CAPITAL UNIVERSITY OF SCIENCE AND
TECHNOLOGY, ISLAMABAD

Cervical Cancer Detection using Deep Neural Networks and Ensemble of Decision Trees

by

Nabeel Ali

A thesis submitted in partial fulfillment for the
degree of Master of Science

in the

Faculty of Engineering
Department of Electrical Engineering

2018

Copyright © 2018 by Nabeel Ali

All rights reserved. No part of this thesis may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, by any information storage and retrieval system without the prior written permission of the author.

Dedicated to my beloved parents for their love, patience and support.Without their consistent encouragement and unconditional support this work would not have been possible. They have been my true motivation and inspiration for the satisfactory completion of my work.

CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY
ISLAMABAD

CERTIFICATE OF APPROVAL

Cervical Cancer Detection using Deep Neural Networks and Ensemble of Decision Trees

by

Nabeel Ali

CE141004

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Tehseen Zia	CIIT Islamabad
(b)	Internal Examiner	Dr. Masroor Ahmed	CUST Islamabad
(c)	Supervisor	Dr. Imtiaz Ahmed Taj	CUST Islamabad

Dr. Imtiaz Ahmed Taj
Thesis Supervisor
May, 2018

Dr. Noor Muhammad Khan
Head
Dept. of Electrical Engineering
May, 2018

Dr. Imtiaz Ahmed Taj
Dean
Faculty of Engineering
May, 2018

Author's Declaration

I, **Nabeel Ali** hereby state that my MS thesis titled "**Cervical Cancer Detection using Deep Neural Networks and Ensemble of Decision Trees**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Nabeel Ali)

Registration No: CE141004

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "*Cervical Cancer Detection using Deep Neural Networks and Ensemble of Decision Trees*" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been dully acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I, as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

Nabeel Ali

Registration No: CE141004

Acknowledgements

I would like to acknowledge certain people who have encouraged, supported and helped me to complete my work at CUST. First, I would like to express gratitude to my supervisor Dr. Imtiaz Ahmed Taj. His guidance, suggestions, discussions and consistent encouragement have helped me to gain the deep insight in the field of machine learning to solve the research problem. Besides my supervisor, I would also like to thank my research group members for their help and support. Last, but certainly not the least, I am thankful to all my friends, who have helped me throughout my stay at CUST and have made my stay a pleasant memory.

Abstract

Automated cervical cancer screening is an efficient cell imaging based cancer detection application of pattern classification that uses liquid-based cytology (LBC) and Pap smear images. LBC and Pap smear images contain cells which can be categorized into "normal" and "abnormal" categories. Screening system uses segmentation approaches for feature extraction for successful classification. However, successful classification depends on how accurately segmentation is done. In this thesis, an auto-assisted cervical cancer screening without prior segmentation of cervical cells is proposed. Transfer learning approach is used for fine tuning of the new Convolutional Neural Network (CNN), i.e. weights of the convolutional and pooling layers of a pre-trained CNN are transferred to new CNN. Fully connected layers of the new CNN are initialed with values from gaussian distribution. New CNN is then fine-tuned on the cervical cell dataset to learn new weights. Performance of the CNN-based screening system is tested on Herlev dataset for two class problem and seven class problem. Herlev cervical cell dataset consist of seven class data, while two class problem is achieved by combining three normal classes i.e. superficial, intermediate and columnar epithelial as normal class and four abnormal classes i.e. mild dysplasia, moderate dysplasia, severe dysplasia and carcinoma as one abnormal class. A distinguished feature of the proposed approach is that, it achieves its objective without getting into conventional segmentation approach for feature extraction. The immediate impact of this approach can be observed on the classification accuracy of the system. Three different classification approaches are used for comparison analysis on the classification accuracies i.e. softmax, SVM and tree ensemble. Classification accuracies of softmax, SVM and tree ensemble for two class problem is 98.8%, 99.10% and 99.23% respectively. For seven class problem, classification accuracies of softmax, SVM and tree ensemble are 97.21%, 98.12% and 98.85% respectively. These results shows that the proposed system yield better performance in all metrics i.e. accuracy sensitivity and specificity than its previous counterparts as the previous best classification accuracies are 98.3% for two class problem and 96.6 for seven class problem.

Contents

List of Figures

List of Tables

Abbreviations

Pap	Papanicalau
FOV	Field of view
LBC	Liquid-based cytology
CNN	Convolutional neural networks
SVM	Support Vector Machine
conv	Convolutional Layer
pool	Pooling Layer
fc	Fully Connected Layer
ANN	Artificial Neural Network
ReLU	Rectified Linear Unit
LSSVM	Least square Support Vector Machine

Chapter 1

Introduction

1.1 Background

Cervical cancer is one of the fatal disease that causes death among women. According to a detailed survey, it was observed that 85% of the cervical cancer patients are found to be infected at initial stage in third world countries. A straight forward reason for this alarmingly high ratio is the unavailability of the required fundamental medical resources in the regions. [1]. Countries with modern medical resources aims at preventing cervical cancer by providing cervical cancer screening systems to detect precancerous cells than can lead to invasive cancer. The Papanicolaou (Pap) smear testing is most commonly used for cervical cancer screening worldwide.

In most developed countries, cervical cancer screening systems has been mostly used for the diagnosis of cancerous and precancerous lesions . The rate of occurring cervical cancer has dropped by 80% since the screening systems are introduced in some Nordic countries [1]. It is dropped by 65% during the last four decades and the occurring of cervical cancer and mortality figures due to cervical cancer are stable over the last decade in Sweden [2]. In third world country Pakistan, cervical cancer is listed third major cause of morality amongst women across all age groups. The mortality figures are high because of ignorance in terms of screening

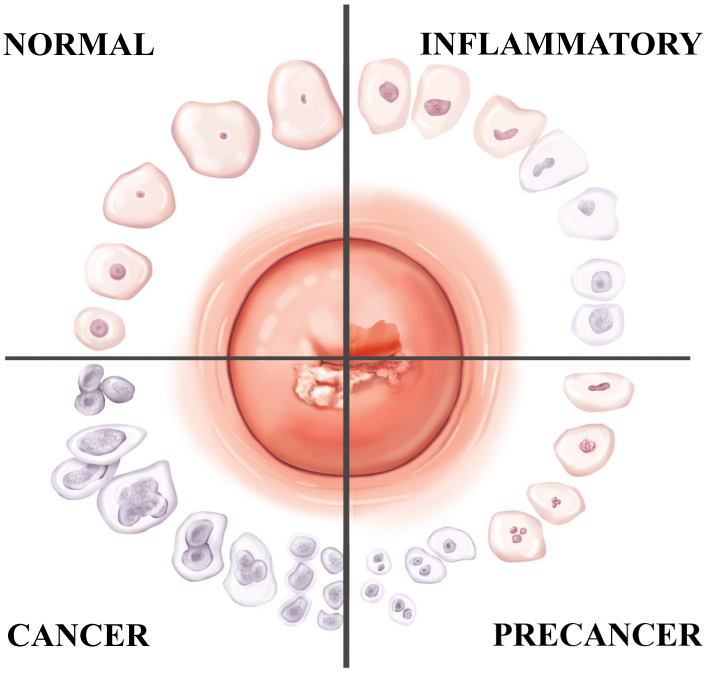

FIGURE 1.1: Cervical cell description

and prevention in Pakistan . According to a detailed survey it was observed that 70% of the cervical cancer are diagnosed at the advanced stage of malignancy which is the cause high mortality figures in the third world country like Pakistan [3].

Cervical cells can be categorized into four types i.e. normal, precancerous and cancerous cells as shown in Figure 1.1. Moderately and mild abnormal cells can go away without any treatment on their own. Precancerous can be removed which can prevent further development of cervical cancer. If precancerous are left untreated, the may turn in to cancerous cells within the time span of 10 to 15 years [3]. Conventional pap smear test the most preferred to diagnose cervical cancer in most of the countries.

1.2 Manual Screening

George N. Papanicolaou devised the Papanicolaou(Pap) smear test in 1928. The main advantage of Pap smear test is the cost effectiveness. It is the most widely

used detection program used for the diagnosis of cervical cancer. George N. Papanicolaou found that abnormal cells can be distinguished within cells collected from the vaginal smear. Dr. Papanicolaou and Dr. Herbert Traut, a gynecologist and pathologist, collaborated and provided scientific proof of the potential of cervix for the identification of cervical changes [4].

1.2.1 Procedure

Conventional Pap smear procedure consist of the following simple steps:[5]

1. A speculum is inserted into vagina to widen the walls of the vagina so that vaginal smear can be viewed.
2. A spatula is used to sample cells from inside and around the vaginal smear.
3. Samples are passed on to a glass slide.
4. Preservatives are applied on to glass slide to preserve the samples.
5. Samples are stained to improve the contrast and enhance the structural patterns to be analyzed by the microscope.

1.2.2 Limitations

Several weeks are required to prepare for the final results of the Pap smear test. Samples are taken from the vaginal smear and send the samples to the pathology lab for visual analysis under the microscope[6]. The process is time consuming and laborious. It requires microscopic examination of hundreds of thousands of cells for the diagnosis of precancerous and cancerous cells. 1 in every 10 to 15 positive cases is missed in conventional screening[7].

There are factors affecting the Pap smear test. Sampling error affects the Pap smear test results if no diagnostic cells make it onto the slide. This is due to human error when heath experts fails to adequately sample the vaginal smear. It also occurs when the spatula fails to transfer precancerous and cancerous cells onto the slide. Interpretation error also affect the Pap Smear test due to negligence of human expert. These factors motivated the research in the field of automated cervical screening systems[6, 7].

1.2.3 Automated Screening

Automated Pap smear screening systems can examine Pap smear slides in a small amount of time with consistent classification results also catering the vulnerabilities of conventional Pap smear test. The following are the ways in which the conventional Pap smear test limitations can be countered[8]:

1. Improve sensitivity and specificity.
2. Reduced workload on pathologists.
3. Reducing the cost and time span.
4. Lowering the percentage of occurring of cervical cancer and mortality figures.

FocalPoint GS Imaging System by BD (Becton Dickinson) and ThinPrep Imaging system by HOLOGIC, Inc are the FDA-approved cervical screening systems available for the diagnosis of cervical cancer [9]. The FocalPoint GS imaging system work both on liquid-based cytology (LBC) slides and Pap smear slides. The system ranks slides based on the likelihood of malignancy and categorize them into four classes: review, no further review, process review and quality control review. No further review can be forwarded without human review. ThinPrep system work only on Liquid biopsy cytology slides. The system takes 22 field of view (FOV) slides from a total of 120 FOV which are of the diagnostic interest. Then these 22 FOV are thoroughly reviewed by the pathologists.

These systems are semi-automated, the pathologists reports the final results on the basis of results generated by these systems. In the past decade, several research has been done for the development of automated system based on image processing and analysis methods[10] . Auromated screening systems works in the following manner i.e.:

1. Segmentation of cervical cell components
2. Extraction and selection of features.
3. Cell classification.

Exact segmentation is a major challenge in cell segmentation because there are other particles like mucus, blood cells are also present in the slide image. The presence of overlapping cells are also a major problem in accurate segmentation which causes the classification accuracy of the system to decrease. To overcome the problem of low classification accuracy due to high segmentation error, deep learning convolutional neural networks (CNN) are used to learn information and discriminative features of raw data automatically [11]. In 2012, CNN have achieved interesting results regarding ImageNet Large Scale Visual recognition challenge[12] . Deep CNN also improved the performance of several medical imaging systems[13] , such as: diseases of lungs, cervical intraepithelial neoplasia, blood vessel extraction, brain tissue extraction and classification and lymph nodes classification, etc. CNN provides superior performance when used to classify cell image data [14]. CNN performs better on large datasets. For cervical cells, there is limited amount of labeled data available. The HERLEV dataset contains only 917 cells [14]. To cater this problem, transfer learning is used which is an effective technique in case of small image datasets. The layers of CNN trained on large scale datasets can be transferred to small scale datasets which reduces over-fitting and achieves high training accuracy for small scale trained networks [12].

1.3 Objectives

The aim of this research is to classify between normal and abnormal (cancerous) cells image shown in Figure 1.1 and extraction of features without going into segmentation of cells.

The objectives are:

1. Increase the number of samples in the small scale Herlev dataset by using augmentation techniques such as rotation and translation of the images.
2. CNN previously trained on ImageNet dataset is used to fine tune a new CNN on cervical cell dataset to learn and extract deep hierarchical features automatically from the raw cervical cell images.
3. Several classifiers are trained on deep featured extracted from the deepest layers of the CNN to evaluate the performance of the system.

1.4 Scope

As described in the objective section, pre-trained CNN model which is trained on ImageNet dataset containing 1000 classes and 1.2 million images is fine-tuned on Herlev Cervical cells dataset to discriminate cell images containing normal and abnormal cells based on deep feature set. Transfer learning is an effective technique to be use with small scale datasets. In the training phase, a pre-trained CNN on ImageNet dataset is used to learn the features of Herlev dataset. In the testing phase, classifier is trained on deep features learned by CNN to get the classification score.

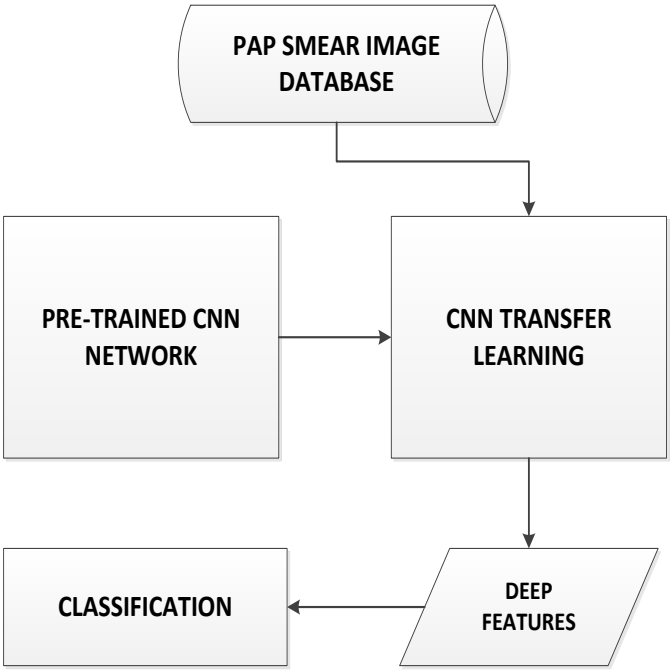

FIGURE 1.2: Scope of the thesis.

1.5 Organization of Thesis

Chapter 1 has provided an overview of manual screening test, including its pros and cons. The basis for the automation screening of Pap smear slides is also discussed.

The phenomena of using CNN to extract deep features based on medical imaging problems is described, addressing the problem of prior segmentation of the cervical cell images. Aims, objectives and scope of this thesis is defined.

Rest of the thesis is arranged as follow:

Chapter 2 provides the literature review. A review of cell segmentation methods of Pap smear images published in the literature. Also a review of feature selection methods using CNN without prior segmentation is presented in the literature review.

Chapter 3 deals with the methodology used for the solution of problem. This chapter explains the training of CNN, feature extraction and selection and classification of cervical Pap smear slide images.

Chapter 4 elaborates experimental methods used in this thesis.

Chapter 5 represents the performance evaluation of the proposed system.

Chapters 6 explains limitations, conclusions and future work regarding the proposed work.

Chapter 2

Literature Review

A review of methods published in the literature for the classification of Pap smear images using segmentation approaches and without using segmentation approaches is presented in this chapter.

2.1 Existing methods for cell classification with prior segmentation

This section provides a review of the existing cell segmentation approaches used for morphological feature extraction of cell components published in the literature. A review of classification algorithms is also presented which are trained on features extracted by using segmentation approaches.

A conventional pap smear slide image may contain cells which are normal, inflammatory, pre-cancerous and abnormal. These cells can be single and overlapping [6] as shown in figure 2.1. Accurate segmentation of cervical cell is vital to the success of automated Pap smear screening system. However, presence of large shape, appearance dissimilarities and cell clusters between abnormal and normal cell nucleus are the major problems in accurately segmenting the cytoplasm and nucleus from the background [15].

Classifying cervical cells into categories with only nucleus features are taken into account yields less performance, segmentation of whole cell components i.e. nucleus, cytoplasm and background components yields superior performance [16] .

Segmentation of overlapping cervical cells is a great challenge. Deformable templates are used for the segmentation of nucleus from cell images containing overlapping cell components. This is done by approximation and refining of locations obtained from images and obtaining local deformations by using deformation models [17]. This work uses single image containing overlapping cells.

Pixel classification is a thresholding technique which is used to separate nucleus background and cytoplasm by assigning gray levels to each pixel [18]. The method works fine for the image containing single cells. In case of overlapping cells, the segmentation error increases. Another method uses cell segmentation from the background of pap image and block wise pixel classification (gray level thresholding) to extract features. Then uses wavelet to carry out comparison between different image features for selection to calculate the performance of multi-spectral textural features from the cell images [19]. Another method uses a feature screening algorithm based on support vector machine's decision boundary resulting from Pap smear images. the cell image is segmented and features are extracted using wavelet transforms. Then the screening algorithm is applied to the initial feature set in order to remove features which are irrelevant to the cancer detection [15].

Seed based region growing technique is used for detecting edges of cell nucleus and cytoplasm[20]. The approach is promising for single cell image containing single cell images as it provides efficient results in segmenting the cell components but in case of overlapping cell images the efficiency of the system degrades. Morphological and textural features combination is used for the extraction of nucleus part in conventional pap smear images, watershed transform is used as an segmentation approach. The performance of the system is analyzed using unsupervised classification (k-mean) and supervised classification (support vector machine) techniques. As the approach work only on the nucleus of the cell image the other particles like

inflammatory cells were not taken into account. However the existence of inflammatory cells is common among the Pap smear slide images. [21]. Therefore, numerous methods are proposed in the literature to cater the problem of identification of inflammatory cell in conventional Pap smear images. The boundary of the nucleus is first detected and segmented from pap smear slide image based on morphological details of the image. Shape, texture and intensity of the nucleus is then extracted. Feature selection scheme is used to select features with back-propagation classifier to mitigate false positive findings of the system. Fuzzy C-Mean clustering scheme is then used to differentiate between nucleus and inflammatory cells present in the pap smear slide image. Sensitivity and specificity values of the system are 95% and 98% respectively [22]. Segmentation accuracy is low in case of overlapping cells and inflammatory cells. To mitigate the effect of inflammatory cells on the performance accuracy, gray level thresholding is used to segment the nucleus and inflammatory cells, a definition of distance rule is used to reduce the number of inflammatory cells from the segmented slide. [23].

Another method that uses rule classification of texture grey level run length matrix GLRLM to extract and classify between nucleus and inflammatory cells present in pap smear images by using rule texture of decision tree (J48). Further color coding scheme is used to differentiate cell components which are nucleus, cytoplasm and inflammatory cells. The system evaluates a specificity of 75.04% and a sensitivity of 49.82% which means still some of the inflammatory cells are classified as nucleus and some of the nuclei are classified as inflammatory cells [24]. Along with nucleus, cytoplasm and background are also important for the better performance of the system. several works in the literature has been done challenge this problem.

Atypical morphological details of cell for the diagnosis of abnormal cells is also of importance. Unsupervised genetic algorithm which relies on genetic operators and uses fisher classification to separate nucleus, cytoplasm and background portions of the cell. Morphological features used for the recognition system are size of the nucleus and cytoplasm, homogeneity texture i.e. intensity mean and variance and deformity of the cell components. The system does not take overlapping cells into account and works only on single cell images [25].

Segmentation of each pixel of the cell image is done in a specific wave band with respect to the intensity, spectrum relationship of each pixel. Features of each pixel of Pap smear single cell image are extracted using cosine correlation analysis. Features used in this work are size of the cell components i.e. nucleus and cytoplasm, wavelet characteristics of cell components and color intensity. The work intended to improve the specificity and sensitivity of the system [26].

Classification of pap cell image depending only on the segmentation of nucleus part neglecting cytoplasm and other cell particles. Feature extraction is based solely characteristics of the nucleus[27].

Gabor filter hybrid with k-mean clustering is used for the classification of cytological cervical images. Gabor filters are applied to each pixel of a cervical image containing normal, inflammatory, pre-cancerous and abnormal cells to detect the textural variation of the cervical cells. These extracted textural variations are localized textural features based on frequency and orientation. features vectors are generated for each pixel and compared with feature variations extracted by gabor filter bank. Segmentation process uses a color coding scheme i.e. red and green for abnormal and normal cells respectively and blue for the background of the pap image. segmentation images are further processed by k-mean clustering algorithm to differentiate between normal and abnormal cells. the classification of the segmented images processed by k-mean clustering is done according to shape of the extracted nuclei. The sensitivity of normal and abnormal cells is 87% and 89% respectively and specificity of the system is 85% [28].

Fuzzy C-mean clustering technique is used as a segmentation technique to automatically segment single cell images into nucleus, cytoplasm and background, After segmentation, features are extracted based on nucleus and cytoplasm segmented from pap images. Six features based on nuclei are used in this works which are total number of pixels in the nucleus region, compactness of the nucleus, length of major axis enclosing nucleus, length of minor axis encloses the nucleus, aspect

ratio of nucleus, homogeneity and nucleus to cytoplasm ratio. the repones of different classifier is analyzed for the system and achieved classification accuracy of 93.78% with artificial neural network as a classifier [29].

All of the methods discussed in this section uses single cell images in which some contains multiple cell components and overlapping cells. All the methods are based on textural and morphological features extraction from cell images. Accurate segmentation of nucleus is a difficult task because of the presence of other cell particles. It can be infer that the performance of the system is high when there are cell image containing single nucleus and a single cytoplasm. In case of overlapping cells i.e. when there are multiple cell components in a single cell image and inflammatory cells, the performance of the system is low as the segmentation error is high in this case. To overcome the problem of high segmentation error in pap smear cytological imagery, deep learning approach is a good candidate as it does not go into segmentation for the extraction of features. this approach undergoes extraction of deep features extracted directly from the raw data. Performance of the system is evaluated by using different classifiers trained on extracted deep features from raw data.

TABLE 2.1: Feature based approaches comparison with prior segmentation

References	**Segmentation Method**	**Dataset**	**Performance Measure**
Bamford et al. [30]	Active Contours	20130 (Single Cell)	99.6% segmentation accuracy
Lezoray et al. [31]	Watershed Transform	209 cells from 10 Pap slide images	2.8% (RGB) and 0.47% (HSI) Seg. Error.
Yang Mao et al. [32]	Edge Detector	124 (Single cell)	Segmentation error of 0.1523 for nucleus and 0.0775 for cytoplasm.
Lin et al. [33]	Edge Detector	10 (Single cell)	Segmentation error of 0.1323 for nucleus.
Bak et al. [18]	Iterative Pixel level Thresholding	1	No Quantitative measures
Garrido et al. [17]	Deformable Template Fitting	3 Pap slide images	No Quantitative measures
Plissitiet et al. [34]	Edge Detector	5617 cells from 38 Pap slide images	Sensitivity: 90.57%(FCM), 69.86%(SVM). Specificity: 75.28%(FCM), 92.02%(SVM)
Rahmadwati et al. [28]	Gabor filter hybrid along with k-Mean clustering	475 labelled Pap cell images	Sensitivity (Normal): 87%. Sensitivity (abnormal): 89% Specificity: 85%
Muhimmah et al. [22]	Fuzzy C-Mean	321 cell images from 20 Pap images	Sensitivity: 95%. Specificity: 98%
Chankong et al. [29]	Fuzzy C-mean	917 Single Cell images containing overlapping cells	Accuracy (ANN): 93.78%

2.2 Existing methods for cell classification without prior segmentation

Most of the techniques proposed in the literature performs segmentation in the local image patches i.e. images with single nucleus, cytoplasm and the background. They are unable to provide higher segmentation accuracy in case of pap images containing overlapping and inflammatory cells. The boundary of cytoplasm cell is usually not smooth, it is the most complicated task to separate the cytoplasm. irregular shape pattern of cytoplasm is also a limitation in achieving segmentation accuracy. The extent of occurring overlapping cytoplasm including nuclei is very high in Pap smear images [35]. The classification accuracy of single cell segmentation of Herlev dataset [36] ranges from 85% [37] to 92% [38]. In case of overlapping cell taken into account, the classification accuracies vary from 87% to 89% [39]. Most of the Pap smear cell classification works assume that cytoplasm and nucleus are accurately segmented, however in real world there are other particles along with cytoplasm and nucleus in the background which are blood, air drying, bacteria and mucus. These particles also affect the segmentation results of the system [27]. High classification accuracies are achieved using the features extracted from segmentation of cytoplasm and nucleus on Herlev Dataset while neglecting other particles present the cell background of a pap smear image. These high accuracies could decrease if the segmentation error due to cell particles other than nucleus and cytoplasm are taken into account [40][41].

Correct segmentation of segmentation and cytoplasm is most important task for achieving high classification accuracy of the system. The major problem of methods using the segmentation of the nucleus, cytoplasm and the background is low segmentation accuracy which is due to non convex nature of the cell particles, overlapping of cytoplasm. It can be seen that cytoplasm and nucleus of one cells are covering the the cytoplasm of other cell, the bacteria particles present on the surface of the cytoplasm and the irregular cell boundary of cytoplasm in Figure 2.1[29].

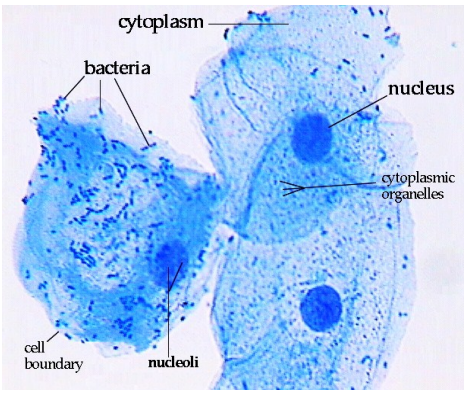

FIGURE 2.1: Pap Slide Image containing overlapping cells [41]

Current Pap smear cervical cell classification systems are limited by feature design and extraction. Current feature types regarding Pap smear cervical cell images are either handcrafted or engineered using The Bethesda System rules. The handcrafted features define the morphological and chromatin patterns characteristics of the Pap smear cell images , either the morphological features or textural features are used or sometime hybrid features having both types of features are used for the classification task [42].

Several efforts has been made to partially remove the dependency on the segmentation of nucleus, cytoplasm and background for feature extraction. A good performance is achieved using a non-linear dimensionality algorithms (Kernel-PCA (K-PCA), Isomap, Locally Linear Embedding (LLE) and Laplacian Eigenmaps) for feature extraction along with supervised learning for classification using support vector machine [43]. Another method proposes block image processing to crop arbitrary image blocks prior to feature extraction and then the cropped blocks are then classified using SVM classifier [44]. The arbitrary cropping of Pap smear slide can leads to cropping of cell images into different blocks.

Deep learning algorithms have achieved good reputation for its ability to learn mid and high level image representations. CNNs are first introduced to medical imaging problems in 1996 by Sahiner et al, in this work regions of interest are extracted either masses or normal tissues from the mammogram [45]. in this work CNN with two hidden layers, input and a output layer was used with bach-propagation

Representation learning achieve good popularity due to deep learning methods. Performance of machine learning methods depends on the representation (or features) of data. Representation learning is referred to set of methods design to discover and learn discriminative information from the data. Representation learning has strong impact on speech processing, natural language processing and object recognition with breakthrough results [11].

Deep CNN have achieved very impressing performance in 2012 ImageNet Large Scale Visual Recognition Challenge. Main objective of the challenge is to classify the 1.2 million high resolution images in ImageNet LSVRC-2012 dataset to 1000 classes [12].

CNN have proved to be providing significantly better performance in various medical imaging applications [13], such as vessel segmentation [46] in fundus images, lymph nodes and lung deceases classification in CT images [43], detection of cervical intraepithelial neoplasia at patient level on the basis of cytological cervical images [47] and segmentation of brain tissues in MRI scan images [48]. CNN have provided significant performances over the classification of cell type images such as human epithelial images and pleural cancer cell images. Transformation of segmentation problem into labeling problem by predicting pixels probability using deep convolutional neural network and contrast level set to improve the accuracy of splitting boundaries of cervical cells [35].

A method uses deformable templates to label cells in pap smear slide images and multi-scale deep convolutional neural networks are then used to extract feature of the labeled cells. the classification accuracy achieved is 94% for the nucleus and 89% for cytoplasm detection [49]. Features are extracted using CNN on herlev dataset, feature selection is done by principle component analysis (PCA) and the classification is done using LSSVM (Least Square Support Vector Machine) and softmax. The maximum classification accuracy achieved with feature selection for herlev dataset is 84.41% and maximum accuracy without feature selection is 88.88% [27].

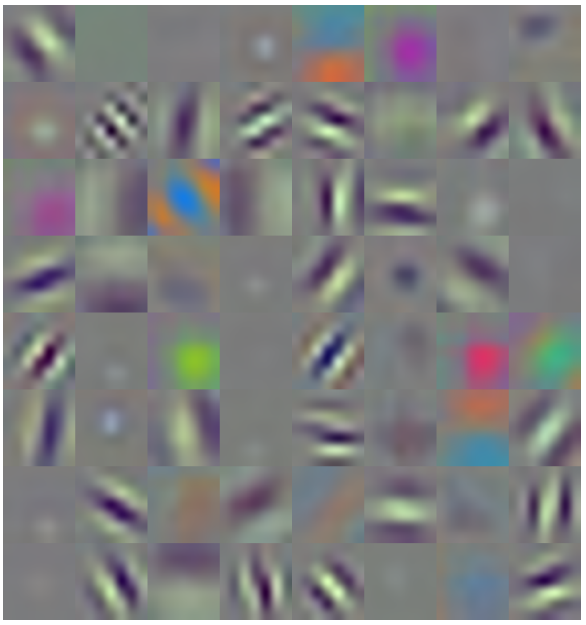

FIGURE 2.2: First Layer weights of a trained network [44]

Transfer learning approach is used for the feature extraction and classification of pap smear images. first layer features of a previously trained network are transferred to a new network, new network is then trained on cervical images to learn features on the last layers of the networks. Features are extracted and softmax regression is used for the classification. classification accuracy of 98.8% is achieved [50]. Large datasets are of very importance to the performance of CNN. Small datasets leads to over-fitting of the network i.e. CNN does not achieve its maximum training accuracy. Transfer learning is a promising approach to be used when the size of the dataset is small. Features learned in first layer of CNN resemble gabor filters or color blobs as shown in figure 2.2. these first layers features are not specific to different datasets and are very generic features. therefore these features are applicable to different datasets. the transition of features from generic to specific occurs by the last layer of the CNN [44]. The weights of first layers i.e. convolutional, pooling and linear unit of the CNN previously trained on ImageNet dataset [51] are transferrable to other medical imaging problem where the size of dataset is small. Several approaches have been used for the classification of cervical cancer based on Herlev cervical cell dataset. Performance comparison of some of the work done in the literature on deep learning for the classification

of cervical cancer on Herlev University dataset is presented in Table2.1 with performance metrics sensitivity (Sens), specificity (Spec), accuracy (Acc), harmonic mean (H-mean), harmonic average (F-measure), area under the curve (AUC) and cross validation (CV).

TABLE 2.2: Hybrid and deep features based approaches without segmentation

References	**Method**	**Dataset**	**Performance Measure**
Jantzen et al. [38]	Pap Smear Benchmarks	Herlev Dataset	Accuracy : 93.6% Sensitivity : 98.8% Specificity : 79.3%
Jantzen et al. [37]	Particle swarm Optimization and nearest neighbor (1nn) Classifier	Herlev Dataset	Accuracy : 96.7% Sensitivity : 98.4% Specificity : 92.2%
Marinakis et al. [52]	Genetic Algorithm based Feature selection and 1-nn Classifier	Herlev Dataset	Accuracy : 96.8% Sensitivity : 98.5% Specificity : 92.1%
Plissiti et al. [39]	Principle Component Analysis for Feature selection and SVM Classifier	Herlev Dataset	Harmonic Mean : 96.9%
Bora et al. [27]	Ensemble Classifier	Herlev Dataset	Accuracy : 96.5% Sensitivity : 99% Specificity : 89.7%
Guo et al. [40]	Local Binary Pattern for feature extraction	Herlev Dataset	Area under the curve : 96.4%
Zhang et al. [50]	Convolutional Neural Networks for deep feature extraction and Classification	Herlev Dataset	Accuracy : 98.3% Sensitivity : 98.2% Specificity : 98.3%

Chapter 3

Feature Extraction and Training Methodology

In this work, Deep Convolutional Neural Networks (CNN) are applied for the purpose of feature extraction of cervical cells in cytological images. The performance of different classifier is evaluated for comparison. Augmented RGB image data i.e. multiple instances of single image are generated by using rotation and translation operations from Herlev dataset centered on nucleus are used to increase the size of cervical image dataset. A pre-trained CNN named Alex-Net trained on a natural dataset named ImageNet dataset is used to fine-tune a new CNN on cervical dataset to learn discriminative information between images containing malignant and benign cells depending on deep hierarchical features. Fine-tuned CNN is then used to classify cell patches roughly centered on nucleus. Classification results are grouped to produce final cell category. Core work of this thesis is to extract deep hierarchical features of Pap smear images without going into prior segmentation and morphological or textural feature extraction and train different classifiers on extracted deep features. The comparison is taken out between different classifiers on the basis of classification accuracy, sensitivity and specificity. In this way the classification accuracy does not get affected by accuracy loss which is caused by inaccurately segmented cell component. This approach consist of two phases, a train and a test phase as shown in Figure 3.1. In the train phase, a CNN previously

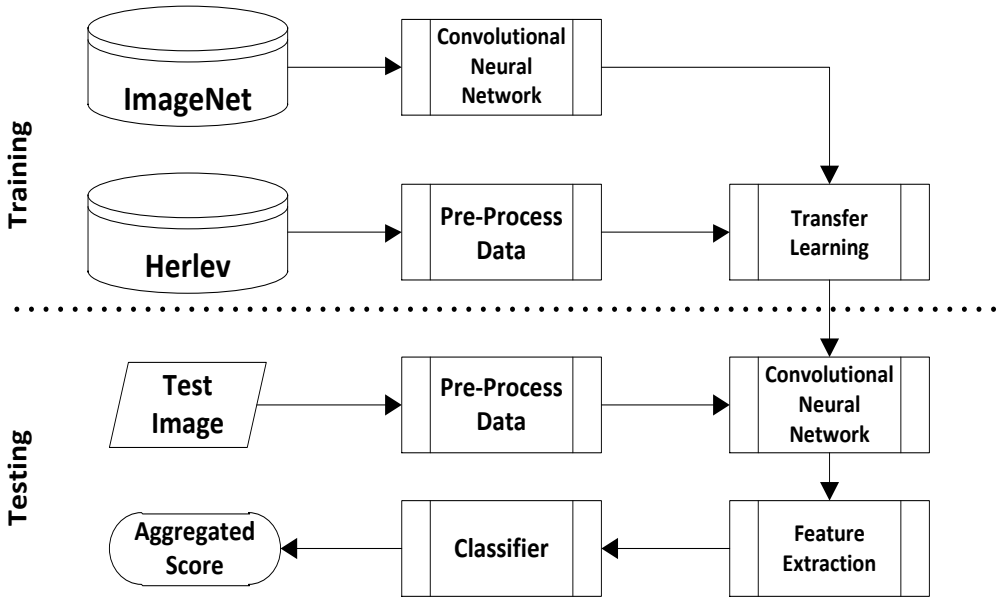

FIGURE 3.1: Architecture of proposed scheme

trained on natural dataset is applied on Pap smear dataset after data preprocessing. Transfer learning is then applied. The parameters of pre-trained CNN are then used to start a new CNN. New CNN is fine-tuned on the augmented Pap smear dataset. In the testing phase, same data preprocessing is applied to the test images and fine-tuned CNN is applied on the test set and malignancy score is obtained by aggregation the output of the fine-tuned CNN.

3.1 Data Preprocessing

Data processing phase includes image patch extraction from the cervical cell dataset and augmentation of data for the CNN.

3.1.1 Image patch extraction

The method does not directly operates on single cell images present in Herlev dataset. To get individual cell, pre-segmentation is required. At least segmentation of cytoplasm is required, which is a challenging and unresolved problem. Malignancies of several categories of cervical cytology are associated with types of malignancies of the nucleus. Thus substantial discriminative information is available in features of nucleus. Image patches of size M x M centered on the nucleus center should be generated as this will embed not only the size and scale information of the nucleus but also the textural information i.e. the information of the cytoplasm in the extracted image patches. Scale and size of the nucleus is a very important discriminative feature between malignant and benign cells. In this thesis, the image patches are extracted by directly translating the centers of nucleus present in Herlev dataset. The method of extraction of image patches is described below.

3.1.2 Image Augmentation

Image data augmentation techniques are used to virtually increase the size of training dataset and reduces overfitting [12]. Data augmentation can be done by linear transformation of the data such as mirroring, scaling, translations, rotation, color shifting unless the information of the object in the image is intact. Data augmentation can also be done during training. For that parallel processing is used, one thread for augmentation and passed on the other thread which is training the CNN. Since cervical cells are invariant of rotations and can be rotated from 0-360 with a step of angle θ. N_r rotations are performed over the dataset, thus increases the number of training samples.

Each cell image is rotated with a step of θ and N_r patches are generated. These patches are centered on the center of the rotated nucleus. The training image patches are shown in Figure 3.3. The high frequency component may degrade due to the rotation of the cell image but malignancy/benign information of the

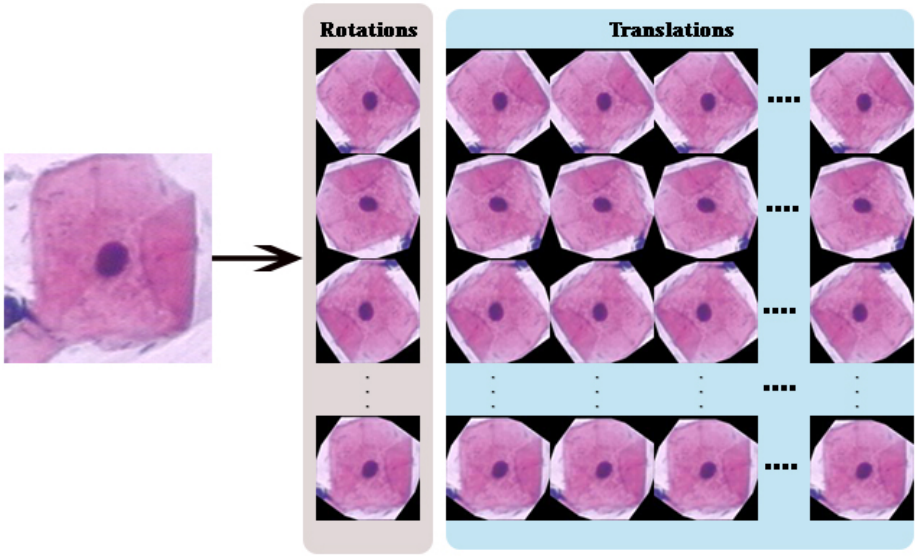

FIGURE 3.2: Data Augmentation using rotations and translations

cell image is still intact. Data augmentation step is of great importance to the success of CNN [53]. It has been proved to be crucial for accuracy improvement of CNN-based cell image classification applications [54] as the training samples in the Herlev dataset are limited. Image patches are zero padded for the regions that lie outside the boundary of the image.

As in practice, the nucleus detected in the cell image dataset are inaccurate. Random translation of each nucleus center N_{th} horizontally and N_{tv} vertically are performed upto d pixels is used to obtain N_{th} and N_{tv} points as the coarsely centered nucleus. Therefore, translated image patches N_{th} and N_{tv} of size mxm are generated to be used as training samples as shown in the Figure 3.3. The detection of inaccurate nucleus center is simulated and number of training samples are increased to be fed to CNN using these translated image patches. Other augmentation techniques like scaling, color shifting and shearing is not used as the color intensity and size of the nucleus is important to classify between malignant and benign cells.

The malignant cells in the Herlev dataset are 3 times more than the benign cells. Classifier tend to be biased towards the majority class i.e. malignant cells. Achieving correct classification score is ideal from medical perspective [38], from medical point of view, incorrect classification i.e. classify benign cells as malignant is not

important. To address the biasness in the data, common solution is to evenly distribute the ratio of positive and negative samples of data [55]. This will improves the convergence rate of training of the CNN. Also the classification accuracy of CNN improves [12]. To balance the training set, higher proportion of benign training samples are generated as compared to malignant training samples.

3.2 Deep Convolutional Neural Networks

A convolutional neural network is a deep learning model consisting of sub-sampling layers, which are convolution, non-linear and pooling layers consecutively followed by more convolutional layers, drop-out layers and the final layers are fully connected layers. The input to the convolutional layer of the CNN is a raw pixel image. In this thesis, the image at the convolutional layer as input is a zero-centered normalized image from the dataset. The images is obtained by subtracting the mean activity over the dataset [12]. Final layer or the classification layer consists of neurons each corresponding to one class. Back-propagation algorithm is used to optimize the weights of the CNN [55]. Proposed scheme is shown in Figure 3.1 showing two CNNs. First network CNN1 is trained on natural dataset i.e. ImageNet, while the second network CNN2 is initialzed with weights of the layers of CNN1. CNN2 is then fine-tuned on the Pap smear dataset containing cervical cell images . Fully connected layer contains abstract information of the image features, these features are extracted from fully connected layer to be fed to classifier. Softmax, SVM and decision tree classifier is used to produce the malignancy score.

3.2.1 Convolutional layer

Convolutional layer of the CNN takes local rectangular image patches extracted from the cervical cell image as an input at first layer $conv1$. The input at $conv1$ is with offset by stride. The spatial information can be preserved by using padding. The subsequent convolutional layers takes feature maps as input. 2-Dimensional

convolution with a filter is performed at the *conv* layers. A non-linearity function is then exerted to the sum$x$ of the convolutions at the convolution layer. A rectified linear unit (ReLU) is used [12] to speed up the training of the CNN presented in Equation 3.1.

$$f(x) = max(0, x) \tag{3.1}$$

Convolution layer shares same filter across the feature map to allow the detection of same patterns in different positions of the feature maps. For different feature maps distinct filters are used.

3.2.2 Pooling layer

The operation of the pooling layers *pool* is to down-sample the feature maps by summarizing the responses of the features in each incoming local rectangular patch. This is done by calculating the maximum activations. The phenomena is named as max-pooling. This makes features become invariant to slight translations in the data. Classification probability for each class is computed by two neurons at the last fully connected layer using softmax regression.

3.2.3 Fully connected layer

Feature maps generated by convolutional and pooling layers are of smaller dimensions as compared to the input images. Feature maps generated at *conv* and *pool* are passed on to several fully connected layers. Feature vector is generated by initial fully connected layers and become more abstract in the deeper fully connected layers.

3.2.4 Training

The weights of the Convolutional neural network (CNN) are initialized from the Gaussian distribution. Stochastic gradient descent is used to compute the gradients of loss to iteratively update the weights at the layers of the CNN during the training process. Batch size of training samples from the dataset is set to 256 training samples per epoch. Initial learning rate for the fully connected last layers of the CNN is set to decrease after specific number of epochs are passed. Momentum is set to 0.9 and L2-regularization or weight decay is set to 0.0005 to reduce over-fitting and speed up learning process [12]. The network is trained for certain epochs and terminated. Several models of CNN are trained, the CNN model having minimum validation loss is used for classification application.

3.3 Training using Transfer learning

Transfer learning is the most promising technique in deep learning when training set size is small and not enough for the network. Smaller datasets lead CNN to over-fitting. A pre-trained CNN can be used as a starting point to learn a new task. Training a network and initializing weights from scratch is very time consuming as compared to fine-tune the CNN by using CNN models that are pre-trained on large-scale image data sets. The weights at the convolutional and pooling layers of pre-traind Alexnet CNN **??** pre-trained on ImageNet dataset (ILSVRC2012) are used as initial layers for the new CNN, which is to be trained on Pap smear dataset. Random weights are initialized to fully connected layers fc. Convolutional and pooling(*convandpool*) layers are transferred to the same locations from pre-trained CNN to new CNN. To fine tune the new CNN layers, learning rate for convolutional and pooling layer is reduced 10 times as compared to the rate used for the training of pre-trained CNN. For fully connected layers of the CNN, same learning rate is initialized as used in pre-trained CNN.

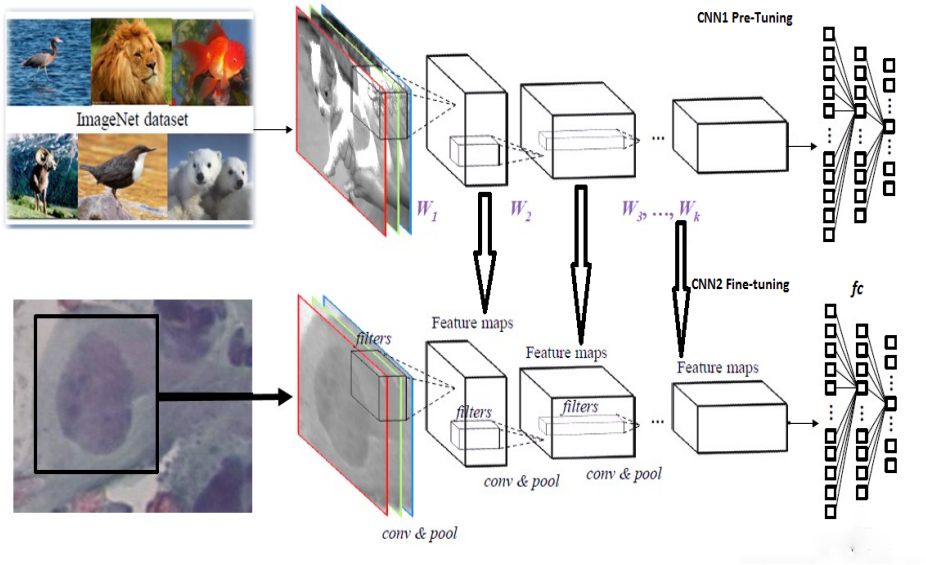

FIGURE 3.3: Layer Transfer Pre-trained CNN to new CNN

3.4 Feature extraction

Convolutional neural networks are feed forward networks as input received on in-between layers are features generated by the previous layers. There are three fully connected layers in the fine tuned network i.e. fc6, fc7 and fc. In this work, the deepest layer of the network i.e. fully connected layer fc7 is selected as the feature vector. fc7 is the last layer before the output layer. The reason for the selection of fc7 is that it contains more specific and abstract details of the pap smear images. Feature vector extracted from fc7 is number of training samples X_t x 4096 dimensional vector. Different classification approaches are used for the comparative analysis of classification accuracy, sensitivity and specificity of the system. The details of the classification approaches are provided in the next chapter.

Chapter 4

Classification Methodology

As described in the previous chapter, Deep hierarchical features extracted from the fully connected layer are used to train classifier on the Herlev Pap smear data.

4.1 Classification

Three classifiers are used for cervical cell classification namely SVM, softmax and Decision tree classifier. The response of each classifier is analyzed individually. Classifiers are trained on the deep features extracted from the fully connected layer of fine tuned convolutional neural network. The performance comparison of the classifiers is based on performance measures i.e. Area under the curve, Accuracy, Specificity and Sensitivity.

4.1.1 Softmax regression

Softmax regression is a generalized form of logistic regression that is used for multi-class classification. Convolutional neural network uses a loss function i.e. cross entropy known as softmax activation function as shown in figure 4.1 at the softmax layer. Softmax provides a non-linear variant of logistic regression [56]. Softmax layer of CNN is shown in figure 4.2

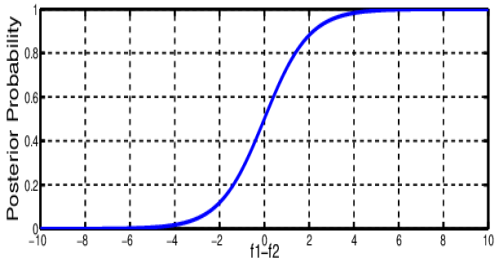

FIGURE 4.1: Softmax activation function

$$P\left(y = j|z^{(i)}\right) = \phi_{softmax}\left(z^{(i)}\right) = \frac{e^{z^{(i)}}}{\sum_{j=0}^{k} e^{z_k^{(i)}}}, \tag{4.1}$$

where we define the network input z as

$$z = w_0 x_0 + w_1 x_1 + ... + w_n x_n = \sum_{i=0}^{n} w_i x_i = w^T x. \tag{4.2}$$

w is the weight vector, x is the feature vector of 1 training sample, and $w0$ is the bias unit. This softmax function computes the probability score that training sample x_i belong to class j given the network z.

The probability score is generated at the softmax layer of CNN which is next to fully connected layer. Cross entropy function is used for the classification at the final layer of the CNN.

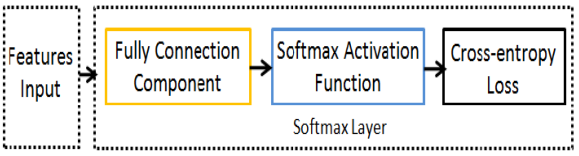

FIGURE 4.2: Softmax layer of Convolutional neural network

4.1.2 SVM

The reason for choosing SVM is its superior performance on medical imaging data. The results of SVM for non-linear boundaries are very impressive[57][58][59]. Support vector machines (SVM) is supervised learning model that uses an optimization method to identify support vectors s_i, weights α_i, and bias b that are used to classify vectors x according to the following equation

$$c = \sum_i \alpha_i k\left(s_i, x\right) + b \tag{4.3}$$

Where k is a kernel function. In the case of a linear kernel, k is the dot product. If $c \geq 0$, then x is classified as a member of the first group, otherwise it is classified as a member of the second group.

Error correcting code classifier is trained using support vector machine. The batch size is set to 256. training set is given to the classifier along with deep hierarchical feature vector using convolutional neural networks. validation data is then fed to the classifier to get the validation accuracy of the Herlev dataset.

Further information on SVM can be found at [60].

4.1.3 Ensemble of Decision trees

Ensemble of Decision trees are classification or regression trees used to predict the response of the data. In order to predict the response, the decisions about the belonging classes are followed from the initial or the root node to the end of the leaf node. Response of the data is contained within the leaf node. Classification tree provides nominal response while regression tree provides numeric responses.

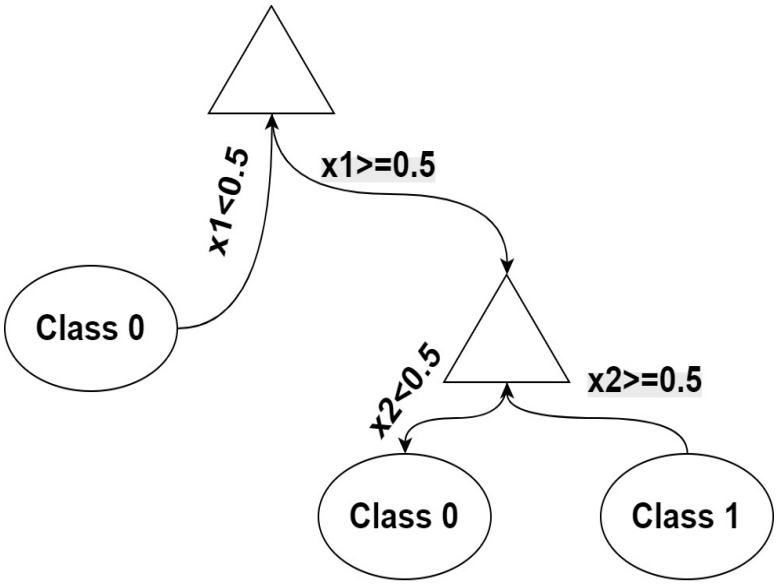

FIGURE 4.3: Decision Tree Classifier

The reason for choosing decision tree ensemble is its exploitation of randomness. For further information decision tree ensemble please refer to [61][62]. As shown in Figure 4.3 .This tree starts with two predictors $x1$ and $x2$. If $x1$ is less than 0.5 it is moved to the left, tree classifies it class 0. If not so the decision is moved to the right, where checks for $x2$ less than 0.5 moves to left node and classifies it as class 0 else classified as class 1.

Ensemble of decision trees is trained on the training data, number of trees are 100 and the batch size is set to 256.

Validation accuracy of the Herlev dataset is evaluated using validation data.

4.2 Testing

Multiple crops of unseen test images [12] are fed to the system and prediction score is generated from classifier for each of the cropped image. The score is then aggregated to achieve the final score [63] as shown in figure 4.5

In the data augmentation step, N_p images are generated which includes rotated and translated images. From each generated image patch, N_{sp} crops are generated

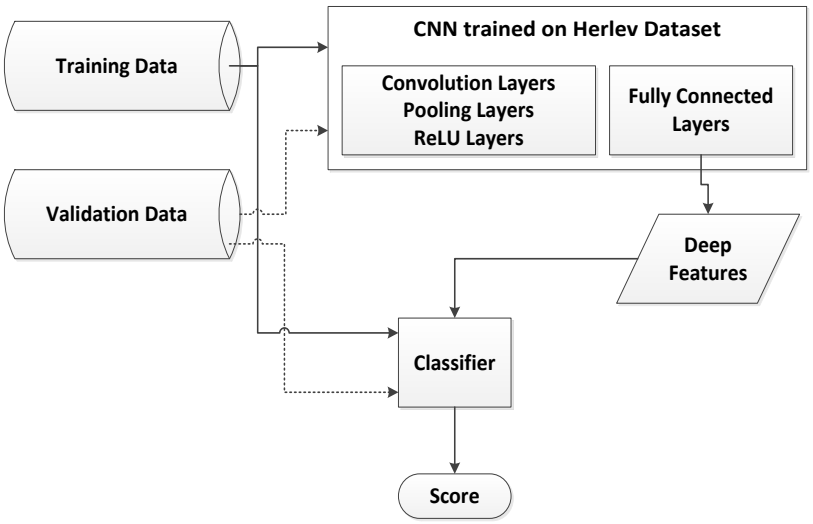

FIGURE 4.4: Classification System of Herlev Pap Smear Images

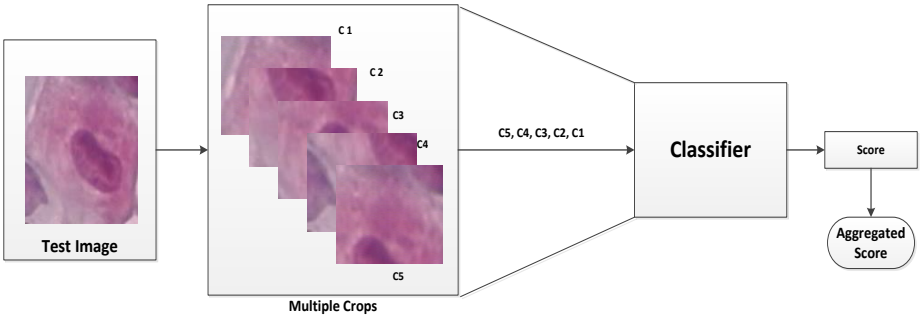

FIGURE 4.5: Multiple Crop testing Scheme

which includes 4 corners, center and their mirrored images. Therefore, for each test image, $N_p x N_{sp}$ are fed to the classifiers. Final score is achieved by aggregating or averaging the scores of $N_p x N_{sp}$ predictions.

Chapter 5

Implementation and Performance Evaluation

5.1 Herlev Dataset

The cervical cell data used for the training and testing of CNN comes from Herlev Pap smear dataset. Herlev dataset is publically available [1] generated at Herlev University Hospital. The cell images are collected by a digital camera and a microscope [38]. The resolution of cell images contained in the datasets is 0.201 um per pixel[36]. Conventional Pap smear staining and processing is used to generate specimens. The Herlev Pap smear dataset contains 917 single cell cervical cell images with ground truth classification and segmentation. The cells are categorized into seven different stage classes. These seven classes are diagnosed by doctors and cytologists to increase the reliability of the diagnosis. Further these seven classes are categorized into two categories i.e. malignant and benign. classes from first class to third class are normal or benign, while fourth to seventh class are abnormal or malignant classes. The classes distribution is shown in Table**??**.some of the examples of normal and abnormal images are shown in Figure 4.1. It can be seen that the size of the nucleus in malignant or abnormal cells is larger than that

[1] http://mde-lab.aegean.gr/index.php/downloads

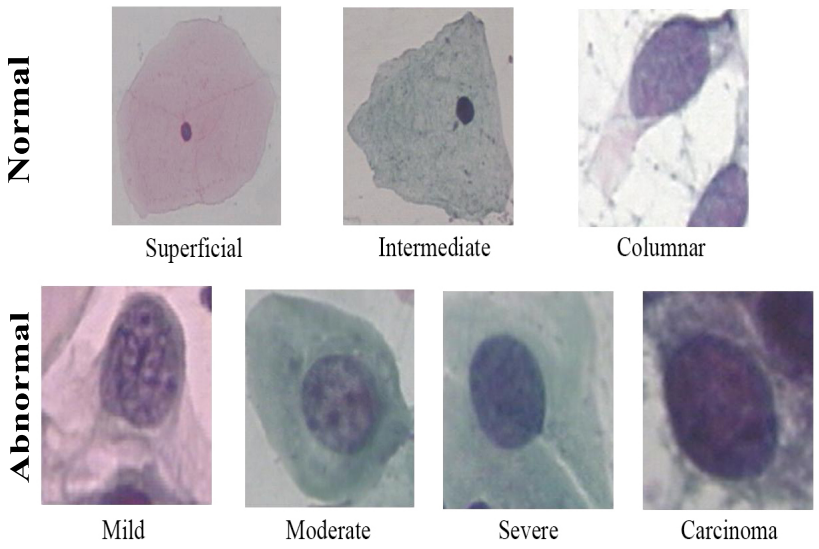

FIGURE 5.1: Normal Vs Abnormal Images in Herlev Dataset.

of normal cells. The challenging task from classification perspective is that the normal columnar cells have nucleus size quite similar to that of severe nucleus, also chromatin distribution is same. For data augmentation to reduce over-fitting and increase the size of the training samples as large samples are crucial for CNN performance, transformed copies of each and every cell image in the original dataset are generated for both malignant and benign cells. For all benign cell in the Herlev dataset, $N_r = 10$ rotations where $\theta = 36°$, $N_{th} = 15$ translations up to 15 pixels horizontally and $N_{tv} = 15$ translations up to 15 pixels vertically are performed. For each abnormal cell in the Herlev dataset, $N_r = 10$ rotations where $\theta = 36°$, $N_{th} = 8$ translations up to 15 pixels horizontally and $N_{tv} = 8$ translations up to 15 pixels vertically are performed. This processing results in 300 image patches for a single cell for normal category and 160 image patches for a single cell for abnormal category. This transformation yields relative normal distribution as the number of samples of abnormal cell images is large as compared to that of normal cell images. The size of the generated image patch is set to $m = 128$ pixels to cover cytoplasm region. These patches are then up-sampled to a size $256x256x3$ using bi-linear interpolation. These up-sampled image patches are fed to the CNN for initiating layer transfer and training[53].

Category	Class	Cell type	Count
Normal	1	Superficial squamous epithelial	74
Normal	2	Intermediate squamous epithelial	70
Normal	3	Columnar epithelial	98
Abnormal	4	Mild squamous non-keratinizing dysplasia	182
Abnormal	5	Moderate squamous non-keratinizing dysplasia	146
Abnormal	6	Severe squamous non-keratinizing dysplasia	197
Abnormal	7	Squamous cell carcinoma in situ intermediate	150

TABLE 5.1: Herlev Dataset

5.2 Architecture and Implementation

Architecture of network is shown in Figure 4.2. The base network CNN1 is pre-trained on ImageNet classification natural dataset. CNN1 contains five convolutional $(conv)$ layers denoted as $conv1 - conv5$, followed by pooling $(pool)$ layers denoted as $pool1, pool2, pool5$ and there are three fully connected (fc) layers as $fc6, fc7 and fc$. $conv$ layers and $pool$ layers are transferred to CNN2 at the same locations. In other words all $conv$ and $pool$ layers are copied from pre-trained network to new network to be trained on cervical cell dataset.Fully $fc6, fc7 and fc$ layers of CNN2 are initialized with values from random gaussian distributions. Configuration of CNN2 is listed in Table 5.2. Local response normalization for $conv1, 2$ is set according to the settings in [12].

Hidden layers are used with rectified linear units activation function. CCN1 and CNN2 shares same structure from $conv1$ to $pool5$. Number of neurons in fc layer of CNN1 and CNN2 are 4096 - 4096 - 1000 and 4096 - 4096 - 7 in case of 7 class problem, while number of neurons in fc layer of CNN2 are 4096 - 4096 - 2 in case of 2 class problem.

5.3 Training

Each image patch from augmented dataset, a 227 x 277 patch is cropped randomly and mean image over the dataset is subtracted to zero-center normalize the image. 227 x 227 is cropped for the reason that CNN1 takes 227 x 227 x 3 image at its input

TABLE 5.2: CNN Configuration

	Filter Size	Channel	Stride	Padding
Input	**-**	**3**	**-**	**-**
First Convolutional layer (conv1)	**11 x 11**	**96**	**4**	**-**
First Pooling Layer (pool1)	**3 x 3**	**96**	**2**	**-**
Second Convolutional Layer (conv2)	**5 x 5**	**256**	**1**	**2**
Second Pooling Layer (pool2)	**3 x 3**	**256**	**2**	**-**
Third Convolutional Layer (conv3)	**3 x 3**	**384**	**1**	**1**
Fourth Convolutional Layer (conv4)	**3 x 3**	**384**	**1**	**1**
Last Pooling Layer (pool5)	**3 x 3**	**256**	**2**	**-**
Full Connected Layer (fc6)	**-**	**4096**	**-**	**-**
Fully Connected Layer (fc7)	**-**	**4096**	**-**	**-**
Last Fully Connected Layer (fc)	**-**	**7 (7 Classes)** **2 (2 Classes)**	**-**	**-**

layer. Stochastic gradient descent (sgdm) is used for the training of CNN2 for 30 epochs. Small batches of image patches are fed to the CNN2, validation accuracy of batches are calculated. The size of Mini-batch is set to 256. Initial learning rate for convolutional and pooling layers is set to 0.0001, which is decreasing with a factor of 10 after every 10 epochs. L2-Regularization or weight decay is set to 0.0005. Momentum is set to 0.9. L2-Regularization and momentum can be tuned to reduce over-fitting of the CNN.

5.4 Testing

In order to test a new unseen image to the system, multiple patches (as for training data) and multiple crops of test image are generated. Abnormal score of each of the crop is generated by the classifier. The abnormal scores of all $N_{test}xN_{crop}$ patches of the test image are aggregated to generate the final score [51]. Patches

of test image $N_{test} = 300$(10 rotations x 30 translations) are generated same as the training images. Further ten cropped images N_{test} (four corner, center of cell i.e. nucleus portion and their mirrored images) are generated from each of the test patch. These $N_{test}xN_{crop}$ image patches are fed to the classifier. The prediction score of the classifier (Softmax regresseion, SVM, Tree Ensemble) is then aggregated to calculate the final score.

5.5 Evaluation Parameters

Evaluation of the cervical cell classification task is done using 5-fold cross validation on Herlev dataset for both two class problem and seven class problem. The performance metrics used for evaluation include accuracy, harmonic mean, area under the curve, specificity and sensitivity. Detail of the performance metrics is provided in Table 5.3. Finally the number of correct classification score is obtained

TABLE 5.3: Performance metrics for analysis

Sensitivity	Number of correctly classified malignant cells.
Specificity	Number of correctly classified benign cells.
Accuracy	Global correct classified cells.
Harmonic-mean	$\mathbf{2} * (\mathbf{Senstivity} * \mathbf{Specificity})/(\mathbf{Senstivity} + \mathbf{Specificity})$
AUC	Area under the curve

for each cell from all the categories in the Herlev dataset.

5.6 CNN Training

CNN2 is fine-tuned on Herlev dataset for 2 class problem and 7 class problem for 30 epochs as illustrated in Figure 5.2 and Figure 5.3 respectively. Validation accuracy of 99.3% and 87.45% is achieved for two class problem and seven class problem respectively. The comparison of softmax, SVM and tree ensemble used in this work is listed in Table 5.4. It can be seen that the performance decision tree is high as compared to SVM and softmax. Validation accuracy is evaluated

TABLE 5.4: Validation Accuracies during CNN training

	Softmax	**SVM**	**Tree Ensemble**
2-Class	99.35	99.8	100
7-Class	87.45	96.20	99.27

at the time of training. The system calculate the validation of the batch which is being fed to the network. The training and validation loss at the training time can be observed in Figure 5.3 for 2 class and 7 class problem. System validation

FIGURE 5.2: Training and Validation Loss during CNN training.

accuracy at the time of training can be viewed in Figure 5.3. Validation accuracy is calculated along with system validation and training loss. After fine tuning of the network, the layers of the network can be analyzed by a test image at the input of the network as shown in Figure 5.5a. At the first layer of fine tuned CNN, the convolutional layer (conv1) features learned are more generic as shown in Figure 5.4. It can be seen that these learned filters at first the convolutional layers contain gradients of different frequencies, blobs and orientations of colors which are of importance for the transfer of weights the cervical cell image.

For a cervical cells, these filters can be visualized for each of the layer of CNN. The features learned at convolutional layer *conv*1 for a test cervical cell is shown

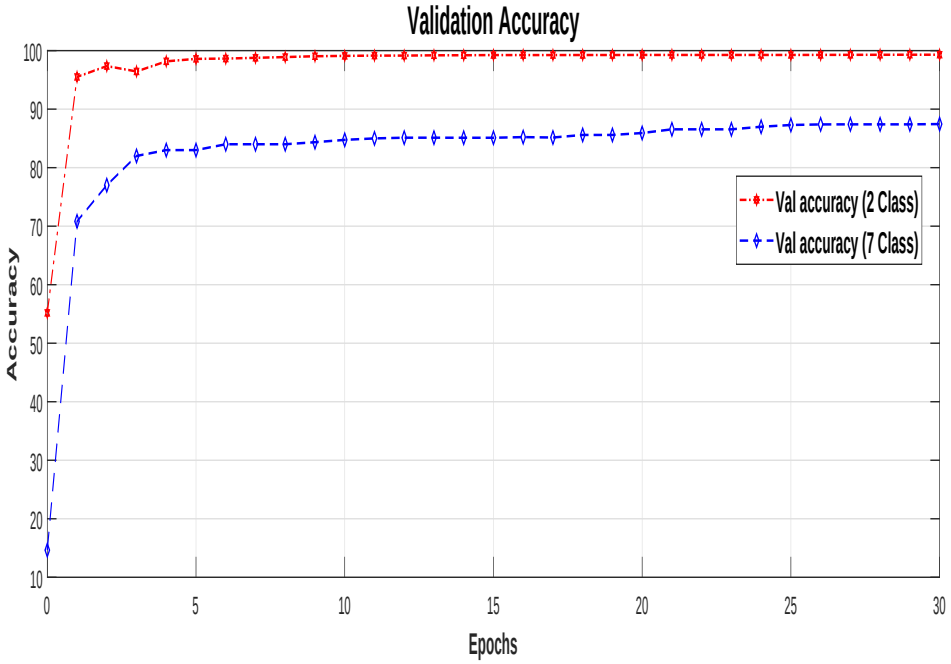

FIGURE 5.3: Validation Accuracy during CNN training.

in Figure 5.5. Features learned at convolutional layers *conv*1*toconv*5 are shown in Figure 5.6. As you get deeper the convolutional layers there are more abstract features and provide more information for the test cervical cell image.

Along with these filters learned at the convolutional layers, the activation or the feature map for the particular cervical cell can be visualized. Activation map for the cell shown in Figure 5.5a is shown in Figure 5.7. Strongest activation for the cervical cell in 5.5a at *conv*5 can be viewed in Figure 5.8. Feature set at fully connected layer is shown in Figure 5.9 and can be observed that features are more abstract as compared to the previous layers. Fully connected layer shows features learned for the seven classes i.e. superficial, intermediate, columnar epithelial, mild dysplasia, moderate dysplasia, severe dysplasia and carcinoma.

Strongest activation for the cell at pooling layer is shown in Figure 5.10. White pixels shows strong positive activation and black pixels shows strong negative activation while gray does not activate strongly. It can be observed that the strongest activation activates negatively on right edges, and positively on left edges. It can be observed that the pooling operation on cervical cell and the previous activation

FIGURE 5.4: Filters learned at First Convolutional layer (conv1).

maps summarizes by highlighting the activated spatial locations. In deeper layers of CNN, features learned are more abstract as shown in Figure 5.9.

5.7 Test analysis

The test set is fed to three classifiers i.e. softmax, SVM and tree ensemble using multiple crop testing. Comparison of performance for the classifiers is shown in Table 5.5. It can be seen that tree ensemble outperforms the classification accuracies of SVM and softmax regression. Malignancy score of several example from the validation set of Herlev dataset is shown in Figure 5.11 and Figure **??** for the abnormal and normal cell images respectively. Classification of a validation cell image is done using tree ensemble classifier. Figure 5.12 shows test cell images that

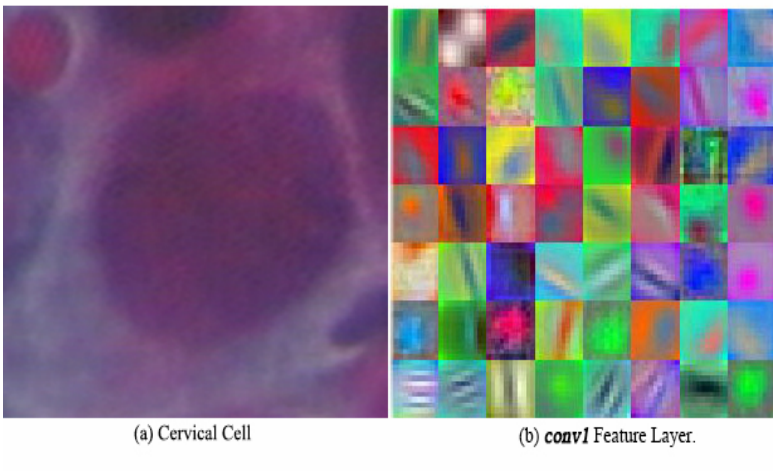

FIGURE 5.5: (a)Input Test Image (b)Feature maps learned at First Convolutional layer.

are misclassified (normal misclassified as abnormal and abnormal misclassified as normal).

TABLE 5.5: Classification Accuracies after Classifier training on deep features

	Softmax	**SVM**	**Tree Ensemble**
2-Class	98.80	99.10	99.23
7-Class	97.21	98.12	98.85

Evaluation parameters of the classification performance i.e. accuracy, harmonic mean, area under the curve, specificity and sensitivity of the fine-tuned CNN in comparison with the previous methods is illustrated in table 5.6. The mean values of accuracy, harmonic mean, area under the curve, specificity and sensitivity from fine-tuned CNN with t-Ensemble classifiers are 99.23%, 99.14%, 99.9%, 99.2% and 99.1% respectively for the two class problem. The mean values of accuracy, harmonic mean, area under the curve, specificity and sensitivity from fine-tuned CNN with t-Ensemble classifiers for seven class problem are 98.85%, 98.77%, 99.8%, 98.8% and 99.74% respectively. These performance metrics outperforms all the previous performance metrics except Sens in the literature.

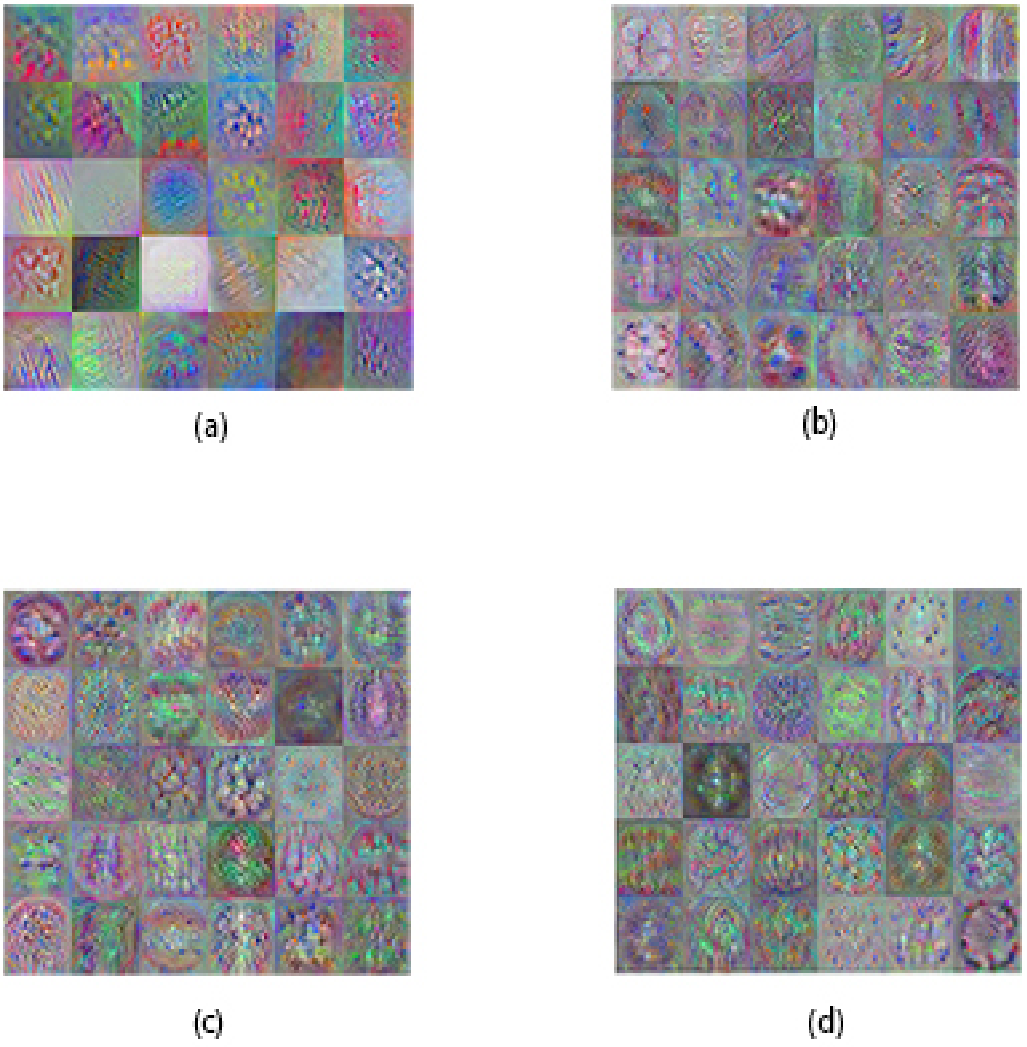

FIGURE 5.6: Features learned at (a). conv2, (b). conv3, (c). conv4 and (d). conv5 for cervical cell

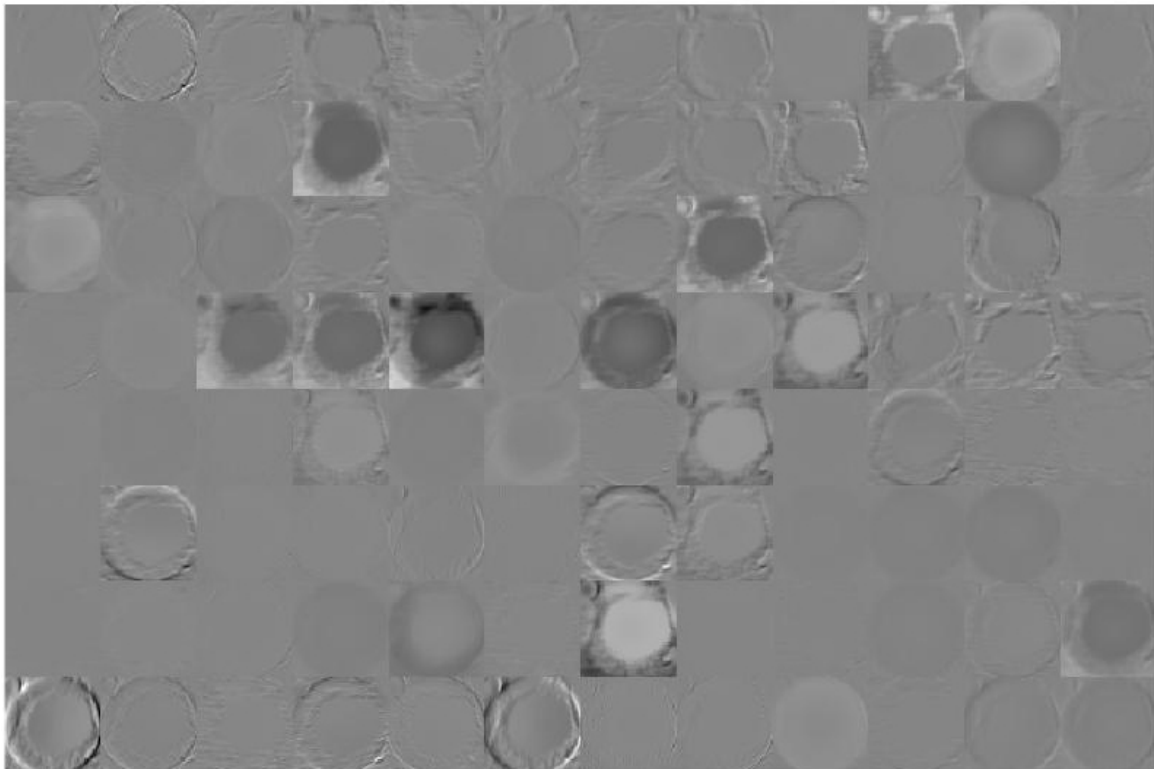

FIGURE 5.7: Montage of deepest Pooling Layer feature maps

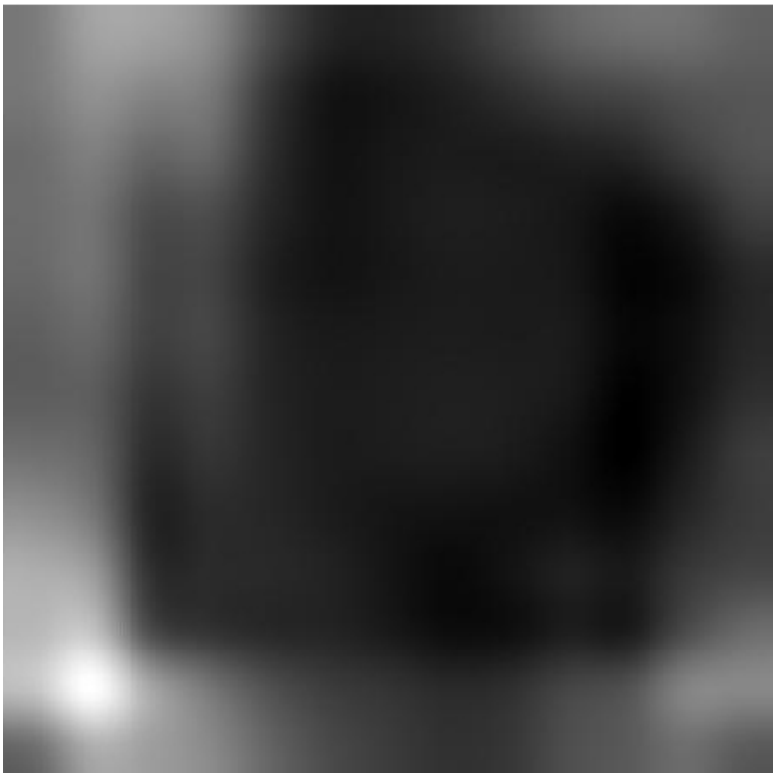

FIGURE 5.8: Strongest activation at deepest Convolutional layer(conv5)

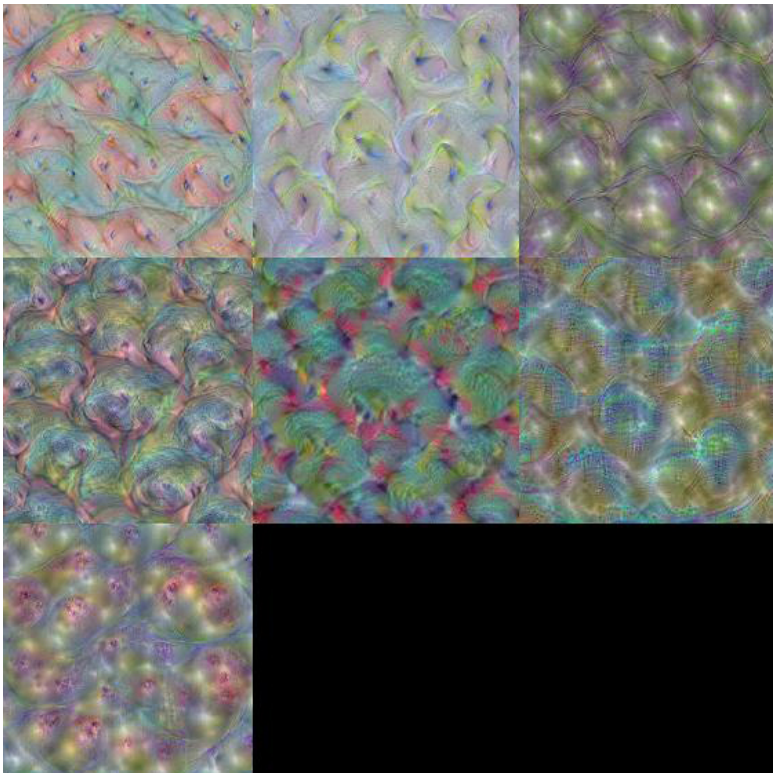

FIGURE 5.9: Fully connected layer feature maps for Seven classes

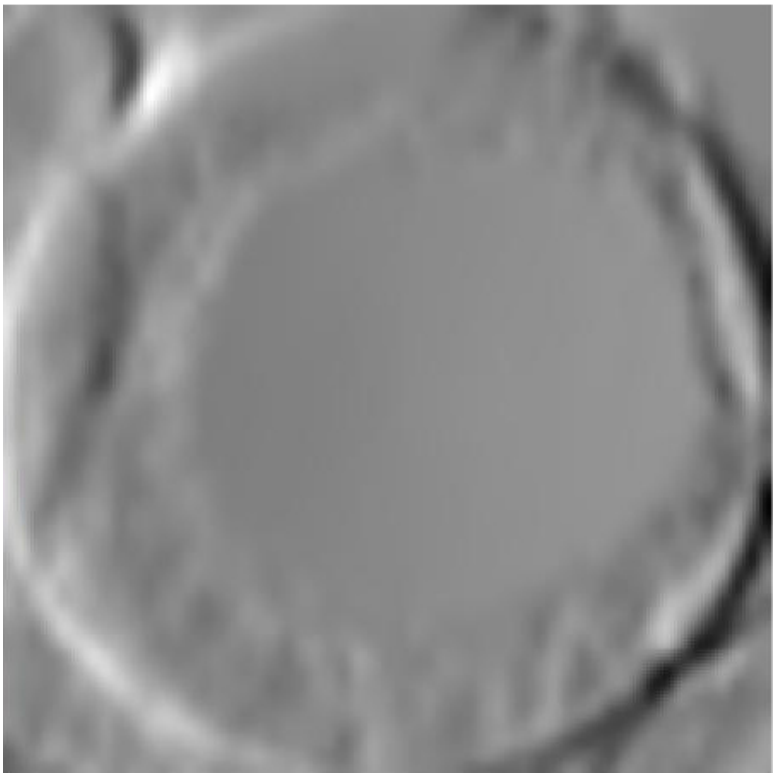

FIGURE 5.10: Strongest activation at deepest pooling layer

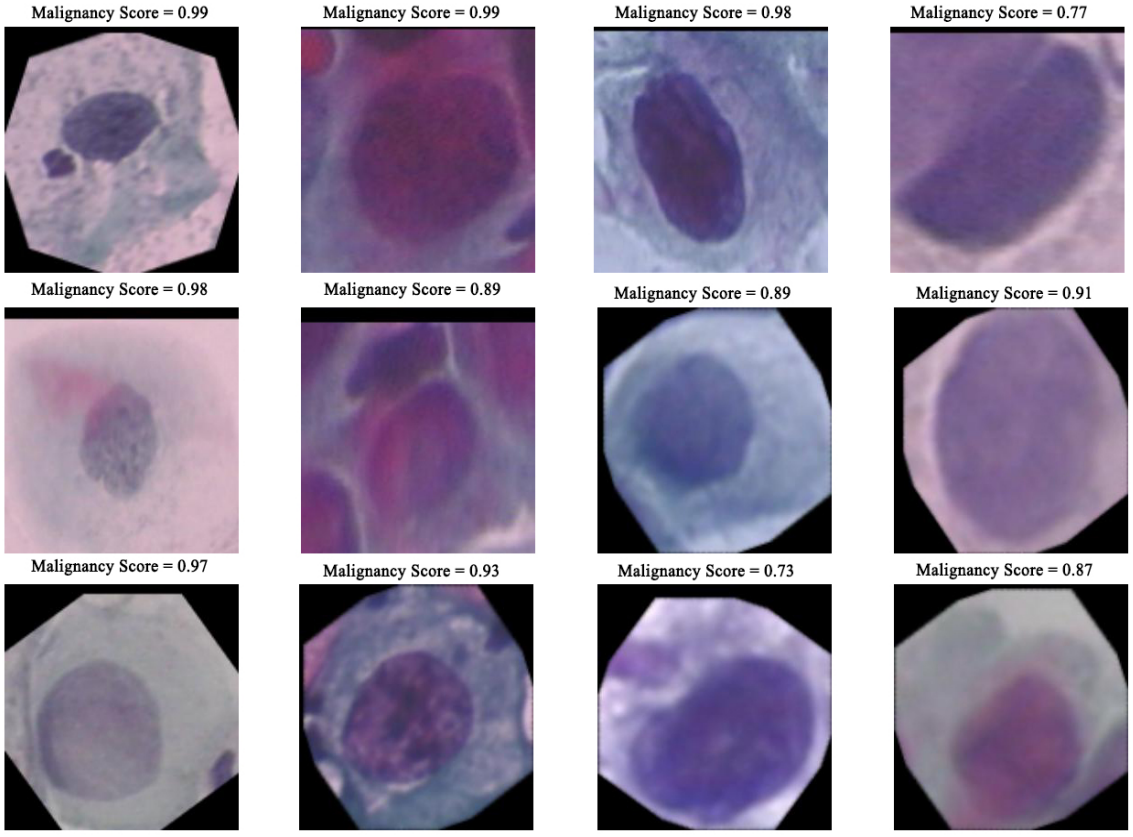

FIGURE 5.11: Correctly classified malignant cells, column 1 to 4 are mild dysplasia, moderate dysplasia, severe dysplasia and carcinoma, respectively.

TABLE 5.6: Hybrid and deep features based approaches without segmentation

References	**Method**	**Dataset**	**Performance Measure**
Jantzen et al. [38]	Pap Smear Benchmarks	Herlev Dataset	Accuracy : 93.6% Sensitivity : 98.8% Specificity : 79.3%
Jantzen et al. [37]	Particle swarm Optimization and nearest neighbor (1nn) Classifier	Herlev Dataset	Accuracy : 96.7% Sensitivity : 98.4% Specificity : 92.2%
Marinakis et al. [52]	Genetic Algorithm based Feature selection and 1-nn Classifier	Herlev Dataset	Accuracy : 96.8% Sensitivity : 98.5% Specificity : 92.1%
Plissiti et al. [39]	Principle Component Analysis for Feature selection and SVM Classifier	Herlev Dataset	Harmonic Mean : 96.9%
Bora et al. [27]	Ensemble Classifier	Herlev Dataset	Accuracy : 96.5% Sensitivity : 99% Specificity : 89.7%
Guo et al. [40]	Local Binary Pattern for feature extraction	Herlev Dataset	Area under the curve : 96.4%
Zhang et al. [50]	Convolutional Neural Networks for deep feature extraction and Classification	Herlev Dataset	Accuracy : 98.3% Sensitivity : 98.2% Specificity : 98.3%
This work	Convolutional Neural Networks for deep feature extraction SVM Classifier	Herlev Dataset	Accuracy : 99.10% Sensitivity : 98.9% Specificity : 99.2%
This work	Convolutional Neural Networks for deep feature extraction and Ensemble tree Classifier	Herlev Dataset	Accuracy : 99.23% Sensitivity : 99.2% Specificity : 99.3%

Classification accuracies of each of the seven cell categories are calculated feeding all the images in all the categories as test to the classifier trained on fine-tuned CNN weights. Classifier shows perfect performance on superficial squamous epithelial (Class 1), intermediate squamous epithelial (Class 2), mild squamous non-keratinizing dysplasia (Class 4) and moderate squamous non-keratinizing dysplasia (Class 5). The performance of the classifier is slightly low on columnar epithelial (Class 3) and severe squamous non-keratinizing dysplasia (Class 6). The classification accuracy of Class 4 and Class 6 is 97.5% and 97.79% respectively. Classification accuracy on squamous cell carcinoma in situ intermediate (Class 7) is 99.20%. The class scores are illustrated in table 5.7

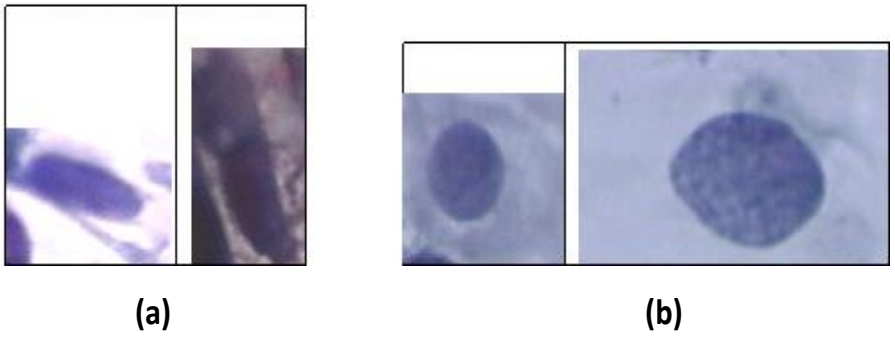

FIGURE 5.12: Misclassified (a)Normal as Abnormal (b) Abnormal as Normal

5.8 Computational complexity

In the training phase, CNN1 is trained on Corei7 machine with clock speed 2.8 GHz, Nvidia 1080Ti GPU and 8 GB of memory on MATLAB R2017b. The average training time of a fine-tuned CNN running for 30 epochs is about 4 hours and 30 minutes for two class problem and 8 hours and 20 minutes. In the testing phase, the system takes 8 seconds to classify a test image into normal and abnormal classes. Using multiple crop testing i.e. $N_{patches}$ x N_{crops} = 3000 classifications and score aggregation, the average time for the testing of one cell image is around 8 seconds.

TABLE 5.7: Herlev Dataset correct classifications.

Category	Class	Cell type	Count	Accuracy
Normal	1	Superficial squamous epithelial	74	100%
Normal	2	Intermediate squamous epithelial	70	100%
Normal	3	Columnar epithelial	98	97.50%
Abnormal	4	Mild squamous non-keratinizing dysplasia	182	100%
Abnormal	5	Moderate squamous non-keratinizing dysplasia	146	100%
Abnormal	6	Severe squamous non-keratinizing dysplasia	197	97.8%
Abnormal	7	Squamous cell carcinoma in situ intermediate	150	99.20%

Chapter 6

Conclusion and Future Work

6.1 Limitations

Despite higher performance of deep learning based cervical cell screening system, it has some limitations. Classification time of testing a cropped single cell image is 8 seconds for the system which is very slow in clinical setting as their are large number of samples from one pap smear slide image to be classified. This limitation can be addressed by neglecting the process of data augmentation step for the test data i.e. 300 image patches else only multiple crop testing can be used for classification problem. Although this increases the speed of the system as it requires only 0.08 seconds for classification while accuracy of the system is compromised by 1.5%. Although classification accuracy of the system on the Herlev dataset is high, the system misclassify few cells from Columnar epithelial class i.e 2.5% of the cells are misclassified, Severe squamous non-keratinizing dysplasia class i.e. 2.2% are misclassified and Squamous cell carcinoma in situ intermediate class i.e. 0.80% are misclassified. This misclassification error is not an ideal case because in real world these severe abnormal cells must not be classified as normal.

6.2 Conclusion

This thesis considers a convolutional neural network based approach for automatic feature extraction of cell image using Deep Convolutional Neural Network and classification of cervical cell image using different classification approaches i.e. softmax, SVM, and ensemble of decision trees.

Coarsely centered nucleus cell image patches are presented to the CNN as the input. Initial weights or features maps are transferred from a pre-trained network to a new CNN for fine tuning on cervical cell dataset. The features learned by the new fine-tuned CNN are extracted and fed to a softmax, SVM and ensemble of decision trees with surrogate splits classifiers along with training data for training. Validation accuracy of the dataset is calculated by feeding validation set to the trained classifier.

In the testing phase, for a single cell image to be test whether malignant or benign, the test image patches are generated same as training data and multiple crop testing is applied on all the patches to generate classification score by the classifier. The score of the classifier is then aggregated by averaging the score of all the crops extracted from the test image to get the final score.

The proposed method yields better classification accuracy i.e. 99.23% (Ensemble of decision trees) and 99.10% (Support vector machines) which is encouraging as compared to the previous best which is 98.6% [50] on Herlev University Pap smear dataset. A significant feature of the proposed model is that it do not require prior segmentation and manual feature extraction mechanisms. By looking into the results produced by the model, it can be argued with a comparatively higher degree of confidence that it can be effectively and efficiently used for the development of auto-assisted screening systems.

6.3 Future Work

Current system performance is evaluated on the images with majority of the cell are individual cells . This research aims to address the problem of feature extraction and classification of overlapping cells. Good classification accuracy is achieved on cell image data using ensemble of trees classifiers. In the future, the effect of inflammatory cells and cell particles (i.e. blood, mucus, bacteria, etc.) on classification accuracy need to be analyzed. The system should avoid the misclassification of inflammatory cells and cell particles as normal and abnormal cells. Specific classifiers relying on deep learning may be used to cater these problems.

References

[1] F. Jacques, S. Isabelle, D. Rajesh, E. Sultan, M. Colin, R. Marise, P. D. Maxwell, F. David, and B. Freddie, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012," *International Journal of Cancer*, vol. 136, no. 5, pp. E359–E386.

[2] F. Bray, B. Carstensen, H. Moller, M. Zappa, M. Zakelj, G. Lawrence, M. Hakama, and E. Weiderpass, "Incidence trends of adenocarcinoma of the cervix in 13 european countries," *Cancer Epidemiol Biomarkers Prev*, vol. 14, pp. 2191–2199, 2005.

[3] S. Batool, S. Sajjad, and H. Malik, "Cervical cancer in pakistan: A review," *The Journal of the Pakistan Medical Association*, vol. 67, no. 7, pp. 1074–1077, 2017.

[4] L. G. Koss, "The papanicolaou test for cervical cancer detection. a triumph and a tragedy," *The Journal Of The American Medical Association*, vol. 261, no. 5, pp. 737–743, 1989.

[5] M. F. Janicek and H. E. Averette, "Cervical cancer: prevention, diagnosis, and therapeutics," *CA: a cancer journal for clinicians*, vol. 51, pp. 92–114, 2001.

[6] J. Sherris, S. Wittet, A. Kleine, J. Sellors, S. Luciani, R. Sankaranarayanan, and M. Barone, "Evidence-based, alternative cervical cancer screening approaches in low-resource settings," *nternational perspectives on sexual and reproductive health*, vol. 35, no. 3, pp. 147–152, 2009.

[7] R. DeMay, "Common problems in papanicolaou smear interpretation," *Archives of Pathology and Laboratory Medicine*, vol. 121, pp. 229–238, 1997.

[8] M. Arbyn, A. Herbert, U. Schenck, P. Nieminen, J. Jordan, E. Mcgoogan, J. Patnick, C. Bergeron, J.-J. Baldauf, P. Klinkhamer, J. Bulten, and P. Martin-Hirsch, "European guidelines for quality assurance in cervical cancer screening: recommendations for collecting samples for conventional and liquid-based cytology*," *Cytopathology*, vol. 18, no. 3, pp. 133–139, 2007.

[9] M. Desai, "Role of automation in cervical cytology," *Diagnostic Histopathology*, vol. 15, no. 7, pp. 323–329, 2009.

[10] E. Bengtsson and P. Malm, "Screening for cervical cancer using automated analysis of pap-smears," *Computational and Mathematical Methods in Medicine*, pp. 1–12, 2014.

[11] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[13] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[14] R. M. Summers, "Progress in fully automated abdominal ct interpretation," *American Journal of Roentgenology*, pp. 1–13, 2016.

[15] J. Zhang and Y. Liu, "Cervical cancer detection using svm based feature screening," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*, C. Barillot, D. R. Haynor, and P. Hellier, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 873–880.

[16] M. E. Plissiti, C. Nikou, and A. Charchanti, "Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 233–241, March 2011.

[17] A. Garrido and N. P. de la Blanca, "Applying deformable templates for cell image segmentation," *Pattern Recognition*, vol. 33, no. 5, pp. 821 – 832, 2000.

[18] E. Bak, K. Najarian, and J. P. Brockway, "Efficient segmentation framework of cell images in noise environments," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, Sept 2004, pp. 1802–1805.

[19] Y. Liu, T. Zhao, and J. Zhang, "Learning multispectral texture features for cervical cancer detection," in *Proceedings IEEE International Symposium on Biomedical Imaging*, 2002, pp. 169–172.

[20] N. Ashidi and N. A. Mat Isa, "Automated edge detection technique for pap smear images using moving k-means clustering and modified seed based region growing algorithm," vol. 13, pp. 45–59, 10 0002.

[21] M. E. Plissiti, C. Nikou, and A. Charchanti, "Combining shape, texture and intensity features for cell nuclei extraction in pap smear images," *Pattern Recognition Letters*, vol. 32, no. 6, pp. 838 – 853, 2011.

[22] I. Muhimmah, R. Kurniawan, and Indrayanti, "Analysis of features to distinguish epithelial cells and inflammatory cells in pap smear images," in *2013 6th International Conference on Biomedical Engineering and Informatics*, Dec 2013, pp. 519–523.

[23] D. Riana, M. E. Plissiti, C. Nikou, D. H. Widyantoro, T. L. R. Mengko, and O. Kalsoem, "Inflammatory cell extraction and nuclei detection in pap smear images," *Int. J. E-Health Med. Commun.*, vol. 6, no. 2, pp. 27–43, Apr. 2015.

[24] D. Riana, D. H. Widyantoro, and T. L. Mengko, "Extraction and classification texture of inflammatory cells and nuclei in normal pap smear images," in

2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), Nov 2015, pp. 65–69.

[25] N. Lassouaoui, L. Hamami, and N. Nouali, "Morphological description of cervical cell images for the pathological recognition," *Proc. of World Academy of Science Engineering and Technology WASET. Org*, pp. 49–52, 2005.

[26] F. Cao, S. Chen, and L. Zeng, "New abnormal cervical cell detection method of multi-spectral pap smears," *Wuhan University Journal of Natural Sciences*, vol. 12, no. 3, pp. 476–480, May 2007.

[27] K. Bora, M. Chowdhury, L. B. Mahanta, M. K. Kundu, and A. K. Das, "Automated classification of pap smear images to detect cervical dysplasia," *Computer Methods and Programs in Biomedicine*, vol. 138, pp. 31 – 47, 2017.

[28] Rahmadwati, G. Naghdy, M. Ros, C. Todd, and E. Norahmawati, "Cervical cancer classification using gabor filters," in *2011 IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology*, July 2011, pp. 48–52.

[29] T. Chankong, N. Theera-Umpon, and S. Auephanwiriyakul, "Automatic cervical cell segmentation and classification in pap smears," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 539 – 556, 2014.

[30] P. Bamford and B. Lovell, "A water immersion algorithm for cytological image segmentation," 03 1998.

[31] O. Lezoray and H. Cardot, "Cooperation of color pixel classification schemes and color watershed: a study for microscopic images," *IEEE Transactions on Image Processing*, vol. 11, no. 7, pp. 783–789, Jul 2002.

[32] S. F. Yang-Mao, Y. K. Chan, and Y. P. Chu, "Edge enhancement nucleus and cytoplast contour detector of cervical smear images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 2, pp. 353–366, April 2008.

[33] C.-H. Lin, Y.-K. Chan, and C.-C. Chen, "Detection and segmentation of cervical cell cytoplast and nucleus," *International Journal of Imaging Systems and Technology*, vol. 19, no. 3, pp. 260–270.

[34] M. E. Plissiti, C. Nikou, and A. Charchanti, "Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 233–241, March 2011.

[35] Y. Song, J. Z. Cheng, D. Ni, S. Chen, B. Lei, and T. Wang, "Segmenting overlapping cervical cell in pap smear images," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, April 2016, pp. 1159–1162.

[36] E. Martin, "Pap-smear classification," Master's thesis, Technical University of Denmark, 2003.

[37] Y. Marinakis, M. Marinaki, and G. Dounias, "Particle swarm optimization for pap-smear diagnosis," *Expert Systems with Applications*, vol. 35, no. 4, pp. 1645 – 1656, 2008.

[38] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," *Nature inspired Smart Information Systems*, pp. 1–9, 2005.

[39] M. E. Plissiti and C. Nikou, "Cervical cell classification based exclusively on nucleus features," in *Image Analysis and Recognition*, A. Campilho and M. Kamel, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 483–490.

[40] Y. Guo, G. Zhao, and M. Pietikinen, "Discriminative features for texture description," *Pattern Recognition*, vol. 45, no. 10, pp. 3834 – 3843, 2012.

[41] L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artificial Intelligence in Medicine*, vol. 49, no. 2, pp. 117 – 125, 2010.

[42] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS'14. Cambridge, MA, USA: MIT Press, 2014, pp. 3320–3328.

[43] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[44] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, Jan 2016.

[45] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, April 2015, pp. 294–297.

[46] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 11, pp. 2369–2380, Nov 2016.

[47] T. Xu, H. Zhang, C. Xin, E. Kim, L. R. Long, Z. Xue, S. Antani, and X. Huang, "Multi-feature based benchmark for cervical dysplasia classification evaluation," *Pattern Recognition*, vol. 63, pp. 468 – 475, 2017.

[48] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Igum, "Automatic segmentation of mr brain images with a convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1252–1261, May 2016.

[49] Y. Song, J. Z. Cheng, D. Ni, S. Chen, B. Lei, and T. Wang, "Segmenting overlapping cervical cell in pap smear images," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, April 2016, pp. 1159–1162.

[50] L. Zhang, L. Lu, I. Nogues, R. M. Summers, S. Liu, and J. Yao, "Deeppap: Deep convolutional networks for cervical cell classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1633–1643, Nov 2017.

[51] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec 2015.

[52] Y. Marinakis, G. Dounias, and J. Jantzen, "Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification," *Computers in Biology and Medicine*, vol. 39, no. 1, pp. 69 – 78, 2009.

[53] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[54] Z. Gao, L. Wang, L. Zhou, and J. Zhang, "Hep-2 cell image classification with deep convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 2, pp. 416–428, March 2017.

[55] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[56] B. Chen, W. Deng, and J. Du, "Noisy softmax: Improving the generalization ability of dcnn via postponing the early softmax saturation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 4021–4030.

[57] W. Lirong, M. Chao, W. Peter, and X. Xiang-Qun, "Linear and nonlinear support vector machine for the classification of human 5-ht1a ligand functionality," *Molecular Informatics*, vol. 31, no. 1, pp. 85–95.

[58] Y. Weng, C. Wu, Q. Jiang, W. Guo, and C. Wang, "Application of support vector machines in medical data," in *2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS)*, Aug 2016, pp. 200–204.

[59] Y. Jiang, Z. Li, L. Zhang, and P. Sun, "An improved svm classifier for medical image classification," in *Rough Sets and Intelligent Systems Paradigms*, M. Kryszkiewicz, J. F. Peters, H. Rybinski, and A. Skowron, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 764–773.

[60] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, July 1998.

[61] S. Bernard, L. Heutte, and S. Adam, "On the selection of decision trees in random forests," in *2009 International Joint Conference on Neural Networks*, June 2009, pp. 302–307.

[62] J. J. Rodríguez, J. F. Díez-Pastor, and C. García-Osorio, "Ensembles of decision trees for imbalanced data," in *Multiple Classifier Systems*, C. Sansone, J. Kittler, and F. Roli, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 76–85.

[63] H. R. Roth, L. Lu, J. Liu, A. Yao, J.and Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1170–1181, 2016.