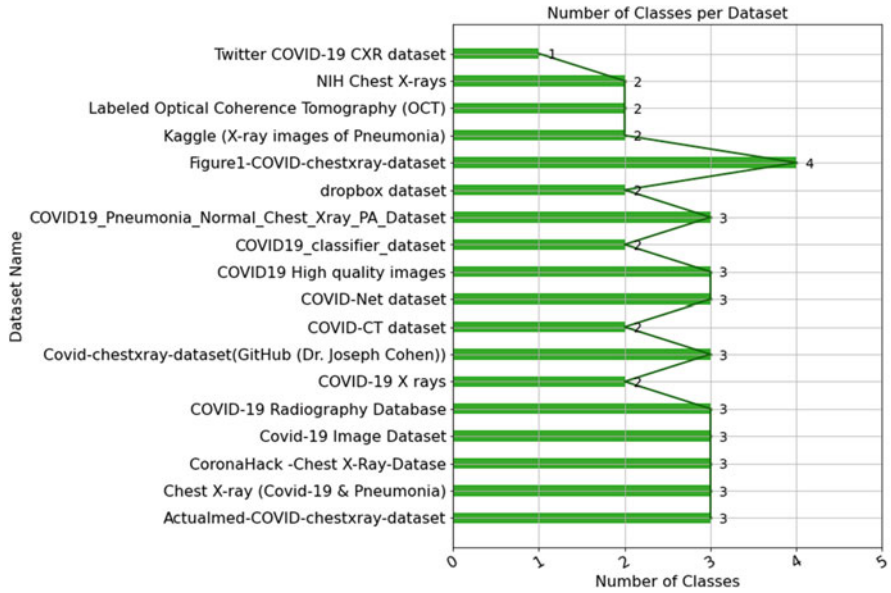


Fig. 3 Number of classes per dataset

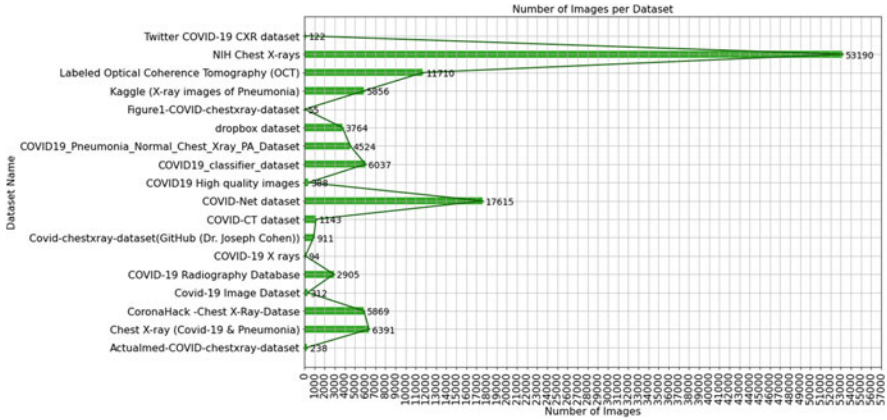


Fig. 4 Number of images per dataset

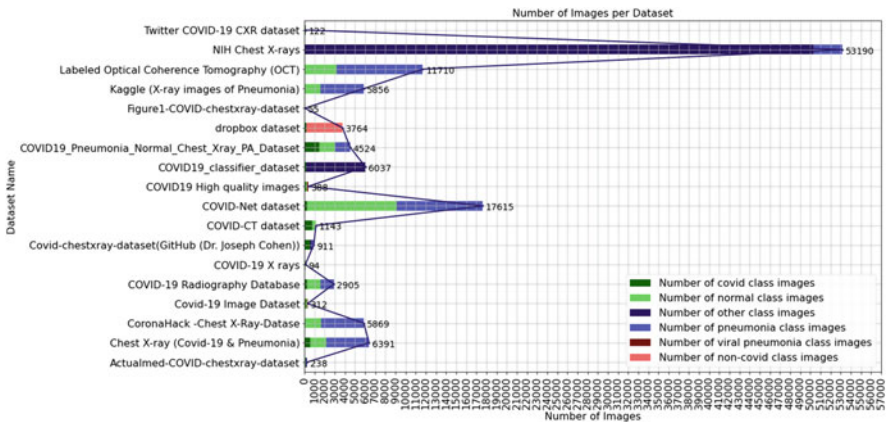


Fig. 5 Number of images per dataset from classes perspective

3 Methods

The cloud-based lung COVID–non-COVID classification system developed in this study consists of two stage classification, where the first stage will classify normal and abnormal lung using chest X-ray images as shown in Fig. 12. Those classified as abnormal lung will then be further classified into COVID and non-COVID lungs. Deep learning approaches, namely, various ResNet and DenseNet neural networks, were investigated in this study. We utilized the Amazon Web Service (AWS) cloud-computing service during the development and the testing of the system.

Fig. 6 Percentage of images per class

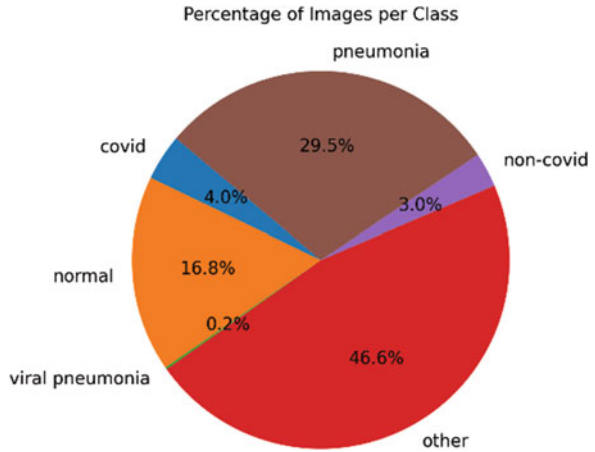
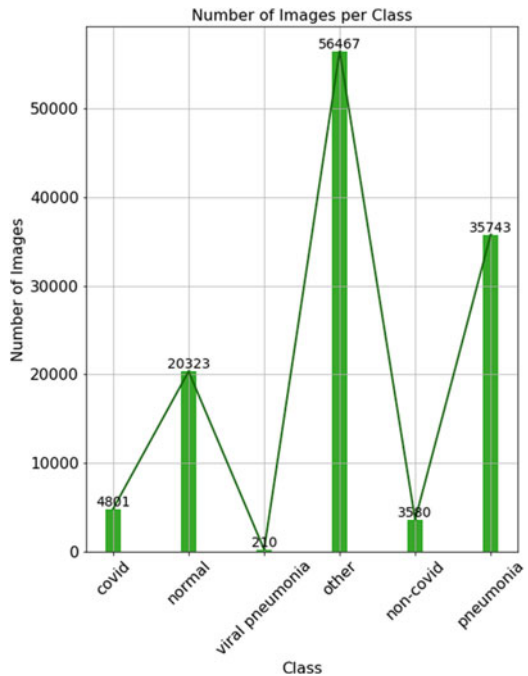


Fig. 7 Number of images per class



3.1 Normal–Abnormal Lung Classification

For normal versus abnormal lung classification stage, a total of 23838 X-ray images have been collected from multiple datasets explained in the data collection section. There were 11838 normal and 12000 abnormal images, where abnormal images consist of 775 COVID-19 X-ray images, 5000 pneumonia X-ray images, 6225 other lung diseases X-ray images. The summary of the dataset used in this experiment is

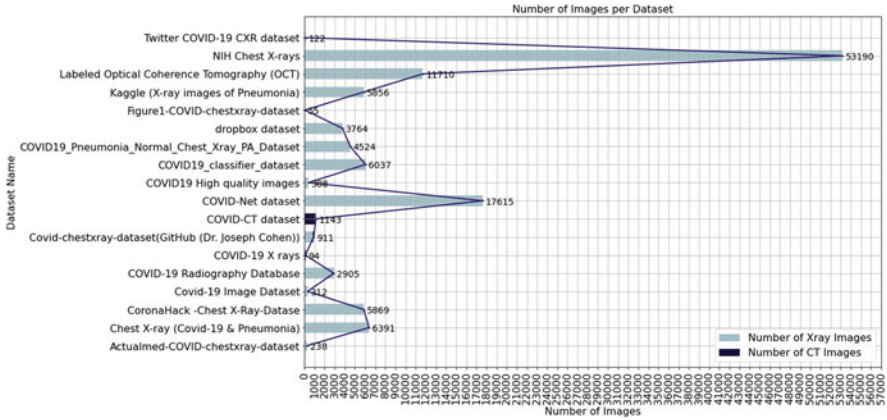


Fig. 8 Number of images per type per dataset

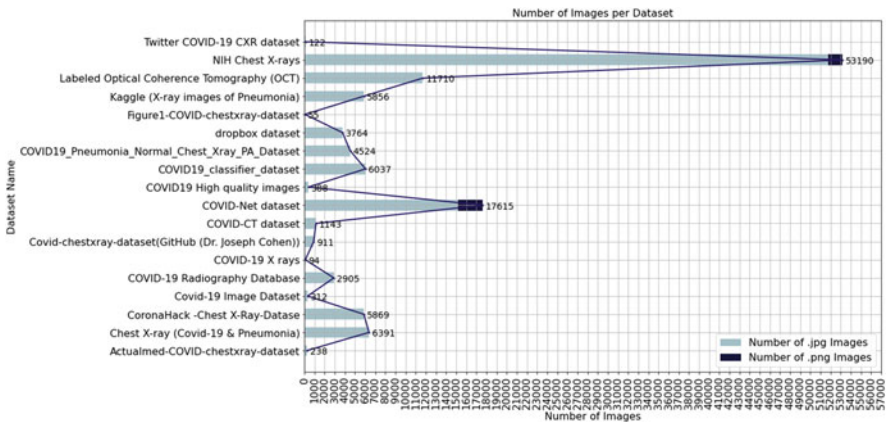


Fig. 9 Number of images per dataset from image extensions perspective

shown in Table 1. All images have JPG extension and have size dimension greater than (224, 224). To build the normal vs. abnormal model, the transfer learning concept has been applied which aims to simply use the acquired knowledge from a problem and use it to solve another problem. ResNet and DenseNet neural networks have been tested to build the model.

For all of the models, the final layer of the network has been removed, and a new layer of two neurons has been added since the problem is a binary classification problem. The training process is split into two stages:

1. After replacing the final layer, we train just the new layer for four epochs.
2. Training the entire model for 200 epochs.

Fig. 10 Percentage of images per extension

Percentage of Images per Extension

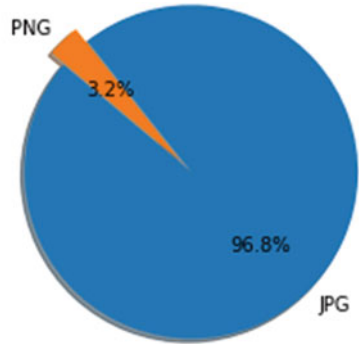
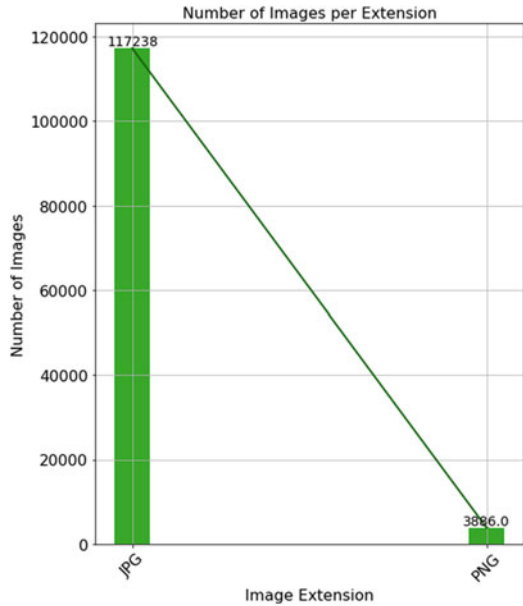


Fig. 11 Number of images per extension



The idea is to avoid getting high gradient values if we train the entire model directly after adding the new layer. In order to train the model, the following data augmentation has been used to increase the number of images: random horizontal flip, random vertical flip, random affine with 180 degrees, and the scale factor in the range of 0.9–1.1]. The Adam optimizer with learning rates of 0.01, 0.001, and 0.0001 has been tested with 64 as the batch size and cross entropy as the loss function. The dataset is split to 80% for training and 20% for testing.

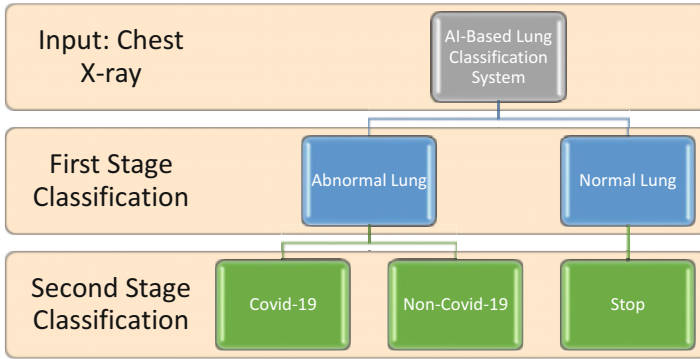


Fig. 12 Flowchart of the cloud-based COVID–non-COVID chest X-ray classification system

Table 1 The dataset used in normal–abnormal lung classification

Dataset name	Normal images	COVID-19 images	Pneumonia images	Other disease images
Actualmed-COVID-chestx-ray-dataset	53	58	–	127
COVID-19 Radiography Database	1341	219	768	–
COVID-19 X-rays	7	71	–	–
COVID-Net dataset	8851	40	–	–
COVID-chestx-ray-dataset (GitHub (Dr. Joseph Cohen))	–	257	–	–
Figure1-COVID-chestx-ray-dataset	3	8		
Kaggle (X-ray images of pneumonia)	1583		4232	
NIH Chest X-rays				6098
Twitter COVID-19 CXR dataset		122		
Total	11838	775	5000	6225

3.2 COVID–Non-COVID Classification

For COVID versus non-COVID lung classification stage, a total of 1548 X-ray images have been collected from multiple datasets explained in the data collection section. The images were divided into 774 COVID and 774 non-COVID, where non-COVID images consist of 374 pneumonia X-ray images and 400 other lung diseases X-ray images. The summary of the data used in this experiment is shown in Table 2. All images have JPG extension and size dimension greater than (224, 224). ResNet and DenseNet neural networks have been tested to build the model.

Transfer learning concept has been applied, which aims to simply use the acquired knowledge from a problem and use it to solve another problem. For all of the models, the final layer of the network has been removed, and a new layer of

Table 2 The dataset used in COVID–non-COVID lung classification

Dataset name	Covid19 images	Pneumonia images	Other diseases images
Actualmed-COVID-chestx-ray-dataset	58	–	–
COVID-19 Radiography Database	219	–	–
COVID-19 X rays	71	–	–
COVID-Net dataset	39	2	–
COVID-chestx-ray-dataset(GitHub (Dr. Joseph Cohen))	257	–	1
Figure1-COVID-chestxray-dataset	8	–	–
Kaggle (X-ray images of pneumonia)	–	–	–
NIH Chest X-rays	–	372	399
Twitter COVID-19 CXR dataset	122	–	–
Total	774	374	400

two neurons has been added since the problem is a binary classification problem. The training process is split into two stages:

1. After replacing the final layer, we train just the new layer for four epochs.
2. Training the entire model for 200 epochs.

The idea is to avoid getting high gradient values if we train the entire model directly after adding the new layer. In order to train the model, the following data augmentation has been used to increase the number of images: random horizontal flip, random vertical flip, random affine with 180 degrees, and the scale factor in the range of 0.9–1.1. The Adam optimizer with learning rates 0.01, 0.001, and 0.0001 has been tested, with 64 as the batch size and cross entropy as the loss function. The dataset is split to 80% for training and 20% for testing.

4 Results and Discussion

4.1 Normal–Abnormal Lung Classification

ResNet18, ResNet34, ResNet50, and ResNet101 were investigated with 0.01, 0.001, and 0.0001 learning rate. The accuracy results for each network are given in Fig. 13. The graph shows that ResNet50 with the learning rate of 0.001 gave the highest accuracy of 88.4% with the 80/20 split for training and testing. We repeat the experiment using the k-fold cross validation method with $k = 5$ for ResNet50, and the average accuracy obtained is 87.62%.

Next, DenseNet neural networks were investigated. In this experiment, DenseNet121, DenseNet169, and DenseNet201 were applied with 0.01, 0.001, and 0.0001 learning rate. The accuracy results for each network are given in Fig. 14. The

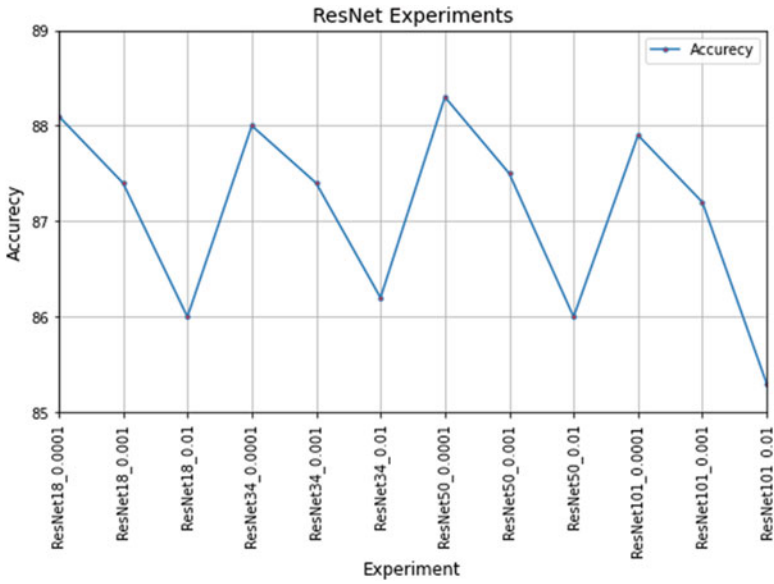


Fig. 13 The accuracy for various ResNet neural networks with 0.01, 0.001, and 0.0001 learning rate for normal–abnormal lung classification

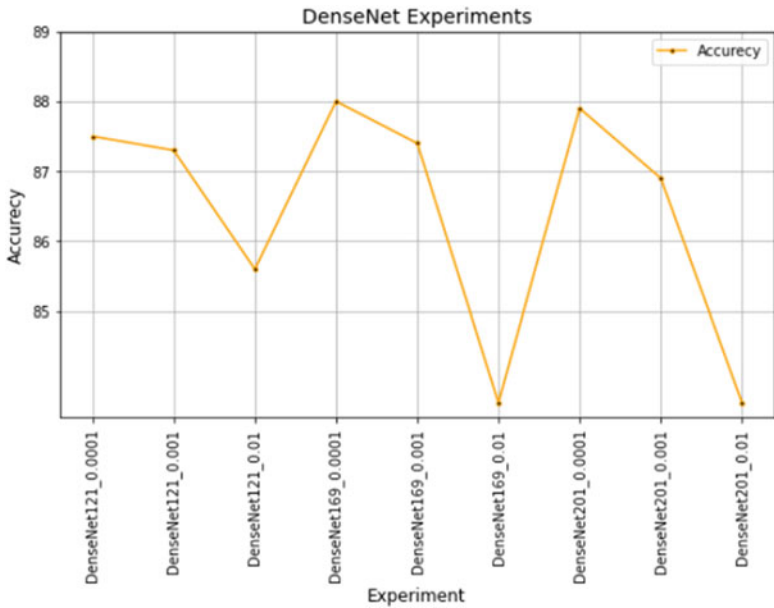


Fig. 14 The accuracy for various DenseNet neural networks with 0.01, 0.001, and 0.0001 learning rate for normal–abnormal classification

graph shows that DenseNet169 with 0.001 learning rate gave the highest accuracy of 88% with the 80/20 split for training and testing. We repeat the experiment using the k-fold cross validation method with $k = 5$ for DenseNet169, and the average accuracy obtained is 87.5%.

The best model chosen for the Normal–Abnormal Lung Classification system is ResNet50 because it achieved the highest test accuracy compared to DenseNet169.

4.2 COVID-Non-COVID Lung Classification

The classified abnormal lung X-ray images will then be fed into the second stage classification to classify COVID and non-COVID cases. The same neural networks were applied, which were ResNet18, ResNet34, ResNet50, and ResNet101 with 0.01, 0.001, and 0.0001 learning rate. The accuracy results for each network are given in Fig. 15. The graph shows that ResNet34 with the learning rate of 0.01, ResNet50 and ResNet101 with 0.01 learning rate gave the highest accuracy of 99.7% with the 80/20 split for training and testing. We repeat the experiment using the k-fold cross validation method with $k = 5$ for ResNet34, ResNet50, and ResNet101, and all of them gave the average accuracy of 99.4%.

Next, DenseNet neural networks were investigated. In this experiment, DenseNet121, DenseNet169, and DenseNet201 were applied with 0.01, 0.001,

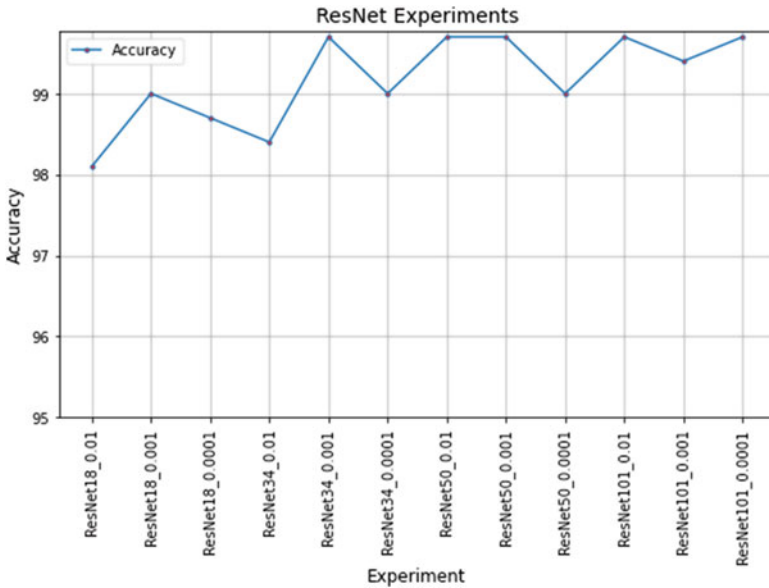


Fig. 15 The accuracy for various ResNet neural networks with 0.01, 0.001, and 0.0001 learning rate for COVID–non-COVID lung classification

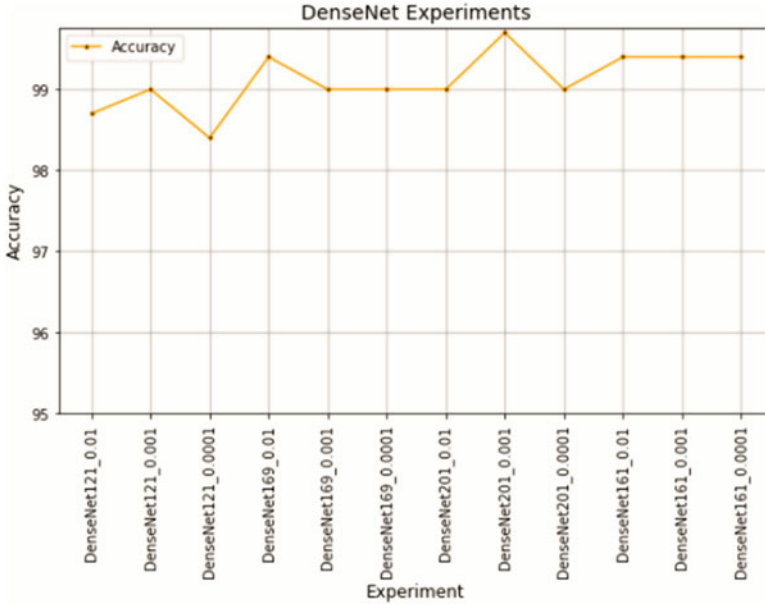


Fig. 16 The accuracy for various DenseNet neural networks with 0.01, 0.001, and 0.0001 learning rate for COVID–non-COVID classification

and 0.0001 learning rate. The accuracy results for each network are given in Fig. 16. The graph shows that DenseNet201 with 0.001 learning rate gave the highest accuracy of 100% with the 80/20 split for training and testing. We repeat the experiment using the k-fold cross validation method with $k = 5$ for DenseNet201, and the average accuracy obtained is 99.6%.

The best model chosen for the COVID–non-COVID Lung Classification system is DenseNet201 because it achieved the highest test accuracy compared to ResNet34, ResNet50, and ResNet101.

4.3 Testing the Cloud-Based COVID Lung Classification System

We tested the proposed systems with the unused dataset. For the first stage (normal–abnormal), transfer learning using ResNet50 deep learning architecture was utilized, which showed 96% accuracy. For the second stage (COVID–non-COVID), transfer learning using DenseNet201 deep learning architecture was developed. We tested using various COVID–non-COVID database available and obtained 70% accuracy. Lower accuracy may be due to a small number of COVID lung X-ray images used in the training. For comparison, Chowdhury et al. [25] followed a binary

classification scheme, normal and COVID using chest X-ray images and they obtained 99.7% accuracy. They also followed a three-class classification scheme, Normal/Pneumonia/COVID-19, with 99.7% accuracy.

Our proposed system, a two-stage lung classification system is able to provide a more systematic differential classification scheme, where normal–abnormal cases are classified first, and at the next stage, the abnormal cases are then classified into COVID and non-COVID cases. For future work, the non-COVID cases will then be classified into various lung diseases such as pneumonia and tuberculosis.

5 Conclusion

This study proposed a two-stage classification approach to classify COVID chest X-ray images. The first stage is to classify normal–abnormal lung, and the second stage is to classify the abnormal lung into COVID and non-COVID. For the first stage, our proposed system obtained 96% accuracy which is on par with other researchers; however, for the second stage, our proposed system only managed to achieve 70% accuracy due to low number of COVID lung X-ray images used in the training.

References

1. A. J. Rodriguez-Morales et al., “Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis,” *Travel Medicine and Infectious Disease*, p. 101623, 2020/03/13/ 2020.
2. C. I. Paules, H. D. Marston, and A. S. Fauci, “Coronavirus Infections -More Than Just the Common Cold,” *JAMA*, vol. 323, no. 8, pp. 707708, 2020.
3. C. Sohrabi et al., “World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19),” *International Journal of Surgery*, vol. 76, pp. 71–76, 2020/04/01/ 2020.
4. WHO Coronavirus disease 2019 (COVID-19) Situation Report – 132, 31st May 2020.
5. Y. K. Lim, O. J. Kweon, H. R. Kim, T.-H. Kim, and M.-K. Lee, “Impact of bacterial and viral coinfection in community-acquired pneumonia in adults,” *Diagnostic Microbiology and Infectious Disease*, vol. 94, no. 1, pp. 50–54, 2019/05/01/ 2019.
6. J. A. Scott, W. A. Brooks, J. S. Peiris, D. Holtzman, and E. K. Mulholland, “Pneumonia research to reduce childhood mortality in the developing world,” *J Clin Invest*, vol. 118, no. 4, pp. 1291–300, Apr 2008.
7. P. Huang et al., “Use of Chest CT in Combination with Negative RTPCR Assay for the 2019 Novel Coronavirus but High Clinical Suspicion,” *Radiology*, vol. 295, no. 1, pp. 22–23, 2020.
8. M.-Y. Ng et al., “Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review,” *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, p. e200034, 2020.
9. H. Liu, F. Liu, J. Li, T. Zhang, D. Wang, and W. Lan, “Clinical and CT imaging features of the COVID-19 pneumonia: Focus on pregnant women and children,” *Journal of Infection*, 2020/03/21/ 2020.
10. J. Ding et al., “Experience on radiological examinations and infection prevention for COVID-19 in radiology department,” *Radiology of Infectious Diseases*, 2020/03/31/ 2020.

11. L. J. M. Kroft, L. van der Velden, I. H. Giron, J. J. H. Roelofs, A. de Roos, and J. Geleijns, "Added Value of Ultra-low-dose Computed Tomography, Dose Equivalent to Chest X-Ray Radiography, for Diagnosing Chest Pathology," *Journal of Thoracic Imaging*, vol. 34, no. 3, pp. 179–186, 2019.
12. N. Chen et al., "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," *Lancet*, vol. 395, no. 10223, pp. 507–513, Feb 15 2020.
13. Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, and U. Rajendra Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, *Computers in Biology and Medicine*, Vol. 121, (103792), June 2020.
14. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, Liu Y, Gevaert O, Wang K, Zhu Y, Zhou H, Liu Z, Tian J. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *European Respiratory Journal* 2019; 53 (3): 1800986.
15. Walsh SL, Humphries SM, Wells AU, Brown KK. Imaging research in fibrotic lung disease; applying deep learning to unsolved problems. *The Lancet Respiratory Medicine* 2020.
16. Angelini E, Dahan S, Shah A. Unravelling machine learning: insights in respiratory medicine. *European Respiratory Journal* 2019; 54 (6).
17. Wang S, Zha Y, Li W, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *European Respiratory Journal* 2020; in press (<https://doi.org/10.1183/13993003.00775-2020>).
18. Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, and U. Rajendra Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, *Computers in Biology and Medicine*, Vol. 121, (103792), June 2020.
19. Seung Hoon Yoo, and et al, Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging, *Front Med (Lausanne)*. 2020; 7: 427.
20. Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, Sahlol AT (2020) New machine learning method for image-based diagnosis of COVID-19. *PLoS ONE* 15(6): e0235187. <https://doi.org/https://doi.org/10.1371/journal.pone.0235187>
21. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *IEEE CVPR* 2017.
22. Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", *Mendeley Data*, V2, doi: <https://doi.org/10.17632/rscbjbr9sj.2>
23. Joseph Paul Cohen and Paul Morrison and Lan Dao and Karsten Roth and Tim Q Duong and Marzyeh Ghassemi, COVID-19 Image Data Collection: Prospective Predictions Are the Future, arXiv:2006.11988, <https://github.com/ieee8023/covid-chestxray-dataset>, 2020
24. Zhao Jinyu, Zhang Yichen, He Xuehai, Xie, Pengtao, COVID-CT-Dataset: a CT scan dataset about COVID-19, arXiv preprint arXiv:2003.13865, 2020
25. M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, M. T. Islam, "Can AI help in screening Viral and COVID-19 pneumonia?" *IEEE Access*, Vol. 8, 2020, pp. 132665 – 132676.

GRU-Based Parameter-Efficient Epileptic Seizure Detection



Ojas A. Ramwala, Chirag N. Paunwala, and Mita C. Paunwala

1 Introduction

Epilepsy is a neurological (central nervous system) disorder that affects the brain. It causes loss of awareness and is characterized by recurrent unprovoked seizures caused by abnormal brain cell activity. It is one of the most common chronic brain diseases, with over 50 million people [1] suffering from it worldwide. It can affect people of any gender and age; however, 80% of them develop epileptic symptoms in childhood and adolescence [2]. Moreover, the risk of premature death in people who have epilepsy is three times higher than in the general population [3]. According to WHO [4], 70% of epileptic patients can be seizure-free if appropriately detected, properly diagnosed, and treated.

The electroencephalogram (EEG) signals are usually utilized to diagnose epilepsy since providing a conclusive diagnosis without EEG analysis is considered to be unfeasible. EEG waveform of a healthy and epileptic person recorded for 100 timesteps at a sampling rate of 173.61 Hz is shown in Fig. 1a, b, respectively.

The manual visual inspection of EEG signals is a strenuous, cumbersome, and enduring process. Automation of the EEG analysis process can reduce errors and diagnosis time and improve neurologists' ability to administer medications. Thus,

O. A. Ramwala (✉)

Electronics Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat, India

C. N. Paunwala

Electronics and Communication Engineering Department, Sarvajanic College of Engineering and Technology, Surat, India

M. C. Paunwala

Electronics and Communication Engineering Department, C. K. Pithawala College Engineering and Technology, Surat, India

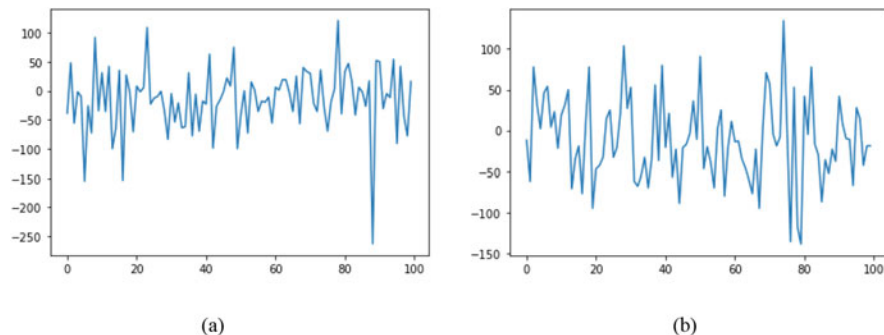


Fig. 1 EEG waveform of (a) healthy person and (b) epileptic patient

this work is intended to reduce the burden on the medical and paramedical fraternity by proposing a deep learning architecture for accurate epileptic seizure detection.

Furthermore, the proposed architecture is parametric efficient and computationally inexpensive and hence can be deployed on low-power computing and lightweight devices, including NVIDIA's Jetson Nano, to develop a low-cost solution for detecting epileptic seizures in the EEG recordings of susceptible patients.

2 Related Work

Several attempts have been made to develop accurate epileptic seizure detection methods. The time-domain-based analysis is one of the most popular epileptic seizure detection methods. Detection of seizures in the time domain requires the histogram-based analysis of the discrete-time sequences. Tracing the consecutive maxima-minima and histogram estimation was proposed [5] for SVM-based classification [6] to detect neonatal seizures in the patients. However, it was limited to only one patient. A body sensor network (BSN) was also developed to detect epileptic seizures based on statistics like variance, entropy, dynamic time warping (DTW)-based auto-correlation with template signals, mean, and zero-crossing rate extracted from time-domain signals.

Frequency domain techniques based on the Fourier Transform phase and magnitude have also been proposed for EEG seizure detection. A phase-slope index [7] to compute the interaction between channels of a multi-channel EEG signal has been utilized to detect seizures. A patient-specific seizure detection method by utilizing frequency-moment [8] signatures has been developed, where moments of the spectra are utilized as features to distinguish between normal and seizure activities.

To tackle the difficulty in characterizing EEG signals' different activities due to their nonlinearity and non-stationarity, a method based on four entropy features [9] for classification: phase entropy (S1 and S2), approximate entropy (ApEn),

and sample entropy (SampEn) was proposed. The extracted features were then fed to seven different classifiers: SVM, Fuzzy Sugeno Classifier (FSC) [10], Probabilistic Neural Network (PNN) [11], K Nearest Neighbours (KNN) [12], Decision Tree (DT) [13], and Gaussian Mixture Model (GMM) [14]. The Fuzzy classifier showcased improved performance.

Wavelets have also been employed for epileptic seizure detection. The fundamental concept behind utilizing the wavelet analysis for EEG seizure detection is extracting discriminating features from appropriate sub-bands. However, determining the appropriate wavelet decomposition level is a significant challenge. Bayesian Linear Discriminant Analysis (BLDA) [15] was proposed as a wavelet-based seizure detection method that depended on fluctuation index and lacunarity, a measure of the homogeneity in the fractal analysis as features. BLDA focuses on reducing the risk associated with the classification decision. Five-level wavelet decomposition methods [16, 17] for seizure detection have also been proposed. The extracted features like energy, relative amplitude, standard deviation, coefficient of variation, entropy, and fluctuation index are then fed to the SVM classifier.

Wavelet Neural Networks (WNNs) [18] have also been implemented by estimating the wavelet transform of EEG signals and extracting the minimum, maximum, and standard deviation of the absolute values of the wavelet coefficients in each sub-band as features that are fed to the trained WNNs to differentiate seizure activities from normal activities. Nevertheless, determining the appropriate wavelet decomposition level and selecting the features from certain sub-bands is a significant challenge in wavelet-based EEG seizure detection.

Several machine learning techniques have been explored for automated seizure detection [19]. Local Binary Pattern (LBP) has also been utilized to classify seizure and seizure-free EEG signals. The utilization of the K-NN-based classifier for seizure detection [20] has also been proposed. However, several LBP-based methods [21, 22] compute LBP at every sample value of the EEG signal. This disadvantage was addressed by detecting a set of stable key points [23] through multiscale EEG analysis and computing LBP at those points only. Though, simplifying the process of detection of key points can enhance the computational performance of the method. Artificial Neural Networks [24] have been combined [25] with Principal Component Analysis (PCA) [26] for the diagnosis of epilepsy. Logistic Regression combined with Artificial Neural Networks has also been developed for epileptic seizure detection using Multi-Layer Perceptron Neural Network [27].

Deep Learning techniques have also been implemented for epileptic seizure detection. Two-Dimensional Convolutional Neural Networks (2D-CNN) have been deployed wherein the one-dimensional (1D) EEG signals are initially transformed into two dimensions by utilizing several visualization methods. A 2D-CNN model based on extracting the temporal and spectral characteristics [28] of EEG signals to learn the seizure's overall structure had been proposed. SeizureNet [29], a deep learning framework based on Convolutional Neural Network and Dense Connections, has also been deployed for epileptic seizure detection. Temporal Graph Convolutional Network (TGCN) [30] has also been introduced for detecting epileptic seizures.

Research on deploying Recurrent Neural Networks for effective epileptic seizure detection is also being done. LSTM- and GRU-based architectures are being developed to detect epileptic seizures from EEG signals. Two LSTM architectures with 3 and 4 layers [31] together with the softmax classifier have been implemented to get satisfactory results. And 5-layer [32] and 3-layer [33] GRU-based deep learning architectures were also implemented to achieve good results for epilepsy detection. Similarly, a 4-layer GRU-based epileptic seizure detection system [34] was also proposed, which involved splitting the input signals into time windows as a pre-processing step.

To achieve accurate epileptic seizure detection, this chapter proposes a Gated Recurrent Unit-based deep learning architecture that circumvents the requirement of any feature-extraction steps and implements a computationally inexpensive model that can be deployed even on a resource-constrained embedded platform.

3 Proposed Method

Efficient epileptic seizure detection necessitates a dedicated deep learning architecture that can be deployed for an accurate diagnosis. This section is intended to formulate the network designed and implemented for reliable epileptic seizure detection. Concepts behind Recurrent Neural Networks have been showcased with mathematical explanation. Gated Recurrent Units that can circumvent the requirement of discrete memory cells to regulate the passage of information in the unit have been elucidated. Explanation of the implemented loss function has also been presented in the subsequent subsections.

3.1 Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are considered to be an augmentation of traditional feedforward neural networks. The activation at each time of the recurrent hidden state of RNNs is dependent on the preceding activation, which explains their ability to handle variable-length sequences. For a sequence, $i = [i_1, i_2, i_3, i_4, \dots, i_N]$, the recurrent hidden state s_n of the RNN is updated, as shown by Eq. 1, where the nonlinear function is denoted by Υ , a composition of the logistic sigmoid with affine transform.

$$s_n = \begin{cases} 0, & \text{if } n = 0 \\ \Upsilon(s_{n-1}, i_n), & \text{else} \end{cases} \quad (1)$$

RNNs can have a variable-length output: $o = [o_1, o_2, o_3, o_4, \dots, o_n]$. For the smooth bounded function f and parameter matrices and vector W , U , and B , the

recurrent hidden state can be updated, as shown in Eq. 2.

$$s_n = f(Wi_n + Us_{n-1}) \quad (2)$$

The probability distribution over the next element is the output of a generative RNN. The conditional probability distribution is modeled, as shown in Eq. 3.

$$p(i_n | i_1, i_2, i_3, i_4, \dots, i_{n-1}) = f(s_n) \quad (3)$$

However, due to the frequently faced vanishing gradients and seldom encountered, albeit severely impactful, exploding gradients issues, gradient-based optimization becomes challenging, and training Recurrent Neural Networks turns out to be complicated. To address this problem, utilizing clipped gradients [35–37] had been proposed; however, due to its dependency on the identical growth pattern between the second- and first-order derivatives, it does not yield a guaranteed [38] solution. Thus, developing an elaborate activation function by utilizing the affine transform and element-wise nonlinearity becomes essential. Gating units have been developed to cater to the solution.

LSTM [39, 40] (Long Short-Term Memory) units comprising of an input gate, an output gate, and a forget gate determine if the existing memory through the input gates is to be kept or not, rather than content overwriting at each timestep. With the internal memory state present in the LSTM, it can hold the previous information that the network has identified before. However, due to the possibility of developing architecture with better computational efficiency with minimal data, Gated Recurrent Units have been utilized in the proposed method.

3.2 Gated Recurrent Units

To develop recurrent units that flexibly comprehend reliances of various time scales, Gated Recurrent Units (GRU) were proposed [41]. To regulate the passage of information inside the unit without having discrete memory cells, gating units are provided in GRUs. GRUs utilize Update Gate and Reset Gate to resolve the issue of vanishing gradients, as shown in Fig. 2.

GRUs devote specialized mechanisms to determine when the hidden state must be updated and when it should be reset. These gates are engineered to be vectors that perform convex combinations with inputs in binary: [0, 1], designed and trained to sustain information.

3.2.1 Reset Gate

The Reset Gate R_n regulates the amount of past information from the preceding time stamps that are to be neglected. By considering the logistic sigmoid function as σ ,

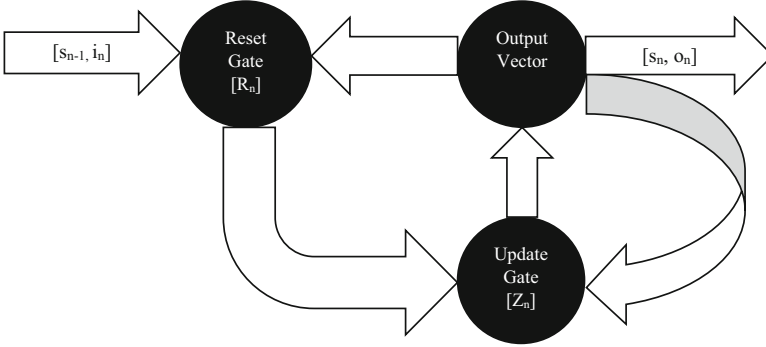


Fig. 2 A Gated Recurrent Unit

the Reset Gate R_n can be calculated by considering the parameters W_{ir} , W_{sr} and B_r , as shown in Eq. 4.

$$R_n = \sigma(i_n W_{ir} + s_{n-1} W_{sr} + B_r) \quad (4)$$

Nonlinearity in the form of \tanh is introduced to constrain the values of the hidden states in the range of $(-1, 1)$ and make the input mean equal to zero. To reduce the impact of the previous states, s_{n-1} is multiplied element-wise with R_n . Whenever the Reset Gate R_n is approximately equal to 1, a traditional RNN is recovered, and with R_n approaching 0, a Multi-Layer Perceptron is enabled. Thus, a previously extant hidden state is set to the default value.

$$\check{s}_n = \tanh(W_z i_n + U_z (R_n \odot s_{n-1}) + B_s) \quad (5)$$

The candidate recurrent hidden state \check{s}_n , which has yet not been incorporated with the action of the update gate, can thus be calculated, as shown in Eq. 5, where \odot denotes element-wise multiplication. The Reset Gate thus recognizes short-term dependencies in the time series.

3.2.2 Update Gate

The Update Gate regulates the amount of past information from the preceding time stamps that are to be transferred. To address the issue of vanishing gradients, the model can continue with all the past information without considering the elimination of any details. By considering the logistic sigmoid function as σ , the Update Gate Z_n can be calculated by considering the parameters W_{iz} , W_{sz} , and B_z , as shown in Eq. 6.

$$Z_n = \sigma(i_n W_{iz} + s_{n-1} W_{sz} + B_z) \quad (6)$$

For the Update Gate Z_n , the degree to which the current recurrent hidden state s_n resembles the previous hidden state s_{n-1} can be determined by performing convex combinations by utilizing the new candidate state \check{s}_n , as shown in Eq. 7.

$$s_n = Z_n \odot s_{n-1} + (1 - Z_n) \odot \check{s}_n \quad (7)$$

As per Eq. 4, whenever the Update Gate Z_n is approximately equal to 1, the previous state is retained, while information from i_n is not considered, and, thus, skipping time step n . Also, whenever the Update Gate Z_n is approximately equal to 0, the current recurrent hidden state s_n proceeds toward the candidate hidden state \check{s}_n . The Update Gate thus recognizes long-term dependencies in the time series.

3.3 Model Architecture

The proposed method implements a Gated Recurrent Unit-based deep learning architecture. The architecture comprises four Gated Recurrent Unit layers and two dense layers. As depicted by the flowchart, the number of nodes in each layer gets reduced by 50 percent, beginning with 512 to 64. The dense layers perform the task of classifying the EEG signals from the features extracted by the GRU layers and provide the output as 1 or 0 for epileptic seizure present or absent, respectively, as shown in Fig. 3.

3.4 Loss Function

The proposed architecture for epileptic seizure detection utilizes the Binary Cross-Entropy Loss, also known as Log Loss. Considering y as the label (1 or 0 for epileptic seizure present or absent, respectively), and $p(y)$ as the predicted probability for the particular class, the loss function can be denoted as shown in Eq. 8.

$$L = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)) \quad (8)$$

The deployed loss function is also known as Sigmoid Cross-Entropy Loss. It is a Sigmoid activation plus a Cross-Entropy Loss. Considering $f()$ as the Sigmoid function and CE as the Cross-Entropy Loss for binary problems, the loss function can be expressed as:

$$f(s_i) = \frac{1}{1 + e^{-s_i}} \quad CE = -t_i \log(f(s_i)) - (1 - t_i) \log(1 - f(s_i))$$

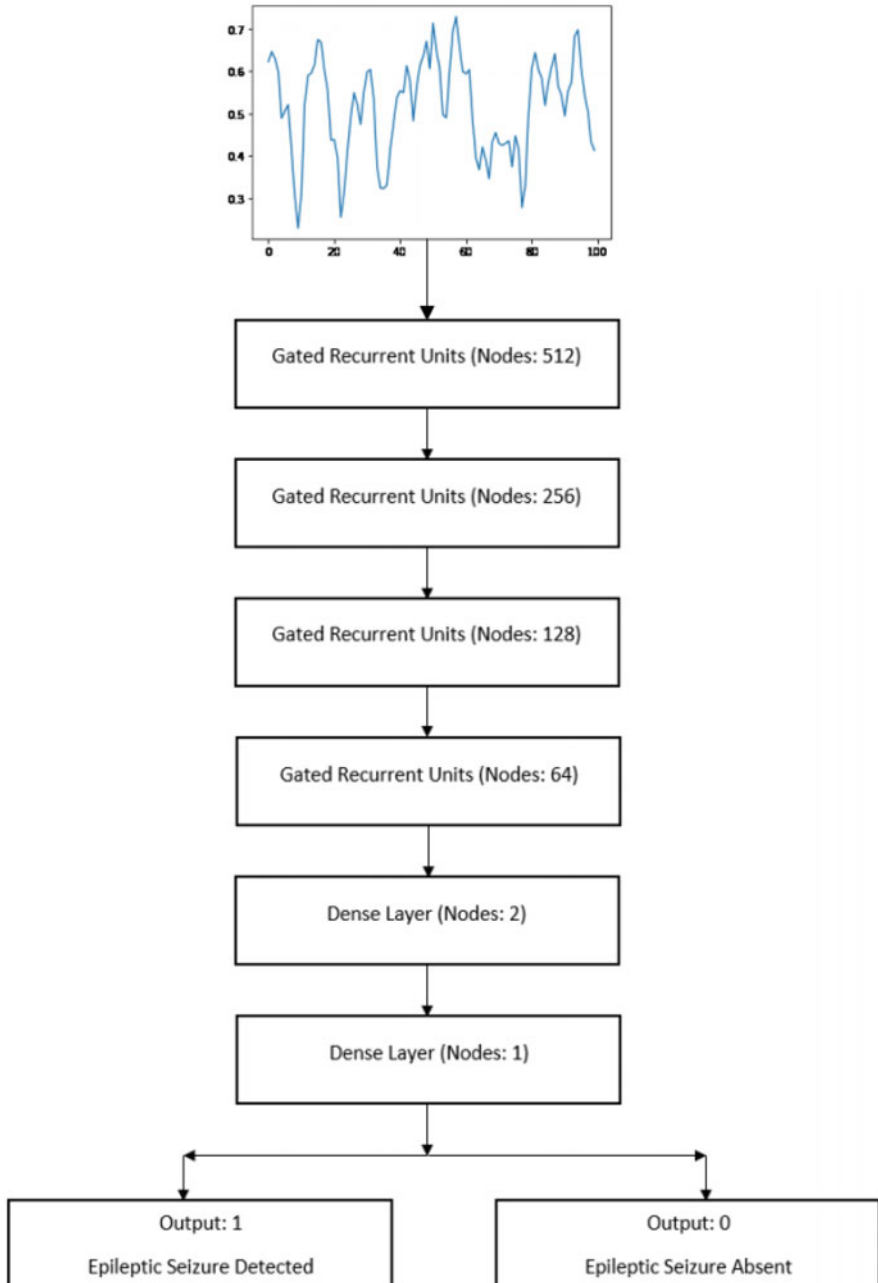


Fig. 3 Proposed architecture for epileptic seizure detection



The Cross-Entropy Loss can be understood as shown in Eq. 9:

$$CE = \begin{cases} -\log(f(s_1)) & \text{if } t_1 = 1 \\ -\log(1 - f(s_1)) & \text{if } t_1 = 0 \end{cases} \quad (9)$$

The gradient concerning the score $s_i = s_1$ can be expressed as shown in Eq. 10:

$$\frac{\partial CE(f(s_i))}{\partial s_i} = t_i (f(s_i) - 1) + (1 - t_i) f(s_i) \quad (10)$$

For $f()$ being a Sigmoid function, the expression can be written as shown in Eq. 11:

$$\frac{\partial CE(f(s_i))}{\partial s_i} = \begin{cases} f(s_i) - 1 & \text{if } t_i = 1 \\ f(s_i) & \text{if } t_i = 0 \end{cases} \quad (11)$$

4 Experiments and Results

This section mentions the dataset utilized to train the proposed deep learning architecture for accurate epileptic seizure detection. Training details have been elaborated along with the implementation of the model on resource-constrained hardware, and the performance of the proposed method by demonstrating several evaluation metrics has been showcased.

4.1 Dataset and the Training Details

The proposed Gated Recurrent Unit-based deep learning architecture must be trained to classify the EEG signals accurately. The proposed model has been trained on the University of Bonn EEG Database [42], which has been sampled at the rate of 173.6 Hz.

The model has been trained, validated, and tested on 320, 40, and 40 samples, respectively. The proposed architecture has been trained with the following parameters on NVIDIA's Tesla P100 [43] GPU having 16 GB RAM:

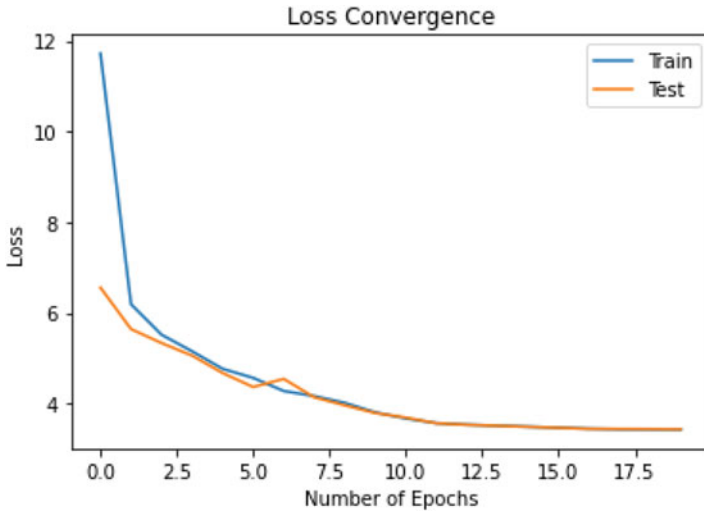


Fig. 4 Convergence of the training and validation loss (best viewed in color)

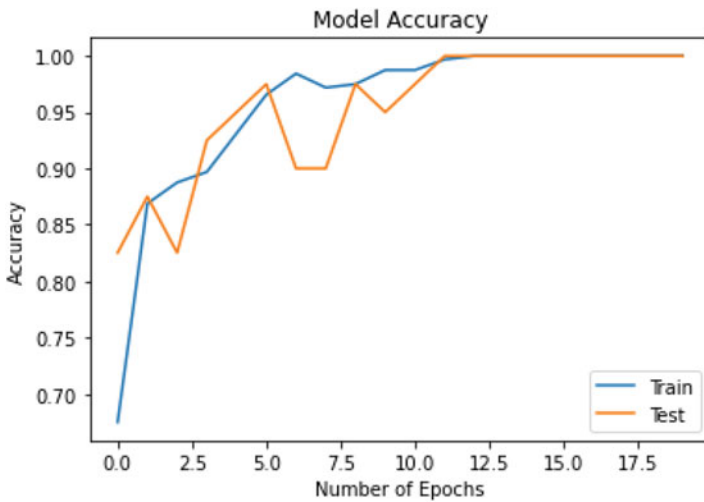


Fig. 5 Improvement in the training and validation accuracy (best viewed in color)

- Epochs: 20.
- Optimizer: NADAM [42–47].
- Batch Size: 4.
- Learning Rate: 0.001.

The convergence of training and validation loss is demonstrated in Fig. 4, and the improvement in the training and validation accuracy is demonstrated in Fig. 5.

Table 1 Confusion matrix

	Normal	Positive
Predicted normal	19 (true positive)	00 (false positive)
Predicted positive	01 (false positive)	20 (true negative)

Table 2 Evaluation metrics: performance of the model

Evaluation metrics	Value
Accuracy	0.9750
Sensitivity	0.9500
Specificity	1.0000
Positive predictive value	1.0000
Negative predictive value	0.9524
False positive rate	0.0000
False negative rate	0.0500
F1 score	0.9744
Matthews correlation coefficient	0.9512

4.2 Hardware Implementation

NVIDIA Jetson Nano [48] is a power-efficient and cost-effective embedded development kit that is extremely useful for inferencing deep learning models. The development board has 4GB of system-wide memory utilized by the CPU and GPU. There are 128 CUDA-supported cores in the GPU.

Since the proposed deep learning architecture has only 1.7 million parameters, it was possible to perform the model’s inference on the resource-constrained embedded platform. The computationally inexpensive Gated Recurrent Unit–based deep learning architecture showcased the real-time performance on NVIDIA Jetson Nano.

4.3 Evaluation Metrics: Performance of the Proposed Architecture

It is imperative to carefully examine the deep learning models deployed for medical diagnosis to ensure a reliable decision-making process. A comprehensive evaluation of the proposed GRU-based architecture has been performed by considering several evaluation metrics. During training, the model with the least validation loss was saved and utilized for testing.

The Confusion Matrix summarizes the model’s prediction results as shown in Table 1, wherein Normal and Positive indicate the absence and presence of seizure, respectively.

A holistic understanding of the proposed method’s performance can be gained by a careful observation of Table 2, which summarizes the values of all the evaluation metrics and highlights that the model yields reliable performance for accurate epileptic seizure detection.

5 Conclusion

This chapter proposes a computationally inexpensive deep learning architecture for efficient epileptic seizure detection to reduce the burden on neurologists by automating the cumbersome process of manually analyzing the EEG signal to detect epilepsy in susceptible patients. A Gated Recurrent Unit-based network has been developed for accurate epileptic seizure detection. A detailed mathematical explanation of the implemented binary cross-entropy loss function and the NADAM optimizer has been presented. The proposed model has been comprehensively evaluated using several metrics by considering all the possible scenarios, including false positive and false negative rates. Experiments demonstrate the reliability of the proposed method and showcase the architecture's parametric efficiency by deploying it on the resource-constrained NVIDIA Jetson Nano embedded platform. A future direction of this work could be studying and implementing other loss functions in place of the binary cross-entropy loss.

References

1. <https://www.who.int/news-room/fact-sheets/detail/epilepsy> (Last Accessed: 17-02-2021)
2. S. Macleod and R. E. Appleton, "Neurological disorders presenting mainly in adolescence," *Archives of disease in childhood* vol. 92, pp. 170–175, 2007.
3. <https://www.who.int/news-room/fact-sheets/detail/epilepsy> (Last Accessed: 17-02-2021)
4. <https://www.who.int/health-topics/epilepsy> (Last Accessed: 17-02-2021)
5. T. P. Runarsson and S. Sigurdsson, "On-line Detection of Patient-Specific Neonatal Seizures using Support Vector Machines and Half-Wave Attribute Histograms," *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, Vienna, 2005, pp. 673–677, <https://doi.org/10.1109/CIMCA.2005.1631546>.
6. Cortes, C. & Vapnik, V., Support-vector networks. *Machine learning*, 1995, pp. 273–297.
7. P Rana, J Lipor, H Lee, WV Drongelen, MH Kohrman, BV Veen, Seizure detection using the phase-slope index, and multi-channel ECoG. *IEEE Trans. Biomed. Eng.* 59(4), 1125–1134 (2012)
8. H Khamis, A Mohamed, S Simpson, Frequency–moment signatures: a method for automated seizure detection from scalp EEG. *Clin. Neurophysiol.* 124(12), 2317–2327 (2013)
9. UR Acharya, F Molinari, SV Sree, S Chattopadhyay, KH Ng, JS Suri, Automated diagnosis of epileptic EEG using entropies. *Biomed. Signal. Process. Control.* 7(4), 401–408 (2012)
10. Z. Deng, L. Cao, Y. Jiang, and S. Wang, "Minimax Probability TSK Fuzzy System Classifier: A More Transparent and Highly Interpretable Classification Model," in *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 4, pp. 813–826, Aug. 2015, <https://doi.org/10.1109/TFUZZ.2014.2328014>.
11. Ancona, F., Colla, A.M., Rovetta, S. et al. Implementing Probabilistic Neural Networks. *Neural Comput & Applic* 5, 152–159 (1997) <https://doi.org/10.1007/BF01413860>.
12. Mucherino A., Papajorgji P.J., Pardalos P.M. (2009) k-Nearest Neighbor Classification. In: *Data Mining in Agriculture*. Springer Optimization and Its Applications, vol 34. Springer, New York, NY], Naive Bayes Classifier (NBC) [Webb G.I. (2011) Naïve Bayes. In: Sammut C., Webb G.I. (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA.

13. Fürnkranz J. (2011) Decision Tree. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA.
14. Reynolds D. (2009) Gaussian Mixture Models. In: Li S.Z., Jain A. (eds) Encyclopedia of Biometrics. Springer, Boston, MA.
15. Y Zhou, Y Liu, Q Yuan, X Li, Epileptic seizure detection using lacunarity and Bayesian linear discriminant analysis in intracranial EEG. *IEEE Trans. Biomed. Eng.* 60(12), 3375–3381 (2013).
16. Y Liu, W Zhou, Q Yuan, S Chen, Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20(6), 749–755 (2012).
17. R Panda, PS Khobragade, PD Jambhule, SN Jengthe, PR Pal, TK Gandhi, Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction, in Proceedings of International Conference on Systems in Medicine and Biology (Kharagpur), pp. 405–408. 16–18 Dec 2010.
18. Z Zainuddin, LK Huong, O Pauline, On the use of wavelet neural networks in the task of epileptic seizure detection from electroencephalography signals. *Proc Comput. Sci.* 11(2012), 149–159 (2012).
19. U. Acharya, S. Sree, G. Swapna, R. Martis, and J. Suri, “Automated EEG analysis of epilepsy: A review,” *Knowledge-Based Systems*, vol. 45, pp. 147–165, 2013.
20. T. Kumar, V. Kanhangad, and R. Pachori, “Classification of seizure and seizure-free EEG signals using local binary patterns,” *Biomedical Signal Processing and Control*, vol. 15, pp. 33–40, 2015.
21. T. Kumar, V. Kanhangad, and R. Pachori, “Classification of seizure and seizure-free EEG signals using local binary patterns,” *Biomedical Signal Processing and Control*, vol. 15, pp. 33–40, 2015.
22. Y. Kaya, M. Uyar, R. Tekin, and S. Yldrm, “1D-local binary pattern-based feature extraction for the classification of epileptic EEG signals,” *Appl. Math. Comput.*, vol. 243, pp. 209–219, 2014.
23. A. K. Tiwari, R. B. Pachori, V. Kanhangad, and B. K. Panigrahi, “Automated Diagnosis of Epilepsy Using Key-Point-Based Local Binary Pattern of EEG Signals,” in *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 4, pp. 888–896, July 2017, <https://doi.org/10.1109/JBHI.2016.2589971>.
24. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics.* 1943;5(4):115–33.
25. S. Ghosh-Dastidar, H. Adeli, and N. Dadmehr, “Principal Component Analysis-Enhanced Cosine Radial Basis Function Neural Network for Robust Epilepsy and Seizure Detection,” in *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 2, pp. 512–518, Feb. 2008, <https://doi.org/10.1109/TBME.2007.905490>.
26. Jolliffe I. (2011) Principal Component Analysis. In: Lovric M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_455
27. Alkan A, Koklukaya E, Subasi A. Automatic seizure detection in EEG using Logistic Regression and artificial neural network. *J Neurosci Methods.* 2005; 148(2): 167–176. <https://doi.org/10.1016/j.jneumeth.2005.04.009>
28. M. S. Hossain, S. U. Amin, M. Alsulaiman, and G. Muhammad, “Applying deep learning for epilepsy seizure detection and brain mapping visualization,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 15, no. 1s, pp. 1–17, 2019.
29. U. Asif, S. Roy, J. Tang, and S. Harrer, “SeizureNet: a deep convolutional neural network for accurate seizure type classification and seizure detection,” *arXiv preprint arXiv:1903.03232*, 2019.
30. I. Covert, B. Krishnan, I. Najm, J. Zhan, M. Shore, J. Hixson, and M. J. Po, “Temporal graph convolutional networks for automatic seizure detection,” *arXiv preprint arXiv:1905.01375*, 2019.

31. M. Golmohammadi, S. Ziyabari, V. Shah, S. L. de Diego, I. Obeid, and J. Picone, "Deep architectures for automated seizure detection in scalp eegs," arXiv:1712.09776, 2017
32. S. Roy, I. Kiral-Kornek, and S. Harrer, "Chrononet: a deep recurrent neural network for abnormal EEG identification," in Conference on Artificial Intelligence in Medicine in Europe. Springer, 2019, pp. 47–56.
33. X. Chen, J. Ji, T. Ji, and P. Li, "Cost-sensitive deep active learning for epileptic seizure detection," in Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, 2018, pp. 226–235.
34. S. S. Talathi, "Deep recurrent neural networks for seizure detection and early seizure detection systems," arXiv preprint arXiv:1706.03283, 2017.
35. Y. Bengio, N. Boulanger-Lewandowski, and R. Pascanu. Advances in optimizing recurrent networks. In Proc. ICASSP 38, 2013.
36. J. Martens, and I. Sutskever. Learning recurrent neural networks with Hessian-free optimization. In Proc. ICML'2011. ACM, 2011.
37. R. Pascanu, T. Mikolov, and Y. Bengio. On the difficulty of training recurrent neural networks. In Proceedings of the 30th International Conference on Machine Learning (ICML'13). ACM, 2013.
38. Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In NIPS Workshop on Deep Learning.
39. Sepp Hochreiter; Jürgen Schmidhuber (1997). "Long Short-Term Memory." *Neural Computation*. 9(8): 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>. PMID 9377276.
40. A. Graves. Generating sequences with recurrent neural networks. arXiv:1308.0850, 2013
41. K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259, 2014.
42. Andrzejak RG, Lehnertz K, Rieke C, Mormann F, David P, Elger CE (2001) Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, *Phys. Rev. E*, 64, 061907
43. <https://www.nvidia.com/en-us/data-center/tesla-p100/> (Last accessed: 17-02-2021)
44. Timothy Dozat, "Incorporating Nesterov Momentum into Adam," ICLR Workshop(1):2013–2016, 2016.
45. Ojas A. Ramwala, Smeet A. Dhakecha, Chirag N. Paunwala, Mita C. Paunwala. "Reminiscent Net: Conditional GAN-based Old Image De-Creasing," *International Journal of Image and Graphics*, 2021. <https://doi.org/10.1142/S0219467821500509>
46. O. A. Ramwala, S. A. Dhakecha, A. Ganjoo, D. Visiya and J. N. Sarvaiya, "Leveraging Adversarial Training for Efficient Retinal Vessel Segmentation," 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2021, pp. 1–6, <https://doi.org/10.1109/ECAI52376.2021.9515093>.
47. O. A. Ramwala, P. Dalal, Parikh, Dalal, M. Paunwala, and C. Paunwala, " Novel Multi-Modal Throat Inflammation and Chest Radiography based Early-Diagnosis and Mass-Screening of COVID-19," *The Open Biomedical Engineering Journal*. <https://doi.org/10.2174/1874120702115010226>
48. <https://developer.nvidia.com/embedded/jetson-nano-developer-kit> (Last Accessed 17-02-2021)

An Object Aware Hybrid U-Net for Breast Tumour Annotation



Suvidha Tripathi and Satish Kumar Singh

1 Introduction

Accurate segmentation of biological structures and micro-structures visualized in digitized biopsy images could assist pathologists in measuring the disease extent. It could also help biomedical researchers around the world by automatically annotating the huge amount of medical images. In recent times, with the arrival of deep learning techniques, the methods for image analysis have advanced rapidly, more so with the easy availability of vast amounts of data such as ImageNet, Cifar100, COCO, etc. for various applications. The architectures trained on such huge amounts of dataset a benchmark for similar applications. However, no state-of-the-art deep learning model except U-Net [29] has been proposed for biomedical image analysis because of less data volume in the medical domain, high variability among datasets, and various modalities to address. The huge variations and less annotated data prevent the generalization of deep learning models. In such a scenario, most of the models proposed in this domain use pre-trained deep architectures or modify existing ones to suit their application. For segmentation problem, U-Net has been used extensively as base architecture [3, 7, 11, 14, 20, 26, 27, 31, 35]. U-Net is basically a semantic segmentation architecture. Semantic segmentation classifies each pixel in a target region as a class or non-class pixel. It is a supervised learning model whose performance depends on how well the ground truth is annotated. Ground truth annotation, however, in the case of multi-structural histopathological images is a tedious task. This is where U-Net gives the advantage as it is known to produce better results even with limited datasets. The U-Net

S. Tripathi (✉) · S. K. Singh
Department of Information Technology, Indian Institute of Information Technology Allahabad,
Prayagraj, Uttar Pradesh, India
e-mail: sk.singh@iiita.ac.in

architecture follows the basic encoder-decoder structure with a contracting path on the left and expanding path on the right. The unique difference between encoder-decoder architecture and U-Net makes the U-Net network more robust with a limited dataset. Besides contracting-expanding modules, U-Net preserves the feature lost during the contracting path by concatenating them with the expanding feature map in the expanding path of the model. The additional features preserve the quality of the segmentation.

The dataset used for our work comprises Whole Slide Images with three annotated tumour classes- benign, invasive, and *In situ*. The classes are very roughly annotated by experts to indicate the presence of the disease. The problem aggravates in the case of invasive carcinoma where it is difficult to draw a boundary to contain the class. High-grade invasive carcinoma could spread across the WSI and does not have an epithelial layer boundary to contain the malignant cells. In such cases, accurate segmentation becomes a challenge. Hence, we aim to develop a rough segmentation framework to indicate the presence of the tumour in the region. Rough segmentation of complex tumour regions is an open research problem. The pathologists themselves do not extensively annotate cells and nuclei to detect a cancerous region, instead they annotated a rough boundary around the suspected region to mark the presence of the disease. These rough boundaries often also contain the cluster of small tumours that are spread across the slide and cannot be bounded separately for each cluster. Hence, such clusters are generally annotated as a single region bounded by a rough annotation. The region of interest thus also, often, contain non-tumorous portions that are found in-between or around the clusters. In such cases, semantic or instance segmentation models often fail to precisely separate objects from the background. Hence, we need to develop models that could know about the possible object location before segmentation.

Therefore, auto-initialized active contours which are initialized using constrained criteria could be useful for such tasks. Active contours [9] are known to predict high-level object shapes by finding the possible boundary of the object depending on both the image features and priors such as length and curvature of the contour and other forces that drive the contour towards the edge of the object. These local priors are selected depending upon the application. Active contours find a minimum energy fit for incrementing contour vertices to the object edges. The energy is defined in such a way that the contours are attracted towards the minima or where the boundaries of the object lie. Active contours have an advantage because they are topology-aware and could also work on high-level image features acquired at low resolution. In our proposed method, due to space constraints, high-resolution WSI regions were downsampled to low-resolution small and even dimension images for input into modified U-Net architecture.

In the proposed work, we tried to amalgamate U-net and active contours to develop an object aware segmentation network for segmenting breast tumour images belonging to three different image categories. We modified the U-Net network using deep network tools such as ResNet [16] and DenseNet [17] blocks to enhance the efficiency of the segmentation. The U-net base acts as a learning framework for active contour priors that are responsible for the length and curvature of the

contours. The priors are learned while the contour moves with each epoch. The active learning model generates polygons close to ground truth instance. The network is inspired by the original work [22] that uses custom CNN and Active Contour Model (ACM). They used structured prediction for optimizing ACM parameters and SSVM (structured SVM) loss for finding optimal parameters. We tested our method using the Intersection over Union (IoU) metric with the original article along with other benchmark methods on the breast tumour dataset. Our method outperformed all the methods used for a similar task.

We have presented a preliminary analysis on the proposed method and a more detailed analysis is our future work. The main contributions of the proposed work are:

- We have applied the state-of-the-art deep structured active contour model on a challenging medical dataset to imitate pathologist annotations for tumours in the breast histopathology dataset.
- For the task we introduced semantic segmentation model U-Net enhanced by ResNet blocks to actively learn local information priors for active contour inference.
- The active contours are initialized through the robust automatic initialization method introduced in the chapter.
- The successful implementation of the method and the comparison with contemporary state-of-the-art methods highlights the importance of integrating traditional methods with deep learning methods for better results.

The rest of the chapter is organized as follows. Section 2 familiarizes readers with state-of-the-art literature for histopathological image segmentation. Followed by Sect. 2, Sect. 3 explains the methodology of the proposed hybrid network with subsections that build the theoretical concepts necessary for understanding the modules of the whole network. Section 4 species the experimental setup required to implement the model followed by results highlighting the comparison with recent benchmark models. Section 5 discusses the complete model with detailed discussions on challenges posed by the dataset and how our model might help to overcome some of those challenges. The chapter ends after Sect. 6 that briefly concludes the findings from the proposed model for breast tumour annotation.

2 Related Work

In digital pathology, tissue-wise labelled data is limited because it is very time consuming and requires expert pathologists. Hence, segmentation outputs are noisy and affect classification performance if the framework is end-to-end. Inconsistency in data acquisition methods also makes a limited amount of data even more useless. Due to these limitations, not much work has been done recently for segmenting tissue-level regions. The authors in [24] have done a similar work in which they have taken four classes of breast biopsy tumours, namely benign, atypia, DCIS, and

invasive. Their work divides WSIs into instances for feeding them into their network for joint segmentation and classification task. The output of their model produces an instance-level segmentation mask and instance-level probability map. The combined discriminative segmentation mask from the two outputs is then used to extract frequency and co-occurrence features which were then fed into MLP for final cancer diagnosis. The strength of their work is that they have used general UNeT [29] architecture for their specific task using simple modifications like adding instance-level probability map to enhance the features of segmentation mask that helped in improving classification accuracy of the final diagnosis. Their dataset was heavily annotated with tissue-level annotation done by 87 pathologists along with extensive substructures annotation by a pathology fellow. Exhaustive annotation is one of their most important strengths that aided in producing less noisy segmentation masks. This also helped them to create an end-to-end learning framework for both of their tasks. However, this is also the main drawback that without the heavy annotations their method would not work. The similar BACH dataset which we have used in the proposed work when tested on their algorithm failed to produce comparable results. Needless to say that other medical data segmentation algorithms like UNeT [29], SegNet [6], FCN [21] also failed to perform well on our dataset due to the same limitation. These pixel classification based segmentation methods have a fundamental limitation when the target object comprises many heterogeneous components. For example, a benign tumour at a low resolution not only consists of nuclei, but structures like papillary, solid, haemorrhagic, and sclerotic growth patterns are also visible. They are surrounded by well-formed one-two layers of epithelium cells and a fibrous sheath of connective tissue. At higher magnification, one can see solid areas composing stromal cells, round nuclei with fine chromatin and rare nucleoli [1]. Such varied structures in one tumour cannot be individually annotated for semantic segmentation and hence treating them as one structure as a whole poses great confusion for such algorithms and therefore, fail to produce good results. The problem with instance segmentation algorithms like MaskRCNN [15] is that it requires a complete object to be present in the image for segmentation and classification. Region Proposal network of MaskRCNN compares the object characteristics as a whole, with a certain threshold, with the learned instances to propose a probable bounding box. Whereas in the case of medical histology images, object characteristics vary widely within classes, and it gets very difficult to learn all types of object priors for smooth detection and classification. Other recent detection and classification methods like FastRCNN [13] and [28] also works on the same theory of instance-level detection and classification. Other works that are similar to ours are the segmentation based method in [23] and saliency map-based method in [12]. Mehta et al. [23] developed a CNN-based method for segmenting breast biopsy images that produces a tissue-level segmentation mask for each WSI. The histogram features they extracted from the segmentation masks were used for diagnostic classification. Geccer et al. [12] proposed a saliency-based method for diagnosing cancer in breast biopsy images that identified relevant regions in breast biopsy WSIs to be used for diagnostic classification.

Due to the rough annotation, our work focuses not only on segmenting the tumour mask from the background but also tried to draw a contour around the mask for better object boundary visualization. Similar works of literature in the past have termed such tasks as contour-aware segmentations [10, 18]. Classically, active contours have been extensively used in histopathology images for segmenting nuclei and cells [2, 33, 34]. But, using ACM with CNN and for larger tissue regions like glands remains under-explored. Recently, Xu et al. [32] segmented nuclei from breast biopsy histopathological images that use CNN for nuclei detection and the detected nuclei act as initialization for active contour-based ellipse fitting over the detected nuclei. Khvostikov et al. [19] trained the CNN model for learning active contour priors for gland segmentation. They also proposed a collision resolution algorithm as a post-processing step to separate overlapping gland objects. The dataset used by them was carefully annotated with crisp gland boundaries. However, in our case, since the annotations are rough and there are no crisp boundaries for tumours, the task of segmentation becomes even more challenging. Therefore, in the proposed work, we have performed a preliminary analysis of our model on the dataset and compared it with recent segmentation benchmarks.

3 Methodology

3.1 Overview

The proposed model segments the roughly annotated breast tumour masks using the hybrid U-Net Active Contour model. The backbone U-net model is modified by adding ResNet blocks. The detailed modified U-Net model is illustrated in Fig. 1. The output map of each upsampling layer is concatenated to produce a concatenated feature map. This feature map is then further processed with convolution layers to produce four active contour priors. Each of the feature prior is then used to calculate active contour energy terms. The active contour energy function equation (refer Eq. 1) has four local priors, i.e., values that weigh the contour energy terms on a per-pixel basis. Hence, the priors as calculated as feature maps which are dynamically learned during backpropagation in an end-to-end model. The block diagram of the proposed model is shown in Fig. 1. We have modified the methodology used in DSAC [22]. The difference is that their method uses CNN to learn Active Contour Model external and internal energy terms [9] while we have used modified U-net architecture to get the deep learning inference (Sect. 3.2). Furthermore, the strategy to use automatic multiple initializations has been expanded so that the initial contour can find the object even if it is not centred (refer to Sect. 3.4). The modified training algorithm is expressed in Algorithm 1.

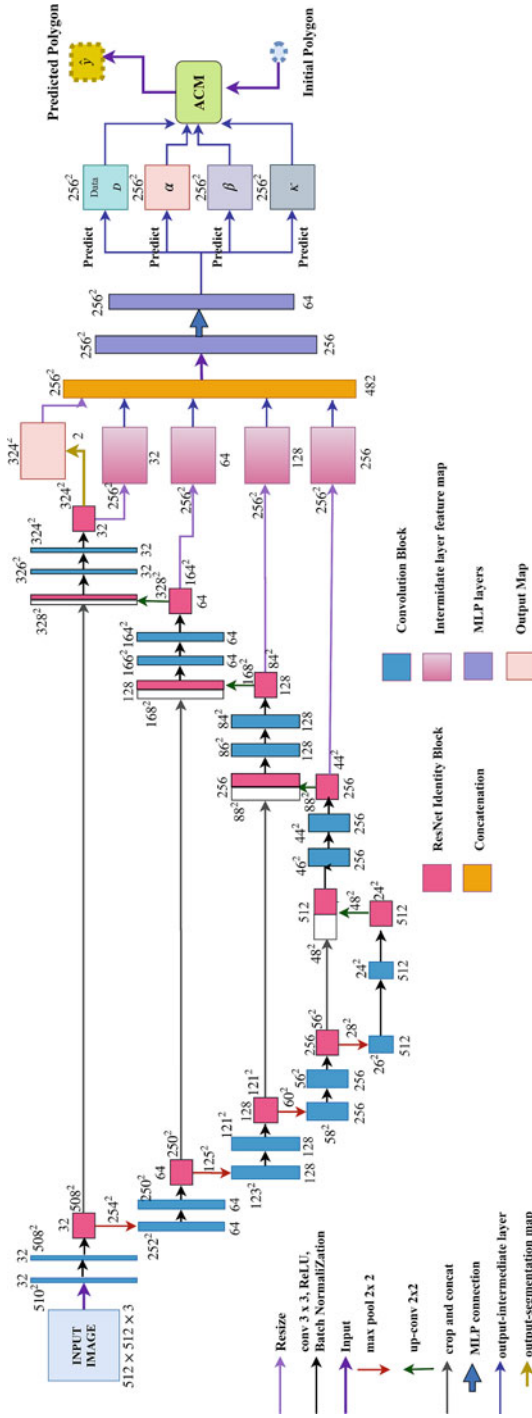


Fig. 1 Block diagram of proposed hybrid segmentation model

Algorithm 1 proposed model training algorithm. The deep learning backbone forwards the feature maps learned to ACM inference at every iteration. Then ACM inference is made for each initialized polygon. The five polygons yield five polygons. A structured loss IoU is calculated for all the five polygons. After the training is complete, the maximum IoU is calculated to yield the final polygon

Data : $\leftarrow X, Y$: image/polygon pairs in the training set
 $\mathbf{y}^0 \leftarrow$ corresponding polygon initializations
for $x_i, y_i \in X, Y$ **do**
 UNet-ResNet inference: $D, \alpha, \beta, \kappa \leftarrow CNN_w(x_i)$
 ACM inference:
 for $j = 1 : n$ **do**
 $\hat{y}_i^j \leftarrow ACM(D, \alpha, \beta, \kappa, y_i^j)$
 $\frac{\partial L^j}{\partial D}, \frac{\partial L^j}{\partial \alpha}, \frac{\partial L^j}{\partial \beta}, \frac{\partial L^j}{\partial \kappa} \leftarrow \hat{y}_i^j, y_i^j$
 Compute $\frac{\partial L^j}{\partial \omega}$ (combined loss) using backpropagation
 Update UNet-ResNet: $\omega \leftarrow \omega - \eta(\frac{\partial L^j}{\partial \omega})$
 Calculate IoU^j
 Determine max IoU and corresponding index

3.2 Modified U-Net Architecture

Original U-net architecture [29] has an encoding and a decoding branch comprising a stack of convolution and deconvolution blocks, respectively. The encoder branch learns input representations while downsampling the input image, whereas the decoder branch recovers the spatial resolution lost during downsampling. The spatial information lost due to the downsampling of the input is added back at the upsampling layer using the skip connections. These skip connections are made between corresponding layers of encoder and decoder branch. We have added ResNet identity blocks after the convolution blocks in each layer on both encoder and decoder branches. The ResNet blocks further help recover spatial information loss in the whole model. In the decoder branch, the output feature maps from each layer are then resized to the output size (256×256) and concatenated to produce the final output feature map. This feature map is then further passed through two-layer MLP with 256 and 64 hidden units to predict four local information priors or weight maps: Data $D(x)$, $\alpha(x)$, $\beta(x)$, and $\kappa(x)$, where x is the input image; $x \in X$. The active contour inference and the prediction of local priors are followed as in the literature [29].

3.3 Active Contour

An active contour [9] is a line or a continuous set of points that move over the image to find the point of minima. In other words, each point in a contour moves around the image so that the energy function is minimized. An active contour can

be represented as a polygon $y = (u, v)$ with L nodes. Let each node s is represented by $y_s = (u_s, v_s)$ with $s \in 1, \dots, L$. The polygon y is then deformed such that the following energy function is minimized.

$$E(y) = \sum_{s=1}^L [D(y_s) + \alpha(y_s) \left| \frac{\Delta_s y_s}{\Delta_s} \right|^2 + \beta(y_s) \left| \frac{\Delta_s^2 y_s}{\Delta_s^2} \right|^2] + \sum_{u,v \in \Omega(y)} \kappa(u, v) \quad (1)$$

$D(y_s)$ is the external energy term indexed by the position $y_s = (u_s, v_s)$ and means the value in function $D(x)$, where x is the input image, where $D(x) \in \mathfrak{R}^{U \times V}$ of size $U \times V$ is the data term, depending on input image x , $x \in \mathfrak{R}^{U \times V \times d}$ and $U \times V \times d$ is the image width, height, and depth, respectively. Both $\alpha(y_s)$ and $\beta(y_s)$ are weights associated with feature maps of dimension $U \times V$ extracted during the CNN training, same as $D(y_s)$. The terms associated with α and β are first-order and second-order derivative of the polygon at y_s defining length and curvature terms. $\sum_{u,v \in \Omega(y)} : \Omega(y)$ is the notation to represent the pixels enclosed by the nodes of polygon y . $\sum_{u,v \in \Omega(y)} \kappa(u, v)$ is the summation of the pixel values of the kappa feature map enclosed within the polygon. This defines kappa energy. We have modified the methodology used in [22]. Their method uses custom CNN to learn Active Contour Model external and internal energy terms (refer to Eq. 1). They used the structured loss to train CNN (explained in Sect. 3.5.1). The complete details of their method, including active contour inference and experimental setup, could be found in their paper [22].

3.4 Automatic Multiple Initialization

We have modified the initialization of active contours by introducing multiple initial contours. There is a total of five contours initialized at four corners and one centre of the image. Each initial contour moves at each active contour inference iteration towards the minimum energy location. At the end of active contour iterations, the IoU over all the five predicted contours is calculated with the ground truth. The maximum IoU we obtain is then projected as the final predicted contour. This modification has been proposed to enhance the detection of objects that are slightly shifted towards the boundary of the image window rather than at the centre. This is explained through the illustration given in Fig. 2.

With multiple initializations of snake contour across the image window, there would be more chances of finding an optimal tumour boundary and minimize the convergence of snake at local maxima. This method is simple and robust for datasets with a single object per image.

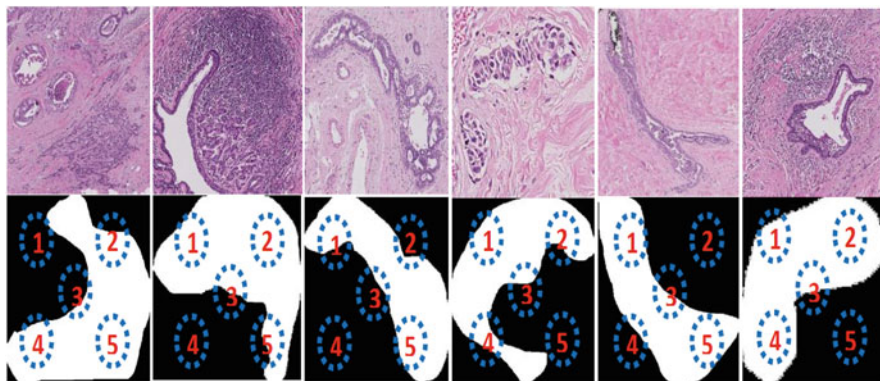


Fig. 2 Representation of modification in DSAC algorithm - Initial Polygon Selection Criterion: The first row comprise the labelled regions and the second row comprise their tumour mask. From left to right read as region label and the first polygon number with which we get the optimal final polygon, respectively: 1. Invasive, 5th, 2. Benign, 2nd, 3. In situ, 5th, 4. Invasive, 1st, 5. Benign, 3rd, 6. Benign, 2nd or 4th

3.5 U-Net Training with Structured SVM Loss

In this method, energy terms are learned on the training sets instead of taking them as constants. Also, the structured loss used in the method is more suited for such complex datasets where both the target domain and the loss are more or less arbitrary. This means that the goal is not a simple target like a label or a number, but possibly a much more complicated object [25]. A non-trivial task of segmenting tumour boundaries is a suitable problem for structured prediction. Here the target mask differs significantly in local features such as size, shape, intensity, colour, and texture. If viewed as a pure segmentation problem, we could see that each possible snake iteration in the training set is provoked by varied values of internal and external forces (i.e., without a uniform pattern or range). This has several drawbacks when the loss is defined by a particular function like softmax, tanh, etc. Therefore, the structured loss which considers the output (the segmentation mask) as a whole and not a set of arbitrary snake points is a more preferable choice. Moreover, structured prediction enables to conform relationship among multiple output variables (snake points) into a model (one output). **Terms:**

X —Image space

Y —Output space

y^i —Positive ground truth polygon corresponding to i th sample image x^i

y —All the negative outputs where $y \neq y^i$ generated after every Active contour inference

\hat{y} —Predicted polygon

\hat{y}^i —Predicted polygon for i th sample

$\Delta(y, \hat{y}) = \frac{y \cap \hat{y}}{y \cup \hat{y}}$ —Task loss function or IoU

Energy $E(y)$ corresponding to output polygon y

Given a collection of ground truth pairs $(y^i, x^i) \in Y \times X, i = 1, 2, \dots, N$, and a task loss function $\Delta(y, \hat{y})$ where $y \in Y \wedge y \neq \hat{y}$, we would like to find CNN parameters ω such that by optimizing Eq. (1) and thus obtaining the inference result for i th sample:

$$\hat{y}^i = \arg \min_{y \in Y} E(y; \omega) \quad (2)$$

one could expect a small $\Delta(y^i, \hat{y}^i)$. The problem becomes:

$$\hat{\omega} = \arg \min_{\omega} \sum_i \Delta(y^i, \arg \min_{y \in Y} E(y; \omega)) \quad (3)$$

3.5.1 Structured SVM Loss

Since $\Delta(y^i, \hat{y}^i)$ could be a discontinuous function, this loss can be substituted by a continuous and convex function such as HINGE LOSS.

$$l(y^i; \omega) = \max(0, \Delta(y^i, y) + E(y^i; \omega) - E(y; \omega)) \quad (4)$$

$$l(y^i; \omega) = \max(0, \max_{y \in Y} (\Delta(y^i, y) + E(y^i; \omega) - E(y; \omega))) \quad (5)$$

In Eq. (7) the energy $E(y; \omega)$ corresponding to output y decreases with every *omega* update such that the difference between energy corresponding to the ground truth $E(y^i; \omega)$ and $E(y; \omega)$ is minimized. And, max over output space Y is taken to maximize the margin between two energies such that when the Energy $E(y; \omega)$ decreases the task loss $\Delta(y^i, y)$ increases.

Now adding l_2 regularization and summing up for all training samples, hinge loss becomes the MAX-MARGIN FORMULATION which is our objective function:

$$L(Y; \omega) = \frac{1}{2} \|\omega\|^2 + C \sum_i \max(0, \max_{y \in Y} (\Delta(y^i, y) + E(y^i; \omega) - E(y; \omega))) \quad (6)$$

where PREDICTION FUNCTION is defined as:

$$\hat{y}^i = \arg \max_{y \in Y} (\Delta(y^i, y) + E(y^i; \omega) - E(y; \omega)) \quad (7)$$

If the constant (energy corresponding to ground truth $E(y^i; \omega)$) is dropped from the above equation, we obtain

$$\hat{y}^i = \arg \max_{y \in Y} (\Delta(y^i, y) - E(y; \omega)) \quad (8)$$

Once we obtain \hat{y}^i then the Objective with stochastic approx. for randomly chosen data point i becomes:

$$L(Y; \omega) = \frac{1}{2} \|\omega\|^2 + C \max(0, \Delta(y^i, \hat{y}^i) + E(y^i; \omega) - E(\hat{y}^i; \omega)) \quad (9)$$

Since $L(Y; \omega)$ is not differentiable, gradients cannot be calculated. Hence, we compute subgradient as:

$$\frac{\partial L(Y; \omega)}{\partial \omega} = \omega + C \frac{\partial}{\partial \omega} \max(0, \Delta(y^i, \hat{y}^i) + E(y^i; \omega) - E(\hat{y}^i; \omega)) \quad (10)$$

where

$$\begin{aligned} & \frac{\partial}{\partial \omega} (\max(0, \Delta(y^i, \hat{y}^i) - E(\hat{y}^i; \omega) + E(y^i; \omega))) \\ &= \begin{cases} \frac{\partial E(y^i; \omega)}{\partial \omega} - \frac{\partial E(\hat{y}^i; \omega)}{\partial \omega} & ; \text{if } E(y^i; \omega) - E(\hat{y}^i; \omega) < \Delta(y^i, \hat{y}^i) \\ 0 & ; \text{if } E(y^i; \omega) - E(\hat{y}^i; \omega) = \Delta(y^i, \hat{y}^i) \\ 0 & ; \text{if } E(y^i; \omega) - E(\hat{y}^i; \omega) > \Delta(y^i, \hat{y}^i) \end{cases} \quad (11) \end{aligned}$$

So, following is the subgradient with respect to ω :

$$\begin{cases} \text{if } E(y^i; \omega) - E(\hat{y}^i; \omega) < \Delta(y^i, \hat{y}^i) & ; \omega + C \left(\frac{\partial E(y^i; \omega)}{\partial \omega} - \frac{\partial E(\hat{y}^i; \omega)}{\partial \omega} \right) \\ \text{else} & ; \omega + 0 \end{cases} \quad (12)$$

Algorithm 2 describes our methods training for U-Net architecture.

4 Experimental Setup and Results

For the segmentation task, the dataset comprises the labelled regions from the WSIs which were resized to (512×512) for segmenting the suspected tumour. In this case, normal patches were not included in the segmentation dataset.

Algorithm 2 U-Net training

X - Image space
 Y - Output space
 y^i - positive ground truth polygon corresponding to i^{th} sample image x^i
 y - all the negative outputs where $y \neq y^i$ generated after every Active contour inference
 N - total number of samples in a batch.
 ω - U-Net parameter to be updated
 \hat{y} - predicted polygon
 \hat{y}^i - predicted polygon for i^{th} sample
 $\Delta(y, \hat{y}) = \frac{y \cap \hat{y}}{y \cup \hat{y}}$ - task loss function or IoU
input number of iterations T , step size η for $t = 1, \dots, T$
Energy $E(y; \omega)$ corresponding to output polygon y
Energy $E(\hat{y}^i; \omega)$ corresponding to predicted polygon \hat{y}^i for i^{th} sample
regularizer C
Initialize $\omega \leftarrow \mathbf{0}$
for $t = 1, \dots, T$ **do**
 for $i = 1, \dots, N$ **do**
 $\hat{y}^i \leftarrow \arg \max_{y \in Y} (\Delta(y^i, y) - E(y; \omega))$
 if $E(y^i; \omega) - E(\hat{y}^i; \omega) < \Delta(y^i, \hat{y}^i)$ **then**
 $v^i \leftarrow \frac{\partial E(y^i; \omega)}{\partial \omega} - \frac{\partial E(\hat{y}^i; \omega)}{\partial \omega}$
 else
 $v^i \leftarrow 0$
 $\omega \leftarrow \omega - \eta(\omega + \frac{C}{N} \sum_i v^i)$

4.1 Dataset Preparation and Usage

The ICIAR BACH 2018 challenge published a breast WSI tumour dataset [5]. The dataset is publicly available and has been first reported in [4]. Other publications using the dataset as benchmark are [30] and [8]. The dataset contains ten annotated WSIs for training. They did not, however, reveal the test annotation. Therefore, we have worked only on the tumour regions extracted from the ten WSIs. The WSI contains three annotated classes- Benign, Invasive, *In situ*. Using the annotation coordinates in ground truth files, we calculated the bounding box dimensions around each annotated tumour region. We then increased the bounding box dimensions of the annotated regions by 40% to increase the background area around the tumour mask. From the ten WSIs, a total of 56 Benign, 100 Invasive, and 60 *In situ* regions were extracted. Each extracted region was of arbitrary dimensions ranging between 20,000 pixels to 196 pixels across height and width. To reduce the computational complexity and make the images of even dimensions, all the 216 regions were resized to 512×512 with 3 RGB colour channels. Figure 3 shows the dataset samples with ground truth and corresponding pathologist annotation in each row, respectively. The polygonal annotation is expected to be achieved through active contour inference that moves towards the edge of the tumour while training on a ground truth mask. From the figure, we could see how arbitrary and heterogeneous shapes are roughly annotated by the pathologist for only detection purpose.

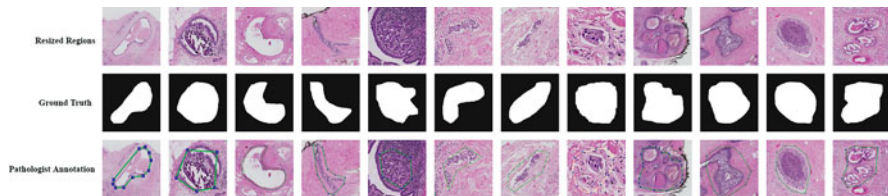


Fig. 3 WSI region samples with their corresponding ground truth mask generated through polygonal annotation by the pathologist [5].

4.2 Experiments

We did several experiments for preliminary analysis of our adopted method for breast tumour segmentation. The main method has been compared with the original DSAC method proposed in [22], FCN16, and SegNet. Some ablation studies have been done to show the variation in results with hyperparameter changes such as the number of layers, optimizer, and learning rate. In the main method, we used Adam Optimizer with a learning rate of 10^{-4} and five layers in an encoder and four layers in the decoder branch. We evaluated the performance of the method on IoU averaged over all test images. IoU is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. The dataset is randomly divided into 150 training images and 66 test images.

The dataset was tested with the original DSAC CNN backbone with and without multiple initializations. The average IoU obtained validated that with multiple initializations, the detection performance of active contour has increased. We then introduced ResNet blocks in the original CNN backbone to test whether there is an improvement with ResNet identity blocks. We observed that the IoU has increased from 59.98% to 61.32% with ResNet blocks in the original CNN backbone architecture followed in original DSAC. The observed results strengthen the choice of including ResNet blocks and multiple initializations in our framework. Further, we replaced the CNN model with U-Net and ResNet blocks with multiple initializations as our final proposed model. The proposed model is tested with semantic segmentation networks like SegNet, FCN16, original U-Net. The original U-Net was then further enhanced with ResNet and DenseNet blocks, respectively. Table 1 shows a comparison between different models with our proposed model. From the observed IoU, we could deduce that the choice of deep learning backbone affects the final performance of the active contour inference over the image. When we compared the results of semantic segmentation models with Active Contour enhanced hybrid semantic models, except the U-Net + ResNet model, we observed incremental improvement with hybrid approaches. The results hence strengthen the idea of annotating medical datasets with such models to imitate pathologist like annotations instead of using semantic segmentation models which are not usually useful in clinical settings. Figure 4 shows the results obtained after semantic

Table 1 Comparative performance evaluation over average IoU on the test set of ICIAR BACH 2018 dataset

Model	Method	Average IoU(%)
Semantic	FCN16	51.72
	SegNet	71.65
	UNet	44.69
	UNet+DenseNet	51.08
	U-Net+Resnet	77.13
Hybrid models	Original DSAC	56.09
	Original DSAC with multiple initializations	58.62
	Original DSAC with multiple initializations and ResBlock	60.07
	U-Net-ResNet-ACM (ours)	76.45

segmentation of dataset test images using state-of-the-art semantic models. Further, the results obtained from hybrid segmentation models is illustrated in Fig. 5.

5 Discussion

Through this work, we aimed to apply the deep learning trained active contour segmentation on a complex histopathology breast tumour dataset. The dataset is roughly annotated by the pathologists for marking tumour regions. The dataset is not explicitly annotated for segmentation purpose. This makes the tasks more challenging. Hence, the base model is enhanced by introducing ResNet blocks for recovering information loss during the downsampling and upsampling operations in the network. The final segmentation results, as shown, prove the sensitivity of the network with ResNet blocks. For marking the boundary of the tumour with a polygon just like the pathologist does, the active contour algorithm with locally learned priors is added to the segmentation model. The active contour is moved to detect the boundaries of the tumour using the strong local priors learned by the U-Net-ResNet deep learning model. Multiple automatic initializations proved to be the critical factor to improve the detection performance of the algorithm.

We started experiments with the new state of the art semantic and instance segmentation algorithms. The results in Table 1 show the comparison of semantic segmentation methods like UNeT, FCN, and SegNet with our modified model application on the histopathology dataset. We could see that pixel-wise classification with pixels as segmenting unit did not work well with our dataset and our intuition about heterogeneity causing misclassification of such a small unit like a pixel is correct. The problem of heterogeneity within objects, their shape, size, coarseness differ so widely in histopathology images that segmentation of coarsely annotated regions of interest like tumours causes either over-segmentation or under segmentation of pixels. For example, in the case of Ductal Carcinoma In Situ

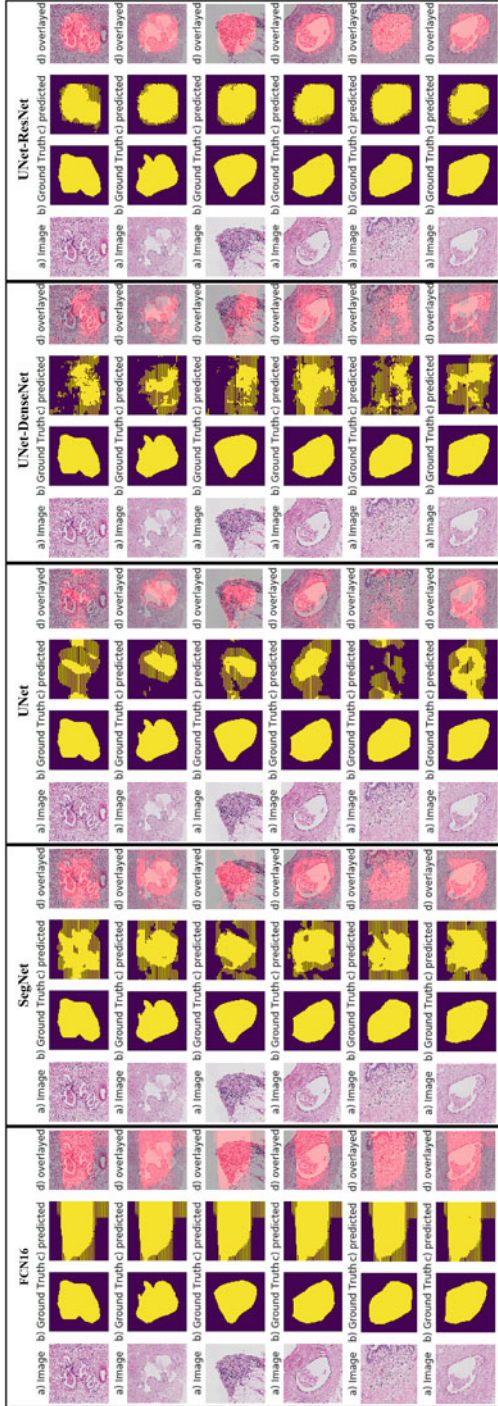


Fig. 4 Comparative semantic segmentation output of different models. The first column in each model shows the original image, the second column shows the ground truth, the third column illustrates the segmented output, and the fourth column shows the overlaid output on original image

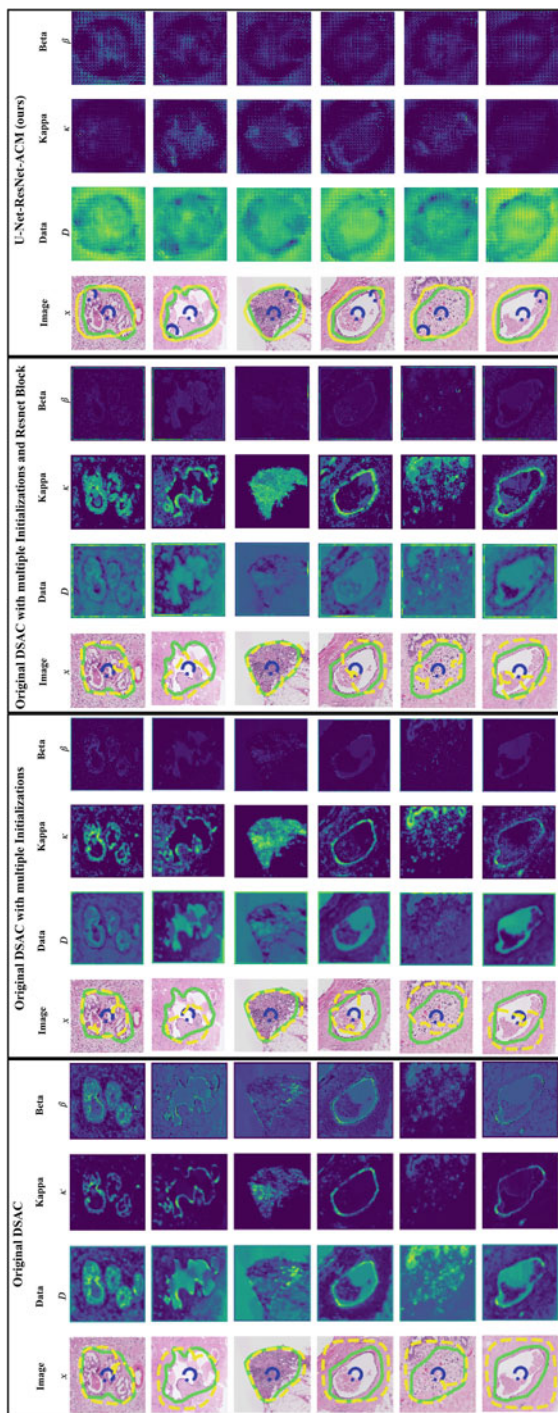


Fig. 5 Comparative hybrid segmentation output of different models. In the first column of each model, the initial contours are in blue and the obtained result in yellow, with the ground truth in green colour. The second column shows the Data term $D(x)$, where the boundary can be seen as a region with lower energy. The third column is the balloon terms κ which highlight the region within the boundary and the fourth column shows the β feature map where we notice the curvature areas in the original image are penalized

(DCIS) of the breast, there are at least five subtypes, namely DCIS: micropapillary, DCIS: Cribriform, DCIS: Cribriform with microcalcifications, and DCIS: Apocrine and DCIS: Comedonecrosis. All these subtypes have different shapes, distribution of nuclei and the presence of substructures. So, without subtype annotations, a computer algorithm treats every subtype as a different class. If the matching instance is not present within that slide, the algorithm fails to recognize the instance of the DCIS tumour. Therefore, without extensive manual annotation of each substructure, semantic and instance segmentation models largely fail in such scenarios. Our method, with active contour inference, increase object awareness through active feature learning and contour displacement within an end-to-end network. Thus, it made the overall network more sensitive to object features. And with it, we achieved our aim of polygonal annotations just as the pathologist would mark for cancer detection in histopathological images.

The proposed work is, however, only tested for the histopathological domain of medical images and need to test for other types of image modalities such as CT, MRI, and Ultrasound images. This poses a constraint for the model's implementation for the analysis of a larger group of modalities. Hence, this work can be extended to other medical image modalities for broader implementation and better integration with the clinical framework. Moreover, our work does not allow the active intervention of pathologists which adds another limitation. In the field of pathological analysis using CAD methods, regular involvement of pathologist is a must to improve the learning of the ML system. Thus, there is a scope to include an active learning framework within the model for improved and reliable performance.

6 Conclusion

We have applied the state-of-the-art deep structured active contour model on a medical dataset to imitate pathologist annotations for tumours in the breast histopathology dataset. For the task, we introduced semantic segmentation model U-Net enhanced by ResNet blocks to learn local information priors for active contour inference actively. The initial active contours acted as object identifiers which helped to improve the network performance for heterogeneous data. The future work would be to enhance the segmentation performance at the WSI level so that the pathologist like annotations could be done for both medical and educational purposes.

Acknowledgments This research was carried out in the Indian Institute of Information Technology, Allahabad and supported by the Ministry of Human Resource and Development, Government of India. We would also like to acknowledge the support and guidance of Dr Hwee Kuan Lee, Principal Investigator, A*STAR Bioinformatics Institute, Singapore. We are also grateful to the NVIDIA corporation for supporting our research in this area by granting us TitanX (PASCAL) GPU.

References

1. Ackerman, L.V., Rosai, J.: The pathology of tumors, part one: introduction, precancerous lesions, benign lesions that resemble cancer. *CA: A Cancer Journal for Clinicians* **21**(3), 162–173 (1971)
2. Ali, S., Madabhushi, A.: An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE transactions on medical imaging* **31**(7), 1448–1460 (2012)
3. Alom, M.Z., Yakopcic, C., Taha, T.M., Asari, V.K.: Nuclei segmentation with recurrent residual convolutional neural networks based u-net (r2u-net). In: *NAECON 2018-IEEE National Aerospace and Electronics Conference*, pp. 228–233. IEEE (2018)
4. Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S.S., Safwan, M., Alex, V., Marami, B., Prastawa, M., Chan, M., Donovan, M., et al.: Bach: Grand challenge on breast cancer histology images. *Medical image analysis* (2019)
5. BACH: [dataset] breast cancer histology images (BACH). <https://iciar2018-challenge.grand-challenge.org/Home/> (2018)
6. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* **39**(12), 2481–2495 (2017)
7. BenTaieb, A., Hamarneh, G.: Topology aware fully convolutional networks for histology gland segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 460–468. Springer (2016)
8. Carvalho, E.D., et al.: Breast cancer diagnosis from histopathological images using textural features and CBIR. *Artificial Intelligence in Medicine* **105**, 101845 (2020). DOI <https://doi.org/10.1016/j.artmed.2020.101845>
9. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Transactions on image processing* **10**(2), 266–277 (2001)
10. Chen, H., Qi, X., Yu, L., Heng, P.A.: Dcan: deep contour-aware networks for accurate gland segmentation. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 2487–2496 (2016)
11. Cui, Y., Zhang, G., Liu, Z., Xiong, Z., Hu, J.: A deep learning algorithm for one-step contour aware nuclei segmentation of histopathological images. *arXiv preprint arXiv:1803.02786* (2018)
12. Geçer, B.: Detection and classification of breast cancer in whole slide histopathology images using deep convolutional networks. *Diss. Bilkent University* **1** (2016)
13. Girshick, R.: Fast r-cnn. In: *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448 (2015)
14. Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.A., Snead, D., Tsang, Y.W., Rajpoot, N.: Mild-net: minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical image analysis* **52**, 199–211 (2019)
15. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969 (2017)
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778 (2016)
17. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708 (2017)
18. Kainz, P., Pfeiffer, M., Urschler, M.: Semantic segmentation of colon glands with deep convolutional neural networks and total variation segmentation. *arXiv preprint arXiv:1511.06919* (2015)
19. Khvostikov, A., Krylov, A., Mikhailov, I., Malkov, P.: Trainable active contour model for histological image segmentation. *Scientific Visualization* **11**(3) (2019)

20. Li, J., Sarma, K.V., Ho, K.C., Gertych, A., Knudsen, B.S., Arnold, C.W.: A multi-scale u-net for semantic segmentation of histological images from radical prostatectomies. In: AMIA Annual Symposium Proceedings, vol. 2017, p. 1140. American Medical Informatics Association (2017)
21. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440 (2015)
22. Marcos, D., Tuia, D., Kellenberger, B., Zhang, L., Bai, M., Liao, R., Urtaşun, R.: Learning deep structured active contours end-to-end. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8877–8885 (2018)
23. Mehta, S., Mercan, E., Bartlett, J., Weaver, D., Elmore, J., Shapiro, L.: Learning to segment breast biopsy whole slide images. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 663–672. IEEE (2018)
24. Mehta, S., Mercan, E., Bartlett, J., Weaver, D., Elmore, J.G., Shapiro, L.: Y-net: joint segmentation and classification for diagnosis of breast biopsy images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 893–901. Springer (2018)
25. Nowozin, S., Lampert, C.H., et al.: Structured learning and prediction in computer vision. *Foundations and Trends® in Computer Graphics and Vision* **6**(3–4), 185–365 (2011)
26. Oda, H., Roth, H.R., Chiba, K., Sokolić, J., Kitasaka, T., Oda, M., Hinoki, A., Uchida, H., Schnabel, J.A., Mori, K.: BESNet: boundary-enhanced segmentation of cells in histopathological images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 228–236. Springer (2018)
27. Qu, H., Riedlinger, G., Wu, P., Huang, Q., Yi, J., De, S., Metaxas, D.: Joint segmentation and fine-grained classification of nuclei in histopathology images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 900–904. IEEE (2019)
28. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, pp. 91–99 (2015)
29. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241. Springer (2015)
30. Roy, K., Banik, D., Bhattacharjee, D., Nasipuri, M.: Patch-based system for classification of breast histology images using deep learning. *Computerized Medical Imaging and Graphics* **71**, 90–103 (2019)
31. Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., et al.: Gland segmentation in colon histology images: The GLaS challenge contest. *Medical image analysis* **35**, 489–502 (2017)
32. Xu, J., Gong, L., Wang, G., Lu, C., Gilmore, H., Zhang, S., Madabhushi, A.: Convolutional neural network initialized active contour model with adaptive ellipse fitting for nuclear segmentation on breast histopathological images. *Journal of Medical Imaging* **6**(1), 017501 (2019)
33. Xu, J., Janowczyk, A., Chandran, S., Madabhushi, A.: A weighted mean shift, normalized cuts initialized color gradient based geodesic active contour model: applications to histopathology image segmentation. In: *Medical Imaging 2010: Image Processing*, vol. 7623, p. 76230Y. International Society for Optics and Photonics (2010)
34. Xu, J., Janowczyk, A., Chandran, S., Madabhushi, A.: A high-throughput active contour scheme for segmentation of histopathological imagery. *Medical image analysis* **15**(6), 851–862 (2011)
35. Zeng, Z., Xie, W., Zhang, Y., Lu, Y.: RIC-Unet: An improved neural network based on Unet for nuclei segmentation in histology images. *IEEE Access* **7**, 21420–21428 (2019)

VLSI Implementation of sEMG Based Classification for Muscle Activity Control



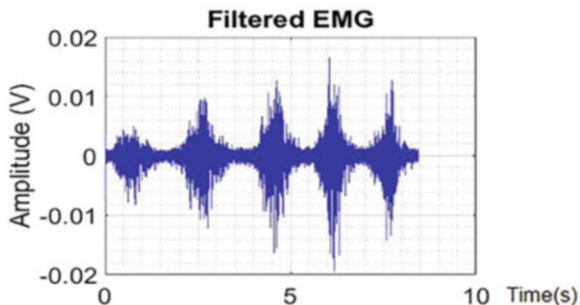
Amit M. Joshi, Natasha Singh, and Sri Teja

1 Introduction

Electromyography (EMG) is a signal which is coming from muscles and is also helpful in analysing and detecting of various activities of the body. EMG is very important bio-potential signal for the prosthesis based applications. Muscle activation takes place whenever a myoelectric signal is generated [1]. The electric activity generated in a human body is processed using electromyography [2].

The electric actuation is analysed for muscle activity using electromyography [3]. Electromyography is also known as myoelectric signal where the actuations are produced in the form of time variant signal [4]. It has very useful and significant information of neuromuscular activities. They are non-stationary, nonlinear and also complex signals [5]. EMG has information which is taken from the features of signal. There are several feature extraction techniques for accurate pattern recognition [6]. The brain is involved in the controlling of the movements of the muscles. Thus electrical activity from the muscle is observed closely for various activities. An action potential will be coming from the brain that passes through nerve fibres and they will stimulate the muscle fibres [7]. Electrical signals are transmitted the motor neurons which cause the muscles contraction for muscle movement [8]. Motion of humans is due to the integration of muscles central nervous system and brain. The effort of brain which is organised and controls 28 major muscles and trunk the limb joints to control the gravity and move the body forward with less energy consumption [9]. The body movements is due to the coordination of muscles with the brain. The muscle of the body performs a particular activity. It will send signals through the CNS [10]. Muscles are innervated in groups

A. M. Joshi (✉) · N. Singh · S. Teja
MNIT, Jaipur, India
e-mail: amjoshi.ece@mnit.ac.in; 2015uec1069@mnit.ac.in; 2015uec1360@mnit.ac.in

Fig. 1 Recorded EMG signal

called ‘Motor Units’. The junction point where the muscle fibres and motor neurons meet is a motor unit [11]. After the motor unit is activated a motor unit action potential is produced. The activities of CNS is performed continuously to generate required force. This action produces motor unit action potential trains which are pre impose for obtaining the EMG signal. A large number of muscles take part in the movement of human body. The total number depends on the activity of the body [12]. For example, in case of the weightlifting such as small stone which involves only few muscles compared to lifting a heavy weight like dumbbell’s. Generally to lift a greater weight the involvement of CNS increase. Hence it results in the increase in amplitudes of EMG signals. The general EMG signal has been shown in Fig. 1.

Electromyography is responsible for generating force, creating movements and allowing us for performing countless other activities for interacting with the world [13]. EMG is useful bio-potential signal has developed for large number of applications. Medically EMG is used as a diagnostic tool for diseases and disorders involving nerves and muscles. It is used mostly for treatments of patients with neuromuscular diseases, low back pain and few muscular diseases. EMG has been used in evaluating the applied research in physiotherapy, sports medicine and rehabilitation. There are some distinct requirements for rehabilitation applications. First, the system must be small and consume low power. Next, a large rehabilitation embedded with a computer device should provide communication with other systems and also providing more features like the possibility of recording data. This functionality have structures which are very diverse in nature and their usage is based on different software libraries. Apart all these applications involving different types of communication (motor control to actuate a device) which are based on software realisation and can be performed by CPU. There are various machine learning algorithms to perform real-time analysis of physiological data [14]. It is computationally very expensive based on amount of data and complexity of the algorithms. Therefore, it is real challenge to have accurate prediction of the movement for controlling the therapeutic device with least latency. The paper presents hardware implementation of classification approaches to have real-time pattern identification using EMG.

The organisation of remaining part of the chapter is as follows: Sect. 2 defines the basic overview of useful electrodes for EMG acquisition. Section 3 explains the EMG based Pattern recognition method for upper limb prosthetic control. Section 4 covers machine learning model for upper limb prosthetic control. Section 5 describes the basic theory of Linear and Quadratic Discriminant Analysis. Section 6 emphasises on VLSI Implementation of LDA and QDA along with their performance measurement.

2 EMG Data Acquisition

The bio electrical activity can be seen using EMG electrodes inside the human body muscle [15]. These electrodes are of three types as: (1) Inserted electrodes, (2) Fine wire electrodes, (3) Surface electrodes. Needle Electrodes and Fine wire Electrodes are of inserted electrodes category. All of these electrodes are explained in detail whereas surface EMG electrodes are used in our experimental analysis to acquire the data.

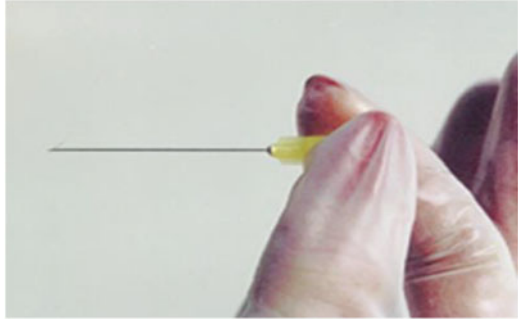
2.1 Needle Electrodes

The needle is used at the surface of the body for signal acquisition (as Fig. 2). The insulated wire is present in the cannula. The quality of the signal is improved by the needle electrodes by using it in conjunction with the other types of electrodes. Needle electrodes are used mainly in neuromuscular evaluations and surgical procedures. The two main advantages of needle electrodes are, (i) During low force

Fig. 2 Practical EMG needle electrode



Fig. 3 EMG fine wire electrode



contractions, it can be able to detect individual MUAPs because of its relatively small pickup area (ii) The electrode can be re-positioned after the insertion.

2.2 Fine Wire Electrodes

This type of electrodes are stiff wire of small diameter with non-oxidising property and insulation (Fig. 3). Ores of Pt, Ag, Ni, and Cr are mostly used. These are very thin and could be inserted in an easier manner. They can be removed from the muscles and cause less pain when compared with the needle electrodes.

2.3 Surface EMG Electrode

These type of electrodes enable a very good method for measuring and detecting the EMG signal which does not involve the introduction of materials into the human system. The current flow from human body into the electrode is done through electrolytic conduction [16]. This electrolytic conduction is made possible with the help of chemical equilibrium formed between human skin and electrodes (Fig. 4). These electrodes that are used for sEMG are quite easy to implement and are very simple. The usage of fine wire electrodes and needles electrodes need to be done under strict medical support and supervision whereas sEMG does not require such kind of observation. sEMG electrodes have used in studies of motor behaviour, recordings of neuromuscular activities, evaluating the performances in sports activities. Along with these the sEMG is mostly used for the detection of muscular activities and also control the device extension for achieving prosthesis for the persons with disabilities and amputees. There are some restrictions due to sEMG. These electrodes are generally applied on superficial muscles so there would be cross talk from the other muscles, this becomes a serious issue when measuring the EMG signal. The stability of their position with respect to the skin must be maintained properly else it results in signal distortion.

Fig. 4 Surface EMG electrode



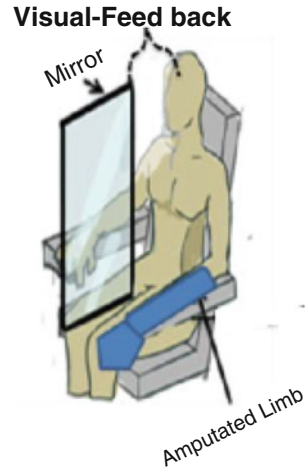
3 EMG Based Pattern Recognition for Upper Limb Prosthesis

EMG has a different and multiple applications in fields of clinical and medicine. It is also used in detecting various diseases involving muscles and nerves and also as an instrument in learning the kinesiology and detecting robotic disorders [17]. Electromyogram signals are used in controlling prosthetic limbs and prosthetic devices, for example, prosthetic hands, lower limbs and arms etc. [18], EMG is useful to measure the perfect functioning and conduction of nerves from the amputees [19].

EMG is recorded and filtered for controlling a prosthetic limb through hands (upper limb) and legs (lower limb) which is capable of interacting with brain and the CNS. It is known as Implantable Myoelectric sensor(IMES) [20]. A series of Implanted myoelectric sensors EMG signal is recorded. The amplitude of the EMG signal falls between the range of zero to ten milliVolts. The EMG signal is mixed by various types of noises when passing through various tissues. Observing the properties and the variations of these unwanted electric signals is very crucial and plays a major role in acquisition. The filtering technique is applied to remove various noises before further processing. These signals are then send to distant places where controller is used for analysing purpose. The allowed data is to be sent to prosthetic limb controller and is done as the system developed allows it and finally it is able to move the prosthetic limb towards the direction in which amputees intended. A USB cable can also be used in monitoring the data from the EMG signal or with the help of external computer connected to the IMES system. This way it can be used in controlling the movement of prosthetic limb using the commands which are coming from the brain and central nervous system. The data acquisition using visual feedback system is shown in Fig. 5.

For classification of hand gestures and movements sEMG signals are used in various studies and implemented successfully in the control of prosthetic limbs and hands for the amputates. For controlling a wearable devices that can assist persons

Fig. 5 Prosthetic limb control



with low muscle mass or the persons who are suffering from sarcopenia sEMG can be used potentially. In position control using sEMG we need to estimate the torque intended of the user in providing sufficient information for efficient control of force of the prosthetic hand or assistive device [21].

4 Machine Learning Models for Prosthetic Control Using EMG Signal

The functioning of the classification of the motion pattern recognition depends on the extraction of various features [22]. This has a major role in the activities recognition from sEMG signal. This method is a way of converting the original raw EMG signal to a vector. The various features of these signals can be categorised into three main domains, namely that of frequency, time and time-frequency features. The time domain are computed with the time-changing amplitude. During the observation process, the signal amplitude depends on types of the muscle and their conditions. Most of the analyses use time domain due to their low computational complexity. For the measurements of these signals no additional transformations are required. The PSD of signals is contained in frequency domain features and are calculated using a periodogram. Time-frequency features are the features that have the information of both time domain and frequency domain. A varying frequency data at various intervals of time can be obtained by these features which in turn results in non-stationary data. The present paper is based on time domain based features extraction method due to less computational burden and for better real-time performance.

4.1 Time Domain Based Feature Extraction

These features are utilised in various fields such as in medical and engineering practices. Time domain features are popular for the purpose of identification of any intention. Due to their quick and easy implementation these features are used for signal classification and these features does not require any transformation of the signal [23]. Most of these are based on raw EMG signal. The variation in statistical properties of EMG signal which is known as the Non-stationary property of EMG signal remained to be the challenge of these category of features because they assume that the data is a fixed signal and more distortions is acquired through it. Most of the burden of computations rely on the signal amplitude but these features have been used most extensively due to their performance for signal classification at low noise environment and low computation time. Total four time domain features are utilised and are discussed below [24].

4.1.1 Mean Absolute Value (MAV)

It is a very well-known feature used in evaluation of EMG signals. It is same as integrated EMG feature used in the detection of surface EMG signal. It is also known as average rectified value (ARV), Integral of absolute value (IAV) or average absolute value (AAV). The first one is basically a reckon of addition of absolute measurement value and measurement of level contraction in the signal. It perceives the mean of the signal amplitude over length of signal as per Eq. (1).

$$mean(\mu) = \frac{1}{N} \sum_{n=1}^N x_n \quad (1)$$

4.1.2 Variance (VAR)

Variance of EMG signal (VAR) is another statistical power tool used to measure EMG signal. Variance is measured as the expectation of average square deviation of random variable from their mean. Variance is also defining as the measure of power density of an EMG signal as in Eq. (2).

$$var = \frac{1}{N-1} \sum_{n=1}^N (x_n - \mu)^2 \quad (2)$$

4.1.3 Standard Deviation (SD)

It is a time domain statistical approach to measure the dispersion of data from its mean (Eq. (3)). It measures the square root of variance by estimating the variation among data points to its mean. If data are outlying from its mean, then it shows the higher deviation within the dataset.

$$std(\sigma) = \sqrt{\frac{1}{M-1} \sum_{m=1}^M (p_m - \mu)^2} \quad (3)$$

4.1.4 Mean absolute deviation (MAD)

It is a statistical approach to find the average interval among each data value of a dataset from its mean (Eq. (4)). It is utilised to find the variations in given data.

$$MAD = \frac{1}{M} \sum_{m=1}^M [P_m - ORT] \quad (4)$$

Once the features have been extracted then they are further applied to recognise the activities using classification algorithms.

5 Pattern Recognition Through Classification with EMG

The activities are being identified through supervised machine learning technique. The classification is the widely used machine learning technique which is helpful to predict the classes for the input data for the pattern recognition [25]. We have used two types of classification techniques which have provided better accuracy for our time domain features as: (1) Linear Discriminant Analysis and (2) Quadratic Discriminant Analysis.

5.1 Linear Discriminant Analysis

It is a supervised machine learning model which is similar to logistic regression and can be used for Classification. LDA can be useful where frequencies of classes are different and their performance is evaluated on randomly generated testing data. LDA gives three benefits over the logistic regression:

- When there is well-separated classes, it would be difficult to estimate the model parameter where LDA would provide efficient classification.
- The Linear Discriminant model is more stable for distribution of homogenous activities.
- LDA would be effective for multi-class classification problems.

LDA has given a better accuracy when compared to ANNs and SVM but in the looking in the perspective of real-time application the training time for is classifier in more. The accuracy which obtained was around 85% with time complexity of around one minute.

The process of algorithm is follows:

- Importing the packages required from sklearn.
- Loading the dataset which is in .csv form using pandas data frame.
- After loading the dataset, we segregate the data into features and labels, since this is a supervised classification.
- Now the crucial part is training the algorithm through LDA. Below are the steps for the same:
 - Calculating the mean vectors.
 - Calculating the Covariance Matrices.

1. The scatter matrix S_w of a class is defined as in Eq. (5)

$$S_w = \sum_{i=1}^{classes} scatter_i \quad (5)$$

where $scatter_i = \sum_{input \in D_i} (input - mean_i)(input - mean_i)^T$ (for every class) and $mean_i$ is the mean vector $mean_i = \frac{1}{total_i} \sum_{input \in D_i} x_k$

2. Covariance matrix for between class, where overall mean is represented by mean, and this of different classes is represented by m_i and N_i are in Eq. (6)

$$Scatter_j = \sum_{classes}^{j=1} Total_j (Mean_j - m)(Mean_j - m)^T \quad (6)$$

- Solving the generalised eigen value problem.
- Choosing K eigen vectors with largest eigen values.
- Transforming onto the new subspace.
- After the training phase, we are onto the testing phase where it randomly selects the data and tests itself.
- Finally we note down the accuracy.

5.2 Quadratic Discriminant Analysis

It is used in statistical classification and machine learning model to separate parameters of classes of events by a polynomial surface. Linear discriminant analysis (LDA) applies the generative approach for classification, i.e., a straightforward method. It is based on the assumption that all the classes have same covariance matrix and each of the class can be modelled by a gaussian distribution. Quadratic discriminant analysis (QDA) is same as LDA but without the assumption that the classes have same covariance matrix, i.e., every class has its own covariance matrix. So the boundary between the classes becomes quadratic. In practice, LDA requires very few computations with which it can estimate classifier parameters. These computations are percentages, matrix inversion and means where QDA has slightly higher computational cost with higher accuracy.

Another classifier which we have used is QDA (Quadratic Discriminant Analysis). This classifier has given a better accuracy when compared to LDA and the important characteristics of this classifier is that it would provide much better accuracy with good real-time performance. The accuracy which obtained was around 90% with time complexity of slightly higher than one minute. The process of algorithm is same as LDA. In QDA, the covariance matrix for every class is different.

5.3 Mathematical Approach for LDA and QDA

LDA assumes that the data of each class is generated by a Gaussian distribution of pdf. It follows a generative approach.

$$P_{(x|Y=y)} = \frac{1}{(2\pi)^{(d/2)} (|\sum_y|)^{\frac{1}{2}}} \exp_x \quad (7)$$

and that the covariance matrix \sum_y is the equal for all the classes:

$$\forall y \in Y, \sum = \sum_y \quad (8)$$

The parameters are estimated as follows. The prior probabilities is the probability for the labelled class in the training data.

$$\forall y \in Y, P(Y = y) = \frac{N_y}{N}, \text{ with } N_y = \sum_{i=1}^N 1_{y_i} = y \quad (9)$$

The means are estimated as:

$$\forall y \in Y, \mu_k = \frac{1}{N_y} \sum_{y_i=y} x_i \quad (10)$$

and the covariance matrix by:

$$\sum = \frac{1}{N - |y|} \sum_{y \in Y} \sum_{y_i=y} (X_i - \mu_y)(x_i - \mu_y)^T \quad (11)$$

This formula comes from the weighted average of the local covariance matrix estimates within each class:

$$\sum = \frac{\sum_{y \in Y} (N_y - 1) \sum_y}{\sum_{y \in Y} (N_y - 1)} \text{ where } N = \sum_{y \in Y} N_y \quad (12)$$

$$\sum_y = \frac{1}{N - |y|} \sum_{y_i=y} (X_i - \mu_y)(x_i - \mu_y)^T \quad (13)$$

By following the same steps Quadratic Discriminant analysis is also obtained but without the assumption that the covariance matrix is common for all classes: each class y has a different covariance matrix \sum_y estimated with the formula. So the boundary between the two classes is not a hyperplane but it is a quadratic surface.

6 VLSI Implementation of a Classification Algorithm for EMG

The conceptual diagram of hardware implementation for the discriminant analysis is shown in Fig. 6. The dataset has been divided into 70% for training while 30% for the testing purpose. In this implementation, we have done the training phase extracted the parameters like mean, covariance, and priori of the training data required for classification. These extracted parameters has to be given as input to the Verilog module. The floating point arithmetic has been performed as per IEEE 754 Single Precision standard.

After implementing those modules, we need to convert the decimal data input into floating point number. This data is given as input to the main module in Verilog using file Input/output concept. The QDA algorithm includes a 2D matrix multiplication but with each value being binary and of 32 bits makes it a 3D multiplication to be performed on the data. Therefore, for the simplification purpose, we have converted the 3D into 2D just with MATLAB. In the next step,, we

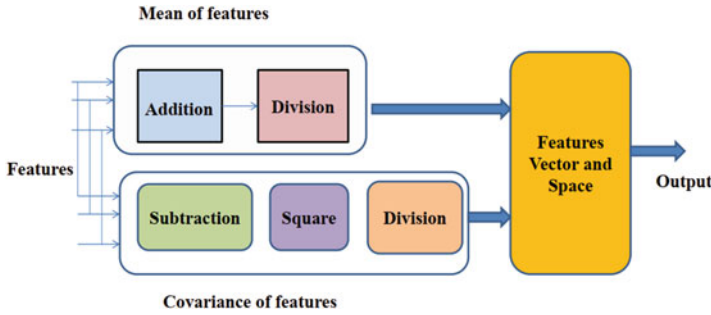


Fig. 6 Conceptual diagram of LDA and QDA implementation

Table 1 Comparison of Performance on various platforms

Implementation	Computational time	Accuracy
Python	75 s	85%
MATLAB HLS	60 s	80%
VLSI	50 s	65–70%

Table 2 FPGA Implementation of LDA and QDA

Resources	Available	LDA	QDA
Slice registers	35,200	240 (0%)	289 (0%)
Slice LUT	17,600	4875 (27%)	5217 (29%)
Bounded IOBs	100	28 (28%)	32 (32%)
Block RAM/FIFO	60	14 (23%)	16 (23%)
BUFG/BUFGCTRLs	32	2 (6%)	3 (9%)

calculate the discriminant functions of all the eight classes. Now, of all those values we find the maximum value. Finally, the function with maximum discriminant analysis is also the predicted class output. Then this module is simulated and output can be observed. The accuracy of LDA has been observed around 80% using python, Matlab HLS synthesis and VLSI. Since the classification needs to be done in real-time, so we consider the computational time required for the different implementations. The computational times for different implementation of QDA algorithm are summarised in Table 1 as above. The reported value of accuracy is average value for both LDA and QDA.

The LDA and QDA are synthesised on Zybo board FPGA using (Zynq 7000 xc7z010clg400 device). The results for utilisation of resources and hardware blocks are shown in Tables 2 and 3, respectively. The performance of timing is also analysed as per Table 4.

VLSI implementation has overall less computational time than others because it helps to exploit the advantage of parallel processing in FPGA up to a full potential [26]. With Verilog HDL being a hardware language, it has been used for the concurrent execution on FPGA [27] and therefore decreasing the computational time. Since the simulation is only for few samples, the accuracy could be improved for higher

Table 3 Hardware Blocks
Utilisation of LDA and QDA

Hardware	LDA	QDA
Multipliers	5	7
Adders/subtractors	12	14
Registers	475	510
Comparators	40	48
Multiplexers	610	676
Xors	18	22

Table 4 Timing analysis of
LDA and QDA

Parameters	LDA	QDA
Minimum period, ns	25	30.3
Maximum clock frequency, MHz	40	33
Minimum input arrival time before clock, ns	1.30	1.38
Maximum output required time after clock, ns	0.51	0.48

number of samples in VLSI implementation. Therefore, VLSI implementation has outperforms other approaches in terms of computational time.

7 Conclusion

The paper presents EMG based pattern recognition for upper limb prosthetic control. Four different time domain features, MAV, VAR, SD, and MAD have been utilised for real-time pattern identification. The two discrimination analysis, LDA and QDA, have been implemented on FPGA for real-time classification. The time complexity analysis have been performed on Software and hardware platform where hardware has better performance with good accuracy. The parallel processing of FPGA would provide edge with concurrent execution therefore increasing the computational time.

References

1. Anders Fougner, Øyvind Stavdahl, Peter J Kyberd, Yves G Losier, and Philip A Parker. Control of upper limb prostheses: Terminology and proportional myoelectric control—a review. *IEEE Transactions on neural systems and rehabilitation engineering*, 20(5):663–677, 2012.
2. Sidharth Pancholi and Amit M Joshi. Electromyography-based hand gesture recognition system for upper limb amputees. *IEEE Sensors Letters*, 3(3):1–4, 2019.
3. SIDHARTH PANCHOLI and AMIT M JOSHI. Intelligent upper-limb prosthetic control (iULP) with novel feature extraction method for pattern recognition using EMG. *Journal of Mechanics in Medicine and Biology*, page 2150043, 2021.
4. Derya Karabulut, Faruk Ortes, Yunus Ziya Arslan, and Mehmet Arif Adli. Comparative evaluation of EMG signal features for myoelectric controlled human arm prosthetics. *Biocybernetics and Biomedical Engineering*, 37(2):326–335, 2017.
5. Erik Scheme and Kevin Englehart. Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use. *Journal of Rehabilitation Research & Development*, 48(6), 2011.

6. Gunter Kanitz, Christian Cipriani, and Benoni B Edin. Classification of transient myoelectric signals for the control of multi-grasp hand prostheses. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(9):1756–1764, 2018.
7. Matthew Dyson, Sigrid Dupan, Hannah Jones, and Kianoush Nazarpour. Learning, generalization, and scalability of abstract myoelectric control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(7):1539–1547, 2020.
8. Kenneth R Lyons and Sanjay S Joshi. Upper limb prosthesis control for high-level amputees via myoelectric recognition of leg gestures. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(5):1056–1066, 2018.
9. Olivier Lamercy, Serena Maggioni, Lars Lünenburger, Roger Gassert, and Marc Bolliger. Robotic and wearable sensor technologies for measurements/clinical assessments. In *Neurorehabilitation technology*, pages 183–207. Springer, 2016.
10. Julian Maier, Adam Naber, and Max Ortiz-Catalan. Improved prosthetic control based on myoelectric pattern recognition via wavelet-based de-noising. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(2):506–514, 2017.
11. Sidharth Pancholi and Amit M Joshi. Improved classification scheme using fused wavelet packet transform based features for intelligent myoelectric prostheses. *IEEE Transactions on Industrial Electronics*, 67(10):8517–8525, 2019.
12. Skyler Ashton Dalley, Huseyin Atakan Varol, and Michael Goldfarb. A method for the control of multigrasp myoelectric prosthetic hands. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(1):58–67, 2011.
13. Kazuo Kiguchi and Yoshiaki Hayashi. Motion estimation based on EMG and EEG signals to control wearable robots. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pages 4213–4218. IEEE, 2013.
14. Jie Liu. Adaptive myoelectric pattern recognition toward improved multifunctional prosthesis control. *Medical engineering & physics*, 37(4):424–430, 2015.
15. Sidharth Pancholi and Amit M Joshi. Portable EMG data acquisition module for upper limb prosthesis application. *IEEE Sensors Journal*, 18(8):3436–3443, 2018.
16. Turker Tuncer, Sengul Dogan, and Abdulhamit Subasi. Surface EMG signal classification using ternary pattern and discrete wavelet transform based feature extraction for hand movement recognition. *Biomedical Signal Processing and Control*, 58:101872, 2020.
17. Sidharth Pancholi and Amit M Joshi. Advanced energy kernel-based feature extraction scheme for improved EMG-PR-based prosthesis control against force variation. *IEEE Transactions on Cybernetics*, 2020.
18. Bernabe Rodríguez-Tapia, Israel Soto, Daniela M Martínez, and Norma Candolfi Arballo. Myoelectric interfaces and related applications: current state of EMG signal processing—a systematic review. *IEEE Access*, 8:7792–7805, 2020.
19. Ahmed W Shehata, Erik J Scheme, and Jonathon W Sensinger. Evaluating internal model strength and performance of myoelectric prosthesis control strategies. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(5):1046–1055, 2018.
20. Joseph L Betthausen, Christopher L Hunt, Luke E Osborn, Matthew R Masters, György Lévy, Rahul R Kaliki, and Nitish V Thakor. Limb position tolerant pattern recognition for myoelectric prosthesis control with adaptive sparse representations from extreme learning. *IEEE Transactions on Biomedical Engineering*, 65(4):770–778, 2017.
21. Meike A Wilke, Cornelia Hartmann, Felix Schimpf, Dario Farina, and Strahinja Dosen. The interaction between feedback type and learning in routine grasping with myoelectric prostheses. *IEEE transactions on haptics*, 13(3):645–654, 2019.
22. Alexander E Olsson, Anders Björkman, and Christian Antfolk. Automatic discovery of resource-restricted convolutional neural network topologies for myoelectric pattern recognition. *Computers in Biology and Medicine*, 120:103723, 2020.
23. Sidharth Pancholi and Amit M Joshi. Time derivative moments based feature extraction approach for recognition of upper limb motions using EMG. *IEEE Sensors Letters*, 3(4):1–4, 2019.

24. Sidharth Pancholi, Prateek Jain, Arathy Varghese, et al. A novel time-domain based feature for EMG-PR prosthetic and rehabilitation application. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5084–5087. IEEE, 2019.
25. Sidharth Pancholi, Amit M Joshi, and Deepak Joshi. A robust and accurate deep learning based pattern recognition framework for upper limb prosthesis using sEMG. *arXiv preprint arXiv:2106.02463*, 2021.
26. Shilpa Thakral, Divya Goswami, Ritu Sharma, Challa Krishna Prasanna, and Amit Mahesh Joshi. Design and implementation of a high speed digital fir filter using unfolding. In *2016 IEEE 7th Power India International Conference (PIICON)*, pages 1–4. IEEE, 2016.
27. Amit M Joshi, Vivekanand Mishra, and Rajendra M Patrikar. Fpga prototyping of video watermarking for ownership verification based on h. 264/avc. *Multimedia Tools and Applications*, 75(6):3121–3144, 2016.

Content-Based Image Retrieval Techniques and Their Applications in Medical Science



Mayank R. Kapadia and Chirag N. Paunwala

1 Introduction

Image retrieval is a method to retrieve images from a vast image database. An image retrieval system is a computer system for browsing, searching, and retrieving images from an extensive digital image database. Image retrieval has become an important research area in computer vision due to the continuous advancement in technology, which caused a significant increase in the number of images worldwide. With the image processing techniques, the storage of these massive numbers of images becomes easy. Therefore, there is a need for an efficient image retrieval technique.

The most common method for retrieving the image from the vast collection of image databases involves tags, annotation, keywords, or short descriptions through text-based image retrieval (TBIR). It is an old method, starting in the 1970s. This technique requires text as an input to search for an image. The very well-known search engines such as Google and Yahoo are the examples, which are using this approach. These prominent search engines are robust and fast but sometimes retrieve irrelevant images. It is shown in Fig. 1.

The irrelevant words in the surrounding textual descriptions and the surrounding text do not entirely describe web images' semantic content. Moreover, text for image labeling and tagging does not portray clearly and definitely what the image represents. Not only this, even the same word can have several meanings in a different context. For example, the word "orange" can be an "orange color" or "orange fruit." This language ambiguity problem has been shown in Fig. 1.

M. R. Kapadia (✉)

C. G. Patel Institute of Technology, Uka Tarsadia University, Bardoli, India

C. N. Paunwala

Sarvajnik College of Engineering and Technology, Surat, India

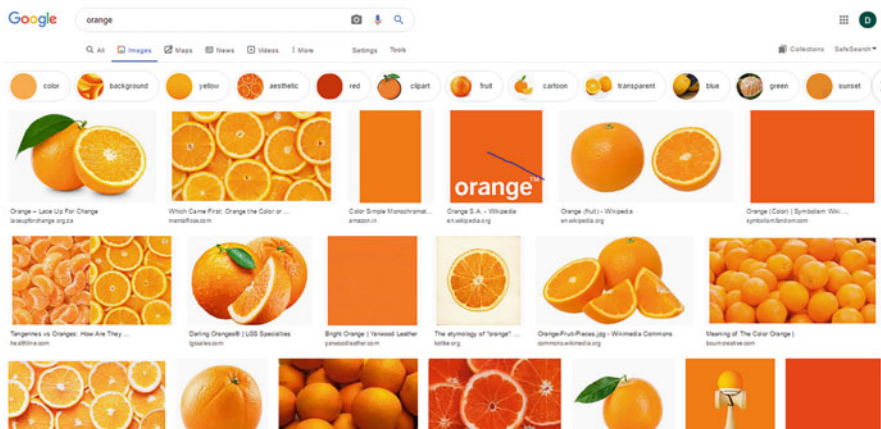


Fig. 1 Text-based image retrieval (TBIR)

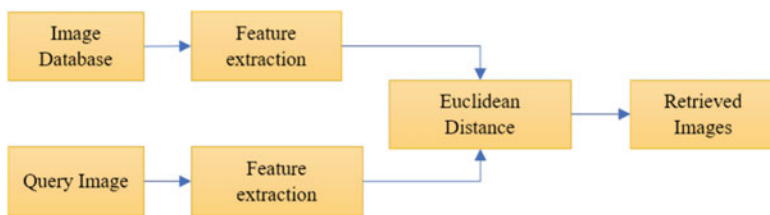


Fig. 2 Block diagram of CBIR system

Sometimes, to mention more than one object, a similar word can be utilized. This problem is known as the polysemy problem. Sometimes the English searcher will not find the image tagged in Gujarati, and a Chinese searcher will not find the image tagged in English. It indicates that text related to images must be matched with query text. The query text must match the language. Another problem is the manual annotation of the massive amount of images. To describe, the image is a highly subjective task. Hence, the traditional TBIR method relies upon a keyword search that contains limitations such as a maximum human resources requirement and dependency requirement. In order to overcome these disadvantages, the content-based image retrieval (CBIR) method has been utilized [1].

In the late 1990s, CBIR was introduced by T. Kato, and later CBIR becomes a very active research area. It has been used as an alternative to TBIR. IBM was the first, who took the initiative by proposing query by image content (QBIC). “Content-Based” means that the search needs image contents instead of metadata such as keywords, tags, or descriptions linked with the image. The term “content” can be features such as color, shape, and texture. CBIR is needed since most web-based image search engines depend only on metadata. Due to this, a lot of junk in the results. The basic block diagram of CBIR is shown in Fig 2.

The CBIR system is beneficial and a useful technique in searching and retrieving images from a vast database. Similar images are retrieved from the image database. The CBIR system derived the feature associated with the entered image, and it will be compared with the features related to the database's images. The system displayed the images whose features are closest to the entered image.

2 Classification of CBIR Techniques

The CBIR techniques are classified into three categories based on the feature extraction and retrieval of the images. Figure 3 shows that CBIR techniques are classified into traditional feature-based techniques, machine-learning-based techniques, and deep-learning-based techniques.

2.1 Feature-Based Technique

Feature extraction is a technique of deriving compressed but semantically valuable information from images. CBIR uses an image's visual contents such as color, shape, texture, and spatial layout to represent the image. At the early stage of the research, a single feature-based image retrieval system was developed. But one feature is not enough to describe the image since the image contains various visual characteristics [2]. Hence, to benefit the image's different visual aspects, the researcher has started the fusion of the features.

2.1.1 Color Features for Image Retrieval

CBIR is the most basic and most crucial method for image retrieval. It is also an essential feature of perception. Color features are not susceptible to scale change, translation, and rotation. It is robust and stable. The computation complexity is minimum in the color feature. It is an extensively used visual feature in the CBIR system.

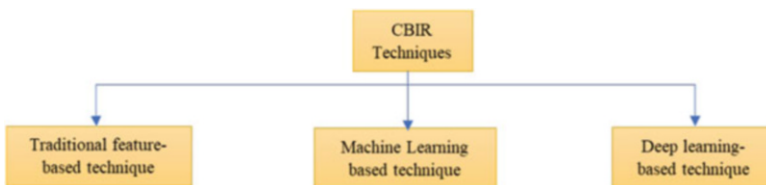


Fig. 3 Classification of CBIR techniques

- **Color Histogram [3–9]**

The color histogram usually represents the color. The color histogram is represented by the bar graph. The height of the bar shows the quantity of the color. The image can be represented by RGB or HSV color space. It is shown in Fig. 4.

Histogram can be calculated either globally or locally. In a global histogram, the histogram is calculated on the whole image. In contrast, in a local histogram, the image is divided into small blocks, and then a histogram is calculated for each block. A drawback of a global histogram is that it will not preserve a particular color's spatial location in an image while possible in a local histogram.

Figure 5 shows the importance of the image's spatial location, which contains the sky and sea. The sea must be at the bottom of the image, and the sky must be at the top of the image.

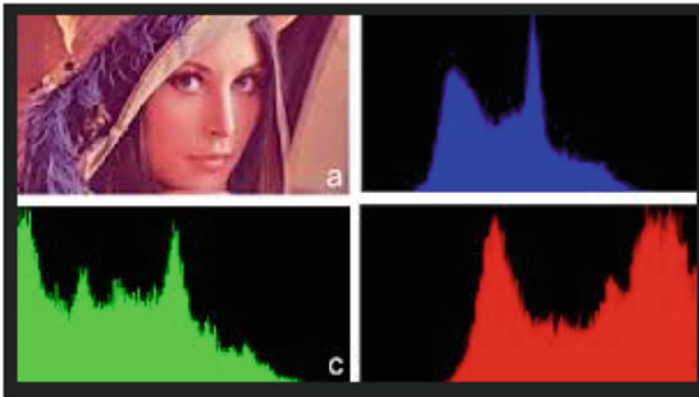


Fig. 4 Color histogram for an image

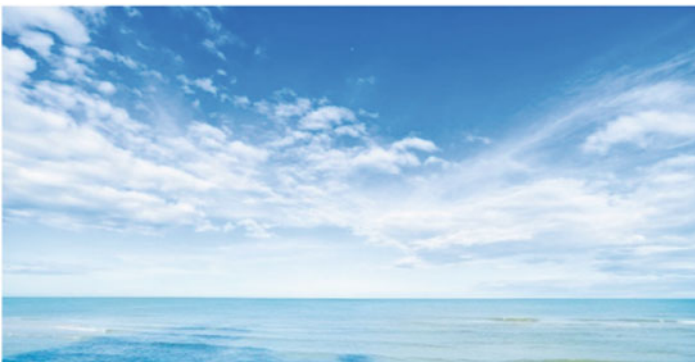


Fig. 5 Importance of spatial locations of sky and sea



Fig. 6 Two semantically different images

Another drawback of the color histogram is that the two semantically different images can assume a very similar color histogram. Also, the same image taken under different lighting conditions may produce a different histogram. It is shown in Fig. 6.

- **Color Moments [2, 3, 10–13]**

Color moments are the statistical moments. Color moments show the probability distributions of colors. When an image contains only an object at that time, these moments are used. The mean and variance have been used to represent the color distribution of the image. Equations (1) and (2) show the color moments, e.g., mean and variance.

$$\mu_i = \frac{1}{N} \sum_{j=1}^N P_{ij} \quad (1)$$

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (P_{ij} - \mu_i)^2}. \quad (2)$$

Figure 7 shows how the color moments have been derived as a feature for the CBIR system.

The significant advantages of the color moments are the small feature vector size and lower computational complexity. Simultaneously, they are unable to encode any spatial information surrounding the color within the image.

2.1.2 Shape Features for Image Retrieval

Human identifies the object by their shapes. The boundary of the object is referred to as the shape of the object. Hence, shape features provide vital information about the

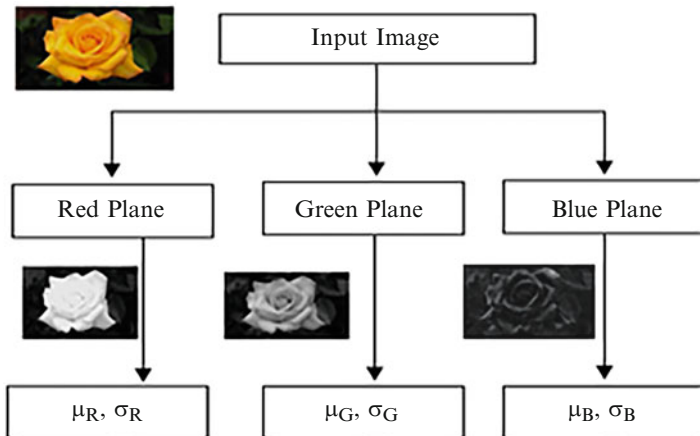


Fig. 7 Color moments for RGB image [13]

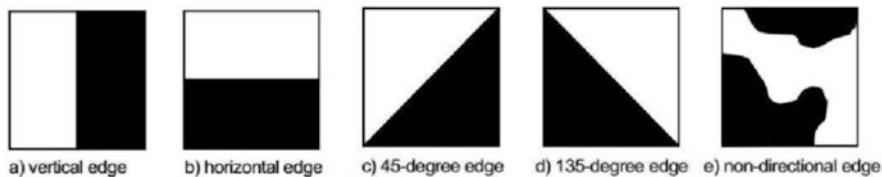


Fig. 8 Five types of edges in EDH

image, which helps for the image retrieval. The shape features are derived using the out boundary or using the entire region of the object. In some cases, the boundary is not evident, which is a limitation of this method.

• **Edge Direction Histogram (EDH) [8–10]**

EDH is the most popular technique to capture the shape features of an image. EDH is used to capture shape features of color images that create a histogram to simulate the edges’ distribution in an image. It determines the local edge distribution in an image. EDH is obtained by partitioning the entire image into small blocks.

The EDH contains five types of edges, e.g., the vertical, horizontal, diagonal, and non-directional edge. It is shown in Fig. 8.

2.1.3 Texture-Feature-Based Image Retrieval

The texture refers to the repetitive pattern. It has valuable information regarding the structural arrangement and its relationship with the surrounding pixels. Various

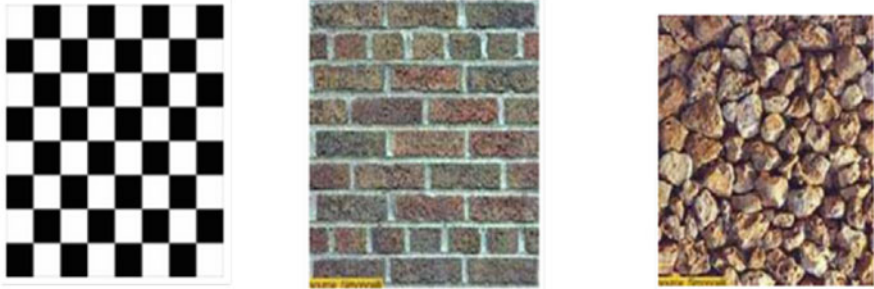


Fig. 9 Sample images of texture

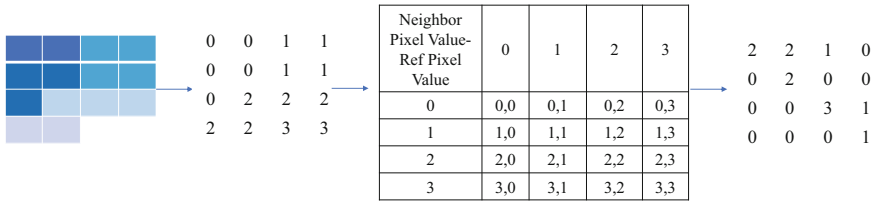


Fig. 10 Process of GLCM calculation

parameters can be extracted as a texture feature. Some samples of texture images are shown in Fig. 9.

• **The Gray-Level Co-occurrence Matrix [3, 13]**

Gray-level co-occurrence matrix (GLCM) is the most popular technique for texture analysis. It is utilized to estimate image properties correlated to second-order statistics. A pixel with a gray-level value i occur horizontally adjacent to a pixel with the value j creates the GLCM. Each element (i, j) in GLCM specifies the number of times that the pixel with a value i occurred horizontally adjacent to a pixel with value j . The process to calculate the GLCM matrix is shown in Fig. 10.

Contrast, homogeneity, energy, and correlation are examples of the features that can be derived from the GLCM. It can be calculated using Eqs. (3), (4), (5), and (6), respectively.

$$\text{Contrast} = \sum |i - j|^2 P(i, j). \tag{3}$$

$$\text{Homogeneity} = \sum_{ij} \frac{1}{1 - (i - j)^2} P(i, j). \tag{4}$$

$$\text{Energy} = \sum_{ij} P(i, j)^2. \tag{5}$$

$$\text{Correlation} = \sum_{i,j} P(i, j) \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]. \quad (6)$$

The contrast is used to find the linear dependency of the neighboring pixels. The homogeneity measures the closeness of the distribution of pixels. The energy is utilized to measure the uniformity of the texture. The energy is ranged from 0 to 1, where 1 represents the constant image. The pixel correlation with its neighbor is calculated using the correlation.

- **Wavelet transform [7, 8, 11, 14]**

Wavelet transform is the most popular technique for texture feature extraction. Wavelet transform has a multi-resolution capability. It divides the image into an approximate and detailed band. So, it provides information regarding the intensity and edges.

- **Gabor filter [7, 9, 11, 12]**

Gabor filters consist of a group of wavelets that capture energy at a specific resolution and orientation. Hence, Gabor filters can catch the local energy of the whole image. The Gabor filter is the most popular technique for texture features. The Gabor filter can produce several texture scales and texture orientations. Gabor filters provide adequate resolution in visual space. Gabor filter helps to capture granularity and repetitive patterns of the surface. It has significantly richer information than color histograms and corresponds to human perception. Gabor filter's drawback is that it is susceptible to transforms such as scaling illumination and viewing angle.

The research has used these color, shape, and texture features for image retrieval. Some researcher has fused these features to improve the retrieval rate of the CBIR system. Table 1 shows the comparative analysis of the various independent or feature combined techniques for the well-known Corel-1K dataset. It is also known as the Wang1000 dataset. This dataset contains 1000 images of 10 different classes.

From Table 1, the average precision rates (APRs) for the individual feature-based techniques, e.g., color moments (CM), GLCM, local color histogram (LCH), global color histogram (GCH), edge descriptor histogram (EDH), discrete wavelet transform (DWT), are below 60%. The combination of the features increases the retrieval rate. The more feature gives more characteristics of the image and helps to improve the retrieval rate. Here, the combination of color histogram (CH) and DWT gives comparatively good results than both the individual techniques. Similarly, the combination of three different features further increases the retrieval rate. The combination of color coherence vector (CCV) and Gabor filter gives the best results among all the mentioned techniques.

As shown in Fig. 11a, an image of the "bus" class is given as a query. The CBIR system retrieved the images of the buses from the image database. Similarly, in Fig. 11b, an image of "flower" is provided as a query, and the CBIR system has retrieved the images of flowers from the image database.

Table 1 Comparative analysis of various techniques for Wang or Corel-1K dataset

Class names	CM [3]	GLCM [3]	LCH [3]	GCH [3]	HSV-CH [8]	EDH [8]	DWT [8]	CH+DWT [8]	CH+EDH+DWT [8]	CCV+Gabor [12]
Africa	29	80	26	28	65	65	55	55	85	90
Beach	23	66	38	50	35	20	25	30	50	70
Building	30	50	46	31	45	50	55	75	75	80
Bus	45	50	44	44	75	85	70	80	100	100
Dinosaur	87	85	91	80	95	90	95	100	100	100
Elephant	32	33	30	37	35	15	35	80	55	100
Flower	70	56	66	68	70	65	70	95	95	100
Horse	67	71	48	64	75	60	85	90	90	90
Mountain	57	67	63	65	75	25	30	40	30	80
Food	37	33	44	42	35	30	65	50	55	80
APR (%)	47.7	59.1	49.6	50.9	60.5	51	58.5	69.5	73.5	89



Fig. 11 (a) Image retrieval for the bus class. (b) Image retrieval for flower class [8]

2.2 Machine-Learning-Based Technique

The machine learning technique is generally used for the classification of the image. Classification is used to categorize the image. Classification is the procedure of assigning a class label to an image. The classification procedure aims to split the whole image database into numbers of classes. The classification techniques are divided into two categories:

- Supervised learning
 In the supervised learning, all the input image is used to train the network and is associated with an output class, e.g., support vector machine, K-nearest neighbor (KNN), and neural network.
- Unsupervised learning

In unsupervised learning, the output class is not given to the network. Hence, the system learns on its own, e.g., K-means clustering.

2.2.1 Support Vector Machine [15–18]

Support vector machines (SVMs) are used to classify both linear and non-linear data. An image is highly non-linear. Hence, features derived from the images are also non-linear. SVM utilizes a non-linear mapping to convert the original training images' feature into a higher dimension. This new dimension searches for the linear optimal separating hyperplane, which separates one class from another. In high dimensions, it is possible to separate two classes using a hyperplane. The SVM identifies this hyperplane using support vectors and margins.

Let $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be the pair of training images' features and a class of the images, respectively. The class of the image y_i can be $+1$ or -1 , according to the class. An infinite number of hyperplanes can be possible to separate the two classes. But the appropriate hyperplane must be selected, which provides the maximum margin between the separating planes. It is shown in Fig. 12.

Here, the weights “ w ” can be adjusted to separate data into two classes. The hyperplanes define the sides of the margin, and it can be written as

$$H1 : w_0 + w_1x_1 + w_2x_2 \geq 1 \text{ for } y_i = +1 \quad (7)$$

$$H2 : w_0 + w_1x_1 + w_2x_2 \leq -1 \text{ for } y_i = -1. \quad (8)$$

Any image located on or above $H1$ is configured as class $+1$, and features of any image located on or below $H2$ are configured as class -1 . Combining the two Eqs. (7) and (8) results in Eq. (9).

$$y_i(w_0 + w_1x_1 + w_2x_2) > 1. \quad (9)$$

Fig. 12 Linearly separable features [13]

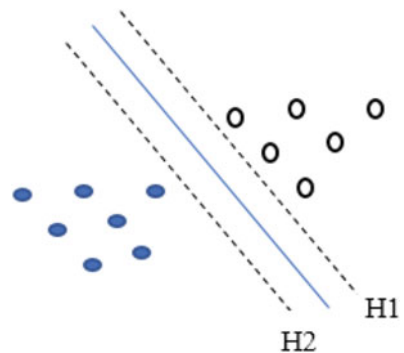
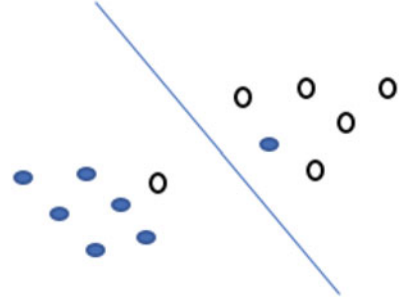


Fig. 13 Linearly non-separable features [13]



The training features located on hyperplanes H_1 or H_2 are named support vectors. It is beneficial for linearly separable features. The non-linear SVM is used when the features of the images cannot be linearly separable. It is illustrated in Fig. 13.

If features are linearly non-separable, then the following kernels are utilized for searching the separating hyperplane.

$$\text{Polynomial: } K(x, y) = \gamma (x^T y + c)^d. \quad (10)$$

$$\text{Radial Bases Function (RBF): } K(x, y) = e^{-\gamma(x-y)^2} \quad (11)$$

$$\text{Sigmoid: } K(x, y) = \tanh(\gamma(x^T y) + c). \quad (12)$$

Here x and y are the input vectors, d is the polynomial degree, γ is the adjustable parameters, and c is the constant term. The adjustable parameter γ plays a vital role in the kernel's performance, so it must be wisely selected for a better result.

SVM is generally preferable for small datasets. It is also effective in high-dimensional space. But for a large dataset, it is difficult to identify the hyperplane even in the high dimension. Also, it will take much time for the large image datasets for the training. It is less effective on noisier datasets with overlapping classes.

2.2.2 K-Means Cluster [19–21]

K-means cluster is one of the simplest unsupervised learning techniques. It is used to classify the images based on the feature's similarity with the "k" centers cluster. Following are the steps for the K-means clustering technique:

- Initially, select the "k" number of cluster centers randomly.
- Calculate the distance between the features of the image and cluster centers.

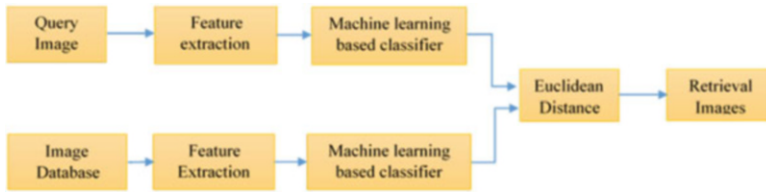


Fig. 14 Machine-learning-based CBIR system

- Based on the distance calculation, assign the image to that cluster whose distance is minimum from all other cluster centers.
- After assigning the image, again calculate the new cluster center.
- Again, calculate the distance between features of the image and newly obtained cluster centers.
- Repeat this procedure till the availability of the features.

K-means can easily be implemented and also work well for the large image dataset. The major drawback of the K-means cluster is to identify the initial value of “K.” It will not work properly for different cluster sizes and densities.

2.2.3 Machine-Learning-Based Image Retrieval System

Machine-learning-based image retrieval system takes the handcrafted features for the training of the model. Once the model is trained, the trained model is used to retrieve similar images from the database. The most popular models in machine learning techniques are the SVM and K-means cluster techniques. Figure 14 shows the machine-learning-based CBIR system.

Here, features are extracted from the database images. These features can be color, shape, or texture features. These handcrafted features will be given to the machine learning model. This model will be trained. The same features are extracted from the query image. These features will be provided to the trained machine learning models such as SVM, K-means, etc. The classifier will give the class of the query image. Later, the Euclidean distance between the feature of the query image and the identified class’s images is calculated. Based on the distance, images of the database are arranged in ascending order, and then most similar images will be retrieved. Table 2 shows the comparative analysis of feature-based techniques and fusion of handcrafted features and machine learning technique SVM. This comparison is made on the Corel-1K dataset, which contains 1000 images of 10 different classes.

As shown in Table 2, feature fusion with machine-learning-based classification provides better results than the without machine-learning-based technique and single feature with machine-learning-based technique.

Table 2 Comparative analysis of feature-based and machine-learning-based techniques

Measurement parameters	Average precision rate (%)
Color moments [3]	47.7
Color moments+SVM [13]	86.2
Color moments+GLCM+SVM [13]	89.6

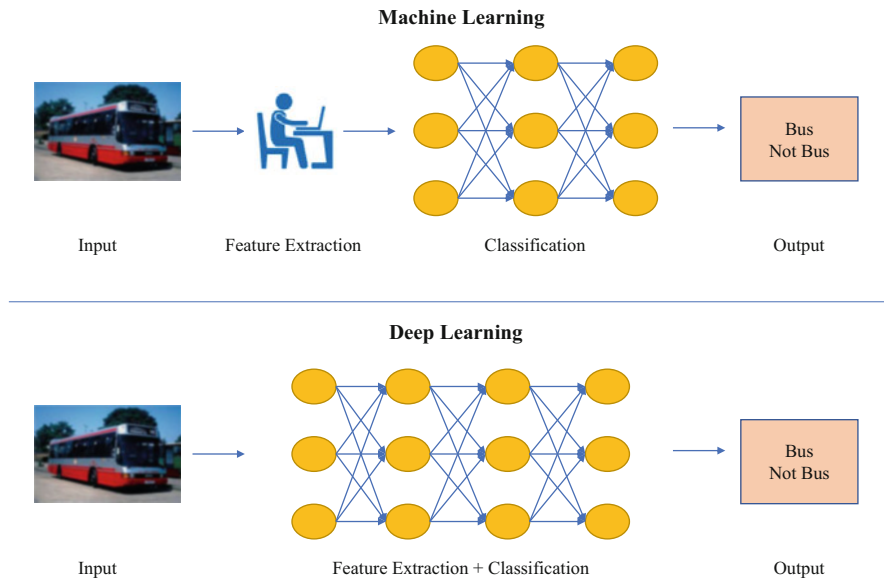


Fig. 15 Machine learning vs. deep learning

2.3 Deep-Learning-Based Technique

Research in the field of the CBIR is increased after the advancement in machine learning algorithms. The CBIR domain gets attention due to the large size of database availability and the availability of high-speed computing hardware such as the Graphics Processing Unit (GPU). The latest neural networks are deep neural networks where the number of hidden layers is more between input and output layers. Initially, hidden layers are used to extract the low-level features such as color, edge, or shape from the images. These features are later on provided to the higher-level hidden layers to derive the high-level features. Hence, such a deep learning algorithm can extract the high-level features from the images, and thus it is possible to overcome the semantic gap problem. The deep learning algorithm requires a massive amount of data for the initial training.

As shown in Fig. 15, in machine-learning-based technology, the features were extracted and then provided to the classifier. The deep-learning-based method is an end-to-end process where high-level features are automatically extracted from the provided data. The deep learning method takes much time for the training data, but

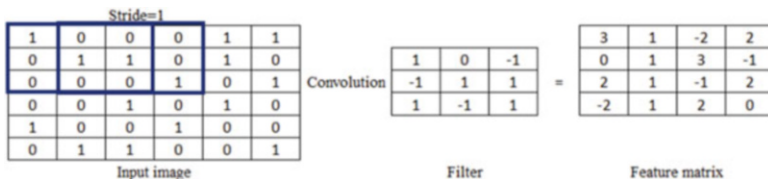


Fig. 16 Convolution operation

once the network is trained, it will take a reasonable time for the testing data. The convolution neural network (CNN) is the most popular network of deep learning technology, widely used for 2-D data-like images.

The CNN has many different layers such as convolution, max pooling, rectified linear unit (ReLU), and fully connected layers. These layers work as the features extractor, which links the input image’s pixel information with its category. The parameters of these layers are tuned and optimized to reduce the misclassification error.

The convolutional layer has a 2-dimensional matrix, which is known as a filter. The convolution layer is made up of neurons, which can learn by modifying the weights and bias. While the neural network is being trained, each filter of this convolution layer is convolved with the input image, performs the dot product operation between filter and input image, and generates the 2-dimensional feature map as a result. It can be mathematically represented by Eq. (13).

$$g(p, q) = f(p, q) * h(p, q) = \sum_n \sum_m f(n, m)h(p - n, q - m). \tag{13}$$

Here, filter “h” is convolved with the fragment of the image “f” centered at (p, q) point and generates (p, q) point of feature matrix g. Figure 16 shows how convolution operation generates feature maps from the input image.

Here, 3×3 filter is applied over the 6×6 input image. The first point of the feature matrix is calculated by dot product as follows: (1*1+0*0+0*−1+0*−1+1*1+1*1+0*1+0*−1+0*1) = 3. The filter is shifted right over the input image by one pixel, known as the stride of one pixel. The operation is repeated for the remaining feature matrix calculation by shifting the filter mask in both directions. The size of the resultant feature matrix is 4×4, as shown in Fig. 16. The appropriate size of the filter and the stride are selected to improve the result’s accuracy. In general, if image size is “i x j,” the filter size is “1 x m,” and “n” number of the filter is being used, then the feature matrix “g” generated by the convolution is given by Eq. (14).

$$g = n \times (i - 1 + 1)x(j - m + 1). \tag{14}$$

ReLU is the non-linear activation function that is used after the convolution process. The ReLU layer is made up of the activation function $f(a) = \max(0, a)$, where “a” is the layer’s input. This layer improves the non-linear characteristics of

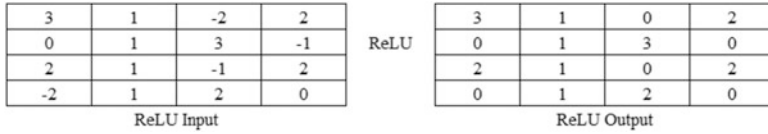


Fig. 17 ReLU operation

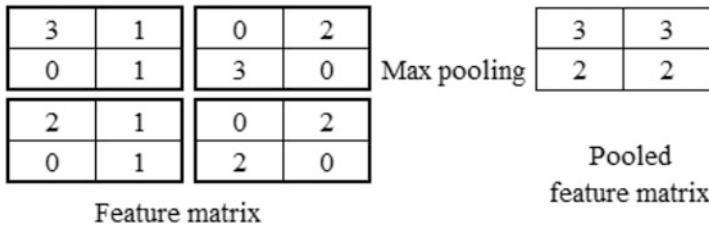
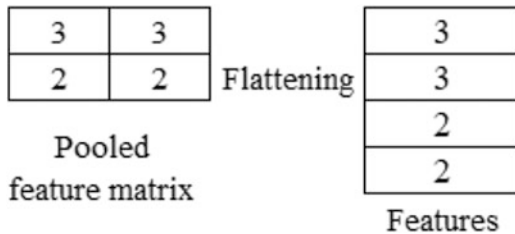


Fig. 18 Max pooling operation

Fig. 19 Flattening operation



the network. Figure 17 shows the ReLU operation on the input matrix of the ReLU function.

Max pooling layers are utilized to decrease the spatial dimension of the feature matrix. It is used to decrease the computational cost by directly reducing the number of parameters. This layer also helps to overcome the chances of the overfitting problem of the network. The $g \times h$ size mask is used over the non-overlapping area of the input matrix. Here the input matrix is reduced by a factor of g and h along with both height and width. This layer will not change the deepness. The maximum activation function provides the output of this layer. Figure 18 shows how the pooling layer reduces the feature matrix.

The fully connected layer needs the one-dimensional vector as an input. The last pooling layer's output is given to the flattening layer, which converts the multidimensional matrix into the one-dimensional vector. Figure 19 shows the operation of the flattening layer.

The fully connected layer contains a full connection. So, the output of the flattening layer is applied as an input to the fully connected layer. The last layer of the fully connected network performs the classification operation because one neuron is available for each class in this layer. It uses the softmax activation function for the classification operation. Figure 20 illustrates the fully connected layer.

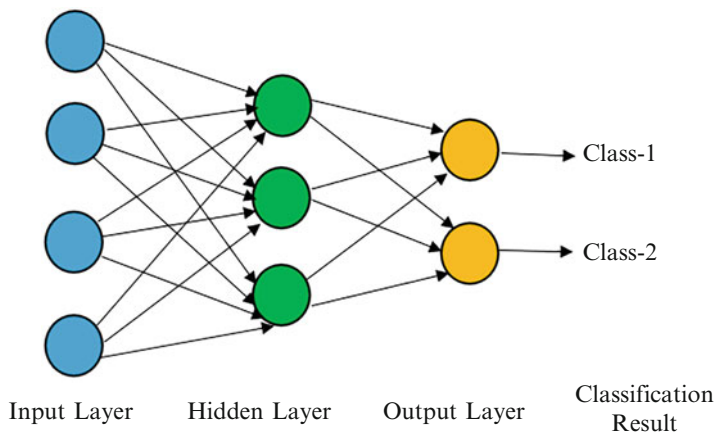


Fig. 20 Fully connected layer

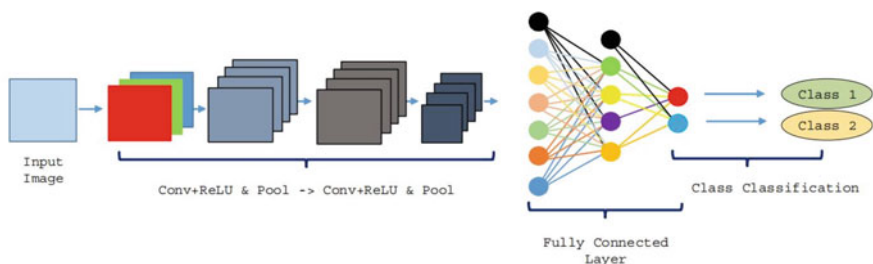


Fig. 21 CNN architecture [22]

Apart from this, some additional layers such as dropout layers are used to overcome the problem of overfitting. In this dropout layer, some neurons are dropped for the training during the training of the network. Hence, mathematical complexity is reduced, and it prevents the overfitting issue of the network. The spatial transformer unit is used to accomplish the geometric transformation of the provided input images. Hence, there is no need to do data augmentation such as translation, scaling, rotation, and skewing manually. The batch normalization layer is used to normalize the layer’s inputs for each mini-batch to stabilize the learning process, which will reduce the number of epochs required to train the network. In this way, batch normalization accelerates the training process and minimizes the generalization error.

Figure 21 shows the basic architecture of the CNN model. There is an “N” number of convolution layers in between inputs and the flattening layers. The primary difference between the NN and CNN is that NN uses the shallow network, while the CNN uses the deep network.

The CNN’s major drawback to train from scratch is that it required a large image dataset to learn the model. It is not easy to initialize the weights. Another disadvantage is that the extensive dataset training required a high-speed processor

such as GPU, which is very expensive. Sometimes it will take a week or a month to train the model from scratch for the large image dataset. The solution to this problem is the transfer learning model.

2.3.1 Transfer-Learning-Based Technique

The model learned for one problem can be utilized for another similar problem. For example, knowledge about the recognition of car can be applied to recognize truck. This concept is known as transfer learning. In transfer learning, the model is not trained from scratch. But, the model that was already trained for one problem will start leaning for the second problem. Some transfer learning models are explained in this section:

- **VGG [23–30]**

Karen Simonyan and Andrew Zisserman found the convolutional network depth's effect on its accuracy [30]. They increased their architecture depth to 16 and 19 layers with tiny (3x3) convolution filters. These models are known as VGG16 and VGG19. The full form of VGG is visual geometry group.

VGG16 has a total of 16 layers in its architecture. It contains thirteen convolutional layers, followed by maximum pooling layers and three fully connected layers. VGG19 has a total of 19 layers in its architecture. It includes sixteen convolutional layers, followed by maximum pooling layers, and three fully connected layers.

- **ResNet [25, 31–34]**

In the CNN network, the researchers observed that “The deeper, the better.” But it has been observed that after some depth, the performance degrades in terms of accuracy. This is a drawback of the VGG network. The solution is residual network (ResNet). It was designed to enable hundreds or thousands of convolutional layers. While the CNN architectures' effectiveness decreases with the additional layers, ResNet performs well with many layers. The basic block of the ResNet model is shown in Fig. 22.

In ResNet, the input to the first layer of the model is also added to its last layer. Due to this, it can solve the vanishing gradient problem. When the network is too

Fig. 22 Residual learning: a building block [33]

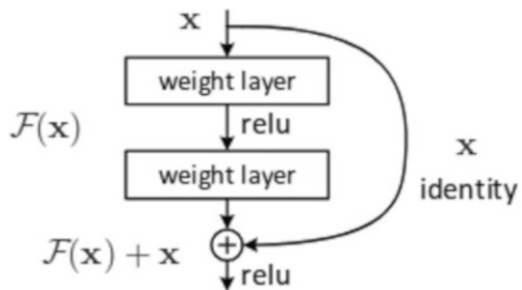
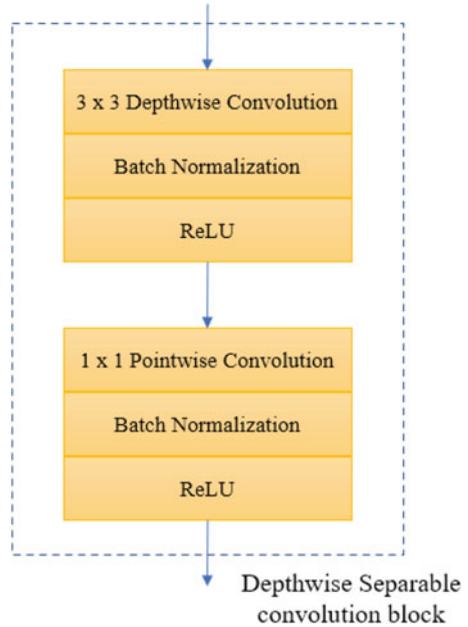


Fig. 23 Concept of depthwise separable convolution



deep, the gradients from where the loss function is calculated quickly shrink to zero after passing from the several layers. Thus, the weights will not be updated, and due to that, the learning process is not being performed. With ResNet, the gradients can pass through the skip connections backward from later layers to initial filters.

- **MobileNet [35, 36]**

The convolution layers are necessary for the deep network, but the computational expense is more. Hence, convolution layers are replaced by depthwise separable convolution in MobileNet. In this network, the convolution layer is divided into depthwise and pointwise convolution operation. It is shown in Fig. 23.

In the normal convolution operation, a kernel is applied to all the input image's channels. It performs the weighted sum operation, combines all the channels' value, and generates a single-channel output. It is shown in Fig. 24.

The MobileNet architecture uses normal convolution only in the first layer. In all the remaining layers, it uses depthwise separable convolution. In depthwise convolution, it performs convolution operation on every channel separately. For example, if the numbers of channels in the input are three, then output channels also remain three. It is shown in Fig. 25a.

A pointwise convolution follows the depthwise convolution. This is the same as a regular convolution but with a 1×1 kernel. It is shown in Fig. 25b. The pointwise convolution is used to combine the output of the depthwise convolution. So, the combination of depthwise and pointwise convolutions is known as depthwise separable convolution. The output dimension of normal and separable

Fig. 24 Normal convolution operation

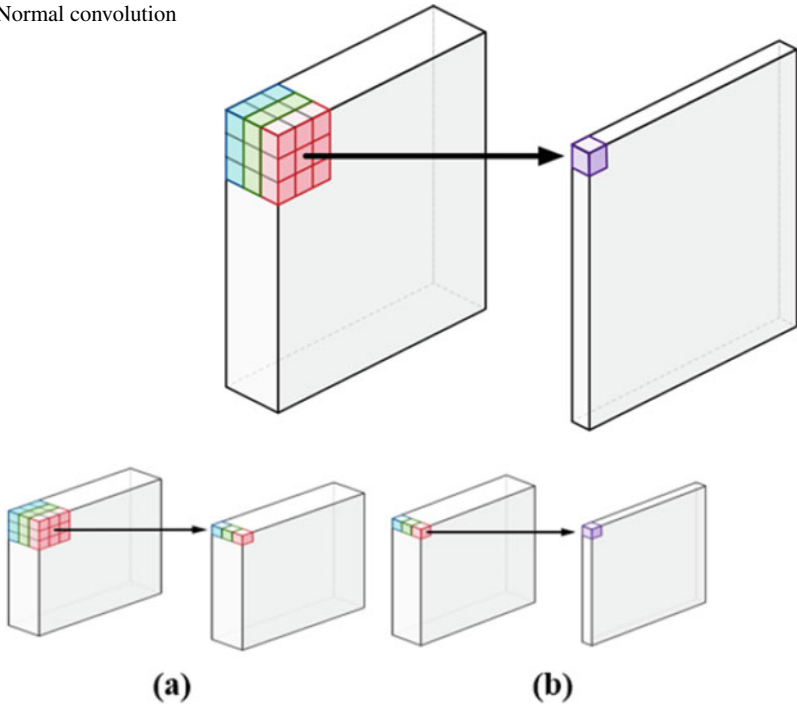


Fig. 25 (a) Depthwise convolution operation. (b) Pointwise convolution operation

convolutions is the same, but the normal convolution required more computation work and needs to train more weights. Due to less multiplication operation, separable convolution is faster and effective than the normal convolution.

- **DenseNet [37]**

In the dense convolutional network (DenseNet), each layer is connected to every other layer in a feed-forward fashion. DenseNets are the next step in increasing the depth of deep convolutional networks. DenseNet is a logical extension of ResNet. The CNN network is deep, and due to that, the information can vanish during the passing from the first input layer to the last output layer. In ResNet, every layer has its weight to learn. Due to that, the numbers of parameters are too large. DenseNet layers are narrow, and it will generate a small set of feature maps. In DenseNets, each layer directly accesses the gradients from the loss function and the original input image. Hence, it solved the problem of effective training.

Figure 26 shows the comparison of normal CNN architecture, ResNet architecture, and DenseNet architecture. As shown in Fig. 26c, in DenseNet architecture, each layer gets extra inputs from all previous layers and passes on its feature maps to all succeeding layers. Here, concatenation is used. Each layer is getting collective information from all previous layers. This architecture solves

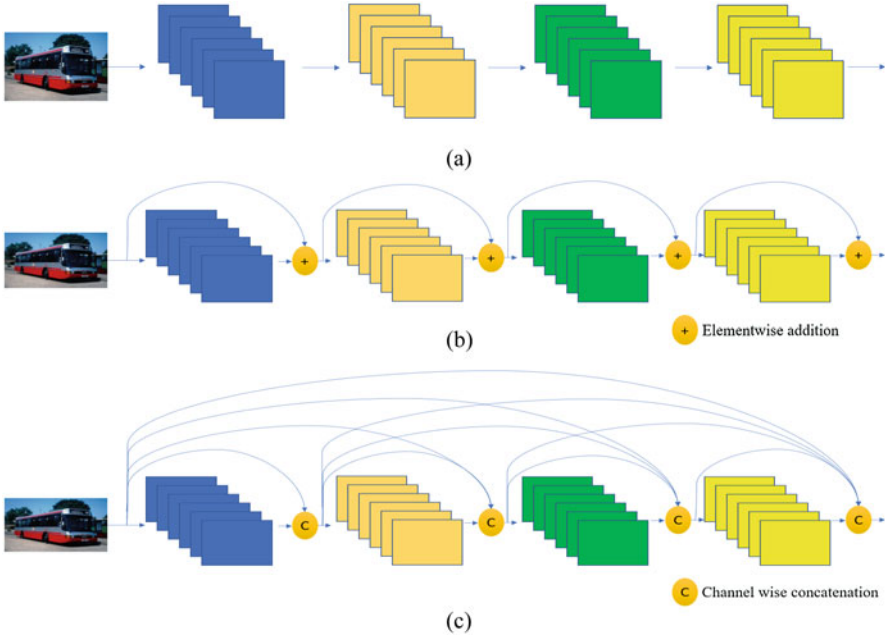


Fig. 26 (a) CNN architecture. (b) ResNet architecture. (c) DenseNet architecture

the problem of gradient vanishing and also reduces the number of parameters to store. It also effectively propagates and reuses the features.

2.3.2 Deep-Learning-Based Image Retrieval System

The deep-learning-based image retrieval system is shown in Fig. 27. In the training process, the database images are used to train the deep learning model. The deep learning model can be any model developed from scratch using CNN or any pre-trained model. In the testing process, the trained deep model identifies the class of the unknown query images. The trained deep model also provides the features of the query image. In the image retrieval process, the Euclidean distance between the database image's features and query image's features is calculated, and images are arranged in ascending order based on the Euclidean distance. The most relevant images are retrieved as a result.

Table 3 shows the comparative analysis of feature-based techniques, machine learning techniques, and deep-learning-based methods. This comparison is made on the Corel-1K dataset, which contains 1000 images of 10 different classes.

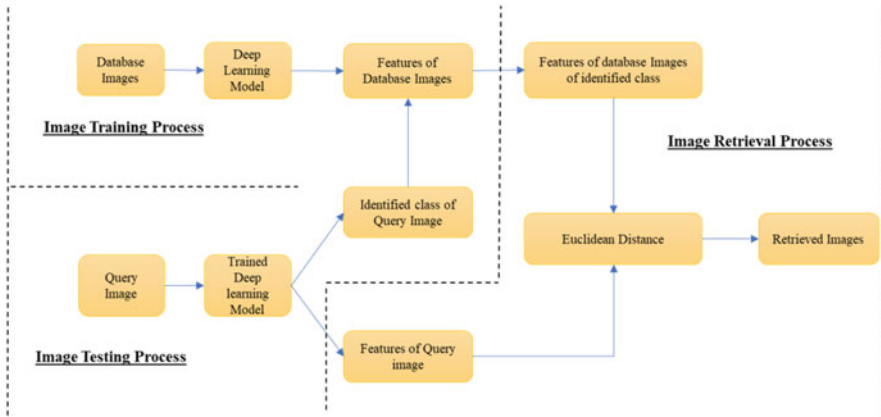


Fig. 27 Deep-learning-based image retrieval system

Table 3 Comparative analysis of feature-based, machine-learning-based, and deep-learning-based techniques

Measurement parameters	Average precision rate (%)
Color moments [3]	47.7
Color moments+SVM [13]	86.2
Color moments+GLCM+SVM [13]	89.6
Multilayer CNN model [38]	95.32

As shown in Table 3, the deep-learning-based multilayer CNN model provides better results than the feature-based and machine-learning-based techniques.

2.4 Application of CBIR System

CBIR system can be utilized in many applications. Online image searching using the image from the extensive image database is one of the CBIR applications. The CBIR system can also help to find details about the unknown entity or unknown places by retrieving similar images and relevant information. CBIR can also be useful for forensic science for retrieving images of the identical body organ from the forensic image depository. Similarly, it is helpful for criminal identification by retrieving the images from the criminal database. CBIR can also be useful for Trademark image registration. The new trademark is compared with the existing trademarks to ensure no risk of confusion. Copyright protection is also a potentially important application area. The CBIR system also finds its application in the domain of architectural and engineering design. The use of stylized 2-D and 3-D models to represent design objects is needed to manage. The designer needs to be aware of previous methods. Hence, the ability to search design archives of earlier examples that are similar or

meet specified suitability criteria can be valuable. The CBIR system is also helpful for the fashion and interior designer. To find fabrics of a specific combination of color and texture is increasingly recognized as a useful tool for the design process. The CBIR system helps teachers and students to retrieve suitable teaching materials. The image of any movie can help to identify the movie or to retrieve the whole movie. The most popular application of the CBIR is content-based medical image retrieval (CBMIR).

2.4.1 Content-Based Medical Image Retrieval (CBMIR)

Due to the medical imaging technology upgradation, many medical images are generated worldwide in many hospitals, clinics, and laboratories. These images can be CT scan, X-ray images, MRI images, etc. These images can be helpful for the doctors to retrieve similar cases from the record. These historical cases can help the doctor to diagnose the diseases and to identify which treatment will be better in the patient's current condition. The image retrieval system for medical images is known as content-based medical image retrieval (CBMIR). The meaning of similar image retrieval is that the images of the same type of disease, severity, treatment, and stage. Medical image retrieval is used for mainly three domains: teaching, research, and diagnostics.

The professors can utilize the large medical image database to search for meaningful and exciting cases to represent among the medical students. The retrieved similar images may have different diagnoses. In this way, students can learn about the various treatments for a similar type of disease. It improves the education quality. In the online-based teaching system, the CBMIR can help students by retrieving similar images and cases. The researcher can also take the help of the CBMIR system for the study purpose and the invention of the new treatment and medicines. The significant advantage in the domain of medical science is the diagnose of the disease. It will help doctors make a clinical decision based on the past treatment by retrieving similar cases and images for the unknown query image. The major challenge in CBMIR systems is the semantic gap between the low-level visual information captured by imaging devices and high-level semantic information perceived by humans [39]. The various medical areas where the CBMIR system can be helpful are described as below:

- **MRI Images of Brain**

- A combination of closed-form metric learning (CFML) and VGG19 model was used to train and retrieve the brain tumor's images. The dataset comprises three types of images: pituitary, meningioma, and glioma tumor. CFML is very efficient memory and computationwise due to its small dimension. It works well for small medical image datasets [40]. The CBMIR system to retrieve the brain tumor's image is shown in Fig. 28.
- Multi-channel 2D CNN model was utilized for the brain image classification. The 3D-CNN model's accuracy was better than the 2D-CNN model, but it

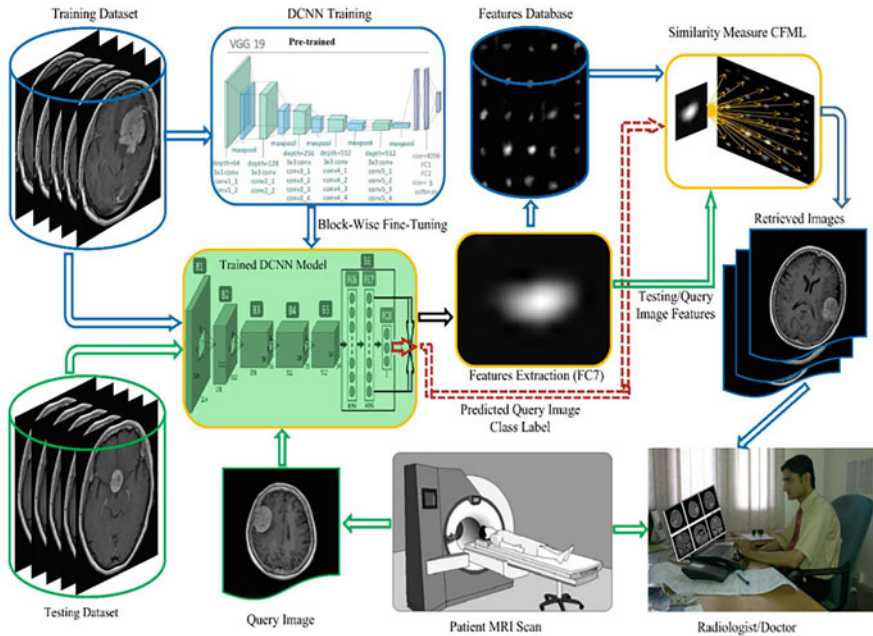


Fig. 28 CBMIR system for brain tumor’s image retrieval[40]

was computationally costly due to extra dimension. M2D-CNN archives better accuracy with fewer numbers of parameters compared to the 3D-CNN model [42].

- CBMIR was useful for the early detection of Alzheimer’s disease (AD). It is an irreversible disorder of the brain related to memory loss. It is usually seen in the elderly and aged people. Even for small datasets, the CapsNet was capable of fast learning. It can efficiently handle the transition and rotation of the image. It was observed that a collective method of a CNN with 3D-autoencoder and 3D-CapsNets increased the detection performance compared to the Deep-CNN method alone [43].

• **Retinal Diabetic Retinopathy**

- Color-histogram-based CBMIR system was used to detect diabetic retinopathy. It helps to prevent blindness. The result has been compared for the HSV and RGB color histogram. The performance of the HSV color histogram is better than the RGB color histogram for diabetic retinopathy detection [44].
- Radial inverse force histograms technique was used to detect the diabetic eyes. CBMIR system helped to diagnosis the retinal disorder to effectively separate the diabetic retinopathy patients [45].
- A combination of multiple pre-trained model can also be utilized to take benefits of individual pre-trained model’s ability. A combination of DenseNet,

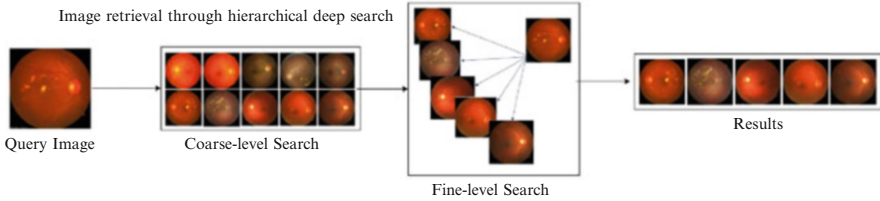


Fig. 29 CBMIR system for retinal image retrieval [41]

ResNet, and VGG16 was used to retrieve similar retinal images for the treatment based on the severity stage of the disease [41]. The CBMIR system for retinal image is shown in Fig. 29.

- **X-ray images**

- IRMA dataset contains 14410 X-ray images of 193 classes. X-ray images were retrieved based on the X-ray image of the body organ. IRMA dataset is challenging because of its imbalance; some image classes have hundreds of examples, whereas others only have a few. The deep CNN network and Radon transform were used to retrieve the similar X-ray images of semantic body parts. Radon transformation is very efficient for shrinking the search space and also for a retrieval system [46, 47].
- Based on the X-ray images, gender of the person can be detected. CNN model was trained from scratch to identify the gender from the X-ray images of the spine. Performance of the pre-trained model DenseNet was better than the CNN model because the bias and weights of the DenseNet models were not random, while the initial weights and bias of the CNN model developed from the scratch were random [48].
- The X-ray image can also help to identify the COVID-19. The CBMIR system was utilized to classify the unknown images into the normal, pneumonia, and COVID-19 classes. The fusion of Xception and ResNet50 pre-trained models was utilized to identify the COVID-19 from the chest X-ray images [49].

- **Skin images**

- The skin disease can also be identified using the CBMIR system. Dermoscopic images were utilized for skin lesion analysis and the detection of melanoma. Skin surface photos were utilized for four diseases, including eczema, heatrash, subitum, and varicella. The multi-channel ResNet50 models were utilized to identify the various skin diseases [50].

Hence, the CBMIR technology can be utilized in many medical areas such as retinal diabetic retinopathy, X-ray images, MRI images for brain tumor detection, COVID-19 detection, mammography, etc.

3 Conclusion and Future Scope

In this chapter, a literature survey regarding different methods of CBIR system using feature-based, machine-learning-based, and deep-learning-based methods is presented. The CBIR domain's primary challenge is the semantic gap, which is the gap between the low-level feature extracted from the image and the high-level concept understood by the human. The feature-based techniques retrieved images based on the handcrafted feature. Identifying the useful feature for a specific dataset is difficult. Hence, the feature-based methods cannot effectively reduce this semantic gap. The machine-learning-based techniques learn from the provided input–output training pair. But it also used handcrafted features for the training of the machine. Hence, it has moderately reduced the semantic gap. The deep learning techniques are an end-to-end process; therefore, it automatically extracts the necessary high-level features from the provided input image. Hence, deep learning methods reduced the semantic gap to a significant extent. Various pre-trained deep learning models are available to reduce the training time. CBIR system has a wide range of applications. CBMIR is one of the applications that get more attention due to the doctor's supportive tool to diagnose the disease. CBIR also has scope in many applications such as online searching, crime prevention, military application, intellectual property, architectural design, engineering design, fashion design, interior design, journalism, advertising, education, training, cultural heritage, and medical diagnosis, and advertising, education, training, cultural heritage, and medical diagnosis.

4 Copyright Statement

Please be aware that the use of this L^AT_EX 2_ε class file is governed by the following conditions.

4.1 Copyright

The Copyright licensed to EAI.

Acknowledgments This class file was developed by Sunrise Setting Ltd, Torquay, Devon, UK. Website: www.sunrise-setting.co.uk

References

1. DEVARAJ, A.F.S., MURUGABOOPATHI, G., ELHOSENY, M., SHANKAR, K., MIN, K., MOON, H. and JOSHI, G.P. (2020) An Efficient Framework for Secure Image Archival and Retrieval System Using Multiple Secret Share Creation Scheme. *IEEE Access* **8**: 144310–144320. <https://doi.org/10.1109/ACCESS.2020.3014346>.

2. AFIFI, A.J. and ASHOUR, W.M. (2012) Content-based image retrieval using invariant color and texture features. *2012 International Conference on Digital Image Computing Techniques and Applications, DICTA 2012* <https://doi.org/10.1109/DICTA.2012.6411665>.
3. BHAGAT, A.P. and ATIQUE, M. (2012) Design and development of systems for image segmentation and content based image retrieval. *Proceedings - 2012 2nd National Conference on Computational Intelligence and Signal Processing, CISP 2012* : 109–113. <https://doi.org/10.1109/NCCISP.2012.6189688>.
4. YUE, J., LI, Z., LIU, L. and FU, Z. (2011) Content-based image retrieval using color and texture fused features. *Mathematical and Computer Modelling* **54**(3–4): 1121–1127. <https://doi.org/10.1016/j.mcm.2010.11.044>.
5. ERKUT, U., BOSTANCI OGLU, F., ERTEN, M., OZBAYOGLU, A.M. and SOLAK, E. (2019) HSV Color Histogram Based Image Retrieval with Background Elimination. *1st International Informatics and Software Engineering Conference: Innovative Technologies for Digital Transformation, IISEC 2019 - Proceedings* <https://doi.org/10.1109/UBMYK48245.2019.8965513>.
6. ANANDABABU, P. and KAMARASAN, M. (2020) An Effective Content Based Image Retrieval Model using Improved Memetic Algorithm. *Proceedings of the 5th International Conference on Inventive Computation Technologies, ICICT 2020* : 424–429. <https://doi.org/10.1109/ICICT48043.2020.9112503>.
7. KOUR, N. and GONDHI, N. (2019) Assessment on various Approaches for Content Based Image Retrieval. *Proceedings of the 3rd International Conference on Inventive Systems and Control, ICISC 2019 (ICISC)*: 225–230. <https://doi.org/10.1109/ICISC44355.2019.9036378>.
8. NAZIR, A., ASHRAF, R., HAMDANI, T. and ALI, N. (2018) Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor. *2018 International Conference on Computing, Mathematics and Engineering Technologies: Invent, Innovate and Integrate for Socioeconomic Development, iCoMET 2018 - Proceedings 2018-January*: 1–6. <https://doi.org/10.1109/ICOMET.2018.8346343>.
9. SHINDE, S. (2018) MULTI-SEQUENTIAL SEARCH. *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*: 973–978.
10. WALIA, E., VESAL, S. and PAL, A. (2014) An Effective and Fast Hybrid Framework for Color Image Retrieval. *Sensing and Imaging* **15**(1). <https://doi.org/10.1007/s11220-014-0093-9>.
11. ARTEMI, M. and LIU, H. (2020) Image Optimization using Improved Gray-Scale Quantization for Content-Based Image Retrieval. *6th International Conference on Optimization and Applications, ICOA 2020 - Proceedings* <https://doi.org/10.1109/ICOA49421.2020.9094507>.
12. SINGH, J., BAJAJ, A., MITTAL, A., KHANNA, A. and KARWAYUN, R. (2018) Content Based Image Retrieval using Gabor Filters and Color Coherence Vector. *Proceedings of the 8th International Advance Computing Conference, IACC 2018* : 290–295. <https://doi.org/10.1109/IADCC.2018.8692123>.
13. KAPADIA, M.R. and PAUNWALA, C.N. (2018) Analysis of SVM kernels for content based image retrieval system. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing, ICECDS 2017* : 1409–1414. <https://doi.org/10.1109/ICECDS.2017.8389676>.
14. BOUSSAAD, L. (2019) Content Based Image Retrieval Using Wavelet Moments and Local Binary Patterns in CIE-Lab Color Space. *2018 International Conference on Signal, Image, Vision and their Applications, SIVA 2018* <https://doi.org/10.1109/SIVA.2018.8661155>.
15. AGRAWAL, S., VERMA, N.K., TAMRAKAR, P. and SIRCAR, P. (2011) Content based color image classification using SVM. *Proceedings - 2011 8th International Conference on Information Technology: New Generations, ITNG 2011* : 1090–1094. <https://doi.org/10.1109/ITNG.2011.202>.
16. ALRAHHAL, M. and SUPREETHI, K.P. (2019) Content-Based Image Retrieval using Local Patterns and Supervised Machine Learning Techniques. *2019 Amity International Conference on Artificial Intelligence (AICAI)* : 118–124.
17. NARASIMHA, Y.R., PAVITHRA, L.K. and SREE, S.T. (2018) Analysis of Supervised and Unsupervised Learning in Content Based Multimedia Retrieval. *2nd International Conference on Computer, Communication, and Signal Processing: Special Focus on Technology and*

- Innovation for Smart Environment, ICCCS 2018 (ICCCSP):* 1–5. <https://doi.org/10.1109/ICCCSP.2018.8452821>.
18. VANI, R., VYAS, T. and TAHILRAMANI, N. (2019) CBIR using SVM, genetic algorithm, neural network, fuzzy logic, neuro-fuzzy technique: A survey. *Proceedings of the 2018 International Conference on Communication, Computing and Internet of Things, IC3IoT 2018* : 239–242. <https://doi.org/10.1109/IC3IoT.2018.8668197>.
 19. CHANG, R.I., LIN, S.Y., HO, J.M., FANN, C.W. and WANG, Y.C. (2012) A novel content based image retrieval system using K-means/KNN with feature extraction. *Computer Science and Information Systems* **9**(4): 1645–1661. <https://doi.org/10.2298/CSIS120122047C>.
 20. JAIN, M. and SINGH, S.K. (2018) An Efficient Content Based Image Retrieval Algorithm Using Clustering Techniques For Large Dataset. *2018 4th International Conference on Computing Communication and Automation (ICCCA)* : 1–5.
 21. SERRANO-TALAMANTES, J.F., AVILÉS-CRUZ, C., VILLEGAS-CORTEZ, J. and SOSSA-AZUELA, J.H. (2013) Self organizing natural scene image retrieval. *Expert Systems with Applications* **40**(7): 2398–2409. <https://doi.org/10.1016/j.eswa.2012.10.064>.
 22. RIAN, Z., CHRISTANTI, V. and HENDRYLI, J. (2019) Content-Based Image Retrieval using Convolutional Neural Networks. *Proceedings - 2019 IEEE International Conference on Signals and Systems, ICSigSys 2019* : 1–7. <https://doi.org/10.1109/ICSIGSYS.2019.8811089>.
 23. ALZU'BI, A., AMIRA, A. and RAMZAN, N. (2017) Content-based image retrieval with compact deep convolutional features. *Neurocomputing* **249**: 95–105. <https://doi.org/10.1016/j.neucom.2017.03.072>.
 24. FU, R., LI, B., GAO, Y. and PING, W. (2017) Content-based image retrieval based on CNN and SVM. *2016 2nd IEEE International Conference on Computer and Communications, ICC 2016 - Proceedings* : 638–642. <https://doi.org/10.1109/CompComm.2016.7924779>.
 25. BHANDI, V. and SUMITHRA DEVI, K.A. (2019) Image Retrieval by Fusion of Features from Pre-trained Deep Convolution Neural Networks. *1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering, ICATIECE 2019* : 35–40. <https://doi.org/10.1109/ICATIECE45860.2019.9063814>.
 26. GUPTA, A., AGARWAL, D., VEENU and BHATIA, M.P. (2019) Performance analysis of content based image retrieval systems. *2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018* : 899–902. <https://doi.org/10.1109/GUCON.2018.8675107>.
 27. RAMANJANEYULU, K., SWAMY, K.V. and RAO, C.H. (2018) Novel CBIR System using CNN Architecture. *Proceedings of the 3rd International Conference on Inventive Computation Technologies, ICICT 2018* : 379–383. <https://doi.org/10.1109/ICICT43934.2018.9034389>.
 28. WANG, L. and WANG, X. (2017) Model and metric choice of image retrieval system based on deep learning. *Proceedings - 2016 9th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics, CISP-BMEI 2016* : 390–395. <https://doi.org/10.1109/CISP-BMEI.2016.7852742>.
 29. LIU, F., WANG, Y., WANG, F.C., ZHANG, Y.Z. and LIN, J. (2019) Intelligent and Secure Content-Based Image Retrieval for Mobile Users. *IEEE Access* **7**: 119209–119222. <https://doi.org/10.1109/access.2019.2935222>.
 30. SIMONYAN, K. and ZISSERMAN, A. (2015) Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* : 1–14. <http://arXiv:1409.1556v6arXiv:1409.1556v6>.
 31. KUMAR, V., TRIPATHI, V. and PANT, B. (2020) Content based fine-grained image retrieval using convolutional neural network. *2020 7th International Conference on Signal Processing and Integrated Networks, SPIN 2020* : 1120–1125. <https://doi.org/10.1109/SPIN48934.2020.9071334>.
 32. DAS, R., KUMARI, K., MANJHI, P.K. and THEPADE, S.D. (2019) Ensembling Hand-crafted Features to Representation Learning for Content Based Image Classification. *2019 IEEE Pune Section International Conference, PuneCon 2019* : 1–4. <https://doi.org/10.1109/PuneCon46936.2019.9105759>.

33. HE, K., ZHANG, X., REN, S. and SUN, J. (2016) Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2016-Decem**: 770–778. <https://doi.org/10.1109/CVPR.2016.90>. <http://1512.033851512.03385>.
34. HE, K., ZHANG, X., REN, S. and SUN, J. (2016) Identity mappings in deep residual networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **9908 LNCS**: 630–645. https://doi.org/10.1007/978-3-319-46493-0_38. 1603.05027.
35. HOWARD, A.G., ZHU, M., CHEN, B., KALENICHENKO, D., WANG, W., WEYAND, T., ANDREETTO, M. *et al.* (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications <http://arxiv.org/abs/1704.04861>. 1704.04861.
36. SANDLER, M., HOWARD, A., ZHU, M., ZHMOGINOV, A. and CHEN, L.C. (2018) MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* : 4510–4520. <https://doi.org/10.1109/CVPR.2018.00474>. 1801.04381.
37. HUANG, G., LIU, Z., VAN DER MAATEN, L. and WEINBERGER, K.Q. (2017) Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* **2017-Janua**: 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>. 1608.06993.
38. KAPADIA, M.R. and PAUNWALA, C.N. (2018) Improved CBIR system using multilayer CNN. *2018 International Conference on Inventive Research in Computing Applications, ICIRCA 2018* : 840–845. DOI 10.1109/ICIRCA.2018.8597199.
39. QAYYUM, A., ANWAR, S.M., AWAIS, M. and MAJID, M. (2017) Medical image retrieval using deep convolutional neural network. *Neurocomputing* **266**: 8–20. <https://doi.org/10.1016/j.neucom.2017.05.025>.
40. SWATI, Z.N.K., ZHAO, Q., KABIR, M., ALI, F., ALI, Z., AHMED, S. and LU, J. (2019) Content-Based Brain Tumor Retrieval for MR Images Using Transfer Learning. *IEEE Access* **7**(c): 17809–17822. <https://doi.org/10.1109/ACCESS.2019.2892455>.
41. WIJESINGHE, I., GAMAGE, C. and CHITRARANJAN, C. (2019) Deep supervised hashing through ensemble CNN feature extraction and low-rank matrix factorization for retinal image retrieval of diabetic retinopathy. *Proceedings - 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering, BIBE 2019* : 301–308. <https://doi.org/10.1109/BIBE.2019.00061>.
42. HU, J., KUANG, Y., LIAO, B., CAO, L., DONG, S. and LI, P. (2019) A Multichannel 2D Convolutional Neural Network Model for Task-Evoked fMRI Data Classification. *Computational Intelligence and Neuroscience* **2019**(i). <https://doi.org/10.1155/2019/5065214>.
43. KRUTHIKA, K.R., RAJESWARI and MAHESHAPPA, H.D. (2019) Erratum: CBIR system using Capsule Networks and 3D CNN for Alzheimer's disease diagnosis (Informatics in Medicine Unlocked (2019) 14 (59–68), (S235291481830176X), (10.1016/j.imu.2018.12.001)). *Informatics in Medicine Unlocked* **16**(August). <https://doi.org/10.1016/j.imu.2019.100227>.
44. KASHYAP, N. and SINGH, D.K. (2017) Color histogram based image retrieval technique for diabetic retinopathy detection. *2017 2nd International Conference for Convergence in Technology, I2CT 2017* **2017-Janua**: 799–802. <https://doi.org/10.1109/I2CT.2017.8226238>.
45. KIMPAN, S., MANERAT, N. and KIMPAN, C. (2018) Diabetic retinopathy image analysis using radial inverse force histograms. *ICIIBMS 2017 - 2nd International Conference on Intelligent Informatics and Biomedical Sciences* **2018-Janua**: 266–271. DOI 10.1109/ICIIBMS.2017.8279708.
46. LIU, X., TIZHOOSH, H.R. and KOFMAN, J. (2016) GENERATING BINARY TAGS FOR FAST MEDICAL IMAGE RETRIEVAL Department of Systems Design Engineering University of Waterloo, Waterloo, ON, Canada N2L 3G1 Centre for Bioengineering and Biotechnology University of Waterloo, Waterloo, ON, Canada N2L 3G1 : 2872–2878.
47. KHATAMI, A., BABAIE, M., KHOSRAVI, A., TIZHOOSH, H.R., SALAKEN, S.M. and NAHAVANDI, S. (2017) A deep-structural medical image classification for a Radon-based image retrieval. *Canadian Conference on Electrical and Computer Engineering* : 17–20. <https://doi.org/10.1109/CCECE.2017.7946756>.

48. XUE, Z., RAJARAMAN, S., LONG, R., ANTANI, S. and THOMA, G. (2018) Gender Detection from Spine X-Ray Images Using Deep Learning. *Proceedings - IEEE Symposium on Computer-Based Medical Systems* **2018-June**: 54–58. <https://doi.org/10.1109/CBMS.2018.00017>.
49. RAHIMZADEH, M. and ATTAR, A. (2020) A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. *Informatics in Medicine Unlocked* **19**: 100360. <https://doi.org/10.1016/j.imu.2020.100360>.
50. GUO, S. and YANG, Z. (2018) Multi-Channel-ResNet: An integration framework towards skin lesion analysis. *Informatics in Medicine Unlocked* **12**(June): 67–74. <https://doi.org/10.1016/j.imu.2018.06.006>

Data Analytics on Medical Images with Deep Learning Approach



S. Saravanan , K. Surendheran , and K. Krishnakumar 

1 Introduction

Data collected in the Healthcare domain were varied in types such as images, sensors text representation, and electronic health records. Temporal data mining and clinical prediction models are found to be the recent advancement in medical healthcare analytics. Multiple data sources in healthcare set the need for a wide variety of techniques drawn toward data analytics. Electronic health records (EHRs) are one of the popular data sources used in the medical field. The EHR is noted to be the digitized form of the patient's medical history. The EHR contains the patient's physician observations, history of the illness, lab data, different radiological reports, graphs, and even billing data. The EHR directly provides an immediate access to the patient's medical data from the individual and the organization on a real-time basis. It also brings much more quality improvement and convenience of patient participation in healthcare management. Also, it increases the accuracy of diagnosis and is a great outcome in the health monitoring system.

Anatomical internal structures in human beings can be projected as high-quality medical images. Biomedical imaging plays a vital role in helping the physician identify the disease and treatment planning. Different medical imaging modalities are involved in retrieving the clinical region as an image or a slice of images. Magnetic resonance imaging (MRI) and computed tomography (CT) are found to be the popular modalities in medical imaging. However, physicians opt to use these imaging data as a primary step to identify the disease or the problem that occurred with the internal human part. The analysis's main phase includes obtaining the quantitative information and making implications over the medical image for more insights on the patient's condition. Medical data analysis has a significant social

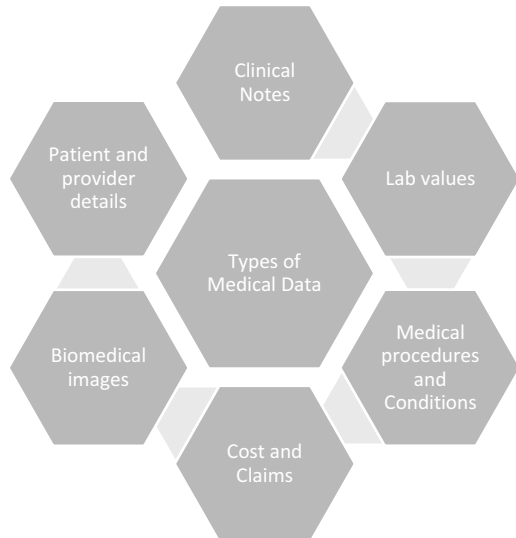
S. Saravanan (✉) · K. Surendheran · K. Krishnakumar
VIT school of Design, Vellore Institute of Technology, Vellore, Tamil Nadu, India

implication as it is the main role in understanding biological systems and solving health problems.

2 Healthcare Data Sources and Analytics

Electrocardiogram (ECG) and electroencephalogram (EEG) are medical data collection instruments that work on sensors that collect signals from several parts of the human body. Retrospective analysis is made using the collected medical sensor data. Biomedical signal analysis is also one of the important data monitoring systems used in the healthcare unit. It measures the signals from different biological sources through different physiological processes. Social media data are found to be an effective representation of the people's psychological behavior, which can be mined to obtain inferences about population health and its monitoring system. Text mining methods were an innovative way of applying new knowledge discovery methods in the medical field. These use the long-duration preservation and mobility methods of digitally available resources for retaining the scientific literature. As an important form of medical data, clinical notes are encoded with information about patients, which is also called the healthcare data's backbone [1]. The set of clinical notes is classically stored as an unstructured data format. Figure 1 details the types of medical data used in healthcare department.

Fig. 1 Types of medical data used in healthcare



3 Background and Motivation

Data analytics is found to be a rapidly growing field in the evolution of healthcare research and practices. As the recent trend toward healthcare data analytics brings in more implementation on hardware and software technologies, it can increase the ease of the data collection process. Among the different healthcare data, medical images were found to be the most used data for diagnosis, assessment, and planning [2]. Medical image data vary from few megabytes (MB) for a single slice of an image to hundreds of megabytes per slice. A huge data storage device is required in order to maintain the digital images for a longer-term. Moreover, it needs an accurate and fast processing algorithm for decision-supporting automation to be used by the data. Image-processing applications like image enhancement, image segmentation, and many more are implemented in those algorithms with learning-based methods. As the data dimension and its size become more, analyzing and interacting with the algorithm to obtain high accuracy with less computational time reflect an effective method involving an efficient technique [3].

Incorporating algorithm-based analysis with a low-dimensional size medical data has the insight to help physicians improve diagnostic accuracy. Combining medical image data with EHR data and genomic data advances accuracy and reduces the time taken to diagnose. As the collection of slices in the medical images is growing exponentially as an instance, the medical image dataset (Image CLEF) had around 60,000 images for experimentation and analysis around the year 2005 and 2007, and it had grown up to 4,50,000 images in the year of 2018, which were stored every day. The vast set of medical data volumes requires an efficient compression technique, overcoming data storage limitations and network bandwidth. A lossy image compression methodology results in data loss that cannot be used in medical image data. For effective retrieval of medical data, a lossless image compression methodology needs to be brought in to maintain fidelity and information preservation, which are very important.

Several transform-based algorithms were proposed to attain the lossless image compression [4] with medical image data. Traditional algorithms such as Wavelet Transform [5], Discrete Cosine Transform [6], JPEG [7] are widely used for medical image data analysis. To achieve the compression of medical image data without degrading the details present over them, a visually lossless or a lossless compression is attained with a curious hybrid combination of algorithms. Machine learning algorithms [8] are efficient in achieving a desired resultant value in medical image data. Automated image analysis methodology using machine learning attains a rich quality improvement [9]. Deep learning methodology brings a state of art accuracy over the machine learning method on facilitating identification or any diagnosis application [10]. Due to the advancement of deep learning algorithms, research has identified that deep learning-based algorithms will be implemented on all the state of day-to-day activities over the next few years.

Deep learning-based medical image compression achieves a tremendous reach through different algorithms such as CNN [11], autoencoders [12], etc. The

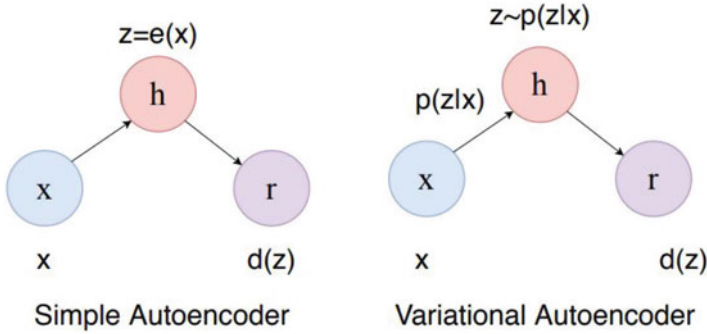


Fig. 2 Architecture of simple autoencoders and variational autoencoders

convolutional neural network algorithm has been implemented in many medical diagnosis experimentations. Autoencoders are found to be efficient in dimensionality reduction applications, which are constructed with three components such as encoders, codes, and decoders. When the input has a higher dimension than the code, then it is said as undercomplete. In contrast, when the input has less dimension than the code, it is said to be overcomplete. A variational autoencoder [13] is found to be a category in the type of autoencoder. Architecture difference comparing the simple autoencoder and the variational autoencoder is depicted in Fig. 2.

Dimensionality reduction is a method for reducing the number of features with some data that can be a subset of the main feature or a combined feature states an encoder's process. Autoencoders [14] and their categories are general neural network architectures combined with an encoder and a decoder to generate a bottleneck to observe the data. During the training process, it is trained to lose a minimal quantity, generally the gradient descent iterations for reducing the reconstruction error of information during the encoding–decoding process. An autoencoder's latent space/code space can be extraordinarily irregular or meaningless due to an autoencoder's overfitting process. Variational autoencoders (VAEs) are a type of autoencoders that solve the latent space irregularity problem. It works on the process of assigning the encoder to return distribution to the latent space. Involving the returned distribution on the architecture achieves a better representation in the code space. When comparing, an autoencoder is found to be a deterministic factor and the variational autoencoder is probabilistic based.

4 Methodology

In our model, we present a solution for the challenges and a novel framework for medical image reconstruction is proposed. For efficient medical image reconstruction, the variational autoencoder is used for testing, and restricted Boltzmann machines architecture for efficient training is implemented in this proposed model.

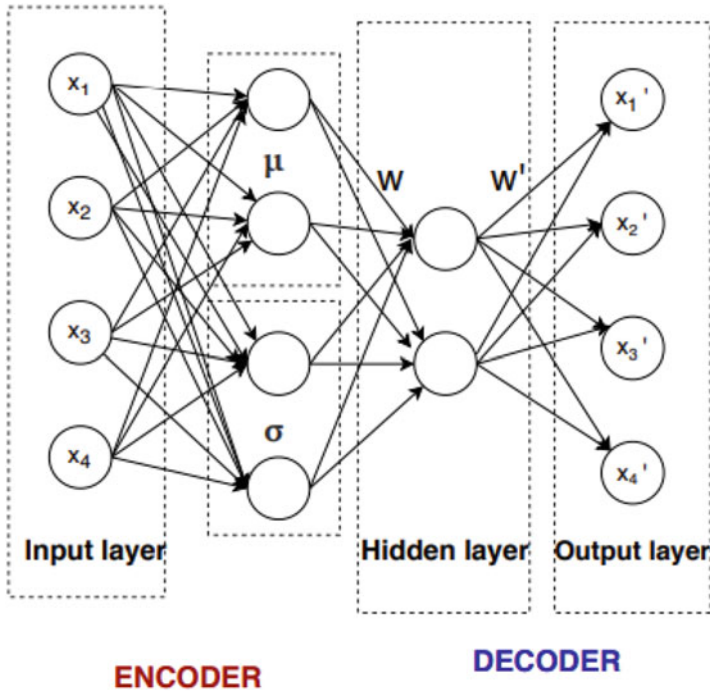


Fig. 3 Single-layer variational autoencoder architecture

A single layer variational autoencoder architecture is depicted in Fig. 3. Variational autoencoders (VAEs) work on the unique method of separating its property, which helps the generative modeling. The hidden layer, also called a latent space, works in continuity and allows random sampling and interpolation. The working module of the VAE includes two vectors of size (n) from the encoder with the vector of means (μ) and vector of standard deviations (σ).

4.1 Problem Formulation

In our model, medical images that are used for various diagnosis problems are considered for significant compression. We reformulate and segment the medical images as region of interest and non-region of interest using the contextual method separation. For the various tasks required in medical images, an automated image analysis tool based on machine learning is needed for improving the quality of image diagnosis with less storage space. Deep learning is an extensively applied technique that provides state-of-the-art accuracy.

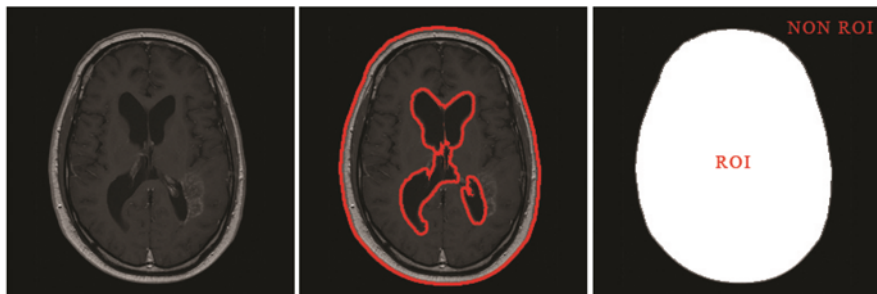


Fig. 4 Contextual selection on the sample medical image

4.2 Contextual Selection of Medical Images

Considered medical images from the medical database are segmented using the region based contour method for region of interest and non-region of interest separation. In our model, the medical image's clinical region is considered a region of interest area, and the background region is considered as a non-clinical region area. The region-based contour method has been implemented to obtain the proper region of the clinical area. The sample medical image considered for the contextual region separation is implemented in Fig. 4. To attain the effectiveness of reducing the computational complexity, it is implemented with hand-drawn free selection by the physicians.

5 Experimentation

5.1 Dataset

Medical image datasets are collected from the Harvard medical images database (<http://www.med.harvard.edu/AANLIB>). Some of the sample images considered for experimentation are illustrated in Fig. 5. The collected images are sampled with a wide variety of CT and MR images and trained the restricted Boltzmann machine architecture with 100 epochs. Different MRI images are collected as illustrated in Fig. 5 and are considered to have the dark shade region, bright shaded, T1 tissue, T2 tissue, bled, and plaque.

The images are tested using the multilayered variational Autoencoder model that is depicted in Fig. 6. In our model, a multilayered neural network is built, which gives an advantage of efficient representation in terms of retrieving the diagnostic details from medical images. From the multilayered VAE, a parameter of a vector is obtained for random variables of length n , with the i -th element of (μ) and (σ) , which represents the mean and standard deviation of an i -th random variable.

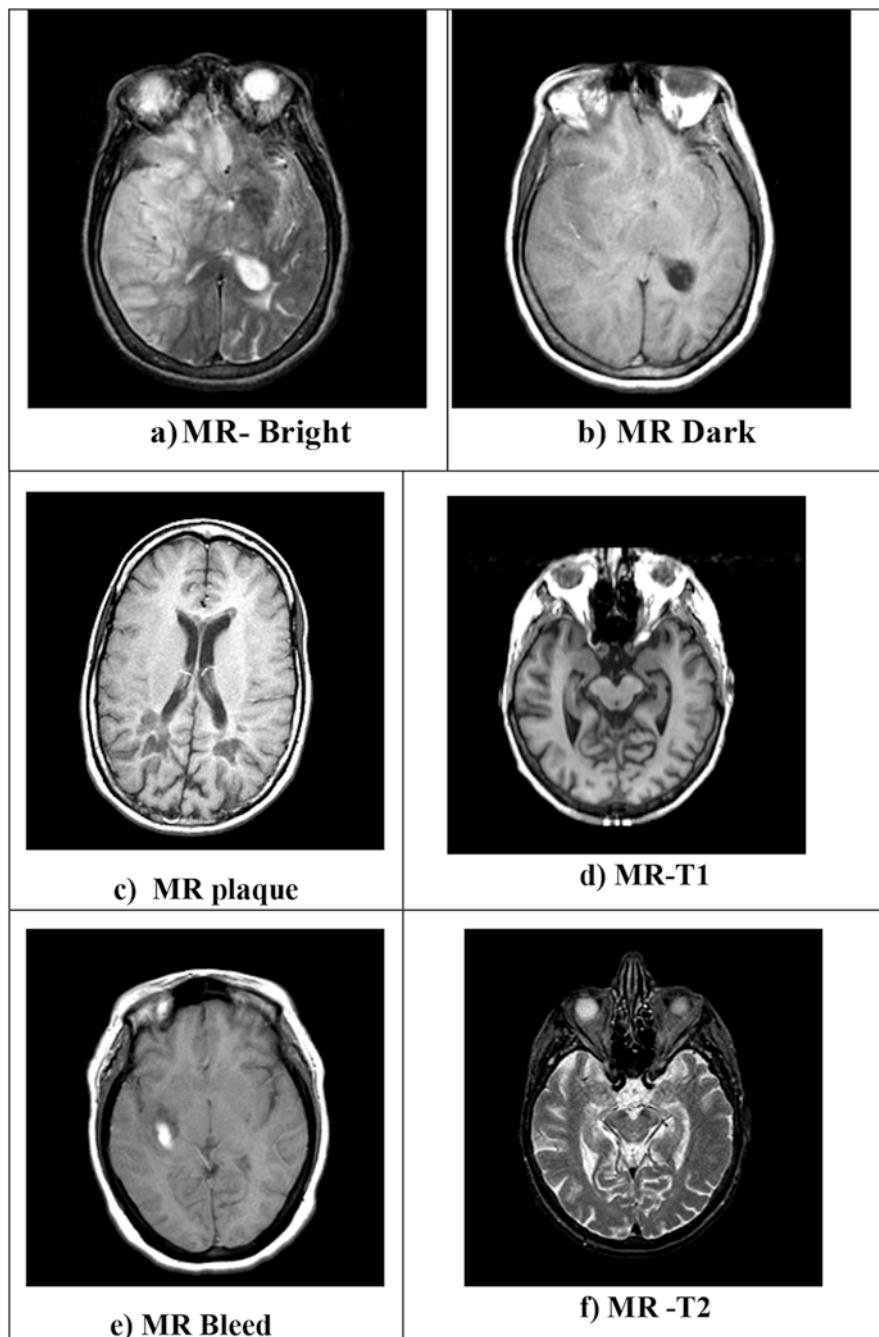
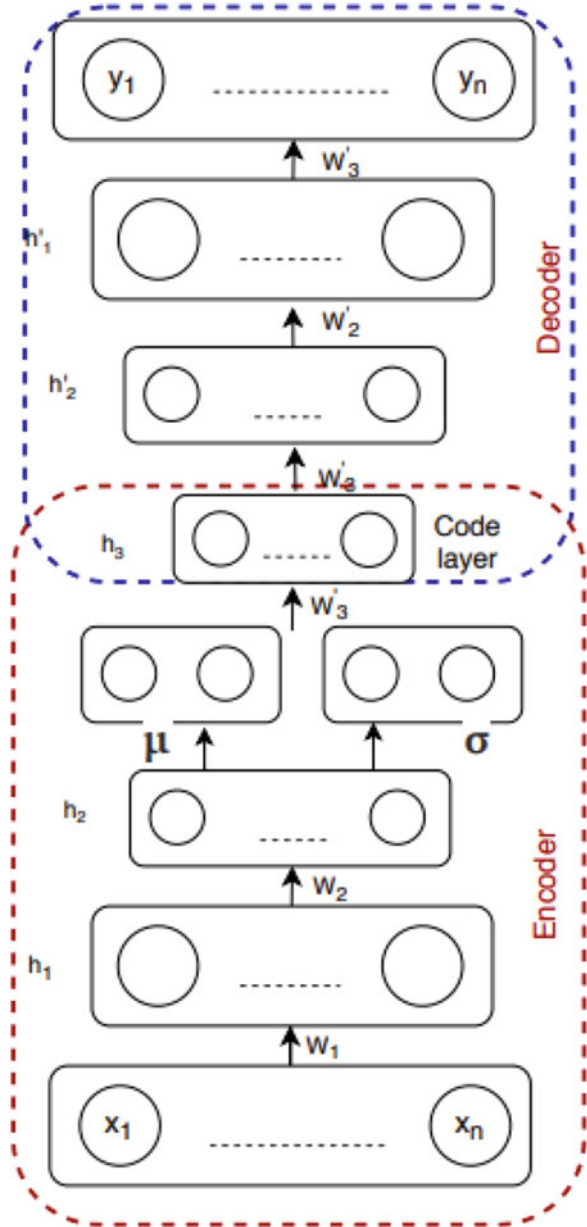


Fig. 5 Sample images from the Harvard medical image dataset

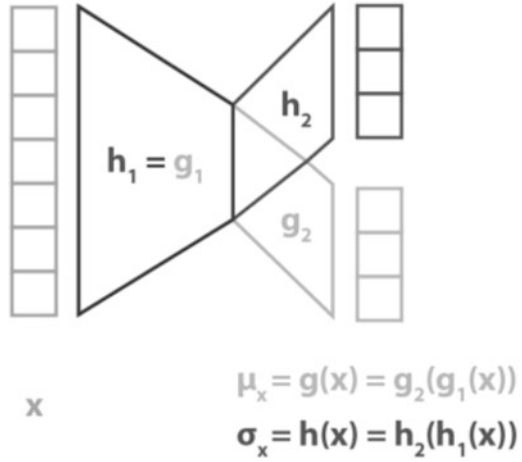
Fig. 6 Multilayered variational autoencoder architecture



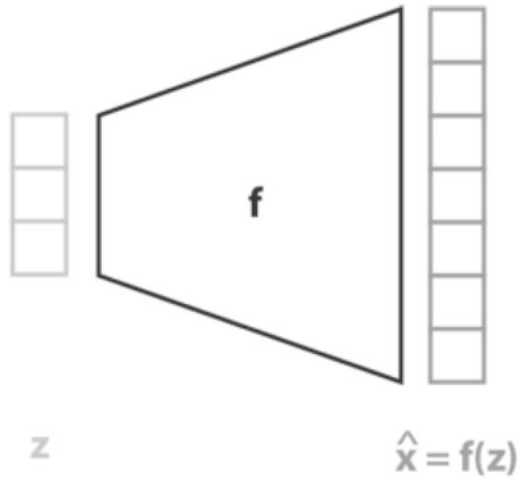
5.2 Model Selection and Training parameters

Figure 7 represents the methodology implemented in VAE and it specifies the process of encoders and decoders with the equations. Mean square error (MSE)

Fig. 7 Encoders and decoders in VAE architecture



a) Encoder



b) Decoder

and perceptual metric MS-SSIM to train the network with the loss function as represented in Eq. 1.

$$\text{Loss} = R + \lambda \times D \tag{1}$$

where D is the distortion measured as $\|x - \hat{x}\|_2$ for MSE or MS-SSIM, R is the entropy of latents \hat{y} and \hat{z} . λ controls the tradeoff between rate and distortion.

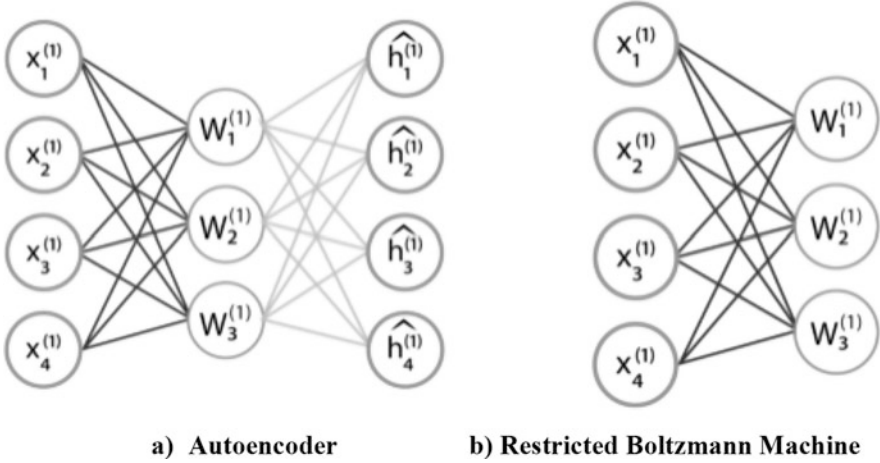


Fig. 8 Structure of an autoencoder with a restricted Boltzmann machine

For the pre-training process, the restricted Boltzmann machine architecture is implemented. We can observe that the added network can exactly enhance the compression quality with a considerable improvement, especially for the multi-scale encoder. However, when compared with the input image, the complexity of the proposed model achieves an enhanced output image with better quality. Figure 8 depicts the comparison process of AE implemented on the medical images to reconstruct using the pre-training architecture called restricted Boltzmann machine architecture.

Figure 9 depicts the pre-training phase with three hidden layers of learning stack of the restricted Boltzmann machine. As a stack is composed to form a single model, the layer copies are removed, and the total inputs coming into the first and second hidden layers are halved. The pre-training algorithm with a restricted Boltzmann machine with three layers is illustrated in Fig. 9.

5.3 Evaluation

Performance of the evaluations carried out on the medical images is calculated as peak signal noise ratio (PSNR), compression ratio (CR), bits per pixel (BPP), structure similarity index (SSIM), and computational time (CT). Equations of the different performance metrics are illustrated as follows:

$$\text{PSNR} = 10 * \log_{10} \left(\frac{255^2}{\sqrt{\text{MSE}}} \right) \quad (2)$$

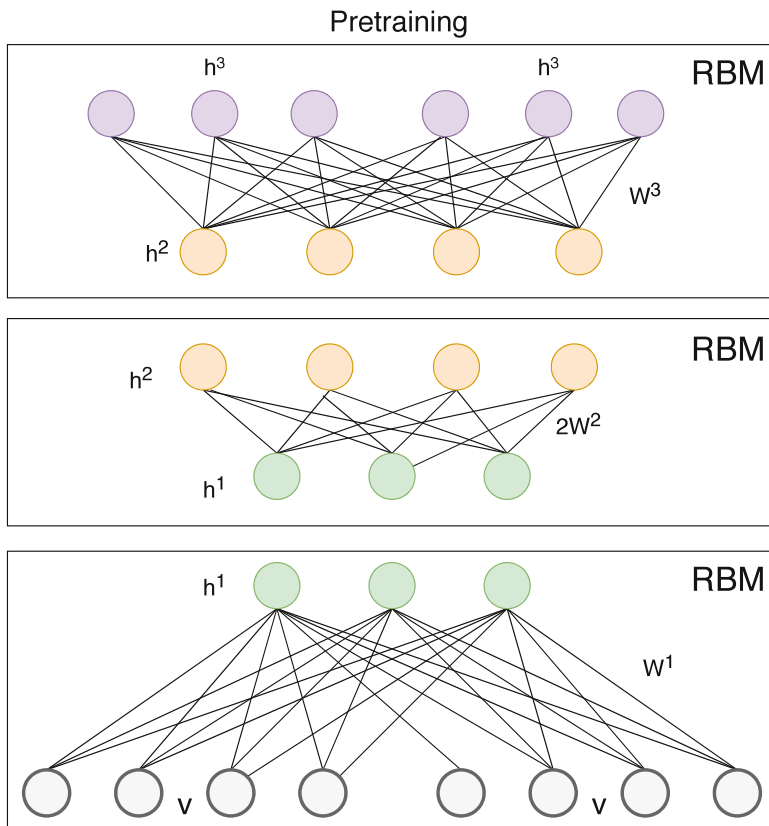


Fig. 9 Pre-training using the stack of RBM

$$MSE = \frac{1}{N} \times \sum_i \sum_j (f(x, y) - F(x, y))^2 \tag{3}$$

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\delta_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\delta_x^2 + \delta_y^2 + C_2)} \tag{4}$$

Based on the testing carried out with the different medical images, the performance of the proposed method is analyzed by comparing the existing architecture like deep autoencoders, back propagation networks, and convolutional neural network values are illustrated in Figs. 10 and 11.

As illustrated in Figs. 10 and 11, the proposed method outperforms when compared with the existing architecture for image reconstruction. The structured similarity index of the proposed method is high with an average of 0.989 using the proposed variational autoencoder. Sample compressed images with the input

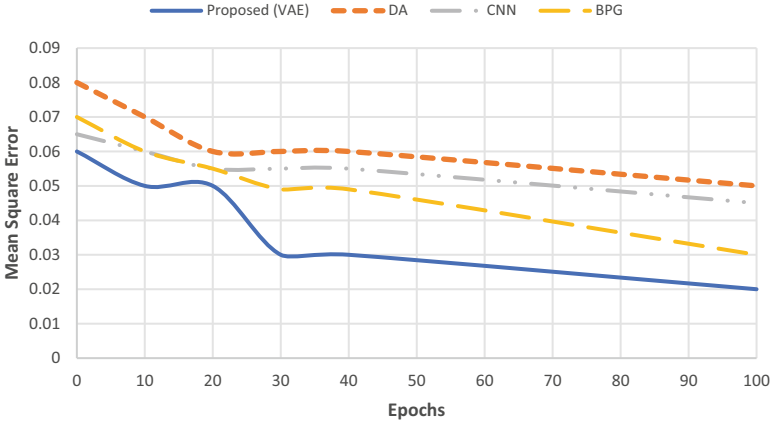


Fig. 10 Performance evaluation comparison among different architectures on MSE with the number of iterations

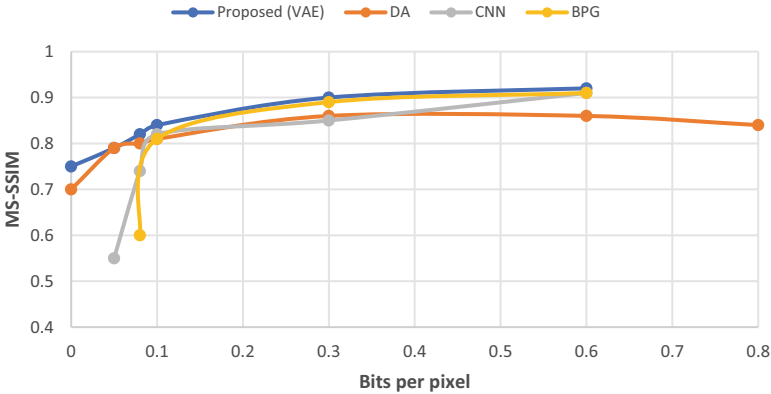


Fig. 11 Bits per pixel with SSIM – Distortion curve with other methods

images are illustrated in Fig. 12. Thus, results show that a variational autoencoder with a restricted Boltzmann machine architecture for pre-training attains an efficient medical image reconstruction.

6 Conclusion

A variational autoencoder proved to be the best autoencoder in achieving an efficient image reconstruction as compared to the existing state of the art architectures like backpropagation network, CNN, and deep autoencoders. The pre-training phase of using the stacked restricted Boltzmann machines helps attain an efficient image

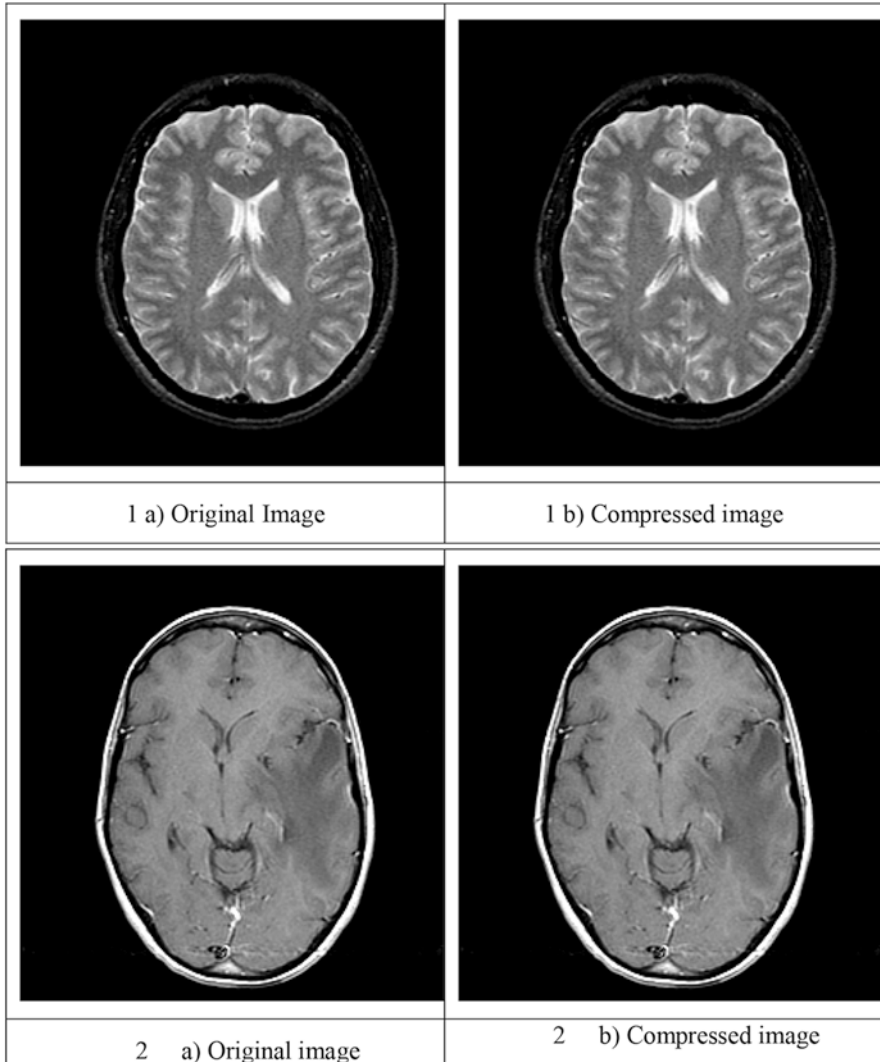


Fig. 12 Sample input medical images with compressed output images using VAE

output quality among the huge set of input data. Performance analysis proved that the proposed method outperforms in terms of PSNR, CR, and SSIM. Even though computational complexity was moderate as compared with the other architectures, resultant output image quality satisfies the subjective and objective analysis on the output images. As a future scope, other type of autoencoders can be tested and trained to analyze the best efficient algorithm for achieving dimensionality reduction.

References

1. O. Kulak, H. G. Goren, and A. A. Supciller, "A new multi criteria decision making approach for medical imaging systems considering risk factors," *Appl. Soft Comput. J.*, vol. 35, pp. 931–941, 2015, <https://doi.org/10.1016/j.asoc.2015.03.004>.
2. R. C. Gonzalez, R. E. Woods, and B. R. Masters, "Digital Image Processing, Third Edition," *J. Biomed. Opt.*, vol. 14, no. 2, p. 029901, 2009, <https://doi.org/10.1117/1.3115362>.
3. T. Kesavamurthy and K. Thiagarajan, "Survey on recent techniques in volumetric medical image compression," *Int. J. Signal Imaging Syst. Eng.*, vol. 6, no. 4, pp. 250–258, 2013, : <https://doi.org/10.1504/IJSISE.2013.056639>.
4. S. Juliet, E. Blessing, and K. Ezra, "A novel medical image compression using Ripplet transform," *J. Real-Time Image Process.*, pp. 401–412, 2016, <https://doi.org/10.1007/s11554-013-0367-9>.
5. Y. W. M. Yusof, A. Saparon, and N. A. Jalil, "A comparative analysis of transformation techniques for the reconstruction of medical images," 2014 2nd Int. Conf. Electr. Electron. Syst. Eng. ICEESE 2014, pp. 95–100, 2014, <https://doi.org/10.1109/ICEESE.2014.7154593>.
6. A. A. Mohammed and J. A. Hussein, "Hybrid transform coding scheme for medical image application," 2010 IEEE Int. Symp. Signal Process. Inf. Technol. ISSPIT 2010, pp. 237–240, 2010, <https://doi.org/10.1109/ISSPIT.2010.5711785>.
7. A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 36–58, 2001, <https://doi.org/10.1109/79.952804>.
8. K. M. Sagayam and D. J. Hemanth, "Computers in Industry ABC algorithm based optimization of 1-D hidden Markov model for hand gesture recognition applications," *Comput. Ind.*, vol. 99, no. April, pp. 313–323, 2018, <https://doi.org/10.1016/j.compind.2018.03.035>.
9. G. Vishnuvarthanan, M. P. Rajasekaran, P. Subbaraj, and A. Vishnuvarthanan, "An unsupervised learning method with a clustering approach for tumor identification and tissue segmentation in magnetic resonance brain images," *Appl. Soft Comput. J.*, vol. 38, pp. 190–212, 2016, <https://doi.org/10.1016/j.asoc.2015.09.016>.
10. T. Dumas, A. Roumy, and C. Guillemot, "Autoencoder Based Image Compression: Can the Learning be Quantization Independent?," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2018-April, no. 1, pp. 1188–1192, 2018, <https://doi.org/10.1109/ICASSP.2018.8462263>.
11. H. Fu et al., "Improved hybrid layered image compression using deep learning and traditional codecs," *Signal Process. Image Commun.*, vol. 82, no. July 2019, p. 115774, 2020, <https://doi.org/10.1016/j.image.2019.115774>.
12. S. Saravanan and S. Juliet, "Deep Medical Image Reconstruction with Autoencoders using Deep Boltzmann Machine Training," *EAI Endorsed Trans. Pervasive Heal. Technol.*, vol. 6, no. 23, 2020, <http://dx.doi.org/10.4108/eai.24-9-2020.166360>.
13. L. Zhou, C. Cai, Y. Gao, S. Su, and J. Wu, "Variational Autoencoder for Low Bit-rate Image Compression," *IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, pp. 2617–2620, 2018.
14. G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science (80-.)*, vol. 313, no. 5786, pp. 504 LP – 507, Jul. 2006, <https://doi.org/10.1126/science.1127647>.

Analysis and Classification Dysarthric Speech



Siddhant Gupta and Hemant A. Patil

1 Introduction

A natural speech production mechanism works with a *synchronized harmony* of different human organs. These organs include lungs, larynx, vocal folds, jaw muscles, tongue, lips, teeth, soft palate, velum amongst others. The output received from such system is a distinct sound which is complex in nature and intelligible by the listeners. However, sometimes a disorder in one or more of the sub-systems results in disruption in the overall speech production mechanism rendering speech unintelligible and difficult to interpret. This gives rise to a completely different class of speech signals, which is impaired in general perception, and may not be analyzed considering normal healthy speech as a basis for comparison.

Dysarthria is one such speech impairment in which the muscles that help in speaking, such as vocal folds, jaw muscles, throat muscles, etc., becomes weak and coordination between them becomes difficult. Dysarthria has been rated amongst one of the most common types of speech impairments. Speech of dysarthric patients can be characterized as slow, slurry, monotonous, unnaturally whispered, etc. or a combination of such symptoms [1]. Analysis and classification of dysarthric speech is finding its applications in fields, such as biomedical speech signal processing [2], and voice-assisted electronic device manufacturing [3]. Dysarthria is directly associated with neurological diseases, such as Parkinson Disease, Cerebral Palsy, etc. Therefore, dysarthric speech analysis can help in the diagnosis and progression mapping of such diseases. However, the characteristics of dysarthric speech are different from that of normal speech. Therefore, it has been found that applications,

S. Gupta (✉) · H. A. Patil

Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar,
Gujarat, India

e-mail: Hemant_patil@daiict.ac.in

such as Automatic Speech Recognition (ASR) systems and Voice Privacy (VP) systems do not perform even considerably in the case of dysarthric speech [4, 5]. Analysis of dysarthric speech can help in development of more robust systems targeted at people suffering from dysarthria. This chapter focuses on understanding dysarthria as a signal processing problem.

Rest of the book chapter is organized as follows. Section 2 presents various types of dysarthria, whereas Sect. 3 presents time-domain and time-frequency domain analysis, such as Linear Prediction (LP) spectrum, Teager Energy Operator (TEO) profiles, spectrograms, waterfall plots, etc., of normal vs. dysarthric speech. Section 4 gives brief details of some standard and statistically meaningful dysarthric speech corpora. A discussion on the application of deep-learning methods in the classification of dysarthric speech from normal speech. Finally, the chapter concludes with potential future research directions.

2 Types of Dysarthria

Dysarthria shares many of its symptoms with the other neurological diseases, such as Aphasia, Dysphasia, and Apraxia [6]. However, it is distinct from these neurological diseases due to the organ of its origin [7]. While Aphasia and Dysphasia effect the ability of an individual to understand and produce speech, and Apraxia results from the damage to the parietal lobe of the brain that is responsible for planning of speech [8]; dysarthria resides in the muscles responsible for the production of speech. Patients with dysarthria do not show any deviations in perceptual processing and planning of speech, as compared to a healthy subject. However, the lack of synchronization amongst muscles causes the output speech to be damaged and unintelligible. This section describes the types of dysarthria that are widely recognized in the field of speech impairments:

2.1 *Spastic Dysarthria*

Spastic dysarthria is caused as a result of some damage to the Central Nervous System (CNS), which includes brain and spinal cord [9]. It is usually accompanied by weakening of muscles and abnormal reflexes in the other regions of the body as well. Hence, phonation is strained-strangled and articulation becomes weak. In addition, mouth opening seems to be restricted and speech is perceived to come from the back of the mouth. Furthermore, jaw jerk, gag reflex, and facial reflexes are also common with the patients of dysarthria.

2.2 Flaccid Dysarthria

Flaccid dysarthria is usually recognized by the difficulties faced by the patients in pronouncing consonants. It is caused by the damage to the Peripheral Nervous System (PNS), which connects brain and spinal cord to the rest of the body [9]. Flaccid dysarthria results in symptoms, such as hypernasality, breathiness in voice, and weak pressure consonants. Depending on which nerves are damaged, it affects phonation, respiration, resonance, and articulation.

2.3 Ataxic Dysarthria

Ataxic dysarthria is caused due to a damage to part of the brain called *Cerebellum* which is responsible for receiving sensory information and regulating movements [9]. It results in imprecise articulation with distorted vowels and inaccurate consonant production, disturbed speech prosody, and abnormal phoneme timing. There is inappropriate stress on syllables, loudness, and the pitch (F_0) of the voice is deviant.

2.4 Hypokinetic Dysarthria

Hypokinetic dysarthria is caused because of the malfunction in the extrapyramidal systems of brain, which consists of areas of the brain responsible for coordination of subconscious muscle movement [9]. It is characterized by reduced pitch (F_0) variation, reduced loudness, variable speaking rate, imprecise consonants, breathy voice, and short rushed of speech. Patients with Hypokinetic dysarthria also have difficulties in swallowing and sometimes observe drooling.

2.5 Hyperkinetic Dysarthria

Hyperkinetic dysarthria is caused due to the damage to the part of the brain collectively known as Basal Ganglia, which is responsible for regulating involuntary muscle movements [9]. It is characterized by abnormal involuntary muscle movements that affect respiration, phonation, and articulatory structure impacting speech quality.

2.6 *Mixed Dysarthria*

Mixed dysarthria represents a heterogeneous group of speech disorders and neurological diseases [9]. Any combination of two or more types of dysarthria (discussed above) results in mixed dysarthria. This type of dysarthria is more common than one kind of dysarthria occurring in a patient. The symptoms of mixed dysarthria can be a mix of the symptoms discussed above.

Though different types of dysarthria are clearly defined, it is often a difficult task to distinguish amongst them because of the negligible perceptual differences in production of output speech. One has to be a trained expert to recognize one dysarthria type from the other and with a substantial probability of error. These experts are called Speech-Language Pathologists (SLPs), and they often use the pre-defined techniques and scales to distinguish between different kinds of dysarthria.

To better understand how dysarthria effects the production of speech and its intelligibility, acoustic analysis of dysarthric speech becomes necessary. Study of different acoustic features can help in distinguishing whether the speech is dysarthric or normal. This can help in early diagnosis of diseases associated with dysarthria.

3 Analysis of Dysarthric Speech

Since in patients with dysarthria, the speech is affected by the weakness in the muscles of the vocal tract system, a dysarthric speech output is very different from the speech output of a normal speaker. In addition, the acoustic features change with different severity of dysarthria from which a patient is suffering from. This section provides a comparative study of different acoustic features between dysarthric speech and normal speech and between different kinds of dysarthric speech.

3.1 *Time-Domain Analysis*

The time-domain waveform of dysarthric speech consists of useful information for the analysis of dysarthria. The pathological defects in the vocal tract system can be observed by looking at the time-domain waveform of dysarthric speech. Figure 1 shows the speech waveform of a normal person and a person suffering from dysarthria uttering the same word. It can be easily observed from the two waveforms that the dysarthric speech waveform is much longer as compared to the normal speech. Moreover, it can be that the variability in the acoustic pressure is more in dysarthric speech. Dysarthric speech also consists of regions of silence which are absent in case of normal speech, implying that these silent regions are not necessary for the speech wave but still exists. These silent regions represent the defects in motor control mechanism in the speech production system, where the vocal

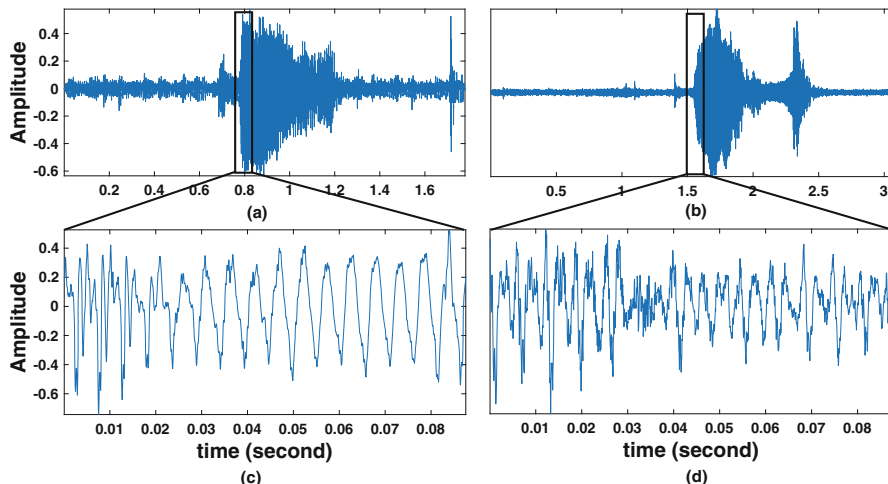


Fig. 1 Time-domain waveform. (a) normal, (b) dysarthric speech, and zoomed waveform for (c) normal, and (d) dysarthric speech

folds involuntary start/stop vibrating creating jitters and shimmers, representing variations in pitch period and volume airflow velocity, repeated across consecutive glottal cycles in the speech waveform.

3.1.1 Fundamental Frequency (F_0)

The Fundamental Frequency (F_0) of a speech signal is the average number of oscillations per second, in Hertz, of the voiced region of the speech. It arises due to the vibrations of the vocal folds which in turn oscillates the air flowing through the vocal tract system. Since the oscillations arise in an organic structure, it consists of some fluctuations, rather than being perfectly periodic, in particular, jitters and shimmers as discussed above. Due to the weakening of vocal fold muscles, a patient with dysarthria has less control over his vocal fold vibrations and hence, these jitters and shimmers are much more significant in dysarthric speech as compared to normal speech and can change the overall nature of the fundamental frequency.

3.1.2 Teager Energy Operator (TEO)

Teager Energy Operator (TEO) is a non-linear operator which helps in the analysis of speech waveform from an energy point of view. For a speech signal $s(n)$, TEO profile is given by [10]:

$$\text{TEO}\{s(n)\} = (s(n))^2 - s(n-1) \cdot s(n+1). \quad (1)$$

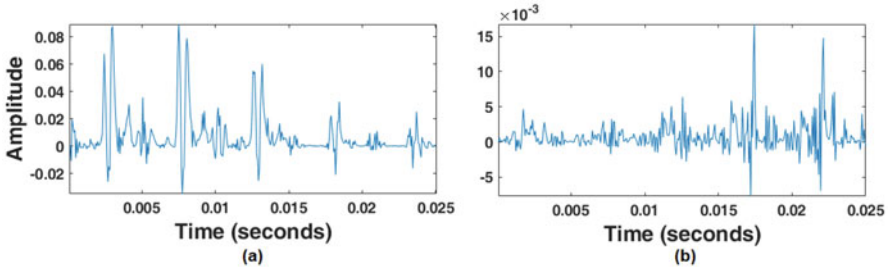


Fig. 2 TEO Profile. (a) normal and (b) dysarthric speech

From TEO, we can observe that three consecutive speech samples are required to find the running estimate of signal energy and thus, it is known to have excellent time-resolution. TEO is very efficient in capturing the non-linearity in the speech, which is captured by the airflow in the vocal tract which will change for dysarthric vs. normal speech. Figure 2 shows the corresponding TEO profile for the normal vs. dysarthric speech. We can observe from Fig. 2 that, as in LP residual, TEO is also highly irregular for dysarthric speech, as compared to normal speech, indicating abnormal changes in pitch period, i.e., T_0 , and, hence, pitch frequency. In particular, TEO gives high energy pulses corresponding to GCIs due to its capability to capture characteristics of impulse-like excitation which are known to have higher signal-to-noise (SNR) ratios.

3.2 Linear Prediction (LP) Residual

Linear Prediction (LP) residual can be a very good method for the analysis of the characteristics of the speech excitation source. LP analysis deconvolves the speech signal into its source excitation and speech system components. For a speech signal $s(n)$, LP residual $[r(n)]$ is given by [11]

$$r(n) = s(n) - \hat{s}(n), \quad (2)$$

where

$$\hat{s}(n) = \sum_{k=1}^p a_k s(n - k), \quad (3)$$

and a_k corresponds to k th Linear Prediction Coefficient (LPC).

The speech production system has its own inertia. The Glottal Closure Instants (GCIs), are the instances when the glottis closes to provide a sudden burst of air pressure through the vocal folds, act as an excitation signal in the form of input

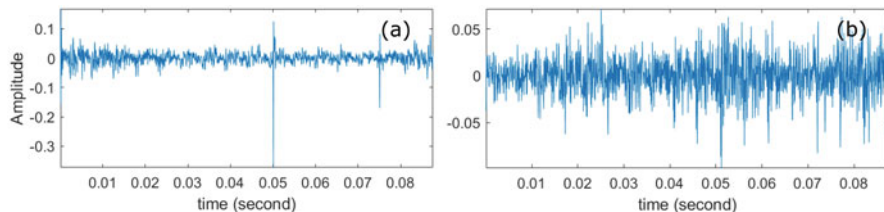


Fig. 3 LP Residual Plot. (a) normal, and (b) dysarthric speech

impulse to the system. Sudden bumps can be observed in LP residual at *periodic* locations. These locations are called GCIs. GCIs denote the time instants, where there is sudden closing of vocal folds, which acts as impulse-like excitation, during speech production. The GCIs have been estimated using Hilbert transform in [12, 13]. Figure 3 shows the LP residual plot of normal *vs.* dysarthric speech. It can be clearly observed from these plots that the LP residual for dysarthric speech is highly irregular as compared to the LP residual plot for the normal speech signal. This shows the abnormality in the dysarthric speech signal in terms of pitch period (T_0) and therefore, pitch frequency (F_0).

3.3 Time-Frequency Analysis

Due to non-stationary nature of speech signals, spectrograms can be used for the time-frequency analysis of a speech signal. A spectrogram is a visual representation of how spectral energy density varies with respect to different frequencies and time instances. The energy in a spectrogram is computed using Short-Time Fourier Transform (STFT) on windowed speech signal. Let $x(n)$ be the input signal. STFT is calculated as :

$$X(\omega, \tau) = \sum_{n=-\infty}^{\infty} x(n) \cdot w(n, \tau) \cdot e^{-j\omega n}, \quad (4)$$

$$X(\omega, \tau) = \sum_{n=-\infty}^{\infty} x(n, \tau) \cdot e^{-j\omega n}, \quad (5)$$

where $x(n, \tau) = x(n) \cdot w(n, \tau)$ is the windowed speech segment. Now spectrogram (spectral energy densities) is obtained by calculating the magnitude square of $X(\omega, \tau)$, i.e.,

$$S(\omega, \tau) = |X(\omega, \tau)|^2. \quad (6)$$

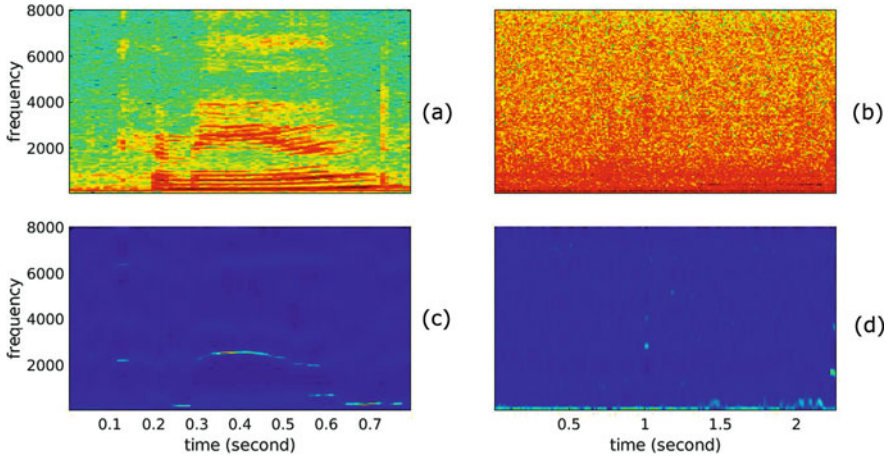


Fig. 4 Comparison of spectrogram for (a) normal and (b) dysarthric speech vs. LP Spectrum for (c) normal and (d) dysarthric speech

It can be seen from Fig. 4 that the energy is concentrated in the lower spectrum of the frequency in contrast with the normal speech, in which the energy is observed to be spread in the high as well as low frequency spectrum. This is due to the fact that abnormal functioning of the speech production results in a great amount of energy loss for higher frequencies [14, 15]. This can be the reason why speaking intelligibly is a challenging task for a person suffering from dysarthria, which usually comes very naturally to the normal speakers. Figure 4 show the plot of Short-Time Fourier Transform (STFT) vs. LP spectrum for the normal vs. dysarthria speech case. Waterfall plot is also shown in Fig. 5 to emphasize the corresponding joint time-frequency characteristics during the production of dysarthric speech. From the waterfall plots, We can observe that the formant structure is severely damaged for dysarthric speech as compared to its normal counterpart, where formant peaks and their evolving structures are clearly visible. Thus, the analysis presented in this section indicates that F_0 , its harmonics, formants, and their structures are severely affected due to dysarthria.

4 Datasets on Dysarthric Speech

4.1 TORGO Database

TORGO Database [16] was developed through a collaboration between the departments of Computer Science and Speech-Language Pathology, University of Toronto; Holland-Bloorview Kids Rehab Hospital, Toronto; and The Ontario Federation for Cerebral Palsy with an aim to develop Automatic Speech Recognition

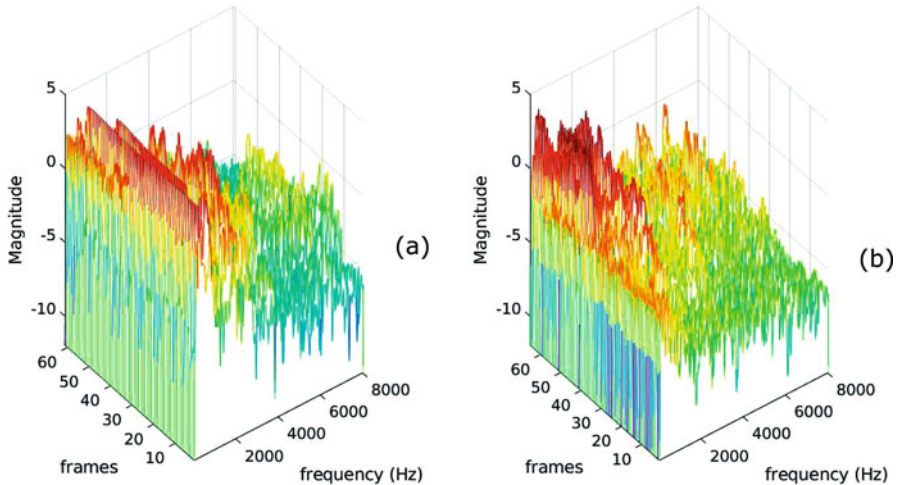


Fig. 5 Waterfall Plot. (a) normal and (b) dysarthric speech

(ASR) systems for people suffering from dysarthria as the unintelligibility in the dysarthric speech results in a Word Error Rate (WER) of as high as 97.5% as compared to a WER of 15.5% for normal speech when both the speech are tested on modern ASR systems trained on normal speech [17].

The database consists of speech samples from 7 dysarthric subjects (4 males and 3 females), between the age 16 and 50, who are chosen by a speech-language pathologist at Bloorview Research Institute, Toronto. Dysarthria of 6 subjects resulted from Cerebral Palsy while 1 subject developed dysarthria due to Amyotrophic Lateral Sclerosis (ALS), which are two of the most common cause of speech impairment [18]. The database also includes speech samples from non-dysarthric subject which were age and gender matched with the dysarthric subjects.

All the subjects were assessed using the standard Frenchay Dysarthria Assessment (FDA) [19] by a speech-language pathologist. The FDA measured 28 perceptual dimension of speech production and articulation which are rated on a 9-point scale. The database is divided into the speech samples of the following categories:

- **Non-Words** consists of 5–10 repetitions of /iy-p-ah, ah-p-iy and p-ah-t-ah-k-ah/, respectively. In addition, utterances with high and low pitch vowels are also recorded, e.g., pronouncing “eee” for 5 s (also used in [20]).
- **Short Words** consists of repetition of English digits 1–10 along with some other words like yes, no, left, right, etc. In addition, 50 words from word intelligibility section of FDA [19] and 360 words from Yorkston-Beukelman Assessment of Intelligibility of Dysarthric Speech (YBAIDS) [21] are chosen. Ten most common words from the British National Corpus were also recorded by the subjects.

- **Restricted sentences** consists of preselected phoneme rich sentences, The Grandfather Passage from Nemours Database [22], 162 sentences from sentence intelligibility section of YBAIDS [21], and 460 sentences from MOCHA database.
- **Unrestricted Sentence** consists of unscripted sentences by the subjects recorded while describing 30 images of interesting situation chosen randomly from Webber Photo Cards: Story Starters Collection.

4.2 *Universal Access (UA) Corpus*

The UA database [23] consists of speech samples from 19 dysarthric subjects (15 males and 5 females) [24]. Each subject is scored on a scale of 0–100% on the basis of their speech intelligibility rated by the human listeners. The recording is done using an eight-channel microphone arrays.

The database was recorded into three blocks of words. Each block consists of a total of 255 words, out of which 155 words are repeated across the blocks while the remaining 100 uncommon words are unique for every block. The repeated 155 words include 10 digits (one, two, etc.), 26 radio alphabets (Alpha, Bravo, etc.), 19 computer commands (enter, delete, etc.) and 100 most common words chosen from the Brown corpus of written English. The 100 common words (naturalization, exploit, etc.) were chosen from children’s novels. In this way, each subject was recorded for a total of 765 speech samples out of which 300 samples are distinct uncommon words and remaining 465 are 3 repetitions of 155 distinct words across the blocks.

4.3 *HomeService Corpus*

The homeService corpus [25] is created as a part of the bigger homeService project, whose objective is to provide the people with speech and motor disabilities with the ability to operate home appliances with voice commands [26]. The project is motivated by the fact that there is a shortage of dysarthric speech data which is recorded in a real life environment within the research community. The project enables its user to operate their home appliances, such as TV, lamps, etc. using voice commands, which are recorded and transferred using a cloud-based environment to a data collection center. The dataset consists of speech data of 5 dysarthric patients (3 males and 2 females). The speech samples were recorded with an 8-channel microphone array at a sampling rate of 16 kHz and consists of two types of speech data. In particular,

- **Enrollment Data**—This data is recorded in a controlled research environment and is used to train the ASR system which the user can use to operate their

Table 1 Comparison of various corporas for dysarthric speech [17], [24], [26]

Dataset	# Speakers	Male/Female	Text material	Dysarthria inducing disease	Application domain
TORGO corpus	7	4/3	Words & Sentences	Cerebral Palsy, Amyotrophic Lateral Sclerosis	ASR
UA corpus	19	14/5	Words	Cerebral Palsy	ASR
HomeService corpus	5	3/2	Voice commands	Cerebral Palsy, Motor-Neuron Disease	Voice assistants

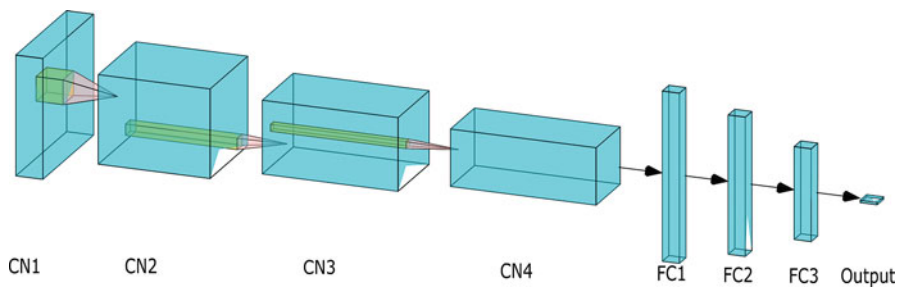


Fig. 6 CNN architecture for classification task. After [27]

appliances. The user was asked to read from a list. Therefore, the speech recorded is less natural. Annotation is done using the transcriptions in the reading list.

- **Interaction Data**—This data is recorded at the house of the users while they control their devices. Identity of each word in this data is not known and is therefore, annotated by the human listeners. The speech in the data is more natural.

5 Classification of Dysarthric and Normal Speech

Recently, there has been significant increase in the popularity of deep-learning based approach to solve complex task by the computers. For that reason, computers are now able to efficiently perform tasks, such as image classification [28–30], image recognition [31, 32], computer vision, etc. [33, 34]. Focus of deep-learning based algorithms are also increasing in Speech recognition and classification. Convolutional Neural Network (CNN) is one such algorithm which can efficiently detect complex pattern from a set of matrices, such as images. It can be used to classify normal vs. dysarthric speech by learning the patterns in the spectrogram of the speech samples (Fig. 6).

5.1 Experimental Setup

For the experiment, spectrograms were obtained for the speech utterances of both normal and dysarthric speakers, keeping a window size of 25 ms and an overlap of 10 ms. These spectrograms were stored in the form of images which is to be fed to network.

Our CNN [35] comprises 4 convolutional layers followed by 3 fully connected layers. Each convolution layer performs a convolution operation with a kernel size of 5x5 keeping step size of 1 and no padding. This convolution operation is followed

by a maxpooling operation with a kernel size of 4×4 . The number of output channels obtained by the 4 convolutions channels are 8, 16, 32, and 64, respectively. The output of the final convolutional layer is given to the fully connected layer. Sigmoid activation is used on the final output from the final fully connected layer which provide us with probabilistic value of the input. The threshold value is kept at 0.5, i.e., if output value is greater than 0.5 then the input will be classified as dysarthric speech, and as normal speech if output is less than 0.5. ReLU activation function is used to activate the hidden layers in the network. In addition, Stochastic Gradient Descent (SGD) is used as optimization algorithm and binary cross entropy is taken as the loss function.

5.2 Dataset Used for This Study

UA corpus [24] is used as the dataset for the experiment. Data from one dysarthric speaker (M07) and one normal speaker (CM01) is used. The data for each speaker was divided into 3 blocks out of which data from block 1 is chosen for training and utterances were taken from mic 3 of the 7-channel microphone array. For testing the accuracy of the model, testing was done with the data from block 3. Specifically, 100 distinct Uncommon Words (UW) were chosen for testing.

5.3 Results and Analysis

It is observed that the model was able to provide an accuracy of 65.68% on the testing data. The performance of the network is effective given the fact that the experiments were performed on a small training set. The model was able to recognize the variability in the spectrogram that differentiates dysarthric speech from normal speech. In addition, it can be said that the model was also able to learn that the low energy that is associated with the spectrogram of the dysarthric speech.

6 Conclusion

In this chapter, we have discussed dysarthria as a speech technology problem. A number of analysis have been done on normal *vs.* dysarthric speech, such as F_0 , TEO profile, LP residuals, spectrograms, and waterfall plot to provide the reader with an insight of the difference between normal *vs.* dysarthric speech. In addition, some widely used datasets are also discussed along with their key features. Furthermore, an experiment has also been presented for the classification of normal and dysarthric speech using a deep neural networks approach based on CNN.

The effectiveness of the classification task is dependent on the training data on which it is trained on. A model trained on the speech sample of a speaker having high severity-level of dysarthria may not be effective for a speaker with low dysarthria severity-level and vice versa. Therefore, this becomes a prominent limitation of the current methods of classification. Furthermore, research on severity-based classification of dysarthric speaker are very limited. In the future, more sophisticated deep neural networks can be used for classification of normal vs. dysarthric speech and classification based on the severity-level of dysarthria.

Acknowledgments The authors would like to thank the authorities at DA-IICT Gandhinagar, India for providing resources and kind support towards the completion of this book chapter. The authors would also like to thank Ms. Priyanka Gupta for providing useful suggestions for further improvement of our this book chapter.

References

1. BENT, T., BAESE-BERK, M., BORRIE, S.A. and MCKEE, M. (2016) Individual differences in the perception of regional, nonnative, and disordered speech varieties. *The Journal of the Acoustical Society of America (JASA)* **140**(5): 3775–3786.
2. CASTILLO-GUERRA, E. (2009) Acoustic study of dysarthria. *International Journal of Biomedical Engineering and Technology* **2**(4): 352–369.
3. BALLATI, F., CORNO, F. and DE RUSSIS, L. (2018) “hey siri, do you understand me?”: Virtual assistants and dysarthria. In *7th International Workshop on the Reliability of Intelligent Environment (WoRIE), Rome, Italy*.
4. DE RUSSIS, L. and CORNO, F. (2019) On the impact of dysarthric speech on contemporary ASR cloud platforms. *Journal of Reliable Intelligent Environments* **5**(3): 163–172.
5. HWANG, Y., SHIN, D., YANG, C.Y., LEE, S.Y., KIM, J., KONG, B., CHUNG, J. et al. (2012) Developing a voice user interface with improved usability for people with dysarthria. In *International Conference on Computers for Handicapped Persons* (Springer): 117–124.
6. FLOWERS, H.L., SILVER, F.L., FANG, J., ROCHON, E. and MARTINO, R. (2013) The incidence, co-occurrence, and predictors of dysphagia, dysarthria, and aphasia after first-ever acute ischemic stroke. *Journal of Communication Disorders* **46**(3): 238–248.
7. MCNEIL, M.R., ROBIN, D.A. and SCHMIDT, R.A. (1997) Apraxia of speech: Definition, differentiation, and treatment. *Clinical Management of Sensorimotor Speech Disorders* : 311–344.
8. GOLDENBERG, G. (2009) Apraxia and the parietal lobes. *Neuropsychologia* **47**(6): 1449–1459.
9. DUFFY, J.R. (2013) *Motor Speech disorders-E-Book: Substrates, differential diagnosis, and management* (Elsevier Health Sciences).
10. KAISER, J.F. (1990) On a simple algorithm to calculate the ‘energy’ of a signal. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Albuquerque, USA* (IEEE): 381–384.
11. ATAL, B.S. and HANAUER, S.L. (1971) Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acoustical Society of America (JASA)* **50**(2B): 637–655.
12. ANANTHAPADMANABHA, T. and YEGNANARAYANA, B. (1975) Epoch extraction of voiced speech. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **23**(6): 562–570.
13. ANANTHAPADMANABHA, T. and YEGNANARAYANA, B. (1979) Epoch extraction from linear prediction residual for identification of closed glottis interval. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **27**(4): 309–319.

14. VÁSQUEZ-CORREA, J.C., OROZCO-ARROYAVE, J.R. and NÖTH, E. (2017) Convolutional neural network to model articulation impairments in patients with Parkinson's disease. In *INTERSPEECH, Stockholm, Sweden*: 314–318.
15. GUPTA, S., PATIL, A.T., PUROHIT, M., PARMAR, M., PATEL, M., PATIL, H.A. and GUIDO, R.C. (2021) Residual neural network precisely quantifies dysarthria severity-level based on short-duration speech segments. *Neural Networks* .
16. Torgo Database. Available at. <http://www.cs.toronto.edu/~complingweb/data/TORGO/torgo.html>. {Last Accessed 23/03/2021 10:05PM}.
17. RUDZICZ, F., NAMASIVAYAM, A.K. and WOLFF, T. (2012) The torgo database of acoustic and articulatory speech from speakers with dysarthria. *Language Resources and Evaluation* **46**(4): 523–541.
18. KENT, R.D. (2000) Research on speech motor control and its disorders: A review and prospective. *Journal of Communication Disorders* **33**(5): 391–428.
19. ENDERBY, P. (1980) Frenchay dysarthria assessment. *British Journal of Disorders of Communication* **15**(3): 165–173.
20. BENNETT, J.W., VAN LIESHOUT, P.H. and STEELE, C.M. (2007) Tongue control for speech and swallowing in healthy younger and older adults .
21. YORKSTON, K.M. and BEUKELMAN, D.R. (1978) A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders* **11**(6): 499–512.
22. MENENDEZ-PIDAL, X., POLIKOFF, J.B., PETERS, S.M., LEONZIO, J.E. and BUNNELL, H.T. (1996) The Nemours database of dysarthric speech. In *Proceeding of Fourth International Conference on Spoken Language Processing (ICSLP), Philadelphia, USA (IEEE)*, **3**: 1962–1965.
23. Universal Access (UA) Corpus. Can be requested at. <http://www.isle.illinois.edu/sst/data/UASpeech/>. {Last Accessed 23/03/2021 10:30PM}.
24. KIM, H., HASEGAWA-JOHNSON, M., PERLMAN, A., GUNDERSON, J., HUANG, T.S., WATKIN, K. and FRAME, S. (2008) Dysarthric speech database for universal access research. In *Ninth Annual Conference of the International Speech Communication Association, Brisbane, Australia*.
25. HomeService Corpus. Can be requested at. <http://www.isle.illinois.edu/sst/data/UASpeech/>. {Last Accessed 23/03/2021 10:38PM}.
26. NICOLAO, M., CHRISTENSEN, H., CUNNINGHAM, S., GREEN, P. and HAIN, T. (2016) A framework for collecting realistic recordings of dysarthric speech-the homeservice corpus. In *Proceedings of LREC 2016 (European Language Resources Association)*.
27. KRIZHEVSKY, A., SUTSKEVER, I. and HINTON, G.E. (2012) Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **25**: 1097–1105.
28. LIU, S., TIAN, G. and XU, Y. (2019) A novel scene classification model combining ResNet based transfer learning and data augmentation with a filter. *Neurocomputing* **338**: 191–206.
29. WANG, F., JIANG, M., QIAN, C., YANG, S., LI, C., ZHANG, H., WANG, X. *et al.* (2017) Residual attention network for image classification. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), Hawaii, USA*: 3156–3164.
30. ZHU, C., SONG, F., WANG, Y., DONG, H., GUO, Y. and LIU, J. (2019) Breast cancer histopathology image classification through assembling multiple compact CNNs. *BMC Medical Informatics and Decision Making* **19**(1): 1–17.
31. HE, K., ZHANG, X., REN, S. and SUN, J. (2016) Deep residual learning for image recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA*: 770–778.
32. LU, Z., JIANG, X. and KOT, A. (2018) Deep coupled resnet for low-resolution face recognition. *IEEE Signal Processing Letters* **25**(4): 526–530.
33. JUNG, H., CHOI, M.K., JUNG, J., LEE, J.H., KWON, S. and YOUNG JUNG, W. (2017) Resnet-based vehicle classification and localization in traffic surveillance systems. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Hawaii, USA*: 61–67.

34. LIU, X., ZHOU, Y., ZHAO, J., YAO, R., LIU, B., MA, D. and ZHENG, Y. (2020) Multiobjective ResNet pruning by means of EMOAs for remote sensing scene classification. *Neurocomputing* **381**: 298–305.
35. FUKUSHIMA, K. and MIYAKE, S. (1982) Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and Cooperation in Neural Nets* (Springer), 267–285.

Skin Cancer Detection and Classification Using DWT-GLCM with Probabilistic Neural Networks



J. Pandu, Umadevi Kudtala, and B. Prabhakar

1 Introduction

In recent days, skin cancer becomes most affected disease of all the types of cancers, and it is divided as benign and malignant. In these two types, malignant is recognized as the deadliest one while comparing with the non-malignant skin cancers [1]. It is known fact that malignant skin cancer affects more people every year and early treatment is really important for the survival of the patients. Inspection of malignant skin cancer needs well-experienced dermatologists. These people use a computer-assisted system for early detection of malignant [2]. More algorithms in deep learning models were used for diagnosis of skin cancer diagnosis. These models are still facing more challenges for achieving the high accuracy rate, and the drawbacks of conventional models should be overcome. This paper proposes a novel skin cancer detection approach. Many research papers have utilized image preprocessing for the identification of the malignant skin cancer at the initial times, which leads to effective treatment. In this way, it is necessary to broaden the span of such essential diagnostic care by arranging efficient frameworks for skin disease classification. Many research papers have utilized image preprocessing for the identification of the malignant skin cancer at the initial times, which leads to effective treatment. Proficient dermatologists have set up the ABCDEs [3, 4] (Asymmetrical shape, Border irregularities, Color, Diameter, and Evolution) as the standardized descriptions to help with visualizing standard features of severe malignant skin cancer cases. One of the main challenges of classifying harmful

J. Pandu · U. Kudtala (✉)

Sreyas Institute of Engineering and Technology, Nagole, Hyderabad, Telangana, India

B. Prabhakar

Department of ECE, JNTUH College of Engineering, Jagtial, Telangana, India

skin injuries is due to sheer proportions of varieties over the different skin tones from people of different ethnic backgrounds. Recently, new accomplishments in the improvement of convolutional neural networks (CNN) [5] have permitted computers to beat dermatologists in skin cancer classification tasks. The following phase is to further improve the accuracy of location of malignant skin lesion. Our strategy for early diagnosis of skin lesion incorporates deep learning which helps us to enhance the accuracy of the automated framework compared to methods. In this work, we proposed our custom network for lesion classification. The major contributions of this research are as follows:

- K-means clustering-based segmentation mechanism is used to identify the cancer region from the input test image.
- The network is trained and tested with the GLCM-based texture features, DWT-based low level features, and statistical color features by using the PNN deep learning model.
- The proposed classification accuracy is compared with the conventional SVM [14] and active contour segmentation methods and gives the better results compared to them.

The remainder of the paper is structured as: Literature survey conducted for the paper is covered in Sect. 2. Section 3 covers the proposed skin cancer detection method, while Sect. 4 describes the environment in which experiments were conducted. In Sect. 5, the results obtained from experimentation and observations are discussed.

2 Literature Survey

There have been several systems developed for detecting malignant skin lesions as early as possible using the dermoscopic images. The dermatologists assess the skin lesions using the “ABCD Rule.” Based on this rule, many methods have been devised to classify dermoscopic images. Researchers have used extracted features and attempted to train diverse machine learning classifiers such as k-NN, SVM [6]. In Refs. [6, 7], authors used very deep and machine learning residual networks to classify the images. In order to cope with degradation and over fitting, first machine learning is applied. Then, the radial basis function network (RBFN) is constructed so that skin lesion segmentation can be accurate. Then, this RBFN and deep residual networks that are used to classify the images are taken together to make a two-stage framework. In Ref. [8], images have been obtained by epi-luminescence microscopy, which enhances the chances of early recognition of skin lesions as malignant or benign. A binary mask is used, and shape and radiometric features are extracted to detect how malignant a lesion is. After that, the CNN classifier is deployed for classifying images as malignant or benign.

In Ref. [9], automatic border detection is performed and then shapes are extracted from these borders. Texture features are then computed using the GLCM and Euclidean distance transform. Images are then classified using the SVM classifier.

Refs. [10, 11] use fractional coefficients of a cosine transformed skin image, which results in better space complexity and optimum performance in malignant skin cancer identification. The ensemble of “SVM-AD Tree-Random Forest” gave the superior performance among all the classifiers used. Researchers are now focusing more on deep learning concepts as there have been significant advancements in deep learning. They are using the neural network ensemble model, very deep residual networks, and artificial neural networks. But they might have certain drawbacks like more processing power is required or more data are required which might be difficult to find as such datasets are not readily available. In Ref. [12], an overview of the most important implementations of malignant skin cancer detection is given and then comparison of the performance of numerous classifiers on the classification of dermoscopy images as benign or malignant is presented. All the existing approaches [13] of skin cancer detection can be grouped in three streams as malignant skin cancer detection with machine learning models using spatial domain features, malignant skin cancer detection with machine learning models using transform domain features, and malignant skin cancer detection using unsupervised neural network models. The transform domain feature-based machine learning models of malignant skin cancer detection are complex. The SVM models are more complex and do need heavy hardware as well as huge dataset for getting trained in malignant skin cancer detection. The spatial domain feature-based machine learning models are simple, faster, and applicable to any size of skin dermoscopy images.

3 Proposed Method

The proposed research work majorly focuses on detection of the following skin cancers such as malignant and benign, respectively. The detailed operation of the skin cancer detection and classification approach is presented in Fig. 1.

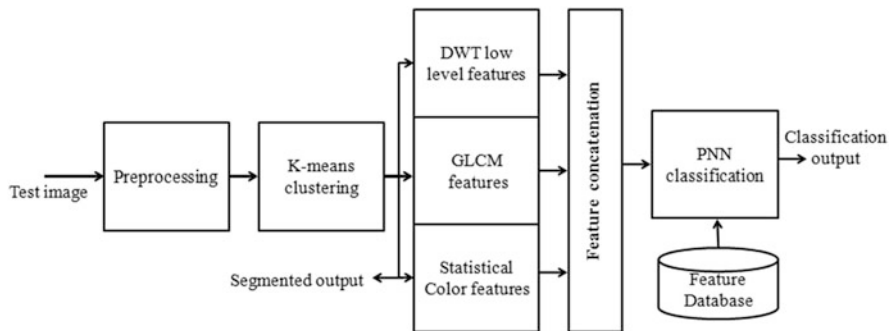


Fig. 1 Skin cancer detection and classification

3.1 Database Training and Testing

Database is trained from the collected images of the “International Skin Imaging Collaboration (ISIC)” archive. ISIC is one of the biggest available collections of quality-controlled dermoscopic images. The dataset consisted of 15 benign and 15 malignant images. All the images are trained using the PNN network model with GLCM features, statistical and texture features. And, a random unknown test sample is applied to the system for detection and classification, respectively.

3.2 Preprocessing

The query image is acquired from the image acquisition step, which includes background information and noise. Preprocessing is required and necessary to remove the above-mentioned unwanted portions. The pre-processing stage is mainly used for eliminating the irrelevant information such as unwanted background part, which includes noises, labels, tape and artifacts, and the pectoral muscle from the skin image. The different types of noise that occurred in the mammogram images are salt and pepper, Gaussian, and speckle and Poisson noise. When noise occurred in an image, the pixels in the image show different intensity values instead of true pixel values.

So by choosing the perfect method in the first stage of preprocessing, this noise removal operation will be performed effectively. Reduction of the noise to a great extent and avoiding the introduction visual artifacts by the analysis of pixels at various scales, sharpening and smoothing filter denoising efforts to eradicate the noise presented in the pixel, as it conserves the image uniqueness, despite its pixel satisfied. These filters can effectively detect and remove noise and thin hairs from the image; then we perform top hat transform for removing the thick hairs. Contrast limited adaptive histogram equalization CLAHE is also performed on the skin lesion to get the enhanced image in the spatial domain. Histogram equalization works on the whole image and enhances the contrast of the image, whereas adaptive histogram equalization divides the whole image and works on the small regions called tiles. Each tile is typically 8*8 pixels, and within each tile, histogram is equalized, thus enhancing the edges of the lesion. Contrast limiting is applied to limit the contrast below the specific limit to limit the noise.

3.3 Image Segmentation

After the preprocessing stage, segmentation of the lesion was done to get the transparent portion of the affected area of skin. On transformation, the K-means clustering method is applied to the image to segment the skin lesion area based

on thresholding. In the K-means clustering algorithm, segmentation is the initial process of this work; at the cluster centers, cost junction must be minimized which varies with respect to memberships of user inputs. Image segmentation is the process of dividing the image into multiple clusters based on the region of interest presented to detect the skin cancer. Regions of interest are portions of skin images, which are used by radiologists to detect abnormalities like micro-classifications (benign and malignant).

The K-means clustering is used in the proposed procedure for segmentation to a certain extent compared to the active counter clustering approach because of its speed of operation while maintaining the highest accuracy. The K-means clustering procedure combines the properties of both possibility and K-means clustering approaches as shown in Fig. 2. Here, the membership functions are generated in the probability-based manner to get better detection. Among those detected tumors, the highest accurate cancer regions are considered as an ROI. The automatic extraction of the ROI is difficult. So, ROIs are obtained through possibility cropping, which are based on location of abnormality of original test images. Here, the membership functions are generated in the probability-based manner to get better detection. Among those detected cancer regions, the highest accurate cancer region is considered as the ROI.

3.4 Feature Extraction

Several features can be extracted from the skin lesion to classify the given lesions. We extracted some of the prominent features, which help us in distinguishing the skin lesions, these are GLCM-based texture features, DWT-based low level features and statistical color features, respectively.

Using the GLCM is a texture technique of scrutinizing textures considering spatial connection of image pixels. The texture of the image gets characterized by GLCM functions through computations of how often pairs of pixels with explicit values and in a particular spatial connection are present in images. The GLCM matrix can be created and then statistical texture features are extracted from the GLCM matrix. GLCM shows how different combinations of pixel brightness values which are also known as grey levels are present in images. It defines the probability of a particular grey level being present in the surrounding area of other grey level. In this paper, the GLCM is extracted first from the image for all three-color spaces, i.e., RGB, CIE L*u*v, and YCbCr. Then, the GLCM matrix is calculated in four directions which are 135°, 90°, 45°, and 0° degrees as shown in Fig. 3. In the following formulas, let a, b be number of rows and columns of the matrix, respectively, $S_{a,b}$ be the probability value recorded for the cell (a, b), and the number of gray levels in the image is "N." Then, several textural features can be extracted from these matrices; extracted textural features are shown in the following equations:

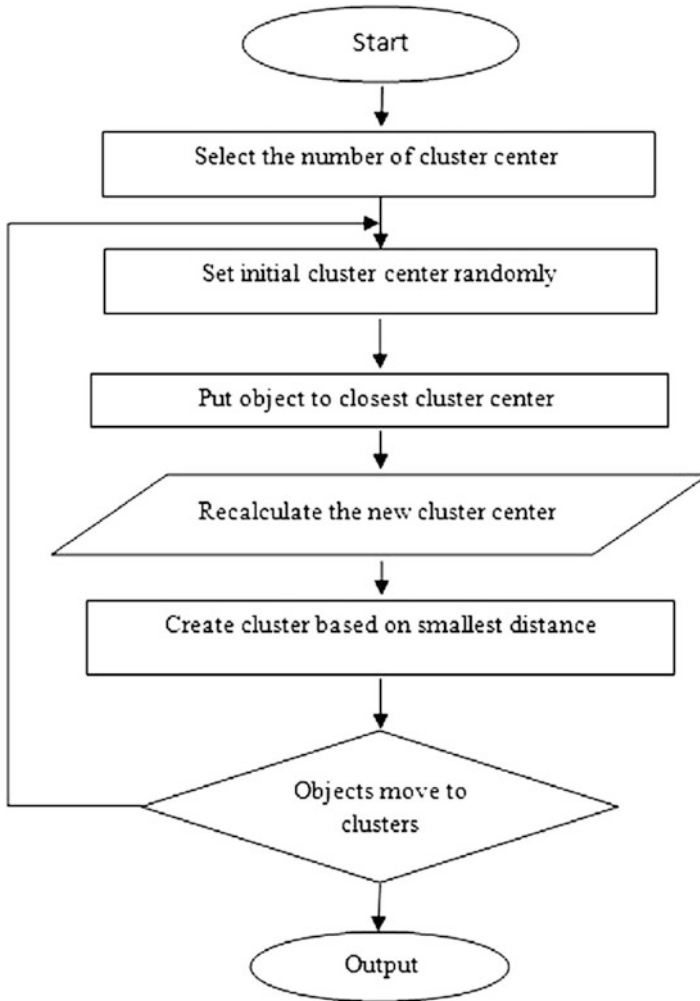


Fig. 2 K-means clustering

GLCM features used are

$$\text{Contrast} = \sum_{a,b=0}^{N-1} S_{a,b}(a-b)^2 \quad (1)$$

$$\text{Homogeneity} = \sum_{a,b=0}^{N-1} \frac{S_{a,b}}{1+(a-b)^2} \quad (2)$$

Fig. 3 Orientations and distance to compute the GLCM

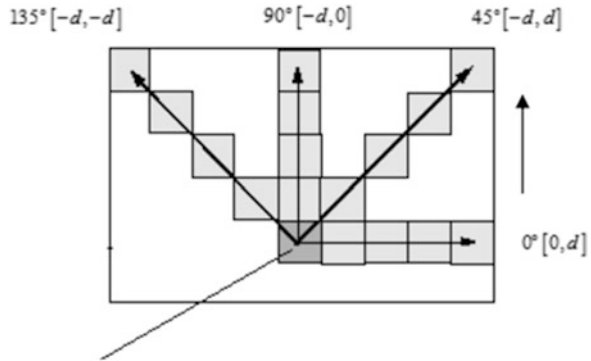


Fig. 4 2-level DWT coefficients

$$\text{Correlation} = \sum_{a,b=0}^{N-1} S_{a,b} \left[\frac{(a - \mu_a)(b - \mu_b)}{\sqrt{(\sigma_a^2)(\sigma_b^2)}} \right] \tag{3}$$

$$\text{Angular Second Moment (ASM)} = \sum_{a,b=0}^{N-1} s_{a,b}^2 \tag{4}$$

$$\text{Energy} = \sqrt{\text{ASM}} \tag{5}$$

Then, 2 level DWT is also used to extract the low-level features. Initially, on the segmented output, DWT is applied, which results in the output as LL1, LH1, HL1, and HH1 bands, respectively. Then entropy, energy, and correlation features are calculated on the LL band. Then, on the LL output band again, DWT is applied and results in the output as LL2, LH2, HL2, and HH2, respectively. Again entropy, energy, and correlation features are calculated on the LL2 band, respectively, as shown in Fig. 4.

And finally, mean and standard deviation-based statistical color features are extracted from the segmented image. They are.

$$\text{Mean } (\mu) = \frac{1}{N^2} \sum_{i,j=1}^N I(i, j) \quad (6)$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\sum_{i,j=1}^N [I(i, j) - \mu]^2}{N^2}} \quad (7)$$

Then, all these features are combined using array concatenation and results in the output as the hybrid feature matrix.

3.5 Classification

Neural networks have been effectively applied across a range of problem domains like finance, medicine, engineering, geology, physics, and biology. From a statistical viewpoint, neural networks are interesting because of their potential use in prediction and classification problems. Developing a PNN is a method that involves the emulation of birth neural scheme.

The neurons are connected in the predefined architecture for effectively performing the classification operation. Depending on the hybrid features, the weights of the neurons are obtained. Then, the relationships between weights are identified using their characteristic hybrid features. The quantity of weights decides the levels of layers for the proposed network. Figure 5 represents the architecture of artificial neural networks. A PNN basically consists of two stages for classification such as training and testing. The process of training will be performed based on the layer-based architecture. The input layer is used to perform the mapping operation on the input dataset; the hybrid features of this dataset are categorized into weight distributions.

The PNN architecture has four hidden layers with weights. The first convolutional 2D hidden layer of the net takes in $224 * 224 * 3$ pixels skin lesion images and applies $96 * 11 * 11$ filters at stride 4 pixels, followed by the class node activation layer and the decision normalization layer. Then, the classification operation was implemented at the two levels of the class node hidden layer. The two levels of the hidden layer hold individual normality and abnormalities of the skin cancer characteristic information. Based on the segmentation criteria, it is categorized as normal and abnormal classification. These two levels are mapped as labels in the output layer. Again, the hidden layer also contains the abnormal cancer types separately; it also holds the benign and malignant cancer weights in the second stage of the hidden layer. Similarly, these benign and malignant weights are also mapped as labels into output layers. When the test image is applied, its hybrid features

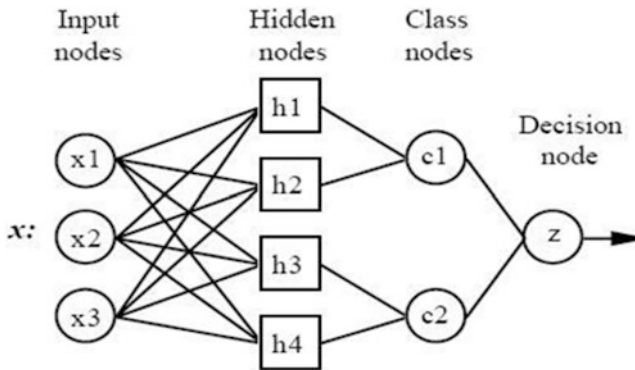


Fig. 5 Layered architecture of the PNN model

are applied for testing purpose at the classification stage. Based on the maximum feature matching criteria utilizing Euclidean distance manner it will function. If the feature match occurred with hidden layer 1 labels, then it is classified as the normal skin image. If the feature match occurred with hidden layer C1 labels with maximum weight distribution, then it is classified as the benign effected cancer image. If the feature match occurred with hidden layer C2 labels with minimum weight distribution, then it is classified as the malignant cancer image.

4 Experimentation Environment

4.1 Dataset

The experiments are done using the MATLAB R2018a tool. ISIC is one of the biggest available collections of quality controlled dermoscopic images. For the implementation of the proposed method, spatial domain, and frequency domain of 30 dermoscopic skin lesion images (15-benign and 15-malignant) have been obtained, respectively, by applying rotations at different angles. Train images of each label have been used to train the PNN architecture with fifty epochs, whereas the remaining 20% is used for testing. The features extracted by the GLCM, DWT future network are used to train the PNN classifier to classify the images into its respective classes. The efficiency of the model can be computed using various performance metrics.

From Fig. 6, it is observed that the proposed method can effectively detect the regions of skin cancers, and it indicates the segmentation done very effectively compared to the active contour approach. Here, TEST-1 and TEST 2 images are considered as the benign, and TEST-3 and TEST-4 images are considered malignant type images, respectively. For the malignant images, the segmentation accuracy is more.


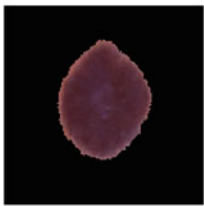
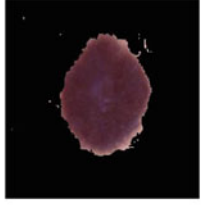

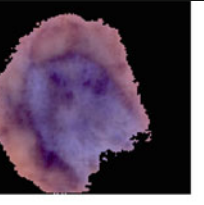
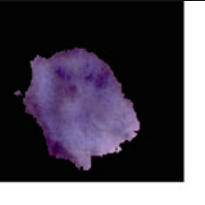
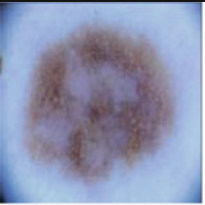
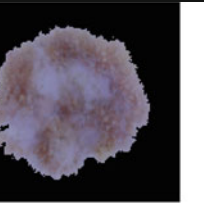




	Input image	Active contour segmented output	K-means segmented output
TEST 1- Benign			
TEST 2- Benign			
TEST 3 Malignant			
TEST 4 Malignant			

Fig. 6 Segmented output images of various methods

4.2 Performance Metrics

For evaluating the performance measure, the proposed method is implemented with the two types of segmentation methods, they are active contour (AC) and k-means clustering, respectively. For performing this comparison, accuracy, sensitivity, F-measure, precision, MCC, dice, Jaccard, and specificity parameters are calculated, respectively.

Table 1 Performance comparison

Metric	Method	Test 1	Test 2	Test 3	Test 4
Accuracy	PNN-AC	0.9157	0.78099	0.85796	0.47765
	PNN-k means	0.99985	0.99715	0.99999	0.99999
Sensitivity	PNN-AC	0.70588	0.90024	0.9166	0.83857
	PNN-k means	0.99931	0.99198	1	1
F-measure	PNN-AC	0.82207	0.68494	0.79395	0.44602
	PNN-k means	0.99965	0.99381	0.99998	0.99998
Precision	PNN-AC	0.98404	0.55275	0.70023	0.30381
	PNN-k means	1	0.99852	0.99997	0.99997
MCC	PNN-AC	0.7869	0.56857	0.70305	0.1835
	PNN-k means	0.99956	0.99198	0.99998	0.99998
Dice	PNN-AC	0.82207	0.68494	0.79395	0.44602
	PNN-k means	0.99965	0.99381	0.99998	0.99998
Jaccard	PNN-AC	0.69789	0.52085	0.65831	0.28702
	PNN-k means	0.99931	0.9877	0.99997	0.99977
Specificity	PNN-AC	0.99564	0.73812	0.83298	0.35685
	PNN-k means	1	0.99956	0.99999	0.99998

Table 2 Accuracy comparison

Method	Test 1	Test 2	Test 3	Test 4
SVM-Linear kernel [14]	0.4	0.40	0.7	0.7
SVM-RBF kernel [14]	0.4	0.45	0.55	0.6
SVM-Polynomial kernel [14]	0.4	0.3667	0.50	0.5667
SVM-5 fold cross validation [14]	0.6	0.55	0.60	0.45
Proposed PNN-AC	0.9157	0.78099	0.85796	0.47765
Proposed PNN-K-means	0.99985	0.99715	0.99999	0.99999

From Table 1 and Fig. 6, it is observed that the proposed K-means clustering method along with the PNN gives the highest performance for all metrics compared to the active counter method.

From Table 2, it is observed that the proposed method gives the highest accuracy for both benign and malignant diseases compared to the various kernels of SVM [14] such as SVM-linear kernel, RBF kernel, polynomial kernel, and 5-fold cross-validation, respectively.

5 Conclusion

This article presented a computational methodology for detection and classification of skin cancer from MRI images using the PNN-based deep learning-based approach. Here, Gaussian filters are utilized for preprocessing, which eliminates

any unwanted noise elements or artifacts innovated during image acquisition. Then, K-means clustering segmentation is employed for ROI extraction and detection of cancerous cells. Then, the GLCM, DWT-based method was developed for the extraction of statistical, color, and texture features from the segmented image, respectively. Finally, the PNN was employed to classify the type of cancer such as either benign or malignant using the trained network model. Thus, upon comparing with state of art works, we conclude that using the PNN is better than the conventional SVM method. In future, this work can be extended by implementing a greater number of network layers into the PNN and can also be applied for other types of benign and malignant cancers.

References

1. Nasiri, Sara, et al. "DePicT Malignant Deep-CLASS: a deep convolutional neural networks approach to classify skin lesion images." *BMC bioinformatics* 21.2 (2020): 1-13.
2. Munir, Khushboo, et al. "Cancer diagnosis using deep learning: a bibliographic review." *Cancers* 11.9 (2019): 1235
3. Kadampur, Mohammad Ali, and Sulaiman Al Riyae. "Skin cancer detection: applying a deep learning-based model driven architecture in the cloud for classifying dermal cell images." *Informatics in Medicine Unlocked* 18 (2020): 100282.
4. Akram, Tallha, et al. "A multilevel features selection framework for skin lesion classification." *Human-centric Computing and Information Sciences* 10 (2020): 1-26.
5. Marka, Arthur, et al. "Automated detection of nonMalignant skin cancer using digital images: a systematic review." *BMC medical imaging* 19.1 (2019): 21.
6. Gaonkar, Rohan, et al. "Lesion analysis towards Malignant detection using soft computing techniques." *Clinical Epidemiology and Global Health* (2019).
7. Hekler, Achim, et al. "Superior skin cancer classification by the combination of human and artificial intelligence." *European Journal of Cancer* 120 (2019): 114-121.
8. Rajasekhar, K. S., and T. Ranga Babu. "Skin Lesion Classification Using Convolution Neural Networks." *Indian Journal of Public Health Research & Development* 10.12 (2019): 118-123.
9. Iyer, Vijayasri, et al. "Hybrid quantum computing based early detection of skin cancer." *Journal of Interdisciplinary Mathematics* 23.2 (2020): 347-355.
10. Roslin, S. Emalda. "Classification of Malignant from Dermoscopic data using machine learning techniques." *Multimedia Tools and Applications* (2018): 1-16.
11. Moqadam, Sepideh Mohammadi, et al. "Cancer detection based on electrical impedance spectroscopy: A clinical study." *Journal of Electrical Bioimpedance* 9.1 (2018): 17-23.
12. Hosny, Khalid M., Mohamed A. Kassem, and Mohamed M. Foad. "Skin cancer classification using deep learning and transfer learning." *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*. IEEE, 2018.
13. Dascalu, A., and E. O. David. "Skin cancer detection by deep learning and sound analysis algorithms: A prospective clinical study of an elementary dermoscope." *EBioMedicine* 43 (2019): 107-113.
14. M. Vidya and M. V. Karki, "Skin Cancer Detection using Machine Learning Techniques," 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2020, pp. 1-5, <https://doi.org/10.1109/CONECCT50063.2020.9198489>.

Manufacturing of Medical Devices Using Artificial Intelligence-Based Troubleshooters



Akbar Doctor

1 Review of Literature

As discussed by Siirtola et al. (2019) [2], we can have 96.5% recognition rates for a supervised learning of an AI agent. Rong Zhang et al. [6] also propose semi-supervised learning, resulting in improved classification accuracy. This in the context of our AI agent would mean to feed it with known troubleshooting solutions and then expect it to provide legit and useful outputs. This definitely looks lucrative but defies the purpose of our agent. The production floor of a complex medical device is always full of surprises and erratic outcomes, very similar to a road traffic as researched by Dinithi Nallaperuma et al. [7]. We, therefore, preferred to have an unsupervised learning of the agent and gauge its outcome accordingly. The AI agent described in this essay works on a similar approach as a look-up table. We have linked our problem XML data to an image. Unlike the study by Iulia Alexandra Lungu et al. [9] wherein the hand recognition symbol algorithm used images, our AI agent used images to identify the errors/warnings in a medical device. Chang Huang et al. [11] described the ill effects of boosting, where the AI agent missed out the benefits of online learning by its offline learnt data. Hence, our AI agent is implemented in an ever learning mode. It updates its knowledge for every novel data input received during production. The agent never forgets the learned knowledge, similar to the approach described by R. Polikar et al. [13]

S Ruping (2001) [1] describes the support vector machine (SVM) for machine learning. This approach, although well-implemented for large- and high-dimension data, does not reap benefits for our application. We do understand that our database of XML files shall grow eventually large enough for the agent to handle with

A. Doctor (✉)

Principal Test Engineer-Fresenius Medical Care, Concord, CA, USA

e-mail: Akbar.Doctor@fmc-na.com

the current look-up table mechanism but for this pilot, we have not implemented SVM. This is something we need to look into for future expansion of our agent's capabilities. Ryszard et al. (1986) [20] implemented a multi-purpose learning system with their TRUNC algorithm. This enables an agent to learn incrementally disjunctive concepts from noisy and overlapping examples. This is true in case of our agent as well due to the fact that one problem on the production floor may have multiple symptoms on the machine and may have multiple solutions. The way our AI agent handles this overlapping knowledge is by reproducing all the available solutions to the operator and lets him choose from the best one. Once the solution works, the agent then specializes its knowledge and hence narrows down the available options for a particular problem. This generalization and specialization cycle is never ending for the AI agent until it has a one–one mapping of problems and solutions. We, therefore, need to pilot this agent for over a year on the floor which should give us better representation of the learnt knowledge.

Swaroop S et al. [4] discuss about the biologically inspired design, where we mimic the existing designs in mother nature to create a manufactured design. It can vary from the propeller to a neural network. Our AI agent also mimics the abstract learning process of a human. It acquires knowledge through experience and senses as pointed out by Langley [5]. But our agent stores the data in the form of XML files as we are in the Pilot state. This makes our data cumbersome to manage and requires more time to come to a conclusion. We plan to look in to the DeeSIL method as described by Eden Belouadah and Adrian Popescu [8] to make our agent more efficient from time efficiency standpoint. Creating a database could also be an alternative.

The works of Ashok K Goel et al. [3] were studied to understand chunking as a way to implement knowledge-based AI. The AI agent described in this chapter learns incrementally. Various works of authors were studied with respect to incremental learning to arrive upon an optimal implementation. Likes of Marc-Schoenauer et al. [10], Giraud-Carrier et al. [14], Boon Keat Puah et al. [15], Chixiao Chen et al. [16], Pramod Sharma et al. [17], Scott H. Clearwater et al. [18], and Chuan-XianRen et al. [19] were evaluated for batch and unsupervised learning of the agent. But this was not apt for our agent. Guorui Feng et al.'s [12] work of incremental learning was also evaluated for this troubleshooting AI agent. The learning employed for the troubleshooting agent of this chapter is a partial supervised learning wherein the operator on the production floor declares a failure which is then used by the agent to learn a new chunk of knowledge. This is described in great detail in Sect. 3 of this chapter.

2 Introduction

Quality is ingrained in the design of a good product. A lot of companies start looking into the quality of their product right during its design phase. This same quality is implemented all the way in production and finally to the release of the product from the facility.

In production, the quality is implemented right at the warehouse where all the raw materials and sub-assemblies are received from the vendors. The incoming inspection process flushes out any discrepancies in the incoming parts which eventually become the part of the end product. Once the parts reach the floor, the sub-assemblies are inspected/tested at every feasible point on the line. This ensures the integrity of the end product. Finally, when the product is fully built, it is tested as part of end-of-line testing. This final testing phase ensures the functional aspect of the product in the field. After passing this particular testing, a product is deemed capable for release. Various testings on the production floor can be summed up as follows:

- Incoming Inspection
 - FAI—First Article Inspection
 - AQL Sampling
- Sub-Assembly Testing
- Functional Test
- End-of-Line Test

Hence, quality is part of the entire production process which is manifested as testing at various levels in production. Quality is of prime importance for any company to avoid loss of business, liability, and warranty cost. All the investment in quality reaps immense amount of benefits when it comes to business and popularity of a product. But for a medical device, quality has a special attribute to it. Quality problems for medical devices may have a negative impact on patients. Hence, when it comes to medical devices, quality directly translates to human lives and well-being. This makes medical device invest more into testing to ensure a quality product output.

Investment in design verification and testing on the production floor are two most important pillars on which the edifice of a quality product is built. Testing starts right in the design phase for a product. For the medical device, the Design Quality Assurance (DQA) team leads and qualifies the efforts for design verification. During a new product development, DQA teams drive the verification efforts which in turn enable the companies to file 510k for the medical device. 510k approval enables medical device companies to sell their product in the US market and elsewhere. Design verification and related testing are important aspects for 510k filing.

After the medical device is in the market, any future design changes also need to go through the DQA team. If the changes to the design are significant, then new 510k may need to be filed with FDA for approval. After the design has been released to manufacturing, R&D (Research and Development) usually releases test specifications which need to be met during the production of the device on the floor.

The manufacturing engineering team takes these test specifications from R&D as an input and devises a test strategy for the production floor. This strategy includes various sub-assembly level testing, inspection, and all the way to end-of-line testing.

As a lot of testing is performed on the production floor, it is very likely that this will also lead to a lot of fall-outs. Medical device industry deals with fall-outs in

a variety of ways. For sub-assembly level fallouts, some industries would discard the part while some would implement a rework. But for end-of-line testing, most of the companies would implement a rework and re-test process as it is not at all economical to discard and scrap an entire product if it fails testing. This is where many a time automation is not possible. Troubleshooting and rework in most of the companies are driven by manual labor. Operators in the troubleshooting area follow a troubleshooting manual/guide and perform diagnostics. These diagnostics can sometimes be very structured and process specific but in many cases, the diagnostics process can also be dependent on tribal knowledge of operators. In such a scenario, losing a skilled laborer may disrupt the normal operations. In many geographic locations where unemployment rate is low, attrition rate may be very high. This leads to high vulnerability of losing skilled labor at regular intervals and hence disruption of production.

3 Method

Troubleshooting is a skill which is acquired by gaining on the job experience. Every one of us remembers the story of a mechanic who charged a hefty amount for fixing a ship by tapping the engine with a hammer. When asked about the amount charged for just one tap of hammer, the mechanic replied that the money was not about tapping the hammer but for the knowledge of where to tap the hammer which fixed the problem.

The above story holds true for repairing/reworking a complex medical device as well. When a new employee joins a company and starts working on the production floor, he/she may not have the skill to perform diagnostics and fix the problem. An operator is trained before starting work on the floor and is provided with a troubleshooting guide. This enables and kick-starts the operators with the initial knowledge required to work on the machines. But for a complex machine with hundreds of sensors and actuators inside, a troubleshooting guide may run from a couple hundreds to thousands of pages. Surfing through such a huge troubleshooting guide and getting acquainted with the process take time and hands-on experience. For such a complex process, an operator may take a couple of months to a year to claim expertise. Hence, this expertise gaining process requires a huge investment from time and money perspective but this can all go in vain if the individual leaves the organization and if a new individual is hired for this position again.

A lot of companies suffer from this issue of high attrition and lose a lot of time and money in training and re-training the workforce. To combat this, a solution is desired which will remove the human dependency from this process. We need a solution which will mimic humans in its learning capability as well as provide a solution based on that knowledge. Hence, this solution should mimic humans and learn the skills and know-how of the chore bit by bit as a human would do. At the same time, when it is presented with a novel problem, it should try everything it knows and try to solve the problem, but if it cannot, then it should learn a new

technique and increase its knowledge. The only solution which can achieve and mimic human intelligence is artificial intelligence. An artificial intelligent agent can be implemented using incremental learning cognitive capability which may chunk knowledge bit by bit and gain troubleshooting experience like a human. A clear advantage of this would be a human-like agent gaining knowledge and retaining it for life time. This will make a company independent of the tribal knowledge of an individual which leaves the company along with the individual.

3.1 AI Agent

Hemodialysis is a complex process, which needs to be performed on patients with end stage renal disease periodically. In dialysis, there are many elements involved like fistula, bloodlines, dialyzer, and the dialysis machine. A dialysis machine, in particular, is a very important and perhaps the most complex element of the dialysis process. A hemodialysis machine performs critical functions of pumping the blood out of the patient, injecting heparin in it to avoid clotting, maintaining its temperature while the blood is out of the patient and flowing through the blood lines and dialyzer. Along with this controlled blood flow, the dialyzer also regulates the flow of the dialysate fluid through the dialyzer. This dialysate is mixed in real time from the acid and bicarbonate solution and its flow and temperature are also maintained by the dialysis machine. Considering this complex functionality, the HD machine consists of various actuators and sensors making it one of the most complex machines to troubleshoot and rework if it breaks.

To troubleshoot such a complex machine is a skill, which is acquired by repetitive execution of the process. Every time a new problem is presented and solved by the troubleshooter, he or she learns a new thing for that instant. Using this paradigm of events, we decided to create an AI agent which will also follow the same pattern.

3.2 Human Cognition and the AI Agent

For humans, we have our cognitive capabilities arising from our senses (see, touch, feel, hear, and smell) and our human brain, which reasons, learns, and stores memory. For this troubleshooter AI agent, we first had to create these capabilities in order to mimic human intelligence. For a dialysis machine, all the errors and fall out information are available on the screen of the HD machine. An operator reads this information and the DHR (device history record) of the machine in order to figure out the problem. Once this input is available to the operator, he/she would surf through the troubleshooting guide to figure out the solution. If a direct solution is available for a given problem, the operator would execute it and repair the machine. If there is no direct solution available, then the operator would consult the engineering and R&D group to figure out the solution and then update the

troubleshooting guide with this new solution. This process is followed for every fallout of testing on the production floor.

To mimic this process, an AI agent needs the following capabilities:

- Visually capture the problem by looking at the screen
- Process the image to identify the errors/alarms/warnings on the screen
- Search solution from a database
- Provide the solution to the operator to execute
- If no solution is available, then learn the new solution and store it in a database (*Learn and Memorize*)

3.2.1 Capture the Image and Identify the Problem

For the AI agent, we provided the visual sensory capability in the form of a camera and python code for processing it. The AI agent is basically a python program which acquires images from a camera mounted on an AIO (all-in-one desktop). This AIO is part of the troubleshooting area on the production floor. The captured image is processed to crop out the error message which satisfies the visual aspect of the AI agent. Figure 1 shows some prospective error messages on the HD machine. The fallouts/repair messages are highlighted as red- and yellow-bordered pop-ups.

The next task is to identify and understand the messages as a human would do. To perform this task, we had two options to choose from:

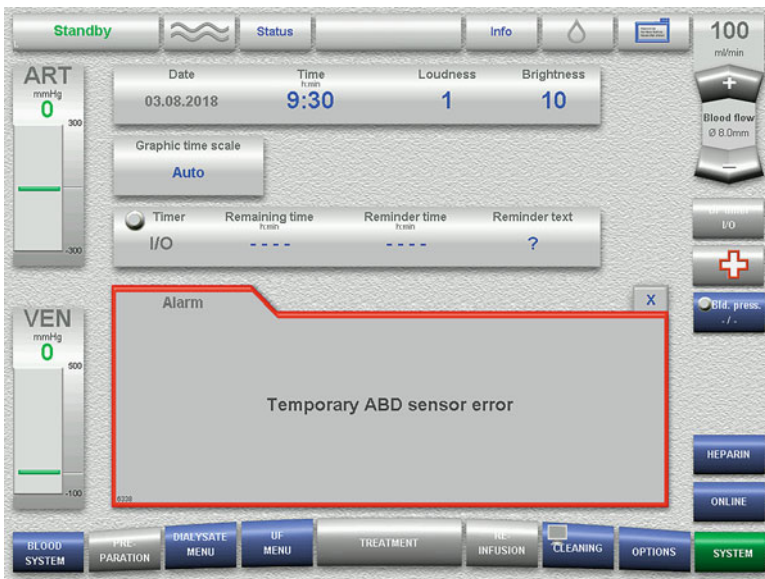


Fig. 1 Fallout/repair messages for the HD machine

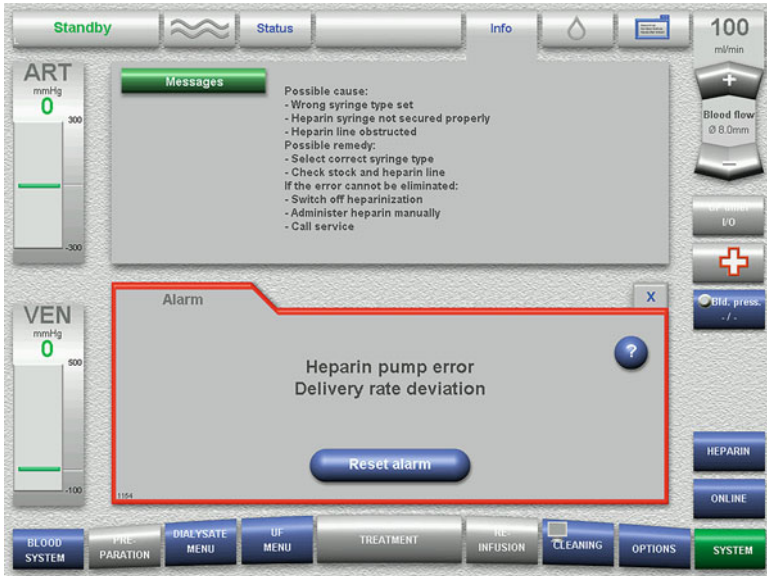


Fig. 2 XOR → Identify the problem

- Image processing and XOR (exclusive OR)
- OCR (optical character recognition)

Image Processing and XOR

For this process, we crop the part of the image which has errors/warnings in it and then perform an XOR with all the available images from R&D. We store all the available images from the R&D group in a secured location which are used to perform this XOR. Once we receive a 98% white image after an XOR, we determine that as the detected error/warning. Hence, by following this process, we are able to identify the error on the HD screen as one of our stored images. This identified image is then used as a pointer toward the potential solution in the database, described in the next section. Figure 2 shows an example of one of this process of XOR.

OCR

For this process, we create a bounding box on the captured image and perform character recognition. This provides us with the exact text of the error/warning which is then used to point toward the database for solutions. This method requires creating a bounding box on all images and creating a dictionary file for the fonts in

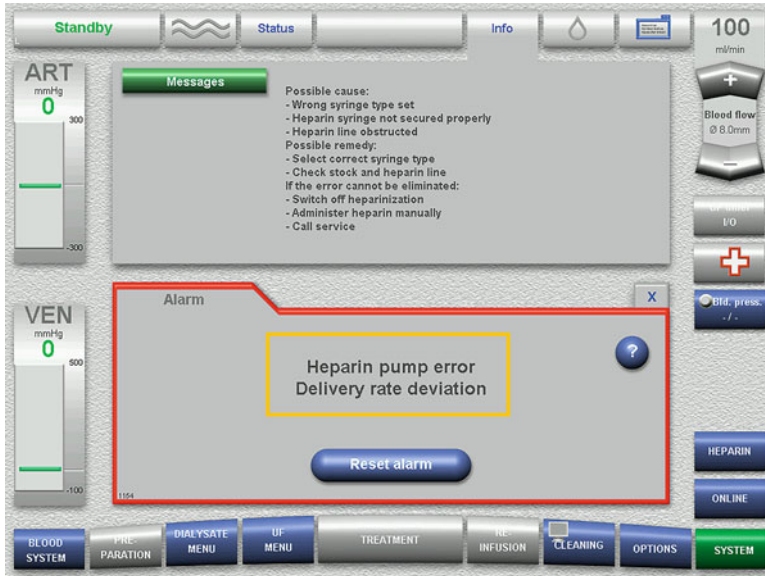


Fig. 3 OCR → Identify the problem

the text as well. Figure 3 shows an example for this method. The yellow box is the bounding box inside the image in which we have the error/warning message.

Both the above methods of identifying the error/warning are equally effective but there are certain pros and cons to both. OCR is low on storage requirements and lower processing time in real time. But it requires a lot of time and effort to create a bounding box for every image inside the image. This is not only labor-intensive for an engineer but it is also very prone to changes in the future. Whenever a change is made by the R&D group on the machine software, an engineer shall need to verify all the position of the bounding box on each image and may need to rectify on few images where R&D had changes in verbiage or position of the error message. Hence, if the OCR method is used, the AI agent will be required to be updated every time with R&D changes related to HD machine software. In any industry, changes are very common. Given this fact, we leaned toward using the XOR method.

The XOR method as described above follows a simple process to identify the image on the HD machine screen without any customization. Hence, for XOR, if R&D changes any image verbiage/message or its position on the screen, we need not update the AI agent at all. We only need to replace the stored HD machine software images with the new images from R&D in the secured location. The AI agent after capturing the HD machine image from the camera will run through all the images in the secured location and identify a particular image, once it finds a 98% match. This method, therefore, is less prone to changes but requires a longer processing time as the AI agent needs to run through the entire set of images in the secured location before it finds a match. As our pilot runs on the production floor,

the AI agent takes around 42 seconds to identify the image. This is good enough for the production floor as an operator would take more than a minute to grab his tools after parking the HD machine in the troubleshooting area.

3.2.2 Search Solution from Database and Provide It to the Operators

For the AI agent to be able to reason, we created a mapping of the stored images to the possible solutions. These solutions are stored as separate XML (extensible markup language) files in a secured location. We did not implement a database during the pilot as playing around with XML files is easier and less programming intensive. This was also preferred as the DHR of the HD machine is available as an XML file.

As described in Sect. 3.2.1, after identifying the image on the HD machine screen, the next step is to search for a solution and provide it to the operators. To do this, the AI agent searches the XML file associated with the identified image. If a solution XML file is available with the identified image, then the AI agent would grab that file from the secured location, decipher it, and present the solution to the operator. If no file is available, then the AI agent would let that operator know that it does not know of any solution and will need to learn from this event.

3.2.3 Learn New Solutions

As described in Sect. 3.2.2, if the agent does not find any XML file, then it pops-up a prompt to the operator to consult the troubleshooting guide or seek assistance from the engineering or R&D group to solve this problem. As in any troubleshooting process, a log is created after the problem has been solved. This log is part of the MES system (Manufacturing Execution System), which stores data in the form of an XML file. As the agent is in learning routine for this particular HD machine, it will reach out to MES to grab the XML file for this particular solution and store it in the secured location. This enriches the database knowledge of the AI agent and enables it to provide solution next time when a similar problem arises in the troubleshooting area.

This process is what we call incremental learning. The AI agent collects knowledge bit by bit very similar to humans and hence becomes more competent as a troubleshooter.

4 Results

The AI agent was implemented outside the production environment. We started off with only five solutions provided to the agent as XML files linked to few images of errors/warnings. The expectation was that the agent will go into learning phase a

lot more in the initial days and will build up the database (linked XML files in the secured folder) as it learns more and more.

We monitored the agent and the secured folder for the first 7 days. Considering the fall-outs on the floor, the agent captured all the learnings and reproduced the solution second time when the problem occurred. This proved the working of the incremental learning concept of our agent. It built up 70% of the troubleshooting guide in relation to the errors/warnings it was exposed to during those 7 days.

The results of the study are promising and motivate us to roll it out to the production floor.

5 Conclusion

The agent shows promising performance in the troubleshooting area. As expected, it makes the troubleshooting process independent of the skills of an operator. Once the AI agent is exposed to a problem and learns the solution, it never forgets it. When presented with the same errors/warnings on the HD screen, the AI agents present the operator with all possible solutions/remedies for those errors in the HD machine.

This excites us to implement this as a full-fledged production process. But for this, the agent will have to be updated to abide by 21-CFR (Code of Federal Regulation) Part 11 and other cGMP guidelines. For this, the agent will need to go through complete CSV validation (Computer Software Validation) before implementing it on the floor.

5.1 Future Prospects

The AI agent for now learns incrementally, which is mimicking human behavior of learning the unknowns. But humans do have another capability, which is extrapolating and guessing a solution when they cannot deduce one. This gives them a creative attribute and helps them to learn from their experiences. We plan to assign this same capability to our AI agent by coding and implementing case-based reasoning in it. This should enable the agent to extrapolate a solution by building on the knowledge it has and by measuring the difference between the problems known to it. This could be easily understood by a simple example of going to a place whose address is unknown to you. Imagine a person knows where restaurant A is. If someone describes to a person the address of restaurant B by mentioning it as 30 feet from restaurant A, then an individual can navigate to restaurant B by having restaurant A as the reference. Here the difference of known and unknown is 30 feet. We can have similar distance in terms of pixel value, problem connection, etc. We will be excited to see how an AI agent will become closer to human intelligence and how it will improve the manufacturing floor further.

Acknowledgement I would like to thank and acknowledge Fresenius Medical Care for the use of pictures of its 5008S Hemodialysis machine (HD). Figure 2 and 3 depict the error/warning messages from the 5008S hemodialysis machine which is used in this paper to explain the detection algorithm for problems on the production floor. The 5008S HD machine is operational in the European markets.

References

1. S. Ruping, "Incremental learning with support vector machines", *Proceedings 2001 IEEE International Conference on Data Mining*, 2001
2. Pekka Siirtola et al. "Incremental Learning to Personalize Human Activity Recognition Models: The Importance of Human AI Collaboration", *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 2019
3. Ashok K. Goel et al. "Artificial Intelligence", *In Cambridge Handbook of Intelligence (3rd Edition)*, R.J. Sternberg & S.B. Kaufman (Editors) pages 1–28, 2011.
4. Swaroop S. Vattam et al. "Compound Analogical Design: Interaction between Problem Decomposition and Analogical Transfer in Biologically Inspired Design", *In Design Computing and Cognition '08*, pages 377-396, 2008.
5. Pat Langley, "The Cognitive Systems Paradigm", *Advances in Cognitive Systems 1*, 2012.
6. Rong Zhang; A.I. Rudnicky, "A New Data Selection Principle for Semi-Supervised Incremental Learning", *18th International Conference on Pattern Recognition (ICPR'06)*, 2006.
7. Dinithi Nallaperuma et al., "Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management", *IEEE Transactions on Intelligent Transportation Systems (Volume: 20, Issue: 12, Dec. 2019)*, 2019
8. Eden Belouadah, Adrian Popescu, "DeeSIL: Deep-Shallow Incremental Learning." *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018.
9. Iulia Alexandra Lungu; Shih-Chii Liu; Tobi Delbruck, "Incremental Learning of Hand Symbols Using Event-Based Cameras", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems (Volume: 9, Issue: 4, Dec. 2019)*, 2019.
10. MarcSchoenauer, MichèleSebag, "Incremental Learning of Rules and Meta-rules", *Proceedings of the Seventh International Conference, Austin, Texas, June 21–23, 1990*, 1990, Pages 49-57, 1990.
11. Chang Huang et al., "Incremental Learning of Boosted Face Detector", *IEEE 11th International Conference on Computer Vision*, 2007.
12. Guorui Feng et al., "Error Minimized Extreme Learning Machine With Growth of Hidden Nodes and Incremental Learning", *IEEE Transactions on Neural Networks (Volume: 20, Issue: 8, Aug. 2009)*, 2009.
13. R. Polikar et al., "Learn++: an incremental learning algorithm for supervised neural networks", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) (Volume: 31, Issue: 4, Nov 2001)*, 2001.
14. Giraud-Carrier, Christophe, "A note on the utility of incremental learning", *AI Communications*, vol. 13, no. 4, pp. 215-223, 2000.
15. Boon Keat Puah et al., "A regression unsupervised incremental learning algorithm for solar irradiance prediction", *Renewable Energy* Volume 164, Pages 908-925, 2021.
16. Chixiao Chen et al., "OCEAN: An on-chip incremental-learning enhanced processor with gated recurrent neural network accelerators", *43rd IEEE European Solid State Circuits Conference*, 2017.
17. Pramod Sharma et al., "Unsupervised incremental learning for improved object detection in a video", *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

18. Scott H. Clearwater et al., "INCREMENTAL BATCH LEARNING", *Proceedings of the Sixth International Workshop on Machine Learning, Pages 366–370*, 1989.
19. Chuan-XianRen et al., "Incremental learning of bidirectional principal components for face recognition", *Pattern Recognition, Volume 43, Issue 1, Pages 318–330*, 2010.
20. Ryszard Michalski, I Mozetic, J Hong, N Lavrac, "The multi-purpose incremental learning system AQ15 and its testing application to three medical domains", *AAAI-86 Proceedings*, 1986

Enhanced Hierarchical Prediction for Lossless Medical Image Compression in the Field of Telemedicine Application



Ketki C. Pathak, Jignesh N. Sarvaiya, and Anand D. Darji

1 Introduction

The evolution in information intensive multimedia-based web application combined with advancement in the field of picture measuring devices plus processing components in cell phone has yielded in compactness of multimedia subject matter pivotal to storage capacity and transmission field. Accurate data with picture compression algorithms are thus essential in order to reduce expenditure coupled with requirements particularly concerning gadgets and systems having constrained assets. Most important techniques are lossy and lossless, for reducing the information. Depending on application, the lossy technique can be used to increase channel throughput, and some advanced lossy algorithms also provide a high compression ratio.

During the image acquisition procedure, there can be higher chances of obtaining damaged medicinal images due to unknown faults or noise from surroundings, which may occur in the capturing device. These damaged medicinal pictures do impact in analysis of the patient [1]. The process involving the elimination of distortion that occurred on damaged pictures is also a challenging work. In telemedicine treatments, medicinal pictures have to be collected previously in order to broadcast them, here arises the requirement of accurate compression algorithms to transfer them on low bandwidth. Image compression algorithms minimize the storage necessity and bandwidth of radiocommunication networks. In comparison

K. C. Pathak (✉)

Sarvajanik College of Engineering and Technology, Surat, India

e-mail: ketki.joshi@scet.ac.in

J. N. Sarvaiya · A. D. Darji

Sardar Vallabhbhai National Institute of Technology, Surat, India

with conventional algorithms for compression, progressive data compression for medicinal conferencing was not quick in the direction of simulating huge data [2].

In lossless compression, the main concern is to preserve the original visual quality with some compromise on image size reduction or on compression ratio. In the medicinal image, the picture condition is the prime concern where minor damage in visual information cannot be acceptable; therefore, the lossless compression algorithms are most suitable in this imaging [3, 4]. We proposed a lossless compression method using Modified Hierarchical prediction with context adaptive encoding.

As per the theory of the color image, every color image is represented as 24 bits, and each color channel is of 8 bits (RGB color channel). For any color image, the principle behind the compression algorithm is application of any reversible color transform on the RGB color channel to minimize the correlation among inter channels, i.e., luminosity Y channels and chrominance C_u and C_v channels. For direct color image compression, it is very essential to choose a proper reversible color transform (RCT) method, as some of the color transformation might not stay exactly reversible owing to damage of exactness in non-integer computations in direct and inverse transforms [5].

A variety of lossless algorithms were previously suggested like lossless JPEG [1], JPEGLS [2], LOCO-I [3], CALIC [4], JPEG 2000 [5], and JPEG –XR [6]. Amongst them, CALIC outperforms rest of the algorithm at the cost of computational complexity. In lossless compression, maximum of the hue changes cannot be applied owing to their invariability by means of numeral calculation. In JPEG 2000, a revertible type color transform (RCT) has been demarcated [6]. There are many research papers proposed on RCT methods. Among them, the RCT transform suggested in [7] is better as it approximates YCbCr transform in a good manner.

For demonstrating higher bit range pictures over lower bit resolution displays, bit resolution needs to be lowered. In lossless, several methods to estimate are dependent on the raster calculation which is fruitless on the higher frequency segment. For fixing this issue, the former study using ordered estimation strategies which consist of the edge-directed forecaster and context adaptive system and are found to be good for such application [8].

The conventional prediction-based image compression algorithm, which utilizes the raster scan-based calculation approach, is probably insufficient at the high frequency regions where sharpness of the edge is much more important. The hierarchical prediction algorithm suggested in Ref. [9] uses pixel-based interpolation, while in the new approach of hierarchical prediction [10], the edge-based adaptive predictor along with context-based adaptive modeling is analyzed.

In our work, first we have proposed two methods based on hierarchical estimation and context adaptive coding and evaluated on color medical images, presented in Sect. 2. Second, we have established a new approach of hierarchical prediction again using two methods represented in Sect. 3. Section 4 gives the comparative result discussion on hierarchical prediction with modified hierarchical prediction.

2 Hierarchical Prediction

Hierarchical prediction is a lossless image compression technique [11–14]. There are many existing prediction strategies intended for lossless compression; however, the fundamental issue is the raster scan estimation technique which can be sometimes ineffective in certain circumstances, chiefly at the high frequency sections.

In hierarchical estimation, it is widely applied in several applications for picture compression. It is meant for picture element interpolation and it is the most precise de-correlation scheme. Directional estimation pixels are predicted by hierarchical decomposition. The hierarchical assembly provides more precise framework patterning of the picture element by means of neighboring picture elements that have been previously programmed; however solitary fundamental data need to be operated for the raster scan scheme [3].

The hierarchical forecast method progresses in a random manner, and restoration of the picture is completed beginning at the lowermost point of the resolution level till reaching the uppermost point. It is utilized in picture element interpolation and it is the utmost effective de-correlation scheme. For hierarchical forecast-based image compression, the image is separated into two sub pictures, an odd sub picture and an even sub picture. In this prediction-based structure, predictors are used.

Figure 1 demonstrates that the pixel in a sample picture X is split into dual sub images, one being even sub image X_e and another odd sub image X_o . Then, X_e gets programmed first and is used to guess the pixels of X_o [1]. A hierarchical prediction structure provides coding efficiency and temporal scalability. The following subsections explain this.

A. Coding Efficiency

Coding proficiency depicts pace and program design methodology and reliability for obtaining a set of instructions meant for applications. It straightaway links through algorithmic productivity and rapidity of runtime performance intended for

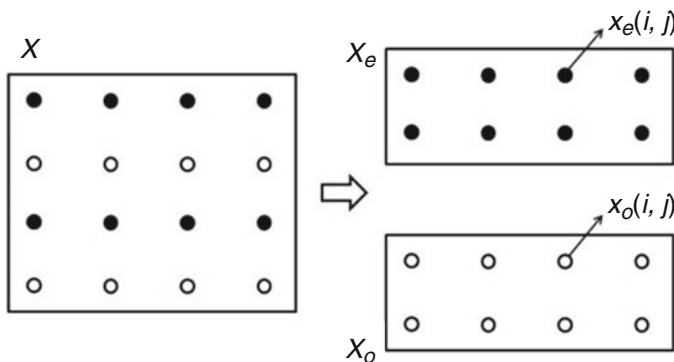


Fig. 1 Sample image and its disintegration [1]

the software system. The coding proficiency decreases the source depletion as well as accomplishment time as much as possible besides a small amount of threat in commercial or functioning circumstances. The different aspects of programming efficiency depend on CPU time, data storage, input/output time, and programming time. The time needed for the CPU for performing the operations that are described in declarations determines the intricacy of the program. Executing and compiling programs require time and space. According to reduce CPU time and to make the program more efficient, one should execute only unavoidable statement, perform calculations for the mandatory observations, and lessen the number of tasks to be executed in a specific statement.

Dropping Input/output duration and CPU use is significant although by means of methods that are productive with respect to the program design duration taken to progress, mend, and authenticate the code that can be even more overvalued. More accuracy is achieved by comprehending right computing rehearses aimed at legibility and maintainability of the code.

B. Temporal Scalability

Time-based quantifiability denotes the capacity to decrease the frame rate of a prearranged bit stream by means of reducing packets, and in so doing, decreasing the bitrates of the stream. While fragments of the stream can be detached by means that the resultant sub stream makes one more acceptable bit stream intended for some target interpreters, and the sub stream signifies the foundation content with a frame proportion lesser than the frame rate of the whole original bit stream, then a cinematic bit stream is called temporal scalability [16].

As shown in Fig. 2, temporal scalability is accomplished via segmenting access elements of bit stream into a temporal base level. One or more temporal improvement coatings along the next property: Let the temporal levels be recognized by a temporal level classifier T, beginning at 0 for the base level, which then amplified one after one to the next. Subsequent to each natural numeral k, bit stream gained by eradicating entirely access units of all temporal layers with a temporal layer identifier T is greater than k and establishes one more effective bit stream for the given decoder.

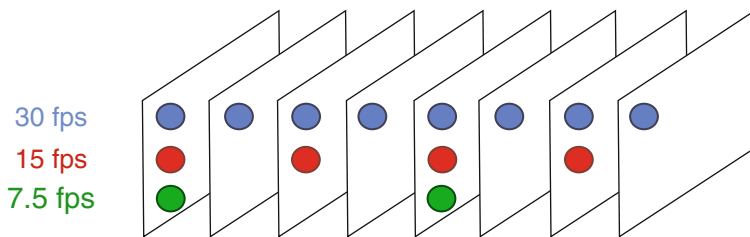


Fig. 2 Temporal scalability [16]

Benefits of Temporal Scalability

1. Bit-rate scalability at access by clipping the frame-rate.
2. Power scalability or MHz at decipherer by picking a lower frame-rate to decrypt.
3. Regressive compatible using prevailing H.264/AVC decipherers – not at all any necessity for an SVC decipherer to decrypt SVC temporal scalability stream. The supplementary header spaces in the stream for assisting the separator can be (and are expected to be) securely disregarded by means of the H.264/AVC decoder. Temporal scalability is easy to accomplish in contrast to other forms of scalability.
4. Negligible or no loss in coding efficiency: For many platforms with an adequate search range, there can be gain in coding efficiency.

2.1 Proposed Method for Hierarchical Prediction

For the compression of images, the RGB image is first altered to YCuCv by means of an alterable color transform technique, and the brightness channel Y channel is determined using a traditional gray scale image compression algorithm. The chrominance channels Cu and Cv resulting out of RCT typically have changed data from Y and are also different from the original color planes R, G, and B. For encoding chrominance channels (Cu and Cv), the hierarchical prediction method is used. Figure 3 shows the general block diagram of the hierarchical prediction scheme for image compression.

Hierarchical prediction and context adaptive coding makes easy uses of left, upper, and lower pixels for pixel estimation. For efficient lossless compression, a hierarchical disintegration system is used, showing a picture element for the input image, where the chrominance image X is divided into two sub images: an even sub image Xe and an odd sub image Xo.

At that point, Xe gets encrypted initially and utilized to expect the picture element in Xo. Additionally, Xe is likewise utilized to estimate the information of prediction errors of Xo. For the compression of Xo pixels using Xe, directional prediction is used to evade large prediction faults adjacent the boundaries.

One of them is carefully chosen as a predictor for $x_o(i, j)$. With these two probable interpreters, the most communal style to encrypting is “mode selection,” where the improved interpreter for the individual picture element is carefully chosen and the mode (horizontal or vertical) too is conveyed as adjacent information. Yet, the vertical predictor is further often more accurate than the horizontal one because upper and lower pixels are used for the “vertical,” whereas just a left pixel is used for the “horizontal.”

For each pixel $x_o(i, j)$ in Xo, the horizontal predictor $x_h(i, j)$ and vertical predictor $x_v(i, j)$ are defined as follows:

$$x_h(i, j) = x_o(i, j - 1) \tag{1}$$

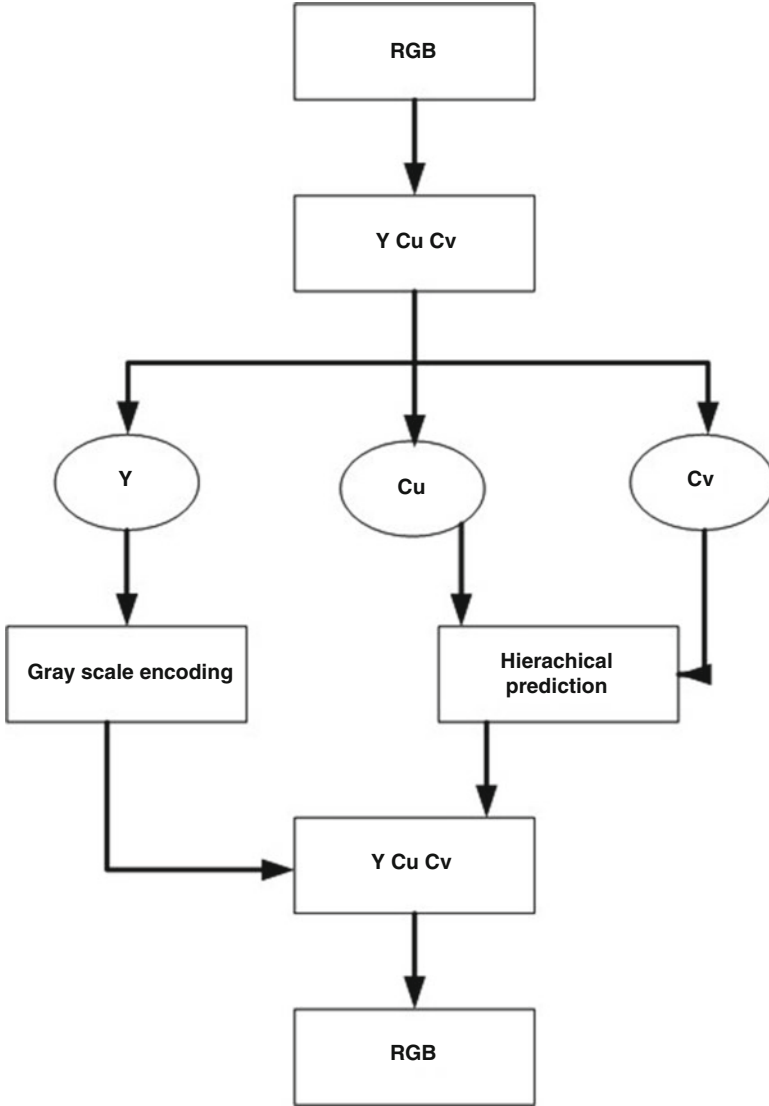


Fig. 3 Proposed block diagram of H.P.C.A

$$x_v(i, j) = \text{round} \left[\frac{x_e(i, j) + x_e(i + 1, j)}{2} \right] \tag{2}$$

The horizontal predictor is more accurate only when there is a strong horizontal edge. Below steps describes the algorithm flow for hierarchical prediction.

Algorithm of Hierarchical Prediction

1. The input RGB color image is transformed into YCuCv color space by RCT.
2. The luminance image Y is encoded by the conventional gray scale encoder.
3. The chrominance images Cu and Cv are encoded using the hierarchical prediction method. The chrominance image is separated into two sub images: an even sub image and an odd sub image. Then even sub image is encoded first and is used to predict the pixels in the odd sub image.
4. A variable for the direction of the edge at each pixel $dir(i, j)$ is defined, which is given either horizontal direction (H) or vertical direction (V).
5. Horizontal predictor and vertical predictor take i and j as argument and returns the left pixel value of average of top and bottom pixel value, respectively.
6. If the direction for left and bottom pixel is H, then it checks the direction for the current pixel. If the current pixel's direction is H, then the value of the left pixel is used as the predicted value, otherwise the average value of top and bottom pixels is used as the predicted value.
7. In the end, we convert the altered YCuCv image back to the RGB image and save the image files.

The hierarchical calculation and context adaptive coding make easy uses of left, upper, and lower pixels for the picture element prediction. We have proposed a method in which luminance channel Y is also encoded by the hierarchical prediction method. For the efficient lossless compression, a hierarchical disintegration system as portrayed in Fig. 4 illustrates the picture element in an input image the luminance and the chrominance image X is divided into two sub images: an even sub image X_e and an odd sub image X_o . Then, X_e is encoded first and is used to predict the pixels in X_o . For the compression of X_o pixels using X_e , steering predictors are employed to avoid large prediction errors near the edges.

The algorithm flow of this hierarchical decomposition method named as HPCA (hierarchical prediction and context adaptive coding) method 1 is explained in the following steps.

Algorithm of Hierarchical Prediction and Context Adaptive Coding Method 1

1. An input RGB color image is transformed into the YCuCv color space by an RCT.
2. The Luminance image is encoded using the hierarchical prediction method. The luminance image is separated into two sub images: an even sub image and an

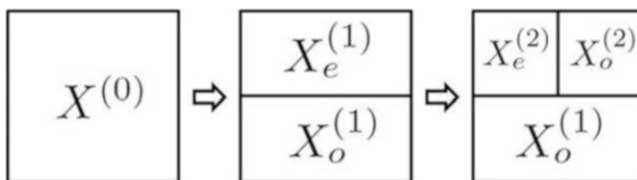


Fig. 4 Illustration of hierarchical decomposition (HPCA method 1)

odd sub image. Then, an even sub image is encoded first and is used to predict the pixels in an odd sub image.

3. *A variable for the direction of the edge at each pixel $dir(i, j)$ is defined, which is given either horizontal direction (H) or vertical direction (V).*
4. *Horizontal predictor and vertical predictor take i and j as argument and return the left pixel value of average of top and bottom pixel values, respectively.*
5. *If the direction for left and bottom pixels is H, then it checks the direction for the current pixel. If the current pixel direction is H, then the value of the left pixel is used as the predicted value, otherwise the average value of top and bottom pixels is used as the predicted value.*
6. *The chrominance images C_u and C_v are encoded using the hierarchical prediction method. The chrominance image is separated into two sub images: an even sub image and an odd sub image. Then an even sub image is encoded first and is used to predict the pixels in an odd sub image.*
7. *A variable for the direction of the edge at each pixel $dir(i, j)$ is defined, which is given either the horizontal direction (H) or the vertical direction (V).*
8. *Horizontal predictor and vertical predictor take i and j as argument and return the left pixel value of average of top and bottom pixel values, respectively.*
9. *If the direction for left and bottom pixels is H, then it checks the direction for the current pixel. If the current pixel's direction is H as well, then the value of the left pixel is used as the predicted value, otherwise the average value of top and bottom pixels is used as the predicted value. In the end, we convert the altered $YCuCv$ image back to the RGB image and save the image files.*

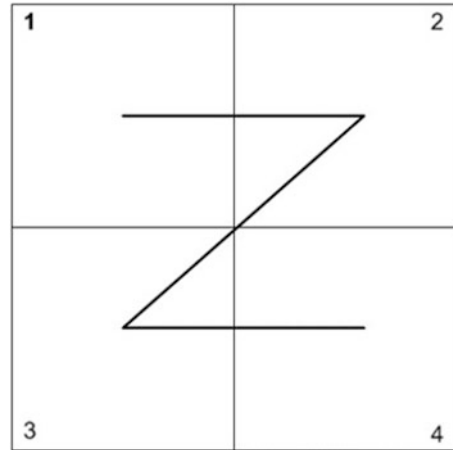
We have proposed the second method for the compression of images, in which the RGB image is first converted to $YCuCv$ by means of the reversible color transform method, and luminance channel Y is programmed by a conventional grayscale image compression algorithm. The chrominance channels C_u and C_v resulting from the RCT usually have different statistics from Y and are also different from the original color planes R , G , and B . For encoding chrominance channels (C_u and C_v), the hierarchical prediction method is used. Each chrominance image is split into four equal parts. The prediction is then taking place in the z-shape order. Top left part is predicted first, then top right and bottom left, and lastly bottom right, which are shown in Fig. 5.

The algorithm flow of this chrominance decomposition method named as HPCA (hierarchical prediction and context adaptive coding) method 2 is explained in the following steps.

Algorithm of Hierarchical Prediction and Context Adaptive Coding Method 2

1. *An input RGB color image is transformed into the $YCuCv$ color space by an RCT.*
2. *The luminance image Y is encoded by the conventional gray scale encoder.*
3. *The chrominance images C_u and C_v are encoded using the hierarchical prediction method. The chrominance image is separated into two sub images: an even sub image and an odd sub image. Then an even sub image is encoded first and is used to predict the pixels in an odd sub image.*

Fig. 5 Illustration of chrominance image decomposition



4. A variable for the direction of the edge at each pixel $dir(i, j)$ is defined, which is given either a horizontal direction (H) or a vertical direction (V).
5. Horizontal predictor and vertical predictor take i and j as argument and return the left pixel value of average of top and bottom pixel values, respectively.
6. If the direction for left and bottom pixel is H, then it checks the direction for the current pixel. If the current pixel's direction is H as well, then the value of the left pixel is used as the predicted value, otherwise the average value of top and bottom pixels is used as the predicted value.
7. Each chrominance image is split into four equal parts. The prediction then takes place in a Z-shaped order. The top left part is predicted first, then top right and bottom left, and lastly bottom right.
8. In the end, we convert an altered YCuCv image back to RGB images and save the image files.

2.2 Result and Discussion of Hierarchical Prediction







Performance parameters are inevitable while trying to compress the images using different technologies. Some of the parameters required to be considered appropriate to measure the effectiveness of any compression algorithm are bits per pixel (BPP), compression ratio (CR), peak signal to noise ratio (PSNR), and threshold value.

In this hierarchical prediction, we have considered performance metrics like BPP, CR, and PSNR. Our proposed schemes mentioned in above algorithms for hierarchical methods are named as HPCA 1 and HPCA 2. Both these approaches exist in the direction of refining the performance of medical image compression. We have conducted experiments on five different medical images, namely, brain, CT scan of heart, shoulder, wrist joint, and chest X-ray to have better understanding of flexibility of our proposed methods. From the experimental results, we can see

Table 1 CR and PSNR values for hierarchical methods

Medical images	Compression ratio (CR)		PSNR (dB)	
	HPCA 1	HPCA 2	HPCA 1	HPCA 2
Brain	36.11	55.75	29.53	28.97
CT scan heart	37.88	56.57	29.59	28.95
Shoulder	37.35	55.91	29.61	28.96
Wrist joint	37.91	58.43	29.58	28.98
Chest X-ray	37.89	54.48	29.5	29.08

Table 2 Visualization of compressed medical images using hierarchical prediction methods

	Original image	YUV image	Compressed image
HPCA method 1			
HPCA method 2			

that HPCA 2 provides with an improved compression ratio compared to HPCA 1 as shown in Table 1. The compression proportion is very little for the prevailing process of HPCA ciphering compared to HPCA 1. Although the proposed HPCA 2 method has better compression proportion, the peak signal to noise proportion is low compared to HPCA 1 for all the specified medical images.

Hence, we can conclude that HPCA 2 is not able to preserve the sharpness of the image, and there is a trade-off between CR and PSNR for the proposed HPCA methods, which does not suffice the motive of medical image compression, i.e., conserving the perceptiveness of the image when reconstructed after compression. We need to have some solution to preserve the sharpness; hence we proposed a modified hierarchical scheme as mentioned in the next section.

Table 1 depicts the compression proportion ranges for both planned hierarchical prediction approaches. The values clearly illustrate that the proposed HPCA 2 has better CR values, whereas HPCA 1 has better PSNR values. Table 2 shows the visualization of YUV and compressed medical images for the proposed HPCA methods.

3 Proposed Method for Modified Hierarchical Prediction

In order to conserve the clarity of initial image and to lessen the bit rates here put forward, a novel method is entitled as modified hierarchical prediction. The calculated image is defined by the horizontal, vertical, and diagonal picture element. By means of the slanting interpreter, we can expect the precise estimated image. Then, an even sub image gets separated stake by stake into an even sub image and an odd sub image. The $X_o(2)$ odd sub image is compressed through which even sub image $X_o(2)$ is created.

These compressed pictures generally have high excellence with distinctness, and while decompressing, we yet again apply the left up, left down, and right up and right down pixels prediction. The proposed block diagram of the modified hierarchical prediction is shown in Fig. 6.

For each pixel $x_o(i, j)$ in X_o , the horizontal predictor $x_h(i, j)$, vertical predictor $x_v(i, j)$, and diagonal predictor $x_d(i, j)$ are defined as follows:

$$x_h(i, j) = x_o(i, j - 1) \quad (6)$$

$$x_v(i, j) = \text{round} \left[\frac{x_e(i, j) + x_e(i + 1, j)}{2} \right] \quad (7)$$

$$x_d(i, j) = \text{round} \left[\frac{x_e(i + 1, j - 1) + x_e(i - 1, j - 1) + x_e(i - 1, j + 1)}{3} \right] \quad (8)$$

Below steps describe the algorithm flow for modified hierarchical prediction.

Algorithm of Modified Hierarchical Prediction

1. The given/sample RGB color image is converted into YCuCv color space by RCT. The luminance image Y is programmed by grayscale image programmers.
2. The chrominance images Cu and Cv are encoded through the modified hierarchical prediction process. The chrominance image is parted into two sub images: an even sub image and an odd sub image. Then an even sub image is programmed first and is used to forecast the pixels in an odd sub image.
3. An inconstant for the direction of the edge at each pixel $dir(i, j)$ is defined, which is given either a horizontal direction (H) or a vertical direction (V).
4. Horizontal predictor and vertical predictor take i and j as argument and return the left pixel value of average of top and bottom pixel values, respectively.
5. If the direction for the left and bottom pixel is H, then it checks the direction for the current pixel. If the current pixel's direction is H as well, then the value of

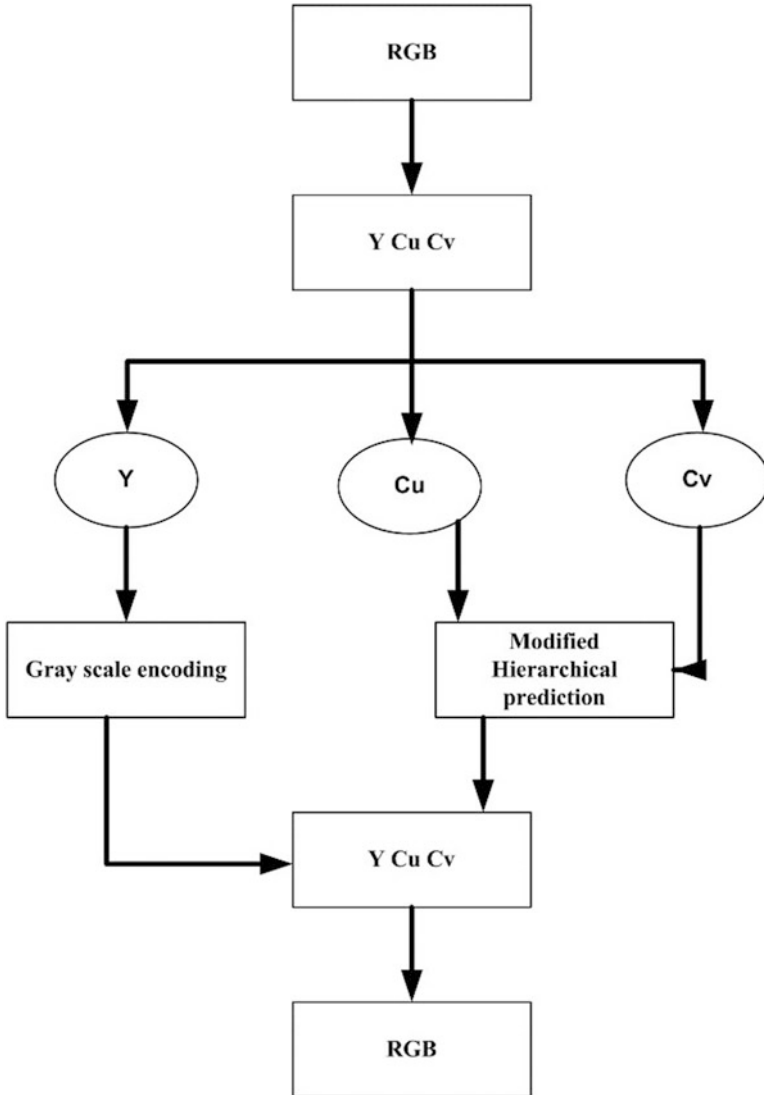


Fig. 6 Proposed block diagram for M.H.P.C.A

the left pixel is used as a predicted value, otherwise the average value of top and bottom pixels is used as the predicted value.

6. *If the top and left pixel direction is H, then the current pixel's direction is calculated. If it is H, then x-horz is used to predict the value; if it is V, x-vert is used to predict the value, otherwise x-diag is used to predict the value. If the current and bottom left pixel direction is V, then x-vert is used to predict the*

value. Lastly, if none of the above conditions are met, x-diag is used to predict the value.

7. *In the end, we convert the altered YCuCv image back to the RGB image and save the image files.*

As modified hierarchical prediction and context adaptive coding makes easy uses of left, upper, and lower pixels for the pixel prediction. We have proposed the image decomposition method in which luminance channel Y is also encoded using the modified hierarchical prediction method.

For efficient lossless compression, a hierarchical decomposition arrangement, as portrayed in Fig. 4, outlines that pixels in an input image represent the luminance, and the chrominance image X is separated into two sub images: an even sub image Xe and an odd sub image Xo. Then, Xe is encoded first and is used to predict the pixels in Xo.

For the compression of Xo pixels using Xe, directional predictors are on duty to avoid large prediction errors near the edges. The algorithm flow of this hierarchical decomposition method named as MHPCA (modified hierarchical prediction and context adaptive coding) method 1 is explained in the following steps.

Algorithm of Modified Hierarchical Prediction and Context Adaptive Coding Method 1

1. *An input RGB color image is transformed into a YCuCv color space by an RCT.*
2. *The luminance image is encoded using the hierarchical prediction method. The luminance image is separated into two sub images: an even sub image and an odd sub image. Then, an even sub image is encoded first and is used to predict the pixels in an odd sub image.*
3. *A variable for the direction of the edge at each pixel $dir(i, j)$ is defined, which is given either a horizontal direction (H) or a vertical direction (V).*
4. *Horizontal predictor and vertical predictor take i and j as argument and return the left pixel value of average of top and bottom pixel values, respectively.*
5. *If the direction for left and bottom pixels is H, then it checks the direction for the current pixel. If the current pixel's direction is H as well, then the value of the left pixel is used as a predicted value, otherwise the average value of top and bottom pixels is used as the predicted value.*
6. *If the top and left pixel direction is H, then the current pixel's direction is calculated. If it is H, then x-horz is used to predict the value; if it is V, x-vert is used to predict the value, otherwise x-diag is used to predict the value. If the current and bottom left pixel direction is V, then x-vert is used to predict the value. Lastly, if none of the above conditions are met, x-diag is used to predict the value.*
7. *The chrominance images Cu and Cv are encoded using the hierarchical prediction method. The chrominance image is separated into two sub images: an even sub image and an odd sub image. Then, an even sub image is encoded first and is used to predict the pixels in an odd sub image.*
8. *Horizontal predictor and vertical predictor take i and j as argument and return the left pixel value of average of top and bottom pixel values, respectively.*

9. *If the direction for left and bottom pixels is H, then it checks the direction for the current pixel. If the current pixel's direction is H as well, then the value of the left pixel is used as the predicted value, otherwise the average value of top and bottom pixels is used as the predicted value. If the top and left pixel direction is H, then the current pixel's direction is calculated. If it is H, then x-horz is used to predict the value; if it is V, x-vert is used to predict the value, otherwise x-diag is used to predict the value. If the current and bottom left pixel direction is V, then x-vert is used to predict the value. Lastly, if none of the above conditions are met, x-diag is used to predict the value. In the end, we convert the altered YCuCv image back to the RGB image and save the image files.*

The predicted image is determined by the horizontal, vertical, and diagonal pixels. Using the diagonal predictor, we can predict the correct predicted image. In the next section, we have proposed the second method for the compression of images, in which the RGB image is first transformed to YCuCv using the reversible color transform method and luminance channel Y is encoded by a conventional grayscale image compression algorithm.

The chrominance channels Cu and Cv resulting from the RCT usually have different statistics from Y and are also different from the original color planes R, G, and B. For encoding chrominance channels (Cu and Cv), the hierarchical prediction method is used. Each chrominance image is split into four equal parts. The prediction then takes place in the z-shaped order. The top left part is predicted first and then top right and bottom left and lastly bottom right parts are predicted as shown in Fig. 5. The algorithm flow of this chrominance decomposition method named as MHPCA (modified hierarchical prediction and context adaptive coding) method 2 is explained in the following steps.

Algorithm of Modified Hierarchical Prediction and Context Adaptive Coding Method 2

1. *An input RGB color image is transformed into the YCuCv color space by an RCT. The luminance image Y is encoded by grayscale image coders.*
2. *The chrominance images Cu and Cv are encoded using the modified hierarchical prediction method. The chrominance image is separated into two sub images: an even sub image and an odd sub image. Then, an even sub image is encoded first and is used to predict the pixels in an odd sub image.*
3. *A variable for the direction of the edge at each pixel $dir(i, j)$ is defined, which is given either a horizontal direction (H) or a vertical direction (V).*
4. *Horizontal predictor and vertical predictor take i and j as argument and return the left pixel value of average of top and bottom pixel values, respectively.*
5. *If the direction for the left and bottom pixel is H, then it checks the direction for the current pixel. If the current pixel's direction is H as well, then the value of the left pixel is used as the predicted value, otherwise the average value of top and bottom pixels is used as the predicted value.*
6. *If the top and left pixel direction is H, then the current pixel's direction is calculated. If it is H, then x-horz is used to predict the value; if it is V, x-vert is used to predict the value, otherwise x-diag is used to predict the value. If the*

current and bottom left pixel direction is V , then x -vert is used to predict the value. Lastly, if none of the above conditions are met, x -diag is used to predict the value.

7. *Each chrominance image is split into four equal parts. The prediction then takes place in the z -shaped order. The top left part is predicted first and then top right and bottom left and lastly bottom right parts are predicted.*
8. *In the end, we convert the altered $YCuCv$ image back to the RGB image and save the image files.*

3.1 Result and Discussion of Modified Hierarchical Prediction

For the lossless image compression, we have proposed schemes on modified hierarchical prediction namely MHPCA 1 and MHPCA 2. The performance metrics used are BPP, PSNR, and CR. The modified hierarchical prediction and context adaptive coding method 1 and the modified hierarchical prediction and context adaptive coding method 2 have been utilized to evaluate the aforementioned performance metrics.

Table 3 clearly shows the visualization of both the modified methods. From investigational outcomes, it is observed that the planned scheme for MHPCA methods produces better PSNR compared to HPCA methods. Table 4 shows the CR and PSNR values for the proposed MHPCA methods. From the table values, it is seen that MHPCA 2 is giving a better compression proportion as well as a better peak signal to noise ratio compared to MHPCA 1. Hence, revised hierarchical prediction and context adaptive coding method 2 does preserve the sharpness of the image along with a high compression ratio.

We have conducted another experiment on the sequence of the chest X-ray medical image to improve the MHPCA 1 method. Table 5 shows BPP, CR, and PSNR, where threshold values are set to $T1 = 2$ and $T1 = 5$. It is seen from the table values that by manipulating the threshold range, we are able to get a better compression ratio with reduced bit rates but at the cost of a less PSNR value. Hence, we can also conclude that MHPCA 1 performs finely when a different data set is provided.

4 Comparative Result Discussion on Hierarchical Prediction with Modified Hierarchical Prediction

In this section, we conclude the overall performance of the combinational methods discussed on hierarchical prediction and context adaptive coding. The performance metric considered is BPP for all the methods. The experimental results clearly state

Table 3 Visualization of compressed medical images using the modified hierarchical prediction methods


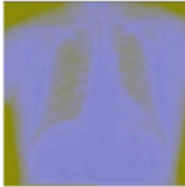


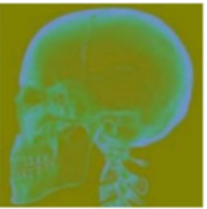

	Original image	YUV Image	Compressed image
MHPCA method 1			
MHPCA method 2			

Table 4 CR and PSNR values for the modified hierarchical methods

Medical images	Compression ratio		PSNR(dB)	
	MHPCA 1	MHPCA 2	MHPCA 1	MHPCA 2
Brain	25.59	40.09	29.42	29.49
CT scan heart	20.54	31.91	29.91	29.91
Shoulder	24.01	37.66	28.89	29.09
Wrist joint	26.64	43	33.16	33.20
Chest X-ray	37.89	54.48	29.50	29.08

Table 5 BPP, CR, and PSNR values of the chest X-ray sequence of images for the modified hierarchical methods with the threshold value set to $T1 = 2$ and $T2 = 5$

Medical images	BPP		CR		PSNR (dB)	
	MHPCA 1	MHPCA 2	MHPCA 1	MHPCA 2	MHPCA 1	MHPCA 2
Chest X-ray_01	2.7753	6.1491	53.6945	37.060	29.2718	29.6641
Chest X-ray_02	2.7316	6.1564	53.7134	36.9493	29.3497	29.7494
Chest X-ray_03	2.8283	6.1980	52.5593	36.2500	28.9720	29.3858
Chest X-ray_04	2.7989	6.3110	54.0320	37.1678	29.1087	29.4961
Chest X-ray_05	2.8548	6.3531	54.1454	37.4912	29.1055	29.5123

that the BPP value is significantly reduced for all medical images by using the proposed MHPCA 2 coding compared to HPCA 1, HPCA 2, and MHPCA 1.

The average bit rate value per pixel is reduced to **2.126** compared to other proposed schemes. Hence, we can justify that the proposed modified hierarchical method 2 is best in preserving the sharpness of the image. Table 6 shows the BPP values for all proposed approaches.

Table 6 Compressed bit rate (bpp) values among all proposed hierarchical predictions

Medical images	HPCA 1	HPCA 2	MHPCA 1	MHPCA 2
Brain	6.68	3.29	12.91	2.96
CT scan heart	6.92	3.53	8.11	1.73
Shoulder	6.86	3.58	14.57	1.93
Wrist joint	6.97	6.49	7.22	1.1
Chest X-ray	6.43	2.91	6.43	2.91
Average	6.772	3.96	9.848	2.126

Table 7 Peak signal to noise ratio (PSNR) values among all proposed hierarchical predictions

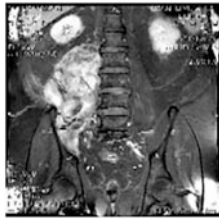
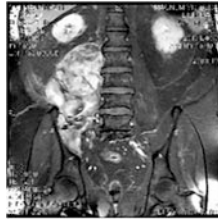
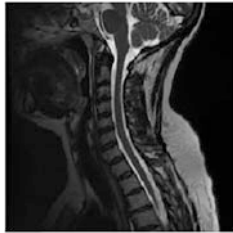

Medical images	Proposed H.P.C.A Method 1	Proposed H.P.C.A Method 2	Proposed M.H.P.C.A Method 1	Proposed M.H.P.C.A Method 2
Abdomen	33.16	48.67	33.26	48.67
Brain 1	37.01	49.10	37.60	50.15
Brain 2	35.38	48.90	35.83	49.98
Brain 3	36.17	48.91	36.60	49.91
Brain 4	35.59	47.83	35.94	49.63
Brain 5	33.58	50.53	34.03	51.46
MRI brain	37.56	55.11	38.95	57.14
Spine	36.71	54.20	38.25	55.52
Wrist joint 1	36.74	54.45	37.91	55.22
Wrist joint 2	36.99	52.63	38.23	54.49
Average	35.88	51.03	36.66	52.21

Table 7 gives the average PSNR values for all proposed approaches of hierarchical prediction, where the experiment is conducted on various medical images, in which the MHPCA 2 method provides an outstanding result compared to other proposed approaches. The sharpness of edges as well as visual quality can be recognized from the images shown in Table 8 and the compression ratio graph of medical images shown in Fig. 7.

We have conducted the experiment on images other than the medical image, which are classical images and are evaluated in terms of BPP for all proposed schemes and compared them with the standard compression methods namely JPEG 2000 and JPEG- XR. Table 9 shows the comparative table with standard compression systems. From the table values, it is seen that the proposed MHPCA methods produce lower BPP than prevailing compression methods.

Figure 8 shows the graph of the compression ratio versus bits per pixel among all proposed hierarchical predictions with existing hierarchical prediction. This graph is plotted for classical image data. It is clearly seen that MHPCA 1 and MHPCA 2 are having higher compression ratio compared to those of the existing HPCA and MHPCA method. Hence, we can conclude that modified hierarchical prediction and context adaptive coding method 2 satisfies all the needs for lossless medical image compression.

Table 8 Visualization of compressed medical images using the proposed hierarchical prediction methods

Proposed method	Method 1	Method 2
HPCA		
MHPCA		

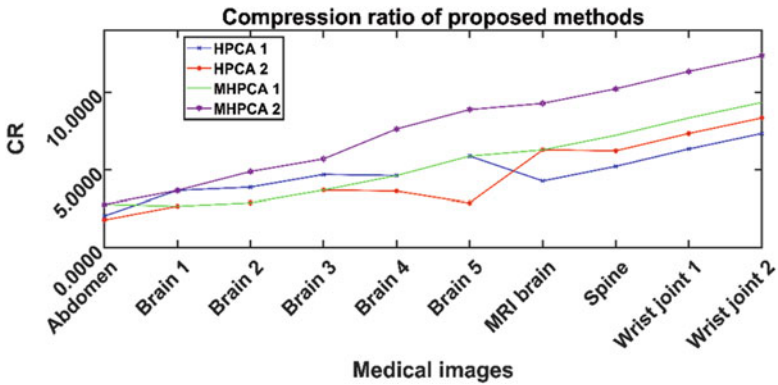


Fig. 7 CR of all proposed hierarchical methods on medical images

Table 9 Comparison of the proposed modified hierarchical method with other standard compression methods in terms of compressed bit rates (bpp) on classical images

Classical images	JPEG 2000	JPEG-XR	MHPCA 1	MHPCA 2
Lena	13.5848	14.0942	18.8383	19.4293
Mandrill	14.8000	15.3245	5.6096	6.0537
Peppers	18.0939	18.2553	3.5644	3.8373
Barbara	11.1612	12.1408	15.7797	16.2833
Average	14.40998	14.9537	10.948	11.4009

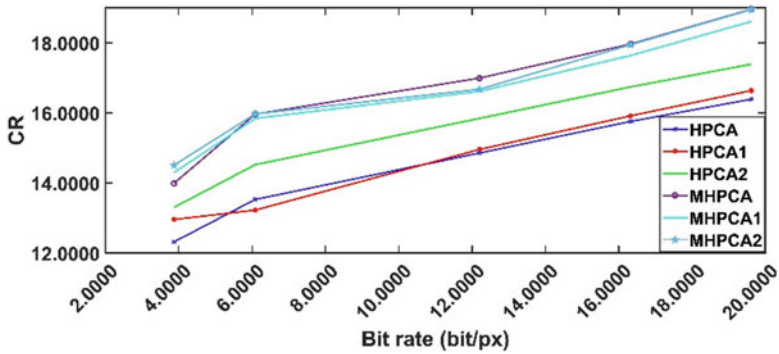


Fig. 8 CR of all proposed hierarchical methods on classical images

Comparison of the proposed HPCA and MHPCA methods with other research paper in term of PSNR is represented in Table 10. From this, we can clearly identify that the proposed methods HPCA 2 and MHPCA 2 give superior values compared to the result of the research paper cited in this table. HPCA 2 and MHPCA 2 yield average PSNR values of **41.70 dB** and **40.71 dB**, respectively, which are high compared to HPCA and MHPCA methods, as described in Ref. [8].

We have compared our proposed methods with other literature on classical images as shown in Table 10. We are not able to compare our proposed methods with other literature on medical images as there is no such literature available, particularly for the dataset we worked on. Hence, we tried to compare our proposed methods only with different approaches which could be seen in the abovementioned tables namely Tables 5, 6, and 7.

5 Conclusion

Modified hierarchical prediction and context adaptive coding scheme is planned along with the hierarchical prediction and context adaptive coding in this work. We have presented results for the modified hierarchical prediction algorithms as well as hierarchical prediction algorithms. The combination of different methods evolved under hierarchical prediction gives the flexibility of understanding our method and acquiring the possibility of obtaining better compression proportion for different images. From all proposed methods, MHPCA method 2 is able to provide a better compression ratio and a better peak signal to noise ratio upon preserving the sharpness of the image. MHPCA method 2 also provides lower bits per pixel. MHPCA 2 performs well on lossless medical image compression. We can clearly identify that the proposed methods HPCA 2 and MHPCA 2 give superior values compared to the result of the research paper cited in Table 10. HPCA 2 and MHPCA 2 yield average PSNR values of 41.70 and 40.71, respectively, which are high compared to HPCA and MHPCA methods, as described in Ref. [8].

Table 10 Comparison of the proposed hierarchical and modified hierarchical methods with other research paper in terms of the peak signal to noise ratio (PSNR) (dB) on classical images

Classical Images	Suresh et al. [8] HPCA	Proposed H.P.C.A Method 1	Proposed H.P.C.A Method 2	Suresh et al. [8] MHPCA	Proposed M.H.P.C.A Method 1	Proposed M.H.P.C.A Method 2
Lena	34.7	37.60	39.79	34.80	36.66	39.05
Peppers	32	32.68	33.95	33	32.68	33.95
Barbara	28	37.96	50.21	28.10	36.98	48.18
Vegetables	33	40.68	42.87	33.39	39.12	41.67
Average	31.92	37.23	41.70	32.32	36.36	40.71

Further, we shall be implementing a hierarchical quadtree structure with variable medical image sequence data to have much higher compression ratio and PSNR value with a low mean square error value.

Appendices

Appendix A: List of Abbreviations

H.P.C.A.	Hierarchical Prediction and Context Adaptive Coding
M.H.P.C.A	Modified Hierarchical Prediction and Context Adaptive Coding
BPP	Bits Per Pixel
CR	Compression Ratio
MSE	Mean Square Error
PSNR	Peak Signal to Noise ratio
PX	Pixel

Appendix B: Formula of Performance Parameters

1. A bits per pixel value of an image is the ratio between data size of image in bits and number of pixels.

$$\text{Bits per pixel (BPP)} = \frac{\text{Datasize of Image (in bits)}}{\text{No. of Pixels}}$$

2. Image compression proportion is demarcated as the ratio between the compressed image and the actual Image.

$$\text{Compression Ratio (\%)} \text{ CR} = 100 - \left[\frac{\text{Compressed image}}{\text{Original image}} \times 100 \right]$$

3. Peak signal to noise proportion is used for the quality measurement between the original and a compressed image.

$$\text{Peak Signal to Noise Ratio : PSNR} = 10 \log_{10} \frac{R^2}{\text{MSE}}$$

where R is the maximum fluctuation in the input image data type. For example, if the input image has a double-precision floating-point data type, then R is 1. If it has an 8-bit unsigned integer data type, R would be 255.

In this equation $L(m,n)$ and $K(m,n)$ are image sizes of original and compressed images, respectively. M and N are the number of rows and columns of an input image, respectively.

4. Thresholding value (T_1 , T_2)

Thresholding is the easiest way of image segmentation. Colored pictures can be applicable with this kind of thresholding values. This style scheme elects distinct threshold meant for the individual of the RGB constituents of pictures and mingles them by means of the AND process. It signifies way toward the camera mechanism and the manner data are stockpiled in the processor, but then the problem faced is that the camera does not match to the way that people identify color [18].

References

1. Seyun Kim and Nam Ik Cho, "Hierarchical prediction and context adaptive coding for lossless color image compression", *IEEE Transactions on image processing*, **2014**, 23(1), 445449.
2. Jinsheng Xiao, Wenhao Li, Guoxiong Liu, Shih-Lung Shaw, and Yongqin Zhang, "Hierarchical tone mapping based on image colour appearance model", *IET Computer Vision*, **2014**, 8(4), 358–364.
3. Seyun Kim and Nam Ik Cho, "Lossless compression of color filter array images by hierarchical prediction and context modeling", *IEEE transactions on circuits and systems for video technology*, **2014**, 24(6), 1040–1046
4. Soo-Chang Pei and Jian-Jiun Ding, "Improved reversible integer-to-integer color transforms", *Image Processing (ICIP), 16th IEEE International Conference on* (pp. 473–476), IEEE, 2009.
5. Shaik Mahaboob Basha and B. C. Jinaga, "An Optimum Novel Technique Based on Golomb-Rice Coding for Lossless Image Compression of Digital Images", *International Journal of Signal Processing, Image Processing and Pattern Recognition*, **2013**, 6(5), 291304.
6. Li Li, Zhu Li, Bin Li, Dong Liu and Houqiang Li, "Pseudo-Sequence-Based 2-D Hierarchical Coding Structure for Light-Field Image Compression", *IEEE Journal of Selected Topics in Signal Processing*, **2017**, 11(7), 1107–1119.
7. Ning Zhang and Xiaolin Wu, "Lossless compression of color mosaic images", *IEEE Transactions on Image Processing*, **2006**, 15(6), 1379–1388.
8. P. Suresh Babu and S. Sathappan, "Efficient lossless image compression using modified hierarchical prediction and context adaptive coding", *Indian Journal of Science and Technology*. **2015**, 8(34), 1–6.
9. Surabhi N and Sreeleja N Unnithan, "Image Compression Techniques: A Review", *International Journal of Engineering Development and Research*, 2017.
10. S. Sathappan and P. Suresh Babu, "Block based prediction with Modified Hierarchical Prediction image coding scheme for Lossless color image compression", *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. IEEE, 2016.
11. Du Liu and Markus Flierl, "Temporal signal basis for hierarchical block motion in image sequences", *IEEE Signal Processing Letters*, **2018**, 25(1), 10–14.
12. A.J. Hussain, Ali Al-Fayadh and Naem Radi, "Image compression techniques: A survey in lossless and lossy algorithms", *Neurocomputing*, **2018**, 300, 44–69.

13. Ruchi Gupta, Mukesh Kumar, Rohit Bathla, “Data Compression-Lossless and Lossy Techniques”, *International Journal of Application or Innovation in Engineering and Management*, 2016.
14. Manjari Singh, Sushil Kumar, Siddharth Singh Chouhan and Manish Shrivastava, “Various Image Compression Techniques: Lossy and Lossless”, *International Journal of Computer Applications*, **2016**, *Image*, 142(6).
15. Lei Guo, Dong Liu, Li Li and Feng Wu, “Hierarchical quadtree-based flexible block ordering in HEVC intra coding”, *Visual Communications and Image Processing (VCIP)*. IEEE, 2016.

Websites

16. “Coding Efficiency” accessed on 27th April 2019, https://www.phusewiki.org/wiki/index.php?title=Coding_Efficiency
17. “Temporal Scalability” accessed on 27th April 2019, <https://www.hhi.fraunhofer.de/en/departments/vca/research-groups/image-videoencoding/research-topics/svc-extension-of-h264avc/temporal-scalability-in-h264avc.html>
18. “Thresholding in Image Processing” accessed on 27th April 2019, [https://en.wikipedia.org/wiki/Thresholding_\(image_processing\)](https://en.wikipedia.org/wiki/Thresholding_(image_processing))

LBP-Based CAD System Designs for Breast Tumor Characterization



Kriti, Jitendra Virmani, and Ravinder Agarwal

1 Introduction

The prominently found cancer among women is breast cancer [25]. The prospects of recovery and survival can be increased if breast cancer is caught at the preliminary stage. For periodic screening of breast cancer, mammography is the preferred choice for women over the age of 40 years [16, 17]. However, ultrasonography has nowadays found prevalent use in breast cancer detection due to (a) ease of availability and use, (b) low cost, (c) ease in detecting obscured tumors (especially, in the case of young women having dense breast tissue), and (d) lack of ionizing radiation (ultrasound is specifically useful for pregnant women as radiations may adversely affect the fetus). Despite these advantages, the quality of the ultrasound images is degraded due to speckle noise and artifacts, which make it laborious for radiologists to make a clear and concise diagnosis. The sample images taken from the standard benchmark database are used in the present work, indicating that the sonographic characteristics exhibited by different types of breast tumors are shown in Fig. 1.

High variability exists among the sonographic appearances of different classes of breast abnormalities, because of which the differential diagnosis of the tumor types

Kriti (✉)
DIT University, Dehradun, Uttarakhand, India
e-mail: kriti@dituniversity.edu.in

J. Virmani
CSIR-Central Scientific Instruments Organisation (CSIR-CSIO), Chandigarh, India
e-mail: jitendravrmani@csio.res.in

R. Agarwal
Thapar Institute of Engineering and Technology, Patiala, Punjab, India
e-mail: ravinder_eeed@thapar.edu

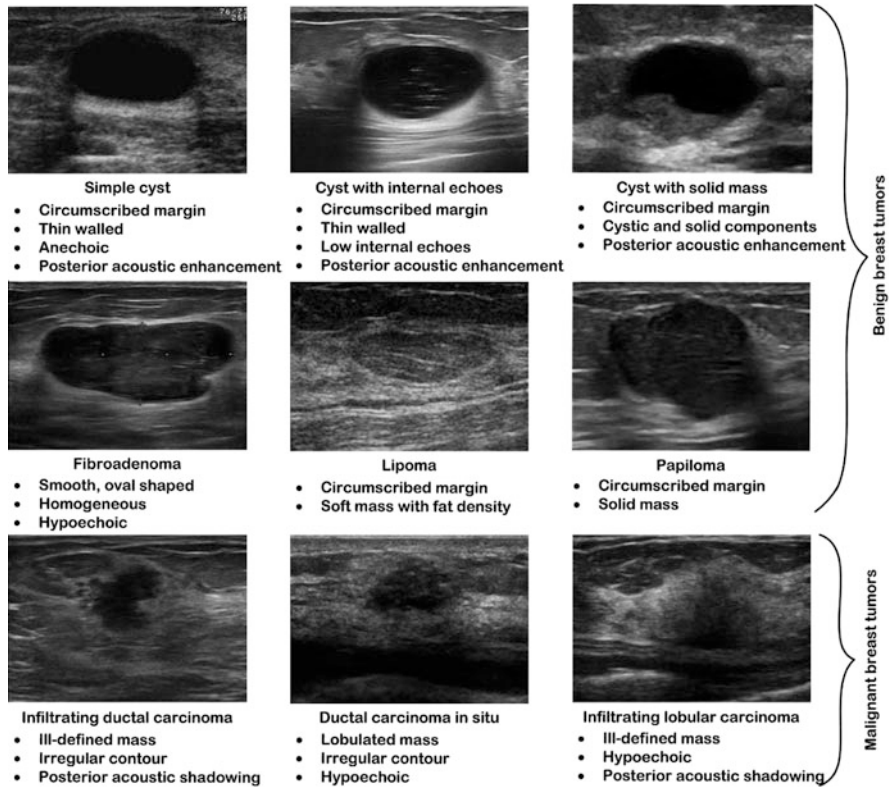


Fig. 1 The sample breast ultrasound images indicating the sonographic characteristics exhibited by different types of tumors

is sometimes difficult even for an experienced radiologist. Accordingly, there has been an increased amount of interest among the researchers for designing efficient computer-aided diagnostic (CAD) systems for breast tumor characterization [1, 4, 6, 12, 15, 27, 28, 30, 35, 36, 39, 41, 47, 50, 57, 58].

Zakeri et al., in 2012, reported an accuracy of 95.0% on a set of 80 images using correlation-based texture features along with morphological features and support vector machine (SVM) classifier. These features have been computed using images segmented by applying deformable parts model-based method. Amin et al., in 2015, reported an accuracy of 99.1% using the SVM classifier and a feature set computed from 112 breast ultrasound images. The feature set contained statistical texture features combined with morphological features. The images were first transformed into a neutrosophic domain, and from the resultant images, statistical texture features have been computed. The optimal features have been selected on the basis of a Chi-square test. Daoud et al., in 2016, implemented a decision fusion-based approach for breast tumors characterization by utilizing the SVM reporting an accuracy of 98.2% using 110 images. For the computation of texture features,

multiple non-overlapping regions of interest (ROIs) have been taken from inside the tumors. The gray level co-occurrence matrix (GLCM)-based texture features have been computed from each ROI, while the morphological features have been computed using the whole tumor. Each computed feature set has been subjected to a two-stage feature selection strategy employing backward selection and minimal-redundancy-maximal relevance (mRMR). Based on texture features, the class membership of each ROI has been determined, and majority voting has been used to obtain the final tumor class. The study also determined the tumor class based on the computed morphological features, and the final tumor class has been obtained by the decision fusion. The study has also experimented with the conventional method of combining the computed texture and morphological sets and reported an accuracy of 90.9%. Takemura et al., in 2009, reported the highest classification accuracy of 100% using an AdaBoost classifier. The optimal feature set has been obtained by the application of sequential forward search (SFS) on a feature set formed by combining statistical texture features and morphological features. Piliouras et al., in 2004, used the SVM classifier and a feature set composed of statistical texture features and morphological features and reported an accuracy of 98.7%. A wrapper-based method of feature selection has been utilized to find out an optimal feature subset for breast tumor characterization. Menon et al., in 2015, used a combination of median, high boost, and sobel filters to pre-process 78 breast images. From the pre-processed images, tumor regions have been segmented using the deformable parts model, and a feature set has been formed by combining statistical texture features and morphological features. The texture features have been computed using the first order statistics (FOS), GLCM and covariance-based methods. An accuracy of 95.7% has been reported using the SVM and an optimal number of principal components (PCs) obtained by applying principal component analysis (PCA) to the computed feature set. Uzunhisarcikli and Goreke, in 2018, reported an accuracy of 99.3% for classifying 153 breast ultrasound images using a type-2 adaptive neuro-fuzzy inference system (ANFIS) and an optimal feature set containing GLCM and morphological features, computed using ROIs extracted using the lesion contour marked by the radiologist and pre-processed using the Gaussian filter and contrast limited adaptive histogram equalization (CLAHE). Nemat et al., in 2018, reported an accuracy of 97.1% using the stepwise logistic regression (SLR) classifier using 104 breast ultrasound images pre-processed by CLAHE and anisotropic diffusion (AD) filter and segmented using watershed transform. From the segmented tumor images, a combined feature set has been formed comprising of texture features computed using Gabor filters and morphological features.

The study carried out by Cheng et al., in 2016, reported a classification accuracy of 82.4% using 520 full images and stacked denoising autoencoders (SDAEs) for breast tumor characterization. Lee et al., in 2018, used 250 full-size ultrasound images pre-processed by CLAHE for breast tumor characterization using the SDAE reporting an accuracy of 83.0%. In a study by Zhang et al., in 2020, 160 full images and 295 tumor images segmented using watershed transform were used for breast tumor characterization using the stacked convolutional autoencoder (SCAE) followed by the softmax classifier, and images were pre-processed using a distance-

transformation coupled Gaussian filter (DTGF). Accuracies of 92.0% and 83.9%, respectively, were reported.

For the CAD system based on original as well as pre-processed ultrasound images, a machine learning-based study has been previously reported by the authors, wherein the original images have been used to compute texture features, while pre-processed images have been used to compute morphological features. Exhaustive experimentation was carried out for assessing the effect of different despeckling filters on the performance of CAD systems for breast tumor characterization and it was validated that texture information is effectively quantified using original images, while efficient morphological features are computed using images pre-processed by the DPAD filter [27].

For computing the features from original or despeckled images, most of the studies have made use of statistical texture features along with morphological features. It should be noted that the performance of local binary pattern (LBP) texture features combined with the morphological features has not been tested for breast tumor characterization using ultrasound images which have otherwise yielded good results in the case of other medical images [13, 24, 38]. However, it has also been noted that the potential of LBP features has been explored by very few studies in the case of breast ultrasound images for quantification of texture [1, 6, 35]. As seen from the authors' previous study [27], it has been experimentally verified that optimal results are obtained when the original images are used for quantifying texture features, and the morphological information of the tumors has been quantified by using images pre-processed by the DPAD filter. Accordingly, four different CAD system designs have been compared in the present work for breast tumor characterization based on LBP texture features computed using original images and morphological features computed using the ultrasound images pre-processed by the DPAD filter.

2 Methodology

The experimental workflow adopted for the design of an efficient LBP-based CAD system for breast tumor characterization is presented in Fig. 2.

2.1 Dataset Description and Bifurcation

The description of the dataset and its bifurcation is presented in Fig. 3.

The images have been taken from an online repository of ultrasound images available at (ultrasoundcases.info). The protocol for the selection of images has been kept same as in the previous study carried out by the authors [27].

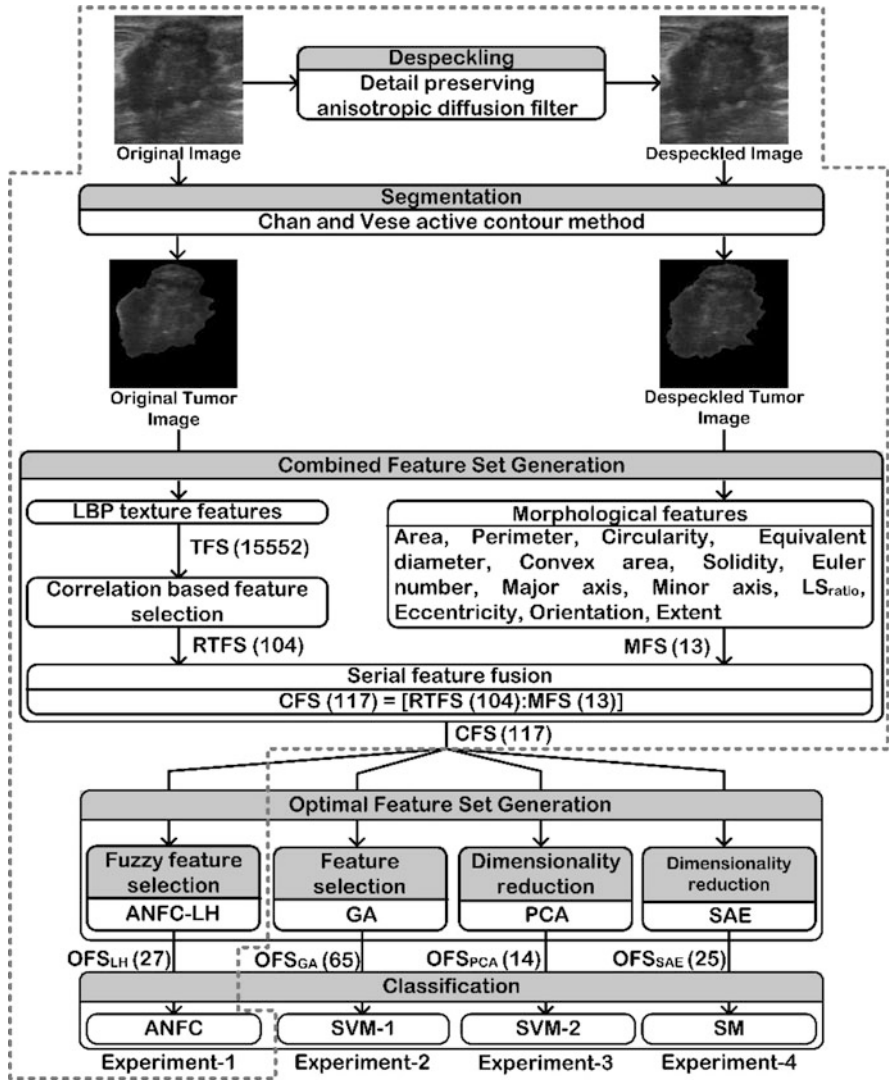


Fig. 2 The experimental workflow adopted for the design of an efficient LBP-based CAD system for breast tumor characterization. (**Note:** *LBP* Local binary patterns, *TFS* Texture feature set, *RTFS* Reduced texture feature set, *MFS* Morphological feature set, *CFS* Combined feature set, *OFS* Optimal feature set, *LH* Linguistic hedges, *GA* Genetic algorithm, *PCA* Principal component analysis, *SAE* Stacked autoencoder, *ANFC* Adaptive neuro-fuzzy classifier, *SVM* Support vector machine, *SM* Softmax)

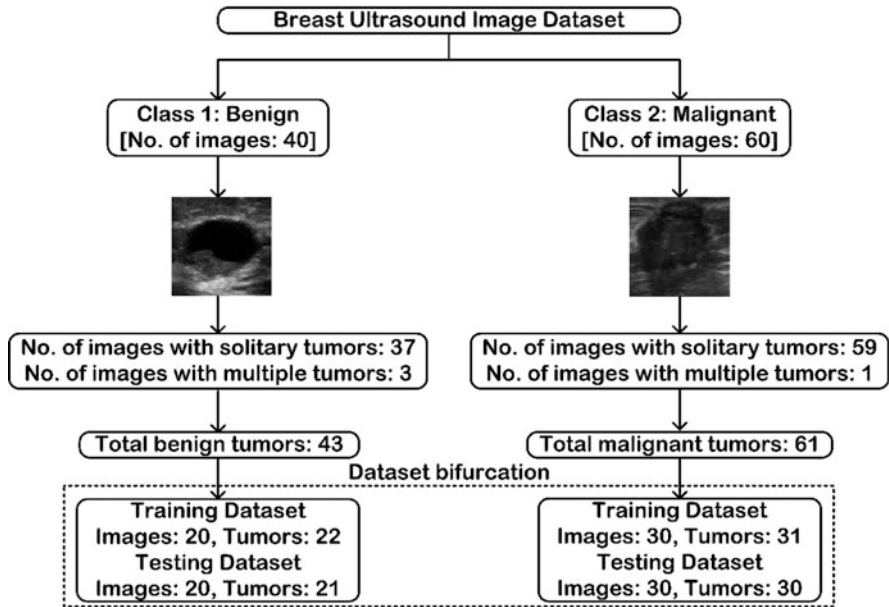


Fig. 3 Dataset description and bifurcation

2.2 Despeckling

The sonographic features of the breast tumors as visible on ultrasound are considered to be significant for clear discrimination between different tumor types. However, the presence of speckle noise and low image contrast severely hampers the ultrasound image quality due to which the diagnosis becomes difficult. The visual interpretation of the radiologist is negatively affected due to low image contrast and the presence of speckle noise as it masks the diagnostically important detailed structures in the image. For filtering out the speckle noise, a controlled despeckle filtering is desired such that homogeneous areas are smoothed and the edge/structure information is preserved, thereby enhancing the image quality, resulting in improved interpretation and increased accuracy of computer-assisted segmentation and classification algorithms [26, 27, 31, 59].

In the authors’ previous study [26], it was validated that the DPAD filter yielded the best performance for efficient edge and structure preservation in images. Thus, in the present work, the DPAD filter [2] has been employed for the pre-processing of breast ultrasound images. For further details on the performance of despeckle filtering algorithms and parameters used for each filter, the readers are directed to [26].

2.3 Segmentation

Segmentation is used to separate out a region of interest from an image with the help of a computer-assisted algorithm. Segmentation techniques have widely been employed in the case of medical images to separate out the tumor region from the background [26]. Out of a myriad of computer-assisted segmentation algorithms, an active contour method has found prevalent use for separating out the tumor region from medical images [14, 26, 32, 34]. In the authors' previous work [26], the performance of the Chan and Vese method of segmentation [10] for extracting the tumors from breast ultrasound images has been assessed. For further details on the effect of despeckle filtering algorithms on segmentation of breast tumors, the readers are directed to [26].

Accordingly, the present work uses the Chan and Vese method to extract out the breast tumors from original images as well as images despeckled by the DPAD filter, with the number of iterations and initial rectangular bounding box (mask) given as the input.

The sample images indicating the tumor contour marked by the participating radiologist and the tumor contour obtained after applying the segmentation algorithm for original and despeckled images are shown in Fig. 4.

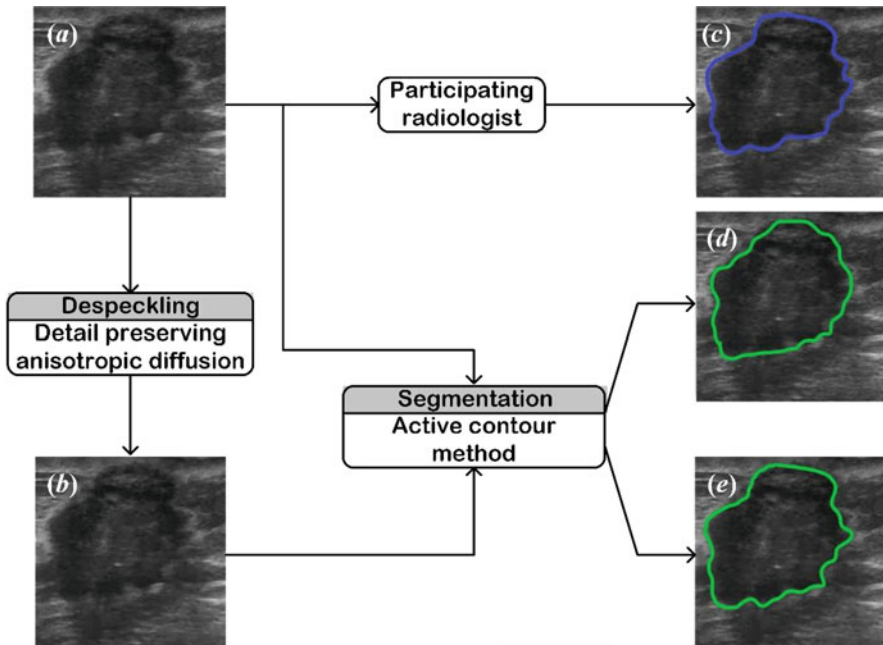


Fig. 4 Application of the segmentation algorithm to ultrasound images. (a) Original malignant image, (b) despeckled malignant image, (c) image indicating a tumor contour marked by the radiologist, (d) original malignant image with the tumor contour obtained using the active contour method, (e) despeckled malignant image with the tumor boundary obtained by the active contour

2.4 Combined Feature Set Generation

Feature extraction converts the perceptible information of an image into mathematical descriptors based on intensity distribution of the image or on the shape of the tumor or on the color features. According to the type of tissue being observed, different sets of features are significantly studied for diagnosis. In the previous study by authors [27], the performance of texture and morphological features extracted using original and pre-processed breast ultrasound images was exhaustively tested, and it was validated that the texture features are aptly quantified through original images, while for extracting morphological features, images pre-processed by the DPAD filter yielded an optimal performance. Accordingly in the present work, original images have been used to extract LBP-based texture features forming a texture feature set (TFS), and the morphological features have been computed using images despeckled by the DPAD filter form a morphological feature set (MFS). The process of combined feature set generation is presented in Fig. 5.

There are a large number of features in the TFS obtained by using LBP which may be redundant, thus to remove these redundant features, correlation-based feature selection (CrFS) has been employed forming a reduced texture feature set (RTFS). Finally, RTFS and MFS have been fused serially, thus forming a combined feature set (CFS).

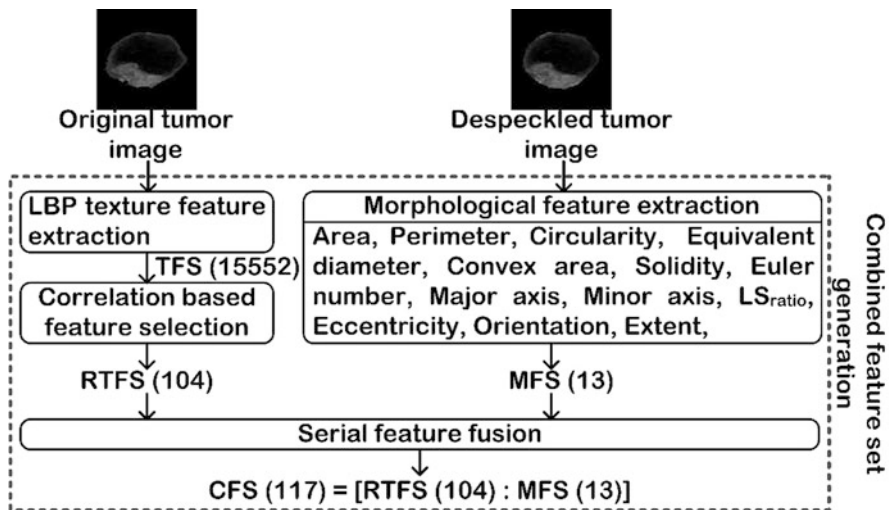


Fig. 5 Combined feature set generation. (Note: LBP Local binary pattern, TFS Texture feature set, RTFS Reduced texture feature set, MFS Morphological feature set, CFS Combined feature set)

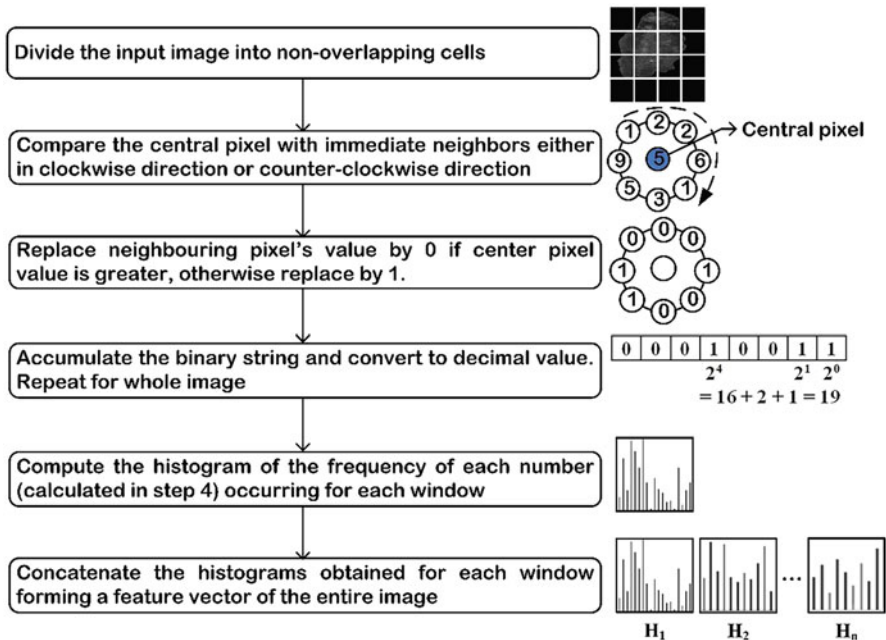


Fig. 6 Steps followed for the implementation of the LBP algorithm. (Note: H_1 Histogram of first window, H_n Histogram of n th window)

2.4.1 LBP-Based Texture Feature Extraction

The texture features are used to quantify gray-level distribution statistics in an image. In the present work, a uniform LBP has been used to compute the texture information from the original ultrasound images. Ojala et al. [40], introduced the LBP for image texture quantification of local neighborhood. In the LBP algorithm, the pixels of an image are labeled using a threshold value for each pixel neighborhood, and the final result is considered as a binary number. The steps followed for computing the LBP feature set are shown in Fig. 6.

The length of the final LBP feature vector is given as:

$$\text{NumCells} = \text{prod} \left(\text{floor} \left(\frac{\text{image size}}{C} \right) \right) = \text{prod} \left(\frac{512}{64} \right) = \text{prod}(8) = 64 \quad (1)$$

$$B:\text{No. of histogram bins} = N \times (N - 1) + 3 = 16(16 - 1) + 3 = 243 \quad (2)$$

$$\text{LBP}(l) = \text{NumCells} \times B = 243 \times 64 = 15552 \quad (3)$$

where C is the cell size and N is the number of neighbors within a radius R . From the segmented original images, a texture feature set (TFS) is formed containing 15,552 LBP features computed at $R = 1$, $N = 16$, and $C = 64$.

2.4.2 Correlation-Based Feature Selection

The redundant features in the computed feature set sometimes degenerate the classification performance of the CAD system. Therefore, it becomes essential that relevant features are selected. Due to feature selection, the training time, overfitting, and complexity of the classification model get reduced.

In CrFS, the best feature subset is chosen on the basis of the correlation coefficient [22, 37, 46]. The optimal feature subset is selected on the basis of the computed score that is used as a threshold and is given as:

$$S = \frac{n \times \overline{r_{cf}}}{\sqrt{n + n(n-1)\overline{r_{ff}}}} \quad (4)$$

where n is the number of features, $\overline{r_{ff}}$ is the mean inter-correlation between features, and $\overline{r_{cf}}$ is the mean correlation between the feature and the class.

In the present work, the TFS having a length of 15,552 is subjected to CrFS yielding RTFS having length 104.

2.4.3 Morphological Feature Extraction

The shape and margin characteristics of a tumor can be efficiently represented by morphological features as they are considered to be clinically significant in discriminating the benign and malignant tumors. The benign tumors have a regular shape (round or oval), while malignant tumors are irregularly shaped. The computed morphological features are: area, perimeter, circularity, equivalent diameter, convex area, solidity, Euler number, length of major axis and minor axis, LS ratio, orientation, eccentricity, and extent of the tumor region [27].

The sample images representing the tumor boundary along with the convex hull boundary, bounding rectangle, and ellipse of the tumor are shown in Fig. 7.

The computed morphological features have been aggregated to form a morphological feature set (MFS).

2.4.4 Serial Feature Fusion

Using feature fusion techniques, a single feature set is obtained by aggregating multiple feature sets. In serial feature fusion, a combined feature set is obtained by simply concatenating different features one after the other (union operation) [3, 53]. In the present work, the CFS is formed by serially fusing the reduced texture

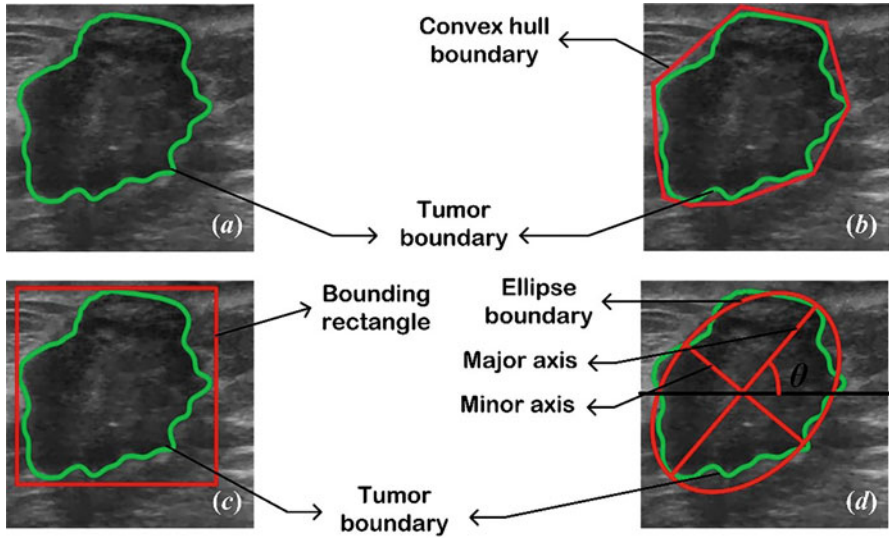


Fig. 7 Sample malignant breast ultrasound image, indicating (a) tumor boundary, (b) tumor boundary and convex hull boundary, (c) tumor boundary and bounding rectangle, (d) tumor boundary and ellipse of the tumor

feature set and morphological feature, i.e., RTFS (104) and MFS (13) are combined serially as: CFS (117) = [RTFS: MFS] (104 + 13).

2.5 Optimal Feature Set Generation

The obtained CFS has been subjected to two feature selection and two feature space dimensionality reduction methods to obtain optimal feature sets. For selecting optimal features, a fuzzy feature selection technique based on an adaptive neuro-fuzzy classifier using the linguistic hedge (ANFC-LH) algorithm, and a feature selection technique based on the genetic algorithm (GA-SVM) has been employed. For finding the optimal attributes, feature space dimensionality reduction methods based on the PCA-SVM algorithm and the stacked autoencoder (SAE) with the softmax classifier (SAE-SM) have been used. The description of the optimal feature sets is given in Table 1.

2.5.1 ANFC-LH Algorithm-Based Feature Selection

In the ANFC-LH algorithm, linguistic hedges (LHs) have been used to bring out the importance of fuzzy rules. The flexibility of fuzzy sets is improved by tuning the

Table 1 Description of optimal feature sets

Original feature set (<i>l</i>)	Feature selection/Feature space dimensionality reduction method	Optimal feature set (<i>l</i>)
CFS (117)	ANFC-LH algorithm	OFS _{LH} (27)
	GA-SVM algorithm	OFS _{GA} (65)
	PCA-SVM algorithm	OFS _{PCA} (14)
	SAE with softmax classifier	OFS _{SAE} (25)

Note: *l* Length of the feature set, *CFS* Combined feature set, *ANFC-LH* Adaptive neuro-fuzzy classifier using linguistic hedges, *OFS_{LH}* Optimal feature set obtained by using the ANFC-LH algorithm, *GA* Genetic algorithm, *SVM* Support vector machine, *OFS_{GA}* Optimal feature set obtained by using the GA-SVM algorithm, *PCA* Principal component analysis, *OFS_{PCA}* Optimal feature set obtained by using the PCA-SVM algorithm, *SAE* Sacked autoencoder, *OFS_{SAE}* Optimal feature set obtained by using the SAE-SM

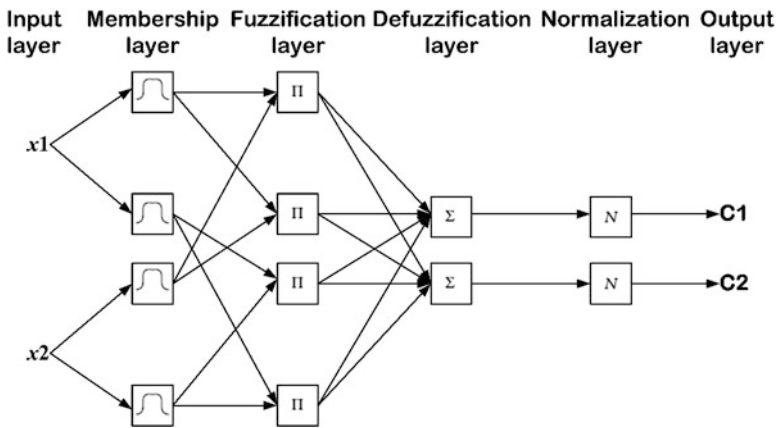


Fig. 8 General layer architecture of ANFC-LH

values of LHs such that the ambiguity of the overlapped classes is removed [8, 9, 54–56]. For further details on the explanation of fuzzy rules and their modification through LHs, refer to [28].

The general architecture of ANFC-LH is presented in Fig. 8 having two inputs in the feature space to be separated into two classes with each input being described by two linguistic variables, thus giving a total of four fuzzy rules.

From the CFS having 117 features, on the basis of linguistic hedge values, a total of 27 optimal features have been selected, thus forming an optimal feature set represented as OFS_{LH}. The relationship between input features and their respective power of the LH value is represented in Fig. 9.

2.5.2 GA-SVM Algorithm-Based Feature Selection

The genetic algorithm is an evolutionary search procedure inspired by the biological evolution model. In GA, the features are represented as binary vectors, and the

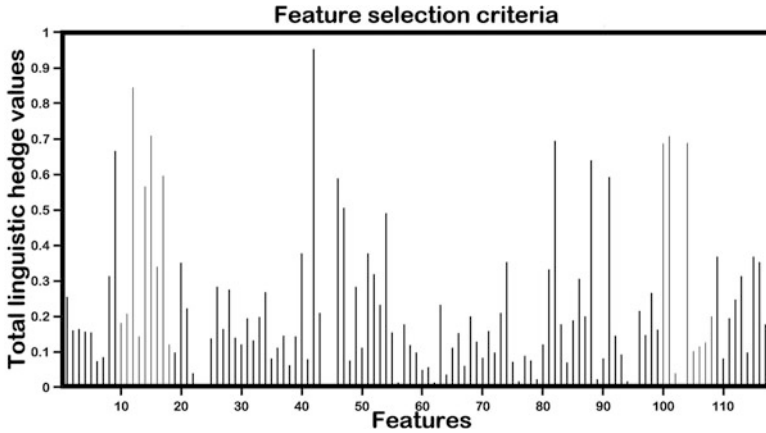


Fig. 9 The relationship between input features and their respective power of the LH value



Fig. 10 General schematic for the implementation of the GA-SVM algorithm. (Note: CFS Combined feature set, OFS_{GA} Optimal feature set obtained by applying the genetic algorithm)

feature search space is considered to be an n -dimensional Boolean space. The GA selects a random set of individuals from the given population and works toward producing the offsprings for the next generation. Three main operators used in GA are selection, crossover, and mutation to create the successive generation using current population on the basis of a fitness function.

This process of fitness-dependent selection is repeated multiple times till an optimal solution is found [20, 45, 52]. In the present work, the initial population is defined as a set of 117-bit binary-coded chromosomes. The classification accuracy obtained by the SVM classifier is then used to evaluate the fitness of a chromosome. The general schematic of the genetic algorithm applied to the CFS to generate an optimal feature set obtained using GA (OFS_{GA}) is shown in Fig. 10.

The different parameters used during the run of GA are number of variables: 117, population size: 200, mutation rate: 0.01, selection function: Roulette, scaling function: Rank, crossover function: single point, and crossover fraction: 0.7. The algorithm is terminated when the maximum iteration count is reached or no improvement is witnessed in the fitness value.

2.5.3 PCA-SVM Algorithm-Based Dimensionality Reduction

The PCA-SVM algorithm is used to find an optimal number of principal components (PCs) to design efficient CAD systems for classification [27, 33, 36, 51]. No significant information is provided by the redundant features present in the CFS, which helps in discriminating the breast tumor types. Therefore, to remove this redundancy, the features in the CFS are converted to optimal attributes using the PCA-SVM algorithm. The optimal number of PCs to be retained has been decided empirically by conducting recurrent experiments and stepping through first few PCs $\in \{2, 3, \dots 15\}$ [26, 51].

2.5.4 SAE-SM Algorithm-Based Dimensionality Reduction

Autoencoders (AEs) come under the class of generative deep models useful for unsupervised learning. The output of the AE is the same as the input. First, the input data are compressed by AEs into a latent space representation, and then the output is reconstructed from this representation. Thus, an AE can be viewed to have two parts: (i) encoder function, (ii) decoder function. The general architecture of an AE is shown in Fig. 11.

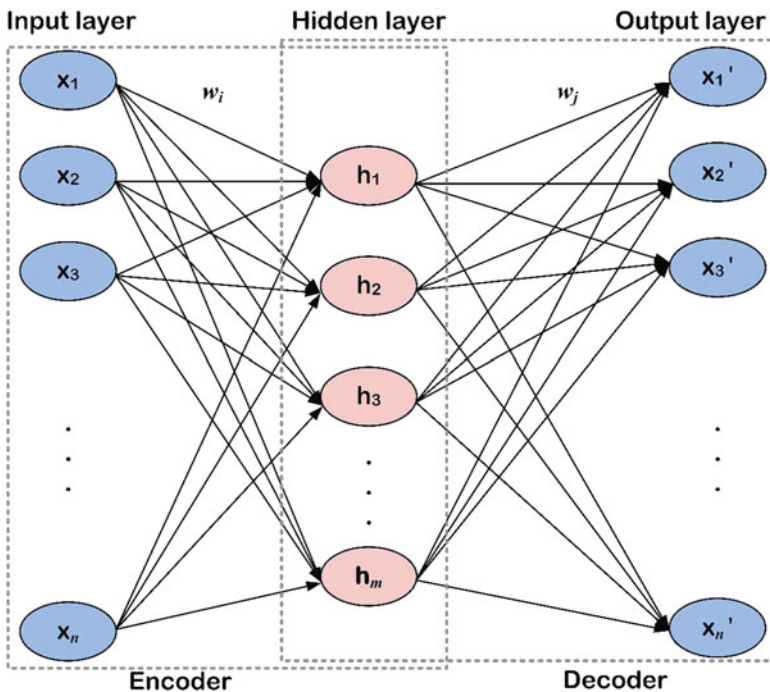


Fig. 11 General architecture of an autoencoder

The encoder maps the input vector x to another vector y using a transfer function h as

$$y = h(w_i x + b_i) \tag{5}$$

where w_i is the weight matrix at the encoder side and b_i is the bias vector.

The function of the decoder is to reconstruct vector y to estimate the original input vector.

$$x' = h'(w_j y + b_j) \tag{6}$$

where w_j is the weight matrix at the decoder side and b_j is the bias vector.

When the encoding layer's output is connected to another encoding layer's input, then the resultant architecture is called a stacked autoencoder (SAE), allowing for several layers of abstraction. The general architecture of an SAE is pictorially shown in Fig. 12.

When multiple hidden layers are used, a greedy layer-wise approach is used for initializing the hidden layers. This process is called pre-training, wherein the input training data train the first hidden layer. The output from the first hidden layer is then used for training the second hidden layer and so on till all the hidden layers of the network have been trained. The hyperparameters of an SAE are the weights and bias of the network, number of hidden layers, and number of hidden units present in each layer. For getting the final network structure used in the present work, a

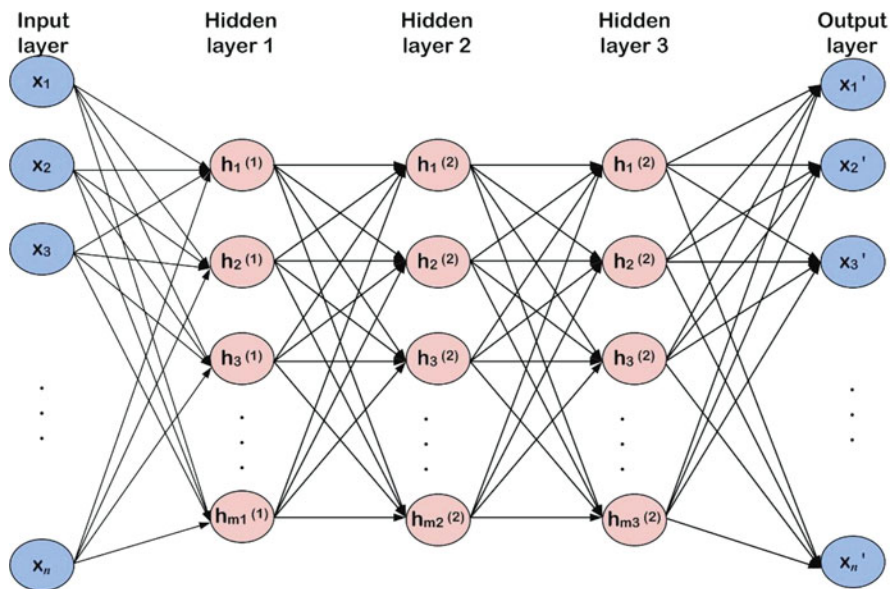


Fig. 12 General architecture of a stacked autoencoder

random search was exhaustively used to select the parameters that produced the best performance [7, 18, 19, 23].

2.6 Classification

Classification is a supervised learning approach, wherein a computer program learns the underlying properties of the input data fed to it and on the basis of the gained knowledge tries to classify the new data instances into discrete classes.

2.6.1 Adaptive Neuro-fuzzy Classifier

A neuro-fuzzy classifier is a combination of the fuzzy inference system and neural networks. In order to deal with imprecise problems, a fuzzy inference system can be used, wherein non-linear functions are approximated using a set of fuzzy IF–THEN rules. However, these systems are unable to adaptively adjust themselves as they cannot learn from their environment. Neural networks have a self-organizing capability and can learn adaptively from their environment. Therefore, the respective advantages of the fuzzy inference system and neural network are integrated in a neuro-fuzzy classifier [21, 29, 43, 44, 48, 50]. The neuro-fuzzy network is made up of multiple nodes interconnected via directional links. The node output depends on the node parameters that can be fine-tuned while training the network in order to minimize the error, making the network adaptive and is thus called an adaptive neuro-fuzzy classifier (ANFC). The layer architecture of the ANFC used is presented in Fig. 13.

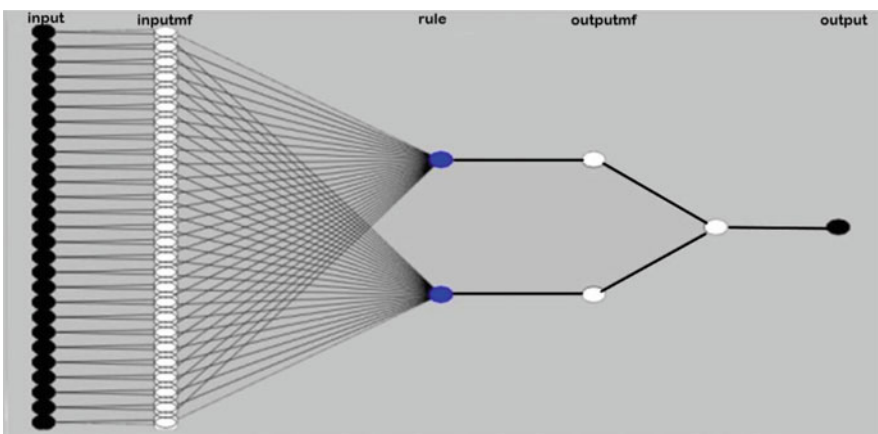


Fig. 13 The layer architecture of ANFC used in the present work

During the training of the ANFC, the center and width values of the Gaussian membership function, power value of LHs, and connection weights between fuzzification and defuzzification layers need to be optimized for finding the optimum fuzzy region. In order to form the fuzzy IF–THEN rules, the initial parameters are obtained using *k*-means clustering. For the parameter optimization, the scaled conjugate gradient (SCG) algorithm has been used as it has shown to produce the lowest error rate for optimization [5].

2.6.2 Support Vector Machine Classifier

This classifier has been a popular tool for a wide variety of machine learning tasks, especially involving the characterization of medical images [20, 27, 36, 51]. In the SVM, a kernel-based approach has been used to separate the instances into disjoint classes using the hyper-plane that maximizes the margin between two classes. Non-linear data from the input feature space have been mapped into a linear higher dimensionality feature space using the Gaussian radial basis function (GRBF) kernel. For the implementation of the SVM algorithm, the LibSVM library has been used [11]. The classification steps along with the optimal values of (C , γ) are shown in Fig. 14. For a detailed study on the working of the SVM classifier, readers are directed to [27, 51].

2.6.3 Softmax Classifier

A softmax classifier is stacked on top of a trained SAE network to represent the class labels of the input data. The number of units in the softmax layer is the same as the classes of the classification problem. The final network thus formed consists of a stack of all the hidden layers and the softmax layer as shown in Fig. 15.

It can be noted that the SAE with three hidden layers stacked with a softmax classifier has been used for breast tumor characterization. The input layer consists of 117 units representing the size of the input feature set. The three hidden layers have 100, 50, and 25 nodes, respectively, and the softmax layer consists of two units corresponding to the benign and malignant classes in the present work. The hidden layers and the softmax layer were trained for 2000 epochs each. This final network is then trained as a whole in a supervised manner to achieve the final classification performance. To fine-tune the network and update the parameters, error back-propagation and scaled conjugate gradient have been used.

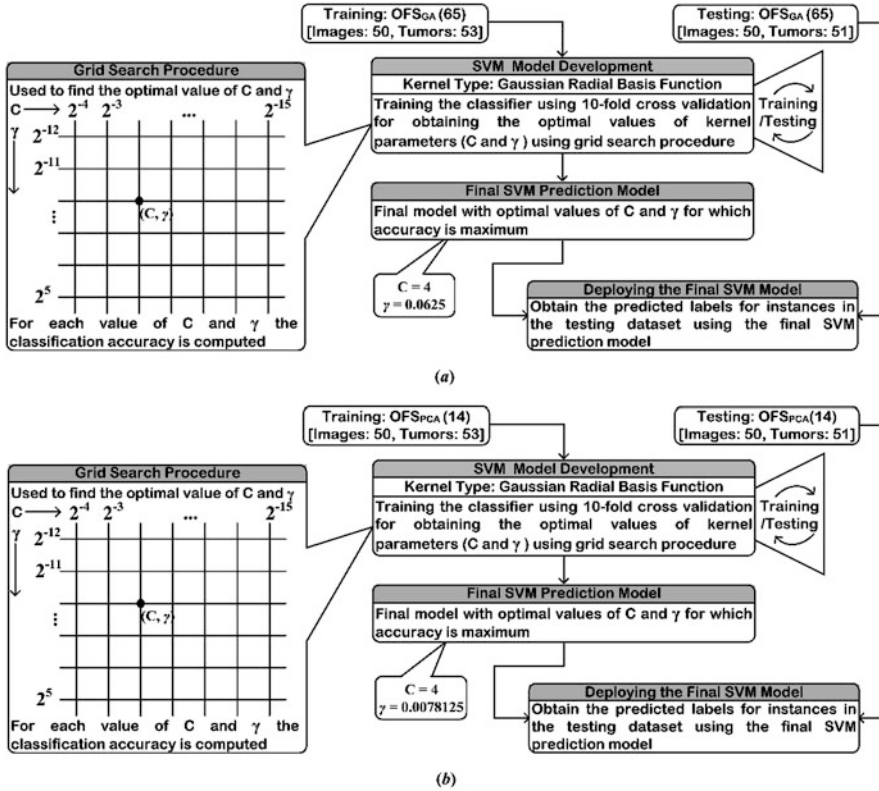


Fig. 14 Steps followed in the SVM algorithm (a) Using optimal feature set generated using GA, (b) using the optimal feature set generated using PCA. (Note: SVM Support vector machine, OFS Optimal feature set, GA Genetic algorithm, PCA Principal component analysis)

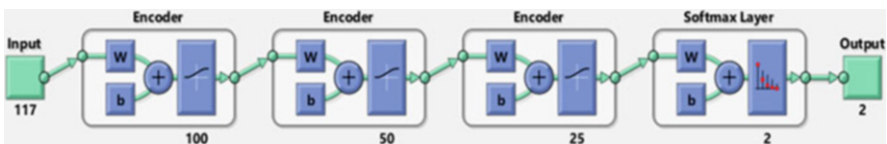


Fig. 15 Architecture of the final network made up of the SAE with the softmax classifier

3 Results and Discussion

3.1 Experiments and Classification Results

The experiments conducted in the present work for breast tumor characterization are described in Table 2.

Table 2 Description of experiments

Experiment No.	CAD system design
Experiment 1	LBP-based CAD system using ANFC-LH
Experiment 2	LBP-based CAD system using GA-SVM
Experiment 3	LBP-based CAD system using PCA-SVM
Experiment 4	LBP-based CAD system using SAE-SM

Note: *LBP* Local binary pattern, *CAD* Computer aided diagnostic, *ANFC-LH* Adaptive neuro-fuzzy classifier using linguistic hedges, *GA* Genetic algorithm, *SVM* Support vector machine, *PCA* Principal component analysis, *SAE* Sacked autoencoder, *SM* Softmax

Table 3 Classification results obtained for the LBP-based CAD system using the ANFC-LH algorithm

Classifier (FS: <i>l</i>)	CM			Acc. (%)	ICA _B (%)	ICA _M (%)
		B	M			
ANFC (OFS _{LH} : 27)	B	19	2	96.0	90.4	100
	M	0	30			

Note: *FS* Feature set, *l* No. of optimal features, *CM* Confusion matrix, *Acc* Accuracy, *ICA* Individual class accuracy, *B* Benign class, *M* Malignant class, *ANFC* Adaptive neuro-fuzzy classifier, *OFS_{LH}* Optimal feature set obtained by using the ANFC-LH algorithm

3.1.1 Experiment 1: LBP-Based CAD System Using the ANFC-LH Algorithm

In this experiment, the CFS (117) containing both texture and morphological features is subjected to the ANFC-LH algorithm generating OFS_{LH} (27) that is further fed to the ANFC for breast tumor characterization. The classification results obtained for the LBP-based CAD system using the ANFC-LH algorithm are shown in Table 3.

The results in Table 3 demonstrate that the 27 optimal features yielded an accuracy of 96.0% with an individual class accuracy (ICA) value of 90.4% for the benign class and 100% for the malignant class. It is worth noting that only two benign instances out of 51 testing instances have been wrongly classified, while all the malignant cases have been classified correctly.

3.1.2 Experiment 2: LBP-Based CAD System Using the GA-SVM Algorithm

In this experiment, the CFS (117) containing both texture and morphological features is subjected to a feature selection technique, GA-SVM algorithm, generating OFS_{GA} (65) that is fed to the SVM classifier for the characterization of breast tumor types. The classification results obtained for the LBP-based CAD system using the GA-SVM algorithm are shown in Table 4.

Table 4 Classification results obtained for the LBP-based CAD system using the GA-SVM algorithm

Classifier (FS: l)	CM			Acc. (%)	ICA _B (%)	ICA _M (%)
		B	M			
SVM (OFS _{GA} : 65)	B	19	2	92.2	90.4	93.3
	M	2	28			

Note: FS Feature set, l No. of optimal features, CM Confusion matrix, Acc Accuracy, ICA Individual class accuracy, B Benign class, M Malignant class, SVM Support vector machine, OFS_{GA} Optimal feature set obtained by using the GA-SVM algorithm

Table 5 Classification results obtained for the LBP-based CAD system using the PCA-SVM algorithm

Classifier (FS: l)	CM			Acc. (%)	ICA _B (%)	ICA _M (%)
		B	M			
SVM (OFS _{PCA} : 14)	B	19	2	94.1	90.4	96.6
	M	1	29			

Note: FS Feature set, l No. of optimal features, CM Confusion matrix, Acc Accuracy, ICA Individual class accuracy, B Benign class, M Malignant class, SVM Support vector machine, OFS_{PCA} Optimal feature set obtained by using the PCA-SVM algorithm

The results in Table 4 demonstrate that the 65 optimal features yielded an accuracy of 92.2% with an ICA value of 90.4% for the benign class and 93.3% for the malignant class. A total of four instances out of 51 testing instances have been wrongly classified.

3.1.3 Experiment 3: LBP-Based CAD System Using the PCA-SVM Algorithm

In this experiment, the CFS (117) containing both texture and morphological features is subjected to a feature space dimensionality reduction technique by using the PCA, generating an optimal feature set OFS_{PCA} (14). The obtained optimal PCs are then further used to train an SVM classifier for the characterization of breast tumor types. The classification results obtained for the LBP-based CAD system using the PCA-SVM algorithm are shown in Table 5.

The results in Table 5 demonstrate that the 14 optimal PCs yielded an accuracy of 94.1% with an ICA value of 90.4% for the benign class and 96.6% for the malignant class. A total of three instances out of 51 testing instances have been wrongly classified.

Table 6 Classification results obtained for the LBP-based CAD system using the SAE-SM algorithm

Classifier (FS: l)	CM			Acc. (%)	ICA _B (%)	ICA _M (%)
		B	M			
SM (OFS _{SAE} : 25)	B	17	4	92.2	80.9	100
	M	0	30			

Note: *FS* Feature set, *l* No. of optimal features, *CM* Confusion matrix, *Acc* Accuracy, *ICA* Individual class accuracy, *B* Benign class, *M* Malignant class, *SM* Softmax, *OFS_{SAE}* Optimal feature set obtained by using the SAE-SM algorithm

3.1.4 Experiment 4: LBP-Based CAD System Using the SAE-SM Algorithm

In this experiment, the CFS (117) containing both texture and morphological features is subjected to a feature space dimensionality reduction technique using the SAE, generating an optimal feature set OFS_{SAE} (25) that is used to train a softmax classifier for the characterization of breast tumor types. The classification results obtained for the LBP-based CAD system using the SAE-SM algorithm are shown in Table 6.

The results in Table 6 demonstrate that the 22 optimal attributes yielded an accuracy of 92.2% with an ICA value of 80.9% for the benign class and 100% for the malignant class. It is also noted that four benign instances out of 51 testing instances have been wrongly classified, while all the malignant cases have been classified correctly.

3.2 Discussion

The classification performance of four LBP-based CAD system designs using (a) ANFC-LH algorithm, (b) GA-SVM algorithm, (c) PCA-SVM algorithm, and (d) SAE-SM algorithm for breast tumor characterization using ultrasound images has been compared. The analytical comparison of the results obtained for each CAD system is shown in Table 7.

From the analytical comparison of the results presented in Table 7, it can be noted that all four CAD system designs have achieved comparable classification accuracy for breast tumor characterization. The highest accuracy of 96.0% has been achieved using the CAD system based on ANFC-LH followed by the CAD system design based on the PCA-SVM algorithm that achieves an accuracy of 94.1%. A comparable classification accuracy of 92.2% has been obtained for the other two CAD system designs, the difference being in the misclassified cases. For CAD systems based on the GA-SVM algorithm, a total of four instances have been wrongly classified out of 51 testing instances, with ICA values for benign and malignant classes being 90.4% and 93.3%, respectively. For the CAD system based on the SAE with the softmax classifier, even though the number of misclassified

Table 7 Analytical comparison of the results obtained for each CAD system designed for breast tumor characterization

Classifier (FS: l)	CM			Acc. (%)	ICA _B (%)	ICA _M (%)
		B	M			
ANFC (OFS _{LH} : 27)	B	19	2	96.0	90.4	100
	M	0	30			
SVM (OFS _{GA} : 65)	B	19	2	92.2	90.4	93.3
	M	2	28			
SVM (OFS _{PCA} : 14)	B	19	2	94.1	90.4	96.6
	M	1	29			
SM (OFS _{SAE} : 25)	B	17	4	92.2	80.9	100
	M	0	30			

Note: *FS* Feature set, l No. of optimal features, *CM* Confusion matrix, *Acc* Accuracy, *ICA* Individual class accuracy, *ANFC* Adaptive neuro-fuzzy classifier, *OFS_{LH}* Optimal feature set obtained by using the ANFC-LH algorithm, *SVM* Support vector machine, *OFS_{GA}* Optimal feature set obtained by using the GA-SVM algorithm, *OFS_{PCA}* Optimal no. of PCs obtained by using the PCA-SVM algorithm, *SM* Softmax, *OFS_{SAE}* Optimal feature set obtained by using the SAE-SM algorithm, *B* Benign class, *M* Malignant class

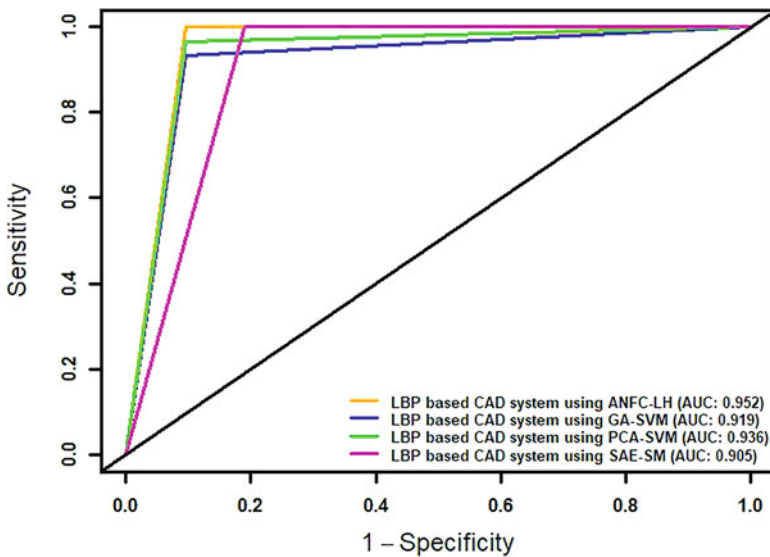


Fig. 16 ROC curves for experiments and their AUC values

instances is 4 out of 51, the advantage of this system is that all the malignant cases are being correctly classified with ICA values for benign and malignant being 80.9% and 100%, respectively.

The ROC curves along with the corresponding area under the curve (AUC) values for the four CAD system designs are presented in Fig. 16. The ROCR library of the R package has been used [42].

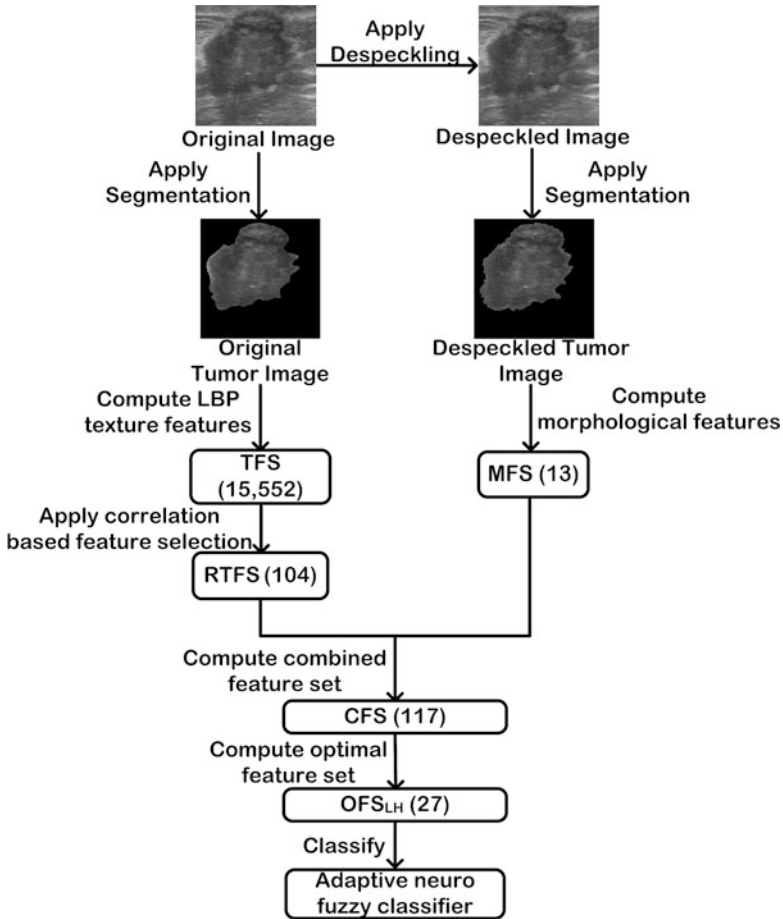


Fig. 17 Proposed ANFC-LH-based CAD system design for breast tumor characterization. (Note: *LBP* Local binary pattern, *TFS* Texture feature set, *RTFS* Reduced texture feature set, *MFS* Morphological feature set, *CFS* Combined feature set, *OFS_{LH}* Optimal feature set obtained by applying the ANFC-LH algorithm)

An ideal CAD system design is the one that increases the sensitivity of correctly identifying the malignant instances; thus the present work proposes the use of the ANFC-LH model for breast tumor characterization as presented in Fig. 17.

4 Conclusion

In the present work, the authors have designed different CAD systems based on different feature sets that are composed of the LBP texture features along with

the morphological features. On the basis of the results obtained for the conducted experiments, it is seen that the CAD system based on the ANFC-LH algorithm yielded an accuracy of 96.0% and an ICA value of 100% for the malignant class utilizing LBP texture features extracted using original breast ultrasound images and morphological features extracted using despeckled images.

The proposed CAD system differs from the other related studies as the majority of the conducted studies consider either original or pre-processed images alone for the analysis of breast abnormalities; however, in the proposed CAD system, the authors have considered a combination of features extracted from both original as well as pre-processed images. These promising results are indicative of the usefulness of the proposed CAD system in routine clinical practice.

References

1. Acharya, U.R., Meiburger, K.M., Koh, J.E.W., Ciaccio, E.J., Arunkumar, N., See, M.H., Taib, N.A.M., Vijayanathan, A., Rahmat, K., Fadzli, F., Leong, S.S., Westerhout, C.J., Astaiza, A.C., Gonzalez, G.R. (2019) 'A novel algorithm for breast lesion detection using textures and local configuration pattern features with ultrasound imagery', *IEEE Access*, Vol. 25, pp. 22829–22842.
2. Aja-Fernández, S., AlberolaLópez, C. (2006) 'On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering', *IEEE Transactions on Image Processing*, Vol. 15, No. 9, pp. 2694–2701.
3. Alivar, A., Danyali, H., Helfroush, M.S. (2016) 'Hierarchical classification of normal, fatty and heterogeneous liver diseases from ultrasound images using serial and parallel feature fusion', *Biocybernetics and Biomedical Engineering*, Vol. 36, No. 54, pp. 696–707.
4. Amin, M., Shahin, A.I., Guo, Y. (2015) 'A novel breast tumor classification algorithm using neutrosophic score features', *Measurement*, Vol. 81, pp. 210–220.
5. Andrei, N. (2007) 'Scaled conjugate gradient algorithms for unconstrained optimization', *Computational Optimization and Applications*, Vol. 38, No. 3, pp. 401–416.
6. Cai, L., Wang, X., Wang, Y., Guo, Y., Yu, J., Wang, Y. (2015) 'Robust phase-based texture descriptor for classification of breast ultrasound images', *BioMedical Engineering OnLine*, Vol. 14, pp. 26–46.
7. Caliskan, A., Yuksel, M.E. (2017) 'Classification of coronary artery disease data sets by using a deep neural network', *The EuRoBiotech Journal*, Vol. 1, No. 4, pp. 271–277.
8. Cetisli, B. (2010) 'Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 1', *Expert Systems with Applications*, Vol. 37, No. 8, pp. 6093–6101.
9. Cetisli, B. (2010) 'Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 2', *Expert Systems with Applications*, Vol. 37, No. 8, pp. 6102–6108.
10. Chan, T.F., Vese, L.A. (2001) 'Active contours without edges', *IEEE Transactions on Image Processing*, Vol. 10, No. 2, pp. 266–277.
11. Chang, C.C., Lin, C.J. (2011) 'LIBSVM: A library of support vector machines', *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 3, pp. 1–27. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. Accessed October 2014.
12. Cheng, J.Z., Ni, D., Chou, Y.H., Qin, J., Tiu, C.M., Chang, Y.C., Huang, C.S., Shen, D., Chen, C.M. (2016) 'Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT', *Scientific Reports*, Vol. 6, No. 1, pp. 1–13.

13. Christiyana, C.C., Rajamani, V. (2012) 'Comparison of local binary pattern variants for ultrasound kidney image retrieval', *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 2, No. 10, pp. 224–228.
14. Cui, J., Sahiner, B., Chan, H.P., Nees, A., Paramagul, C., Hadjiiski, L.M., Zhou, C., Shi, J. (2009) 'A new automated method for the segmentation and characterization of breast masses on ultrasound images', *Medical Physics*, Vol. 36, No. 5, pp. 1553–1565.
15. Daoud, M.I., Badir, T.M., Al-Najar, M., Alazral, R. (2016) 'A fusion based approach for breast ultrasound image classification using multiple-ROI texture and morphological analyses', *Computational and Mathematical Methods in Medicine*, Vol. 2016, pp. 1–12.
16. Fletcher, S.W., Elmore, J.G. (2003) 'Mammographic screening for breast cancer', *New England Journal of Medicine*, Vol. 348, No. 17, pp. 1672–1680.
17. Gardezi, S.J.S., Elazab, A., Lei, B., Wang, T. (2019) 'Breast cancer detection and diagnosis using mammographic data: systematic review', *Journal of Medical Internet Research*. Vol. 21, No. 7, e14464. <https://doi.org/10.2196/14464>.
18. Hassen, T.M., Elmogy, M., Sallam, E.S. (2017) 'Diagnosis of focal liver diseases based on deep learning technique for ultrasound images', *Arabian Journal for Science and Engineering*, Vol. 42, No. 8, pp. 3127–3140.
19. Hegde, R.B., Prasad, K., Hebbar, H., Singh, B.M.K. (2019) 'Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images', *Biocybernetics and Biomedical Engineering*, Vol. 39, No. 2, pp. 382–392.
20. Huang, C.L., Wang, C.J. (2006) 'A GA-based feature selection and parameters optimization for support vector machines', *Expert Systems with Applications*, Vol. 31, No. 2, pp. 231–240.
21. Isen, E., Boran, S. (2018) 'A novel approach based on combining ANFIS, genetic algorithm and fuzzy c-means methods for multiple criteria inventory classification', *Arabian Journal for Science and Engineering*, Vol. 43, No. 6, pp. 3229–3239.
22. Jain, I., Jain, V.K., Jain, R. (2018) 'Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification', *Applied Soft Computing*, Vol. 62, pp. 203–215.
23. Jia, W., Muhammad, K., Wang, S.H., Zhang, Y.D. (2019) 'Five-category classification of pathological brain images based on deep stacked sparse autoencoder', *Multimedia Tools and Applications*, Vol. 78, No. 4, pp. 4045–4064.
24. Keramidas, E.G., Iakovidis, D.K., Maroulis, D., Karkanis, S. (2007) 'Efficient and effective ultrasound image analysis scheme for thyroid nodule detection', in Kamel M, Campilho A (Eds.), *Image Analysis and Recognition, ICIAR*. Springer, Heidelberg, pp. 1052–1060.
25. Key, T.J., Verkasalo, P.K., Banks, E. (2001) 'Epidemiology of breast cancer', *The Lancet Oncology*, Vol. 2, No. 3, pp. 133–140.
26. Kriti., Virmani, J., Agarwal, R. (2019) 'Assessment of despeckle filtering algorithms for segmentation of breast tumors from ultrasound images', *Biocybernetics and Biomedical Engineering*, Vol. 39, No. 1, pp. 100–121.
27. Kriti., Virmani, J., Agarwal, R. (2019) 'Effect of despeckle filtering on classification of breast tumors using ultrasound images', *Biocybernetics and Biomedical Engineering*, Vol. 39, No. 2, pp. 536–560.
28. Kriti., Virmani, J., Agarwal, R. (2020) 'Deep feature extraction and classification of breast ultrasound images', *Multimedia Tools and Applications*, Vol. 79, No. 37, pp. 27257–27292.
29. Kumar, I., Bhadauria, H.S., Virmani, J., Thakur, S. (2017) 'A hybrid hierarchical framework for classification of breast density using digitized film screen mammograms', *Multimedia Tools and Applications*, Vol. 76, No. 18, pp. 18789–18813.
30. Lee, C.Y., Chen, G.L., Zhang, Z.X., Chou, Y.H., Hsu, C.C. (2018) 'Is intensity inhomogeneity correction useful for classification of breast cancer in sonograms using deep neural network', *Journal of Healthcare Engineering*, Vol. 2018, pp. 1–10.
31. Loizou, C.P., Pattichis, C.S. (2008) *Despeckle filtering algorithms and software for ultrasound imaging: Synthesis lectures on algorithms and software for engineering*, Claypool publishers, San Rafael, CA, USA.

32. Lu, R., Shen, Y. (2006) 'Automatic ultrasound image segmentation by active contour model based on texture' in 1st International Conference on Innovative Computing, Information and Control, Beijing, China, pp. 689–692.
33. Maiti, D. (2008) Dimension reduction and classification using PCA and factor analysis—a short overview. [Accessed: August 2018]. Available at <https://pdfs.semanticscholar.org/presentation/c95b/a112cfbbb2842ad3edb21c80acf1871bef82.pdf>.
34. Marcomini, K.D., Caneiro, A.A.O., Schiabel, H. (2014) 'Development of a computer tool to detect and classify nodule in ultrasound breast images', in Aylward, S., Hadjiiski, L.M., (Eds.), *Medical Imaging 2014: Computer-Aided Diagnosis*, SPIE, pp. 90351O-1–90351O-9.
35. Matsumoto, M.M.S., Sehgal, C.M., Udupa, J.K. (2012) 'Local binary pattern texture-based classification of solid masses in ultrasound breast images', in *Proceedings of SPIE 8320 Medical Imaging 2012: Ultrasonic Imaging, Tomography and Therapy*, San Diego, California, USA, pp. 83201H.
36. Menon, R.V., Raha, P., Kothari, S., Chakraborty, S. (2015) 'Automated detection and classification of mass from breast ultrasound images', in *Proceedings of National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics*, Patna, India, pp. 1–4.
37. Michalak, K., Kwasnicka, H. (2006) 'Correlation-based feature selection strategy in classification problems', *International Journal of Applied Mathematics and Computer Science*, Vol. 16, pp. 503–511.
38. Nanni, L., Lumini, A., Brahnam, S. (2010) 'Local binary patterns variants as texture descriptors for medical image analysis', *Artificial Intelligence in Medicine*, Vol. 49, No. 2, pp. 117–125.
39. Nemat, H., Fehri, H., Ahmadinejad, N., Fragi, A.F., Gooya, A. (2018) 'Classification of breast lesions in ultrasonography using sparse logistic regression and morphology-based texture features', *Medical Physics*, Vol. 45, No. 9, pp. 4112–4124.
40. Ojala, T., Pietikainen, M., Maenpää, M. (2002) 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 971–987.
41. Piliouras, N., Kalatzis, N., Dimitropoulos, N., Cavouras, D. (2004) 'Development of the cubic least squares mapping linear-kernel support vector machine classifier for improving the characterization of breast lesions on ultrasound', *Computerized Medical Imaging and Graphics*, Vol. 28, No. 5, pp. 247–255.
42. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria; 2013. <http://www.R-project.org/>.
43. Rawat, J., Singh, A., Bhadauria, H.S., Virmani, J., Devgun, J.S. (2017) 'Classification of acute lymphoblastic leukaemia using hybrid hierarchical classifiers', *Multimedia Tools and Applications*, Vol. 76, No. 18, pp. 19057–19085.
44. Rawat, J., Singh, A., Bhadauria, H.S., Virmani, J., Devgun, J.S. (2018) 'Leukocyte classification using adaptive neuro-fuzzy inference system in microscopic blood images', *Arabian Journal for Science and Engineering*, Vol. 43, No. 12, pp. 7041–7058.
45. Sahiner, B., Chan, H.P., Wei, D., Petrick, N., Helvie, M.A., Adler, D.D., Goodsitt, M.M. (1996) 'Image feature selection by a genetic algorithm: Application to classification of mass and abnormal breast tissue', *Medical Physics*, Vol. 23, No. 10, pp. 1671–1684.
46. Singh, B.K. (2019) 'Determining relevant biomarkers for prediction of breast cancer using anthropometric and clinical features: A comparative investigation in machine learning paradigm', *Biocybernetics and Biomedical Engineering*, Vol. 39, No. 2, pp. 393–409.
47. Takemura, A., Shimizu, A., Hamamoto, K. (2009) 'Discrimination of breast tumors in ultrasonic images using an ensemble classifier based on the AdaBoost algorithm with feature selection' *IEEE Transactions on Medical Imaging*, Vol. 29, No. 3, pp. 598–609.
48. Ubeyli, E.D. (2009) 'Adaptive neuro-fuzzy inference systems for automatic detection of breast cancer', *Journal of Medical Systems*, Vol. 33, No. 5, pp. 353–358.
49. Ultrasoundcases.info [online] <http://ultrasoundcases.info/category.aspx?cat=67> (Accessed 15 July 2016).

50. Uzunhisarcikli, E., Goreke, V. (2018) 'A novel classifier model for mass classification using BI-RADS category in ultrasound images based on Type-2 fuzzy inference systems', *Sadhana*, Vol. 43, No. 9, pp. 138.
51. Virmani, J., Kumar, V., Kalra, N., Khandelwal, N. (2013) 'PCA-SVM based CAD system for focal liver lesions using B-mode ultrasound images' *Defence Science Journal*, Vol. 63, No. 5, pp. 478–486.
52. Yang, J., Honavar, V. (1998) 'Feature subset selection using a genetic algorithm', in Liu, H., Motoda, H., (Eds.), *Feature Extraction, Construction and Selection*, Springer, Boston, MA, pp. 117–136.
53. Yang, J., Yang, J.U., Zhang, D., Lu, J.F. (2003) 'Feature fusion: parallel strategy vs. serial strategy', *Pattern Recognition*, Vol. 36, No. 6, pp. 1369–1381.
54. Zadeh, L.A. (1975a) 'The concept of a linguistic variable and its application to approximate reasoning-I', *Information Sciences*, Vol. 8, No. 3, pp. 199–249
55. Zadeh, L.A. (1975b) 'The concept of a linguistic variable and its application to approximate reasoning-II', *Information Sciences*, Vol. 8, No. 4, pp. 301–357.
56. Zadeh, L.A. (1975c) 'The concept of a linguistic variable and its application to approximate reasoning-III', *Information Sciences*, Vol. 9, No. 1, pp. 43–80.
57. Zakeri, F.S., Behnam, H., Ahmadinejad, N. (2012) 'Classification of benign and malignant breast masses based on shape and texture features in sonography images', *Journal of Medical Systems*, Vol. 36, No. 3, pp. 1621–1627.
58. Zhang, E., Seiler, S., Chen, M., Lu, W., Gu, X. (2020) 'BIRADS features oriented semi-supervised deep learning for breast ultrasound computer-aided diagnosis', *Physics in Medicine and Biology*, Vol. 65, No. 12, <https://doi.org/10.1088/1361-6560/ab7e7d>.
59. Zhang, J., Wang, C., Chang, Y. (2015) 'Comparison of despeckled filters for breast ultrasound images', *Circuits, Systems and Signal Processing*, Vol. 34, No. 1, pp. 185–208.

Detection of Fetal Abnormality Using ANN Techniques



Vidhi Rawat, Vibhakar Shrimali, Alok Jain, and Abhishek Rawat

1 Introduction

Accurate fetal parameter measurement plays a significant role in obstetrics and gynecology for the assessment of proper fetal development. Nowadays, many women suffer from high-risk pregnancy. In this case, continuous and accurate monitoring is required for the proper diagnosis of fetal health. But the accurate evaluation of fetal growth during pregnancy is difficult with ordinary 2D US Images [1]. Recent image processing techniques involving the artificial neural network (ANN) can improve this important aspect of obstetrics and gynecology.

The role of intelligent techniques is very important to detect fetal abnormality in ultrasound images. Fetal abnormality is detected on the basis of fetal biometric parameters [2–6]. These fetal biometric parameters are extracted and measured through segmentation techniques. The extracted features from 2D US images using the segmentation technique form the database of the ANN network. The artificial

V. Rawat (✉)

Department of Electrical Engineering, IES Collage of Technology, Bhopal, India

V. Shrimali

Department of Electronics and Communication Engineering, G. B. Pant Government Engineering College, Delhi, India

A. Jain

Department of Electronics and Instrumentation Engineering, Samrat Ashok Technological Institute, Vidisha, India

A. Rawat

Electrical Engineering, Institute of Infrastructure Technology Research and Management (IITRAM), Ahmedabad, Gujarat, India

e-mail: Arawat@iitram.ac.in

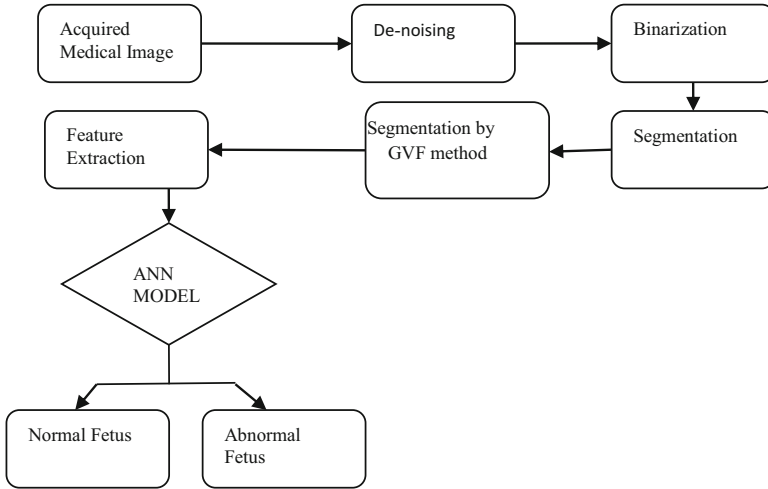


Fig. 1 Block diagram of the proposed approach

neural network (ANN) model is designed for the analysis and classification of fetus status. Figure 1 indicates the block diagram of the proposed approach.

Many diverse neural network designs are offered for biomedical imaging applications. Out of them, a feed-forward network is found to be the most successful one for medical imaging. In this approach, the neurons in each layer are solitarily associated with neurons in the next layer. Signals or information passes through the input layer, hidden layer, and then to the output layer. In this network, normally, a back propagation (BP) algorithm is applied to adjust each neuron's weight and bias values.

In this algorithm, the value of the weight is iteratively modifying based on the error. Error is the comparison between the actual and the target output value. In the network, real alteration of weights is done using a gradient descent algorithm. The neural model is applied to fetus images in two phases. In the first phase, data collection of fetal biometric parameters by the segmentation technique; then in the second phase, a suitable neural network learning algorithm is applied on those data for fetal classification. There are several learning algorithms that can be applied in the detection of fetal abnormality.

2 Artificial Neural Network

ANN is a key intelligent technique used for nonlinear mapping based on the human brain. There are several ANN techniques that can be applied for fetal abnormality detection applications. In this chapter, feed-Forward back-propagation neural network (FFBPNN) has been applied for the detection of fetal abnormality

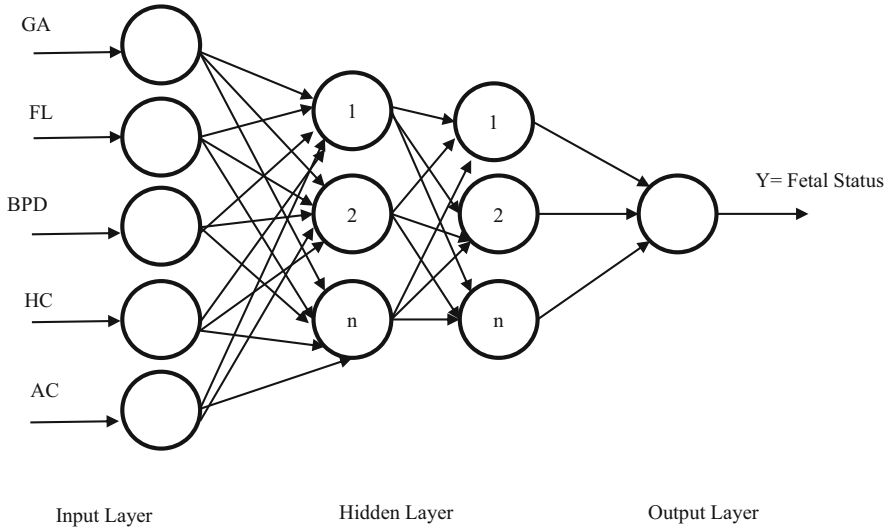


Fig. 2 Proposed feed-forward neural network architecture

because the back-propagation learning approach has been found suitable for fetal abnormality detection. Network FFBPNN is one of the prevalent methods in the area of ANNs. The multilayer perceptron (MLP) involves three or more layers within which each layer is completely associated with the next layer. Basically, MLP is a FFBPNN model that maps input data onto output data [7]. These are unequal connections in which each connection may have a diverse weight. These weights of particular connections encrypt the information of a network. The network works out weights that can be used to regulate the output from a node that is subsequently fed by the next layer.

In the FFBPNN, the input layer consists of the number of I/P neurons, which is equal to the number of selected specific features (GA, FL, BPD, HC, and AC). The output layer determines the anticipated output class. The intermediate hidden layer may rise the fitness of the FFBPNN with nonlinear systems as shown in Fig. 2.

In the literature, various types of back-propagation learning algorithms are described [8–11]. But, it is difficult to find the best one for a specific problem. Generally, Levenberg Marquardt (LMBP), Scale Conjugate Gradient (SCG), Bayesian Regularization (BR), Broyden Fletcher Goldfarb Shanno Quasi-Newton Back-propagation (BFGS), and Conjugate Gradient Back-propagation with Polak-Ribière updates (CGP) are used in the FFNN. But literature shows that LMBP, SCG, and BR are successfully applied in biomedical applications, while the rest of the methods have not been found to be suitable for medical imaging applications. So, all fetal biometric parameters are trained with BPNN and LMBP [12–14], SCG [15], and BR algorithms [16].

In this chapter, the use of the artificial neural network (ANN), especially the feed-forward and back-propagation architecture, for fetus abnormality detection has

been made. In the first phase, two biometric parameters, AC and HC, have been used to detect the abnormality of the fetus by a back-propagation algorithm [17]. Furthermore, five biometric parameters are involved in fetus abnormality detection by LMBP, SCG, and BR algorithms [18]. The proposed neural model will help the radiologist in early and accurate detection of fetal abnormality. The simulation result shows close confirmation with real-time radiologist observations.

3 Experimental Setup for Fetal Abnormality Detection

Fetal images in the DICOM format, with marked and unmarked FL, BPD, HC, and AC region, were obtained from the Peoples Medical College, Bhopal (M.P), India. Then, all processing has been done with the MATLAB release 2.7 version using image processing and the neural network toolbox. Median filters were applied initially to these raw images for preprocessing [19]; then segmentation and measurement of fetal parameters like FL, AC, and HC [20, 21] by the GVF algorithm were performed. All fetal biometric parameters were used to train with the ANN for classification [22].

3.1 Description of Data

Total 500 fetal biometric data of 11–40 weeks have been used for training the neural network. Some fetal biometric data have been extracted from US images, and the remaining data were taken from the radiologist and from the patient's records. The BP neural network has been used on 50 normal and abnormal fetuses for training fetus data and recognizing the fetal status. The result shows that the proposed method can classify the fetus abnormality effectively. Tables 1 and 2 show the value of head and abdominal circumferences corresponding to gestational age. These results are applied to the ANN model for training and validation. All 500 fetal biometric data have been used for training the neural network with LMBP, SCG, and BR algorithms. Fetal biometric features like FL, BPD, AC, and HC have been used for fetal abnormality classification; in all 500 data, some data are given in Table 3.

Standard deviations (std.) and statistical mean are deliberately used to normalize the data. Discrepancies of these parameters provide tentative insinuation about the abnormality. The statistical feature of the fetus is given in Table 4. Statistical mean and standard deviations can be found by the succeeding set of equations:

$$\bar{x} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N x_{i,j} \quad (1)$$

Table 1 Examples of the values corresponding to the head circumference with gestational age

S. no	Gestational age (weeks)	Head circumference (mm)
1	12	80
2	14	108
3	16	128
4	18	120
5	20	170
6	22	188
7	24	220
8	25	231
9	26	200
10	30	210
11	32	288
12	34	305

Table 2 Examples of the values corresponding to the abdominal circumference with gestational age

S. no	Gestational age (weeks)	Abdominal circumference (mm)
1	12	63
2	14	84
3	15	96
4	16	106
5	18	131
6	20	151
7	24	201
8	25	212
9	26	223
10	30	262
11	32	283
12	34	305

where M, N represents a spatial row-column variable.

$$\text{std} = \sqrt{\frac{1}{(N - 1)(M - 1)} \sum (x_{i,j} - \bar{x})} \tag{2}$$

4 Results Using Back-Propagation Algorithm

Total 50 US images have been used for the proper estimation of the training and testing of the ANN model. The data set was acquired by appropriately preprocessing of US images and then computing the structures from images. The actual output and the target value of the network are compared and the error has been calculated as

Table 3 Fetal biometric parameters of different images

S. no.	Gestational age (weeks)	Femur length (mm)	Biparietal diameter (mm)	Head circumference (mm)	Abdominal circumference (mm)
1	12	08	15	80	63
2	14	12	24	108	84
3	15	14	32	128	96
4	16	23	38	170	106
5	18	34	49	188	131
6	20	42	57	208	151
7	24	48	53	220	201
8	25	49	66	231	212
9	26	54	71	200	223
10	30	59	79	210	262
11	32	63	80	288	283
12	34	70	89	305	305

Table 4 Statistical features regarding the normalization of the fetal biometric parameters

Fetal biometric parameters	No. of sample	Mean	Std.	Max	Min
GA	500	25.48	5.929	32	11
FL	500	44.57	25.09	125	01
BPD	500	58.75	28.26	150	10
HC	500	214.87	90.23	425	32
AC	500	208.03	93.18	410	18

given in Tables 5 and 6. The error minimization curve for each epoch of the neural network during learning is shown in Fig. 3. As we increase the number of neurons, the mean square error reduces. The final parameters of the successfully trained neural network are shown in Table 7. Experimental fetal US images with numerical data findings have been divided by an expert radiologist into two states of fetus like: abnormal and normal fetus. Investigational outcomes showed good accuracy and efficiency of the algorithm in clinical applications. After the appropriate training of the ANN model, normal and abnormal fetuses can be rapidly distinguished.

5 Results Using LMBP, SCG, and BR Algorithm

A total of 500 fetal data were applied for the assessment of the training and testing of the ANN models. Fetal data are separated into three sets, correspondingly for training, 70% and 15% each for testing and validation. For proper learning, the FFNN with a back-propagation algorithm is used. Two hidden layers and five input nodes are applied in the proposed ANN model. The ANN model architecture is given in Fig. 4.

Table 5 Target and actual output with the error of HC

S.no	Gestational age(weeks)	Head circumference (mm)	Target output (1- normal 0-abnormal)	Actual output	Error
1	12	80	0	0.047	-0.047
2	14	108	1	0.789	0.211
3	16	128	1	0.656	0.343
4	18	120	0	0.021	-0.021
5	20	170	1	0.612	0.387
6	22	188	1	0.797	0.202
7	24	220	1	0.923	0.076
8	25	231	1	0.894	0.105
9	26	200	0	-0.129	0.129
10	30	210	0	-0.125	0.125
11	32	288	1	0.752	0.247
12	34	305	1	0.912	0.087

Table 6 Target and actual output with the error of AC

S. no	Gestational age (weeks)	Abdominal circumference (mm)	Target output (1- normal 0-abnormal)	Actual output	Error
1	12	63	0	0.027	-0.027
2	14	84	1	0.6784	0.3216
3	15	96	1	0.7568	0.2432
4	16	106	1	0.7881	0.212
5	18	131	1	0.7456	0.2544
6	20	151	1	0.6879	0.3121
7	24	201	1	0.8765	0.1235
8	25	212	1	0.8190	0.181
9	26	223	0	0.2340	-0.234
10	30	262	0	0.245	-0.245
11	32	283	1	0.8764	0.1236
12	34	305	1	0.8224	0.1776

Table 7 Neural network final training parameters

Input nodes	2
Hidden nodes	100
Output nodes	1
Learning rate	0.4
MSE	0.0001
Iterations	6000
Training time (seconds)	2015
Run time (seconds)	0.01

A total of three training algorithms are applied to the given input, targets, and the number of neurons. The target matrix consisted of zero and one; zero for fetal

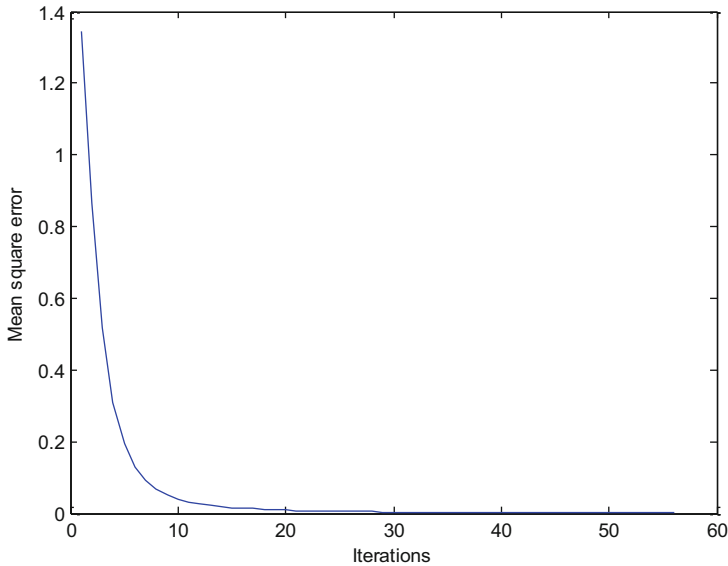


Fig. 3 The neural network learning curve

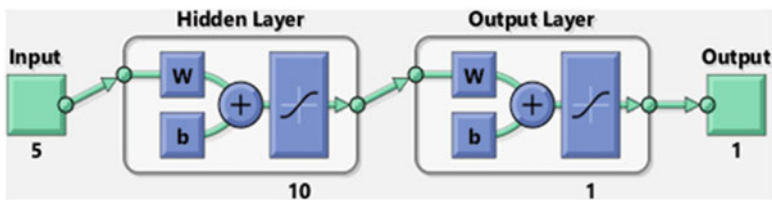


Fig. 4 Block diagram of the FFNN architecture

abnormality and one for the normal fetus. The normal and abnormal status of the fetus can be detected by the output of the ANN model. Every algorithm is applied at least ten times, and the best performance results of LMBP, SCG, and BR algorithms are saved, as shown in Figs. 5, 6, and 7. These figures also show the number of epochs required corresponding to a certain mean square error.

Training performance shows a mean square error with the corresponding epoch. Low MSE gives optimized weights and biases values.

The final neural network parameters, like the number of nodes, the number of neurons, and mean square error (MSE), are given in Table 8. Percent of accuracy is given by:

$$\% \text{ of Accuracy} = \frac{\text{accurate cases}}{\text{total cases}} \times 100 \tag{3}$$

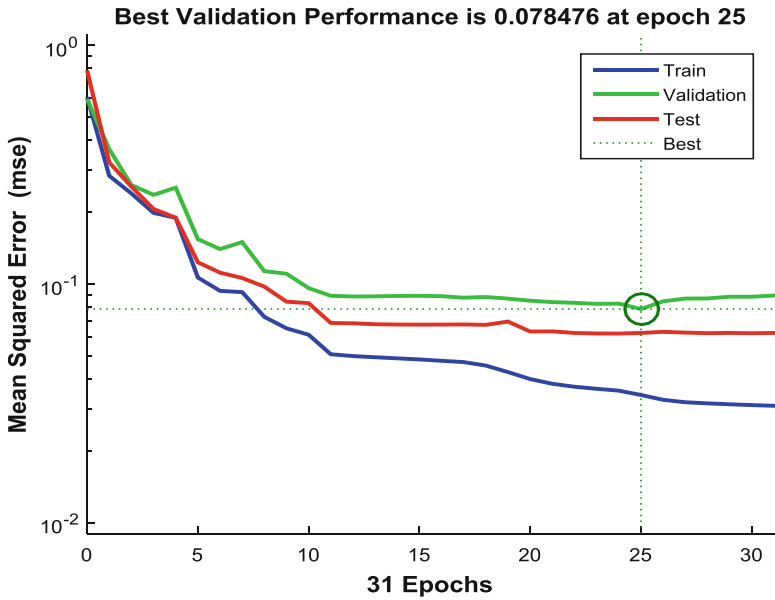


Fig. 5 Training performance of the Levenberg–Marquardt algorithm

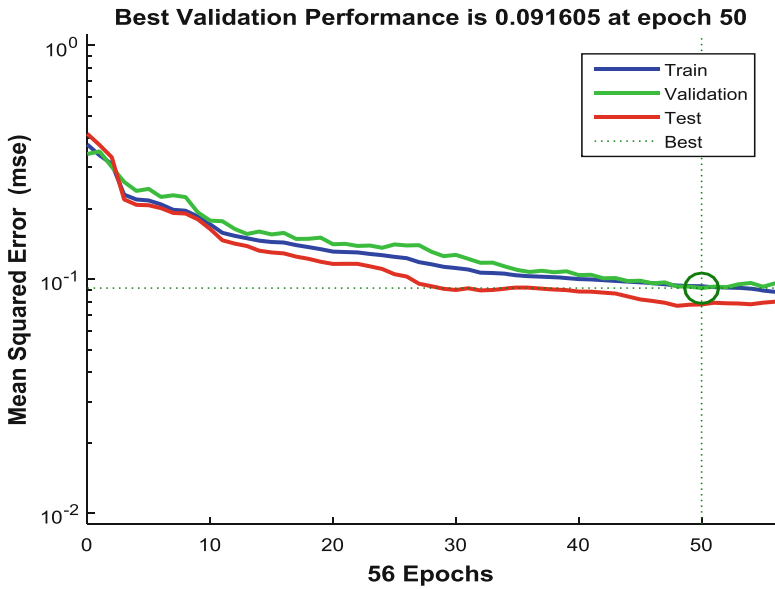


Fig. 6 Training performance of the SCG algorithm

The complexity of the system depends on the number of neurons and hidden layers. The data set is randomly classified into two fetus states (abnormal and

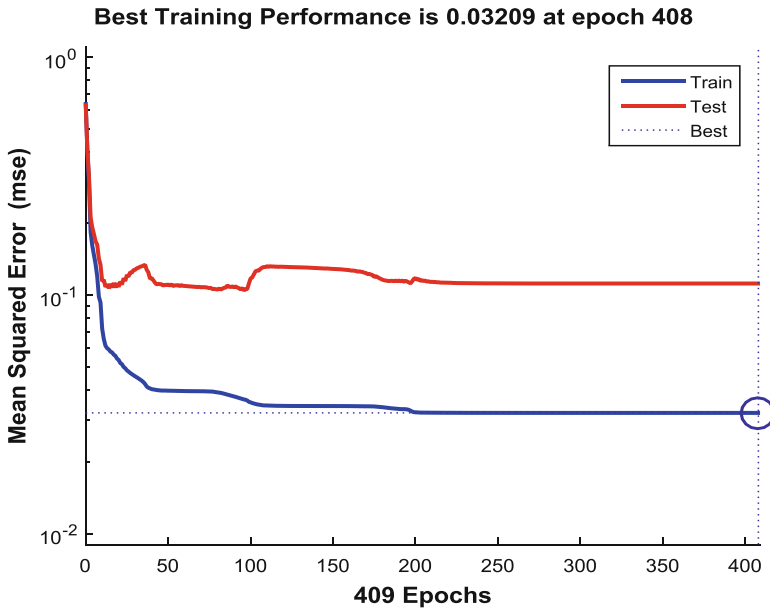


Fig. 7 Training performance of the Bayesian regularization algorithm

Table 8 Performance comparison of algorithms

Algorithm	No. of neurons	Epochs	MSE	Percent of accuracy
BR	10	409	0.03209	100
SCG	20	50	0.09165	60
LMBP	10	31	0.0789	68

normal) after taking proper comments from expert radiologists. These data are applied for training, validation, and testing. The results found in the testing phase are cross-verified with expert radiologists. Proper training of the ANN model can separate the fetus status as the case may be.

6 Comparison of LMBP, SCG, and BR Algorithms

The fetal biometric parameters have been used to train these three algorithms, and the comparison is based on their respective performance. The output, MSE, and training time of each network are calculated. Computational speed and MSE are the key parameters on the basis of which comparison has been made, as given in Table 8. The training methods are completed on the accomplishment of itemized iterations or on the failure of validity, or on the accomplishment of the performance target. Results show that the output of a network with ten neurons is good in the case

of LMBP and BR algorithms and with 20 neurons in the SCG algorithm case. The target of training is to achieve an optimal weight of each connection of neurons using LMBP, SCG, and BR algorithms. After that validation and testing of the proposed ANN, the model is verified with AC, BPD, HC, and FL of 28–32 weeks of fetus.

As per Table 8, the Bayesian regularization algorithm provides an anticipated output with 10 neurons and the SCG algorithm with 20 neurons. In view of complexity and speed, lesser neurons are desirable. Hence, the Bayesian regularization algorithm is good for the fetal parameter analysis. The mean square error has been found to be minimum in the Bayesian regularization algorithm and the highest in the case of the SCG algorithm. Testing results of the neural model show that the Bayesian regularization algorithm has the highest accuracy in detecting fetal abnormality.

7 Conclusion

In the first part, BPNN-based neural models have been applied for the analysis and detection of fetal abnormality using two biometric parameters. Furthermore, in the second part, LMBP, SCG, and BR algorithms in feed-forward back-propagation modes have been applied using five fetal biometric parameters. This approach could offer a real-time and accurate analysis of fetus status and, thus, improved quality of life. The BPNN is suitable for two fetal parameters but its computational speed is low with five fetal parameters.

Hence, the fetal abnormality is detected using three neural network learning algorithms LMBP, BR, and SCG of the FFNN and is compared for the best assessment of fetus growth. Performance of the training algorithm is case sensitive and closely dependent on training parameters. A wide range of training parameter variations has been made, which tested the performance for each case. The result shows that the Bayesian regularization algorithm is the most suitable choice for fetal abnormality detection. Therefore, the Bayesian regularization algorithm-based neural model could offer the best prediction for the patient in real-time, accurate fetal status, and hence improved quality of life is possible. This neural model is reliable and highly adaptable in any environment. It can be incorporated into the software of US machines to provide the best prediction.

References

1. C. W. Hanna and A. B. M. Youssef, "Automated measurements in obstetric ultrasound images", Proceedings of IEEE International Conference on Image Processing, Santa Barbara, CA, Vol. 3, pp. 504–507, 1997.
2. J. J. Cronan, "Ultrasound: Is there a future in diagnostic imaging?", Journal of the American college of Radiology, Vol. 3, Issue 09, pp. 645–646, 2006.

3. M. W. Miller, A. A. Brayman, J. S. Abramowicz, "Obstetric Ultrasonography: A biophysical consideration of patient safety the "rules" have changed", *American Journal of Obstetric Gynecology* Vol. 179, Issue 01, pp. 241–254, 1998.
4. R. Sanders and A. James, "The Principles and Practice of Ultrasonography in Obstetrics and Gynecology", 3rd edition, Appleton-Century-Crofts, Connecticut, Ch. 9 & 10, 2011.
5. S. L. Bridal, J.M. Correias, A. Saied, and P. Laugier, "Milestones on the road to higher resolution, quantitative, and functional ultrasonic imaging", *Proceeding IEEE*, Vol. 91, No. 10, pp. 1543–1561, Oct. 2003.
6. D. C. L.L Ferreira, E. Hardy, M. J.O. Duarte, A. Faúndes, "Termination of Pregnancy for Fetal Abnormality Incompatible with Life: Women's Experiences in Brazil" *Report Health Matters*. Vol. 13, No 26, pp.139–46, Nov. 2005.
7. Priddy K.L., Keller P.E., "Introduction in Artificial Neural Networks: An Introduction", Bellingham Wash., St. Bellingham USA, pp.1–12, 2005.
8. Xinxing Pan, B. Lee and Chunrong Zhang, "A comparison of neural network backpropagation algorithms for electricity load forecasting", 2013 IEEE International Workshop on Intelligent Energy Systems (IWIES), pp. 22–27, Vienna, 2013.
9. S. Haykin, "Neural Networks: A Comprehensive Foundation", 2nd. Prentice-Hall; Englewood Cliffs, NJ: 1999.
10. B. Kröse, P. V. D. Smagt, "An introduction to Neural Networks", The University of Amsterdam; Amsterdam: 1996.
11. C.G. Christodoulou, M. Georgiopoulos, "Application of Neural Networks in Electromagnetics", Artech House; MA, USA: 2001.
12. Levenberg Keneth "A Method For the Solution Of Certain Nonlinear Problems In Least Squares", *Quarterly of Applied Mathematics*, Brown University, Vol. 2, pp. 164–168, 1944
13. Marquardt D., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters", *SIAM Journal on Applied mathematics*, Vol.11, No.2, pp.431–441, June 1963.
14. Hagan M.T. and M. Menhaj, "Training Feed Forward Network with the Marquardt Algorithm", *IEEE Transactions on Neural Network*, Vol. 5, No.6, pp. 989–993, 1999.
15. Martin Fodslette Møller, "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning Neural Networks", *Neural Networks*, Vol. 6, Issue 4, pp. 525–553, 1993.
16. Zhao Yue, Zhao Songzheng and Liu Tianshi, "Bayesian regularization BP Neural Network model for predicting oil-gas drilling cost", 2011 International Conference on Business Management and Electronic Information, pp. 483–487, Guangzhou, 2011.
17. Vidhi Rawat, Alok Jain, Vibhakar Shrimali, Abhishek Rawat, "Automatic Detection of Fetal Abnormality using head and abdominal circumference", *ICCCI 2016, Part II, LNAI*, Vol. 9876, pp. 525–534, 2016. (https://doi.org/10.1007/978-3-319-45246-3_50).
18. Ashwaq Qasem, Siti Norul Huda Sheikh Abdullah, Shahnorbanun Sahran, Tengku Siti Meriam Tengku Wook, Rizuana Iqbal Hussain, Norlia Abdullah, Fuad Ismail, "Breast Cancer Mass Localization Based On Machine Learning", *IEEE 10th International Colloquium on Signal Processing and its Applications*, pp. 31–36, Kuala Lumpur, 2014.
19. Vidhi Rawat, Alok Jain, Vibhakar Shrimali, "Investigation and Assessment of Disorder of Ultrasound B-mode Images", *International Journal of Computer Science and Information Security*, Vol.7, No.2, February 2010.
20. Vidhi Rawat, Alok Jain, Vibhakar Shrimali "Analysis and Assessment of Ultrasound Images for Fetal Biometry Using Morphological Operators", 5th Indian International Conference on Artificial Intelligence (ICAI-11), pp. 1271–79, Tumkur, India, Dec. 14–16, 2011. (ISBN: 978-0-9727412-8-6).
21. Vidhi Rawat, Alok Jain, Vibhakar Shrimali, "Analysis of Ultrasound Images for Assessment of Gestational Sac using Gradient Vector Force", *International Journal of Biomedical Engineering and Technology*, Inderscience, Vol. 12, No 4, pp. 221–233, 2013. (<https://doi.org/10.1504/IJBET.2013.057650>).
22. Vidhi Rawat, Alok Jain, V. Shrimali and Sammer Raghuvansi, "Neural Modeling of Fetal Biometric Parameters for Detection of Fetal Abnormality", *IETE Journal of Research*, pp. 1–13, Jan 2019. (<https://doi.org/10.1080/03772063.2019.1565948>)

Machine Learning and Deep Learning-Based Framework for Detection and Classification of Diabetic Retinopathy



V. Purna Chandra Reddy and Kiran Kumar Gurralla

1 Introduction

Over the last decade, the analysis of high-resolution colour digital photography has received the attention of researchers. Due to different modalities of images, the investigation of digital images is a challenging task [1]. With the help of a conventional digital camera, images of the retina of an eye can be easily captured. High-quality data of the retina's appearance can then be preserved. In the long run, it has been observed that storage, retrieval and transmission without degrading the image quality are not feasible [2]. DR is one of the more recent applications of retinal digital imaging. It is the most common cause of vision loss among working-age people. In recent times, the UK government has suggested that diabetic patients above 12 years old get annual eye screenings using digital retinal photography [3]. It is a unique feature that these images may be captured from anywhere, regardless of time and place. Henceforth, quality assurance of those captured images must be ensured, which is also an integral component of the screening programs. The main causes of diabetes are physical inactivity, increased levels of obesity and ageing. The report states that diabetes rates are predicted to slowly increase from 2.8% of the population in 2000 to 4.4% by 2030. Figure 1 presents the detailed view of regions of a human eye affected by DR.

Due to the increasing diabetic population, the quality assurance of these images must be ensured within a limited time and rapidly investigated with the help of pattern recognition algorithms [4, 5]. It is observed that the process of identifying the lesion data and their decision variables in a retinal image is a non-trivial task, due to several sets of operations involved in low-level as well as high-level image

V. Purna Chandra Reddy (✉) · K. K. Gurralla
Department of ECE, NIT Andhra Pradesh, Tadepalligudem, Andhra Pradesh, India
e-mail: kirankumargurralla@nitandhra.ac.in

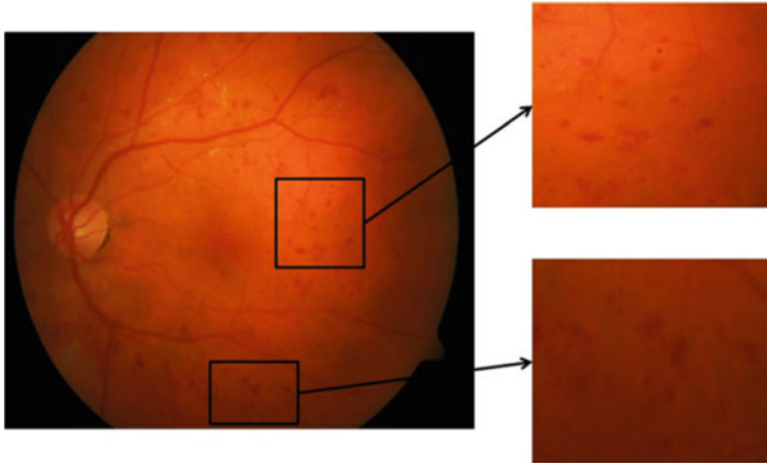


Fig. 1 Detailed view of DR-affected regions

processing [6]. In the algorithmic operation performed on digital retinal images, it composes of different stages that are linked together to achieve efficient outputs.

1.1 Symptoms of DR

As we know that microaneurysms and neovascularisation, intra-retinal haemorrhages, exudates, red lesion, area, perimeter, width and branching angles, etc. are important DR clinical geometrical and haemodynamic features [7]. Let us see detailed data of DR features.

- (a) *Microaneurysm*: It is the deformation formed near the walls of the blood vessels. It is represented as balloon-shaped that degraded the walls of the vasculature [8]. Several microaneurysms were formed for finding the levels (or) stages of DR. Their default size is of 1–3 pixels.
- (b) *Haemorrhage*: It is formed due to the overflow of blood from infected capillaries. It is further divided into three categories, namely, dot, flame, and blot. Each category has distinguished itself from its red spots. It has a larger size than MA, and it has a flame shape. When the blood vessels damage with nerve fibres, then haemorrhages are easily formed [9].
- (c) *Hard exudate*: It is in bright yellow- (or) white-coloured objects formed near the retinal region. It has a waxy outlook with sharp edges over the background from blood vessels. It is also developed due to the leakage of blood from veins and its nearby vessels.
- (d) *Soft exudate*: It occurs due to the occlusion of the arteriole. Due to the reduced blood flow near the retinal region, it causes ischemia near the retinal nerve

fibre layer [10] that causes ganglion cells around the retinal regions. This debris detection is a trivial task.

1.2 Types and Levels of DR

Based on the formation of several microaneurysms and haemorrhages near the retinal images, there exist different stages of DR [11] which are explained as follows:

- Normal: If the signs/symptoms of the DR are absent, it falls under the class of normal.
- Mild: If the signs/symptoms of the microaneurysms are present, it falls under the class of mild.
- Moderate DR: If the signs/symptoms of the microaneurysms and haemorrhages are presently less than 20 in each quadrant, it falls under the class of moderate DR, i.e. hard exudates.
- Severe DR: If the signs/symptoms of the microaneurysms and haemorrhages are presently more than 20 in each quadrant, it falls under the class of severe DR, i.e. soft exudates.

The proliferates of DR are mostly characterized by neovascularization (NV), which is formed from infected blood vessels [12]. Henceforth, investigation of MA and NV is of prime importance in finding out the type of lesions and differentiating exudates from non-exudates.

1.3 DR-Computer-Aided Diagnosis (CAD)'s Perspective

By analyzing the fundus eye images, the level of DR shall be obtained. An intelligent process of detecting and classifying the DR via image processing methods is an integral part of the medical image system [13]. To obtain the information from low, middle and high images, some image processing techniques on segmentation and localization are performed over OD, blood vessels, microaneurysms, haemorrhages, vessels branching angles and other features. Optical features of microaneurysms, haemorrhages and vessels are extracted for the experimental purpose. Based on the size of blood vessels, the classification algorithms are used. Several CAD systems [7] are invented to detect, classify and predict the DR and its level. With the assistance of previously selected features space, the classification of DR is then made. A possible combination set of features helps define the grade of DRs. Several CAD models have been available in the literature to detect DR and classify its level of lesion [14]. Feature extraction and classification algorithms are the key indicators of the algorithmic part. The efficiency of the feature extraction process is directly proportional to the efficiency of the classification algorithm.

2 Detection of DR Using Image Processing

The detection process of DR using image processing techniques consists of stages like pre-processing, segmentation, feature extraction and classification stages as shown in Fig. 2 [15]. The pre-processing stage depicts the normalization of the fundus images like brightness, contrast enhancement, etc. The segmentation stage portrays the segmenting of the infected retinal regions from deformed blood vessels, nerves and pathologic lesions. The task of feature extraction is to estimate the required quantitative information of the feature space [16]. Finally, the selected features are given as an input to classification algorithms and thus, the performance of the systems is computed. Quality of the fundus images alters due to various factors such as different movements of the eye, the opacity of an eye, pupils' changes, brightness and intensity of an eye. Henceforth, pre-processing of images is a mandatory task. The following are the steps of image pre-processing explained below [17]:

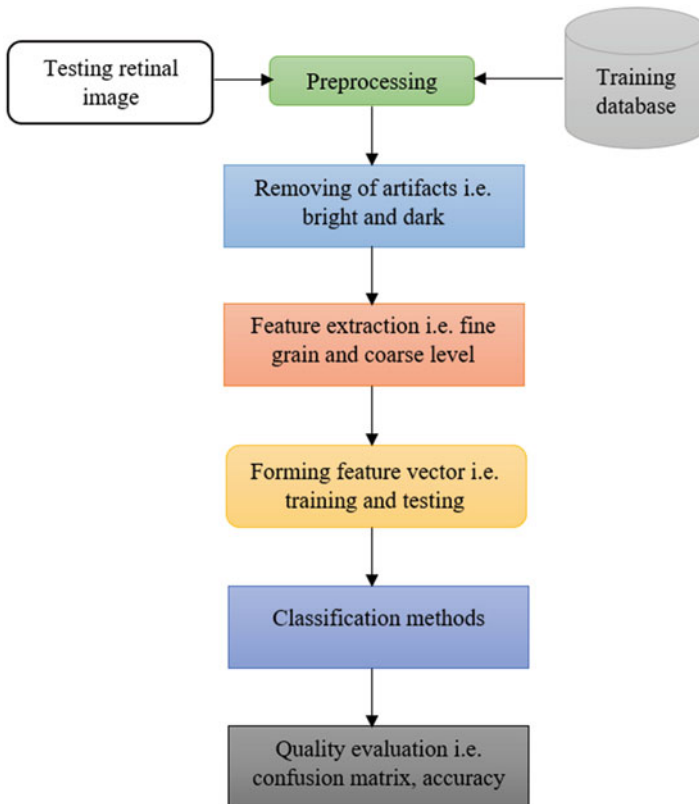


Fig. 2 General block diagram for the detection and classification of DR images

- **RGB space:** Here, the green plane surface depicts the prominence of the blood vessels, and noise is represented as red and blue pixels. These red and blue pixels are improved for efficient segmentation outcomes. In this, the RGB image is transformed into several spaces to pre-process the acquired image [18].
- **Enhancement of contrast and denoising models:** It is helpful to improve the quality of an image. It is purely based on enhancing the images. Sample algorithms consist of the Contrast Limited Adaptive Histogram Equalization (CLAHE), mode, median and Gaussian filters.
- **Segmentation:** Here, some morphological operations are performed over an image via mathematical operators [19].
- **Structural elements:** Grey-level images are performed by analyzing their structural elements. These also differ in terms of size, shape and orientation. Here, all sorts of the circular, elliptical and linear structures were analyzed.
- **Dilation:** It is the easiest operator that translates the pixel values.
- **Erosion:** It depends on the structure of the element. It just erodes the foreground pixels, like the background pixels. It is mostly used for preserving the edges and boundaries detection.
- **Opening:** It is used for shape filters and eliminates smaller objects.
- **Closing:** It is the dual operator, which preserved the background region on its matching shapes.

3 Related Work

The section presents reviews of existing techniques suggested by various researchers. Nadeem Salamat et al. [20] presented a review of DR using retinal images. They reviewed about merits and demerits of 79 algorithms that are subjected to the detection and classification of DR. It is observed that feature selection acts as the main key indicator to classify the process. In [21], authors presented a CANet algorithm that detects both the DR and diabetic macular oedema (DME) grading. The relationship between diseases at the image level was analyzed using deep network features. Two modules, disease-specific and disease-dependent features were collected and trained for classifiers. Authors in [22] presented a deep learning model that detected DR at various stages. Prior research stated that low accuracy was obtained during the classification stage. Therefore, convolutional neural network (CNN) models like Resnet50, Inceptionv3, Xception, Dense121 and Dense169 were employed to select the relevant features, and then the classification tasks were processed. In [23], a binocular Siamese-like CNN via a transfer learning process has been addressed. With the help of Inception V3, instead of feeding a single eye as an input, a group of fundus images was given. By doing so, the authors have seen remarkable results in classification accuracy.

In [24], authors presented a multiple instance learning (MIL) model, which leveraged the discontinued information in the image annotations, which inclined the error rate. Here, the joint optimization encoding scheme was also explored

using pathological images with the medium features set. It depicted the importance of the decision-making process. In [25], a super-pixel multi-feature classification on digital colour fundus images is presented, where 19 multi-channel intensity features are taken to characterize the candidates, and classification results have maximized between the class-scatter and the within-class scatter. Authors in [26] explored the non-proliferative framework to detect the symptoms of retinopathy using neural networks. A novel morphological algorithm was designed to segment the retinal lesions of an eye. Then, a set of 19 features is collected and learned via artificial neural networks. Finally, the outputs are classified into normal, mild, moderate and severe. With the help of a back-propagation network, features are easily trained for the ANN. In [27], authors discussed the significance of converging the local features. They introduced automatic detection using hybrid sampling and boosting classifiers that discriminated the microaneurysm from non-microaneurysm candidates. It is observed that the convergence of intensity and shape descriptor features developed different modalities. An automatic classification system on red lesions in longitudinal fundus images is addressed in [28]. The authors have found that due to the differentiation in the illumination and the contrast of a retinal feature lowered the performance of the system. Here, the SVM classifier was employed to train the intensity and shape features. Due to small changes in the lesions, the classification rate may vary, which causes high modalities. In [29], authors presented a Multi-Sieving deep learning algorithm that detected the retinal microaneurysm via hybrid text/image mining. They provided a semantic solution between images and diagnostic information. It also additionally detected the unbalanced microaneurysm via a CNN that leveraged all sorts of supervised information. It aimed to remove the research gap of low-level image features for diagnostic information.

In [30], authors suggested a leakage detection model for diabetic and malarial retinopathy. Authors have observed the image modalities in fluorescein angiography. Here, saliency maps were generated between intensity and compactness of the images which helped for differentiating the superpixel of all images. An averaging operator is a unique feature employed in the generation of saliency maps. If two saliency maps share a similar pixel-wise multiplication operator, then it is considered as leakage regions. Furthermore, graph-cut segmentation models were employed over different saliency cues. In [23], a Siamese-like CNN for the detection of DR is explored. Initially, a Siamese-like architecture was trained using transfer learning. Then, whole binocular fundus images were given as an input to the network model and their correlation was studied. Along with this inception, V3 was also used for discovering the pre-processing modules of the binocular designs. Authors in [31] presented a reliable detection model that localized the microaneurysms via singular spectrum analysis. Here, dark objects are removed using the filtering process. The images are analyzed via multiple directions via a singular spectrum model. Then, the correlation coefficient is estimated between the observed profile and the actual profile measured from its shape scale factors. In [32], a microaneurysm detection model using PCA and machine learning methods is presented. Images are arranged into 25 by 25 patch pixels and classified via the random forest, neural networks and support vector machine. The system has reduced

the dimensionality of the inputs. In [33], the morphology mean shift algorithm, which recognized the exudates of retinal images, was presented. The images are pre-processed by normalization, contrast enhancement and the removal noise. Finally, the mean shift process is applied for defining the coarse information analysis. At last, the morphology algorithm is used for classification that contains exudates pixels. Recently, a deep multiple instance learning mode by image level annotation features is presented in [34]. It helped to learn the features via improvement of DR images via lesions analysis. Some features space was independent of instances. Authors in [35] presented a model to label the DR images by using the modified AlexNet architecture. The severity of the diseases is found and then classified into its level using SoftMax and the Rectified Linear Activation Unit (ReLU). The system has improved the learning parameters.

4 Comparative Analysis

In this section, a comparative table is developed based on the studies discussed above. The table is developed based on the performance evaluation parameters and the merits and drawbacks of the methods (Table 1).

5 Experiments and Results

5.1 Datasets

For performing the various kinds of experiments, Indian Diabetic Retinopathy Image Dataset (IDRID) along with the ISBI 2018 sub-challenge 2 dataset is considered. It consists of 5 class DR with 5 levels of grades for classification operation ranging from 0 to 4 along with DME grading ranging from 0 to 2. The dataset contains 103 images available for testing purpose, and 413 images are considered for training operation.

5.2 Performance Evaluation

Table 2 demonstrates the performance of various classifiers with different quantitative evaluation. The accuracy of the CNN is 85.48%; this is higher than the other classifiers. The precision rate of a Naive Bayes classifier is 94.31%, and the CNN framework achieves 95.16%, which is higher than the other classifiers. The recall rate for the Naïve Bayes classifier is 92.16%, which is also higher than the other classifiers, and the CNN ensemble learner attains the value of 93.33%.

Table 1 Comparative analysis of various DR detection and classification approaches from the literature

Author name and year	Details of dataset	Performance measures	Limitations
Li et al., 2019 [21]	ISBI 2018 IDRid challenge dataset and Messidor dataset.	Joint accuracy, precision, recall and F1 score were estimated. The performance measures of DR are joint accuracy (85.1%), recall (92.0) and F1 measure (91.2). Likewise, performance measures of DME are recall (70.8%) and F1-measure (72.4%).	During multi-task learning models, an increased number of features are dropped out to achieve a better precision rate. During lesion segmentation, DR takes more time for domain labelling.
Qummar et al., 2019 [22]	Kaggle dataset.	The proposed model is explored on five classes. Class 0 dictates recall (0.97), precision (0.84), specificity (0.40) and F1-score (0.90). Class 1 dictates recall (0.80), precision (0.51), specificity (0.99) and F1-score (0.15). Class 2 dictates recall (0.41), precision (0.65), specificity (0.95) and F1-score (0.50). Class 3 dictates recall (0.51), precision (0.48), specificity (0.98) and F1-score (0.49). Finally, class 4 recall (0.56), precision (0.69), specificity (0.99) and F1-score (0.62).	Though imbalanced data issue is resolved, the augmentation process of each dataset is expensive in terms of cost and time while selecting the features from fundus images.
Zeng et al., 2019 [23]	Kaggle DR competition is provided by EyePACS, which is composed of different imaging conditions.	Results are stated under two models, the monocular model has yielded AUC (0.940), sensitivity (77.4%) and specificity (63.5%), whereas the binocular model has yielded AUC (0.951), sensitivity (82.2%) and specificity (70.7%).	It is sensitive towards the costs and higher classification error rate.
Costa et al., 2017 [24]	Messidor dataset contains 1200 colour fundus images for the period of 2005 and 2006.	The new loss function was estimated from implicit information. The system has achieved 93% AUC.	It is observed that during image labelling, some lesions and their features are ignored due to complex features.

Zhou et al., 2017 [25]	DiaretDB1 and e-ophtha EX datasets were used.	System has achieved performance on contextual features, sensitivity (87.12%), specificity (94.01) and 0.9587 (AUC).	The exudates of images altered the tone of the illumination that weakened the performance of the classification system. Growing of irrelevant regions was also enhanced.
Al-Jarrah et al., 2017 [26]	Fundus images from DIARETDB1, PDR DIARETDB0 and databases.	The system has achieved accuracies of 96.6% and 89.9%.	Hidden layers are composed of eight nodes that make the learning process simpler. Yet, the morphological operations consume more time.
Dashbozorg et al., 2018 [27]	Retinopathy online challenges (ROC) dataset, e-ophtha MA, RC-RGB-MA, RC-SLO-MA and DiaRetDB1.	The system has yielded an average sensitivity score of 0.471. Likewise, AUC analysis has yielded 0.798 (e-ophtha), 0.785 (E-ophtha-MA), 0.732 (RC-RGB- MA) and 0.698 (ROC).	Variability in image resolution is not analyzed. The computational burden of the feature candidate extraction has increased the time taken for the feature selection process.
Adal et al., 2017 [28]	DR screening programs at Rotterdam eye hospital which composed of 81 diabetic eyes for years 2012 and 2013 were taken for study purpose.	It has yielded a detection sensitivity rate (98%). It has reduced the false alarm rate by about 18% with low SNR regions of 19%.	Due to lowered image quality, the feature extraction process still takes training time. It also leads to lowered clinical significance.
Dai et al., 2018 [29]	Clinical datasets are acquired from DR patients.	The system has achieved recall (0.878), precision (0.997), accuracy (0.961) and F-score (0.934).	High false positive rate was observed. When the patch size increases, some unbalanced issues of image modalities are not focussed.
Zhao et al., 2016 [30]	Two publicly available datasets have been taken for experimental purpose, namely, malarial retinopathy and DR. DR data were collected from the vision and image processing laboratory and 25 MR images were collected from Liverpool Reading Center at St Paul's eye unit, Royal Liverpool University Hospital and the Department of eye and Vision Science, University of Liverpool.	Evaluation metrics such as sensitivity, accuracy, specificity and AUC were analyzed to measure the performance of the system. By varying the learning rate of the class model, the above-mentioned metrics were assessed. MR images have achieved sensitivity (0.90), accuracy (0.90), specificity (0.94) and AUC (0.92) and likewise, DR images have sensitivity (0.76), accuracy (0.88), specificity (0.88) and AUC (0.85)	During the feature extraction process, over-segmentation and over-intensive computation processes are observed.

(continued)

Table 1 (continued)

Author name and year	Details of dataset	Performance measures	Limitations
Wang et al., 2016 [31]	Retinopathy online challenge (ROC), DiaretDB1 2.1 and Moorfields eye hospital datasets.	The system has achieved the F-score of 0.464. False positive rate is estimated to range between 13% and 16%.	Due to the low contrast and blurry outline, the background edges are not preserved. Some low probabilities have degraded the subsets of candidate's functions.
Cao et al., 2018 [32]	Diabetic retinopathy database: Calibration level 1 (DIARETDB1).	Area under the receiver operating characteristic (ROC) curve (AUC) (0.985) and the F-measure (0.926) was assessed.	While designing network classifiers, some features are removed for MA detection models. The green contrast plane surface is not analyzed during pixel patching.
Morales et al., 2015 [36]	ARIA, STARE, E-OPHTHA and DIAGNOS.	The system has achieved 0.990 (TNR) using SVM, 0.897 (TNR) using random forest, 0.881 (TNR) using adaboost and 0.984 (TNR) using neural networks. Likewise, RGB has influenced 0.990 (TNR) using SVM, 0.897 (TNR) using random forest, 0.881 (TNR) using adaboost and 0.984 (TNR) using neural networks. Overall, the usage of LBP features has achieved TPR (1.00) and TNR (0.990) compared to the features like LCP and wavelet transform.	Segmentation of the lesions has consumed more time. Different aliasing effects have degraded the validation sets.
Wisaeng et al., 2018 [33]	A raw dataset and DIARETDB1 public dataset were used for the experimental study.	The system has achieved sensitivity (98.40%), specificity (98.13%) and accuracy (98.35%).	In some cases, uneven contrast has limited the feature learning process. Some mathematical operations degraded the image quality.
Zhou et al., 2018 [34]	Kaggle and Messidor datasets.	The system has achieved the ROC curve of 0.925 on the kaggle dataset and 0.960 on Messidor. Finally, the detection of DR lesions has been performed with the F-score of 0.924 and the precision of 0.863.	Interpolation of pixels using the nearest neighbour has maximized the loss function. Likewise, gradients have maximized the analysis of the irrelevant regions.

Xu et al., 2018 [37]	Grampian diabetes database.	The system has achieved 94% sensitivity and 93% specificity for the image analysis method.	Pathological risk factors are not identified properly, if their dynamic weight alters. Lowered prediction labels are observed when using the training model.
Wang et al., 2019 [38]	Fundus photography of 1589 images were collected in a DR screening project.	Different graders were analyzed and performance is measured as AG3r (91.2%), AG4r (88.0%), AG5r (85.8%), AG6r (85.0%) and AG7r (81.6%).	Though good annotation capability was achieved, the process of predicting the feature at an earlier phase is quite feasible. The annotation error is high when the prediction of image modalities changes.
T. Shanthi and R.S Sabeenian, 2019 [35]	Messidor database.	The system has obtained classification accuracies of different DR stages, as stage 1 (96.2%), stage 2 (95.6) and stage 3 (96.6%).	The learning process of features training decreased the classification error rate. The complexity of the convolutional layer has depressed the networks.
Yazdanyar et al., [39]	STZ-induced diabetic Mice	A preclinical study was performed to show the effects of intravitreal CD34+ BMSCs. The system has optimized the viability and sustainability of the human cells, even under high concentration of CD34+ cells.	A limited set of in vivo was taken for the clinical study.
Liu et al., 2019 [40]	60,000 images were collected from a raw dataset.	The designed WP-CNN has obtained 94.23% of accuracy, 90.24% sensitivity and 0.9823 AUC. The designed model has met the objectives of redundancy reduction and improved the congestion speed.	Convergence speed while combining different convoluted layers has increased the testing error rate. It is also stated that ResNet 101 and DenseNet 121 have higher redundant features, and SeNet-101 has avoided some affected lesions.

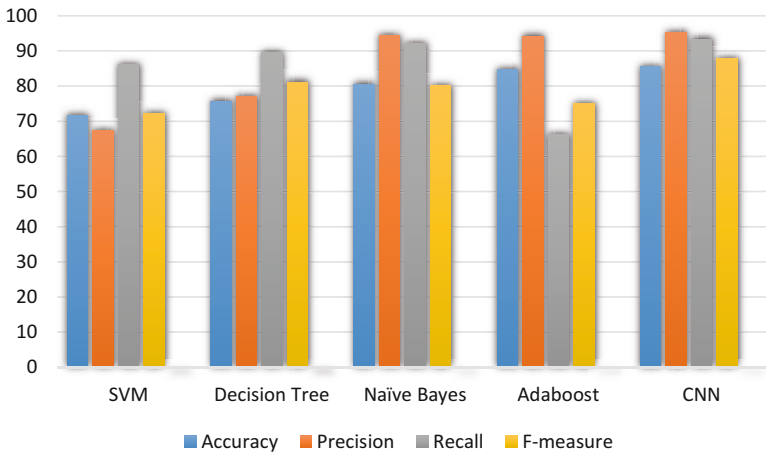
(continued)

Table 1 (continued)

Author name and year	Details of dataset	Performance measures	Limitations
Kamble, 2020 [41]	130 DIARETDB0 and 89 DIARETDB1.	The system has obtained, for 130 DIARETDB0, 71.2% of accuracy, 0.83 of sensitivity and 0.043 of specificity and for 89 DIARETDB1, 89.4% of accuracy, 0.94 of sensitivity and 0.16 of specificity.	Radial basis function is incorporated in ROI extraction. Though there is no weight link in the summation layer, some features were not linked for computational purpose. Histological features of retinal images were also avoided.
Mirshahi et al., 2019 [42].	11 eyes of six patients with no severe proliferative DR (NPDR) data were collected.	The foveal avascular zone (FAZ) is the main variable that helps to easily evaluate the optical coherence tomography angiography (OCTA). When the retinal thickness increases, the density of the vascular zone also increases which shows that the flow surface area remains unchanged.	It is analyzed with a limited set of patients. The metabolic condition of the patients was not discussed.
Butt et al., 2019 [43]	EyePACS datasets were composed of 35,126 images.	The designed multi-channel CNNs have achieved 97.08% accuracy.	High resolution images were omitted. Lesions with larger area were also omitted. Lack of supporting multi-class problems.
Jebaseeli et al., 2019 [44]	STARE, DRIVE, HRF, REVIEW and DRIONS fundus image datasets were taken for the experimental study.	The system has obtained the segmentation accuracy of 99.49%, 80.61% of sensitivity and 99.54% of specificity.	Though the system has reduced the computational time for all datasets, the connectivity among different lesions is not ensured, i.e. non-vessel pixels were taken for the study to increase the convergence rate.

Table 2 Performance analysis, comparison with the existing methodologies

Method	Accuracy	Precision	Recall	F-measure
SVM [25]	71.62	67.3	86.03	72.1801
Decision Tree [27]	75.66	77	89.49	81.1452
Naïve Bayes [35]	80.53	94.31	92.16	80.0425
Adaboost [41]	84.69	94.18	66.12	75.0446
CNN [42]	85.48	95.16	93.33	87.71

**Fig. 3** Graphical representation of quality parameters obtained using various classification methods

The F-measure and the mean square error also relatively provide 87.71% and 0.105%, respectively. From this analysis, it is clear that both Naïve Bayes and Adaboost outperform the other classifiers, and also the CNN ensemble classifier provides better performance than the other classifiers. The graphical representation is presented in Fig. 3.

6 Conclusion and Future Scope

DR is the prime reason behind the vision loss among the working-age population. Medical practitioners have identified that the changes in clinical, geometrical and haemodynamic features are the main causes of DR. The distortions in features like blood vessel area, exudates, microaneurysm, haemorrhages and neovascularization, etc. are the main symptoms of DR. This paper is a review of techniques of low, middle and high-level vision for detecting and classifying DR. Several recent works are collected and their techniques are reviewed. Simultaneously, numerical comparative analysis of their suggested techniques is also discussed. It is observed

from the review that further clinical research on detecting/predicting DR needs to be performed. Feature extraction and classification are an integral part of the detection algorithms for earlier detection of DR.

As future work, we are planning to design an efficient recommendation framework from the DR detection process. Though a variant study has been conducted to detect DR by statistical features of abnormal blood vessels, the concept of neovascularization methodology is least focussed while observing abnormal blood vessels. Since it includes the analysis of high-risk statistical features, it will be a challenging and innovative task for researchers.

References

1. A. Saéed, A. R. S. Abdulaziz, A. S. Mohammad. Effective optic disc detection method based on swarm intelligence techniques and novel pre-processing steps. *Applied Soft Computing*, 2016, 49, pp.146–63.
2. S. B. Akhade, V. U. Deshmukh, S. B. Deosarkar. Automatic optic disc detection in digital fundus images using image processing techniques. *International Conference on Information Communication and Embedded Systems (ICICES2014)*, Feb. 2014, pp.1–5.
3. M. U. Akram, K. Shehzad, A. K. Shoab. Identification and classification of microaneurysms for early detection of diabetic retinopathy. *Pattern Recognition* 2013 January;46(1):107–16.
4. A. Sharib, S. Desiré, M. A. Kedir, G. Luca, C. Edward, P. K. Thomas, M. Fabrice. Statistical atlas-based exudate segmentation. *Computerized Medical Imaging and Graphics* 2013, 37(5), pp. 358–68.
5. A. Mohammad, A. S. Abdulaziz, A. Saéd. Optic disc detection in retinal fundus images using gravitational law-based edge detection. *Medical Biology Eng. Comput.*2017, 55(6): pp. 935–948
6. A. Javeria, S. Muhammad, Y. Mussarat, A. Hussam, F. S. Lawrence. A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions. *J Comput. Sci.* 2017;19:153–64.
7. A. Shahab, S. Haldun. A new supervised retinal vessel segmentation method based on robust hybrid features. *Biomed Signal Process Control*, 2016, pp: 1–12.
8. A. Basit, F. M. Moazam. Optic disc detection and boundary extraction in retinal images. *Applied Optics*, 2015, 54(11), pp.3440–3447.
9. B. Renátó, T. János, H. András. A review on automatic analysis techniques for color fundus photographs. *Computer Structure Biotechnology Journal*, 2016, 14, pp. 371–384
10. B. Sangita. Automatic segmentation of optic disk in retinal images. *Biomed Signal Process Control*, 2017, 31, pp. 483–498.
11. F. Calivá, G. Leontidis, P. Chudzik, A. Hunter, L. Antiga, B. Al-Diri. Hemodynamics in the retinal vasculature during the progression of diabetic retinopathy. *Journal for Model Ophthalmology*, 2017, 1(4), pp. 6–15.
12. D. Baisheng, X. Wu, B. Wei. Optic disc segmentation based on variational model with multiple energies. *Pattern Recognition*, 2017, 64, pp. 226–35.
13. D. s Jyotiprava, B. Nilamani. A thresholding-based technique to extract retinal blood vessels from fundus images. *Future Computer Informatics Journal*, 2017, 2(2), pp. 103–109.
14. D. N. Sekhar, D. H. Sekhar, D. Mallika, M. Saurajeet. An effective approach: Image quality enhancement for microaneurysms detection of non-dilated retinal fundus image. *Procedia Technology*, 2013, 10, pp. 731–737.
15. E. Decenciére, et al., Machine learning and image processing methods for teleophthalmology. *IRBM* 2013, 34(2):196–203.

16. M. M. Fraz, J. Waqas, Z. Saqib, M. H. Mian, A. B. Sarah. Multiscale segmentation of exudates in retinal images using contextual cues and ensemble classification. *Biomed Signal Process Control* 2017, 35, pp.50–62.
17. R. G. Ramani, B. Lakshmi. Retinal blood vessel segmentation employing image processing and data mining techniques for computerized retinal image analysis. *Biocybernet and Biomedical Engineering*, 2016, 36(1), pp.102–118
18. O. J. Ignacio, P. Elena, D. F. Mariana, B. B. Matthew. An ensemble deep learning-based approach for red lesion detection in fundus images. *Computer Methods and Programs in Biomedicine*, 2018. , 153, pp. 115–127
19. V. Roberto, S. A. Gildardo, E. F. M. Luis, S. Humberto, G. Elizabeth. Retinal vessel extraction using lattice neural networks with dendritic processing. *Computer methods and programs in biomedicine*, 2015, 58, pp. 20–30.
20. N. Salamat, M. M. S. Missen and A. Rashid. Diabetic retinopathy techniques in retinal images: A review, *Artificial Intelligence In Medicine*, 2019, 97. pp. 168–188
21. X. Li, X. Hu, L. Yu, L. Zhu. CANet: Cross-disease Attention Network for Joint Diabetic Retinopathy and Diabetic Macular Edema Grading, *IEEE Transactions on Medical Imaging*, 2019, 7, pp.150530–150539.
22. S. Qummar, F. G. Khan, W. Jaddon. A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection, *IEEE access*, 2019, 7, pp. 150530–150539.
23. X. Zeng, H. Chen, and W. Ye. Automated Diabetic Retinopathy Detection Based on Binocular Siamese-like Convolutional Neural Network, *IEEE access*, 2019, pp. 30744–30753.
24. P. Costa, A. Galdran, A. Smailagic, and A. Campilho. A Weakly-Supervised Framework for Interpretable Diabetic Retinopathy Detection on Retinal Images, *IEEE access*, 2018, 6, pp. 18747–18758.
25. W. Zhou, W. du, Y. Yi. Automatic Detection of Exudates in Digital Color Fundus Images Using Superpixel Multi-Feature Classification, *IEEE access*, 2017, 5, pp. 17077–17088.
26. M. A. Aljarrah and H. Shatnawi. Non-proliferative diabetic retinopathy symptoms detection and classification using neural network, *Journal of Medical Engineering & Technology*, 2017, 41(6), pp. 498–505.
27. B. Dashtbozorg, J. Zhang, F. Huang, and B. M. terHaarRomeny. Retinal Microaneurysms Detection using Local Convergence Index Features, *IEEE Transactions on Image Processing*, 2018, 27(7), pp.3300–3315.
28. K. M. Adal, P. G. van Eten, J. P. Martinez, K. W. Rouwen, K. A. Vermeer. An Automated System for the Detection and Classification of Retinal Changes Due to Red Lesions in Longitudinal Fundus Images, *IEEE Transactions on Biomedical Engineering*, 2017, pp. 1382–1390.
29. L. Dai, R. Fang, H. Li, X. Hou, B. Sheng, Q. Wu, W. Jia. Clinical Report Guided Retinal Microaneurysm Detection with Multi-Sieving Deep Learning, *IEEE Transactions on Medical Imaging*, 2018, 37(5), pp. 1149–1161.
30. Y. Zhao, Y. Zheng, Y. Liu, J. Yang, Y. Zhao, D. Chen and Y. Wang. Intensity and Compactness Enabled Saliency Estimation for Leakage Detection in Diabetic and Malarial Retinopathy, *IEEE Transactions on Medical Imaging*, 2017, 36(1), pp.51–63.
31. S. Wang, H. L. Tang, L. I. Al-turk, Y. Hu, S. Sanei. Localising Microaneurysms in Fundus Images Through Singular Spectrum Analysis, *IEEE Transactions on Biomedical Engineering*, 2017, 64(5), pp. 990–1002.
32. W. Cao, J. Shan, L. Li. Microaneurysm Detection Using Principal Component Analysis and Machine Learning Methods, *IEEE Transactions on NanoBioscience*, 2018, 17(3), pp. 191–198.
33. K. Wisaeng and W. Ngiamviboo. Exudates Detection Using Morphology Mean Shift Algorithm in Retinal Images, 2019, pp. 11946–11958.
34. L. Zhou, Y. Zhao, J. Yang, Q. Yu, X. Xu. Deep multiple instance learning for automatic detection of diabetic retinopathy in retinal images, *IET journals on image processing*, 2018, 12(4), pp. 563–571.

35. T. Shanthy & R.S. Sabeenian. Modified Alexnet architecture for classification of diabetic retinopathy images, *Computers and Electrical Engineering*, 2019, 76, pp. 56–64.
36. S. Morales, K. Engan, V. Naranjo and A. Colomer. Retinal Disease Screening through Local Binary Patterns, *IEEE journals of biomedical and health informatics*, 2015, pp.184–192.
37. J. Xu, X. Zhang, H. Chen, J. Li, J. Zhang, L. Shao and G. Wang. Automatic Analysis of Microaneurysms Turnover to Diagnose the Progression of Diabetic Retinopathy, *IEEE access*, 2018, 6, pp. 9632–9642.
38. J. Wang, Yujingbai and B.Xia, Feasibility of Diagnosing Both Severity and Features of Diabetic Retinopathy in Fundus Photography, *IEEE access*, 2019, pp. 102589–102597.
39. A. Yazdanyar, P. Zhang, C. Dolf and S. S. Park, Effects of intravitreal injection of human CD34 bone marrow stem cells in a murine model of diabetic retinopathy, *Experimental Eye Research*, 90, 2019.
40. Y. P. Liu, Z. Li, C. Xu and J. Li, Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network, *Artificial Intelligence In Medicine*, 99, 2019.
41. V. V. Kamble and R. D. Kokate. Automated diabetic retinopathy detection using radial basis function, *International Conference on Computational Intelligence and Data Science*, 167,2020.
42. A. Mirshahi, F. Ghassemi, K. Fadakar, H. R. Esfahani. Effects of panretinal photocoagulation on retinal vasculature and foveal avascular zone in diabetic retinopathy using optical coherence tomography angiography: A pilot study, *Journal of current ophthalmology*, 31, 2019.
43. M. M. Butt, G. Latif, D. N. F A.Iskander and A. H. Khan, Multi-channel Convolutions Neural Network Based Diabetic Retinopathy Detection from Fundus Images, *International Learning & Technology Conference 2019*, 163, 2019.
44. T. Jemima Jebaseeli, C. Anand Deva Durai and J. Dinesh Peter, Retinal blood vessel segmentation from diabetic retinopathy images using tandem PCNN model and deep learning based SVM, *International journal for light and electron optics*, 199, 2019.

Applications of Artificial Intelligence in Medical Images Analysis



Pushpanjali Gupta and Prasan Kumar Sahoo

1 Introduction

The visual representation of a human perception, which can be depicted through two-dimensional (2D) or three-dimensional (3D) display is referred to as an image. Image could be a photograph captured by camera or a hologram created using lenses. The image can be captured using different devices such as microscopes, telescopes, cameras, lenses, mirrors, etc. Based on the capturing device, the image could be static or moving, where a static image is obtained from a single frame. In contrast, the moving images are obtained from multiple frames, and therefore are called video. However, when an image is stored and handled using digital devices such as computer, the image termed as “digital image” is organized as a finite 2D array of picture elements called pixels. Each pixel of a digital image represents a number or set of numbers, describing the gray level intensity or color intensity, where the row and column of 2D array correspond to the vertical and horizontal spatial location of the pixels in the image, respectively [1]. Digital images have several characteristics such as the type of image that could be black and white, where the pixels values are either 0 or 1, determined based on the illumination of light on the pixel of image. Another type of image is the colors image consisting of

P. Gupta

Department of Computer Science and Information Engineering, Chang Gung University,
Guishan, Taiwan

e-mail: D0521006@cgu.edu.tw

P. K. Sahoo (✉)

Department of Computer Science and Information Engineering, Chang Gung University,
Guishan, Taiwan

Department of Neurology, Chang Gung Memorial Hospital, Linkou, Taiwan

e-mail: pksahoo@mail.cgu.edu.tw

three colors, red, green, and blue (RGB), used in computer monitors and scanners or image consisting four colors cyan, magenta, yellow, and black (CMYK), which are generally associated with color printers.

1.1 Medical Images

The digital images can also be obtained from non-optical sources such as X-ray, ultrasound, electromagnetic radio wave, where instead of light, the intensity of X-rays or sound or radio waves are recorded. The conveniences provided through storage, instantaneous flow, and digital images transfer have led to the wide use of digital imaging in the medical field. The digital medical images can be obtained using several diagnostic imaging tools such as X-rays machines, ultrasound machines, Computed tomography (CT) machines, Magnetic Resonance Imaging (MRI) machines, Positron Emission Tomography (PET) machines, etc. The medical images obtained using the diagnostics tools reveal the internal part of the human body to the medical practitioner for diagnosis of diseases, examining of injury and deciding the treatment procedures [2]. The advancements of health care system and increase in the availability of medical imaging equipment have led to the global increase in the quantity and quality of the medical images. However, the medical images obtained are highly unstructured data, which makes it difficult for inexperienced medical practitioner to derive value out of such unstructured data. Consequently, the use of artificial intelligence (AI) in diagnostic medical imaging is extensively suggested [3], where AI can be applied to denoise the raw digital data produced during the scan and automatically recognize complex patterns in imaging data to provide physicians with insights on patients' medical needs.

1.2 Artificial Intelligence

As soon as there was successful progress made towards the possibility of loading and scanning the medical images into the computer, there have been many researches carried out for digital image processing consisting of edge and line detector filters, and region growing models from 1970s to the 1990s. Further, the rule-based systems were also modeled to solve particular tasks. However, in case of complex objects such as in medical field, it is complicated job to build a rule-based model for such complex objects with a large number of parameters. As a result, the uses of machine learning and deep learning in computer-aided diagnosis have gained importance in determining the objects of interest in medical images, such as organs and lesions, which are mostly irregular objects. There are various ML algorithms which can be used for classifying or clustering the objects of interest in medical images. In traditional computer vision approach, using standard feature input such as intensity the lesions can be separated from the organ. However, in ML

model such as support vector machine (SVM), multilayer perceptrons (MLP) and random forests, the manually extracted features of lesions are fed to the ML models for training the model to differentiate the lesions optimally. Nonetheless, feature-based ML requires feature engineering, which is a tedious and error prone job. Besides, the feature extraction is a challenging task, as the manual procedure may not have the distinguishing power that is sufficient for classifying objects of interest [4, 5]. The end-to-end ML approach termed as deep learning, created by Geoffrey Hinton in 2007, can automatically learn high-level representations of objects from large numbers of data instead of using a set of handcrafted features. The deep learning was not recognized widely, until the introduction of convolutional neural network (CNN), a deep learning based approach, which won victory in ImageNet, the best-known computer vision competition. In short, the CNN, an approach of deep learning takes the image as input and gives the output in the categories of classes such as lesion and no lesion in an image. This form of learning is made possible with multiple nonlinear layers used to acquire a high-level representation of image without feature engineering [6, 7].

Rest of this chapter is organized as follows. In Sect. 2, various medical image modalities used for diagnosing and monitoring diseases are discussed. Various image processing methods used for making the medical images suitable for AI-based analysis are elucidated in the Sect. 3. In Sect. 4, surveys of related works used for performing AI-based analysis, especially classification on medical images of different modalities are discussed. The concluding remarks are given in Sect. 5.

2 Medical Imaging

Medical imaging, also called diagnostic imaging refers to several technologies used for producing images of internal structure of the human body for facilitating accurate diagnosis, intervention, prognosis, assessment of injury, and function of some organs or tissues. Efficient decision making crucially depends on correct diagnosis when disease diagnosis or prevention, as well as curative and palliative care are considered. Although the clinical judgment of practitioner may be sufficient prior to treatment of some medical conditions, the use of medical imaging confirms the clinical judgment. Furthermore, the medical imaging assists in the correct assessment of diseases for providing proper treatment and follow-up strategies. In clinical context, medical imaging is roughly equivalent to radiology which uses radiation to diagnose and treat diseases. Nonetheless, other techniques such as soundwaves or radio waves also can be used to view tissues. Moreover, other medical imaging techniques also include endoscopy and colonoscopy, where a flexible instrument is equipped with a camera for obtaining image. Although, in recent years images of removed tissues or organs are also digitalized for medical study, such procedures are still considered pathology instead of medical imaging. However, in this chapter the digital pathology is considered as one of the medical imaging modalities.

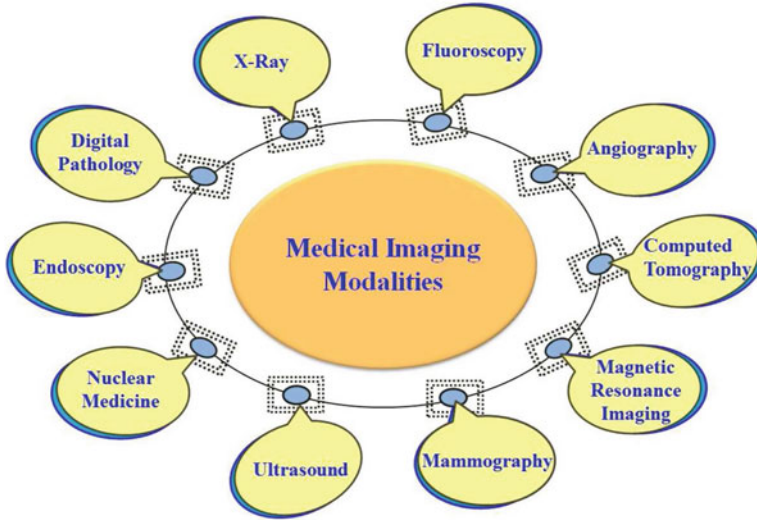


Fig. 1 Various modalities in medical imaging

2.1 *Imaging Modalities*

The imaging modalities refer to the different types of medical imaging techniques, as shown in Fig. 1, those utilize certain physical mechanisms such as sound, light, or electromagnetic wave to detect patient's internal signals that reflect either the anatomical structures or physiological events. Due to the numerous varieties of disease and abnormalities affecting all regions of the human body, it is scientifically impossible to use a single imaging modality for providing the uniquely desired understanding and/or discrimination of the disease type or abnormality. As a result, the different imaging modalities instead of being considered as substitute, work as complementary, providing a powerful and synergistic armamentarium of clinical diagnostic, biomedical, and therapeutic research capabilities. Although the different imaging modalities have significant disparity in scale and/or characteristics features, each of them has the potential to significantly advance the practice of medicine.

2.1.1 **Conventional X-Ray**

X-ray is the oldest and most commonly used imaging modality, as shown in Fig. 2a, which uses ionizing radiation to visualize patients' internal structures by sending beams of x-rays through the body [8]. Based on the density of tissues the beams are absorbed at different levels. Considering the anatomical locations of different parts of body, the x-ray can be used for analyzing the abnormalities in skeletal systems, lungs, teeth, digestive system or any ingested substance. However, conventional

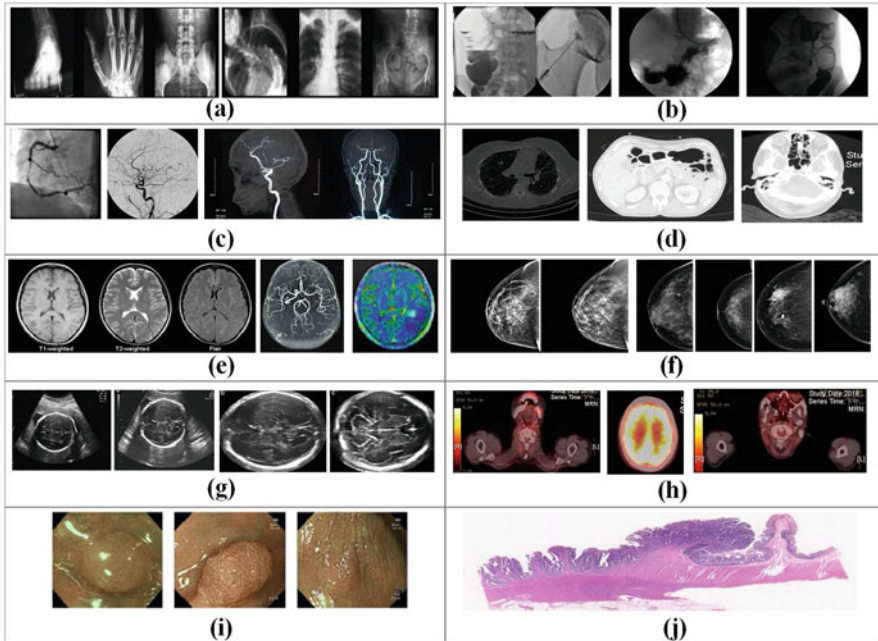


Fig. 2 Examples of various modalities in medical imaging, (a) X-ray [9], (b) Fluoroscopy [10, 11], (c) Angiography [12–14], (d) CT, (e) MRI [15, 16] (f) Mammography [17, 18], (g) Ultrasound, (h) Nuclear medicine, (i) Endoscopy, and (j) Digital pathology

x-ray produces static images. In order to have moving images, another form of imaging modality, namely fluoroscopy can be used.

2.1.2 Fluoroscopy

The fluoroscopy, as shown in Fig. 2b, uses x-rays at a lower dose, to have a real-time visualization of body structures. During the procedure, contrast media such as iodine, barium, and air are used to view movement of tissues, or to guide a medical intervention such as pacemaker insertion or joint replacement/repair. The common clinical applications of fluoroscopy are Barium studies, where barium meal, barium swallow and barium enema are used for evaluation of the gastrointestinal tract. Moreover, other applications include the evaluation of fistulae, and reduction of fractures under image guidance. In addition, the fluoroscopy is also used to monitor the throat during the process of food swallowing.

2.1.3 Angiography

The angiography is used for visualizing the inside or lumen of blood vessels, especially the veins, arteries, and the heart chambers. In this imaging modality, contrast media is injected into the blood vessels for the study, and x-rays are used for visualization of obstruction. The common clinical applications of angiography are diagnosis of aneurysms, particularly the intracranial aneurysm, Fig. 2c, diagnosis of obstructive vascular disease, diagnosis of bleeding vessels, diagnosis of arteriovenous malformations, image guided interventional procedures, and assessment of the vascularity of malignant tumors. Conventionally, the angiography was performed employing simply angiography or digital subtraction angiography (DSA). With the advancement of technology, current procedure of angiography replaces conventional x-rays with Computed Tomography (CT) scan, Magnetic Resonance (MR) Imaging scan, where the angiography is clinically termed as CT angiography (CTA) and MR angiography (MRA), respectively. In the case of both CTA and MRA, the images can be reconstructed in 3D to have better visualization of vessels and accompanying pathology can be viewed from different angles.

2.1.4 Computed Tomography (CT)

The CT scan is a noninvasive diagnostic procedure which uses multiple x-ray images captured from different angles and reconstructed for the creation of 2D and 3D images on the film. The images produced during this procedure are referred to as non-contrast CT scan (NCCT) or C-images. The elements that make up the image are displayed as a 2D pixel, where each pixel carries a value for density or attenuation, represented by a Hounsfield Unit (HU). In addition to being noninvasive procedure, CT scan can also be used in scenario where contrast media can also be injected during a CT study to distinguish structures of similar density in the body. The contrast media contains iodine, a substance which can block x-rays thus allowing the proper visualization of tissue of interest. The images obtained from contrast CT scan are termed as CCT or C+ images. The most common clinical applications of CT imaging are in brain for both with and without contrast, head and neck CT, chest/mediastinum CT, abdominal and pelvic CT, Fig. 2d. In addition the CT is preferred for urography, colonography, angiography for determination of complications if any. As discussed earlier, when CT scan is preferred for performing angiography, it is referred to as CTA. On the other hand, when CT scan is used for examining the blood flow in the blood vessels, it is called CT perfusion (CTP), which is functional imaging technique. The CTP is primarily used for determining the blood flow volume in heart and brain.

2.1.5 Magnetic Resonance Imaging (MRI)

The magnetic resonance imaging technique in radiology is used to visualize the detailed internal structure of human body using magnetic radiation. Based on the dominant influence on the appearance of tissues, the MRI are divided into different sequences, namely T1 weighted, T2 weighted, diffusion weighted imaging (DWI), fluid attenuated inversion recovery (FLAIR), and Apparent Diffusion Coefficient (ADC); those are ordered combination of radiofrequency pulses and gradient pulses designed to obtain the data for the formation of an image. In order to have a detailed study of smaller organs or vessels, the MRI technique may be combined with intravenous contrast media injection where gadolinium based contrast agents are used in the MR image. The common applications of MRI include brain MRI with diffusion studies, Fig. 2e, spinal MRI, neck MRI, cardiac MRI, chest/mediastinal MRI. The abdominal MRI is used for assessment of liver, spleen, kidneys, and extremities for joints, muscles, and bone disorders. Although MRI is safe for patients, injuries or death may be caused to patients with metallic implants, as the foreign metallic objects may cause injury through projectile motion into the magnet. When MRI scan is used for determining the health of blood vessels, as discussed in angiography, it can be referred as MRA. Similarly, when it is required to determine the blood flow at capillary level in tissue, using functional imaging method, perfusion studies can be done using MRI, thereby referring to the imaging method as magnetic resonance perfusion (MRP). Like CTP, MRP is also primarily used for study of perfusion in brain tissues.

2.1.6 Mammography

The mammography uses low energy x-rays for imaging of breast tissue specifically. Using mammography in practice, standardized views of the breasts can be obtained for the assessment of breast lesions along with detection of early breast cancer. During the procedure, each breast is compressed against the film and examined separately to obtain maximum visualization of calcification or masses. The common clinical applications of mammography include screening mammography to detect early cancer in asymptomatic women, diagnostic mammography to obtain the image of breast for diagnosis of a previously identified suspicious breast lesion, Fig. 2f, surveillance mammography to assess recurrence of malignancy in women with known breast cancer, and needle localization to obtain tissue samples from breast masses that appear suspicious on screening or diagnostic mammography, and tumor marking for surgery.

2.1.7 Ultrasound

This imaging modality utilizes high-frequency sound waves instead of x-rays for obtaining cross-sectional images of the body. The ultrasound could be conventional

or Doppler. The Doppler ultrasound uses Doppler shift phenomenon for vascular studies. Besides, the ultrasound is a cheap procedure, which does not require ionizing radiation. However, the outcome of this easy and safe procedure is operator dependent. The most common clinical applications are abdominal ultrasound which is performed to visualize the pathology and anatomy of liver, gallbladder, spleen, kidneys, lymph nodes, retroperitoneum, and abdominal structures. The pelvic ultrasound is carried out to assess the urinary bladder, reproductive organs, prostate, vascular structures, lymph nodes, and adnexal masses. In case of cardiovascular ultrasound including echocardiography, the assessment of heart, and peripheral vascular structures is made, whereas the intraluminal ultrasound assesses the gastrointestinal tract and the blood vessels. The transfontanelle ultrasound helps visualize the intracranial structures before the closure of the fontanelles in pediatric patients. The obstetric ultrasound is primarily used to examine the fetus and related structures in pregnant women as shown in Fig. 2g.

2.1.8 Nuclear Medicine

This imaging modality involves the inhalation or injection of radioactive tracers to visualize the various organs. The radioactive tracer or radiopharmaceutical tracers is formed by the addition of a radioactive isotope with a pharmaceutical specific to the part of the body being examined. The image obtained in this technique is taken using a gamma camera that captures the gamma radiation emitted by the radioactive tracers injected into a patient body. In the gamma camera, there is radiation sensitive crystal which detects the distribution of the tracer within the patient's body. All the detected distribution of tracers are collected and converted to a digital format to produce 2D or 3D images on the monitor. This form of imaging is referred to as positron emission tomography (PET) scan, which mainly focuses on disease on cellular level. In recent years, the PET scan is combined with CT (PET-CT) or MRI (PET-MRI) to obtain macro and micro level detailed study of the organ to be studied. The common clinical application of nuclear medicine includes bone scan to assess metabolic activity of the bones, specifically for oncology staging, arthritis, and fractures, as shown in Fig. 2h. The myocardial perfusion scan application is used to compare the blood flow to the myocardium at exercise and rest allowing the differentiation of infarction and ischemia. In the renal scan, the perfusion and drainage of the kidneys are determined to calculate the renal function. In lung scan, the PET allows comparison of ventilation and perfusion of the lungs to diagnose pulmonary embolism; and in thyroid scan the assessment of appearance and functionality of thyroid gland can be made.

2.1.9 Endoscopy

The endoscopy imaging modality involves the insertion of long, thin, flexible tube called endoscope, to view the insides of different organs of human body; for

instance, colon as shown in Fig. 2i. The endoscope comprises a camera and light, which are responsible for capturing the images of insides of organs [19]. During endoscopy the endoscope can be put into the body through mouth, anus, urethra, and sometimes through a small incision made near the organ of interest. In recent years, new endoscopy techniques are included such as virtual endoscopy and capsule endoscopy. In virtual endoscopy, the endoscope is not inserted into the body; instead CT scans are obtained for thin segment of the body and images are reconstructed using computer to obtain complete view of insides of organ. However, in case of capsule endoscopy, a small vitamin-capsule sized camera is swallowed by patient. The camera takes pictures of insides of esophagus, stomach, and intestine. Later, after 8 hours the camera is excreted and doctors review the images.

2.1.10 Digital Pathology

In recent years, with the advancement of cost-effective whole slide scanners, tissue histopathology slides are now digitized and stored in digital image form [20]. The availability of sophisticated, high-performance imaging and analysis platform can rapidly replace traditional paradigm of a pathologist-microscope with pathologist-large flat screen panel to view and rapidly analyze digitized tissue sections. As a result, the digital pathology uses computer workstations to view digital whole slide images (WSIs), as shown in Fig. 2j, obtained from scanner such as NanoZoomer digital slide scanner, Aperio digital pathology slide scanner, Glissando digital pathology scanner, etc. These scanners scan the glass microscope slides of tissues with high resolution and produce the image of size greater than 1 Gigabytes. Along with the image file, the associated metadata, including the specimen ID, specimen type, patient information, and the relevant staining information are also obtained in digital format. Digital pathology and image analysis can be used to have greater diagnostic accuracy, reproducibility and standardization of inclusion criteria and prediction of outcomes. Currently, there is no standardization in the file format for digital pathology. However, the stored digital pathology images can be retrieved using the different image management systems for viewing, annotating, and analyzing images. The common applications of digital pathology are the clinical laboratory areas including histology, cell biology, medical biology, hematology, and most importantly oncology. However, digital pathology has not yet received Food and Drug Administration (FDA) approval in broader areas. The transition from pathologist-microscope to pathologist-flat screen panel is a major step towards the growth and development of pathology as a scientific discipline.

2.2 Image Formats

The images obtained using different imaging modalities, including conventional radiography, ultrasound, CT, MRI, fluoroscopy, angiography, mammography,

endoscopy, and nuclear medicine are stored in PACS (Picture Archiving and Communication System). The PACS is a medical technology used primarily to store significantly largest of image files in the routine course of diagnosing and treating the patients. In PACS, the images can be stored in both 2D and 3D form with slice thickness 1mm, 3mm, or 5mm, using a standard protocol named DICOM (Digital Imaging and Communications in Medicine) which is used by medical professionals for the transmission and management of medical images and related data. Nevertheless, in case of the digital pathology there is no standard for storing the images obtained by digital pathology slides scanner. As a result, the format of images obtained is proprietary file format. For instance, the digital pathology image of WSI obtained using a Nano Zoomer Digital Pathology Scanner is in the proprietary file format, NanoZoomer Digital Pathology (NDP) image, which is vendor dependent; therefore requires special software to view the image [21].

When obtaining image using different modalities for different parts of human body, the human is exposed to different dose of radiation, which could have harmful effect in the body. In order to measure the impact of intensity of radioactive emissions on the human tissues and health, sievert (Sv) and millisievert (mSv) are used. The Sv and mSv are used for estimating “equivalent dose” by comparing imaging procedures, taking into consideration the biological effect of radiation, which varies with the type of radiation used and vulnerability of the exposed body tissue. Similarly, in case of MRI, instead of using radiation, magnets are used for acquiring image. The strength of magnet used in MRI is represented as 1.5T or 3T, where T stands for Tesla, the unit of measurement for determining the strength of magnet. Currently, in hospitals the MRI scanner uses magnetic strength of 1.5T or 3T. In October 2017, MRI scanner with magnetic strength 7T was cleared for clinical use, and in December 2017, MRI scanner with magnetic strength 10.5T was approved for clinical use in United States and Europe. Nonetheless, it is not widely available [22].

3 Image Processing Methods

When the images are obtained in different modalities such as CT and MRI, they may contain artifacts in the form of noise, which must be removed to improve the quality of the image. In addition, the image data obtained are highly unstructured. Therefore, different image processing methods such as smoothing [23], sharpening [24], and morphological techniques can be applied on the images as discussed below.

3.1 Smoothing

The image represented in the form of matrix of pixels may contain random variation in the pixel content called noise, which might be caused during acquisition, transmission, and/or digitization process. Although, all the noise cannot be removed altogether, the noise can be reduced using the *smoothing* process. One way of smoothing the image for noise reduction is to normalize the neighboring pixels value and *averaging* the pixel by converting the image to grayscale. In case of averaging, each pixel is replaced by the average of its neighboring pixels in the considered sliding window. Another way of replacing the pixels value in the sliding window is to replace the pixel value with median of its neighboring pixels value.

As given below, there are few predefined filters that can be applied for achieving the image smoothing.

$$\begin{pmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{pmatrix}, \begin{pmatrix} \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \end{pmatrix}, \begin{pmatrix} 0 & \frac{1}{8} & 0 \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ 0 & \frac{1}{8} & 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \end{pmatrix}$$

3.2 Sharpening

When it is required to have enhanced edges, image sharpening technique is applied using high-pass filters. The sharpening focuses on removing blur and de-hazing an image while emphasizing on the texture of the image. While sharpening an image, the resolution and acutance of an image is considered. The resolution represents the size of an image in pixels. The greater the number of pixels, the higher the resolution and sharpness of the image. On the other hand, acutance refers to as the subjective measurement for contrast of an edge. Edges that have more contrast appear to be more defined to the human eye. Therefore, sharpening defines the details of an image, especially the small details. It is often applied to overcome the blurring effect introduced by capturing device during image acquisition to increase the legibility, and focus on certain areas.

There are few predefined high-pass filters that can be used for producing the sharpening effect as discussed below.

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & -2 & 1 \\ -2 & 1 & -2 \\ 1 & -2 & 1 \end{pmatrix}$$

3.3 Morphological Transformations

The morphological transformations [25] are some simple image processing techniques mainly applied in images. In image processing, the morphological operations are used for extracting information about shapes and structures of objects in the image. The morphological operation needs two inputs, firstly the input image and secondly the kernel or structuring element based on which the nature of operation such as erosion, dilation, opening, closing, etc. is decided. The *structuring element* is a small matrix of pixels, which is positioned to slide over at all possible locations in the image and the corresponding neighboring pixel values are compared. The *size* of the structuring element is specified by the dimension of matrix and the *shape* of the structuring element is specified by the spatial organization of the pixel values. In image processing, some morphological operations check if the structuring element *fits* exactly to the image, while others check if structuring elements *intersects* the neighboring pixels in the image. Some of the morphological transformations are described below.

3.3.1 Erosion

Similar to soil erosion, the erosion [26] erodes away the boundaries of foreground objects. When the kernel slides through the image, all the pixels near the boundaries are discarded depending on the size of the kernel. The erosion is useful for removing the small unwanted noises in the output image and detaching two connected components. When a structuring element Y is used to erode an input image X , the erosion can be denoted as $X \ominus Y$. In order to produce a new image Z , the erosion checks if the structuring element Y *fits* in to the input image X , repeating for pixel coordinates (x,y) . The erosion operation removes the foreground structures that are smaller in size than that of the considered structuring element. Therefore, it helps in reducing the noisy connection between two foreground objects. When the unwanted noisy pixels are removed, the resultant output is the sharpened object in the image. However, the erosion also reduces the size of the region of interest.

3.3.2 Dilation

The dilation [27] acts as opposite of erosion. It increases the size of the foreground object. Generally, the erosion is followed by dilation to increase the area of the foreground object, which was shrunk due to erosion. It is also useful in connecting broken parts of an object in the image. When a structuring element Y is used to dilate an input image X , the dilation can be denoted as $X \oplus Y$. The new image Z is produced when structuring element Y hits the input image X repeating for all coordinates (x,y) . Contrary to the erosion operation, the dilation operation adds a layer of pixels to the outer and inner boundaries of the region of interest, which

results in filling the holes within a region and reducing the gaps between different regions where smaller gaps are filled in.

3.3.3 Opening

The opening procedure involves the erosion followed by dilation in one step [28]. Opening opens up the gap between the regions connected by a thin connectivity of pixels. During opening, any regions that are not eliminated by the erosion operation are restored to their original size by the use of dilation. The opening of an image X by element Y is denoted as $X \circ Y = (X \ominus Y) \oplus Y$. The opening operation is idempotent. Once opening is applied to an image X with a structuring element Y , subsequent opening with the same structuring element has no further effect on the image X . Therefore, $(X \circ Y) \circ Y = X \circ Y$. However, opening smoothes the boundary of the object and eliminates the thin bulges appearing on the boundary of the object. Opening is dual of closing; closing the background pixels with a particular structuring element Y is equivalent to opening the foreground pixels with the same structuring element Y . When opening is applied to an image, all pixels those are belonging to the foreground object and are covering entirely by the structuring element will be preserved, failing to which the foreground object will be eroded away. After the erosion, the new boundaries of the foreground are formed in such a way that the structuring element Y fits inside the image X .

3.3.4 Closing

The reverse of the opening is closing, where dilation is followed by the erosion. The closing of image X with the structuring element Y is denoted as $X \bullet Y = (X \oplus Y) \ominus Y$. The closing is suitable when it is required to close the small holes in the inside of the foreground object along with preservation of the background regions those have a similar shape of structuring element. The most common application of closing is to fill in the small holes in the regions of interest. If only dilation is applied to fill in the holes, all regions of pixels are distorted indiscriminately. In order to reduce such distortion effect, dilation is followed by erosion in the closing operation. The effect of closing can be observed when a structuring element Y slides-over each foreground region of interest. For every pixel point in the image X , it is checked if the structuring element X can touch the background point without being inside of any part of the foreground region. If it is possible, the pixel value of the considered point is set to the background, else it is set to the foreground region of interest. Once the closing is carried out covering all points in an image, application of closing in an image X with the same structuring element Y will have no effect. Therefore, like opening, closing is also idempotent operator $(X \bullet Y) \bullet Y = X \bullet Y$.

4 Medical Image Analyses

There are various applications of artificial intelligence in the field of medical data analysis such as classification, detection, and segmentation. Out of various applications, our scope includes the classification of medical images. The classification focuses on classifying objects of interest within an image into two or more classes. When using traditional image classification, the low-level and mid-level features are extracted from an image and then, a ML based trainable classifier is used for performing label assignment. However, with the advancement of research in machine learning, the high-level features in deep learning classifiers have proven to be superior to handcrafted low-level features. In convolutional neural network (CNN) [29], end-to-end training is performed combining both feature extraction and classification networks.

In 2012, the deep learning based image classification achieved a milestone when AlexNet [30] achieved top-5 error of 15.3%, resulting the top performance in the annual competition of ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). The ILSVRC considers roughly 1000 images in each of 1000 classes from the ImageNet dataset. Overall, the training set consists of around 1.2 million images, validation set consists of 50,000 images, and testing set consists of 150,000 images. ImageNet is a dataset comprising over 15 million categorized high-resolution images belonging to roughly 22,000 categories [31]. Later on, a very deep convolutional network, known as Visual Geometry Group network [32] gained popularity after becoming 1st Runner up in ILSVRC-2014. The top position was achieved by GoogLeNet [33]. Later, the ResNet [34] won 1st position in ILSVRC-2015. All these popular deep learning algorithms have obtained state-of-the-art classification accuracy in ImageNet.

4.1 Radiography Image

The deep learning based image analysis of x-ray images are widely discussed in different literatures as follows. In the work [35], a deep CNN model called Decompose, Transfer and Compose (DeTraC) is developed for the determination of COVID-19 in chest x-ray (CXR) images. The designed model comprised of several phases, where the first phase consisted of the pretrained CNN model, required for extracting deep local features from each input image. This is followed by the class decomposition layer, which is primarily used for simplifying the local structure of the data distribution. In the second phase, the training is performed using a popular optimization method, namely gradient descent. Finally, a class-composition layer is used to refine the resultant classification of the images. The class decomposition and composition components were appended before and after knowledge transformation from an ImageNet pretrained CNN model. The class decomposition layer partitioned all the classes belonging to the image dataset into

k number of independent subclasses. Each of the subclasses was assembled back using the class-composition component for obtaining the resultant classification of given image dataset. In the stage I, a pretrained CNN (AlexNet, VGG19, ResNet, GoogLeNet, and SqueezeNet) model was used for feature extraction for constructing a deep feature space from input CXR images followed by use of principal component analysis for dimension reduction. In the stage II, transfer learning was adapted where the final classification layer of an ImageNet pretrained CNN model was fine-tuned to classify the decomposed classes. In the stage III, the labels associated with decomposed classes were predicted, and final classification were refined using error-correction criteria. The main contribution of DeTraC was the way the algorithm dealt with the most challenging problem of data irregularities, class imbalance. There were 19 cases, comprising both normal and SARS cases used for this experiment, with 80 samples of normal CXR images (4020 * 4892 pixels) and CXR images containing 105 samples of COVID-19 and 11 samples of SARS with 4248 * 3480 pixels. All the experiments were carried in MATLAB [36] 2019a on a 3.7 GHz Intel(R) Core(TM) i3-6100 Duo, Nvidia Corporation and 8 GB RAM. The DeTraC model achieved 98.23% accuracy and AUC 0.96 with VGG19 pretrained model.

The CheXNet [37] is a popular model with 121 layers trained on 112,120 frontal view x-ray images of 30,805 unique patients with 14 diseases; namely atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, pneumothorax, consolidation, edema, emphysema, fibrosis, pleural thickening and hernia. The model achieved F-score of 0.435 as compared to F-score of 0.387, achieved by radiologist. The work used CNN for the identification of vessel regions in angiography images, which is crucial for the early diagnosis of the coronary artery disease. A fixed-size window was used for extracting patches from the input image. The position of central pixel of the patch whether it was located inside or outside the vessel was also indicated in the output. Finally, the trained CNN was used to divide the image in two regions comprising the vessel and the background. In [38], DarkCovidNet was designed for identification of COVID-19. There were total 1125 images used consisting of 500 Pneumonia, 125 COVID-19(+), and 500 No-Findings. The model achieved 98.08% accuracy and 87.02% accuracy, for binary and three-class classification, respectively. The designed deep learning model is built with 17 convolutional layers. For each input image “X” and kernel “K,” the convolution operation in the convolutional layer can be performed using the Eq. 1.

$$X(X * K)(i, j) = \sum_m \sum_n K(m, n)X(i - m, j - n) \quad (1)$$

where * is used to apply the convolution operation. The activation function used after convolution is the Leaky ReLU calculated using Eq. 2.

$$f(x) = \begin{cases} 0.01x & x < 0 \\ x & x \geq 0 \end{cases} \quad (2)$$

When considering MR images, there have been several studies performed for different organs of the human body. For instance, in [39], 3D deep learning convolutional neural network architecture was presented for brain extraction of MR images. The proposed method can be used for both non-enhanced and contrast-enhanced scans. The model used brain tumor dataset and achieved dice coefficient score of 95.02. A different study was made for prostate cancer management considering the MR images [40]. The study aimed to assess the early changes in femoral heads in prostate cancer patients. The T1, T2 and Apparent Diffusion Coefficient (ADC) sequences were used to extract 34 radiomics features. Sixty femoral heads were analyzed. Postradiotherapy, it was observed that there were no changes in features extracted from ADC. However, the features obtained from T1 and T2 had significant changes postradiotherapy with p -value < 0.005 .

In the study made in [41], 446 lesions for hepatocellular carcinoma (HCC) and hepatic hemangioma (HH) were considered from 369 patients. 1029 radiomics features were obtained from T2 weighted imaging and DWI. There were four classifiers used, namely decision tree, random forest, K-nearest neighbors and logistic regression. The best model logistic regression obtained testing AUC of 0.89, sensitivity of 0.822 and specificity of 0.714, which was better than less experienced radiologist, where AUC was 0.702, $p < 0.005$. A combined study was made for handcrafted feature analysis and deep learning analysis considering brain tumor MR images [42]. Using the handcrafted classifier; the model achieved 96.10% accuracy. With only deep learning model the accuracy achieved was 97.8% accuracy. However, when both deep learning and handcrafted feature were combined, the model achieved 99.3% accuracy.

4.2 *Ultrasound Image*

Considering the use of ultrasound image, there have been several works performed in the field of artificial intelligence. One such work focusses on fecal retention assessment, where absence or presence of rectal feces was analyzed in 42 patients [43]. Among 42 patients, 31 patients had the presence of rectal feces, and 11 had no feces in rectal area. considering the presence of feces in rectum, further classification of positive feces cases was also performed. Therefore, three-class classification was performed using deep learning for identifying classes such as absence of feces, hyperechoic area, and strong hyperechoic area in the rectum. The designed model achieved sensitivity of 100% and specificity of 100% in binary classification. However, when three-class classification was performed, the model achieved 85.7% accuracy in identifying strong hyperechoic area and 88.2% accuracy in identifying hyperechoic area.

Another important analysis was performed in [44], where robust pattern classifier was built for diagnosing children with posterior urethral valves (PUV). The developed multiple-instance learning model was designed to distinguish 71 children with unilateral hydronephrosis from 86 children with PUV. Total number of ultrasound

images obtained were 3504 images in sagittal view, and 2558 images in transverse view. The designed multiple-instance deep learning model provided extraction of information features automatically based on kidney ultrasound images. The model achieved AUC of 0.961 ± 0.026 , with sensitivity of 0.873 ± 0.120 , and specificity of 0.986 ± 0.032 .

In order to have deep learning based classification of thyroid nodules using ultrasound images, retrospective study was performed considering 1040 cases which consisted of 1841 benign nodules and 1393 malignant nodules [45]. Although, transverse and longitudinal views were scanned for each patient, only longitudinal view ultrasound images were chosen for further investigation, since the transverse view contained small background for images with thyroid nodules. In addition, the images were taken from same ultrasonic system. The whole dataset was divided in the ratio of 80:20 for training and testing, respectively, using pretrained VGG16 model with fine-tuning. The performance of deep learning based model was compared with radiomics-based methods which used 302 features such as intensity difference, gray scale histogram, Gabor filters, and wave features extracted manually from region of interest, marked by radiologists. The radiomics method used SVM as the classifier. The designed model achieved an accuracy of 74.69%, whereas the radiomics method achieved an accuracy of 66.81%, proving the better performance of deep learning based model.

The advancement of technology has made it possible to study fetal brain in ultrasound images. Considering the use of deep learning for fetal brain analysis, the work [46] proposed the classification of normal and abnormal sonographic images in standard axial planes. The work used dataset collected from affiliated hospital of Sun Yat-Sen University, China. There were 12,780 cases of women who underwent prenatal examination between 18 and 32 weeks of pregnancy with twins or singleton. The dataset contained 12,682 ultrasound images from 10,251 normal cases in standard axial neurosonographic (SAN) planes, obtained following International Society of Ultrasound in Obstetrics & Gynecology guidelines. In similar method 2529 abnormal cases consisting of 3277 images were also obtained with abnormality like neural tube defect, lissencephaly, midline structural anomaly, space-occupying lesion, ventriculomegaly, microcephalus, intracranial hemorrhage, holoprosencephaly or posterior fossa anomaly. In addition, 3D ultrasound dataset were also included consisting of 1922 3D volume dataset from 961 normal cases, and 4843 3D volume dataset from 1051 cases with abnormality. Before classifying the images as abnormal or normal, segmentation was performed to remove distracting areas and find possible candidates for region of interest. The region of interest candidates were classified into normal and abnormal, and the localization of abnormal region was performed using heat maps, and overlay images. The dataset were augmented using random rotation of angle between 0° and 60° , vertical or horizontal flipping for simulating different fetal positions. All training procedures were implemented using Keras with TensorFlow as backend with four Nvidia GeForce GTX 1080Ti graphics processing units. The classification model achieved an average accuracy of 96.3%, with sensitivity of 96.9%, specificity of 95.9% and

AUC was 0.989. The use of AI in prenatal ultrasound can assist sonologists in earlier and better diagnosis of fetal abnormality.

4.3 Endoscopy Image

Considering the use of AI in endoscopy imaging, there have been few works proposed for identification and detection of polyps in colon or rectum using CNNs. For instance, 41 cases of colon endoscopies consisting of 190 colon lesions images collected from February 2015 until October 2016 were used for training CNNs to assist in cT1b diagnosis [47]. The retrospective study used unenhanced colon endoscopy white light images with dimension 520*520 pixels. The considered cases included 14cTis cases with endoscopic resection and 14cT1a and 13cT1b cases with surgical resection. The types of lesions analyzed were protruding, flat, and recessed. Caffe was used as implementation framework and AlexNet was used as the CNN. Augmentation of data in the form of oversampling was performed to avoid impartiality in image numbers. The model achieved sensitivity, specificity, and accuracy of 67.5%, 89.0%, and 81.2%, respectively, with AUC 0.871 in the determination of cT1b. The work tried to minimize the workload of endoscopists by developing quantitative diagnostic support system. Another work aimed to develop a fully automatic algorithm for detection and classification of hyperplastic and adenomatous colorectal polyps. The work used transfer learning application which utilizes features learned from nonmedical dataset with 1.4–2.5 million images using deep CNN. There were 1104 endoscopic non-polyp images taken under both white light and narrow band imaging (NBI) endoscopy.

In addition, 826 NBI endoscopic polyp images consisting of 263 hyperplasia images, and 563 adenoma images were also considered. The proposed method contributed with sensitivity of 87.6%, precision of 87.34%, and accuracy of 85.9% in the identification of polyps those are adenomatous but have been incorrectly judged as hyperplasia and therefore, the automatic algorithm can assist endoscopists in timely resection of polyp at an early stage before the polyps develop into invasive cancer [48]. The proposed work aimed to construct an AI system which can be used for accurate automatic detection and classification of colon polyps using retrospective images obtained during colonoscopy. The work used 16,418 images, where the training was performed using 4752 colon polyps and 4013 images of normal colon and rectum. The performances of model were validated using remaining 7077 colonoscopy images. The model achieved sensitivity of 92% and positive predicted value (PPV) of 86%, wherein case of white light images the sensitivity and PPV were 90% and 83%, respectively, and in case of narrow band images the sensitivity and PPV were 97% and 98%, respectively [49].

4.4 Digital Histopathology Image

When considering digital pathology, authors in [50] used CNN to extract the features automatically and the extracted features were used for classification of breast and colon cancer into benign and malignant tumor. The CNN model consisted of 5 layers architecture similar to LeNet and achieved 99.74% accuracy for binary classification. The works in [51, 52], and [53] focused on deep learning analysis for omitting the feature engineering and performing the classification of colorectal cancer into benign and malignant based on tumor differentiation, classifications of brain tumor and colorectal tissues into normal and abnormal considering 717 patches, using AlexNet architecture. The work in [51] achieved 99.6% accuracy, [52] achieved 97.5% accuracy for classification and the work in [53] achieved 96% accuracy with VGG16 architecture. A different contribution was made in [54], where the authors attempted to predict the 5 years Disease Free Survival in case of patients with CRC. The work used VGG16 for feature extraction and LSTM for predicting the 5 years survival probability. The work achieved an accuracy of only 69% when performing the DFS prediction directly from the image. Recently, in [55], the authors trained CNNs and RNNs on WSI of stomach and colon for performing multiclass classification into three categories, namely adenoma, adenocarcinoma, and non-neoplastic. They achieved AUCs up to 0.99 and 0.97 for gastric adenoma and adenocarcinoma, respectively. While, on the other hand for colonic adenoma and adenocarcinoma, they achieved AUC 0.99 and 0.96, respectively.

5 Conclusions

Applications of artificial intelligence are the most useful use case for medical image analysis in healthcare. AI could be used to decode the complex unstructured image data obtained from different imaging modalities such as CT, MRI, endoscopy, etc. It can extract meaningful information for better decision making with better precision in applications such as cancer detection and pneumonia diagnosis. Since medical imaging technology is advancing abruptly; human cannot keep pace with the advancement. As a result, the use of machine learning can process the information faster than human to provide accurate and contextual information to the experts. It can improve the jobs of the clinicians by enabling them to focus on most complex cases, resulting in acceleration of the diagnostic and treatment process as a whole. Consequently, medical image analysis can move away the trend from reactive healthcare and shift to support the informed and predictive digital healthcare. In this chapter, the use of AI in medical image analysis is discussed when different imaging modalities such as CT, X-ray, MRI, and nuclear medicine are considered. We present review of related literatures showing the use of AI considering different modalities such as analysis of CT or endoscopy images to design the prognosis models for the diseases like cancer, tumor, and diabetes.

Acknowledgments This work was supported in part by the Ministry of Science and Technology (MOST), Taiwan, under Grant number 110-2221-E-182-008-MY3.

References

1. Patin F. An introduction to digital image processing. online]: <http://www.programmersheaven.com/articles/patin/ImageProc.pdf>. 2003.
2. Maintz JA, Viergever MA. A survey of medical image registration. *Medical image analysis*. 1998 Mar 1;2(1):1–36.
3. Tang X. The role of artificial intelligence in medical imaging research. *BJR| Open*. 2019 Nov;2(1):20190031.
4. Russell S, Norvig P. Artificial intelligence: a modern approach.
5. Gupta P, Chiang SF, Sahoo PK, Mohapatra SK, You JF, Onthoni DD, Hung HY, Chiang JM, Huang Y, Tsai WS. Prediction of colon cancer stages and survival period with machine learning approach. *Cancers*. 2019 Dec;11(12):2007.
6. Le Cun Y, Bengio Y, Hinton G. Deep learning. *nature*. 2015 May;521(7553):436–44.
7. Onthoni DD, Sheng TW, Sahoo PK, Wang LJ, Gupta P. Deep Learning Assisted Localization of Polycystic Kidney on Contrast-Enhanced CT Images. *Diagnostics*. 2020 Dec;10(12):1113.
8. Inskip PD, Ekblom A, Galanti MR, Grimelius L, Boice Jr JD. Medical diagnostic x rays and thyroid cancer. *JNCI: Journal of the National Cancer Institute*. 1995 Nov 1;87(21):1613–21.
9. Available online: https://www.natgeokids.com/wp-content/uploads/2018/07/drkarl_xray.jpg
10. Available online: <https://health.ucdavis.edu/radiology/mymri/myexam/myxray/FLUOROS COPY>
11. <http://www.southend.nhs.uk/your-services/diagnostic-therapeutic-services/radiology/fluoroscopy>
12. Available online: <http://www.concordcardiology.com.au/tests-treatments/coronary-angiography/>
13. Available online: https://commons.wikimedia.org/wiki/File:Cerebral_Angiogram_Lateral.jpg
14. Available online: <https://www.ucsfhealth.org/medical-tests/magnetic-resonance-angiography>
15. Available online: <https://casemed.case.edu/clerkships/neurology/Web%20Neurorad/t1t2flair brain.jpg>
16. Available online: <https://www.imagilys.com/perfusion-imaging-MRI/>
17. Available online: <https://media.npr.org/assets/img/2014/06/24/mammography>
18. Available online: <https://physicsworld.com/a/ai-rivals-human-radiologists-at-breast-cancer-detection/>
19. Forrest JH, Finlayson ND, Shearman DJ. Endoscopy in gastrointestinal bleeding. *The Lancet*. 1974 Aug 17;304(7877):394–7.
20. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities.
21. Abels E, Pantanowitz L, Aeffner F, Zarella MD, van der Laak J, Bui MM, Vemuri VN, Parwani AV, Gibbs J, Agosto-Arroyo E, Beck AH. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *The Journal of pathology*. 2019 Nov;249(3):286–94.
22. Mettler Jr FA, Huda W, Yoshizumi TT, Mahesh M. Effective doses in radiology and diagnostic nuclear medicine: a catalog. *Radiology*. 2008 Jul;248(1):254–63.
23. Ramponi G. The rational filter for image smoothing. *IEEE Signal Processing Letters*. 1996 Mar;3(3):63–5.
24. Schavemaker JG, Reinders MJ, Gerbrands JJ, Backer E. Image sharpening by morphological filtering. *Pattern Recognition*. 2000 Jun 1;33(6):997–1012.
25. Soille P. Morphological image analysis: principles and applications. Springer Science & Business Media; 2013 Mar 14.

26. Tambe SB, Kulhare D, Nirmal MD, Prajapati G. Image processing (IP) through erosion and dilation methods.
27. Jawas N, Suciati N. Image inpainting using erosion and dilation operation. *International Journal of Advanced Science and Technology*. 2013 Feb;51:127–34.
28. Soille P. Opening and closing. In *Morphological Image Analysis 2004* (pp. 105–137). Springer, Berlin, Heidelberg.
29. O’Shea K, Nash R. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458. 2015 Nov 26.
30. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 2012;25:1097-105.
31. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A largescale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition 2009 Jun 20* (pp. 248–255). IEEE.
32. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014 Sep 4.
33. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition 2015* (pp. 1–9).
34. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition 2016* (pp. 770–778).
35. Abbas A, Abdelsamea MM, Gaber MM. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. arXiv preprint arXiv:2003.13815. 2020 Mar 26.
36. Higham DJ, Higham NJ. *MATLAB guide*. Society for Industrial and Applied Mathematics; 2016 Dec 28.
37. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, Lungren MP. CheXnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225. 2017 Nov 14.
38. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*. 2020 Apr 28:103792.
39. Kleesiek J, Urban G, Hubert A, Schwarz D, Maier-Hein K, Bendszus M, Biller A. Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. *NeuroImage*. 2016 Apr 1;129:460-9.
40. Abdollahi H, Mahdavi SR, Shiri I, Mofid B, Bakhshandeh M, Rahmani K. Magnetic resonance imaging radiomic feature analysis of radiation-induced femoral head changes in prostate cancer radiotherapy. *Journal of cancer research and therapeutics*. 2019 Mar 1;15(8):11.
41. Wu J, Liu A, Cui J, Chen A, Song Q, Xie L. Radiomics-based classification of hepatocellular carcinoma and hepatic haemangioma on precontrast magnetic resonance images. *BMC medical imaging*. 2019 Dec 1;19(1):23.
42. Hasan AM, Jalab HA, Meziane F, Kahtan H, Al-Ahmad AS. Combining deep and handcrafted image features for MRI brain scan classification. *IEEE Access*. 2019 Jun 13;7:79959-67.
43. Matsumoto M, Tsutaoka T, Nakagami G, Tanaka S, Yoshida M, Miura Y, Sugama J, Okada S, Ohta H, Sanada H. Deep learning-based classification of rectal fecal retention and analysis of fecal properties using ultrasound images in older adult patients. *Japan Journal of Nursing Science*. 2020 May 11:e12340.
44. Yin S, Peng Q, Li H, Zhang Z, You X, Fischer K, Furth SL, Fan Y, Tasian GE. Multi-instance deep learning of ultrasound imaging data for pattern classification of congenital abnormalities of the kidney and urinary tract in children. *Urology*. 2020 May 20.
45. Wang Y, Yue W, Li X, Liu S, Guo L, Xu H, Zhang H, Yang G. Comparison Study of Radiomics and Deep Learning-Based Methods for Thyroid Nodules Classification Using Ultrasound Images. *IEEE Access*. 2020 Mar 12;8:52010-7.
46. Xie H, Wang N, He M, Zhang L, Cai H, Xian J, Lin M, Zheng J, Yang Y. Using deep learning algorithms to classify fetal brain ultrasound images as normal or abnormal. *Ultrasound in Obstetrics & Gynecology*. 2020 Jan 7.

47. Ito N, Kawahira H, Nakashima H, Uesato M, Miyauchi H, Matsubara H. Endoscopic diagnostic support system for cT1b colorectal cancer using deep learning. *Oncology*. 2019;96(1):44-50.
48. Zhang R, Zheng Y, Mak TW, Yu R, Wong SH, Lau JY, Poon CC. Automatic de-tection and classification of colorectal polyps by transferring lowlevel CNN features from nonmedical domain. *IEEE journal of biomedical and health informatics*. 2016 Dec 5;21(1):41-7.
49. Ozawa T, Ishihara S, Fujishiro M, Kumagai Y, Shichijo S, Tada T. Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. *Therapeutic Advances in Gastroenterology*. 2020 Mar;13
50. Xu J, Luo X, Wang G, Gilmore H, Madabhushi A. A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing*. 2016 May 26;191:214-23.
51. Chen H, Qi X, Yu L, Dou Q, Qin J, Heng PA. DCAN: Deep contour-aware networks for object instance segmentation from histology images. *Medical image analysis*. 2017 Feb 1;36:135-46.
52. Kainz P, Pfeiffer M, Urschler M. Segmentation and classification of colon glands with deep convolutional neural networks and total variation regularization. *PeerJ*. 2017 Oct 3;5:e3874.
53. Xu Y, Jia Z, Wang LB, Ai Y, Zhang F, Lai M, Eric I, Chang C. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC bioinformatics*. 2017 Dec;18(1):1-7.
54. Ponzio F, Macii E, Ficarra E, Di Cataldo S. Colorectal cancer classification using deep convolutional networks. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies 2018 (Vol. 2, pp. 58-66)*.
55. Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, Walliander M, Lundin M, Haglund C, Lundin J. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Scientific reports*. 2018 Feb 21;8(1):1-1.

Intelligent Image Segmentation Methods Using Deep Convolutional Neural Network



Mekhla Sarkar and Prasan Kumar Sahoo

1 Introduction

The modern era of Artificial Intelligence (AI) has emerged as an absolute revolutionary maneuver that has become a fundamental constituent in all present-day software. AI incorporates those computational mechanisms that can mimic human-like intelligence and can minimize human intervention. These mechanisms mainly involve Machine Learning (ML) and Deep Learning (DL) methodologies. But, DL-based methods have gained much popularity owing to their superior performance concerning model accuracy with the handling of large-sized data. Besides, the DL method does not require hand-crafted feature engineering techniques, unlike traditional ML algorithms. These DL characteristics have greatly influenced the development of robust architectures applicable across multiple domains, such as image data analysis, natural language processing, grid signal analysis, etc. However, in this article, we will mainly focus on image data analysis using DL.

The extensive involvement of the camera in day-to-day life, as noticed in the case of autonomous driving, surveillance maintenance, or healthcare management, generates a bulk amount of data in the form of images. Images are nothing but an artifact representing and depicting human perception of any physical object/objects in either two-dimensional (2D) or three-dimensional (3D) form. These image data

M. Sarkar

Department of Computer Science and Information Engineering, Chang Gung University, Guishan, Taiwan

P. K. Sahoo (✉)

Department of Computer Science and Information Engineering, Chang Gung University, Guishan, Taiwan

Department of Neurology, Chang Gung Memorial Hospital, Linkou, Taiwan

e-mail: pksahoo@mail.cgu.edu.tw

require continuous analysis for extracting meaningful information. But, manual inspections of these data are challenging, laborious, and time-consuming. Besides, human intervention is erroneous and subjected to inter and intra-observer variability. Thus, computerized digital image processing becomes vastly essential. DL methods are used tremendously in solving problems related to digital image processing such as image classification, colorization, detection, localization, segmentation. Amidst several DL techniques, such as DCNN, LSTM (Long Short Term Memory), Deep Boltzmann Machines (DBM), Recurrent Neural Network (RNN), Auto Encoder (AE), Deep Neural Networks (DNN), particularly DCNN, from 2012 onwards, have received much attention owing to the proposed AlexNet. AlexNet has successfully reduced the error rate from 25.8% to 16.4% [1] in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) dataset. The DCNN models have the potential ability to learn complex non-linear functions along with self-evaluation over end-to-end training of features generates superior results in comparison to the traditional ML algorithms. The self-evaluation and learning method helps in capturing data variations through backpropagation, where DCNN tries to minimize the differences among ground truth and predicted data. The in-built invariance of DCNN generates advisable results for image classification tasks but failed for dense prediction tasks as in image segmentation due to the presence of undesired spatial information [2].

Image segmentation is the approach for separating the digitized image into multiple visually distinct sub-regions having similarities in properties, such as color, texture, gray level, brightness, and contrast which have a semantic meaning for the given problem. Formally, image segmentation can be defined as [3]

the partition of an image into a set of non-overlapping regions whose union is the entire image.

Therefore, the main goal of image segmentation lies in simplifying the image constituents, into a more eloquent and easily comprehensible form. Mathematically, the segmentation of an image can be represented as in Eq. 1 [4]

$$\bigcup_{i=1}^{N_s} S_i = I, S_i \cap S_j = \emptyset, i \neq j \quad (1)$$

The application of image segmentation varies widely. For example, in Computer-Aided Diagnosis (CAD) for tumor lesions, tumor cell segmentation from the non-tumor cell becomes the initial step. In the case of Context-Based Image Retrieval (CBIR) methods, segmentation helps in extracting the essential and informative features which are relevant to the query. Image segmentation also plays an active role in self-driving cars, video surveillance, augmented reality systems, robotic cognizance, scene parsing, etc. 3D reconstruction, a rudimentary outline, and an object appearance detector can also be constructed from the segmentation result using interpolation algorithms such as marching cubes, tetrahedrons, and so on. Therefore, the diverse problem statements coupled with a wide variety of datasets require different image segmentation models. These models have distinct

network architectures. Therefore, this chapter focuses on some popular DL models commonly used in solving image segmentation tasks.

2 Image Segmentation

Image segmentation assists in extracting the Region of Interest (RoI) present within an image. The extracted region contains information about the RoI boundaries as well as their location in the image. This information is necessary for object recognition. For instance, a food identification system requires the knowledge of each item present on a plate. The identified food aids in nutrition value calculation which is very helpful for people suffering from obesity. Therefore, segmentation becomes the primary driving component for solving a wide range of image processing tasks. Despite being several years of research, segmentation models still require thorough structural improvement to achieve condensing performance than human analytics. Thus, before proceeding further in structural analysis, comprehensible knowledge regarding different available datasets, the purpose of segmentation, typical operations used in DL models for image segmentation, and performance metrics towards the successful evaluation of the segmentation model, is needful.

2.1 Image Database Domain Types

The plethora of datasets, competitions, and challenges have greatly influenced and encouraged researchers to propose miscellaneous state-of-the-art segmentation architecture which can be applicable across a wide range of domains. These domains include distinct challenges which are disjoint from each other. The categorized domains are conveyed below in Fig. 1.

- **Natural scenes (NS):** NS databases used for scene recognition-related tasks, mainly contains details of labeled photographs of street view, mountain areas, indoor scenes like in-home, museum, etc. For example, a scene-centric new database called “*Places*” from MIT Computer Science and Artificial Intelligence Laboratory offers 2.5 million images of 205 scene categories along with labels of each category [5]. Other popular NS dataset includes Berkley Segmentation Dataset, PASCAL VOC, Microsoft COCO, MIT Scene Parsing Data, etc. [6].
- **Medical Imaging Modalities (MIM):** MIM describes dataset which contains information related to the human body. These images are typically used to visualize the interior of the body without surgical intervention. Corresponding datasets are brain-related neuro-modalities: Computerized Tomography (CT), Magnetic Resonance Imaging (MRI), scan images, Digital Subtraction Angiography (DSA), and so on; liver tumor segmentation; breast cancer histology

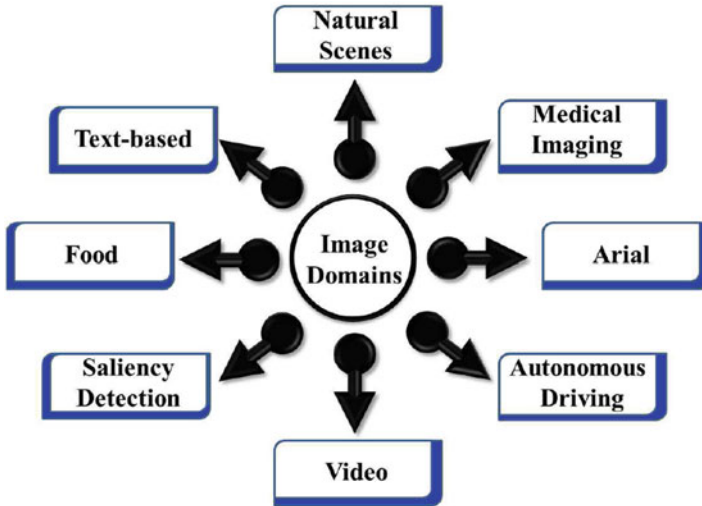


Fig. 1 Describes the different available image domain types for image segmentation

images, etc. The notable datasets are BraTS (Brain Tumor Segmentation Challenge 2020: Data), LiTS (Liver Tumor Segmentation Challenge), DRIVE (Digital Retinal Images for Vessel Extraction), SCR (Segmentation in Chest Radiographs) [6].

- **Arial Imaging (ArI):** ArI mainly includes satellite images or drone images. Therefore, ArI can be grouped into Satellite Imagery (SI) and Drone Dataset (DD). The significant dataset corresponding to SI and DD are:
 - **SI:** SI covers images from satellite. Some notable databases include COWC (Cars Overhead with Context) [7], Microsoft Canadian Building Footprints [7], DSTL Satellite imagery Feature Detection [7], DeepGlobe [6], Google Open Street Map [6].
 - **DD:** DD describes images captured through drones and a few remarkable datasets of DD are Stanford Drone Dataset [7], Vertical Aerial Photography [7], etc.
- **Video Segmentation Dataset (VSD):** Apart from images, video sequences can also be segmented. DL model trained on video data can be applied in surveillance, object motion detection. The popular VSD dataset is YouTube-Video Object segmentation, VSB100 (Video Segmentation Benchmark), etc. [6].
- **Salient Object Detection Database (SODD):** The term “saliency” refers to the “most noticeable” or “most featured” or “most important” object present in an image. The rose in Fig. 2a(A) and the white and pink daffodils in Fig. 2b are the most noticeable elements. Thus, rose and daffodils can be regarded as the “salient” objects with respect to Fig. 2a and b.

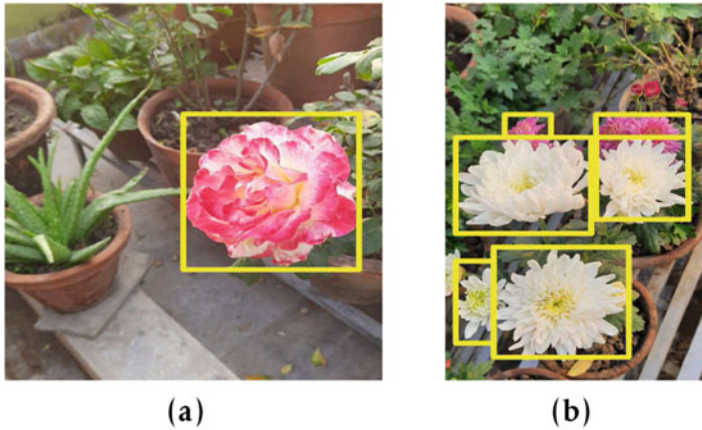


Fig. 2 Signifies the presence of salient objects in image. (a) Image with one salient object. (b) Image with multiple salient objects

- **Food Database (FD):** This database mainly consists of different food images which are being served generally in the restaurants or the individual food habit of human beings as seen in UNIMIB2015 Food Database [8], UEC Food 256 Dataset [9], and so-forth.
- **Text-based Image Segmentation Database (TISD):** TISD mainly consists of scene text (Street View Text Dataset, ICDAR Robust Reading competitions, etc.), machine-printed documents (PRImA Layout Analysis Dataset, IMPACT Database, etc.), graphical documents (Braille Dataset, Trademarks image dataset), handwritten documents (UNIPEN database, Rimes Database, so-forth) [10].

2.2 Purpose of Image Segmentation

Image segmentation has ushered in a revolutionary change in providing flexibility in image editing to concluding robust prediction in uncertain, real-time robotic environments. The low granular level analysis provides more meaningful insights and an easy understanding of digitized images. Therefore, the various fields where image segmentation has been adopted considerably, have been explained below:

- **RoI Detection:** The detection of RoI plays a fundamental role in recognition systems, like, human face recognition systems. Human faces have distinct color range along with varied texture among different races. Therefore, in a complex scenario like in a busy street or crowded place, it becomes very difficult to identify each person separately using physical methods. Another example could be, due to the Covid-19 pandemic situation, many public places have

restricted social gatherings to a definite limit. But, the manual headcount is not possible when the place is overcrowded. In such scenarios, segmentation helps by separating the human faces from the background. The segmentation results obtained then can be used by an automated person verification system or tallying the headcounts. The similar application includes fraud detection, fingerprints identification, brake light detection.

- **Biomedical Image Segmentation:** Image segmentation has made an immense impact on the CAD system. For instance, in the case of vessel segmentation, particularly for the retina, where the vessel structure gets affected due to diseases such as diabetics, muscular degeneration due to age, glaucoma. Manual artery delineation in such cases becomes a strenuous and tedious job for the physician. Thus, AI-enabled CAD helps in easy vessel structure analysis.
- **Real-time Road Segmentation:** This type of real-time-based application is mainly useful for traffic control systems, autonomous cars, which require prompt image analysis. Delay in image processing in such scenarios can be fatal. Besides, road quality assessment can be done where the physical investigation is not always feasible.
- **Food Identification:** The increasing trends towards individual health conciseness have triggered the calculation of nutritional value present in a single food serving. The naked human eye fails to predict food quantity along with the accurate nutritional value. In such cases, segmentation defines the boundaries of each food item. Thereby, assists the necessary algorithms in performing automated analysis.
- **Text Recognition:** “Text” segmentation from the image is the key mechanism for applications where it is mandatory to detect the presence of possible manuscripts in the text area. It is also widely used in handwriting detection and identification. Besides, house plate number, street number, electronic meter readings are some popular applications.

The varied application has been explained diagrammatically in Fig. 3 with the help of images collected from the MS-COCO dataset [11] for RoI segmentation, Food item detection, DRIVE dataset [12] for biomedical image segmentation, and CamVid [13] for road surface segmentation.

2.3 Operations involved in Image Segmentation

Typically, DL-based segmentation models involve several operations such as convolution, pooling, etc. which will be discussed in this section. Understanding each of these operations is necessary to decipher the requirement of different segmentation models and their evolution over the years. The different operations in each layer, which include convolution, atrous convolution, transposed convolution, max-pooling, upsampling, pixel-wise softmax layer, and so on used on those models, have been explained with minimal complex mathematics along with a



Fig. 3 Purpose of image segmentation in application point of view

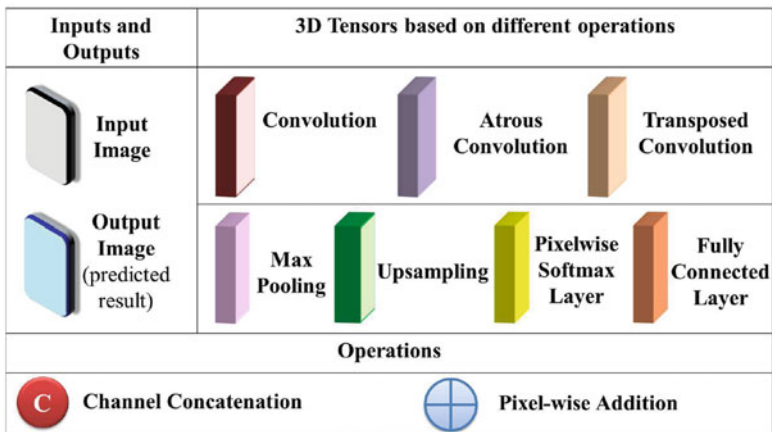


Fig. 4 Graphical representations of different operations in each layer for subsequent DL-based segmentation architecture analysis

common graphical appearance as indicated in Fig. 4. Furthermore, the knowledge will help in designing DL-based customized segmentation architectures.

Generally, segmentation models involve two distinct networks: encoder network and decoder network, as illustrated in Fig. 5. The encoder networks act as feature extractors. It downsamples the spatial resolution of the input image while increasing the channel number in the feature maps. Subsequently, the decoder network functions as a shape generator by upsampling back the obtained feature maps from the encoder network back to the initial input image size. Thus, the principal purpose

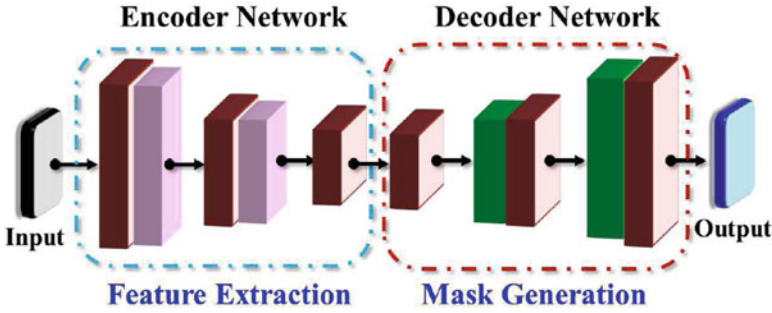


Fig. 5 Displays the typical DL-based image segmentation model architecture

of the decoder network lies in generating a segmentation mask. Therefore, the decoding operation expresses the inverse working procedure of encoding.

- **Convolution Layer:** Convolution is a product (*) of the pixel value at any position of image or feature map and kernel (filter) which is depicted as an array of numbers. The output (or feature map) $o(t)$ is described by Eq. 2 where input $y(t) \in Y$ has undergone convolution operation using a filter or kernel $\omega(c)$ [14].

$$o(t) = (y * \omega)(t) \tag{2}$$

When it takes only integer values, the discretized convolution is given by Eq. 3:

$$o(t) = \sum c y(c) * \omega(t - c) \tag{3}$$

This type of convolution is also known as a one-dimensional convolution operation. Moreover, one-dimensional convolution is very effective in extracting features from a fixed-length segment of the overall dataset, where the location of features in the segment is not important. A two-dimensional convolution operation can be defined as in Eq. 4 for input $y(r, s) \in X$ and a kernel $\omega(c, d)$ [14].

$$o(t) = \sum c \sum d [y(c, d) * \omega(r - c, s - d)] \tag{4}$$

By using the commutative law, Eq. 4 can be modified to Eq. 5 after flipping the kernel

$$o(t) = \sum c \sum d y(r - c, s - d) * \omega(c, d) \tag{5}$$

- **Rectified Linear Unit Layer:** Generally, each convolution layer is followed by a Rectified Linear (ReLU) layer which converts negative input values to zero.

It helps in avoiding vanishing gradient problem [14]. Mathematically, it can be expressed as in Eq. 6

$$f(y_i) = \max(0, y_i) \quad (6)$$

where $y_i \in y$ is the value of a particular location in input feature map.

- **Atrous Convolution Layer:** The term “atrous” has been coined from the French word àtrous which means “hole.” Atrous Convolution (AtC) is also known as “hole convolution,” or “dilated convolution” [2] is based on dilation rate (r). “ r ” signifies the hole (space) among the values present inside a kernel. For example, a kernel of 3×3 having r as 2 will have a field view of 5×5 , but, AtC will only consider 9 values from the kernel. Thus, AtC offers a broader field view with the same computational cost. And, AtC is mainly popular with real-time segmentation where latency along with accuracy is vital as in the case of autonomous driving cars.
- **Pooling Layer:** The pooling operation is done by sliding a two-dimensional filter over each channel of the feature map and then summing up the features lying within the covered region. It minimizes the feature map size (height and width, but not depth) along with the number of parameters required for calculation. There are three types of pooling available:
 - Max-pooling (most frequently used)
 - Average-Pooling
 - L2-normalization

Max-pooling works by extracting the maximum input value inside a filter and discards the rest of the values. Whereas, Average-pooling takes the average of all elements present inside a filter. Max-pooling effectively summarizes the strongest activations over a neighborhood. However, if the position of the feature is important in any particular analysis, then Average-pooling gives better results compared to Max-pooling.

The series of successive convolution and pooling operations in the encoder results in a subsequent reduction in the spatial dimensions of the input image as it goes deeper and deeper, thereby creating an abstract representation of the input image. This reduced abstract feature is useful for performing image classification tasks. But, image segmentation requires output to be of the same dimension of the input image having distinct spatial information instead of mere feature maps. This led to the concept of upsampling whose main purpose is to bring back the resolution of the current (reduced) feature map to the resolution of the previous layer. Several methods exist like nearest neighbors, bi-linear interpolation, bed of nails, max un-pooling, transposed convolutions. Among the several methods available for upsampling, the decoder mainly uses max un-pooling and transposed convolutions.

- **Max-Pooling Layer:** During the pooling operation, a matrix is created for storing the location of maximum value within the filter. The un-pooling operation

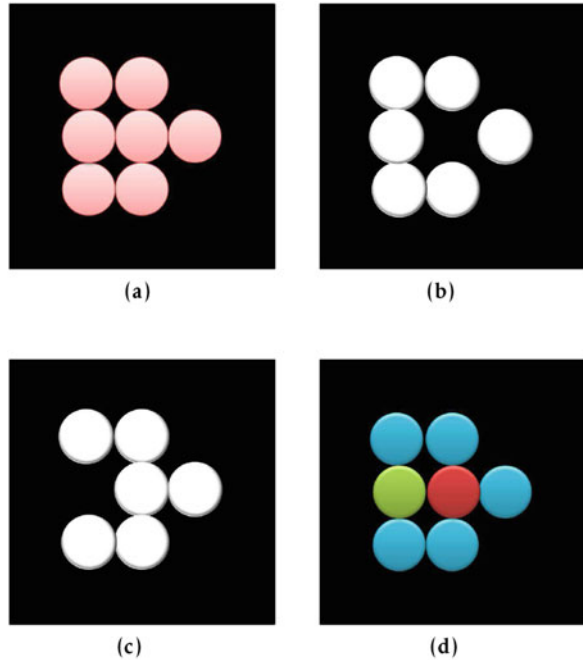
uses those locations of max pooled values and inserts the pooled value in the original place, with the remaining elements being set to zero. Therefore, by tracing the original locations, un-pooling captures example specific to structures with strong activations back to image space. Also, it can reconstruct the detailed structure successfully.

- **Transposed Convolution Layer:** Transposed convolutions are also known as “fractionally strided convolutions,” “deconvolutions,” or “upconvolutions” [15]. It operates by switching the forward and backward passes of a convolution. Although kernel defines a convolution, but forward and backward pass calculation determines direct convolution or transpose convolution. For example, direct convolution calculates the product of forwarding pass (C) and backward pass (C^T) with kernel ω . However, using the same kernel ω , transposed convolution is calculated as the product of C^T and $(C^T)^T = C$. Practically, Transposed Convolution operation can be thought of as the gradient of convolution operation concerning its input [15].
- **Pixel-wise Softmax Layer:** The softmax activation function is normally present in the last layer of the decoder network. The objective of softmax is to normalize the output vector (containing “M,” real values) to a probability distribution over the predicted output classes so that, the sum of the resultant vector (containing “M,” real values) always sums to 1. Applying softmax pixel-wise helps in generating the segmented output in the same size as the input image.
- **Fully Connected (FC) Layer:** FC layers present only in the decoder network of some DL-based segmentation models like DeepMask, Mask R-CNN (Mask Recurrent Convolutional Neural Network). FC layers help in pixel-wise classification and localization. For DeepMask, the FC layers do not contain ReLu (activation function) as ReLu neglects the negative value and works only on positive values that may generate poor segmentation results. However, Mask R-CNN uses FC layers for predicting the output mask.

2.4 Performance Metrics for Segmentation Models

Normally, the number of classes present in an image characterizes image segmentation as the binary-class image (if only a single RoI class present along with background class) and multi-class image (if more than one RoI class exists along with background class). Thus, in this section, we delineate the commonly used metrics used for evaluating the performance of segmentation models concerning binary and multi-class image segmentation. Theoretically, a segmentation model can be evaluated based on different characteristics including accuracy, speed, memory, time, and power consumption [16]. However, speed and time are directly dependent on hardware associated with the experiment. Besides, the visual quality of the segmented output is still considered an important factor in deciding the best segmentation model considering any given problem. The different quantitative metrics available in favor of evaluating segmentation models are described below

Fig. 6 Describes the overlapping scenario for binary-class image segmentation model performance analysis. In Fig. 6d, “cyan,” indicates TP, “red,” shows FP, “green,” signifies FN and “black,” specifies TN. (a) Input image. (b) Ground truth of input image. (c) Predicted output of input image. (d) Overlapped result of ground truth and predicted image



with the help of Fig. 6. Figure 6a describes the input (binary-class) image where Fig. 6b indicates the corresponding ground truth image of Fig. 6a and the final predicted image after segmentation is Fig. 6c. Furthermore, to visualize the overall performance of the binary segmentation model, Fig. 6d has been formed by overlapping ground truth images on the predicted output. The purpose of overlapping the image is to identify the wrongly predicted regions in the output image. However, different colors have been introduced to explain the incorrect predictions. For instance, “cyan” signifies True Positive (TP) meaning that the region has been correctly predicted as RoI, while “green” indicates False Negative (FN) as the RoI region has been mistaken as background. On the other hand, False Positive (FP) marked as “red” specifies those background regions which has been misidentified as RoI. Finally, “black” reveals True Negative (TN) demonstrating background class. The model performance assessment metrics are discussed further using the TP, FP, FN, TN.

2.4.1 Accuracy

The accuracy metrics signifies the correlation among the appropriately segmented regions and the ground truth regions. For binary-class segmentation, accuracy can thus be explained through the Eq. 7.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

2.4.2 Pixel-Wise Accuracy (PA)

It detects the proportion of correctly classified pixels over total number of pixels present in an image. Thus, PA for $N + 1$ classes (“N” indicates number of foreground classes and “1” indicates the background class) can be described as [17] in Eq. 8.

$$PA = \frac{\sum_{i=0}^N P_{ii}}{\sum_{i=0}^N \sum_{j=0}^N P_{ij}} \quad (8)$$

where indicates the number of pixels of class i predicted as belonging to class j . However, PA produces inappropriate results when the dataset is unbalanced (that is, when a single class dominates largely in entire image) [17].

2.4.3 Mean Pixel Accuracy (MPA)

MPA is mainly used for multi-class segmentation model analysis. MPA is the measure of class-wise correctly segmented pixels which is then averaged over the combined classes (N) present, as shown in Eq. 9.

$$MPA = \frac{1}{N + 1} \sum_{i=0}^N \frac{P_{ii}}{\sum_{j=0}^N P_{ij}} \quad (9)$$

2.4.4 Precision

Precision, also known as positive predictive value, describes the relation between TP and all element classified as positive [17]. It can be defined as in Eq. 10.

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

2.4.5 Recall

Recall, also known as sensitivity, indicates the correctness of the predicted segmented mask for each class [17]. Recall is defined as in Eq. 11.

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

2.4.6 Specificity

Specificity, also known as selectivity or true negative rate, signifies the ratio of negatives which has been segmented correctly. Specificity can be described as in Eq. 12

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (12)$$

Moreover, for binary image segmentation, the background class signifies the negative class. As a result, specificity and pixel accuracy are generally ignored [18].

2.4.7 Intersection over Union (IoU) and Dice Coefficient (Dice)

Popular metrics based on overlap are particularly IoU or the Jaccard Index (Jaccard), F1-score or Dice, and Mean IoU (MIoU). However, in the case of segmentation, the Dice score is generally written instead of F1-score. For, binary-class segmentation analysis, IoU/Jaccard and Dice can be calculated as in Eq. 13 and Eq. 14, respectively.

$$\text{IoU} = \text{Jaccard} = \frac{TP}{TP + FP + FN} \quad (13)$$

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \quad (14)$$

However, in case of multi-class segmentation, given any predicted segmented class matrix (A) and the ground truth class matrix (B), IoU or Jaccard [19] can be defined as Eq. 15, Dice can be defined as Eq. 16.

$$\text{Jaccard}(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (15)$$

$$\text{Dice}(A, B) = \frac{|A \cap B|}{|A| + |B|} \quad (16)$$

Moreover, Jaccard and Dice can be related by the Eq. 17. Dice can also be defined in terms of precision and recall using Eq. 18 [17]. And MIoU can be described as the average IoU overall classes [19].

$$\text{Jaccard} = \frac{\text{Dice}}{2 - \text{Dice}} \quad (17)$$

$$Dice = \frac{2 * Precision * Recall}{Precision + Recall} \quad (18)$$

3 Types of Image Segmentation

There exist different types of image segmentation. But, these distinct segmentation types prevail due to the presence of diversified disproportionate contents such as trees, humans, and so on, present in an image. Thus, in this section, we will look over different contents of an image, current image segmentation types. Furthermore, we will also investigate some popular existing segmentation models developed so far.

3.1 Segmentation Concepts

The basic segmentation concepts are based on components present within an image. The entire image components can be broadly classified into two classes namely: “things,” and “stuff.” The “things,” and “stuff,” are determined on the basis of five characteristics: (1) shape, (2) size, (3) parts, (4) instances, and (5) textures [20]. A contrasting study will be performed in Table 1 on the basis of point of differences (PoD) between “things,” and “stuff.”

However, there are some scenarios such as huge crowd where both “things,” and “stuff,” classes can be recognized as one class. Furthermore, the “things,” and “stuff” are explained using Fig. 7a [11] where “things,” class indicates the person, car, street light, bench, etc., as shown in Fig. 7b, and “stuff,” class considers the sky, grass, and road as in Fig. 7c.

Based on the image contents (“things” and “stuff”), image segmentation types can be widely categorized into Semantic Segmentation (SS), Instance Segmentation (IS) and Panoptic Segmentation as shown in Fig. 8.

Table 1 Comparative study to differentiate “things,” and “stuff” classes in image

PoD	“Things”	“Stuff”
Shape	Objects are of specific shape. (E.g.: cat, car)	Objects are nebulous. (E.g.: river, grass)
Size	Distinct with moderate variation in size	Irregular with high variation in size
Parts	Possesses identifiable parts	Has inconspicuous parts
Instances	Countable in nature	Uncountable in nature
Textures	Have varying textures. (E.g.: skin)	Have consistent textures. (E.g.: sky, tree) [21]

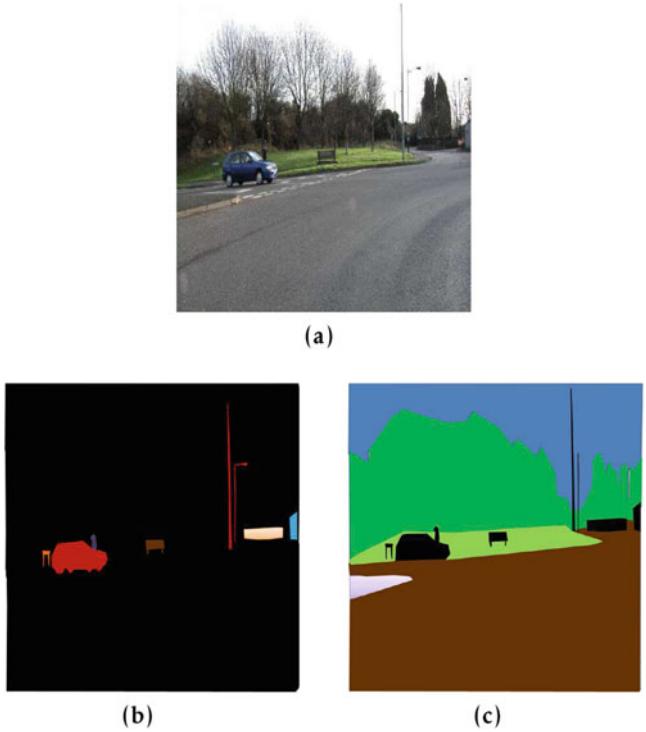


Fig. 7 The components of an image. (a) Input image. (b) “things” in input image. (c) “stuff” in input image

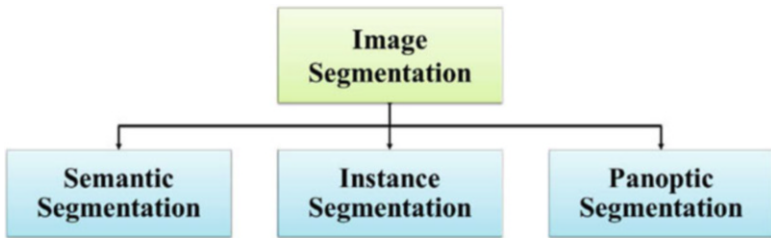


Fig. 8 Signifies the image segmentation types

SS refers to the task where each pixel of an image is associated with a class label for easy differentiation between “things,” and “stuff,” contents of an image. However, SS does not differentiate multiple instances of the same class. On the other hand, IS detects and delineates each object of interest in the image. Specifically, IS is concerned with the individual distinction of “things.” PS, on the other hand, is the unification of both SS and IS. In other words, PS associates each pixel with its class label and instance number. Therefore, each “stuff” classes and the “things” classes

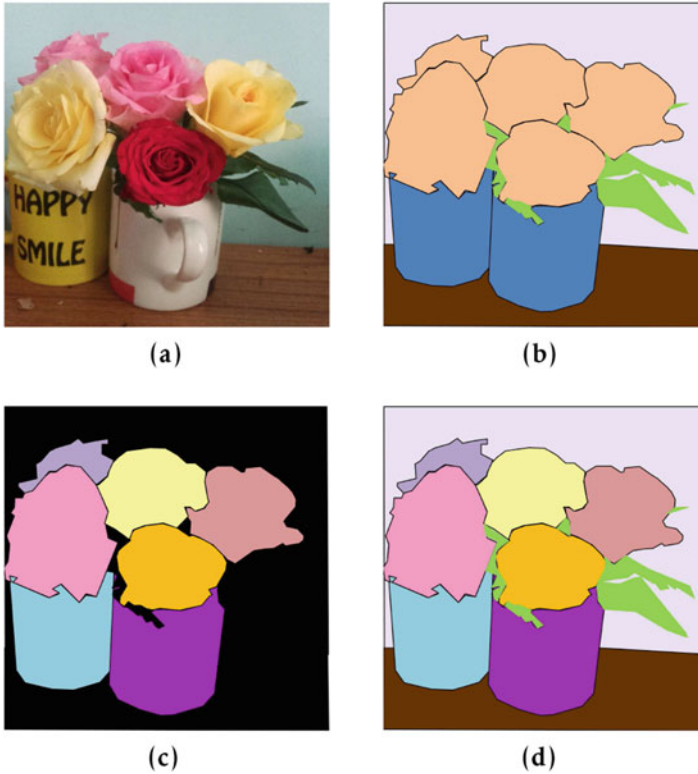


Fig. 9 Explains the visible differences in output for SS, IS, PS, respectively. (a) Input image. (b) Output of SS. (c) Output of IS. (d) Output of PS

are separated and multiple individuals of the “things” classes are also separated from each other [11]. Moreover, the scope of this chapter is limited to SS and IS only.

The various classes generate different results with respect to SS, IS, and PS. Therefore, an input image Fig. 9a after performing SS separates “things” (roses and mugs) and “stuff” (wall, leaves, and table) classes and generates output as in Fig. 9b. But, in the case of IS, only the individual “things” (roses and mugs) got segmented as shown in Fig. 9c and PS shows both the “things” (individual roses and mugs) and “stuff” (wall, leaves, table) classes as explained in Fig. 9d.

3.2 Semantic Segmentation

Let us consider $y \in Y = \mathbb{R}^{H \times W \times 3}$ is composed of a set of pixels I with constant cardinality $|I| = N$, $x_s \in X_s = \mathbb{R}^{H \times W \times C}$ denotes the corresponding ground truth of y , where C is the number of classes present in the ground truth and X^N ,

denotes the output space which is the product set of N -tuples with elements in a label space X . Here, H and W represent the height and width of the image, respectively. Mathematically, the goal of semantic segmentation is to assign each pixel $y_i \in y$, a label $x_j \in X$ [21] based on $x_j \in x_s$. Thus, a fully convolutional neural network trained in a supervised manner with Y produces softmax output volume of size indicating predicted semantic class probabilities [22, 43]. Therefore, SS can be stated as labeling each pixel of an input image in the output image.

Several progressions have been made in SS so far. Some existing popular semantic segmentation models have been explained according to their years of development. For instance, in 2015, FCN (Fully Convolutional Neural Network) [16], encoder-decoder based symmetric UNet [16] and SegNet [16] were developed. In 2016, attention-based FusionNet [23], asymmetric DeepLab [16] and faster inference concluder for low-latency based operation, ENet (Efficient Neural Network) [16] were proposed. And during 2017, AdaptNet was proposed where the architecture of ResNet-50 was adapted in which an additional convolutional kernel was introduced before the first convolution layer of ResNet. AdaptNet also instigated the idea of the convoluted mixture of deep experts (CMoDE) fusion scheme [16]. Besides, FC-DenseNet (Fully Convolutional DenseNet) [16], RefineNet [16], WNet [24], RedNet [25], LinkNet [25] also evolved during 2017. FC-DenseNet was formed by the addition of skip-connection and upsampling layer to the already established Densely Connected Convolutional Networks (DenseNets). RefineNet, on the other hand, was developed using residual connection design. It also contains Residual Convolution Unit (RCU), Multi-resolution fusion, and Chained residual pooling. Whereas in 2018, LadderNet [26], BiseNet (Bilateral Segmentation Network) [27], ERNNet (Edge Loss Reinforced semantic segmentation network) [28] came up. Besides, O-Net [29], LED Net [19], Fickle Net [19] emerged in 2019, while, T-Net [30] (Fully tensorized FCN architecture), MED-Net (Multiscale Encoder-Decoder Network) [31] and EF-Net (Enhancement and Fusion Network) [32] was developed in 2020. The timeline of the evolution of the SS models has been described in Fig. 10.

3.2.1 Fully Convolutional Networks

The success of convolutional networks in performing classification tasks lead to the proposal of FCN in 2015, [16] for pixel-wise SS of an image. FCN considered contemporary classification networks such as AlexNet, VGG net, and GoogLeNet as the backbone. FCN accepts the input image of random size and generates an output of corresponding input size with efficient inference and learning. However, for preserving the contextual spatial information within an image, the output of the upsampling layer was fused with the previous layer's output as shown in Fig. 11. The author has further considered three variations of FCN, namely FCN-8, FCN-16, FCN-32 based on the fusion layers. However, object boundaries are seen as poorly localized in the segmentation results of FCNs as FCN considers usage of

Fig. 10 Timeline of different SS architectural development

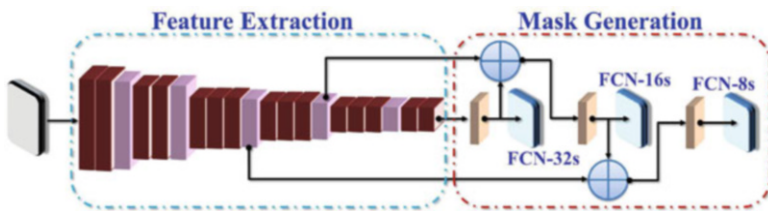
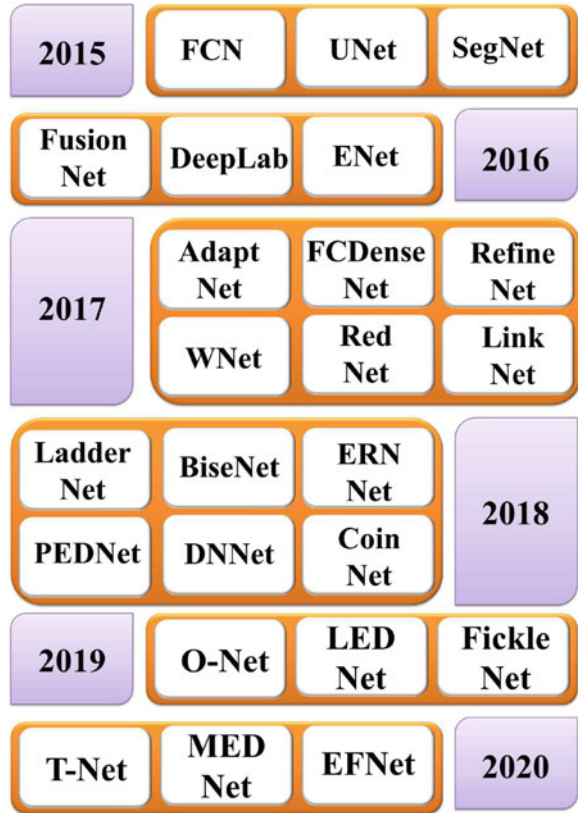


Fig. 11 FCN architecture in graphical form

many pooling layers and large receptive fields that produces low spatial resolution and blurring in the deep layers.

3.2.2 SegNet

SegNet [16], is a popular encoder-decoder-based SS model which uses max pooled indices from the encoder to perform non-linear upsampling in the decoder as shown

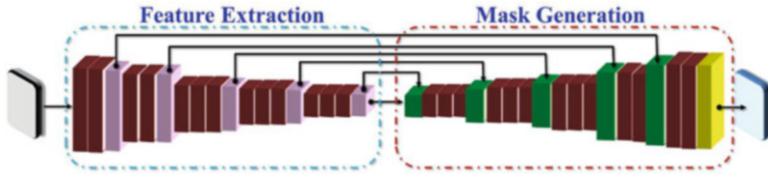


Fig. 12 Graphical architecture of SegNet

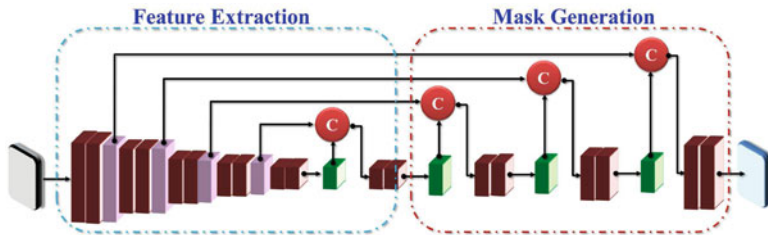


Fig. 13 Graphical representation of UNet architecture

in Fig. 12. Each convolution layer in the encoder is followed by a Rectified Linear (RL) layer and Batch Normalization layer. They have used the VGG-16 network in the encoder where they have removed the fully connected layers to further reduce the number of parameters.

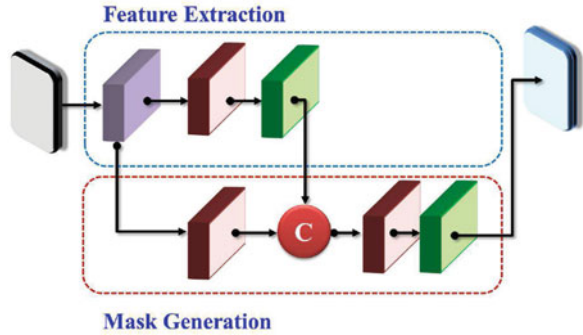
3.2.3 UNet

In 2015, UNet was proposed [16] for biomedical image segmentation which consists of a contracting path to capture context information which is followed by a symmetric expanding path for precise localization. Instead of using max pooled indices only, as in case of SegNet, UNet concatenates previous tensor information and upsampling tensor instead of element-wise addition as shown in Fig. 13.

3.2.4 DeepLab

Apart from symmetric SS models, there exists some asymmetric encoder-decoder SS networks such as E-Net [16], DeepLab [16], etc. Nevertheless, the performance of pixel-wise segmentation suffers greatly because of kernel size. Smaller kernel size fails to capture the contextual information present in an image, whereas large kernel size produces slower results due to the presence of numerous trainable parameters. Drawbacks of smaller kernel sizes are outperformed in classification problems with the introduction of pooling layers which increases the sensory area of each kernel. However, additional pooling layers decrease the sharpness in the segmented output image. To counteract these issues, DeepLab family, proposed and

Fig. 14 DeepLabv3+ model explained graphically

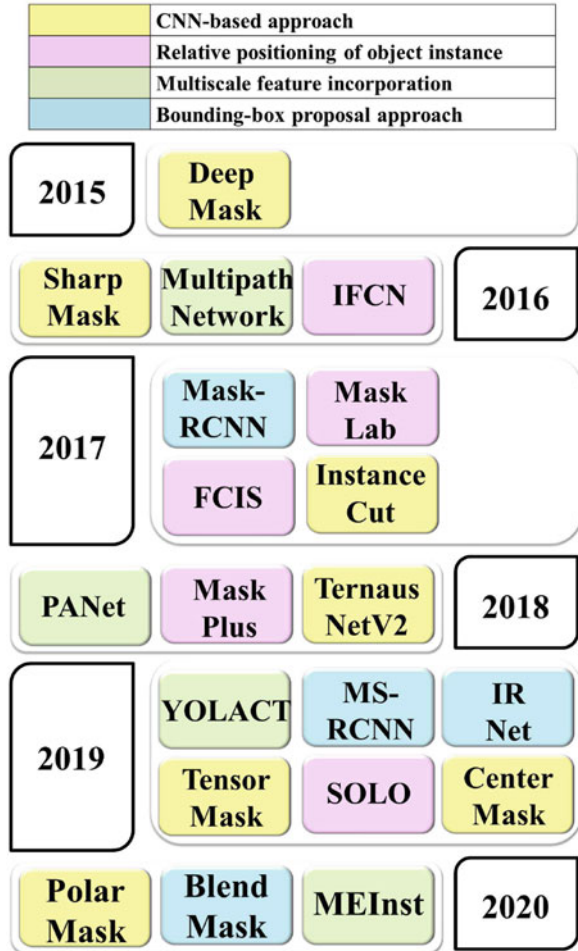


open-sourced by Google, has introduced the concept of AtC, Conditional Random Field (CRF), and spatial pyramidal pooling layers in DeepLabv1. For improving the performance of DeepLabv1 further, DeepLab v2 was proposed which focuses mainly on the problem of the existence of objects at multiple scales. Here, they have fused the output of multiple AtC operations obtained by varying sample rates. But, the architecture fails to capture object boundaries precisely. However, DeepLabv3, adopts depth-wise separable convolution for enhancing the segmentation output further along with increasing computational efficiency. Even so, the most popular DeepLab architecture is DeepLabV3+ as shown in Fig. 14, which uses Xception architecture as the backbone in the encoder part coupled with removing max-pooling operations with depth-wise separable convolutions [2].

3.3 Instance Segmentation

Let $y \in Y = \mathbb{R}^{H \times W \times 3}$ be an image and $i \in I = \{1, \dots, H\} \times \{1, \dots, W\}$ indicates a pixel. The goal of instance segmentation is to map the image to a collection $A_y = \{A_1, \dots, A_{k_y}\} \subset 2^I$ of image regions, where each region represents the occurrence of an object of interest. The symbol $A_0 = I - \cup_k A_k$ denotes the background [33]. IS can be broadly categorized into CNN-based approach, relative positioning of object instance, multiscale feature incorporation and bounding-box proposal approach. CNN-based approach includes models such as DeepMask [34], SharpMask [35], Instance Cut [34], TerausNetV2 [36], TensorMask [34], CenterMask [34], PolarMask [34]. While, IFCN (Instance-sensitive Fully Convolutional Neural Network) [37], MaskLab [38], FCIS (Fully Convolutional Instance-aware Semantic Segmentation) [39], MaskPlus [40], SOLO (Segmenting Objects by Location) [34] falls under relative positioning of object instance. However, Multipath Network [34], PANet (Path Aggregation Network) [34] and MEIst (Mask Encoding for Single Shot Instance Segmentation) [41] uses multiscale feature incorporation for performing IS. Finally, Mask R-CNN [34], MS-R-CNN (Mask Scoring R-CNN)

Fig. 15 Timeline of IS architecture development



[42] and BlendMask [34] uses bounding-box proposal approach. The gradual advancement of IS algorithms has been explained in Fig. 15.

3.3.1 DeepMask

Facebook AI Research (FAIR) has proposed DeepMask [34] as shown in Fig. 16 based on traditional feed-forward neural network architecture callable of performing multitasking. Here, DeepMask focuses on generating a segmentation proposal region instead of a bounding box as in the case of object detection. The feature extraction part consists of VGGNet where the last FC layer along with the last max-pooling layer has been removed. The architecture consists of two approaches: mask generation and score generation. Mask generation consists of an FC layer

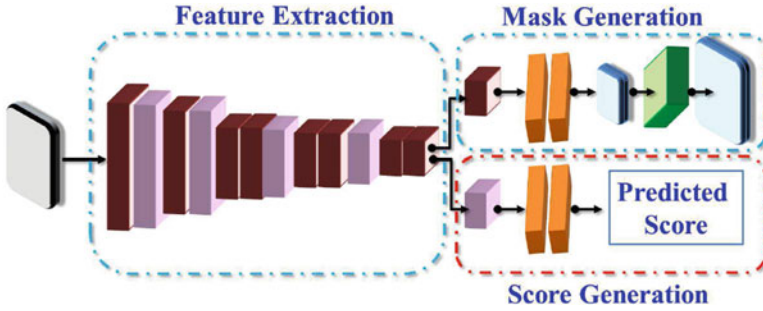


Fig. 16 DeepMask architecture in details

along with upsampling layer for generating a class-agnostic segmentation mask. While the score generation network assigns a score relative to the probability of the patch-containing object. Thus, specifying a patch as an input, DeepMask generates a segmentation mask along with showing the probability of the mask being centered on a full object present in the input patch. It is successful in generating “coarse object masks,” but fails in “pixel-accurate segmentations,” due to the presence of bi-linear upsampling in the architecture.

3.3.2 Mask R-CNN

The concept of Instance segmentation can also be summarized as object detection followed by semantic segmentation. Particularly, Mask R-CNN is one of the most remarkable architecture in this consideration. Mask R-CNN is basically an extension of Faster R-CNN where pixel-level object-specific binary classification was performed in parallel for providing accurate segmentation results. Mask R-CNN works in two stages after feature extraction. The first stage produces proposals regarding the probable regions containing any object. Whereas the second stage focuses on predicting the object class, refining the generated bounding-box, and finally generating the segmented mask for the object based on the proposal in the first stage. The detailed transition of Mask R-CNN [34] has been explained in Fig. 17.

4 Research Challenges in Image Segmentation

In spite of several decades of research work, automated segmentation methods have received much attention owing to the existing challenge across different domains. The challenges are a resultant of several factors.

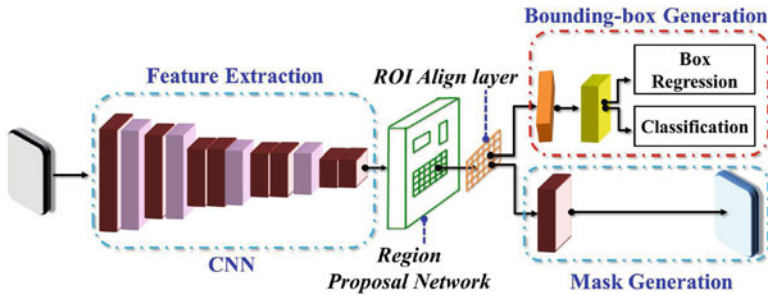


Fig. 17 Mask R-CNN explained using graphics

4.1 Complexity and Computation of DL Segmentation Models

DL-based segmentation models involve a huge number of operations for concluding inference. As a result, those models require sufficiently large memory and computation speed. This makes DL models inappropriate for deployment on resource-constrained devices such as embedded systems. Besides, the compression operation or one-shot decomposition operation decreases feature map redundancy. But, the feature map redundancy reduction may reduce model prediction accuracy for some architecture. Therefore, an adequate investigation is required which can minimize the number of operations without affecting model accuracy [16].

4.2 Variability in Object Appearance

Obscured RoI boundaries, inconsistent object shape, or similar adjacent area generates low-quality segmentation results. The presence of image artifacts, varied imaging conditions due to low image contrast, noise and other pathological reasons such as partial volume effect creates additional problems for segmentation in the medical domain. However, in the case of scene parsing, the unpredictable mobility and location of objects (humans, animals, and cars) also produce improper segmentation.

4.3 Requirement for Huge and Highly Defined Labeled Dataset

Present-day DL models require a sufficiently large dataset along with respective high-quality labeled data in their training phase. But, data annotations being laborious and time-consuming, these large-scale data annotations are not easily available. Thus, DL models try to use semi-supervised or weakly supervised datasets

but fail to achieve good performance. Moreover, manually annotated data often suffer from; intra and inter-observer variability as it relies completely on expert knowledge, resulting in erroneous output [16].

4.4 Overfitting Issue

Most of the available datasets are relatively small resulting in an overfitting issue. The overfitting issue arises due to less number of training images in any particular domain as DL models fail to capture all useful information. Generally, data augmentation (augmentation methods: rotation, translation, horizontal and vertical flipping, etc.) is used to solve the overfitting problem because it increases the training dataset by generating synthetic images. But, data augmentation sometimes produces poor segmentation results. Thus, data augmentation is not a solution optimal solution.

4.5 Class Imbalance

Class imbalance occurs when data of a particular class is more prevalent than other classes. For example, in the case of brain tumor segmentation using MRI, brain tumor only appears in very few slices. Therefore, the segmentation model, if not trained with sufficient brain tumor images, will generate segmentation results containing background class only.

4.6 Issues with Real-Time Segmentation

Real-time segmentation is mainly applicable for autonomous driving, mobile computing, robot interaction, where execution time plays a crucial role in evaluating the performance of DL models. However, most of the segmentation models are time-consuming. To reduce the time several CNN models have been proposed where they have tried to effectively use convolution operations. But, more research is still required to improve the accuracy of the model in the real-time analysis along with maintaining well trade-off between accuracy and execution time [16].

5 Conclusion

Regardless of the development of substantial segmentation architectures, image segmentation still remains a challenging task because of the existing challenges

in terms of preserving contextual information, loss calculation over the predicted image, lack of well-defined labeled dataset, diversified domains along with different RoI, and complexity of different algorithms. Based on the state-of-the-art model performance, image segmentation has been classified broadly into three categories: SS, IS, and PS. SS classifies object features present in an image, IS identifies each instance of an object whereas PS classifies as well as identifies each object instance. However, the success of the model depends largely on the dataset and corresponding annotations, loss function, hyper-parameters, data pre-processing methods. In this chapter, we have tried to highlight different types of available datasets, segmentation methods, existing state-of-the-art DCNN models, and finally existing challenges. Image segmentation has a vast application area, and thus it becomes difficult to conclude any optimal DCNN model for solving every problem. Moreover, some domain-specific fine-tuning might result in producing near-optimal segmentation results for some problems. Altogether, this article provides different existing segmentation models which might help researchers in further proceedings.

Acknowledgments This work was supported in part by the Ministry of Science and Technology (MOST), Taiwan, under Grant number 110-2221-E-182-008-MY3.

References

1. Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*. 2020 Dec;53(8):5455-516.
2. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE transactions on pattern analysis and machine intelligence*. 2017 Apr 27;40(4):834-48.
3. Haralick RM, Shapiro LG. *Computer and robot vision*. Reading: Addison-Wesley; 1992 Feb.
4. Chouhan SS, Kaul A, Singh UP. Image segmentation using computational intelligence techniques. *Archives of Computational Methods in Engineering*. 2019 Jul;26(3):533-96.
5. Zhou B, Lapedriza A, Xiao J, Torralba A, Oliva A. Learning deep features for scene recognition using places database.
6. Ghosh S, Das N, Das I, Maulik U. Understanding deep learning techniques for image segmentation. *ACM Computing Surveys (CSUR)*. 2019 Aug 30;52(4):1-35.
7. Daniel Smith. 15 Drone Datasets and Satellite Image Databases for Machine Learning <https://lionbridge.ai/datasets/15-best-aerial-image-datasets-for-machine-learning/>
8. Ciocca G, Napoletano P, Schettini R. Food recognition: a new dataset, experiments, and results. *IEEE journal of biomedical and health informatics*. 2016 Dec 7;21(3):588-98.
9. Kawano Y, Yanai K. Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In *European Conference on Computer Vision 2014 Sep 6* (pp. 3-17). Springer, Cham.
10. TC-11 Online Resources. <http://tc11.cvc.uab.es/datasets/>
11. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft coco: Common objects in context. In *European conference on computer vision 2014 Sep 6* (pp. 740-755). Springer, Cham.
12. Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*. 2004 Apr 5;23(4):501-9.

13. Brostow GJ, Fauqueur J, Cipolla R. Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*. 2009 Jan 15;30(2):88-97.
14. Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. *IEEE Access*. 2017 Dec 29;6:9375-89.
15. Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*. 2016 Mar 23.
16. Lateef F, Ruichek Y. Survey on semantic segmentation using deep learning techniques. *Neurocomputing*. 2019 Apr 21;338:321-48.
17. Minaee S, Boykov YY, Porikli F, Plaza AJ, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021 Feb 17.
18. Jun TJ, Kweon J, Kim YH, Kim D. T-net: Nested encoder–decoder architecture for the main vessel segmentation in coronary angiography. *Neural Networks*. 2020 Aug 1;128:216-33.
19. Taghanaki SA, Abhishek K, Cohen JP, Cohen-Adad J, Hamarneh G. Deep semantic segmentation of natural and medical images: A review. *Artificial Intelligence Review*. 2021 Jan;54(1):137-78.
20. Caesar H, Uijlings J, Ferrari V. Coco-stuff: Thing and stuff classes in context. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018* (pp. 1209-1218).
21. Tighe J, Lazebnik S. Finding things: Image parsing with regions and per-exemplar detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition 2013* (pp. 3001-3008).
22. Cermelli F, Mancini M, Bulò SR, Ricci E, Caputo B. Modeling the Background for Incremental Learning in Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020* (pp. 9233-9242).
23. Huang HY, Zhu C, Shen Y, Chen W. Fusionnet: Fusing via fully-aware attention with application to machine comprehension. *arXiv preprint arXiv:1711.07341*. 2017 Nov 16.
24. Xia X, Kulis B. W-net: A deep model for fully unsupervised image segmentation. *arXiv preprint arXiv:1711.08506*. 2017 Nov 22.
25. Jiang J, Zheng L, Luo F, Zhang Z. Rednet: Residual encoder-decoder network for indoor rgb-d semantic segmentation. *arXiv preprint arXiv:1806.01054*. 2018 Jun 4.
26. Zhuang J. Laddernet: Multi-path networks based on u-net for medical image segmentation. *arXiv preprint arXiv:1810.07810*. 2018 Oct 17.
27. Yu C, Wang J, Peng C, Gao C, Yu G, Sang N. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European conference on computer vision (ECCV) 2018* (pp. 325-341).
28. Liu S, Ding W, Liu C, Liu Y, Wang Y, Li H. ERN: edge loss reinforced semantic segmentation network for remote sensing images. *Remote Sensing*. 2018 Sep;10(9):1339.
29. Luke GP, Hoffer-Hawlik K, Van Namen AC, Shang R. O-Net: A convolutional neural network for quantitative photoacoustic image segmentation and oximetry. *arXiv preprint arXiv:1911.01935*. 2019 Nov 5.
30. Kossaiji J, Bulat A, Tzimiropoulos G, Pantic M. T-net: Parametrizing fully convolutional nets with a single high-order tensor. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2019* (pp. 7822-7831).
31. Kose K, Bozkurt A, Alessi-Fox C, Gill M, Longo C, Pellacani G, Dy J, Brooks DH, Rajadhyaksha M. Segmentation of Cellular Patterns in Confocal Images of Melanocytic Lesions in vivo via a Multiscale Encoder-Decoder Network (MED-Net). *arXiv preprint arXiv:2001.01005*. 2020 Jan 3.
32. Chen Q, Fu K, Liu Z, Chen G, Du H, Qiu B, Shao L. EF-Net: A novel enhancement and fusion network for RGB-D saliency detection. *Pattern Recognition*. 2020 Nov 4:107740.
33. Novotny D, Albanie S, Larlus D, Vedaldi A. Semi-convolutional operators for instance segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV) 2018* (pp. 86-102).
34. Hafiz AM, Bhat GM. A survey on instance segmentation: state of the art. *International Journal of Multimedia Information Retrieval*. 2020 Jul 3:1-9.

35. Pinheiro PO, Lin TY, Collobert R, Dollár P. Learning to refine object segments. In European conference on computer vision 2016 Oct 8 (pp. 75-91). Springer, Cham.
36. Iglovikov V, Seferbekov SS, Buslaev A, Shvets A. TerausNetV2: Fully Convolutional Network for Instance Segmentation. In CVPR Workshops 2018 Jun 1 (Vol. 233, p. 237).
37. Dai J, He K, Li Y, Ren S, Sun J. Instance-sensitive fully convolutional networks. In European Conference on Computer Vision 2016 Oct 8 (pp. 534-549). Springer, Cham.
38. Chen LC, Hermans A, Papandreou G, Schroff F, Wang P, Adam H. Masklab: Instance segmentation by refining object detection with semantic and direction features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018 (pp. 4013-4022).
39. Li Y, Qi H, Dai J, Ji X, Wei Y. Fully convolutional instance-aware semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017 (pp. 2359-2367).
40. Xu S, Lan S, Qi Z. MaskPlus: Improving Mask Generation for Instance Segmentation. In The IEEE Winter Conference on Applications of Computer Vision 2020 (pp. 2030-2038).
41. Zhang R, Tian Z, Shen C, You M, Yan Y. Mask encoding for single shot instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020 (pp. 10226-10235).
42. Huang Z, Huang L, Gong Y, Huang C, Wang X. Mask scoring R-CNN. In Proceedings of the IEEE conference on computer vision and pattern recognition 2019 (pp. 6409-6418).
43. Subhani MN, Ali M. Learning from Scale-Invariant Examples for Domain Adaptation in Semantic Segmentation. arXiv preprint arXiv:2007.14449. 2020 Jul 28.

Artificial Intelligence Assisted Cardiac Signal Analysis for Heart Disease Prediction



Prasan Kumar Sahoo, Sulagna Mohapatra, and Hiren Kumar Thakkar

1 Introduction

In the current age of digitalization, the individual workload is increasing. Simultaneously, personal health is degrading drastically due to the unhealthy lifestyle, irregular sleeping pattern and unnecessary stress. According to the World Health Organization [1], cardiovascular disease (CVD) is one of the leading causes of death among the people across all age groups. In case of a healthy person, the series of cardiac events such as opening and closing of the heart valves, blood flow into vessels, and contraction-relaxation of ventricular walls are occurred in a predefined order or at a regular interval. In contrast, in case of any coronary heart disease (CHD) like myocardial ischemia, infarction, arrhythmias, etc., hamper the normal cardiac activities and can lead to heart attack if those abnormalities are not detected in an early stage. There is a huge need for constant monitoring of cardiac abnormality symptoms like dizziness, nausea, and chest pain as shown in Fig. 1. However, those symptoms are not well differentiable and sometimes misdiagnosed as the normal disease. Mostly, the irregular heartbeats known as ectopic heartbeats

P. K. Sahoo (✉)

Department of Computer Science and Information Engineering, Chang Gung University, Guishan, Taiwan

Department of Neurology, Chang Gung Memorial Hospital, Linkou, Taiwan

e-mail: pksahoo@mail.cgu.edu.tw

S. Mohapatra

Department of Computer Science and Information Engineering, Chang Gung University, Guishan, Taiwan

H. K. Thakkar

Department of Computer Science and Engineering, SRM University, Amaravati, India

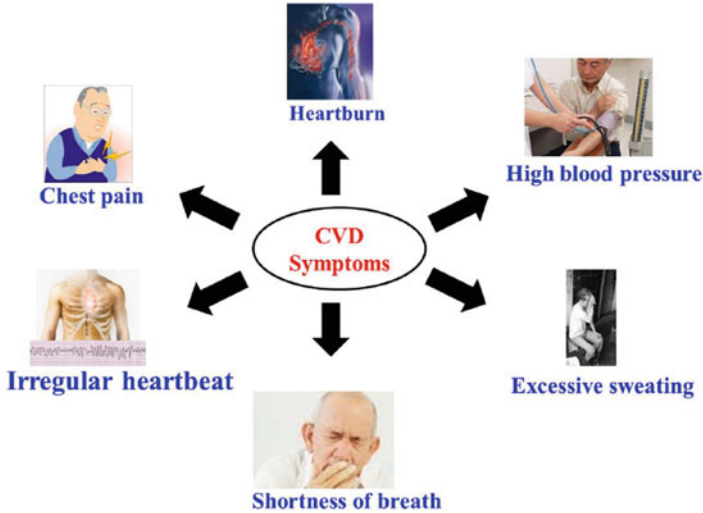


Fig. 1 Common symptoms of cardiovascular disease

appear periodically without showing any serious indication and are often getting unnoticed [2].

1.1 Diagnosis of Heart Diseases

In clinical diagnosis, the cardiac irregularities are mostly determined either in time series signal or image format. Although different imaging methods such as Magnetic Resonance Imaging (MRI), Computerized Tomography (CT) scan, Echocardiography (Echo), Nuclear myocardial perfusion scan, etc., provide reliable and accurate outcomes related to cardiac abnormalities, the acquisition method is time-consuming, labor-intensive, expertise-based and costly [2, 3]. In contrast, diagnosis through the Electrocardiography (ECG), Seismocardiogram (SCG), and Ballistocardiogram (BCG) those represent the activity of the heart in the form of signals are inexpensive, faster, and easily doable.

1.1.1 Diagnosis Through ECG Cardiac Signal

To monitor the physiological activities of the heart, ECG is considered as a well-accessible and widely adopted inter-mediator, where the signals can be easily captured through body sensors [2–5]. Those signal patterns help the clinicians to verify various cardiac electrical activities such as the movement of heart valves, blood circulation into ventricles, suppression-relaxation of ventricle walls, etc.

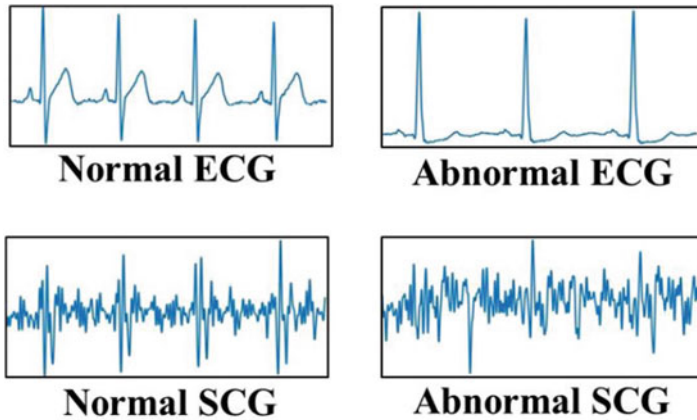


Fig. 2 Normal Vs. abnormal ECG and SCG signals [3]

However, based on the collected evidence, ECG provides the least information on the functioning of various cardiac events related to complex CVD problems such as ischemia, arrhythmias, and infarction that distorts the predefined cardiac sequence as well as the rhythm [2, 3]. In contrast, such deformities even at a mild intensity level, can be well determined using different cardiac mechanical activity recording modalities like BCG and SCG. An example of normal vs. abnormal cardiac ECG and SCG signals is represented in Fig. 2.

1.1.2 Diagnosis Through SCG and BCG Cardiac Signal

The dual modalities like BCG and SCG can measure the vibrations generated from various cardiac mechanical activities in successive heartbeats [6–11]. The SCG data can be acquired from an inexpensive accelerometer sensor placed over a person's body without direct contact with the skin. As far as BCG is concerned, the data acquisition or recording is quite different in terms of sensor placement. For example, BCG sensors can be wrist-worn for a regular interval of data acquisition or can be fitted in objects like bed, chair, weighing scale to acquire the signal at a specific interval [7]. Further, previous literate studies [6–9] have given more attention to SCG than the BCG as there is growing confidence in the accuracy and applicability of SCG in clinical practice. Moreover, the SCG is much popular in research and implementation in compared to BCG as it is inexpensive, non-invasive, convenient and adopts hassle-free data collection methods through wearable devices [10, 11].

Apart from the individual analysis, many researchers have designed the cardiac abnormality prediction models considering both electrical and vibrational signal recording modalities, i.e., ECG and SCG, respectively. Like ECG, SCG signals can be collected through single or multiple channels based on the number of deployed sensors in the body [2, 3]. In case of single channel, the body sensors are only

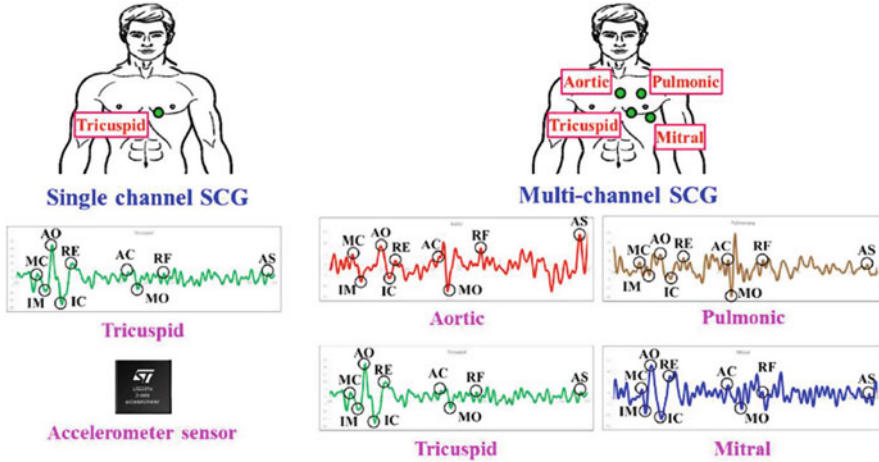


Fig. 3 Single channel Vs. multi-channel SCG signal

deployed in one location of the heart, i.e., *Tricuspid* valve. For multi-channel SCG signals, the sensor data are collected from different locations such as *Aortic* valve, *pulmonic* valve, *Tricuspid* valve, and *Mitral* valve as shown in Fig. 3.

1.2 Early Detection Through Smart IoT Devices

Recently, the Internet of Things (IoT) has made tremendous progress in healthcare, especially for providing early warning via various smart wearable devices such as smart watch, smart t-shirt, smart belt, etc., in case of any cardiac issues [12]. Those smart IoT devices are attached with a person’s body for continuous monitoring of the heart rate, rhythm, and state of the heart muscle tissue. Currently, one can see the electrical activity of the heart on the screen of the smart watches like Apple smart watch, Fitbit Sense or Samsung’s ECG-packing. In addition, its associated APP will help to know whether the rhythm of heartbeat is normal or abnormal [13]. Besides, the deadly atrial fibrillation (abnormally fast heartbeat) condition can be easily identified using such powerful smart devices. The usability of those smart systems is too simple where a person only needs to press the touch-sensitive button for 30 sec to know the behavioral pattern of the heartbeat. Apart from the smart wristwatch solution, the researchers in [14] have developed a smart shirt named S-WEAR using the concept of AI for monitoring cardiac ECG. The intelligent framework is responsible to collect the data from the body-embedded sensors and send them to the S-BOX database via different low-power wide-area network (LPWAN) technology such as LoRa, NB-IoT, and Sigfox. The S-WEAR and S-BOX both are having Artificial Intelligence (AI) components for quick local prediction and determination of anomalous using historical data. Upon finding any abnormalities, the analyzed

data and its corresponding values are transmitted to the consulting person through the APIs.

1.3 Role of AI in Cardiac Abnormality Detection

In recent years, Artificial Intelligence (AI) has been widely adopted in the medical field for developing the clinical expert system by processing a large amount of data and inferring meaningful outcomes. Those inferred clinical results assist clinicians in making accurate and quicker decisions for better treatment and follow up strategies [15]. With the involvement of AI learning models in the form of Machine Learning (ML), Deep Learning (DL), the workload of the cardiologist is decreased and at the same time computation efficiency and disease diagnosis accuracy are increased. The conventional methods of cardiac abnormality detection are mostly time-consuming, error-prone, and experience-based. The powerful, intelligent ML algorithms such as Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR) are used for the analysis of the influential clinical parameters and their combinations for the design of prognosis models. Those intelligent frameworks will be helpful for disease prediction, medical knowledge extraction, outcome prediction, therapy planning, patient support, and data management [16]. DL is a part or subset of ML comprises various advanced algorithms such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Auto Encoder (AE), Deep Boltzmann Machines (DBM), Deep Neural Networks (DNN), etc. [17], which operate on multiple levels of abstraction, and automatically extract features from a large set of medical images assisting doctors for faster diagnosis of the cardiovascular diseases.

The rest of this chapter is organized as follows. A detailed study of cardiological signal acquisition methods is introduced in Sect. 2. Section 3 highlights the analysis of a wide range of mathematical and intelligent AI models for cardiac abnormality prediction considering both ECG and SCG signals. Summary of the literature related to the current development in the cardiac data analysis is presented in Sect. 4 and concluding remarks are made in Sect. 5.

2 Cardiac Signal Data Acquisition

For continuous monitoring of the cardiac activities, the ECG and SCG data need to be recorded continuously either using smart wearable devices, body sensors or in the laboratory via different hardware modules [18]. Generally, the ECG and SCG body sensors transmit their data to any cardiac health analytic platform (HAP) through Body Area Network (BAN) as shown in Fig. 4. In wearable sensors, the sensing cardiac data are transmitted to an upper layer for analysis and feedback via various wireless communication technologies such as Bluetooth, Near

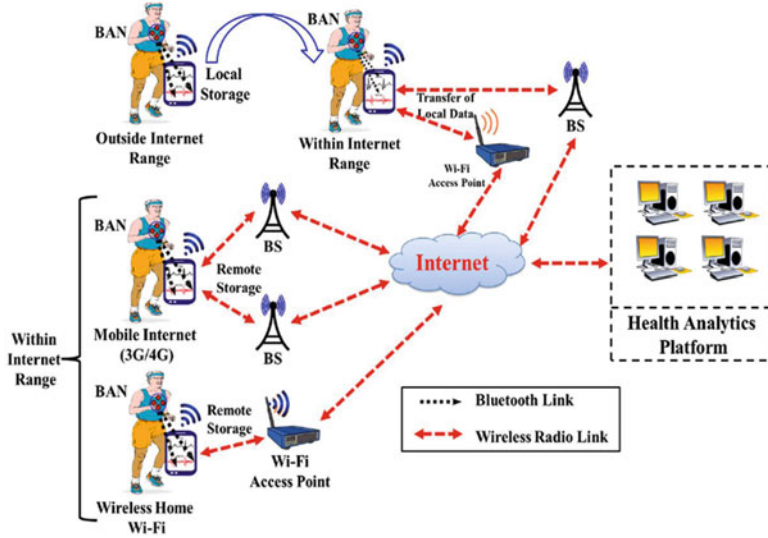


Fig. 4 Communication framework for cardiological signal transmission [3]

Field Communication (NFC), ZigBee, Cellular, and Wi-Fi [3]. Normally, a data communication module acts as a local gateway and is responsible for coordinating the communication between the BAN and HAP. To transmit the cardiac data without any interruption and with utmost reliability, the authors in [3] have considered two scenarios based on the user’s current location such as (a) Within Internet range: if the user is within the Wi-Fi connectivity or cellular network coverage; (b) Outside the Internet range: if the user is within the communication range of Bluetooth but out of range from the Wi-Fi access point or cellular coverage. To prevent any data loss in case of an Internet outage scenario, the raw data are stored in the user’s private smart phone or tablet device temporarily. The stored data will be transmitted immediately upon establishing the internet connectivity. Currently, there are lots of open cardiological data sources are available for analysis and experiment purposes.

2.1 Data Acquisition Through Wearable Gadgets

Recently, there are several small and energy-efficient wearable devices such as smart band, smart belt, smart cloth and smart helmet as shown in Fig. 4 are available [3], which can retrieve the cardiological data during different physical activities and predict the abnormality. The smart devices are well-equipped and beneficial for monitoring purposes such as heart rate or blood pressure. However, they have limited accuracy in complex cardiac abnormality prediction such as ischemia, arrhythmias as they do not consider the location-specific data.

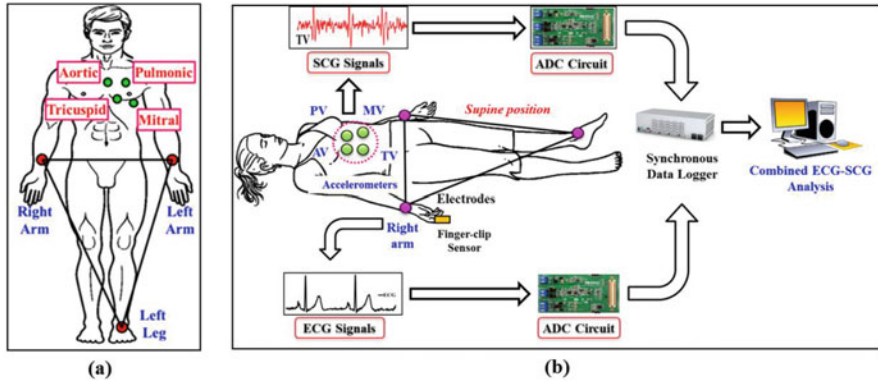


Fig. 5 Location of body sensors and data acquisition framework [2, 3]. (a) Location of ECG & SCG sensors. (b) ECG/SCG data acquisition method

2.2 Data Acquisition Through Body Sensors

To acquire qualitative and accurate cardiological data, the appropriate selection of the body sensor’s location is highly important. Hence, it is necessary to pay extra attentiveness while choosing the sensor location during data collection. Generally, the location of the sensors is selected under the supervision of an expert cardiologist. In case of ECG data collection as shown in Fig. 5a, the body sensors in form of electrodes are placed on the left arm, right arm, and left leg, respectively [3]. On the other hand, four accelerometer sensors are placed at different valvular auscultation sites such as Aortic, Pulmonic, Tricuspid, and Mitral valves for multi-channel SCG data collection as depicted in Fig. 5a [3, 19].

2.3 Data Acquisition in Laboratory

An architectural view of single-channel ECG and SCG data collection procedure in the laboratory is shown in Fig. 5b [2]. In order to collect single-channel SCG signal, the accelerometer sensing module LIS331DLH from STMicroelectronics is placed on the Tricuspid valve region, whereas the single electrode H135SG Covidien from Biomedical Instruments is located at the right arm to acquire the ECG data. Besides, the bandpass filter with frequency 0.5Hz–50Hz is applied to obtain the required SCG and ECG signals at a sampling frequency of 100Hz. Further, the communication between ECG/SCG sensing module and the Analog-to-Digital convert (ADC) circuit is established through the micro-controller system. In addition, the primary job of the attached synchronous data logger is to further amplify and filter the concurrent signals transmitted from the ADC circuit. Sometimes, there is a high chance of signal distortion due to mechanical noise during the signal transmission,

which can be taken care of using nonlinear filters. During the entire data acquisition procedure, it is necessary to monitor the subject's stability by constantly monitoring the heart and respiration rate. In the laboratory, generally, the respiration rate is checked manually, whereas the heart rate is observed using a fingertip sensor such as PAH8001EI-2G.

2.4 Open Source Data

Currently, several public data sources related to ECG/SCG based cardiological data analysis are available apart from the real experimental data. The open availability of data helps the researchers to expedite the research and innovation. The popular publicly available data sources for ECG/SCG data analysis are MITDB database [4], European ST-T Database [20], and PhysioBank [21] that comprises several other databases such as MIT-BIH Arrhythmia, QT database, American Heart Association ventricular arrhythmia, INCART, Long-Term-ST, UCI Machine Learning: Arrhythmia dataset, Breathing and Seismocardiograms (CEBS) dataset. Each database contains thousands of ECG/SCG data with several features, attributes and correlations for precise prediction of any cardiological abnormalities. In addition, several universities such as Huazhong University, iRhythm Technologies/Stanford University, University of California, other institutions like Telehealth Network of Minas Gerais, Mayo Clinic, Geisinger, Health eHeart Study, China Physiological Signal Challenge 2018 and Cleveland Clinic from USA are published their data banks open and available to make the AI-based analysis more reachable and applicable [22].

3 Cardiac Signal Data Analysis

The conventional methods of cardiac data analysis for predicting the abnormalities are mostly tedious, time-consuming, error-prone and intuition-based. Therefore, many researchers have developed novel ways of automatic heart disease prediction using robust mathematical and intelligent AI models [2, 3, 23–30]. In this section, several methods related to feature point delineation and abnormality prediction from ECG and SCG signal data are discussed as follows.

3.1 ECG Signal Data Analysis

Although, the signal pattern of ECG is not complicated, the detection of the signal irregularity considering those feature points is challenging especially when the abnormality is mild. It is always difficult for the experts to determine those

depression and elevation points manually, which is also time-consuming and tedious. Therefore, mathematical algorithms and AI models need to be developed for automatic ECG based feature points delineations and cardiac health monitoring.

3.1.1 Mathematical Model Based Prediction

To predict the cardiovascular abnormalities from the ECG, it is necessary to extract the valuable feature points that can help in correct differentiation of normal and abnormal cardiac activities.

1. Feature Point Delineation in ECG Signals: Generally, ECG represents different activities of the heart in terms of electrical signals. Commonly, the normal functioning of the heart is determined via five important points described as P, Q, R, S, and T delineated from a complete ECG cycles as shown in Fig. 6a. The accurate determination of any cardiac abnormalities highly depends on the precise and accurate delineation of the successive changes in the above mentioned primary points. However, the correct identification of those important points is really challenging due to the presence of signal artifacts and the variability in the position of the points in the ECG plot. Therefore, currently many researchers have proposed automatic methods for detection of those five meticulous points and the deviation of the ECG signal continuity [2, 3, 23].

The authors in [3], have developed a novel way of automatic feature point extraction using the sliding window (SW) concept. The $SW(X)$ contains a set of feasible data points as $X = \{P, Q, S, T\}$ considering the point R as the reference, which is having the maximum positive amplitude in the ECG cycle. The points like P, Q, S and T are fetched from consecutive $R - R$ duration considering the referenced $SW(X)$. After the capture of those primary points, another set of markers comprising set of onset range $\{Range_{onset}(P), Range_{onset}(QRS), Range_{onset}(T)\}$ and set of end range $\{Range_{end}(P), Range_{end}(QRS), Range_{end}(T)\}$ are retrieved. Those marker points are derived from set of candidate points like $P_{onset}, QRS_{onset}, T_{onset}$

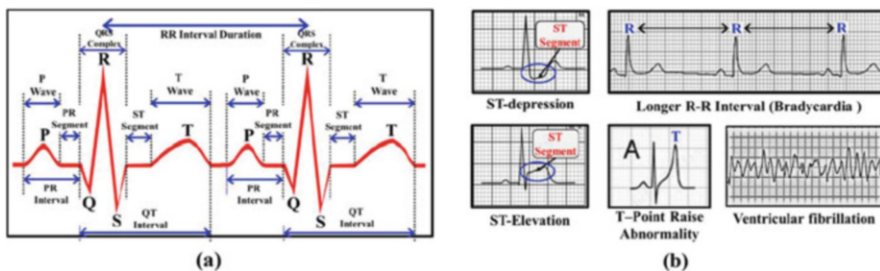


Fig. 6 Example of various cardiac abnormalities in ECG [3]. (a) Normal ECG. (b) Abnormal ECG

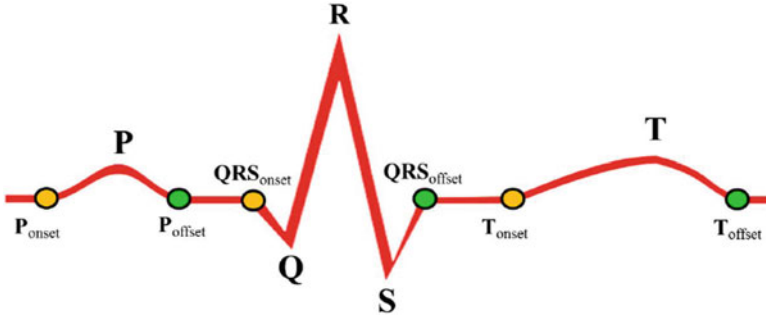


Fig. 7 Example of different onset and end points of ECG [3]

and set of end points P_{end} , QRS_{end} , T_{end} noticed in normal ECG cycles as shown in Fig. 7. To design a robust cardiac abnormality detection method, other valuable features are designed considering the location and order of the feature points in terms of interval such as PR Interval, QT Interval, RR Interval and segments, i.e., PR Segment, ST Segment, respectively. The retrieved value from those intervals is compared with the standard values to find out any abnormality. Different abnormal cardiac activities considering various features of ECG data represented in Fig. 6b. For example, the normal PR Interval lies in between the range 100 ms to 200 ms in case of a healthy patient, however, the interval value longer than 200 ms signifies the first degree of heart blocking and shorter than 100 ms indicates pre-excitation syndrome. Besides, different depression and elevation features of ECG signal curve such as ST Depression, ST Elevation, T Point raise Abnormality, Longer RR Interval and Ventricular fibrillation as shown in Fig. 6b are also considered during the analysis. Considering the amplitude, areas and angles of those depressed or elevated points, the authors in [3] have designed novel mathematical models to detect the cardiac abnormality.

Another notable feature generation and selection method is proposed in [2], where different ECG data points P , Q , R , S , and T are collected in the form of vectors V_{ecg} , where the sampling rate S_r and mean heart rate H_r are already known. Each retrieved ECG feature point in V_{ecg} signifies particular amplitude value in terms of millivolts (mV). Those feature points are selected from consecutive cardiac cycles, where the feature points of one cardiac cycle is separated from another with a cardiac cycle length CL . The value of the CL can be calculated as $CL = \frac{1}{H_r} * S_r$. Especially, the H_r parameter is retrieved continuously considering the successive RR interval duration. In the process of feature point extraction, first the candidate R point is detected considering the following analysis. In the first step, the peak with maximum amplitude ζ_{pt} is identified from all the considered cardiac cycles. In the next phase, the obtained peaks with amplitude value greater than $\sigma R_{pt} \times \zeta_{pt}$ are selected and defined as

cnR_{pts} . The other feature points such as Q_{rg} , P_{rg} , S_{rg} , and T_{rg} are calculated considering the candidate R point. The points, which are appeared before R is formulated as $(R_{pt} - Y, R_{pt})$ and the points appeared after R defined as $(R_{pt}, R_{pt} + (Y + \alpha))$. Here, Y signifies the set of data points, which is equal to the normal signal wave duration of the related feature points. For example, for the formulation of feature point Q , the value of Y is defined as $= \frac{\Delta QRS_{wv} * S_r}{2}$. Here, ΔQRS_{wv} signifies the time duration of the QRS wave in case of normal case. The constant α is added as a precautionary measure to minimize the estimation error of Y . In the similar way, the rest of the feature points are extracted considering their minimum (e.g., points Q and S) and maximum peak characteristics (e.g., points P and T) as shown in the Fig. 8. In addition, the authors have also considered the set of start points $\{P_{onset}, QRS_{onset}, T_{onset}\}$, and set of end points $\{P_{offset}, QRS_{offset}, T_{offset}\}$ corresponding to the feature points P wave, QRS wave and T wave, respectively.

A real-time ECG feature points R and QRS detection method is proposed in [23] considering the concept of derivative filters. The designed system consists of several modules such as R -point detection module, then a module for noise or artifact detection, a detected $R - R$ point connection module, and finally an elimination module for inaccurate RR determination. First, the referenced R point is delineated considering two types of first order derivative filter such as $[1, -1]$ and $[-1, 1]$. The filter $[-1, 1]$ is responsible to detect the rising interval of QRS wave while the sudden falling interval is determined by the filter $[1, -1]$. Through the individual multiplication of each derivative filter, the P and T waves along with the noise is suppressed with an enhancement of QRS interval. Finally, the normalized max filter is applied with those multiplied derivative filter outcomes to detect the QRS interval and R peak by keeping the threshold value $= 0.3$.

2. Detection of Abnormalities in ECG Signals:

The cardiac abnormalities from the ECG data can be identified considering prominent features such as ST Depression, ST Elevation, T Point raise those appeared in the ECG curve [3].

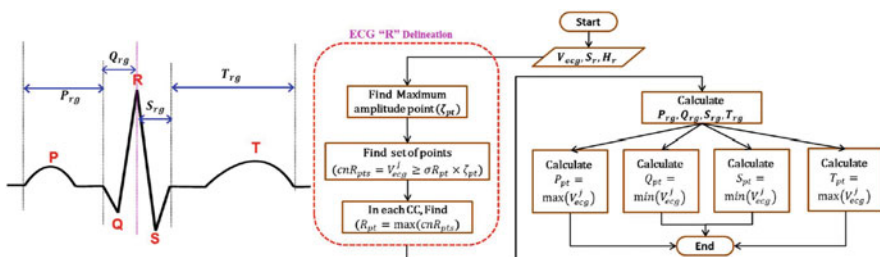


Fig. 8 Example of important points delineation in ECG wave

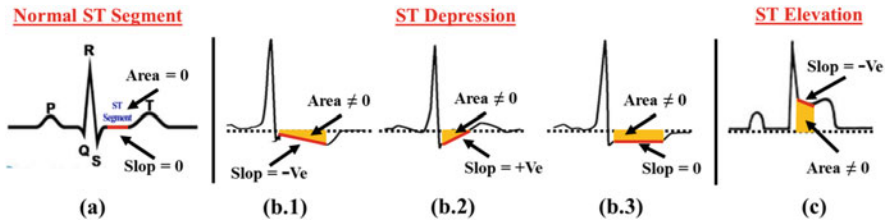


Fig. 9 Abnormality detection in *ST* Segment (a) Normal *ST* Segment; (b) *ST* Depression; (c) *ST* Elevation [3]

• **ST Segment Abnormality Detection:**

In case of a normal ECG, the *ST* segment is defined as an isoelectric flat portion joining the offset part (end part) of the *S* wave, i.e., *J* point with onset point of *T* wave (start point) as shown in Fig. 9a. Generally, the slope of the *ST* portion in a normal ECG is equal to 0 as it is a flat isoelectric line. However, in case of any abnormal circumstance, the *ST* segment bow towards downward (depression) or upward (elevation) direction with a changing in the default slope value to -ve ($m_{st} < 0$) or +ve ($m_{st} > 0$), respectively. Different types of *ST* depression and *ST* elevation examples are shown in Fig. 9b.1–b.3.

The m_{st} value is calculated considering the angle θ between the standard line and *ST* segment, which is defined as $m_{st} = \tan(\theta)$. With the use of this slope concept, the physicians can ascertain majority of *ST* segment related abnormalities. However, in some cases the changes in the slope value is almost negligible (remains zero) in spite of *ST* depression or elevation as shown in Fig. 9c. To handle such type of scenario, the authors have used another parameter, i.e., the area (σ_{st}), which is obtained due to curvature of *ST* segment with respect to the baseline. The area (σ_{st}) can be calculated by applying the mathematical tool of definite integration between *J* point and onset point of *T* wave as formulated in Eq. 1.

$$\sigma_{st} = \int_a^b f(x)dx \tag{1}$$

Here, $f(x)$ signifies the nature of *ST* segment curve while *a* and *b* are points equivalent to offset part of the *S* wave and the onset point of *T* wave.

• **T-Wave Abnormality Detection:**

The cardiac abnormalities like coronary ischemia, hyperkalemia and left ventricle hypertrophy disorder are mostly occurred due to ventricle re-polarization. In ECG, this re-polarization process can be determined considering the morphology of *T* wave as shown in Fig. 10b.1–b.3 in the form of *T* point raise, flattened *T* wave and inverted *T* wave, respectively. It is very difficult to uncover those *T* wave deviations using manual approach, therefore, the authors have developed a novel approach combing three primary

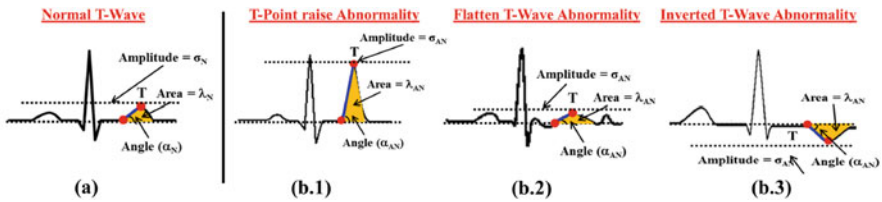


Fig. 10 Different abnormality detection from T Wave (a) Normal T-Wave; (b) Various T Wave abnormalities [3]

Table 1 The predefined normal values for different ECG wave types [3]

Notation	Definition
$D_{P_{wave}}$	Normal P wave Duration (80 ms)
$A_{P_{wave}}$	Normal P wave amplitude (0.1 mm, 0.2 mm)
$D_{QRS_{wave}}$	Normal QRS wave Duration (80 ms, 100 ms)
$A_{QRS_{wave}}$	Normal QRS wave amplitude (≤ 1 mm)

factors such amplitude, area, and angle related to T wave. It can be observed from Fig. 10b.1, the values of amplitude, area, and the associated angle are abnormally higher in comparison to the values obtained from normal T wave. In contrast, a relative smaller amplitude, area and angle are obtained due to flatten T waves in Fig. 10b.2 and the inverted T-wave can be marked when amplitude and angle of the T-wave gives value < 0 as shown in Fig. 10b.3.

• **RR Interval Abnormality Detection:**

The accurate determination of RR interval plays an important role for cardiac abnormality detection as in case of any irregularity the duration of RR interval goes longer. The usual RR Interval duration in a healthy heart ranges between 600 ms to 1000 ms. Any RR Interval duration longer than 1000 ms and shorter than 600 ms can be classified as abnormal. Although, during some activities like sleeping or sitting the RR interval duration is greater than 1000 ms, which is quite normal. However, such values in case of walking or running are not acceptable.

• **Other Abnormalities Detection:**

It is also observed that any changes in amplitude and the duration of the P and QRS wave triggered the cardiac abnormalities. Those amplitude and duration parameters can be calculated as $A_{P_{wave}}$, $A_{QRS_{wave}}$ and duration $D_{P_{wave}}$, $D_{QRS_{wave}}$ considering the time stamp and location of the primary feature points P, Q, R, and S. Whenever the calculated values either related to the amplitude or duration differs significantly corresponding to the normal values given in Table 1, the abnormality associated with the concern wave is noted.

• **Abnormality Detection in ECG Morphology:**

The authors in [2], have developed a mathematical model to detect the cardiac abnormality considering the reference amplitude and duration of the

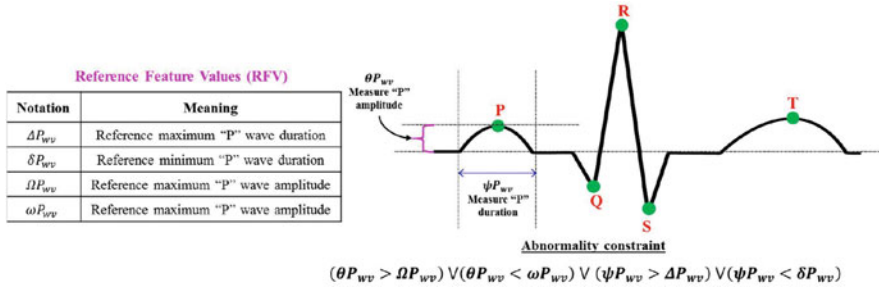


Fig. 11 ECG P point abnormality detection

feature points related to P wave, QRS complex, and T wave. In this work, instead of monitoring the individual cardiac cycle, the authors considered a group of k -number of cardiac cycles to design a robust cardiac abnormality detection method without being misguided with false abnormalities or bad signal quality. First, in each cardiac cycle, different potential parameters such as signal amplitude, wave segment, and interval duration related to each primary feature point is measured and compared with the corresponding normal values. Let us consider the methods of identifying the abnormality in the feature point P . In order to find out any deviation from the normal value, the current captured value of amplitude θP_{wv} and wave duration ψP_{wv} are compared with the reference minimum (ΩP_{wv}) and maximum (ωP_{wv}) value of amplitude as well as wave duration (i.e., ΔP_{wv} and δP_{wv}), respectively, as represented in Fig. 11. An abnormal morphology counter named as cP_{wv} is maintained related to P wave and the value of the counter is increased by one in each time upon finding any of the measured amplitude θP_{wv} or wave duration ψP_{wv} falls outside the reference normal values. In the similar way, the abnormality related to Q , R , S , and T wave can be calculated considering the morphology of QRS_{wv} , T_{wv} and RR_{inv} .

3.1.2 AI-Based Abnormality Prediction

In order to minimize the workload of the cardiologists and to increase computational efficiency, computer-aided systems have been developed using the concept of AI. Generally, the intelligent system is embedded with powerful ML or DL algorithms that can extract the hidden features and convert those salient features into useful insights that will benefit for heart disease prediction and severity analysis. Besides, as CVD is considered as one of the deadly chronic disease, an accurate and faster diagnosis is necessary.

An effective intelligent model has been developed for the prediction of heart disease using the concept of ML in [24]. The overall framework of the proposed AI model is given in Fig. 12. In the proposed work, the authors have used open data set

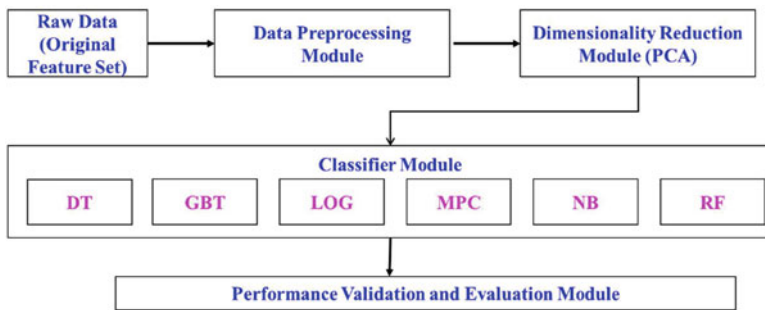


Fig. 12 Proposed ML framework for heart disease classification

named “Heart Disease Dataset” from the UCI Machine Learning Repository. The considered dataset contains in total 74 independent features such as blood pressure, heart rate and many other features related to ST depression, etc. Those data which are extracted from different races like Cleveland, Hungarian, and a combination of both called CH (Cleveland-Hungarian). After the necessary data collection, the raw data are preprocessed as it contains missing values, duplicate records and irrelevant data which might affect the prediction accuracy. Therefore, to make a complete and relevant data set, different cleaning process is performed such as (i) the categorical values like Yes or No changed into numerical 1 and 2 (ii) the null values are included with a unique level (iii) any zero value is changed to null and so on.

In the next step, the high dimensionality of the feature set, i.e., 75, is reduced to 13 important features using the concept Principal Component Analysis (PCA). In the ML, the high dimensionality of the data set is one of the biggest problems. The huge number of multi-variant features results unnecessary memory wastage and sometimes leads to overfitting issues. For dimensionality reduction, the authors have used PCA where the eigenvalue factor is considered as the criteria for effective feature selection. Any component value with an eigenvalue greater than 1.00 is included in the selected feature list. In contrast, any components with eigenvalues less than 1.00 were deleted from the analysis as their contribution is less. In the final step, for the prediction of the heart disease, the selected features are given as input to six different classifiers such as Decision Tree (DT), Gradient-boosted Tree (GBT), Logistic Regression (LOG), Multilayer Perceptron (MPC), Naïve Bayes (NB) and Random Forests (RF), respectively. It is found that among all the considered classifiers, the RF is performed better, with 98.7% accuracy.

The ML algorithms are used in all the fields of medicine, such as drug discovery, clinical decision making, significantly altering the way medicine is practiced. However, the ML algorithms use the numeric data as input; therefore, all types of data must be converted to numeric form. This is achieved by extracting the features manually from the data, which is tedious, manual, and error-prone. In addition to this, when only known features are extracted manually, some important unknown feature might be lost such as the spatial relationship between two adjacent pixels.

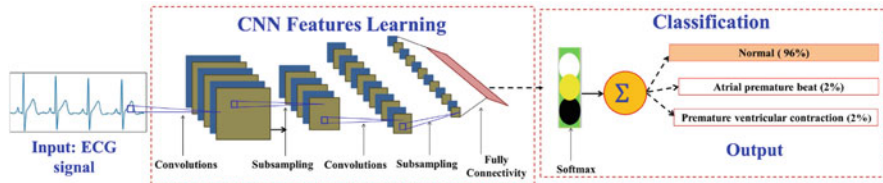


Fig. 13 CNN framework for cardiac abnormality detection

Therefore, the automatic feature extractor, DL is used to overcome the limitations of traditional ML algorithms.

Upon understanding the benefits of the DL, the authors in [25] have developed a CNN-based classification network that can classify the cardiac abnormality in different classes such as “normal,” “atrial premature beat,” and “premature ventricular contraction ” using ECG. The authors have considered 48 numbers ECG recordings collected from the PhysioNet database. The proposed CNN network comprises convolutional layers, batch normalization layers, pooling layers and finally, the Softmax classification layer. In each layer, the Rectified Linear Units (ReLU) is used as the activation function. As shown in Fig. 13, the collected ECG image is given as input to the designed CNN model.

The hidden layers perform the convolution, subsampling, pooling and finally, the output is forwarded to a fully connected layer. Here, all neurons in the fully connected layer vote for whether the input image is “normal” or suffered from “atrial premature beat” or “premature ventricular contraction”. After that, the voting is passed through the Softmax classification layer to classify a patient’s percentage being normal or abnormal. Even with the fewer number of the dataset and simple CNN network, the proposed model is achieved 98.33% mean accuracy for the validation set.

3.2 SCG Signal Data Analysis

The signals obtained from ECG sensors are enriched with in-depth cardiac information. However, there is a need for a detailed study related to different SCG features and their correlation for cardiac abnormality prediction. Currently, several mathematical and AI models have been developed to extract the candidate features automatically and to identify the morphological abnormalities in the retrieved cardiac SCG signals.

3.2.1 Mathematical Model Based Prediction

It is difficult for the physicians to draw any outcome from the raw SCG signal that requires proper annotation considering the cardiac cycle’s beat-by-beat. Besides, proper identification of the candidate feature points can result in efficient abnormality prediction as those features will ultimately use in various intelligent models. The manual annotation of feature points is tedious, experience-based, and time-consuming process. Therefore, there is a requirement for an automatic feature extraction paradigm.

1. Feature Point Delineation in SCG Signals:

A mathematical model for delineation of the feature point considering single-channel SCG is discussed in [2]. The feature points are collected in the form of a vector and is defined as V_{scg} with known sampling rate S_r and mean heart rate H_r . For SCG data analysis, generally nine important points related to various cardiac mechanical activities such as peak of atrial systole (AS), closing of mitral valve (MC), isovolumic movement (IM), opening of aortic valve (AO), isovolumic contraction (IC), peak of rapid systolic ejection (RE), closing of aortic valve (AC), opening of mitral valve (MO), and peak of rapid diastolic filling (RF) are considered as shown in Fig. 14.

Out of those primary points, the data point AO is considered as the candidate points as it exhibits high amplitude in a cardiac cycle. At first, the maximum amplitude peak ζ_{pt} from the set of cardiac cycles is located. After that, for each cardiac cycle a set of candidate peaks are located as $cnAO_{pts}$ with amplitude more than $\sigma AO_{pt} \times \zeta_{pt}$. Considering AO as the candidate point, a sliding window $SW(X)$ is derived, where $X = \{AS, MC, IM, IC, RE, AC, MO, RF\}$. After the delineation of AO in each cardiac cycle, the remaining feature points are captured for each $AO - AO$ duration using the sliding window $SW(X)$.

In the next phase, individual feature points (X_{rg}) is identified according to their morphological maxima (+ve) or minima (-ve) characteristic.

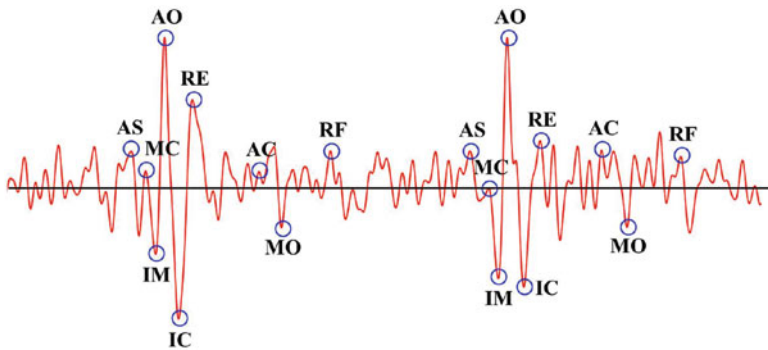


Fig. 14 Delineation of SCG feature points

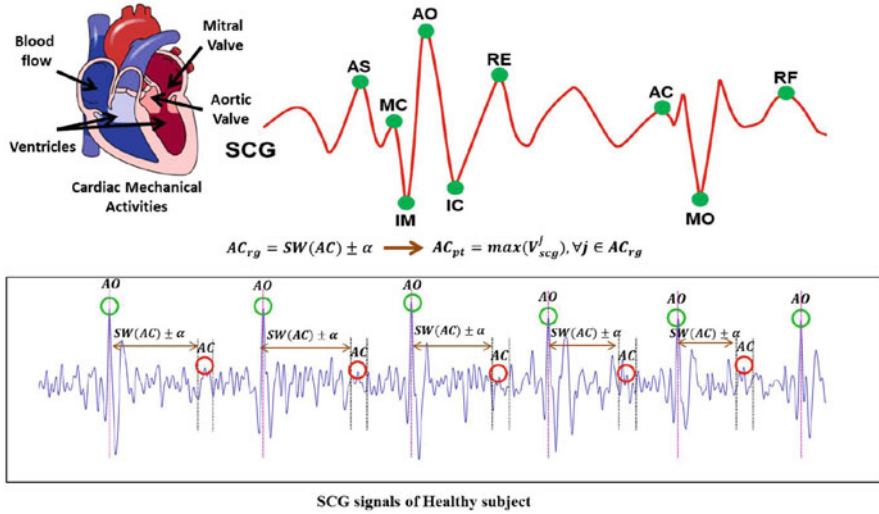


Fig. 15 SCG AC point abnormality detection

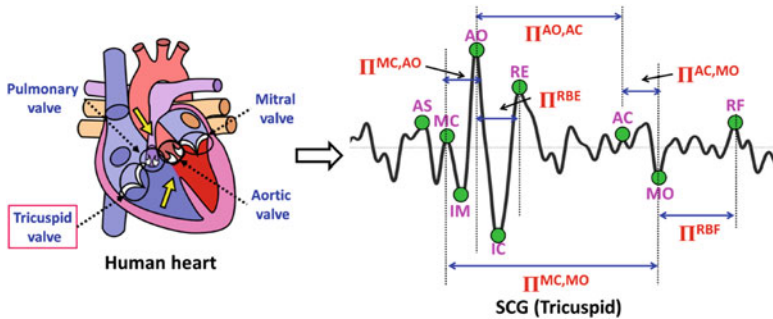


Fig. 16 Candidate feature variables of SCG derived from TV location [2]

Here, X_{rg} is defined as the range of points to locate point X, where $X \in \{AS, MC, IM, IC, RE, AC, MO, RF\}$. Let us consider an example as shown in Fig. 15, where to delineate the feature point AC, probable data points (AC_{rg}) with respect to AO at distance $SW(AC)$ is formulated. Finally, the signal peak with maximum amplitude falls within the range of $AC_{rg} + \alpha$ is nominated as one of the candidate feature point (AS_{pt}) as depicted in Fig. 15.

2. **Abnormality Detection in Single-Channel SCG Signals:**

To determine the cardiac abnormalities from the single-channel SCG signal generated from Tricuspid valve, the authors in [2] have considered six SCG feature variables (FV_{SCG}) such as $\Pi^{MC,AO}$, $\Pi^{AO,AC}$, $\Pi^{MC,MO}$, $\Pi^{AC,MO}$, Π^{RBE} , and Π^{RBF} as shown in Fig. 16. As described in Table 2, the mentioned

Table 2 Description of single-channel SCG signal’s feature variables [2]

Notation	Meaning
$\Pi^{MC,AO}$	Time Duration from closing of mitral to opening of aortic
$\Pi^{AO,AC}$	Time duration between opening and closing of aortic
$\Pi^{MC,MO}$	Time duration between closing and opening of mitral
$\Pi^{AC,MO}$	Time duration from closing of aortic to opening of mitral
Π^{RBE}	Time duration of ventricle blood ejection
Π^{RBF}	Time duration of diastolic blood filling
$FV_{scg} = \{\Pi^{MC,AO}, \Pi^{AO,AC}, \Pi^{MC,MO}, \Pi^{AC,MO}, \Pi^{RBE}, \Pi^{RBF}\}$	

feature variables signify the duration of different cardiac mechanical activities in a cardiac cycle.

The entire process of cardiac abnormality detection is divided into two phases such as estimation phase and evaluation phase. The estimation phase is mainly focused on the time series analysis of data in order to smoothen any kind of signal fluctuation and measure the accurate trend of six feature variables. In the analysis, the behavioral changes factor such as respiration and body movement is also added by assigning certain weights in a decreasing order to the cardiac cycles. The weighted moving average duration $WavgD_i$ and weighted moving standard deviation duration $WstdD_i$ are quantified for individual feature-variable- i , where $i \in FV_{scg}$. At the end of the η number (let) of cardiac cycles, the calculated value of $WavgD_i$ and $WstdD_i$ are used as the decision maker variables for early cardiac abnormality detection. The derivations of estimation and evaluation phases are given details in [2].

3. Abnormality Detection in Multi-channel SCG Signals:

In order to predict the cardiac abnormality in multi-channel SCG, six SCG features such as D_{MC-AO} , D_{RBE} , D_{AO-AC} , D_{MC-MO} , D_{RBF} and D_{AC-MO} are identified as shown in Fig. 17 based on the order and position of nine SCG important points [3]. The six candidate feature points represent the different Cardiac Mechanical Activities (CMAs), such as opening and closing duration of the aortic and mitral valve, time for systolic blood ejection and diastolic blood filling, etc. In the case of a healthy person, the CMAs takes place in a regular pattern. However, in a patient with coronary diseases such as myocardial ischemia, infarction and arrhythmias, irregular pattern of CMAs are found with significant changes in the regular time interval of different SCG features. In Table 3, the definition of each SCG feature and their involvement in different cardiac activities are listed.

Unlike ECG, SCG does not carry any predetermined time interval for the signal waves. Therefore, for each SCG feature i , where $1 \leq i \leq 6$, a reference value of duration, i.e., D_i is estimated considering $\delta > 0$ (let us say $\delta = 20$) number of cardiac cycles. A reference moving average duration $\mu(D_i^k)$ and reference moving standard deviation $\sigma(D_i^k)$ is estimated from δ number of cardiac cycles, where $1 \leq i \leq 6$ and $1 \leq k \leq \delta$. The derivations of $\sigma(D_i^k)$ and $\mu(D_i^k)$ are given

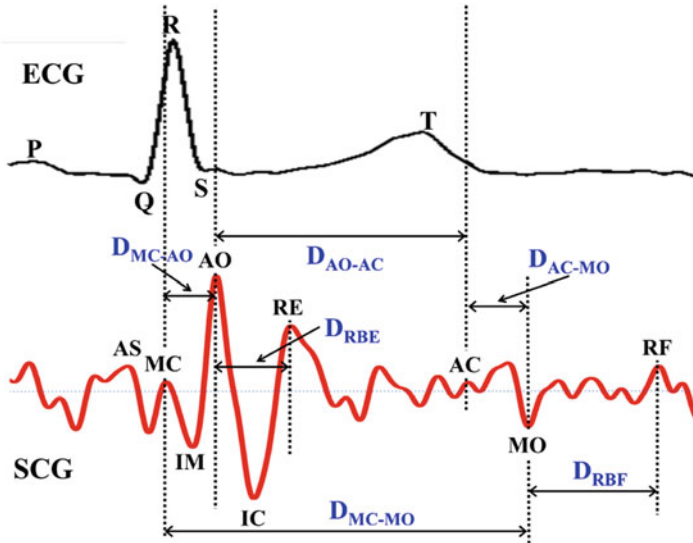


Fig. 17 Identification of SCG multi-channel feature points [3]

Table 3 Definition and cardiac activities associated with multi-channel SCG features [3]

Notation	Cardiac activities
D_{MC-AO}	Time interval from closing of mitral valve to opening of aortic valve
D_{AO-AC}	Time interval between opening and closing of aortic valve
D_{MC-MO}	Time interval between opening and closing of mitral valve
D_{AC-MO}	Time interval from closing of aortic valve to opening of mitral valve
D_{RBE}	Time interval of systolic blood ejection
D_{RBF}	Time interval of diastolic blood filling

in [3]. In case of multi-channel SCG, the general estimation way of $\mu(D_i^k)$ and $\sigma(D_i^k)$ are extended for multi-location SCG values that are placed in different valves such as TV , MV , PV , and AV , respectively.

For the SCG signals obtained from different locations, the reference moving average duration $\mu(D_i^k)_{TV}$, $\mu(D_i^k)_{AV}$, $\mu(D_i^k)_{MV}$, $\mu(D_i^k)_{PV}$ and the reference moving standard deviation $\sigma(D_i^k)_{TV}$, $\sigma(D_i^k)_{AV}$, $\sigma(D_i^k)_{MV}$, $\sigma(D_i^k)_{PV}$ could be calculated [3]. Finally, $\widehat{\mu(D_i^k)}$ and $\widehat{\sigma(D_i^k)}$ are obtained by averaging the value of $\mu(D_i^k)_{TV}$, $\mu(D_i^k)_{AV}$, $\mu(D_i^k)_{MV}$, $\mu(D_i^k)_{PV}$ and $\sigma(D_i^k)_{TV}$, $\sigma(D_i^k)_{AV}$, $\sigma(D_i^k)_{MV}$, $\sigma(D_i^k)_{PV}$, respectively, which is given in [3]. The final estimated values such as $\widehat{\mu(D_i^k)}$ and $\widehat{\sigma(D_i^k)}$ are considered as the decision makers to determine the abnormalities in successive cardiac cycles. In the evaluation phase, each individual

SCG feature D_i^j is compared to the range $(\widehat{\mu}(D_i^k) + \widehat{\sigma}(D_i^k), \widehat{\mu}(D_i^k) - \widehat{\sigma}(D_i^k))$ during each subsequent cardiac cycles j . If the measured value of D_i^j falls out the range of $(\widehat{\mu}(D_i^k) + \widehat{\sigma}(D_i^k), \widehat{\mu}(D_i^k) - \widehat{\sigma}(D_i^k))$ for any cardiac cycle j , the concerned i th SCG feature is considered as potential outlier and the considered j th cardiac cycle is taken as potential abnormality.

3.2.2 AI-Based Abnormality Prediction

Currently, the concept of ML and DL is very much popular in healthcare especially in biomedical signal processing, including SCG analysis to design various predictive models. Intelligent algorithms can be employed to recognize the underlying salient features automatically without the need for any cumbersome manual feature extraction. Considering SCG as the primary modality, various ML models such as SVM, LR can be used in detection of coronary heart diseases in a cardiac cycle. However, the efficiency of a robust model solely depends on effective feature annotation and extraction.

1. Automatic SCG Feature Annotation:

Unlike ECG, the SCG morphology is highly complicated and incomprehensible. Generally, the SCG possesses high inter variability of signal types among multiple subjects. In addition, the signal generated from SCG due to vibration is highly susceptible to external noise. Besides, it is always challenging to distinguish the peaks of candidate feature points accurately. In some cases like the feature point set $\{AO, RE\}$ and $\{IM, IC\}$, exhibit similar amplitude as shown in Fig. 14, hence difficult for the accurate feature delineation. Therefore, it is highly essential to design concrete automatic feature annotation method using AI that can learn the underlying feature from dynamic SCG signal [26].

The authors in [26] have developed an AI-based automatic SCG annotation framework, which is broadly divided into three phases such as preprocessing, training, and testing, as shown in Fig. 18. The preprocessing phase is focused on the identification of candidate features and peaks of SCG signal. During the training phase, based on the retrieved features, the selected classifiers are trained and learned the hidden features. Finally, in the testing phase, the undesired peaks (candidate points) are filtered out using the well-trained ML classifier. In order to train the classifiers, the authors have considered three morphological features as *Amplitude*, *Time of appearance* and *Count* derived from the SCG signal. The feature amplitude is defined with *+ve* (high) or *-ve* (low) value based on the maxima or minimal signal wave nature. For example, as shown in Fig. 14, the point *AO* and *AC* in one cardiac cycle are associated with high and low amplitude, respectively. Similarly, the feature *Time of appearance* is measured taking *AO* as the candidate. Based on the defined rule, the points which appeared before *AO* are assigned with *-ve* value and the coming points after *AO* assigned with *+ve* value in each cardiac cycle. The innovative feature count is described

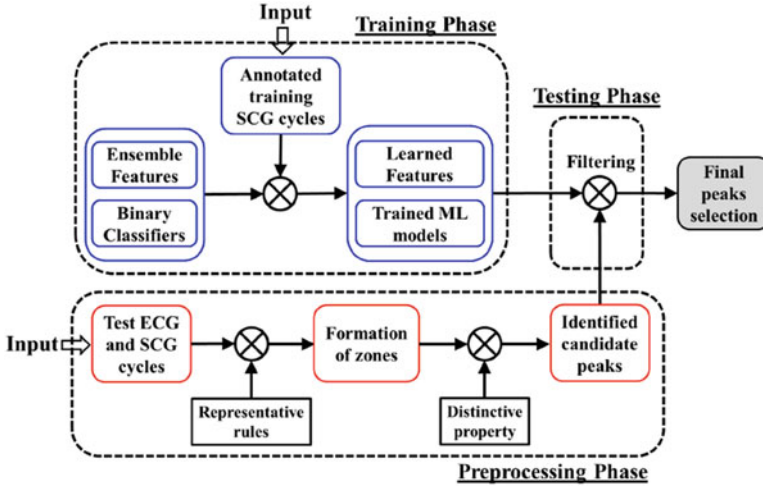


Fig. 18 Overview of SCG annotation framework [26]

as the number of *Upslopes* and *Downslopes* away from *AO* at a particular SCG peak.

In the preprocessing phase, to reduce the search area and to determine the candidate peaks, each cardiac cycle is divided into three zones as Zone 1, Zone 2, and Zone 3, respectively. The respective zones are constructed by comparing the electrical morphology of ECG with the mechanical morphology of SCG by mapping the associate points as shown in Fig. 19. The purpose of Zone 1 is to identify the feature point *AS* by aligning the ECG P_{offset} to QRS_{onset} with SCG. The other primary SCG points such as *MC*, *IM*, *IC*, and *RE* are detected by placing the ECG QRS_{onset} to T_{onset} with SCG. Finally, the other SCG points *AC*, *MO*, *RF* are determined by mapping the ECG T_{offset} to P_{onset} of the next cycle as represented in Fig. 19. After the division of zones and detection of the feature points, associate candidate peaks are identified as *max_peak* (e.g., *AS*, *MC*) or *min_peak* (e.g., *IM* and *IC*) based on the maxima or minima.

For the training purpose, three machine learning classifiers, such as NB, SVM, and LR, are employed by the authors to learn the morphological changes in signals and to capture the irregularity of the signal peaks. The reason to select NB, SVM, and LR is that they are less computer-intensive and highly robust against overfitting, making it beneficial for continuous cardiac monitoring. The considered ML classifiers require only a few parameters to tune, which helps in faster training and quick learning. An example of automatic candidate feature point detection using LR classifier is described as follows. Let us consider there are k numbers of training data samples are for a SCG peak $x \in SCGPs$, the set $Z_x = \{z_x^i | i = 1, 2, \dots, k\}$, where $z_x^i \in Z_x$ is an i th training data sample. Each training sample z_x^i is comprised of feature vector $F_x^i = \{C_x^i, T_x^i, A_x^i\}$. A logistic linear function $\mathcal{L}(z_x^i)$ can be defined as shown in Eq. 2.

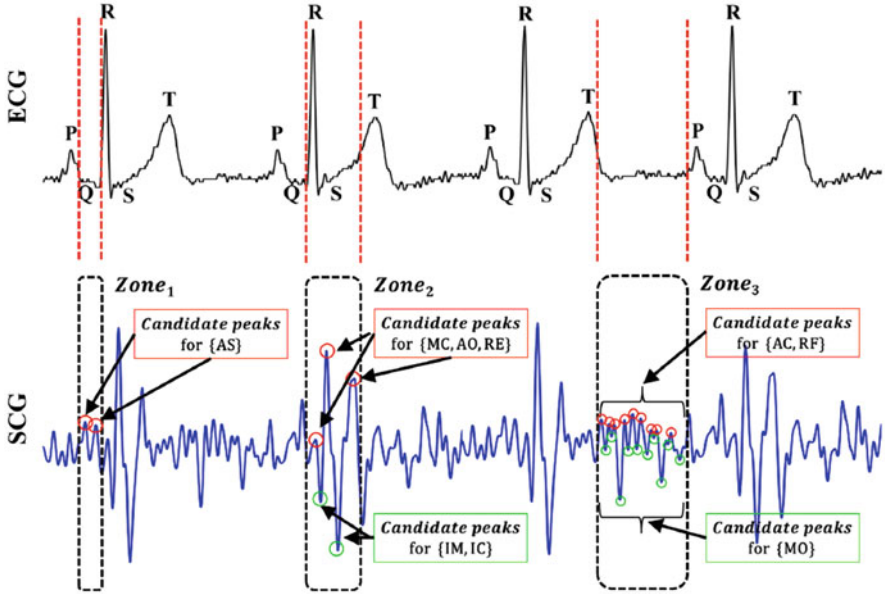


Fig. 19 The zones formation and candidate peaks identification [26]

$$\begin{aligned} \mathcal{L}(z_x^i) &= \beta_0 + (\beta_1 \times C_x^i) + (\beta_2 \times T_x^i) + (\beta_3 \times A_x^i) \\ &= \beta^T \times F_x^i, \forall z_x^i \in Z_x \end{aligned} \tag{2}$$

Here, $\beta_0, \beta_1, \beta_2,$ and β_3 are the earning parameters related to specific classifiers.

$$\sigma(\mathcal{L}(z_x^i)) = \sigma(\beta^T F_x^i) = \frac{1}{1 + \exp^{-(\beta^T F_x^i)}} \tag{3}$$

For testing purpose, let us consider there are n number of candidate peaks (i.e., test data samples) for an SCG peak x , which is defined as $CP_x = \{cp_x^1, cp_x^2, \dots, cp_x^n\}$. Using the trained logistic regression model defined in Eq. 3, to predict the likelihood of each candidate peak $cp_x^i \in CP_x$ to be classified in class *selected* considering the Eq. 4

$$\begin{aligned} p(selected|cp_x^i) &= \sigma(\mathcal{L}(cp_x^i)) = \sigma(\beta^T F_x^i) \\ &= \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 C_x^i + \beta_2 T_x^i + \beta_3 A_x^i)}} \end{aligned} \tag{4}$$

One candidate peak with maximum likelihood out of the n is selected as desired SCG peak x as shown in Eq. 5.

$$x = \max \left(p(\text{selected}|cp_x^1), p(\text{selected}|cp_x^2), \dots, p(\text{selected}|cp_x^n) \right) \quad (5)$$

2. Cardiac Abnormality Prediction:

When huge numbers of complex multi-parametric biosignal data are generated, it becomes a tedious job to correlate those parameters for deciding the cardiac condition of a person. Any error in diagnosis could be fatal for the patients. Therefore, potential ML and DL algorithms are used in all the fields of biomedical signal analysis to predict any kind of abnormalities rapidly and to assist the cardiologist in making faster treatment decisions. For multiple or binary classification, Artificial Neural Networks (ANNs) is the best option as it works pretty well if data is of huge size. ANNs are also a nonlinear model which makes it easy to use and understand, compared to statistical methods. ANNs are especially useful if the outputs are inter-related. ANNs learn on how to associate each of the inputs with the corresponding output, by modifying the synaptic weights of connections between neurons. The ANN, which is employed in this study consist of an input layer, two hidden layers, and an output layer.

In order to classify absence (*binary 0*) or presence (*binary 1*) of the Cardiac Artery Disease (CAD), the authors in [27] have used the concept of ANN considering the SCG modality. The authors have extracted 48 features (F_1, F_2, \dots, F_{48}) during three activities like *rest, immediate post-exercise and recovery* considering six candidate points such as *MC, AO, AC, MO, RE, and AS*, respectively. For example, *systolic interval Q to AC* is considered as one of the parameters and this interval value is recorded for the three actions like *rest, immediate post-exercise and recovery* to generate three features such as F_1, F_2 , and F_3 . In a similar way, by considering 16 SCG signal parameters for the three activities, 48 numbers of features are recorded. For the model training and testing in total, a population of 114 patients are considered, out of which 57 are diagnosed with CAD and the rest 57 are normal category. Those extracted features are given as input to the ANN composing input layer, an output layer and intermediate hidden layer for feature learning, as shown in Fig. 20.

In order to get the correct output, the back-propagation learning is employed where the weights (w_{xa}, w_{ab}, w_{by}) are updated. The Boolean values for each of the two outputs are responsible for predicting the presence or absence of CAD. The proposed model has achieved impressive results with 80% sensitivity and 70% specificity on the unseen testing set.

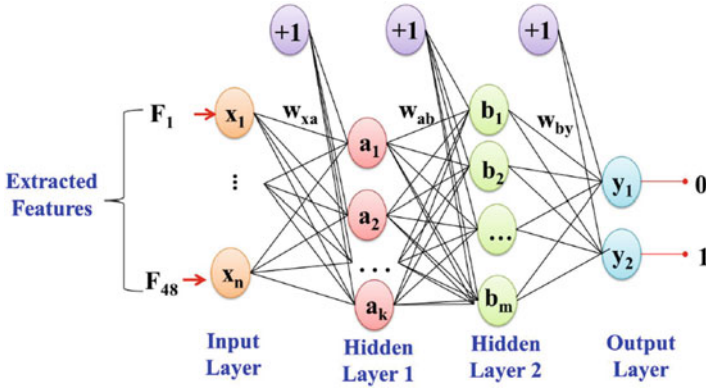


Fig. 20 ANN to predict the presence or absence of CAD

3.3 Combined Analysis of Multichannel ECG and SCG Signals

The prediction of cardiac abnormalities considering only a single modality of cardiac signal type, either ECG or SCG, is not sufficient as each signal type has its limitations. Although the ECG signal analysis is efficient for daily monitoring of cardiac activities, critical heart diseases like ischemia, angina, and blockage usually do not appear in the ECG. Unlike ECG, SCG can give depth knowledge related to mechanical functionalities of the heart. However, the analysis using SCG is not well-studied especially, the extraction of different features in case of abnormalities related to hypertensive heart disease. Therefore, there is a need for a detail combined study of electrical and mechanical activities of the heart through ECG and SCG. This combined study might bring out some additional information that will help early prediction of abnormalities and continuous monitoring of heart disease.

3.3.1 Mathematical Model Based Prediction

A combined analysis of ECG and multi-channel SCG is performed in [3], where the authors have designed a probability-based mathematical model to predict the abnormalities as represented in Fig. 21. In order to determine the severity over a period of time, the probability of abnormality as shown in Fig. 22 for a Group of π number (where $\pi \geq 1$) of Cardiac Cycles (GCCs) is considered. In the initial phase, the abnormality related to individual cycles is calculated, which is then averaged over π numbers of GCCs. The maximum probability value considered for any abnormality is 0.5 irrespective of ECG and SCG modality. In SCG, the maximum probability value of 0.5 is distributed over the four channels such as AV, MV, PV, TV and assigned $\frac{1}{4} \times 0.5 = 0.125$ to individual channels. During the prediction of the abnormality, the probability value of ECG and the SCG is summed up. As shown in Fig. 22c, in the cardiac cycle 3 (CC_3), the output probability of ECG is = 0.5;

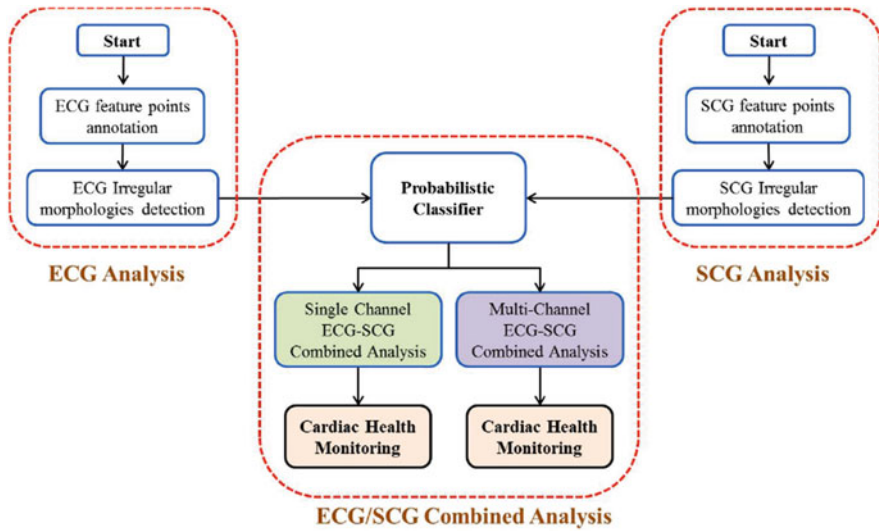


Fig. 21 Proposed framework for combined analysis of ECG and multi channel SCG [3]

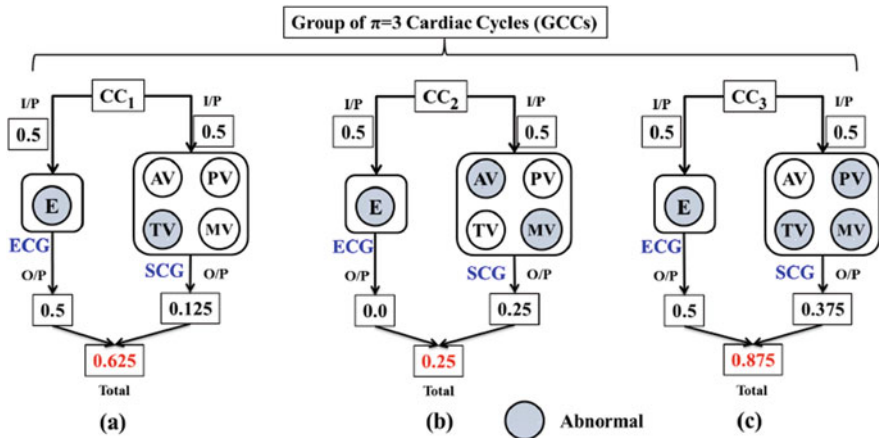


Fig. 22 Combined analysis of ECG and multichannel SCG [3]

however, in case of SCG the probability value = $0.125 + 0.125 + 0.125 = 0.375$ is obtained from affected *TV*, *PV*, and *MV* channels. In the next step, the probability of abnormality for GCCs (P_{GCCs}) is determined by averaging the attained value from each individual cycle over GCC ($\pi = 1$) i.e., $\frac{0.625+0.25+0.875}{3} = 0.58$ (Fig. 22). In the final step, to categorize the abnormality as mild or severe, the P_{GCCs} value is compared with predefined threshold values β_M and β_S , where $\beta_M, \beta_S \in [0, 1]$. The mild severity is confirmed if $\beta_M \leq P_{GCCs} \leq \beta_S$ and in case of higher severity the $P_{GCCs} > \beta_S$.

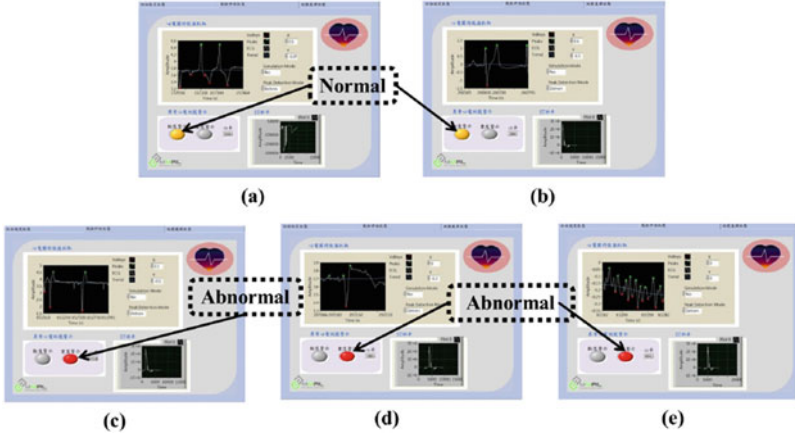


Fig. 23 Abnormality visualization module [3]. (a) ST-depression. (b) T-wave raise. (c) Bradycardia. (d) ST elevation. (e) Ventricular fibrillation

For visualization of the abnormality, the authors in [3] have developed an early warning system comprised of dual LEDs to predict mild and severe abnormality. Based on the probabilistic-based combined analysis, upon finding mild abnormality the Yellow LED glows up along with the attached motor’s vibration. Similarly, in severe cardiac health abnormality, the red LED will flash with a buzzer sound alert as shown in Fig. 23.

3.3.2 AI-Based Abnormality Prediction

The design of robust heart disease prediction model requires beat-by-beat analysis combing the outcome of both ECG and SCG as depicted in Fig. 24. The conventional methods mostly depend on the hand-crafted mathematical features, which is time-consuming, tedious, labor-intensive as they are trial-and-error based. Besides, there is a high chance of mistakes in the manual calculation of the abnormal probability especially a higher number of cardiac cycles. Therefore, there is a high requirement for intelligent models that can correlate the multiple parameters from both the ECG and SCG modalities to produce more accurate outcomes within less time.

An intelligent Naïve Bayes probability model is designed to combine the morphological features of both ECG and SCG by the authors in [2] to classify each cardiac cycle to be normal or abnormal. To construct the probabilistic model, the authors have considered a set of ECG features represented as $FV_{ecg} = \{\Theta X_{wv}, \psi X_{wv}, \psi RR_{inv}\}$ and set of SCG features defined as $FV_{scg} = \{\Pi^{MC,AO}, \Pi^{AO,AC}, \Pi^{MC,MO}, \Pi^{AC,MO}, \Pi^{RBE}, \Pi^{RBF}\}$, where $X \in \{P, QRS, T\}$. The terms Θ and ψ signify the value of wave amplitude and duration. The probability of an ECG cardiac cycle (Let us say k) to be normal or abnormal

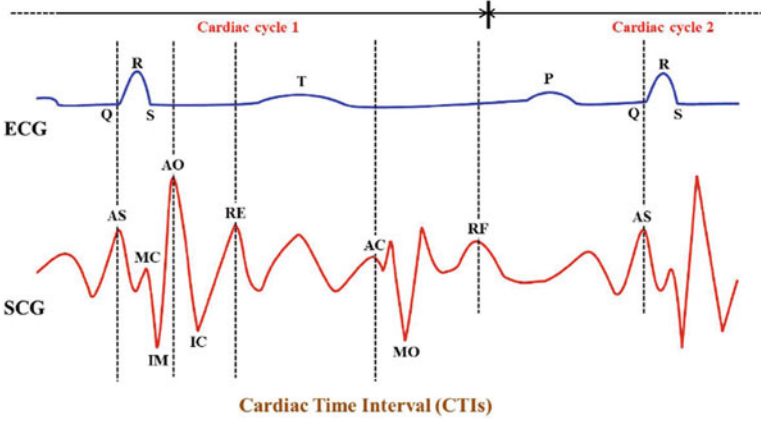


Fig. 24 Feature point mapping considering both ECG and SCG

can be defined in Eq. 6.

$$p(\varphi_l | FV_{ecg}^k) = \frac{p(\varphi_l) \times p(FV_{ecg}^k | \varphi_l)}{p(FV_{ecg}^k)}, \text{ where } l \in \{1, 2\} \quad (6)$$

Here, the output class *normal* and *abnormal* is signified by $\varphi_{l=1}$ and $\varphi_{l=2}$, respectively. The probability of k th cardiac cycle to be *normal* or *abnormal* for a given ECG feature set FV_{ecg}^k is determined by the variables $p(\varphi_{l=1} | FV_{ecg}^k)$ and $p(\varphi_{l=2} | FV_{ecg}^k)$. The $p(\varphi_l | FV_{ecg}^k)$ can be simplified as shown in Eq. 7.

$$p(\varphi_l | FV_{ecg}^k) = p(\varphi_l | \Theta X_{wv}^k, \psi X_{wv}^k, \psi RR_{inv}^k) \quad (7)$$

Each feature $x_i \in FV_{ecg}^k$ is considered to be conditionally independent to every other features $x_j \in FV_{ecg}^k$ for $j \neq i$ under conditional independence assumption the Naïve Bayes theorem. Based on the maximum a posteriori decision rule, the ECG classifier can be defined as given in Eq. 10.

The normal or abnormal classification of SCG signal can be derived using the Naïve Bayes conditional probability classifier as given in Eq. 8.

$$\Gamma_{scg} = \operatorname{argmax}_{l \in \{1,2\}} p(\varphi_l) \times \prod_{i=1}^{i=7} p(y_i | \varphi_l) \quad (8)$$

Here, Γ_{ecg} and Γ_{scg} is assigned with class label φ_l for some l (i.e., binary 0 or 1) based on the maximum a posteriori probability. Accordingly, if the output of both the ECG and SCG modality is abnormal (e.g., binary '1'), then based on Eq. 9, the concerned cardiac cycle is classified as abnormal with the rule, e.g., $1 \wedge 1 = 1$.

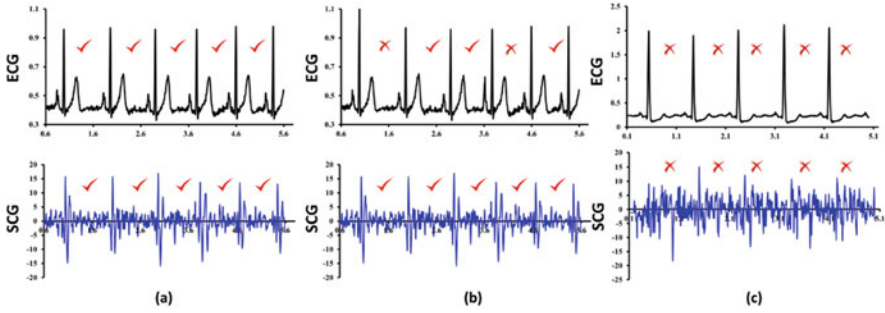


Fig. 25 Combined performance evaluation of ECG and SCG modalities [2]. (a) ECG/SCG normal morphology detection. (b) ECG abnormal, SCG normal morphology detection. (c) ECG/SCG abnormal morphology detection

$$J_{outcome} = CC_{ecg}^k \wedge CC_{scg}^k \tag{9}$$

The outcome from any modality normal (binary 0) means the corresponding cardiac cycles are normal. Here, CC_{ecg}^k and CC_{scg}^k signify individual classification results of k th cardiac cycle of ECG and SCG. The authors in [2] have performed rigorous experiments to determine the cardiac health of a person considering the analysis of both ECG and SCG modalities as shown in Fig. 25. From Fig. 25a, it is observed that the morphology of the cardiac cycle in both ECG and SCG is normal, resulting in the patient is normal, i.e., Cardiac Abnormality Index (CAI) = 0. Although, the ECG shows cardiac abnormalities, based on the SCG, the cardiac rhythm is normal with CAI = 0 in Fig. 25b. It could be concluded that the patient is normal and the ECG might be affected by external noise. In the third case as the morphology of both ECG and SCG show abnormality (CAI = 1) as shown in Fig. 25c, it indicates that the person suffers from severe cardiac abnormalities.

$$\Gamma_{ecg} = \underset{l \in \{1,2\}}{\operatorname{argmax}} p(\varphi_l) \times \prod_{i=1}^{i=6} p(x_i | \varphi_l), \text{ where } x_i \in FV_{ecg}^k \tag{10}$$

4 Comparison of Related Literature

Currently, several literature have proposed the cardiac abnormality diagnosis using ECG or SCG modalities. A detailed comparative study of different related works along with their limitations is presented in Tables 4 and 5. The analysis of feature point-based prediction without using AI is presented in Table 4, whereas Table 5

Table 4 Summary of related works on feature point based cardiological data analysis

Feature point based cardiological ECG signal data analysis					
Purpose	Data source	Feature points	Features for abnormality detection	Results	Limitations
Detection of real-time R and QRS points ECG [23]	MIT-BIH	Q, R, S	Non-contaminated R peak, QRS and RR interval	Sensitivity (%): 99.8, True Positive (TP) (%): 99.8, Detection Error Rate (DER): 0.39	The contamination of Q and S points due to noise are not considered
Classification of QRS morphology using ECG [31]	QT, MITDB	Q, R, S	qR, qRs, Rs, R, RS, rSR', rR', RSr', rS, rSr', Qr, QS, QR, qrS, qS, QRs, Qrs	Accuracy: 0.99, Winning (%): 91.62	Abnormalities related to P and T-wave are not considered
Early Detection of Atrial Fibrillation Based on ECG Signals [32]	MIT-BIH	P, Q, R, S	PQ, QS, RR interval	A-Fib detection rate: ~ 0.85	Only a single cardiac abnormality scenario is considered
Detection of QRS complex in real-time using ECG [33]	BITalino ECG sensor	Q, R, S	R-peak, QRS complex and Q-peak	-	Only the abnormality related to QRS complex is considered ignoring the P and T feature points
Detection of R peaks using ECG for Arrhythmia [34]	MITDB, FDB	Q, R, S	R-peak, Non-QRS, QRS and RR interval	Sensitivity (%): 99.82, TP (%): 99.80, DER (%): 0.38	Single feature point for only a particular abnormality is considered

Feature point based cardiological SCG signal data analysis

	IM, AO, AC	Slopes of IM_AO and AC_ACM interval	Detection rate per individual: 0.32	Requires complex procedures of external envelope formation to delineate fiducial points
Delineation of the fiducial points of the SCG Signal [11]	High-sensitivity accelerometer	IM, AO, AC		
Detection of aortic valve opening phase using SCG signal [35]	Self built SCG acquisition circuit	AO, AC	Sensitivity (%): 97.37, Accuracy (%): 95.1, DER (%): 5.18	Only three fiducial points are delineated that limits the applicability
Detection of heartbeat using SCG [36]	CEBS, Custom dataset	IM, AO	Sensitivity (%): 98.5 (CEBS) and 99.1 (Custom), Precision (%): 98.6 (CEBS) and 97.9 (Custom)	Cannot differentiate the normal and abnormal heartbeat
Identification of systolic time Intervals in SCG [37]	Accelerometer	AO, AC	-	Cannot distinguish desired and undesired SCG peaks
Automatic annotation of high frequency periodic SCG signal [38]	SFU_GYM	IM, IC	Overall IM detection accuracy (%): 97.87, AC detection accuracy (%): 98.42	Only two specific SCG peaks are considered instead of all nine

Table 5 Summary of related works on AI-based cardiac data analysis

AI-based cardiological ECG signal data analysis					
Purpose	Data source	No. of considered classes (or) features	Model(s) used	Results	Limitations
DL-based classification of heartbeat type using ECG [39]	MITBIH	Class oriented scheme: 15, subject-oriented-scheme: 5	CNN +LSTM	Accuracy (%): 98.27, Sensitivity (%): 99.95, Precision (%): 60.93	Higher computation complexity, only arrhythmias cardiac abnormality is considered
DL-enabled detection of cardiovascular disease using ECG [40]	Nanjing Medical University	18 categories	CNN	Sensitivity (%): 99.8, TP (%): 99.8, DER: 0.39	Low precision due to higher false negatives
DL-powered classification of rhythm types and quality using ECG [41]	IEEE Signal Processing Cup 2015	Rhythm types: 2 (sinus, atrial fibrillation), quality types: 3 (poor, acceptance, excellent)	DeepBeat	Sensitivity: 0.98, Specificity: 0.99, F1 score: 0.93	Only one type of cardiac rhythm (albeit) type is studied
DL-based delineation of ECG signal [42]	QTDB (training and validation), MITDB (testing)	4 (No wave, P wave, T wave, QRS)	DENS-ECG (CNN + Bidirectional LSTM)	Sensitivity (%): 97.95, Precision (%): 95.68	Higher computational complexity and mostly the ECG signal from arterial fibrillation is investigated

ML-assisted automatic diagnosis of cardiovascular disorder using ECG [43]	Computing in Cardiology Challenge 2020	Extracted features: Laws texture energy measure, Histogram of oriented gradients, Gabor wavelet transform, Gray level Co-occurrence matrix	Randomized neural network	Accuracy (%): 97.0	Time-consuming and requires skillful feature engineer
AI-based cardiological SCG signal data analysis					
DL-enabled detection and analysis of heartbeats in SCG [18]	CEBS	1 (Delay between each R-peaks and IM/AO complexes)	Variational Auto Encoder (VAE)	Sensitivity (%): 98.5, Precision (%): 98.6	Only IM/AO features points are considered instead of nine that limits the applicability
DL-based cardiac activity monitoring using SCG [30]	CEBS	2 (R-peak, heart rate variability (HRV))	Deep Fully CNN	Sensitivity: 0.98, Positive Predicted Value: 0.98	Only the R-peak location is considered to detect the HRV
ML-assisted automatic annotation of SCG feature points [44]	CEBS	IM, AO and extracted features: Amplitude, Time of appearance	LR, SVM, RF, DT, Gaussian NB (GNB)	Overall accuracy (%): 88.5 (LR), 87.25 (SVM), 85.75 (DT), 89.25 (RF), 88.75 (GNB)	Does not conclude the best among all the methods

is summarizes the work of various ML and DL prognosis models related to the coronary diseases.

5 Conclusions

The application of ECG and SCG in coronary disease prediction is vast but much needed for the early abnormal sign determination. Efficient mathematical models and AI algorithms are designed for relevant cardiac feature extraction automatically, where the manual method of delineation is tedious. Moreover, the combined analysis of both ECG and SCG modality for each cardiac cycle reduces the individual error chances and increases the prediction confidence. The design of early warning modules can be useful for sending emergency alerts to the concerned person so that life can be saved. However, there is also a need for extensive research especially to analyze the combination of cardiac signals and images considering a person's physiological factors.

References

1. Alwan A. Global status report on noncommunicable diseases 2010. World Health Organization; 2011.
2. Sahoo PK, Thakkar HK, Lin WY, Chang PC, Lee MY. On the design of an efficient cardiac health monitoring system through combined analysis of ECG and SCG signals. *Sensors*. 2018 Feb;18(2):379.
3. Sahoo PK, Thakkar HK, Lee MY. A cardiac early warning system with multi channel SCG and ECG monitoring for mobile health. *Sensors*. 2017 Apr;17(4):711.
4. Chen J, Valehi A, Razi A. Smart heart monitoring: Early prediction of heart problems through predictive analysis of ecg signals. *IEEE Access*. 2019 Aug 27;7:120831-9.
5. Lyon, Aurore, et al. "Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances." *Journal of The Royal Society Interface* 15.138 (2018): 20170821.
6. Inan OT, Migeotte PF, Park KS, Etemadi M, Tavakolian K, Casanella R, Zanetti J, Tank J, Funtova I, Prisk GK, Di Rienzo M. Ballistocardiography and seismocardiography: A review of recent advances. *IEEE journal of biomedical and health informatics*. 2014 Oct 7;19(4):1414-27.
7. Mora N, Cocconcelli F, Matrella G, Ciampolini P. A Unified Methodology for Heartbeats Detection in Seismocardiogram and Ballistocardiogram Signals. *Computers*. 2020 Jun;9(2):41.
8. Zanetti J. Seismocardiography: A new technique for recording cardiac vibrations. concept, method, and initial observations. *Journal of cardiovascular technology (New York, NY)*. 1990;9(2):111-8.
9. Wahlstrom J, Skog I, Handel P, Khosrow-Khavar F, Tavakolian K, Stein PK, Nehorai A. A hidden Markov model for seismocardiography. *IEEE Transactions on Biomedical Engineering*. 2017 Jan 9;64(10):2361-72.
10. Zakeri V, Akhbardeh A, Alamdari N, Fazel-Rezai R, Paukkunen M, Tavakolian K. Analyzing seismocardiogram cycles to identify the respiratory phases. *IEEE Transactions on Biomedical Engineering*. 2016 Oct 26;64(8):1786-92.

11. Khosrow-Khavar F, Tavakolian K, Blaber A, Menon C. Automatic and robust delineation of the fiducial points of the seismocardiogram signal for noninvasive estimation of cardiac time intervals. *IEEE Transactions on Biomedical Engineering*. 2016 Oct 12;64(8):1701-10.
12. Al-Turjman F, Nawaz MH, Ulusar UD. Intelligence in the Internet of Medical Things era: A systematic review of current and future trends. *Computer Communications*. 2020 Jan 15;150:644-60.
13. ECG wearables: How they work and the best on the market. (2020) [Online], Available <https://www.wearable.com/health-and-wellbeing/ecg-heart-rate-monitor-watch-guide-6508> (accessed 13 September 2020).
14. Balestrieri E, Boldi F, Colavita AR, De Vito L, Laudato G, Oliveto R, Picariello F, Rivaldi S, Scalabrino S, Torchitti P, Tudosa I. The architecture of an innovative smart T-shirt based on the Internet of Medical Things paradigm. In 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA) 2019 Jun 26 (pp. 1-6). IEEE.
15. Mesko B. The role of artificial intelligence in precision medicine, 2017.
16. Intel AI Academy, Artificial Intelligence and Healthcare Data”, [Online], Available: <https://software.intel.com/en-us/articles/artificial-intelligence-and-healthcare-data>, Accessed February 2020.
17. Litjens, Geert, et al. “A survey on deep learning in medical image analysis.” *Medical image analysis* 42 (2017): 60-88.
18. Mora N, Cocconcelli F, Matrella G, Ciampolini P. Detection and analysis of heartbeats in seismocardiogram signals. *Sensors*. 2020 Jan;20(6):1670.
19. Lin WY, Chou WC, Chang PC, Chou CC, Wen MS, Ho MY, Lee WC, Hsieh MJ, Lin CC, Tsai TH, Lee MY. Identification of location specific feature points in a cardiac cycle using a novel seismocardiogram spectrum system. *IEEE journal of biomedical and health informatics*. 2016 Oct 25;22(2):442-9.
20. Hadjem M, Salem O, Nait-Abdesselam F. An ECG monitoring system for prediction of cardiac anomalies using WBAN. In 2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom) 2014 Oct 15 (pp. 441-446). IEEE.
21. Lyon A, Minchol A, Martez JP, Laguna P, Rodriguez B. Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. *Journal of The Royal Society Interface*. 2018 Jan 31;15(138):20170821.
22. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nature Reviews Cardiology*. 2021 Feb 1:1-4.
23. Bae TW, Kwon KK. Efficient real-time R and QRS detection method using a pair of derivative filters and max filter for portable ECG device. *Applied Sciences*. 2019 Jan;9(19):4128.
24. Garate-Escamila AK, El Hassani AH, Andres E. Classification models for heart disease prediction using feature selection and PCA. *Informatics in Medicine Unlocked*. 2020 Jan 1;19:100330.
25. Avanzato R, Beritelli F. Automatic ECG diagnosis using convolutional neural network. *Electronics*. 2020 Jun;9(6):951.
26. Thakkar HK, Sahoo PK. Towards automatic and fast annotation of seismocardiogram signals using machine learning. *IEEE Sensors Journal*. 2019 Nov 1;20(5):2578-89.
27. Poliac MO, Zanetty JM, Salerno D, Wilcox GL. Seismocardiogram (SCG) interpretation using neural networks. In *Computer-Based Medical Systems-Proceedings of the Fourth Annual IEEE Symposium* 1991 Jan 1 (pp. 288-289). IEEE Computer Society.
28. Yao J, Tridandapani S, Auffermann WF, Wick CA, Bhatti PT. An adaptive seismocardiography (SCG)-ECG multimodal framework for cardiac gating using artificial neural networks. *IEEE journal of translational engineering in health and medicine*. 2018 Oct 8;6:1-1.
29. Murat F, Yildirim O, Talo M, Baloglu UB, Demir Y, Acharya UR. Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review. *Computers in biology and medicine*. 2020 Apr 8:103726.

30. Suresh P, Narayanan N, Pranav CV, Vijayaraghavan V. End-to-End Deep Learning for Reliable Cardiac Activity Monitoring using Seismocardiograms. arXiv preprint arXiv:2010.05662. 2020 Oct 12.
31. do Vale Madeiro JP, Marques JA, Han T, Pedrosa RC. Evaluation of mathematical models for QRS feature extraction and QRS morphology classification in ECG signals. *Measurement*. 2020 May 1;156:107580.
32. Ahmed N, Zhu Y. Early detection of atrial fibrillation based on ECG signals. *Bioengineering*. 2020 Mar;7(1):16.
33. Rodriguez-Jorge R, De Leon-Damas I, Bila J, Skvor J. Internet of things-assisted architecture for QRS complex detection in real-time. *Internet of Things*. 2021 Jun 1;14:100395.
34. Bae TW, Lee SH, Kwon KK. An Adaptive Median Filter Based on Sampling Rate for R-Peak Detection and Major-Arrhythmia Analysis. *Sensors*. 2020 Jan;20(21):6144.
35. T. Choudhary, M. K. Bhuyan, and L. N. Sharma, A novel method for aortic valve opening phase detection using SCG signal, *IEEE Sensors Journal*, vol. 20, no. 2, pp. 899-08, 2020.
36. Cocconcelli F, Mora N, Matrella G, Ciampolini P. High-Accuracy, Unsupervised Annotation of Seismocardiogram Traces for Heart Rate Monitoring. *IEEE Transactions on Instrumentation and Measurement*. 2020 Jan 17;69(9):6372-80.
37. Shafiq G, Tatinati S, Ang WT, Veluvolu KC. Automatic Identification of Systolic Time Intervals in Seismocardiogram. *Sci Rep*. 2016 Nov 22;6:37524.
38. F. Khavar, K. Tavakolian, A. P. Blaber, J. M. Zanetti, R. Rezaei, and C. Menon, Automatic annotation of seismocardiogram with high-frequency precordial accelerations, *IEEE Journal of Biomedical Health Informatics*, vol. 19, no. 4, pp. 1428-1434, 2015.
39. Shi H, Qin C, Xiao D, Zhao L, Liu C. Automated heartbeat classification based on deep neural network with multiple input layers. *Knowledge-Based Systems*. 2020 Jan 5;188:105036.
40. Zhang X, Gu K, Miao S, Zhang X, Yin Y, Wan C, Yu Y, Hu J, Wang Z, Shan T, Jing S. Automated detection of cardiovascular disease by electrocardiogram signal analysis: a deep learning system. *Cardiovascular Diagnosis and Therapy*. 2020 Apr;10(2):227.
41. Torres-Soto J, Ashley EA. Multi-task deep learning for cardiac rhythm detection in wearable devices. *NPJ digital medicine*. 2020 Sep 9;3(1):1-8.
42. Peimankar A, Puthusserypady S. DENS-ECG: A deep learning approach for ECG signal delineation. *Expert Systems with Applications*. 2021 Mar 1;165:113911.
43. Ertugrul OF, Acar E, Aldemir E, Oztekin A. Automatic diagnosis of cardiovascular disorders by sub images of the ECG signal using multi-feature extraction methods and randomized neural network. *Biomedical Signal Processing and Control*. 2021 Feb 1;64:102260..
44. Rai D, Thakkar HK, Singh D, Bathala HV. Machine Learning Assisted Automatic Annotation of Isovolumic Movement and Aortic Valve Closure using Seismocardiogram Signals. In 2020 IEEE 17th India Council International Conference (INDICON) 2020 Dec 10 (pp. 1-6). IEEE.

Early Lung Cancer Detection by Using Artificial Intelligence System



Fatma Taher

1 Chapter 1: Introduction

According to the World Health Organization, Lung cancer is the major cancer killer in both men and women [1]. For the early diagnoses of lung cancer, chest x-ray and sputum cytology techniques are used [2]. Therefore, by employing the sputum color images, a new CAD system is developed for the prediction of lung cancer. A computer-aided diagnosis system is a computing system meant for automated detection and diagnosis of abnormalities in medical images [3]. A successful system will have two main advantages:

1. Reduce time and resources dedicated to manual examination.
2. The amount of data that could be handled by CAD-systems would be much higher compared to the cases that can be investigated manually.

CAD detection is different from CAD diagnosis where the detection is part of the diagnosis. Thus, CAD detection usually refers to the process of automatically detecting the diseased areas. CAD diagnosis is contained in the classification part where the system can classify the subjects to normal and abnormal classes after applying a well-known classification algorithm such as rule-based, artificial neural network or support vector machine [4]. The procedures of the CAD system are represented by the analysis and the diagnostic parts. In the analysis part, the region of interest (ROI) is extracted that includes the nuclei and cytoplasm regions, and image processing techniques are also used. In the diagnosis part, diagnostic rules are applied to detect the abnormal cases based on these rules. Due to the noisy and cluttered background patterns of the sputum images, automatic detection of

F. Taher (✉)

College of Technological Innovation, Zayed University, Dubai, United Arab Emirates
e-mail: Fatma.Taher@zu.ac.ae

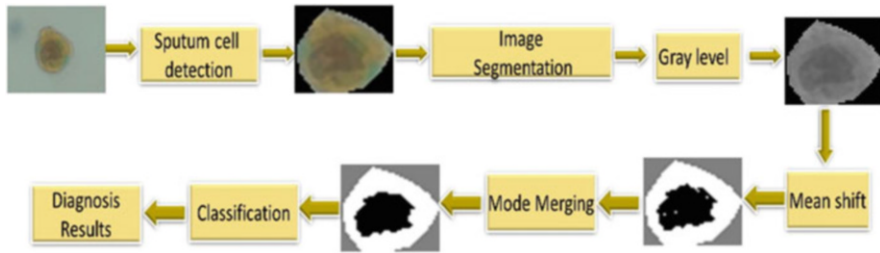


Fig. 1 Diagnosis process of the CAD system

the cancerous cells becomes highly problematic. The proposed methods introduce a new system for early detection, segmentation, and extraction of normal and cancer cells. 100 sputum color images were used for the testing of new system, and its results were evaluated based on the sensitivity, precision, specificity, and accuracy. The entire diagnosis process of the proposed CAD system is depicted in Fig. 1.

2 Chapter 2: Cell Detection and Extraction

Color information is used for the detection of cancer cell. In this work, we used rule-based algorithm and the Bayesian classification.

2.1 Rule-Based Algorithm

This algorithm is based on a heuristic rule-based on the chromatic disparity between the sputum cell and the background in the stained images.

The rule used to extract sputum cell pixels for the blue dye-stained image is explained as follows. Let $I(x, y)$ be an image pixel:

$$\text{If } (B(x, y) < G(x, y) + \theta) \text{ then } I(x, y) \text{ is sputum else } I(x, y) \text{ is non sputum} \quad (1)$$

where $B(x, y)$ and $G(x, y)$ represent the pixel blue and green values in the RGB color space and θ is a threshold parameter set empirically. Figure 2 depicts the result of applying Eq. (1) to the raw image stained with blue dye (Fig. 2a). The nuclei and cytoplasm are not correctly detected as depicted in Fig. 2b. However, a bunch of debris cells are identified. This happens because of incorrect value of the threshold parameter. On the other hand, the image in Fig. 2c depicts the correct result, after determining the appropriate threshold value.

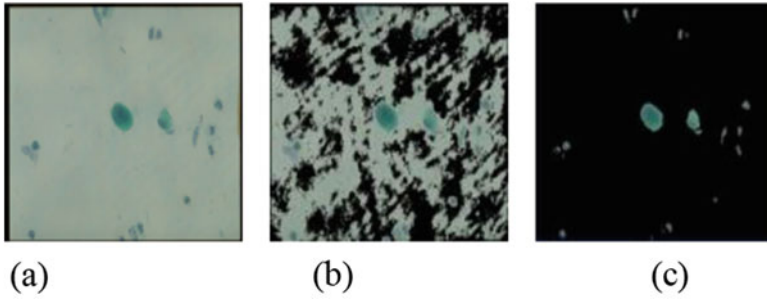


Fig. 2 (a) Blue dye-stained image. (b) The nuclei and cytoplasm are not correctly detected. (c) The correct result

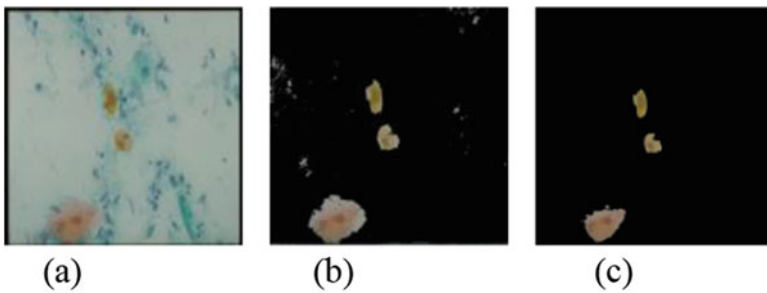


Fig. 3 (a) Image stained with red dyes, (b) the result of applying Eq. (2) with erroneous threshold value, (c) the result of applying same equations with appropriate threshold value

Between the sputum cells and the background, red color dye is used; the rule is as follows:

Let $I(x, y)$ be an image pixel

$$\text{If } ((2 * G(x, y) + \theta) < (R(x, y) + B(x, y))) \text{ then } I(x, y) \text{ is sputum else } I(x, y) \text{ is non - sputum} \tag{2}$$

where $G(x, y)$, $R(x, y)$, and $B(x, y)$ represent the pixel green, red, and blue values in the RGB color space and θ is a threshold parameter set empirically. Figure 3 depicts the result of applying Eq. (2). Figure 3b shows that the nuclei and cytoplasm are not correctly detected. Figure 3c depicts the correct result.

2.1.1 Experiments

The performance of the proposed rule-based algorithm can be analyzed by conducting a series of experiments. In this study, 100 images were used. Ground-truth data were obtained manually. For testing, the rule-based algorithm was applied to the

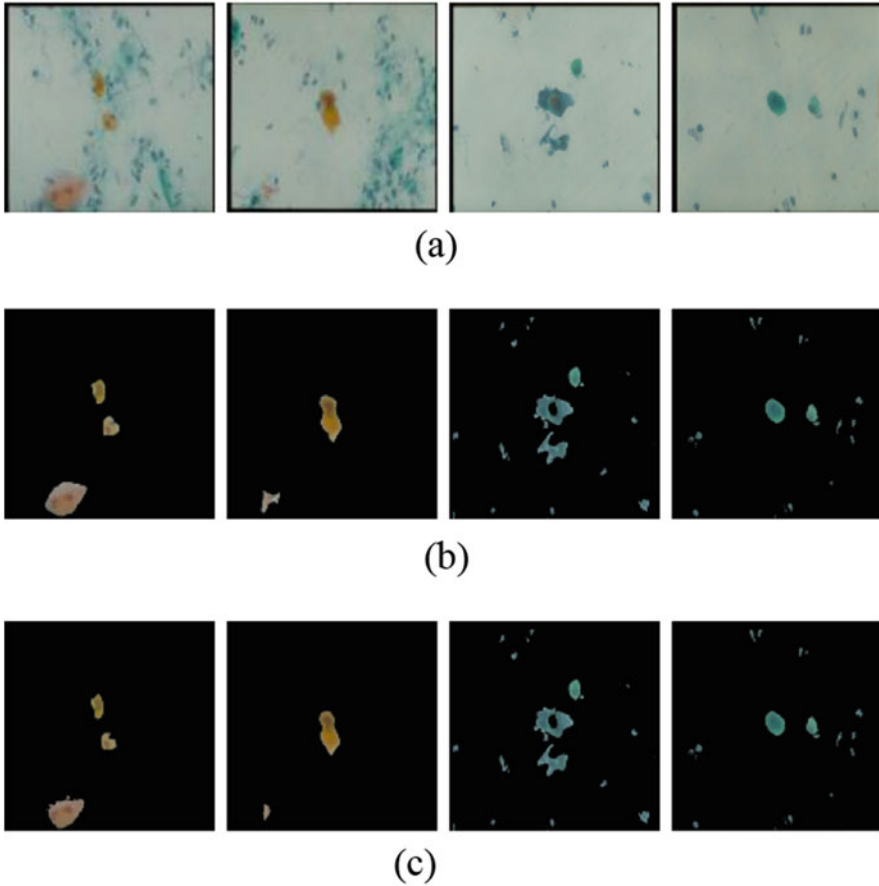


Fig. 4 (a) Raw images, (b) ground-truth images, (c) proposed rule-based results

test images, and then the resulted image was compared to the ground-truth data. Figure 4 depicts the result of applying the proposed rule-based algorithm, where it can be seen that the sputum cells are detected with reasonable accuracy. Figure 4a depicts the raw images, Fig. 4b depicts the ground-truth images which contain the sputum cells, and Fig. 4c depicts the result of applying the proposed rule-based algorithm. The ground-truth cells are employed for evaluating the rule-based resulted images for correctly detecting the ROI. After that, the correctly detected pixels will be obtained. Therefore, after performing a series of trial-and-error tests using an optimization program, the threshold θ parameter was determined which is explained in [5]. The values were found to be in the range from -35 to -15 . The segmentation becomes more selective as the threshold θ value increases.

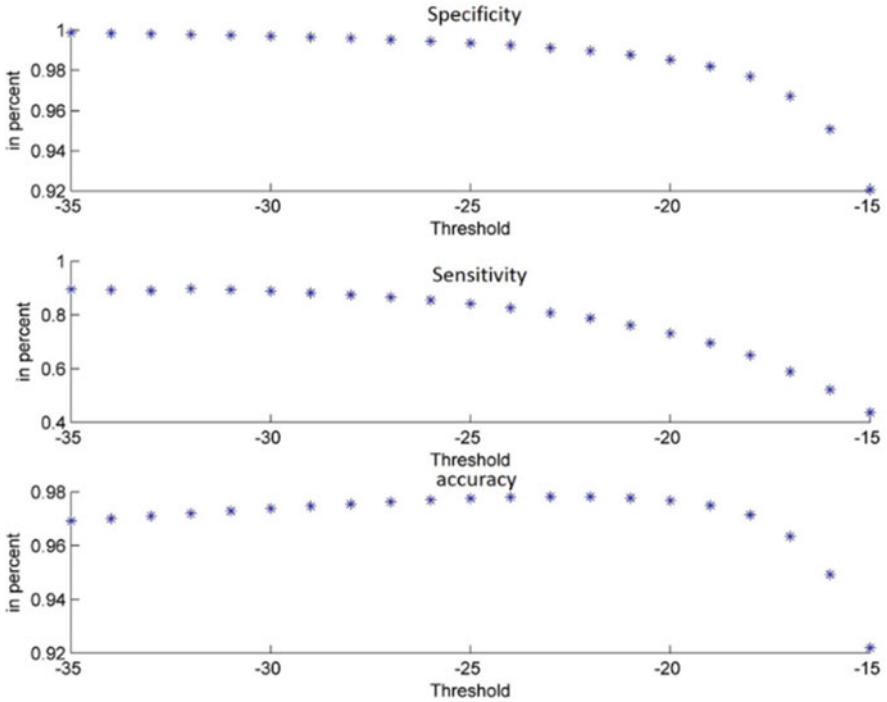


Fig. 5 The percentage of the sensitivity, specificity, and accuracy of the rule-based algorithm

Figure 5 depicts the percentage of the specificity, sensitivity, and accuracy mapped in the function of the threshold θ during the detection and extraction processes. This algorithm achieves best sensitivity of 82%, specificity of 99%, and accuracy of 98%.

2.2 Bayesian Classification

A probabilistic method is used to address the cell detection problem based on the Bayesian classification [6]. Instead of using trial-and-error testing, the threshold parameters are estimated using a systematic method.

While considering this method, a pixel x is used as a part of the sputum region if:

$$p(bg|x) < p(sp|x) \tag{3}$$

where sp denotes the sputum and bg the background, respectively.

Using the classification approach, Eq. (3) can be written as:

$$\frac{\mu_{sp} p(bg)}{\mu_{bg} p(sp)} < \frac{p(x|sp)}{p(x|bg)} \quad (4)$$

where μ_{sp} denotes the loss weight attained when the sputum class is selected and μ_{bg} , when background is selected. $p(bg)$ is the prior probabilities of the background, and $p(sp)$ is the prior probabilities of sputum classes. The following equations are used for estimation of these parameters:

$$p(sp) = \frac{T_{sp}}{T_{sp} + T_{bg}} \quad (5)$$

$$p(bg) = \frac{T_{bg}}{T_{sp} + T_{bg}} \quad (6)$$

where T_{sp} denotes the number of sputum cell pixels and T_{bg} denotes the background pixels. The setting of the ratio $\lambda = \frac{\mu_{sp}}{\mu_{bg}}$ is explained in [7]. Figure 6 shows samples of sputum cell extraction results. Samples of sputum cell detection are depicted in Fig. 6a–d.

2.2.1 Experiments

The performance of Bayesian classifier can be evaluated by conducting a series of experiments with the histogram analysis in terms of color representation and color quantization on the detection of sputum cell. The performance of the system was compared in terms of sensitivity, specificity, and accuracy as were explained earlier. We also used receiver operating characteristics (ROC) curves for the performance assessment. For the training, manually obtained data was used. Then the Bayesian classifier was used to test images. Sputum image data partition was used for training and testing as explained in [7]. In the detection process, the ROC curves are computed for the four-color representations for five histogram resolutions, and then to test images, Bayesian classifier was applied. Across all the resolutions, we found that the HSV and the RGB maintains a strong performance.

2.3 Bayesian Classifier vs. Rule-Based Algorithm

After trying the rule-based algorithm and a Bayesian classification technique, we found that the Bayesian classifier with the analysis of the histogram outperforms the

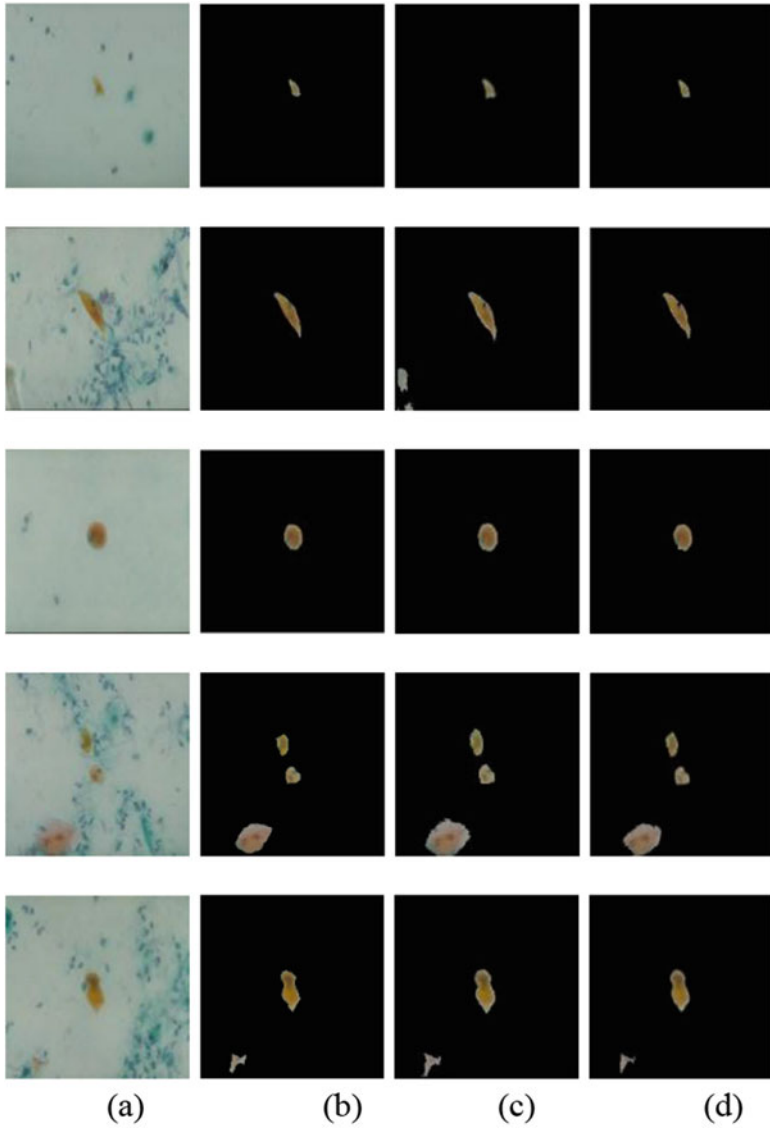


Fig. 6 Examples of detection of sputum cell. (a) Input images. (b) Manually obtained data. (c) Using the Bayesian classifier with $\lambda = 2$, and (d) with $\lambda = 7$

rule-based technique. While considering the color quantization, as the color space resolution increases, accuracy of the classification will also increase. Finally, the Bayesian technique showed better results which is highly satisfying.

Fig. 7 Performance accuracy of the color spaces

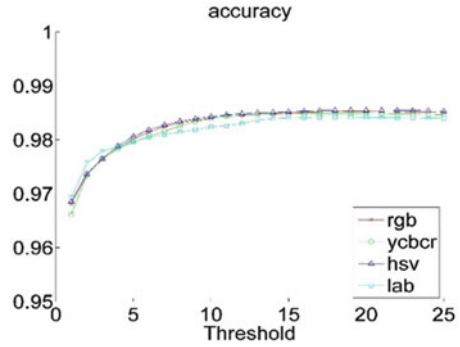
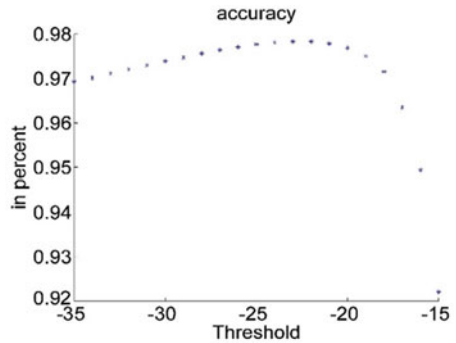


Fig. 8 Accuracy measurement of the rule-based algorithm



Quantitatively, Fig. 7 shows the accuracy criterion variation for the Bayesian classifier for different color spaces for 64-histogram resolutions as a function of the ratio λ .

Figure 8 depicts the accuracy measurement of the rule-based method. The rule-based algorithm and the Bayesian classification exhibits 98% accuracy and 99% specificity. Its sensitivity is 82% and 89%, respectively. This shows that the Bayesian classification has showcased better results.

By comparing the ROC-curves obtained in both methods, it is clear that the Bayesian technique outperforms the other.

3 Chapter 3: Image Segmentation

The mean shift algorithm is the segmentation method used in our proposed system. In this method, the candidate solutions in the feature space are shifted towards the maximum density points by using this algorithm. In our experiment, the pixel's gray level and the pixel spatial coordinates define the feature space. An appropriate kernel is required to find the desired modes for obtaining good density estimation. The kernel (known as the Parzen window technique in the pattern recognition

literature [7]) was used to find the modes in our distribution. For a distribution with d dimensions and a set of data points x_i with $i = 1 \dots n$ is distributed over that space.

The density function is given as follows:

$$f(x) = \frac{1}{n} \sum_{i=1}^n |H|^{-\frac{1}{2}} K_H(x - x_i) \tag{7}$$

where the number of cell pixels is denoted as n and x_i defines the feature vector. The profile of the kernel function is denoted as K_H . The Epanechnikov kernel is one example of such kernel [8]. In this work, we used the normal kernel. The bandwidth matrix is represented as H . The normal function in Eq. (7) can be written as:

$$f(x) = \frac{1}{n} \sum_{i=1}^n (2\pi)^{-3/2} |H|^{-1/2} e^{-(x-x_i)^T H^{-1}(x-x_i)} \tag{8}$$

3.1 Mean Shift Procedure

By using the mean shift algorithm, the local maxima in the feature space are obtained. We need to find the modes and their distribution first. Across the convergence path, the different mode locations are given as follows:

$$y_{j+1} = \frac{\sum_{i=1}^n x_i g\left((y_j - x_i)^T H^{-1}(y_j - x_i)\right)}{\sum_{i=1}^n g\left((y_j - x_i)^T H^{-1}(y_j - x_i)\right)} \quad j = 1, 2, \dots \tag{9}$$

where the kernel function derivative is denoted as g .

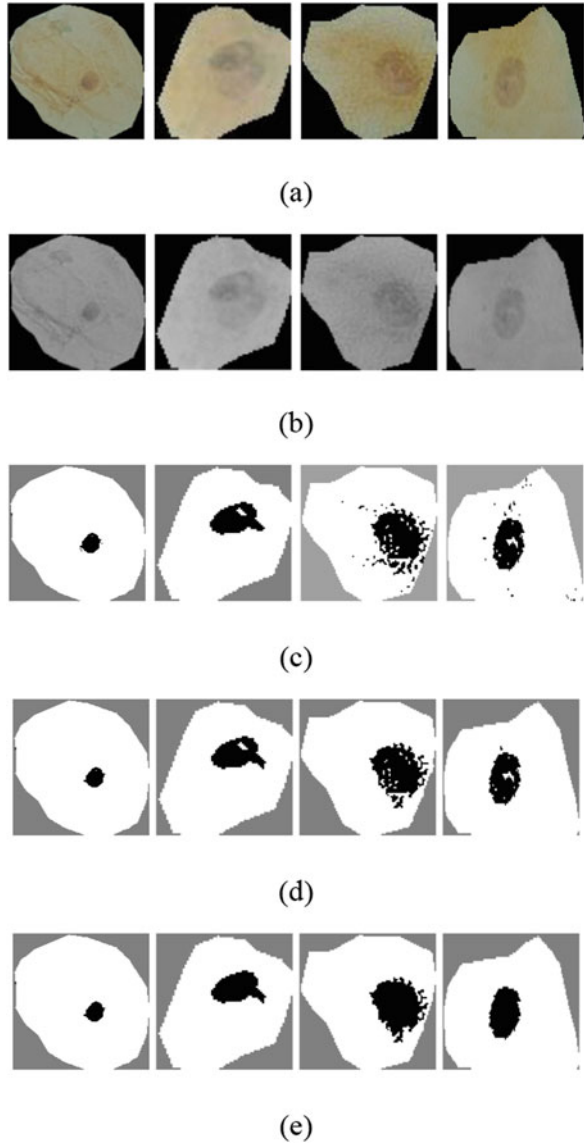
The procedure is composed of the following steps:

1. Select a starting point.
2. Set starting point. Apply kernel Eq. (8) with starting center.
3. Update the candidate center, according to the mean shift Eq. (9).
4. Repeat steps 3–5 until convergence.

The different stages of mean shift segmentation are depicted in Fig. 9. The sputum cells are shown in Fig. 9a. The contrast of the images is enhanced in Fig. 9b. Figure 9c shows the final results. We found that non-compact regions are produced as a result of this. Therefore, region merging is performed. Firstly, the largest connected patches are extracted from each region. Then, region merging is performed and it is calculated as follows:

$$\text{Dist (mode)} = \frac{1}{n} \sum_{\rho_{\text{mode}}} \left| \begin{pmatrix} x_{\text{centre}} \\ y_{\text{centre}} \end{pmatrix} - \begin{pmatrix} x_p \\ y_p \end{pmatrix} \right| \tag{10}$$

Fig. 9 Different stages of mean shift segmentation. (a) Sputum cells. (b) Contrast of the image is enhanced. (c) Mean shift segmentation. (d) Mode merging. (e) Region refinement



Then the nucleus is considered as the mode with the minimal distance. A rule-based region merging (Fig. 9d) is performed to get a connected nucleus. Then, region refinement is done [9] as depicted in Fig. 9e.

Table 1 Performance matrix

Performance measurements	HNN (%)	Gray mean shift (%)	Gray-space mean shift (%)
Sensitivity	73.77	92.7	93.40
Precision	70.31	82.50	88.21
Accuracy	65	85	87.11

3.2 Experiments

A series of experiments are conducted to evaluate our cell segmentation method. The performance was evaluated by comparing the mean shift segmentation results with the ground-truth data. We used the following assessment criteria for performance measurement: sensitivity, precision, and accuracy. In our experiments, the Hopfield neural network (HNN) [10] is used for evaluating the performance of the mean shift in gray-level feature space. Table 1 shows the performance matrix.

From the table, it is clear that the best performance is shown by the gray-space mean shift and the performance of HNN is low when compared with the other methods. Therefore, the gray-level density estimation is the perfect method for nucleus segmentation.

4 Chapter 4: Feature Extraction

Different features [11] are extracted after detecting the nucleus and cytoplasm area in the cell which helps in the diagnostic process. The ability of the CAD system to identify the normal and abnormal cells is the major issue faced by any CAD systems. We can solve this problem by using the correct features. These features are explained as follows. The first feature is the NC ratio, which is computed as [11]:

$$\text{NC ratio} = \frac{\text{Area (Nucleus)}}{\text{Area (Cytoplasm)}} * 100 \quad (11)$$

The extracted nuclei and cytoplasm samples are depicted in Fig. 10a. The black and white areas in Fig. 10b represent the nucleus and the cytoplasm respectively. The nucleus area is shown in Fig. 10c. The next feature is the nucleus perimeter defined by:

$$P (\text{Nucleus}) = \int_t \sqrt{x^2(t) + y^2(t)} dt \quad (12)$$

where the parameterized contour point coordinates are represented as $x(t)$ and $y(t)$. Figure 10d shows the nuclei cell perimeter (green color). The mean value [12] of the nucleus is the next feature used which is calculated as follows:

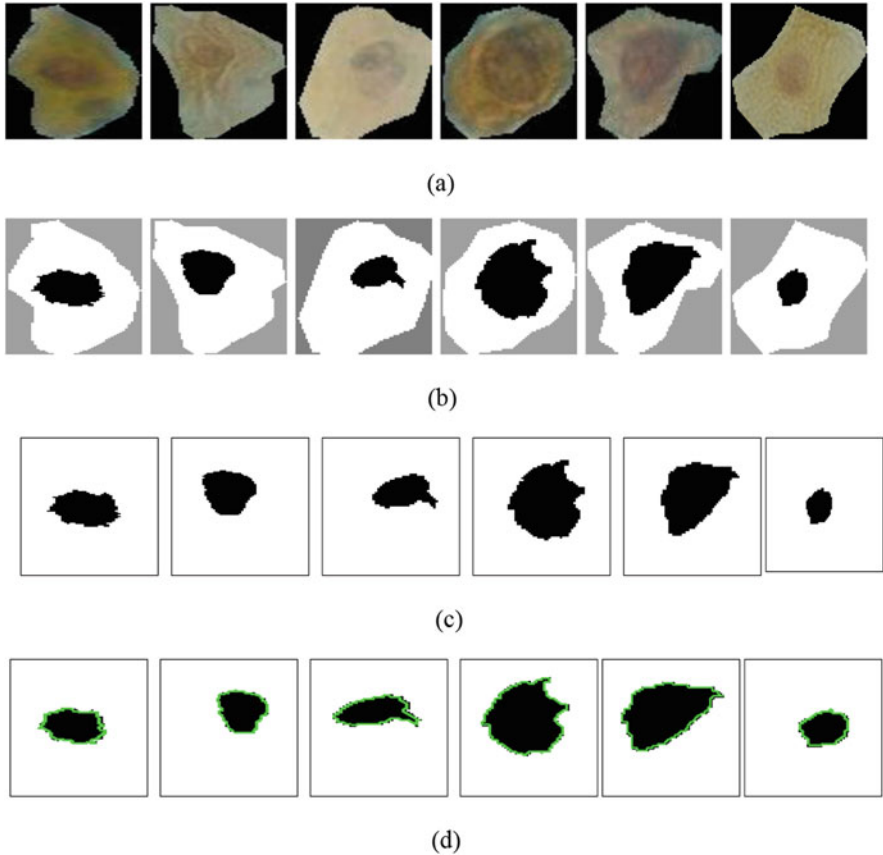


Fig. 10 Samples of the extracted nuclei and cytoplasm. (a) Input image, (b) nucleus and cytoplasm extraction, (c) nucleus area, and (d) the perimeter of the nucleus area (green color)

$$\text{Mean (Nucleus)} = \frac{\sum_{i=1}^N \text{Intensity}(i)}{\text{Area (Nucleus)}} \quad (13)$$

where the intensity color value is given as i and N denotes the number of pixels in the nucleus region. Figure 11 shows the value of mean intensity for both cancerous and non-cancerous cells. In the figure, benign cells are represented in red color and malignant cells as blue. BD refer to the benign density and MD, malignant density. In our system $\theta = 128$. Sometimes, misclassification [13] occurs. In this case, the intensity feature cannot be considered alone. The curvature is the next feature defined as the rate of change in the edge direction. Adjacent tangent line segments will define the curvature at a single point in the boundary. In order to find the curvature at that point of intersection, and difference between slopes of two adjacent straight-line segments is measured [14]. The slope is obtained as:

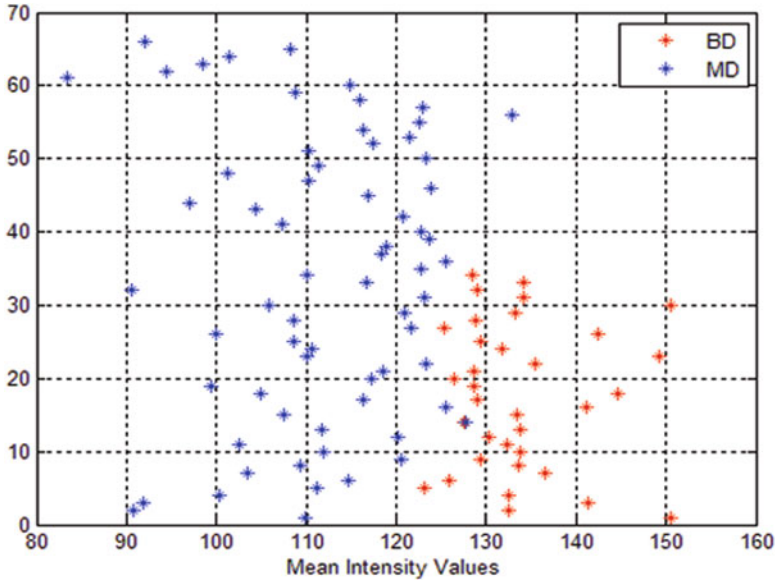


Fig. 11 The nuclei cells intensity variances

$$\varphi(t) = \tan^{-1} \left(\frac{\dot{y}(t)}{\dot{x}(t)} \right) \tag{14}$$

where the derivatives of $x(t)$ and $y(t)$ are denoted as $\dot{x}(t)$ and $\dot{y}(t)$. For each point in the nucleus contour, difference between adjacent slopes ($\delta\theta$) is computed.

If malignant cells are considered, $\delta\theta$ will be higher than the estimated threshold i.e. 50. Figure 12 shows the benign cell curvature extraction ($\delta\theta$). The benign sputum cell is shown in Fig. 12a, d depicts the boundary. Figure 12e depicts the curvature. Figure 13 shows the malignant cell curvature extraction. The fifth feature is called the circularity:

$$\text{Circularity} = \frac{4\pi \text{Area}(\text{Nucleus})}{\text{Perimeter}(\text{Nucleus})^2} \tag{15}$$

When the circularity value is higher, the cells in cleavage are normally round. On the other hand, it will be lower as the normal-growing cells are irregular. The Eigen ratio is the last feature [15]. In our system, irregular cells are long, which is having a high Eigen ratio. Therefore, identifying cancerous and non-cancerous cells using this feature is easy by an appropriate threshold value. The Eigen ratio is calculated as follows [16]:

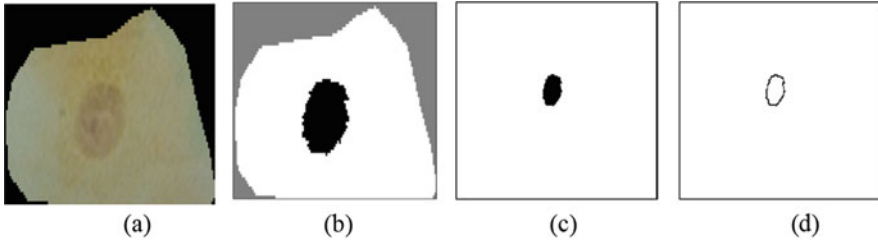


Fig. 12 A benign cell curvature extraction. (a) Sputum cell. (b) Nucleus and cytoplasm segmentation. (c) Extraction of nucleus. (d) Nucleus boundary

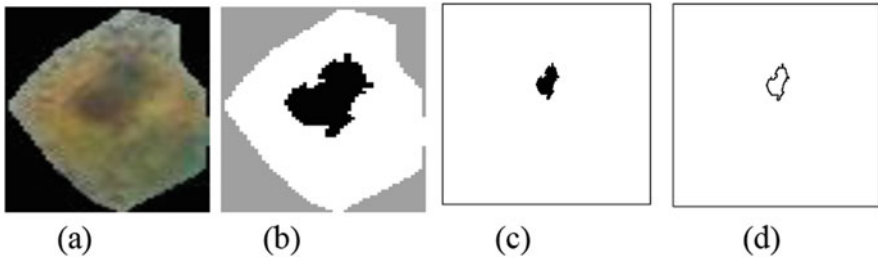


Fig. 13 A malignant cell curvature extraction. (a) Sputum cell. (b) Nucleus and cytoplasm segmentation. (c) Extraction of nucleus. (d) Nucleus boundary

$$\text{Eigen_ratio} = \frac{\frac{a}{b} + \frac{b}{a}}{2} \quad (16)$$

where the eigenvalues of the covariance matrix C are defined as (a, b) [17]:

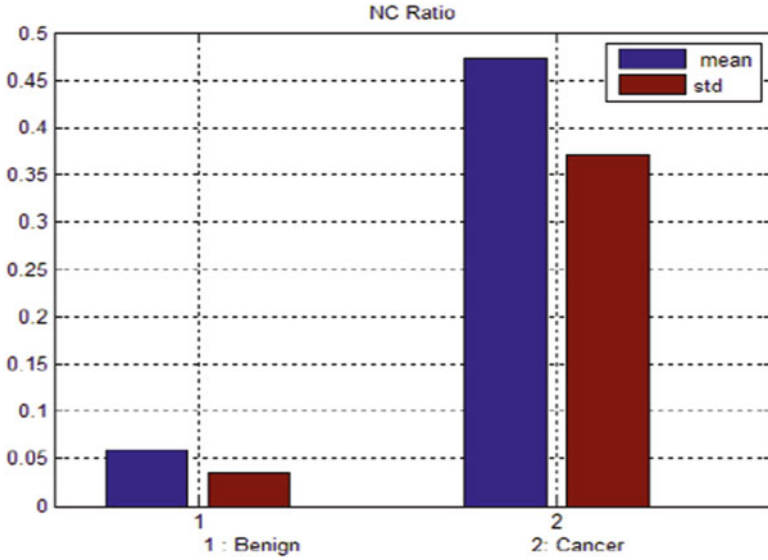
$$C = \frac{1}{N} \sum_{i=1}^N p_i p_i^T, \quad (17)$$

where p_i denotes the point in the nucleus area. The distribution of cell in the nucleus region in both directions (horizontal and vertical) is denoted by the eigenvalues (a, b) . The mean and standard deviation (std) for the benign and malignant cells are explained in [12].

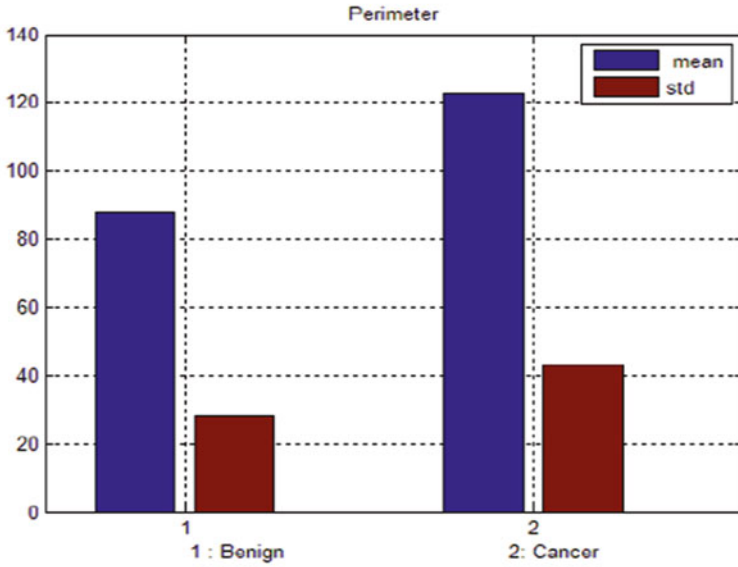
The mean and standard deviation bar charts for cancerous and non-cancerous cells for all the features explained are shown in Figs. 14, 15, and 16.

5 Chapter 5: Classification

Classification is a critical task for computer-aided diagnosis system (CAD), because it is the last step in the CAD system. Therefore, the best outcomes are obtained

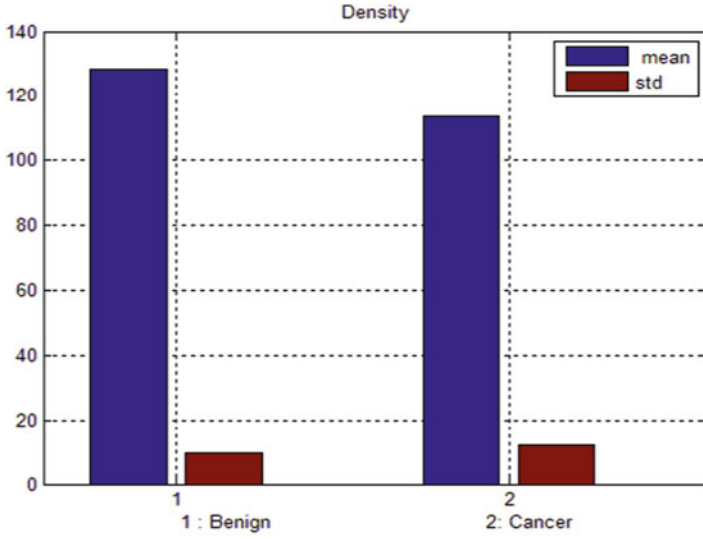


(a)

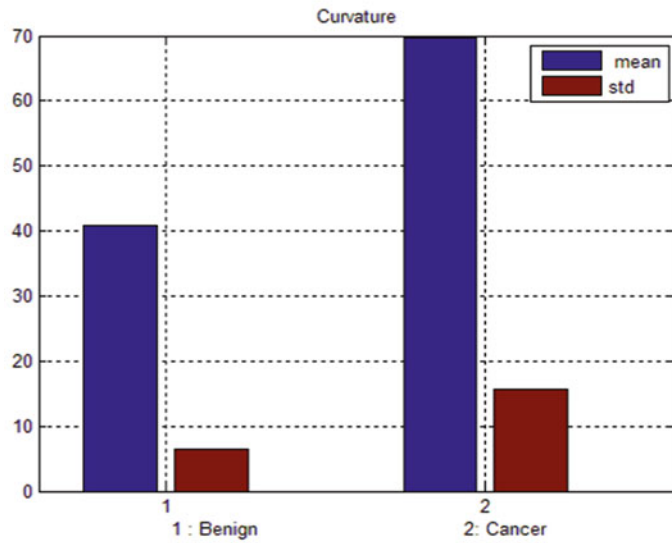


(b)

Fig. 14 Mean and standard deviation for the (a) NC ratio and (b) perimeter feature

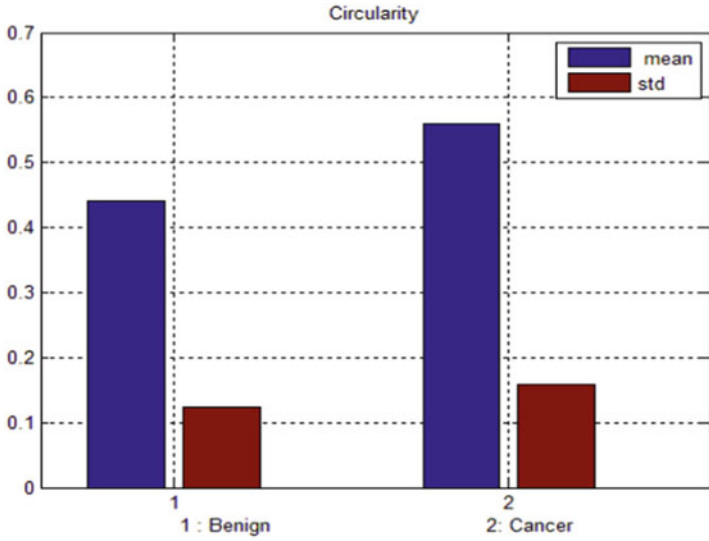


(a)

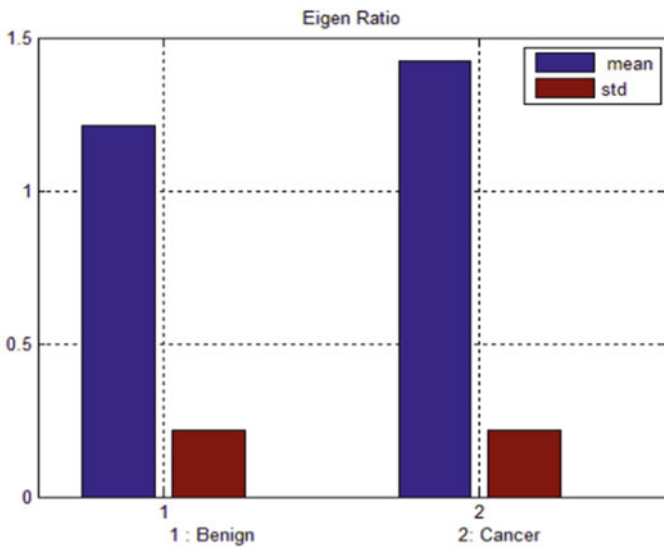


(b)

Fig. 15 Mean and standard deviation for the (a) density feature and (b) curvature feature



(a)



(b)

Fig. 16 Mean and standard deviation for the (a) circularity feature and (b) Eigen ratio feature

depending upon the features that are available. Basically, classification is considered as the heart of pattern recognition and the basis of any diagnostic system. In this work, different classification techniques such as rule-based, artificial neural network (ANN), and support vector machine (SVM) are explained and compared their performances.

5.1 Rule-Based Method

For classification, one of the widely used technique is rule-based method [18]. A set of rules have been used in this CAD system to identify the cancerous cell regions and eliminate those with non-cancerous cell regions. Six diagnosis rules are implemented based on the medical knowledge which is explained in [5]. In the experiments, rule-based method performance is evaluated using individual rules and combined rules.

5.2 Artificial Neural Network

A series of experiments were conducted to evaluate our cell segmentation method. The performance was evaluated by comparing the mean shift segmentation results with the ground-truth data [12]. We used the following assessment criteria for performance measurement: sensitivity, precision, and accuracy.

Artificial neural network (ANN) is one of the significant methods used in the medical field. In this proposed CAD system, to the input data sets, neural network-supervised learning is applied [19]. ANN input data has been normalized in the range of 0–1. ANN algorithm is explained in detail in [12]. The flow chart for the ANN analysis is shown in Fig. 17.

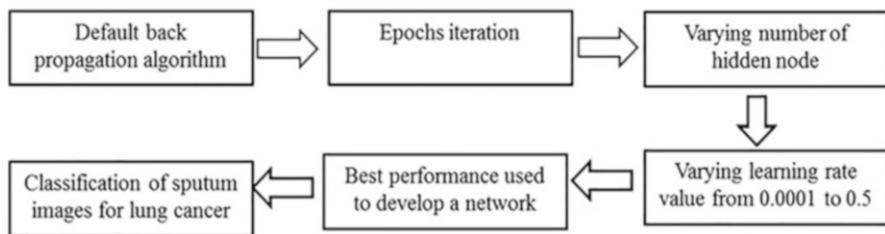


Fig. 17 Flowchart for ANN analysis

Fig. 18 SVM learning approach

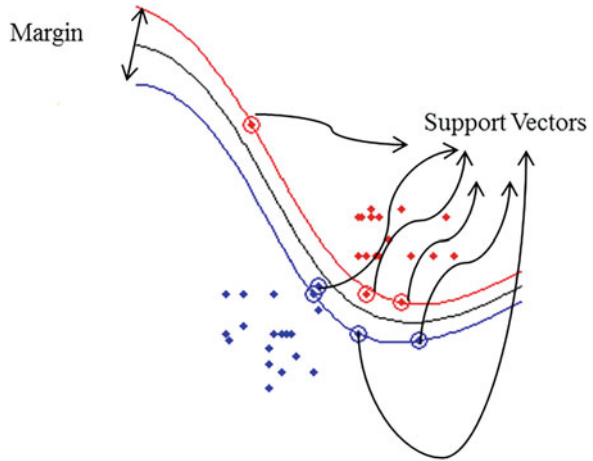


Table 2 Performance matrix

Performance/rules	Rule1	Rule2	Rule3	Rule4	Rule5	Rule6	Rule-based
Sensitivity	94%	89%	91%	92%	94%	86%	89%
Precision	95%	95%	86%	86%	87	92%	87%
Specificity	91%	92%	71%	71%	74	86%	79%
Accuracy	93%	90%	84	85%	87%	86%	85%
Error	7	10	16	15	13	14	15

5.3 Support Vector Machine

Support vector machine (SVM) was first introduced by Vapnik [20]. In order to segregate the data used for training, an optimal separating hyperplane (OSH) is defined. It also used a supervised learning approach [21]. As SVM simultaneously minimize the empirical risk, it is also known as maximum margin classifiers. SVM algorithm is explained in detail in [12]. Figure 18 depicts the SVM learning approach.

5.4 Experiments

A series of experiments were conducted to assess the classification techniques explained previously. The three classifiers discussed above were applied to the input data sets [22]. For these classifiers, sensitivity, specificity, and accuracy have been computed.

Table 2 depicts the results of rule-based classifier. Highest accuracy of 94% is obtained by using rule1.

The next experiment was assigned to the ANN classifier. Tenfold cross-validation [23] was used to validate the output results [24]. The data sets are divided into ten

Fig. 19 Performance matrix of ANN

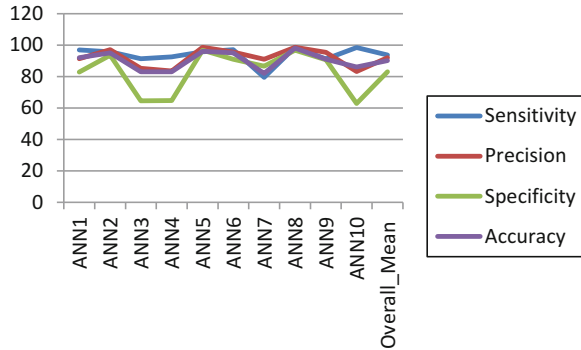
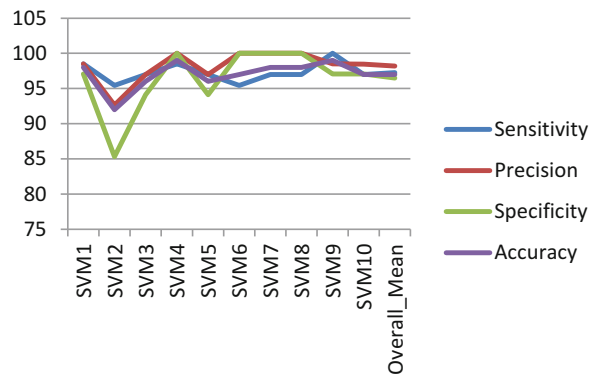


Fig. 20 SVM performance criteria



blocks which consist of the detected features. For all hold out blocks, the testing and training of the system is done on the remaining blocks. The performance can be evaluated by averaging the results over all test blocks. By varying the number of hidden nodes and number of epochs, the best optimized ANN is obtained. The performance criteria [25] of ANN are shown in Fig. 19.

The last experiment is assigned to the SVM classifier. Here also, tenfold cross-validation is used for training and testing. The performance results [26] for SVM are depicted in Fig. 20.

5.5 Comparing Rule-Based, ANN, and SVM Classifiers

After trying different classification techniques: rule-based method, ANN, and SVM classifiers are used for the classification of the sputum cells. We found that the SVM outperforms other classifiers. For the classification of cells into cancerous and non-cancerous, SVM classifier shows the best performance and less classification errors which proves to be a stable and reliable technique for our CAD system. The performances of all the classifiers discussed are shown in Table 3.

We also compared the performance of the rule-based, ANN, and SVM classifiers using the ROC curves [27] with respect to decision threshold as parameter. Figure

Table 3 Performance measurements of the classifier techniques

Performance	Rule-based	ANN	SVM
Sensitivity	89%	94%	97%
Precision	87%	92%	98%
Specificity	79%	83%	96%
Accuracy	85%	90%	97%
Error	15	10	3

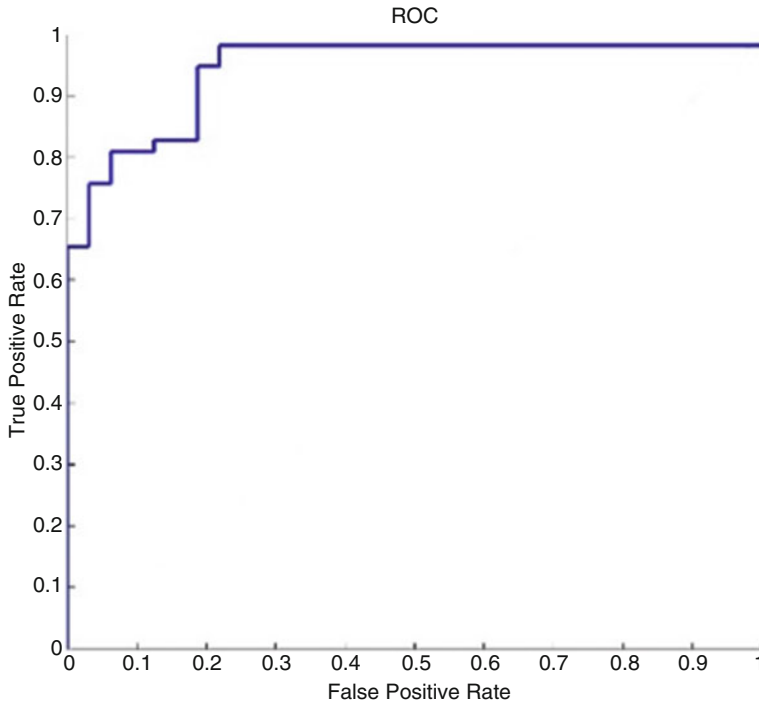


Fig. 21 ROC curve obtained using rule-based classification

21 shows the ROC curves obtained from rule-based, Fig. 22 for ANN and Fig. 23 for SVM classifier. From the ROC curves, SVM shows a clear superiority with highest accuracy.

6 Chapter 6: Performance Evaluation with Previous CAD Systems

The new proposed CAD system uses sputum color images for the prediction of lung cancer at its earlier stage. The proposed CAD system is compared with the previous one as explained in [28]. Our CAD system shows better results with an accuracy and

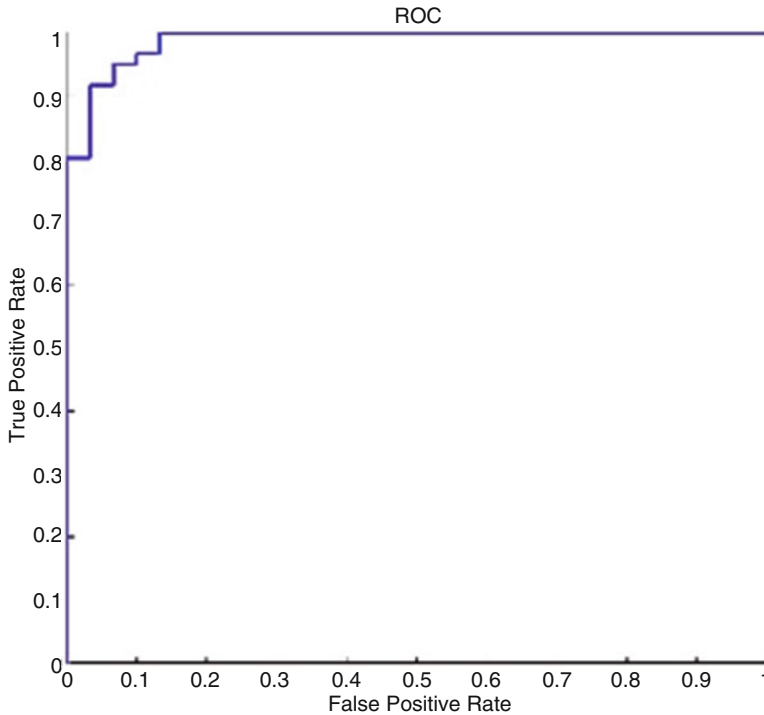


Fig. 22 ROC curve obtained using ANN classification

sensitivity of 97%, precision of 98% and specificity of 96% when compared with the previous CAD system with an accuracy of 85%, sensitivity of 93%, precision of 86%, and specificity of 70%.

7 Chapter 7: Conclusions and Future Works

The following conclusions can be drawn:

1. A new CAD system for detecting lung cancer has been developed and tested successfully on 100 sputum images. The proposed CAD system analyzes the sputum images for classification. The detection and extraction of the sputum cell becomes more accurate if the color space resolution is high. The Bayesian classification achieved an accuracy of 98%. The mean shift approach exhibits better performance with an accuracy of 87% when compared with the HNN technique in the segmentation process. The performance of SVM is found to be superior compared to other classifiers in the detection of cancerous and non-cancerous cells with an accuracy of 97%. The proposed CAD system obtained better accuracy, sensitivity, precision, and specificity.

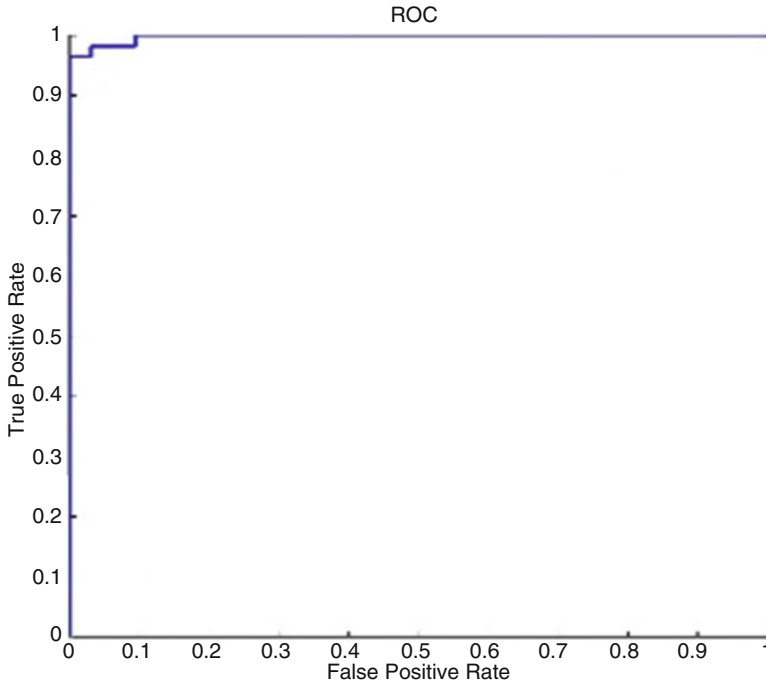


Fig. 23 ROC curve obtained using SVM classification

2. During the implementation of the CAD system, several challenges were faced such as irregularities in the cytoplasm. The current techniques are not suitable for some types of sputum cells, such as dysplastic cells. Therefore, further methods are needed for diagnosing these cells. Active contour method can be used for the segmentation in order to solve this problem. Otsu’s method is also preferred. These issues can be overcome by using a large data set.

Acknowledgments First of all, I would like to thank Allah the most gracious, the most merciful for giving me the strength, patience, power, and endurance during my work.

I would like to express my sincere appreciation and gratitude to my co-authors, Prof. Hussain Al-Ahmad and Prof. Naoufel Werghi, who guided and helped me whenever I was in need. Their constant encouragement, support, enthusiasm, immense knowledge, and invaluable suggestions helped me to bring this work to a successful end. I will certainly keep very good memories of the time we have spent together in discussing this work.

I am grateful to Zayed University Higher Management and to the Office of Research for supporting my research. Special thanks go to H. E. Noura Al Kaabi, the Minister of Culture and Youth for the United Arab Emirates and the president of Zayed University, and Dr. Michael Allen the assistant provost for Research for their immense help, continuous support, and encouragement.

I am deeply and forever indebted to my family. Words cannot express how grateful I am to my mother who took great care of my children during my study; she deserves special mention for her inseparable support and prayers; without her help, this work would have not been possible.

Words fail me to express my appreciation to my husband Mr. Ahmed AlShamsi, whose dedication, love, and persistent confidence in me has taken the load off my shoulder. Special thanks to my kids Alyazia, Mahra, and Mariam and to my sons Mohammed and Sultan; they did suffer from being away from their mom for a long period. I wish they would appreciate this work and feel proud of their mother when they grow up.

References

1. "Cancer Facts & Figures 2021." [Online]. Available: <http://www.cancer.org/research/cancerfactsfigures/cancerfactsfigures/cancer-facts-figures-2021>. [Accessed: 25-Feb-2021].
2. F. Taher, N. Werghi, H. Al-Ahmad and C. Donner, "Extraction and Segmentation of sputum cells for Lung Cancer Early Diagnosis", *Algorithms Journal of Machine Learning for Medical Imaging*, pp. 512–531, vol. 6, August 2013.
3. F. Taher and R. Sammouda, "Lung cancer detection by using artificial neural network and fuzzy clustering methods," *2011 IEEE GCC Conference and Exhibition (GCC)*, 2011, pp. 295–298.
4. M. Sammouda, R. Sammouda, N. Niki, and K. Mukai, "Segmentation and analysis of liver cancer pathological color images based on artificial neural networks," in *1999 International Conference on Image Processing, 1999. ICIP 99. Proceedings*, 1999, vol. 3, pp. 392–396 vol.3.
5. F. Taher, N. Werghi and H. Al-Ahmad, "Rule Based Classification of Sputum Images for Early Lung Cancer Detection", *Proceedings of IEEE*, pp. 29–32, 2015.
6. F. Taher, N. Werghi, H. Al-Ahmad, "Automatic Sputum Color Image Segmentation for Lung Cancer Diagnosis", *KSII Transactions on Internet and Information System*, pp. 68–80, vol. 7, no. 1, January 2013.
7. Werghi, N. Donner, C. Taher, F. Al-Ahmad, H. , "Detection and segmentation of sputum cell for early lung cancer detection", In *Proceedings of the 2012 19th IEEE International Conference on Image Processing (ICIP), Orlando, FL, USA*, 30 September–3 October 2012; pp. 2813–2816.
8. R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification", New York: Wiley, 2001.
9. Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, 1995.
10. R. Sammouda and F. Taher, "Comparison of Hopfield Neural Network and Fuzzy Clustering in Segmenting Sputum Color Images for Lung Cancer Diagnosis," *WSEAS Trans. Biol. Biomed.*, vol. 3, no. 11, pp. 629–637, 2006
11. K. Fukunaga, "Introduction to statistical pattern recognition", Boston: Academic Press, 1990.
12. F. Taher, Naoufel Werghi and Hussain Al-Ahmad, "Computer Aided Diagnosis System for Early Lung Cancer Detection", *Algorithms*, vol. 8, no. 4, pp. 1088–1110, Nov. 2015.
13. K.-S. Chan, C.-G. Koh, and H.-Y. Li, "Mitosis-targeted anti-cancer therapies: where they stand," *Cell Death Dis.*, vol. 3, no. 10, p. e411, Oct. 2012.
14. M. S. Nixon and A. S. Aguado, "Feature extraction & image processing for computer Vision", Oxford: Academic Press, 2012.
15. E. R. Davies, "Computer and machine vision theory, algorithms, practicalities", *Waltham, Mass.: Elsevier*, 2012.
16. A. E. Huque, "Shape analysis and measurement for the HeLa cell classification of cultured cells in high throughput screening", Höskolan, 2006.
17. H. Anton and C. Rorres, "Elementary linear algebra: applications", version. New York: Wiley, 2005.
18. P.-E. Danielsson, Q. Lin, and Q.-Z. Ye, "Efficient Detection of Second-Degree Variations in 2D and 3D Images," *J. Vis. Commun. Image Represent.*, vol. 12, no. 3, pp. 255–305, Sep. 2001.
19. M. X. Ribeiro, C. Traina, C. Traina, and P. M. Azevedo-Marques, "An Association Rule-Based Method to Support Medical Image Diagnosis with Efficiency," *IEEE Trans. Multimed.*, vol. 10, no. 2, pp. 277–285, 2008.

20. V. N. Vapnik, "Statistical learning theory", New York [u.a.]: Wiley, 1998.
21. M. L. Astion and P. Wilding, "The application of backpropagation neural networks to problems in pathology and laboratory medicine," *Arch. Pathol. Lab. Med.*, vol. 116, no. 10, pp. 995–1001, Oct. 1992.
22. P. Dayan and L. F. Abbott, "Theoretical neuroscience computational and mathematical modeling of neural systems", Cambridge, Mass.: Massachusetts Institute of Technology Press, 2001.
23. B. Schölkopf and A. J. Smola, "Learning with Kernels: support vector machines, regularization, optimization and beyond", Cambridge [Mass.]: London, 2002.
24. X. Ye, X. Lin, J. Dehmeshki, G. Slabaugh, and G. Beddoe, "Shape-Based Computer-Aided Detection of Lung Nodules in Thoracic CT Images," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 7, pp. 1810–1820, Jul. 2009.
25. N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods", Cambridge, U.K.; New York: Cambridge University Press, 2012.
26. L. Weruaga and B. Kieslinger, "Tikhonov training of the CMAC neural network," *IEEE Trans. Neural Netw.*, vol. 17, no. 3, pp. 613–622, 2006.
27. K. Kancharla and S. Mukkamala, "Early lung cancer detection using nucleus segmentation based features," in 2013 *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2013, pp. 91–95.
28. F. Taher and R. Sammouda, "Identification of Lung Cancer Based on Shape and Color," in *4th International Conference on Innovations in Information Technology, 2007. IIT '07*, 2007, pp. 481–485.

An Optimal Model Selection for COVID-19 Disease Classification



Pramod Gaur, Vatsal Malaviya, Abhay Gupta, Gautam Bhatia, Bharavi Mishra, Ram Bilas Pachori, and Divyesh Sharma

1 Introduction

Coronaviruses belong to a large family of viruses that are in turn responsible for causing respiratory infections ranging from a mild cold to some severe diseases such as middle east respiratory syndrome (MERS) [1] and severe acute respiratory syndrome (SARS) [2]. The most recently discovered coronavirus causing coronavirus disease is named COVID-19. The coronavirus disease COVID-19 turned out to be a global pandemic [3, 4] as declared by World Health Organization (WHO) as on 11th March, affecting more than 188 countries and causing deaths of more than 5,02,000 people worldwide as on date 29th July. The outbreak was first seen in Wuhan, China in December '19. Multiple parameters have been studied by [5] to impact analysis of COVID-19 disease. The virus is currently reported to transmit between people in close contact making it a type of communicable disease [6]. Also, the current research as of July 2020 reported that virus droplets [7] generated while taking remains active for around ten minutes. This communicable nature of COVID-19 makes it an even harder disease to tackle making pandemic global, social, and

P. Gaur (✉)

Department of Computer Science, Birla Institute of Technology & Science, Dubai, UAE

V. Malaviya · A. Gupta, G. Bhatia · B. Mishra

Department of Computer Science & Engineering, The LNMIIT, Jaipur, India

e-mail: 17ucc003@lnmiit.ac.in; 17ucs058@lnmiit.ac.in; bharavi@lnmiit.ac.in

R. B. Pachori

Department of Electrical Engineering, Indian Institute of Technology Indore, Indore, India

e-mail: pachori@iiti.ac.in

D. Sharma

Department of Cardiology, Western Health and Social Care Trust, Altnagelvin Hospital, Londonderry, UK

economic disruption. Common symptoms include cough, fever, fatigue, breathing problems followed by loss of senses of taste and smell [8, 9]. There is no known vaccine or cure so the primary treatment is only supportive therapy. With the regular increase in the spread, the main problem faced by doctors and medical staff is the unavailability of a quick reporting mechanism that can report a person's COVID report. With the currently available equipment, it takes around 24 h after the sample is given, and because of virus communicable nature, a suspected person has to be quarantined making quarantine centers to be completely filled. Hence, the world is currently facing a great depression with an overload of work on medical staff and doctors.

This chapter covers the variation in performance of COVID-19 identification model with change in optimizer and architecture of models. Stochastic gradient descent and Adam [10] optimizer have been used for the study with architectures such as basic convolutional neural network (CNN), residual networks (ResNet), and densely connected networks (DenseNet). This chapter revolves around the quest of finding the right set of parameters, optimizer algorithm fitting in the set of architectures. It will cover the change in performance with smallest details such as hyper-parameter value, to the change in optimizer algorithm, covered across all the model architectures and applying different deep learning techniques to cook a model with best performance.

The aims of this chapter are:

1. To overcome the less sensitivity of RT-PCR, chest CT images are used in this chapter to detect and diagnose COVID-19.
2. To identify the best deep learning models for classification of COVID-19 (+) and COVID-19 (−) patients.
3. To perform optimization of various hyper-parameters such as learning rate, batch size for mini-batch gradient, a beta for momentum, RMSprop (root mean square propagation) in Adam algorithm, and weight decay.
4. The proposed model has been utilized for feature extraction by changing and by optimizing its learned weights and learning rate on the ImageNet dataset in a convolutional neural structure.
5. To compare the proposed work with other state-of-the-art methods in terms of various performance metrics such as accuracy, F1 score, AUC measure, sensitivity, and specificity.

The rest of the chapter is organized starting with related works followed by explanation of dataset, optimization algorithms, deep learning architectures (CNN, ResNet, and DenseNet), transfer learning, and finally results and discussion with conclusion followed by references ending the chapter.

2 Related Works

Several approaches have been used for classifying COVID-19 positive patients through chest CT scan [11–15]. Deep learning along with some image pre-processing is the most popular way that is being used. In year 2020 [16], research group in their work applied CNN for feature selecting and fully connected network for classifying COVID-19 patients. They build a transfer learning neural network that was based on an inception network [17]. Their network can be divided into parts where the first part uses the inception network for generating feature vectors from the images and the second part for prediction and classification. Their model was able to achieve a total accuracy of 83% with 80.5% specificity and 84% sensitivity for validation. Although with external testing, the total accuracy dropped to 73%. Similarly, another research group in year 2020 [18] adopted various deep transfer learning models for the detection of COVID-19 from chest CT scan images. The dataset used for this work consists of 742 CT images that were organized into 3 folders train, validation, and test. In their work, they compared AlexNet [19], VGGNet16 [20], VGGNet19 [21], GoogleNet [22], and ResNet50 [23] for classification of COVID-19 and claimed ResNet50 to be the best for classification if data augmentation was applied. In their work, they claimed 82.91% testing accuracy with ResNet50. Without data augmentation, they claimed a testing accuracy of 67.34% for AlexNet, 72.36% for VGGNet16, 76.88% for VGGNet19, 75.38% for GoogleNet, and 76.38% for ResNet50.

3 Dataset

For this study, we are using a publicly available SARS-COV-CT dataset that contains 1252 CT scan images of positive and 1230 negative SARS-CoV-2 (COVID-19) CT scan images. This data was collected from real patients in hospitals in Sao Paulo, Brazil. The details of each patient are skipped for each patient due to the data privacy of patients. The dataset is also available on Kaggle (<https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset>).

Figure 1 shows some sample images from the dataset where the first two rows show CT scan images suffering from COVID-19 and the last two show normal CT scan images with no COVID-19. These images are randomly picked from the dataset itself.

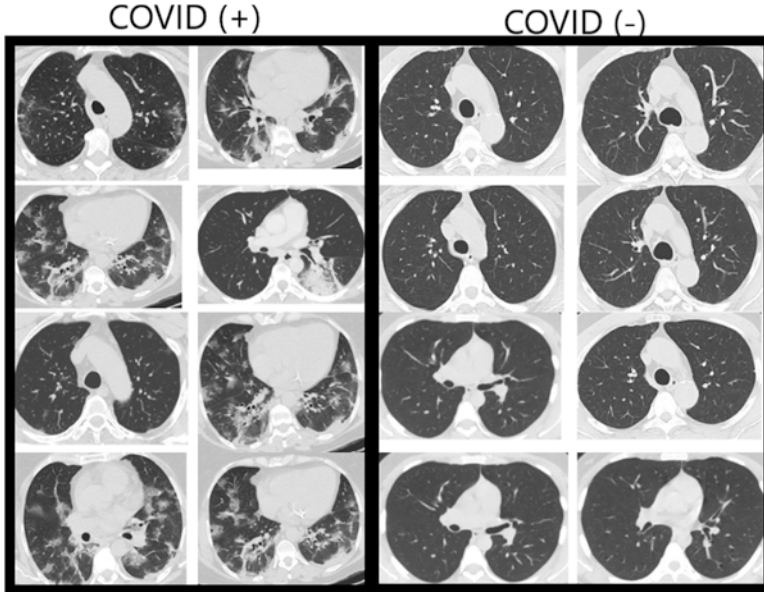


Fig. 1 Sample images of CT scans from dataset

4 Optimization Algorithms

4.1 Stochastic Gradient Descent (SGD) [24]

SGD is used because gradient descent is slow when working on a big dataset with many parameters. It is especially useful when there is redundancy in data because it looks at only one sample or a small subset or mini-batch at a time for each step. The main objective of SGD is to minimize the sum of squared residuals. In SGD, a random point is used to calculate parameters until the minimum point is reached. Whenever a new point is encountered, parameters are calculated again, and then it is multiplied with the learning rate after that new parameters are acquired using the difference between the old parameters and the newly calculated one to get the newly updated parameters. This process is repeated until the gradient is almost zero. Due to random point selection, the computation is reduced significantly. Equations (1, 2) are used to update parameters where W represents the weight and b represents bias value for each layer weight and bias updated with gradients of weight (dW) and bias (db).

$$W = W - \alpha * dW \quad (1)$$

$$b = b - \alpha * db. \quad (2)$$

4.2 Adam

It is a method that computes adaptive learning rate for each algorithm. It is a combination of AdaGrad and RMSprop [25] algorithm. Momentum is calculated using Eqs. (3, 4); in these equations, Adam keeps an average of previous gradients (dW , dB) with β_1 hyper-parameter for weight momentum (vdW) and bias momentum (vdb). Adam is used for various deep learning projects, and some recent works are [10, 26]. RMSprop is calculated using Eqs. (5, 6); in these equations, Adam keeps an average of past squared gradients and stores in the variables RMSprop weight (sdW) and RMSprop bias (sdb). Usually, in initial time, they are biased toward 0. For this, bias correction is done using Eqs. (7, 8, 9, 10), where t is the number of iterations, and now the corrected momentum value and RMSprop value for weights and bias are stored in vdW^c , vdb^c , sdW^c , and sdb^c . After this, with use of corrected momentum and RMSprop values, the final values of weight and bias are calculated using Eqs. (11,12), and here the value of ϵ is taken as 10^{-8} for this study. ϵ prevents from divide-by-zero exception.

$$vdW = (\beta_1 * vdW) + (1 - \beta_1 * vdW) \tag{3}$$

$$vdb = (\beta_1 * vdb) + (1 - \beta_1 * vdb) \tag{4}$$

$$sdW = (\beta_2 * vdW) + (1 - \beta_2 * vdW) \tag{5}$$

$$sdb = (\beta_2 * vdb) + (1 - \beta_2 * vdb) \tag{6}$$

$$vdW^c = \frac{vdW}{(1 - \beta_1)^t} \tag{7}$$

$$vdb^c = \frac{vdb}{(1 - \beta_1)^t} \tag{8}$$

$$sdW^c = \frac{sdW}{(1 - \beta_2)^t} \tag{9}$$

$$sdb^c = \frac{sdb}{(1 - \beta_2)^t} \tag{10}$$

$$W = W - \frac{\alpha * vdW^c}{\sqrt{sdW^c + \epsilon}} \tag{11}$$

$$b = b - \frac{\alpha * vdb^c}{\sqrt{sdb^c + \varepsilon}}. \quad (12)$$

5 Deep Learning Architectures

5.1 CNN

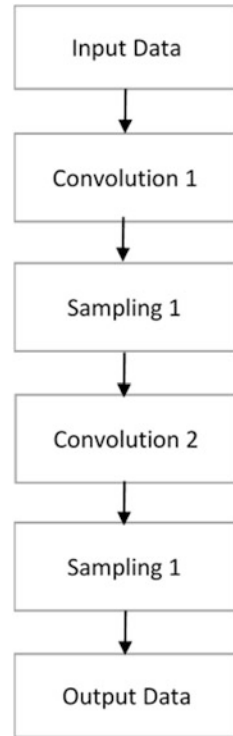
Convolutional neural network (CNN) [27] is an artificial neural network that is most commonly used for image analysis. It is basically an artificial neural network with some type of specialization to detect patterns and make sense from them. CNN consists of an input layer that receives the raw input, followed by some hidden layers known as convolutional layers (Fig. 2), which is also referred to as a learning layer that is from where its name derived from. A convolution layer neuron receives the input, applies some transformation (convolutional operations Fig. 2) to it, and transfers it to the next layer. A convolutional operation is the dot product between the filter and image whose size is the same as that of the filter. As we go deep into the layers, the filters are able to detect more sophisticated objects. For activation, many methods are used, two of them are sigmoid [28] and rectified linear unit (ReLU). ReLU is a threshold layer that applies $\max(0, x)$ as its activation function [29] for spatial dimension reduction of the data max-pooling layer is used [30]. To normalize the data between 0 and 1, a normalized exponential function in the softmax layer is used. After all the hidden layers, an output layer is there that provides the output of the CNN along with loss function and labels.

5.2 Residual Neural Network (ResNet)

The ResNet [31, 32] was proposed in 2015 and is one of the most famous architectures. In ResNet, skipping a step and taking a shortcut method came out to be different than the previously existing architectures. Due to this feature, it can help in training huge layers faster. The basic version of ResNet includes double or triple skips. These skips consist of nonlinearities and normalization (batch normalization).

5.3 Densely Neural Network (DenseNet)

To overcome the problem of gradient vanishing due to passage through many layers in CNN, DenseNet was developed [33]. DenseNet is a deeper version of CNN that gives more accurate and efficient results. DenseNet helps in reducing the number of parameters, and also it encourages the reusability of the features. In DenseNet, each

Fig. 2 CNN architecture

layer is connected to another layer in a feed-forward method. In DenseNet, feature maps from previous layers are concatenated onto the inputs of future layers, which makes a very deep feature map in the spatial resolution. Each layer receives signals from all its preceding layer; this input is connected channel-wise, which makes each layer thinner and provides computational efficiency. To perform downsampling on the feature map, pooling layers are used. The pooling layer reduces the feature map size. DenseNet strengthens the feature propagation. There are different types of DenseNet architectures—DenseNet 121, DenseNet 169, DenseNet 201.

6 Transfer Learning

It is a technique in which a model is designed for one task and then reused as the starting point for the other tasks. Figure 3 elaborates the difference between traditional machine learning techniques and transfer learning. In the proposed study, since there is a limited amount of dataset and hardware architecture, pre-trained transfer learning models trained for classification on similar tasks are used. Transfer learning can be used in two ways by either fine-tuning or freezing the gradients. In the fine-tuning approach, the gradients of pre-trained models are also updated,

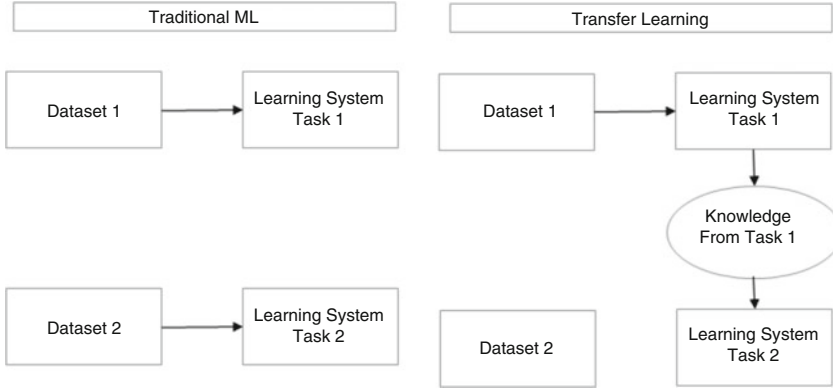


Fig. 3 Transfer learning

while in freezing the gradients other than the top layer that is newly added to the layer (which are not pre-trained, added for customization for the task) therefore in the training procedure in the top layer are updated. In this chapter, sufficient dataset size and also hardware requirements were provided, which allowed us to use the fine-tuning approach. Usage of transfer learning in this study is a case of inductive transfer learning.

7 Learning Rate Scheduler

With constant learning rate, training models to global minima in loss could be very difficult. It is quite intuitive to speed up the process at first, while distance is large at first, and as distance to target reduces, speed should also be reduced to prevent overstepping. This idea is implemented with the help of learning rate schedulers [34, 35]. In this project, ReduceLROnPlateau has been used for modifying learning rate as model reaches toward plateau, and learning rate is reduced by a factor of 2–10 as the performance (loss is used to monitor model performance in this study) stagnates with patience in terms of the number of epochs.

8 Results

8.1 Experimental Setup

The baseline setup used for this study is a mobile GPU RTX 2060. PyTorch and OpenCV libraries are used for implementing the algorithm, on Python 3.6.5.

8.2 Training

For every experiment, accuracy, recall, F1 score, and AUC metrics are used for the evaluation of different methods in this study. In this study, different methods are applied to find the best procedure for the identification of COVID-19 patients using CT scans. Data augmentation is employed, and each image was resized to (256, 256), which undergoes random crop of size (224, 224) and horizontal flip for populating the dataset while training procedure. Basic CNN was trained on the dataset with binary cross-entropy loss and Adam optimizer. The training procedure for each model for COVID-19 identification faced a major setback due to the lack of good-quality large dataset. To overcome this problem, transfer learning was applied, and pre-trained models of architecture AlexNet [19], VGG-19 [36], ResNet50 [23], ResNet101 [37], DenseNet121 [38], DenseNet169 [39], and DenseNet201 [40] were fine-tuned to get the best results. For loss, binary cross-entropy loss was used in combination with Adam optimizer for faster convergence, with a learning rate of $1e-4$ and value of betas equal to 0.9 and 0.999, and weight decay of $1e-5$ for preventing the model from overfitting. ReduceLROnPlateau was used as a learning rate scheduler for better convergence.

8.3 Results

In this chapter, a comparison between the performance of the different state-of-the-art deep learning techniques is used to create models for the identification of the COVID-19 using CT scan images. The above-mentioned models were used for this comparative study. The basic results of any model consist of true positive (TP), false positive (FP), false negative (FN), and true negative (TN). The evaluation metrics used for the study were accuracy, recall, F1 score, and AUC. Accuracy is the closeness of the measurements to ground truth, and it is calculated using the following formula (13).

The recall (14) is also used as an evaluation metric for this study, and in binary classification, recall is called sensitivity. The recall is the ratio of correctly predicted positive values to the actual positive values. The recall is an important criterion for deep learning in the field of healthcare. The best model used in the field of healthcare miss-classifies actual positive patients as negative the minimum number of times. Every miss-classification of actual positive patients can create dangerous scenarios even more in terms of COVID-19 disease that spreads in a very rapid manner. Precision (15) is the measure to detect if the false positives are high or not.

AUC measures how well a model is able to distinguish classes. It also tells about the separability of the positive and negative classes. F1 score (16) is the function of precision and recall and is calculated to check the balance between precision and recall.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (13)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (14)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (15)$$

$$\text{F1-score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \quad (16)$$

All the results achieved in this study are represented in Table 1. The comparison is based on the accuracy achieved, AUC, F1 score, recall, and precision.

The best results were achieved by the DenseNet169 with an accuracy of 96%, AUC 0.99, F1 score 0.96, recall 99%, and a precision of 94%. DenseNet169 got the best overall performance by maintaining precision too with recall, and other models achieved excellent recall scores but lacked in terms of precision. The worst results were achieved by CNN with an accuracy of 82% and a F1 score of 84. ResNet50 and ResNet101 have achieved the accuracies of 87% and 90%, AUC as 0.96 and 0.97, recall as 96.5% and 97.5%, F1 scores of 0.87 and 0.90, precision as 85% and 83%, respectively. Other two DenseNet variants 121 and 201 have achieved accuracy between 90% and 91%, AUC between 0.97 and 0.98, recall between 97% and 99%, and precision around 85% in both the variants (Table 2).

Proposed method has found significant increase in performance from other related works. With the help right of a combination of set of hyper-parameters, optimizer algorithm, model architecture, with of help transfer learning, study has found performance jump of around 10% from the state-of-the-art methods.

Table 1 Results computed with the proposed method

Model name	Accuracy (in %)	AUC	F1 score	Recall (in %)	Precision (in %)
CNN	82	0.90	0.84	85	82
ResNet50	87	0.96	0.87	96.5	85
ResNet101	90	0.97	0.90	97.5	83
DenseNet121	90	0.97	0.90	97	85
DenseNet169	96	0.99	0.96	99	94
DenseNet201	90	0.98	0.92	99	84.5

Table 2 Comparison among the existing methods and the proposed method

Model	Accuracy (%)	F1 Score	AUC
DenseNet-169[30]	79.5	0.76	0.90
Transfer learning [30]	87.1	0.88	0.95
Proposed method	96	0.96	0.99

The best results are shown in bold

There are numerous signal decompositions studied to decompose physiological signals. Few of the modalities to record the physiological signals are electrocardiogram (ECG), electroencephalogram (EEG), and electromyogram (EMG). To decompose these physiological signals, there are advance signal decomposition techniques such as empirical mode decomposition (EMD) [41], multivariate EMD (MEMD) [42], EWT [43], and 2-D EWT[44]. Filtering techniques based on the EMD and MEMD methods are used in brain–computer interface (BCI) to handle the inherent non-stationarity nature of the data [45–50]. In future, it will be interesting to study these advance decomposition techniques with the existing models to improve the performance of the system.

9 Discussion

COVID-19 is a global pandemic and therefore also a center of concern for the whole world. During this period, all the doctors are working day and night to resolve this situation; meanwhile, every help that can automate the most vital process can ease the work of all already exhausted doctors and all the front line warriors for the past 7 months. CT scans are the key to automation in the identification of disease-infected patients, and deep learning is the way to achieve this task. Image processing fields have been dominated by the deep learning (DL) models for many times now. Great research opportunities are presented by the development of image processing and DL for the classification of COVID-19 disease using images of CT scans on the chest area of a person. In this chapter, an extension of the state-of-the-art deep learning models for the detection of COVID-19 using chest CT scans is performed, and a comparative study is performed to evaluate the working of different models.

CNN is one of the key factors in the revolution in the deep learning field; using a CNN model in COVID-19 identification training, an accuracy of 85% was acquired and the final accuracy of 82%. But shallow networks were not enough to extract finer details of CT scan, deep CNNs need to be used to overcome this problem. But with vanishing gradients problem and losing many important features in the way deep CNN was not as effective to its potential to overcome this, residual networks should be used. Therefore, ResNets are used in the study to improve performance, and a pre-trained version enabled faster convergence and experience from the previous task to even boost more performance. ResNet is a state-of-the-art techniques with the first to introduce the importance of the residual network in deep neural networks. ResNet50 and ResNet101 are used to identify COVID-19 disease using a CT scan of the chest area of the patient.

ResNet50 and ResNet101 converged at 16 epoch at 0.98 and 0.97, yet they showed clear signs of overfitting, validation accuracy improved till 8–12 epochs up to 90%, and an average around 90% with ResNet101 getting a better score than ResNet50 and later accuracy was declined for both architectures due to overfitting. Figure 4 represents the accuracy vs. epochs curve for ResNet model.

Fig. 4 ResNet, the smoothed curve for training accuracy vs. epochs

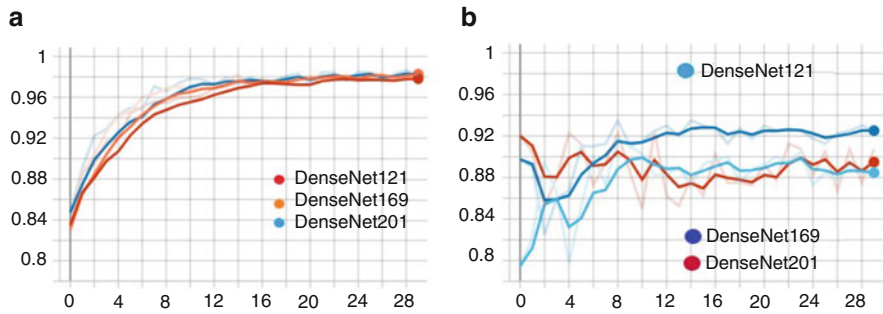
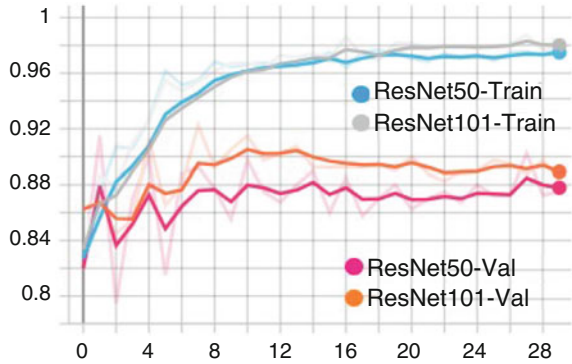


Fig. 5 DenseNet, the smoothed curve for accuracy vs. epochs. (a) Training accuracy vs. epochs. (b) Validation accuracy vs. epochs

DenseNet architecture has been proven to be better in performance as well as fast with significantly less computational parameters. This study has found out DenseNet variations to give the best performance among the rest of the models. Figure 5a represents the evolving performance with epochs for models of DenseNet architectures (i.e., DenseNet121, DenseNet169, DenseNet201) during the training period, while the validation period is covered by Fig. 5b. All the DenseNet variants showed similar promising results, while training, with DenseNet169 and DenseNet201 taking lead in showing convergence with little delay DenseNet121, also converged.

But the case during validation was slightly different. Pre-trained models are used for faster convergence, and DenseNet169 is the best model that converged at 12 epochs with validation accuracy, topping up to 94% and followed by DenseNet121, DenseNet201 at 90%.

Figure 6 shows the performance in an unaltered stage to better represent the history in an accurate fashion. With DenseNet169 scoring the highest overall performance among all the other models when compared to the rest of the architectures and its variants, it will prove to be a vital architecture for further study in future work. With the current settings, it scored an average performance of 93%. Adam

Fig. 6 DenseNet169: accuracy vs. epochs (un-smoothened)

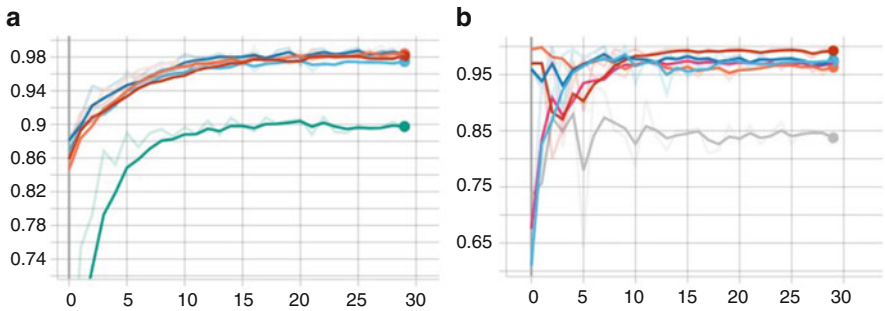


Fig. 7 Recall curve for all the models in training phase (a) and validation phase (b). (a) Training phase: recall vs epochs. (b) Validation phase: recall vs epochs

optimizer with binary classification loss and ReduceLRonPlateau as learning rate scheduler under PyTorch framework DenseNet 169 was able to shine up among the rest.

Deep learning in the medical field also has to focus on recall. If a COVID-19 positive patient is given a green flag, then a spontaneous and hazardous situation follows with him infecting all the other person who comes in contact and hence starting an unstoppable chain. All the models have been designed to focus on recall in recall vs. precision trade-off. Almost all except basic CNN have shown recall above 95% and up to 99% during the training phase represented in Fig. 7a, and for validation Fig. 7b. The validation phase has also maintained recall above 95% and DenseNet201 with the best recall of 99.95% for data containing CT scan images of chest area for 200 COVID-19 positives and 200 COVID-19 negative patients.

Selection of just architecture is not enough to get best model; more fine-tuning has to be done to get better performance. DenseNet169 model will undergo different fine-tuning techniques. For better performances, different optimization algorithms can be applied for better and quicker convergence. Study has experimented with SGD and Adam optimizer, results can be seen in Fig. 8, and accuracy vs. epochs plot shows the difference between the two optimizer algorithms. Adam has performance

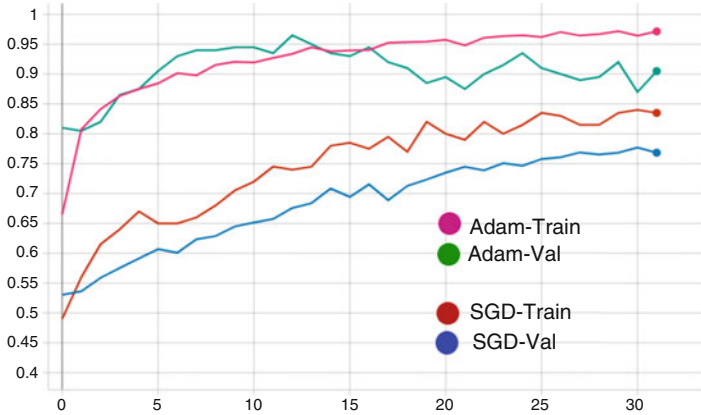


Fig. 8 Adam optimizer vs. SGD optimizer

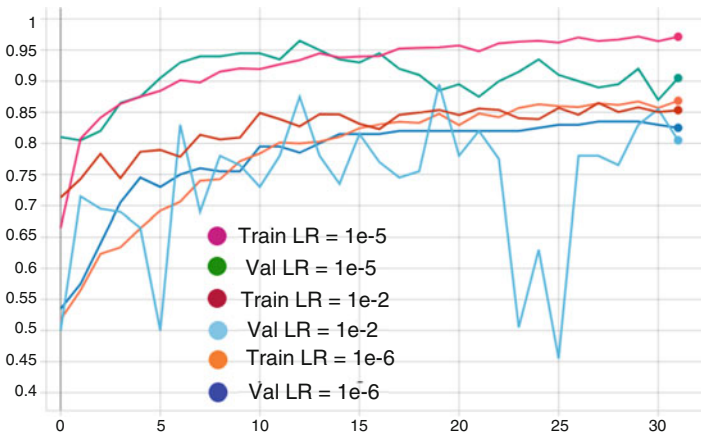


Fig. 9 Adam optimizer with different learning rates

improvement of 5–10% from SGD optimizer and quicker convergence; SGD shows traces of no convergence in 32 epochs, while Adam optimizer found convergence at 15th epoch.

Hyper-parameters also play a big role in DL; some of the most important parameters are learning rate, batch size for mini-batch gradient, beta for momentum, RMSprop in Adam algorithm, and weight decay. Figure 9 shows the change in performance of model with different learning rates. Optimal value of learning rate is most important for any training session in DL. Model is trained on 1e-2, 1e-5, 1e-6, and big learning rate does not allow model to find global minima due to big jumps taken by big learning rate, while too small learning rates have risks for sticking to local minima rather than exploring for global minima. After fine-tuning

in DenseNet169 architecture model with Adam optimizer, and 1e-5 learning rate, a final accuracy of around 96% has been achieved.

10 Conclusion

In this task, the fine-tuning of DL models for COVID-19 disease identification is performed. The architectures that were evaluated include ResNet 50, ResNet 101, DenseNet 121, DenseNet 169, and DenseNet 201. From this experiment, we can conclude that DenseNet 169 tends to yield the best results when classifying COVID-19 CT scan images with a testing accuracy of 96% for the 20th epoch, beating the rest of the architectures. Therefore, DenseNet 169 is a good choice for COVID-19 identification with the CT scan images. However, DenseNet requires a sensible computing time to achieve best in class classification, so future research needs to be done for improving its performance time.

References

1. A. Zumla, D. S. Hui, S. Perlman, Middle east respiratory syndrome, *The Lancet* 386 (9997) (2015) 995–1007.
2. J. Peiris, Y. Guan, K. Yuen, Severe acute respiratory syndrome, *Nature medicine* 10 (12) (2004) S88–S97.
3. A. Lee, Wuhan novel coronavirus (covid-19): why global control is challenging?, *Public health* 179 (2020) A1.
4. A. Tavakoli, K. Vahdat, M. a. Keshavarz, Novel coronavirus disease 2019 (covid-19): An emerging infectious disease in the 21st century, *Iranian South Medical Journal* 22 (6). <https://doi.org/10.29252/ismj.22.6.432>.
5. H. Raj, R. K. Mishra, Data analysis of novel coronavirus based on multiple factors, in: 2020 Seventh International Conference on Information Technology Trends (ITT), 2020, pp. 135–139. <https://doi.org/10.1109/ITT51279.2020.9320887>.
6. J. Chin, et al., Control of communicable diseases manual.
7. L. Bourouiba, Turbulent Gas Clouds and Respiratory Pathogen Emissions: Potential Implications for Reducing Transmission of COVID-19, *JAMA* 323 (18) (2020) 1837–1838. <https://doi.org/10.1001/jama.2020.4756>.
8. Gautret, et al., Lack of nasal carriage of novel corona virus (HCoV-EMC) in French Hajj pilgrims returning from the Hajj 2012, despite a high rate of respiratory symptoms, *Clinical Microbiology and Infection* 19 (7) (2013) E315–E317.
9. X. Jin, J.-S. Lian, J.-H. Hu, J. Gao, L. Zheng, Y.-M. Zhang, S.-R. Hao, H.-Y. Jia, H. Cai, X.-L. Zhang, et al., Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (covid-19) with gastrointestinal symptoms, *Gut* 69 (6) (2020) 1002–1009.
10. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.
11. X. Wang, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, C. Zheng, A weakly-supervised framework for covid-19 classification and lesion localization from chest CT, *IEEE Transactions on Medical Imaging*.

12. D. Singh, V. Kumar, M. Kaur, Classification of covid-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks, *European Journal of Clinical Microbiology & Infectious Diseases* (2020) 1–11.
13. S. R. Nayak, D. R. Nayak, U. Sinha, V. Arora, R. B. Pachori, Application of deep learning techniques for detection of covid-19 cases using chest X-ray images: A comprehensive study, *Biomedical Signal Processing and Control* 64 (2020) 102365.
14. P. K. Chaudhary, R. B. Pachori, Automatic diagnosis of covid-19 and pneumonia using FBD method, in: *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2020, pp. 2257–2263.
15. P. K. Chaudhary, R. B. Pachori, FBSED based automatic diagnosis of covid-19 using X-ray and CT images, *Computers in Biology and Medicine* (2021) 104454.
16. S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, et al., A deep learning algorithm using CT images to screen for corona virus disease (covid-19), *MedRxiv*.
17. X. S. Poma, E. Riba, A. Sappa, Dense extreme inception network: Towards a robust CNN model for edge detection, in: *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 1923–1932.
18. M. Loey, G. Manogaran, N. E. M. Khalifa, A deep transfer learning model with classical data augmentation and CGAN to detect covid-19 from chest CT radiography digital images.
19. M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal, V. K. Asari, The history began from AlexNet: A comprehensive survey on deep learning approaches, *arXiv preprint arXiv:1803.01164*.
20. L. Wang, Y. Xiong, Z. Wang, Y. Qiao, Towards good practices for very deep two-stream ConvNets, *arXiv preprint arXiv:1507.02159*.
21. W. Wu, D. Sun, Multiple deep CNN for image annotation, in: *Tenth International Conference on Graphics and Image Processing (ICGIP 2018)*, Vol. 11069, International Society for Optics and Photonics, 2019, p. 110691S.
22. P. Ballester, R. M. Araujo, On the performance of GoogLeNet and AlexNet applied to sketches, in: *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
23. L. Wen, X. Li, L. Gao, A transfer convolutional neural network for fault diagnosis based on resnet-50, *Neural Computing and Applications* (2019) 1–14.
24. L. Bottou, Large-scale machine learning with stochastic gradient descent, in: *Proceedings of COMPSTAT'2010*, Springer, 2010, pp. 177–186.
25. T. Tieleman, G. Hinton, Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude, *COURSERA: Neural networks for machine learning* 4 (2) (2012) 26–31.
26. Z. Zhang, Improved Adam optimizer for deep neural networks, in: *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, IEEE, 2018, pp. 1–2.
27. S. Albawi, T. A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network, in: *2017 International Conference on Engineering and Technology (ICET)*, IEEE, 2017, pp. 1–6.
28. J. Han, C. Moraga, The influence of the sigmoid function parameters on the speed of backpropagation learning, in: *International Workshop on Artificial Neural Networks*, Springer, 1995, pp. 195–201.
29. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, in: *Advances in neural information processing systems*, 2012, pp. 1097–1105.
30. J. Zhao, Y. Zhang, X. He, P. Xie, Covid-CT-dataset: a CT scan dataset about covid-19, *arXiv preprint arXiv:2003.13865*.
31. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
32. S. Targ, D. Almeida, K. Lyman, Resnet in resnet: Generalizing residual architectures, *arXiv preprint arXiv:1603.08029*.
33. F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, K. Keutzer, DenseNet: Implementing efficient ConvNet descriptor pyramids, *arXiv preprint arXiv:1404.1869*.

34. J. Konar, P. Khandelwal, R. Tripathi, Comparison of various learning rate scheduling techniques on convolutional neural network, in: 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), IEEE, 2020, pp. 1–5.
35. J. Dodge, S. Gururangan, D. Card, R. Schwartz, N. A. Smith, Show your work: Improved reporting of experimental results, arXiv preprint arXiv:1909.03004.
36. L. Wen, X. Li, X. Li, L. Gao, A new transfer learning based on VGG-19 network for fault diagnosis, in: 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, 2019, pp. 205–209.
37. R. U. Khan, X. Zhang, R. Kumar, E. O. Aboagye, Evaluating the performance of ResNet model based on image recognition, in: Proceedings of the 2018 International Conference on Computing and Artificial Intelligence, 2018, pp. 86–90.
38. L. Sarker, M. M. Islam, T. Hannan, Z. Ahmed, Covid-DenseNet: A deep learning architecture to detect covid-19 from chest radiology images.
39. R. A. Aral, Ş. R. Keskin, M. Kaya, M. Hacıömeroğlu, Classification of TrashNet dataset based on deep learning models, in: 2018 IEEE International Conference on Big Data (Big Data), IEEE, 2018, pp. 2058–2062.
40. S.-H. Wang, Y.-D. Zhang, Densenet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 16 (2s) (2020) 1–19.
41. N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, H. H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* 454 (1971) (1998) 903–995.
42. N. Rehman, D. P. Mandic, Multivariate empirical mode decomposition, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 466 (2117) (2010) 1291–1302.
43. J. Gilles, Empirical wavelet transform, *IEEE transactions on signal processing* 61 (16) (2013) 3999–4010.
44. J. Gilles, G. Tran, S. Osher, 2D empirical transforms. Wavelets, ridgelets, and curvelets revisited, *SIAM Journal on Imaging Sciences* 7 (1) (2014) 157–186.
45. P. Gaur, K. McCreadie, R. B. Pachori, H. Wang, G. Prasad, An automatic subject specific channel selection method for enhancing motor imagery classification in EEG-BCI using correlation, *Biomedical Signal Processing and Control* 68 (2021) 102574.
46. P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, R. B. Pachori, H. Wang, A sliding window common spatial pattern for enhancing motor imagery classification in EEG-BCI, *IEEE Transactions on Instrumentation and Measurement* 70 (2021) 1–9.
47. P. Gaur, R. B. Pachori, H. Wang, G. Prasad, A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry, *Expert Systems with Applications* 95 (2018) 201–211.
48. P. Gaur, R. B. Pachori, H. Wang, G. Prasad, An Automatic Subject Specific Intrinsic Mode Function Selection for Enhancing Two-Class EEG-Based Motor Imagery-Brain Computer Interface, *IEEE Sensors Journal* 19 (16) (2019) 6938–6947.
49. P. Gaur, K. McCreadie, R. B. Pachori, H. Wang, G. Prasad, Tangent space features-based transfer learning classification model for two-class motor imagery brain-computer interface, *International journal of neural systems* 29 (10) (2019) 1950025.
50. P. Gaur, R. B. Pachori, H. Wang, G. Prasad, A multivariate empirical mode decomposition based filtering for subject independent BCI, in: 27th Irish Signals and Systems Conference (ISSC), IEEE, 2016, pp. 1–7.

Index

A

Adaptive-neuro fuzzy classifier (ANFC), 235, 242, 246–247, 249, 252
AI agent, vi, 195, 196, 199–204
Amputees, 110, 111, 119
Anonymization, v, 4–5, 10, 15, 17, 18, 20, 21
Artificial intelligence (AI), v–vii, 58, 195–204, 287–305, 309, 311, 314, 337–370, 373–395
Artificial neural network (ANN), vii, 75, 115, 185, 259–269, 276, 360, 361, 373, 390–394, 404
Autoencoder, 36, 145, 155–158, 160, 162–165, 233, 235, 242, 244, 245, 249
Automatic Speaker Verification (ASV), 2–5, 15, 19, 21, 22, 24

B

Background appearance, 299
Bayesian, 374, 377–380, 394
Binary cross-entropy loss, 79, 84, 407
Bits per pixel (BPP), 162, 164, 215, 221–224, 227
Breast cancer, v, 31–55, 90, 231, 293, 305, 311
Breast ultrasound images, vii, 232–234, 236, 238, 241, 254

C

Cancer detection, v–vii, 31–55, 183–194, 231, 305, 373–395

Cardiovascular disease (CVD), 337–339, 341, 350, 368
Chest X-ray (CXR), v–vi, 58, 59, 61, 66, 67, 70, 146, 216, 221–223, 300, 301
Chunking, 196
Classification, v–vii, 32–36, 38–41, 45–50, 55, 57–71, 74, 75, 89, 100, 107–119, 125–146, 167–180, 183–195, 233, 236, 240, 243, 244, 246–251, 260, 262, 271–284, 289, 300–305, 310, 317, 325, 327, 330, 351, 352, 360, 364, 366, 368, 373, 374, 377–380, 384, 386–395, 399–413
Cognition, 199–203
Color images, 128, 208, 213, 214, 217, 219, 220, 373, 374, 393
Compression, v, vii, 36–38, 44–45, 53, 54, 155, 157, 162, 207–228, 331
Content-Based Image Retrieval (CBIR), vi, 123–147, 310
Content-based medical image retrieval (CBMIR), 144–147
Convolutional neural network (CNN), v, vii, 41, 45–50, 58, 89–91, 94, 96, 99, 136, 138, 139, 141–146, 155, 156, 163, 164, 178, 179, 184, 275–277, 281–283, 289, 300–302, 304, 305, 309–333, 341, 352, 368, 369, 400, 401, 404, 405, 407–409, 411
Cosine loss, v, 39, 46–47, 55
COVID-19, vi, vii, 57–71, 146, 300, 301, 313, 399–413

D

- Data analytics, vi, 153–165
- Deep learning (DL), v–vii, 5, 35, 55, 58, 61, 69, 70, 74–76, 79, 81, 83, 84, 87, 91, 93, 99, 100, 125, 135–143, 147, 153–165, 168, 178, 183–185, 193, 271–284, 288, 289, 300–303, 309, 341, 400, 401, 403–405, 407, 409, 411
- Dermoscopic imaging, 184, 185, 191
- Diabetic retinopathy, vii, 145, 146, 271–284
- Discrete wavelet transform (DWT), vi, 130, 131, 187, 189, 191, 194
- DL methods, vi, 89, 135, 147, 155, 168, 309, 310
- Dysarthria, 3, 167–171, 174, 175, 177, 179, 180

E

- Electrocardiogram (ECG), 154, 338–341, 343–352, 355, 357, 358, 361–366, 368–370, 409
- Electroencephalogram (EEG), 73–76, 79, 81, 84, 154, 409
- End-of-line, 197, 198
- Epilepsy, 73, 75, 76, 84

F

- Feature extraction, v, vi, 5, 43–45, 76, 107, 113–114, 125, 130, 187–190, 238–240, 273, 274, 279, 284, 289, 300, 301, 305, 329, 330, 345, 353, 357, 361, 383–386, 400
- Fetal parameters, 259, 262, 269
- Fetus abnormality, vii, 259–269, 304

G

- Gated Recurrent Unit (GRU), vi, 73–84
- Genetic algorithm (GA), 235, 241–243, 248, 249, 261, 264
- Gigapixel images, 34, 36, 37

H

- Haemorrhages, 272, 273, 283
- Hemodialysis, 199
- Hierarchical prediction, vi–vii, 207–228
- Histopathology, v, 31–55, 89, 91, 100, 103, 295, 305

I

- Image segmentation, vii, 58, 89, 155, 187, 228, 309–333, 380–383
- Impaired speech, 167
- Instance segmentation, 88, 90, 103, 322, 328, 330

K

- K-means clustering, 132–134, 184, 187, 188, 192, 194, 247

L

- Linear prediction, v, vi, 4, 7–20, 168, 172–173
- Local binary pattern (LBP), vii, 75, 231–254, 280
- Lung cancer, vii, 373–395

M

- Machine learning (ML), v–vii, 32, 34, 35, 75, 103, 109, 112–116, 125, 131–135, 143, 147, 155, 157, 184, 185, 234, 247, 271–284, 288, 289, 300, 305, 309, 310, 341, 344, 350–352, 357–360, 369, 370, 405
- Manufacturing, vi, 167, 195–204
- Medical image analysis, 287–305
- Medical images, vi, vii, 87, 103, 144, 153–165, 207–228, 234, 237, 247, 273, 288, 289, 296, 300, 341, 373
- Microaneurysms, 272, 273, 276, 283
- ML methods, 35, 155, 276
- Morphological features, vii, 232–235, 238, 240, 241, 249–251, 253, 254, 363

N

- NADAM, 82, 84
- Neural network, 55, 61, 62, 65, 68–71, 131, 135, 158, 185, 190, 196, 246, 260–262, 264–266, 269, 276, 280, 329, 369, 390, 401
- NVIDIA Jetson Nano, 74, 83, 84

P

- Panoptic Segmentation, 322
- Parameter efficient, vi, 73–84
- Principal component analysis (PCA), 75, 233, 235, 242, 248–250, 276, 301, 351

Probabilistic neural network (PNN), vi, 75, 183–194
 Prosthetic, 109, 111–114

R

Real-time, vi, 57, 58, 83, 108, 112, 115, 116, 118, 119, 153, 202, 262, 269, 313, 314, 332, 347, 366
 Retinal images, 146, 271–273, 277, 282, 312
 Reversible color transform, 214

S

Segmentation, vii, 38, 58, 87–92, 95, 97, 99–103, 155, 184, 187, 191, 192, 194, 228, 236, 237, 259, 260, 262, 273–276, 278–280, 282, 300, 303, 309–333, 374, 376, 380–383, 386, 390, 394, 395
 Seismocardiogram (SCG), 338, 344
 Seizures, vi, 73–84
 Semantic gap, 135, 144, 147
 Semantic segmentation, 87–90, 99, 101, 103, 322, 324–328, 330
 Skin cancer, vi, 183–194
 Speaker de-identification, 15, 19, 20

Speech analysis, 167
 Stacked autoencoder (SAE), 235, 241, 242, 245, 247–249, 251
 Support vector machine (SVM), 74, 75, 115, 131–135, 143, 184, 185, 193–196, 232, 233, 235, 242, 243, 247–250, 252, 276, 280, 283, 289, 303, 341, 357, 358, 369, 373, 390–395

T

Troubleshooting, 195, 196, 198–200, 203, 204
 Tumour segmentation, 99, 311, 332

U

Ultrasound imaging, vii, 103, 232–234, 236–239, 241, 251, 254, 259, 302–304

V

Voice privacy, v, 1–25, 168

W

Weakly supervised learning, 35–37, 51