

CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY,  
ISLAMABAD



# A Hybrid Sine-Cosine Binary Chimp Optimization Algorithm for Feature Selection Applications

by

Hassan Mohyuddin

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

in the  
Faculty of Engineering  
Department of Electrical Engineering

2023

Copyright © 2023 by Hassan Mohyuddin

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.

*I dedicate this thesis to my parents, my family and my wife, whose unwavering love, support, and encouragement have been the driving force behind my academic journey. Their belief in my abilities and sacrifices have made this achievement possible. This work is a tribute to their unwavering commitment and serves as a reflection of the values they instilled in me.*



## CERTIFICATE OF APPROVAL

### A Hybrid Sine-Cosine Binary Chimp Optimization Algorithm for Feature Selection Applications

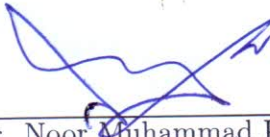
by

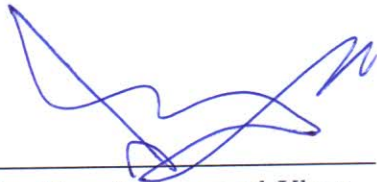
Hassan Mohyuddin

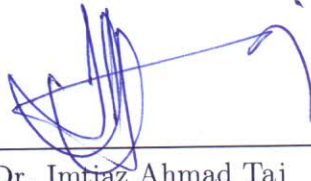
(MEE213003)

### THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Adil Masood	Military College of Signals
(b)	Internal Examiner	Dr. Fazal ur Rehman	C.U.S.T
(c)	Supervisor	Dr. Noor Muhammad Khan	C.U.S.T

  
\_\_\_\_\_  
Dr. Noor Muhammad Khan  
Thesis Supervisor  
November, 2023

  
\_\_\_\_\_  
Dr. Noor Muhammad Khan  
Head  
Dept. of Electrical Engineering  
November, 2023

  
\_\_\_\_\_  
Dr. Imtiaz Ahmad Taj  
Dean  
Faculty of Engineering  
November, 2023

## *Author's Declaration*

I, **Hassan Mohyuddin** hereby state that my MS thesis titled “**A Hybrid Sine-Cosine Binary Chimp Optimization Algorithm for Feature Selection Applications**” 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.

A handwritten signature in black ink, appearing to read 'Hassan', with a long, sweeping underline that extends to the right.

(**Hassan Mohyuddin**)

Registration No: MEE213003

---

## *Plagiarism Undertaking*

I solemnly declare that research work presented in this thesis titled “**A Hybrid Sine-Cosine Binary Chimp Optimization Algorithm for Feature Selection Applications**” is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly 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.



(Hassan Mohyuddin)

Registration No: MEE213003

## *Acknowledgement*

I am profoundly grateful to Allah (S.W.A), the Almighty, for His boundless blessings and guidance throughout my research journey. His unwavering support, divine intervention, and grace have been the driving force behind my accomplishments. I am indebted to my parents and my family, whose love, unwavering belief in my abilities, and continuous prayers have provided me with the strength and motivation to overcome challenges and achieve success. Their unwavering support has been an invaluable gift. I extend my heartfelt appreciation to my thesis supervisor, Dr. Noor Muhammad Khan, for his invaluable guidance, mentorship, and expertise. His unwavering dedication and insightful feedback have played a pivotal role in shaping my research and academic growth. Additionally, I express my gratitude to the faculty members and staff of Department of Electrical Engineering, CUST for fostering a nurturing academic environment and providing resources that have enriched my learning experience. Their commitment to excellence and intellectual growth has been truly inspiring.



(Hassan Mohyuddin)

## *Abstract*

The utilization of biomedical data for various applications, such as disease diagnosis, drug discovery, and treatment optimization, has become increasingly prevalent in recent years. In this context, the selection of relevant features from large-scale biomedical datasets plays a pivotal role in enhancing the accuracy and efficiency of data analysis and modelling. This thesis explores the significance of feature selection in biomedical data analysis and introduces a novel hybrid meta-heuristic algorithm tailored to address this critical issue.

The research begins by highlighting the challenges posed by the complexity and dimensionality of biomedical data, emphasising the need for effective feature selection methods to alleviate the curse of dimensionality and improve the interpretability of machine learning models. A comprehensive review of existing feature selection techniques and their limitations serves as a foundation for the development of an innovative hybrid meta-heuristic algorithm that combines the strengths of multiple optimization techniques to efficiently identify and select the most informative features from biomedical datasets. A combination of the Chimp Optimization Algorithm (ChOA) and Sine-Cosine Algorithm (SCA) is developed. This hybrid approach aims to strike a balance between exploration and exploitation, ensuring the discovery of relevant features while maintaining computational efficiency. To evaluate the efficacy of the developed algorithm, a comparative analysis is conducted with existing feature selection methods, including Particle Swarm Optimization (PSO), Grey Wolf Optimization (GWO), Whale Optimization Algorithm (WOA), Genetic Algorithm (GA), and Chimp Optimization Algorithm (ChOA). The research focuses on the algorithm's ability to improve classification accuracy and reduce the dimensionality of biomedical data, ultimately enhancing the interpretability of machine learning models. To assess the algorithm's real-world applicability, a series of experiments are conducted using eight distinct datasets taken from the University of California, Irvine (UCI) Machine Learning repository and the Mendely Data: Data for Gene Selection repository. These datasets represent a range of large multi-modal biomedical data, reflecting the diversity and complexity of real-world healthcare scenarios. The experiments employ popular classification techniques, including k-Nearest Neighbours (KNN) and Random Forest (RF), to rigorously test the algorithm's performance across different biomedical data domains. The findings of this thesis contribute



to the advancement of feature selection methodologies in biomedical data analysis, offering a promising solution to the challenges posed by high-dimensional datasets. The results of the comparative analysis demonstrate the superiority of the proposed hybrid meta-heuristic algorithm in terms of feature selection efficacy, showcasing its potential to significantly enhance the accuracy and interpretability of machine learning models for biomedical applications.

# Contents

<b>Author's Declaration</b>	<b>iv</b>
<b>Plagiarism Undertaking</b>	<b>v</b>
<b>Acknowledgement</b>	<b>vi</b>
<b>Abstract</b>	<b>vii</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xii</b>
<b>Abbreviations</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Big Data . . . . .	2
1.3 Machine Learning . . . . .	2
1.4 Feature Selection in Machine Learning . . . . .	4
1.5 Meta-Heuristic Algorithms . . . . .	5
1.5.1 Filter Based Algorithms . . . . .	5
1.5.2 Embedded Methods . . . . .	6
1.5.3 Wrapper Based Algorithms . . . . .	6
1.6 Thesis Contributions . . . . .	7
1.7 Thesis Outline . . . . .	9
<b>2 Literature Review</b>	<b>10</b>
2.1 Deep Learning (DL) Based Feature Selection . . . . .	10
2.2 Meta-heuristic Algorithms . . . . .	12
2.2.1 Physics Based Algorithms . . . . .	13
2.2.1.1 Simulated Annealing (SA) . . . . .	13
2.2.2 Human-Social Interaction Based Algorithms . . . . .	14
2.2.2.1 Group Teaching Optimization Algorithm (GTOA) . . . . .	14
2.2.2.2 Tabu Search (TS) . . . . .	14
2.2.3 Biology Based Algorithms . . . . .	15
2.2.3.1 Particle Swarm Optimization (PSO) . . . . .	15
2.2.3.2 Genetic Algorithm (GA) . . . . .	17

---

2.2.3.3	Grey Wolf Optimization (GWO) . . . . .	18
2.2.3.4	Whale Optimization Algorithm (WOA) . . . . .	19
2.2.3.5	Chimp Optimization Algorithm (ChOA) . . . . .	20
2.2.4	Mathematics Based Algorithms . . . . .	24
2.2.4.1	Sine-Cosine Algorithm (SCA) . . . . .	24
2.2.5	Variants for Metaheuristic Algorithms . . . . .	26
2.2.5.1	Hybrid MHAs . . . . .	26
2.2.5.2	Improved MHAs . . . . .	27
2.2.5.3	Multi-objective Algorithms . . . . .	29
2.2.6	Limitations of MHAs . . . . .	31
2.3	Research Gap . . . . .	33
2.4	Problem Statement . . . . .	33
2.5	Thesis Objectives . . . . .	34
2.6	Research Methodology . . . . .	34
<b>3</b>	<b>Binary Sine-Cosine Chimp Optimization Algorithm (BS-ChOA)</b>	<b>35</b>
3.1	Feature Selection Using BS-ChOA . . . . .	39
3.2	Complexity Analysis . . . . .	41
<b>4</b>	<b>Results and Discussions</b>	<b>42</b>
4.1	Dataset Preprocessing . . . . .	43
4.2	Experimental Setup . . . . .	43
4.3	Results . . . . .	44
4.3.1	Average and Best Metrics . . . . .	44
4.3.2	Transformation Functions . . . . .	62
4.3.3	Convergence Curve . . . . .	63
4.4	Discussion . . . . .	64
<b>5</b>	<b>Conclusion</b>	<b>65</b>
5.1	Future Work and Limitations . . . . .	66
	<b>Bibliography</b>	<b>67</b>

# List of Figures

1.1	Process Diagram for Feature Selection Process . . . . .	4
2.1	Attacking Strategy of Chimps[71] . . . . .	21
2.2	Particle Movement in SCA [75] . . . . .	25
3.1	Chimp Movement in the search space . . . . .	37
3.2	a) 2D Movement matrix of Chimps b) 3D Movement matrix of Chimps [71]. . . . .	37
3.3	Flow Diagram: Feature Selection Using BS-ChOA . . . . .	40
4.1	Average and Best Accuracy . . . . .	45
4.2	Average and Best Accuracy . . . . .	46
4.3	Average and Best Precision . . . . .	48
4.4	Average and Best Precision . . . . .	49
4.5	Confusion Matrix 11 Tumour Dataset . . . . .	50
4.6	Confusion Matrix Prostate Tumour Dataset . . . . .	51
4.7	Confusion Matrix Prostate Cancer Dataset . . . . .	52
4.8	Average and Best F1-Score . . . . .	54
4.9	Average and Best F1-Score . . . . .	55
4.10	Average and Best Number of Selected features . . . . .	57
4.11	Average and Best Number of selected features . . . . .	58
4.12	Average Metrics for All datasets . . . . .	60
4.13	Average Metrics for All datasets . . . . .	61
4.14	Average number of features selected using s and v TF . . . . .	62
4.15	Average Metrics for s and v TF . . . . .	63
4.16	Convergence curve for BS-ChOA with simple average for 10 iterations . . . . .	63
4.17	Convergence curve for BS-ChOA with weight average for 10 iterations . . . . .	63

# List of Tables

4.1	Dataset Details	43
4.2	Mean and Best Accuracy	47
4.3	Average and Best Precision	53
4.4	Average and Best F1-Score	56
4.5	Average and Best Selected Features	59

# Abbreviations

<b>AI</b>	Artificial Intelligence
<b>ANN</b>	Artificial Neural Network
<b>BChOA</b>	Binary Chimp Optimization Algorithm
<b>CFS</b>	Correlation-based feature selection
<b>ChOA</b>	Chimp Optimization Algorithm
<b>CNN</b>	Convolutional Neural Network
<b>DLFS</b>	Deep Learning-based Feature Selection
<b>FS</b>	Feature Selection
<b>FSA</b>	Feature Selection Algorithms
<b>GA</b>	Genetic Algorithm
<b>GWO</b>	Grey-Wolf Optimization
<b>IG</b>	Information Gain
<b>IoT</b>	Internet of Things
<b>KNN</b>	K-Nearest Neighbour
<b>LDA</b>	Linear Discriminant Analysis
<b>MHA</b>	Meta-Heuristic Algorithm
<b>MI</b>	Mutual Information
<b>ML</b>	Machine Learning
<b>NFL</b>	No Free Lunch
<b>PCA</b>	Principal Component Analysis
<b>PSO</b>	Particle Swarm Optimization
<b>RFE</b>	Recursive Feature Elimination
<b>RF</b>	Random Forrest
<b>RNN</b>	Recurrent Neural Network
<b>SCA</b>	Sine- Cosine Algorithm

**SVM** Support Vector Machines

**WOA** Whale Optimization Algorithm

# Chapter 1

## Introduction

### 1.1 Background

In today's interconnected world, the proliferation of digital gadgets, internet platforms, and technical breakthroughs has resulted in an unparalleled explosion of data generation. Large volumes of data have been generated in industries such as healthcare, banking, and social media making it difficult to store, analyze, and efficiently predict outcomes from data with such variable and complex structures. Data generated by online transactions, emails, social media interactions, sensors, and mobile phone applications is becoming increasingly challenging to store and manage using traditional database systems. Every day, roughly 2.5 Quintilian bytes of data are created in cyberspace as a result of the transition of manual operations into automated solutions [1]. To conduct a comprehensive analysis on a large scale, it is necessary to collect varied data from numerous sources. Smartphones, cameras, and wearable devices, for example, can collect data on location, movement, and activity (the data is usually connected in space and time, such as traffic sensors located on the same stretch of road), whereas Internet of Things (IoT) devices, such as smart thermostats and security cameras, can collect data pertaining to environmental conditions and activities in homes and buildings. Storing every piece of data from every source is effectively redundant because the same segment of data is being stored but from various sources [2]. Furthermore, the data acquired is typically not in any predefined format, making analysis of such unstructured data impossible. Moreover, the introduction of cloud computing has made it simple and cost-effective to store and handle enormous volumes of data. The social media platforms Facebook, Twitter, and



Instagram create massive volumes of data about their users' interests, preferences, and interactions. Online platforms such as Netflix, Amazon, YouTube, and Spotify collect information about their users' purchases, viewing and listening habits, and preferences. Additionally, the rise of e-commerce and online marketplaces has resulted in massive volumes of data being created about customer behaviours and preferences.

## 1.2 Big Data

The collection of enormous amounts of data and the huge benefits seen from its usage gave rise to the phenomenon known as "Big Data". In recent times the utilization of big data has emerged as a pivotal force driving innovation, efficiency, and informed decision-making across a multitude of sectors. With the exponential growth of interconnected devices, online interactions, and data-generating systems, the potential of big data to unveil insights and patterns hidden within immense information repositories has become increasingly evident [3]. From personalised marketing strategies and predictive maintenance in industries to healthcare diagnostics and urban planning, the strategic application of big data analytics is revolutionising the way organizations and individuals harness information to address challenges and seize opportunities in our complex and dynamic world. However, simply collecting a huge volume of data alone would not improve systems; it is not only the quantity of data but also its quality that influences a system's output [4]. Poor-quality data can have a negative impact on decision-making and business operations since it can lead to erroneous and unreliable results. Such effects can have a substantial impact on the effectiveness and success of the organization. As a result, improving data quality is critical in order to avoid risks and make sound decisions based on reliable information [5].

## 1.3 Machine Learning

The advancements in machine learning (ML) and artificial intelligence (AI) have led to increased data gathering and examination. ML models need a substantial amount of data to be trained and enhanced, and with the increasing accessibility of data storage and processing, it is easier to collect, store, and analyse large amounts of data. Furthermore, businesses are using data analytics to gain insights into consumer behaviour and

preferences, leading to further increased data storage. The need to examine a significant amount of data has led to the development of more sophisticated and accurate ML models, which in turn leads to more data being generated. The symbiotic relationship between big data and advances in ML has brought about a new era of unimaginable possibilities in the field of modern technology. ML approaches have risen to the challenge of extracting useful patterns and insights from this information flood as the volume and diversity of data created continue to grow [6]. The fusion of big data and ML has paved the way for predictive analytics, natural language processing (NLP), image recognition, and personalised recommendations. From classic algorithms like decision trees and linear regression to cutting-edge methods like deep learning and neural networks. This synergy has not only transformed industries, but has also enabled researchers and practitioners to unravel complicated intricacies, allowing previously unthinkable feats of automation, optimization, and cognitive computing to be realised [7].

Contradictory to the above, defective sensors and missing data points warrant a regular small-scale data update and are also critical issues for big data analytics. A fraudulent credit card, for example, must be warned and retained for future transactions. Another challenge for real-time analysis is data privacy and ownership, because there are rigorous regulations restricting the use of sensitive data, limiting the applications and methodologies that can be studied. As a result, the emergence of machine learning, which is expected to minimise processing time and data storage, has certain negative consequences in terms of data analysis.

One problem associated with managing mixed big data is automatically creating sufficient metadata to characterise the recorded information. Knowing how much data to save and how to use it is critical since only actionable and relevant data should be saved. A primary focus for data science engineers is the subset dataset stored from a big data platform that can be used later to reconstitute the original data. Materials used in semiconductor-based storage systems have become scarce as natural resources have depleted. To avoid the problems caused by the expanding amount of data, academics have concentrated their efforts on creating and keeping only valuable data. This is accomplished through the use of multiple effective incremental ingestion approaches for feature selection.

## 1.4 Feature Selection in Machine Learning

ML approaches can be used to turn vast, unstructured datasets into smaller, more valuable ones. Dimensionality reduction strategies simplify the model and eliminate issues like noise accumulation, misleading correlations, and unintentional homogeneity, which can lead to decreased processing costs [8]. While collecting raw data subsets, combining the human experience with automatic feature extraction provides for a quicker modelling approach. For example, in an Artificial Neural Network (ANN), the buried layer nodes generate internal representations that are equivalent to built features. Even though the hidden layers in an ANN network can be automated, the hyper-parameter values, such as the number of hidden layer neurons, are often adjusted by a human because each data set has a different ANN model. Feature selection algorithms (FSAs') are becoming increasingly important in a wide range of applications such as computer vision, object identification, computer vision, voice and speech recognition, data analysis, information processing, and machine learning, in bio-informatics. In the pursuit of extracting meaningful insights and enhancing the performance of machine learning models, the strategic selection of relevant features from a vast dataset has become a cornerstone of success. FS techniques, ranging from classical methods to advanced algorithms, play a pivotal role in identifying the most informative attributes that drive predictive accuracy and reduce dimensionality [9].

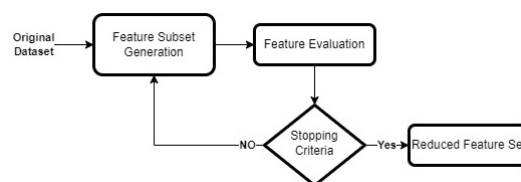


FIGURE 1.1: Process Diagram for Feature Selection Process

Principal Component Analysis (PCA) [10], a powerful dimensionality reduction technique, aims to capture the maximum variance within data by transforming it into a new set of uncorrelated features. Linear Discriminant Analysis (LDA) is a statistical technique that projects high-dimensional data onto a lower-dimensional space while maximising the separation between different classes, making it a powerful tool for classification and dimensionality reduction tasks [11]. Alongside these cornerstones, a multitude of classical techniques such as Recursive Feature Elimination (RFE), Chi-Square,

and Information Gain work cohesively to curate feature subsets that not only enhance model interpretability but also mitigate the risk of over-fitting, ultimately yielding models that are both efficient and effective in tackling real-world challenges. A new field of algorithms in this respect is meta-heuristic algorithms. A literature survey indicates numerous feature selection strategies, each having merits and demerits.

## 1.5 Meta-Heuristic Algorithms

Meta-heuristic algorithms (MHAs') are a form of modern algorithms that have been utilised to tackle a range of optimization problems, including feature selection. These algorithms are very effective for handling computationally intensive problems, as they can identify near-optimal solutions by exploring the search space with heuristics and stochastic processes [12]. One of the primary benefits of utilising MHAs' for feature selection is their ability to handle complex datasets. This is because MHAs' can explore a huge number of possible solutions efficiently, allowing them to discover salient features even when working with high-dimensional data [13]. MHAs' for feature selection are mainly classified into three types: filter-based algorithms, embedded algorithms, and wrapper-based algorithms.

### 1.5.1 Filter Based Algorithms

A widely used feature selection technique is filter-based algorithms, in which each feature's significance or relevance is evaluated independently of the classification model based on a criterion [14]. Mutual information (MI), Information Gain (IG), and Correlation-based Feature Selection (CFS) are examples of often employed criteria. The MI criteria assess the connection between a feature and a class label. The concept is that a trait that is strongly related to the class label is thought to be more informative. MI is a non-parametric measure that may be used for any feature and class label [15].

IG is a mutual information variant used in decision tree algorithms. It calculates the reduction in entropy of the class label after using the feature to partition the data [16]. CFS is a filter-based method that measures the correlation between a class and a feature. The correlation coefficient, which assesses the linear relationship between two variables, is the foundation of CFS. It is a well-known method that is used in a wide array of

applications, including bio-informatics and text classification. It is a simple method that may be applied to both discrete and continuous data. While these strategies are easy to implement and efficient, they may not always be effective. As an example, the CFS method only considers the relevance of each individual feature to the class label, and cannot find complex relationships between multiple classes and class labels [17]. Furthermore, these methods do not consider the specific learning algorithm that will be employed and may not be optimised for the performance of that algorithm. Another concern with filter-based feature selection is that it is sensitive to criterion selection, which can result in variable outcomes depending on the criterion chosen. For example, reciprocal information and information gain are two alternative criteria that can be used to assess the importance of a feature, but they may yield different outcomes [18]. Furthermore, these approaches can be sensitive to outliers, leading to a bias towards specific traits [19].

### 1.5.2 Embedded Methods

Embedded methods (EM) are another class of FSAs' based on the optimization criterion embedded in the classifier's training process [20]. These methods optimise the criterion by modifying the classifier parameters, such as feature weights in linear classifiers or split parameters in decision trees [21]. LASSO [22] and Random Forest (RF) [23] are some of the commonly used methods. These strategies necessitate greater computing power than the filter and wrapper approaches, but they can produce better outcomes by taking into account the link between the features and the class label [24]. One of the main problems with EM-based FSA is that they necessitate the use of a specific classifier or algorithm, which might result in inefficient feature selection if the classifier or algorithm is not well suited to the dataset. Additionally, EM-based FSAs' fail to account for data distribution, resulting in insufficient feature selection due to outliers, perhaps resulting in a bias towards specific features [25].

### 1.5.3 Wrapper Based Algorithms

Wrapper-based algorithms are another powerful class for feature selection. The aforementioned methods assess the efficacy of a classifier by employing various subsets of features. The wrapper approach is a search strategy that employs an iterative process

of selecting or removing features based on the performance of the classifier [26]. The two wrapping approaches that are commonly used in practice are Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) [27]. The wrapper strategy necessitates a higher allocation of processing resources compared to the filter method. However, it has the potential to yield enhanced outcomes by considering the interplay between features and the class label. One of the issues associated with this particular aspect of feature selection pertains to computational costs, as it necessitates the repeated training of a classifier or algorithm using various subsets of data. The management of massive datasets and high-dimensional data can be a substantial challenge. Furthermore, the requirement for a designated algorithm imposes constraints on its adaptability to diverse datasets and tasks [28].

In addition, there are also methods that are based on the use of clustering techniques for feature selection [29]. Such techniques originated from the idea of grouping similar features together and selecting the most representative features from each group [30]. Commonly used clustering-based feature selection methods include k-means and hierarchical clustering [31]. These methods can achieve better performance by considering the relationships between features and the class label, and by identifying the most relevant features in each group. There are other statistical test-based feature selection approaches [32]. These approaches evaluate each feature's statistical significance in regard to the class label [33]. The chi-squared test, t-test, and ANOVA are examples of commonly used statistical-based feature selection procedures. These methods can outperform others by taking into account the links between features and class labels and identifying the most relevant features based on statistical significance.

## 1.6 Thesis Contributions

This thesis presents feature selection approaches that employ metaheuristic algorithms to reduce the dimensionality of datasets with various sizes. The primary objective of this thesis is to investigate the effectiveness of combining two algorithms, namely the Binary Chimp Optimization Algorithm (BChOA) and the Sine Cosine Algorithm (SCA). This research is motivated by the notable achievements reported in the existing literature on hybrid metaheuristic implementation [34–38]. The performance of the proposed technique addressed the constraints of the BChOA in effectively solving high-dimensional

feature selection. This study proposes a hybrid adaptation of the Binary Chimp Optimization metaheuristic approach, which is combined with the SCA. The effectiveness of the BChOA technique in locating global minima is boosted by the optimization of its ability to navigate the solution search space of the underlying issue landscape, resulting in improved outcomes. The algorithm under consideration is referred to as the Hybrid Sine Cosine – Chimp Optimization Algorithm (BSChOA). The proposed methodology involves the utilization of the SCA as a subordinate component, working in conjunction with the BChOA to enhance the efficiency of global solution exploration by reducing the time frame required. The domain of metaheuristic algorithms is undergoing ongoing development as scholars increasingly acknowledge the necessity for a broad spectrum of methodologies. The continual pursuit of new and novel metaheuristic optimization algorithms persists in accordance with the No Free Lunch (NFL) principle, which asserts that no single approach can universally solve all optimization problems. The continuous endeavour is motivated by the recognition that innovative algorithms are essential for effectively tackling the distinct challenges presented by diverse optimization tasks. The key contributions of this thesis in this regard are as follows:

1. A FS model is developed with the objective of generating a subset feature dataset. The primary goals of this model are to minimise the storage requirements of extensive data, eliminate redundant features, and improve the accuracy of the classification model.
2. A novel hybrid MH model, referred to as BSChOA, has been devised by integrating the ChOA and the high-performing SCA. This integration aims to harness the exploration and exploitation capabilities of both algorithms, thereby improving the overall search performance and yielding superior outcomes for the feature selection problem.
3. The search technique employed by the ChOA is improved with the incorporation of the SCA within the mathematical model framework.
4. The performance of the proposed feature selection technique is assessed by employing KNN and RF-based classifiers on a total of 8 datasets. These datasets consist of multi-class unbalanced high-dimensional biomedical datasets taken from UCI and Mendeley datasets for Gene Selection repositories. Furthermore, a comprehensive

comparative analysis is conducted to assess the performance of the aforementioned algorithm in comparison to conventional approaches.

## 1.7 Thesis Outline

This chapter provides a concise examination of the underlying rationale for utilising FS based dimensionality reduction strategies in addressing the challenges associated with the analysis of large-scale datasets. [Chapter 2](#), examines previous studies on feature selection methods that rely on deep learning models. It also explores the associated difficulties and the application of metaheuristic algorithms in the process of feature selection. Furthermore, the thesis also provides an in-depth analysis of several implementations of metaheuristic algorithms. [Chapter 3](#) introduces the mathematical model of the BS-ChOA, taking into account the cost/fitness function, system update policy, and computational complexity. In addition, this chapter provides an explanation of the unique strategy for feature selection using variance-based methods, as well as the artificial neural network model that is based on the BS-ChOA algorithm. In [Chapter 4](#), the method is evaluated against contemporary metaheuristic algorithms using multiple biomedical datasets. The evaluation is conducted based on several performance metrics, including accuracy, precision, F1-score and time complexity. A comprehensive analysis of the obtained outcomes, demonstrating the superiority of the proposed BSChOA and the variance-based feature selection strategy over conventional methodologies is also presented in this chapter. [Chapter 5](#) concludes the study by providing a comprehensive overview of the research conducted and presents valuable perspectives on potential future avenues for exploration.



## Chapter 2

# Literature Review

### 2.1 Deep Learning (DL) Based Feature Selection

The utilization of deep learning techniques has gained significant traction within the machine learning domain, particularly in the context of FS, as seen by their growing popularity in recent years. The use of deep neural networks (DNN) to find and pick the most pertinent and informative features for a given dataset is a novel approach known as Deep Learning based Feature Selection (DLFS). According to [18], FS methods based on DL can acquire intricate data representations. This ability enables them to discern characteristics that hold significance for a given task or classifier. Furthermore, it has been observed that DL-based techniques for FS possess the capability to take into account the intricate relationships and interactions among features. This characteristic has the potential to enhance the precision and accuracy of the feature selection process [39].

One of the primary benefits associated with DLFS techniques is its capability to effectively manage high-dimensional datasets and large-scale data collections. The rationale behind this is that deep learning-based methods for feature selection can acquire knowledge about the underlying structure of the data, enabling them to discern pertinent features even in scenarios involving data with a large number of dimensions. Moreover, it has been shown that DLFS techniques possess the capability to consider the underlying data distribution, resulting in enhanced accuracy in the process of feature selection [25].

Auto-encoders (AE) are widely recognised as one of the prevailing approaches for feature selection in deep learning. AE are a type of neural network that undergoes training to recreate the input data by utilising a representation of reduced dimensionality. The concealed layers of the auto-encoder can be perceived as a condensed depiction of the input data, serving as a method for selecting features. The utilization of the AE-based feature selection technique has been observed in diverse fields, such as medical diagnosis [40], text classification [41], and image recognition [42]. Xu *et al.* in [43] introduced an AE- based approach for feature selection in the context of drug response prediction. The approach employed a hybridization of the random forest technique and the Boruta algorithm to demonstrate the efficacy of the proposed model.

The Deep Belief Networks (DBNs) are another robust method for feature selection in the context of deep learning. DBNs are a type of hierarchical generative model that acquires a layered representation of the input data. These networks consist of many Restricted Boltzmann Machines (RBMs) arranged in a stacked manner. DBNs have demonstrated successful applications in diverse domains, including speech recognition [44] and drug development [45]. DBN-based feature selection algorithms have been widely employed in the domain of bio-informatics for gene expression data analysis and bio-marker development [46, 47]. An illustrative instance may be seen in the study conducted by Ding *et al.* [48], whereby deep neural networks were effectively employed to identify pertinent traits and categorise medications into therapeutic classes using their transcriptional profiles. Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are widely employed deep learning architectures for feature selection across many domains. CNNs have demonstrated a high level of effectiveness in the field of image analysis and have gained significant popularity for their ability to extract features and perform classification in many computer vision applications. Furthermore, CNNs have demonstrated their efficacy not only in the realm of image classification but also in the field of feature selection across several domains, including natural language processing and speech recognition. In [49] a multi-modal CNN was devised to extract and choose pertinent characteristics from both image and text data to identify counterfeit news.

RNNs conversely, are specifically engineered to effectively analyse sequential data and are frequently employed in various domains such as natural language processing, speech recognition, and time series analysis. RNNs possess the capability to acquire knowledge

of temporal dependencies present within the data, enabling them to discern significant aspects that hold relevance in the prediction of forthcoming values. In a recent study conducted by [50], the researchers introduced a cost-sensitive deep active learning approach that utilises a double-deep neural network (double-DNN) incorporating three fundamental neural network architectures (1D CNNs, LSTM, and GRU). The objective of this approach is to enhance the detection of epileptic seizures while minimising the need for extensive sample labelling. The suggested methodology demonstrates a reduction in the number of labelled samples by a maximum of 33% and 80% when compared to uncertainty sampling and random sampling, respectively. Both CNN and RNN have been integrated with other deep learning architectures to enhance the performance of feature selection. In a recent work by [51], the researchers devised a hybrid CNN-RNN model to discern pertinent elements from electroencephalogram (EEG) data to forecast the commencement of seizures. The model demonstrated superior accuracy compared to conventional feature selection techniques, although employing a reduced number of features. Although DL/ML-based feature selection algorithms have demonstrated effectiveness, they are not without their accompanying problems. One of the primary obstacles lies in the necessity for a considerable volume of data to effectively train deep neural networks, a predicament that can become problematic when confronted with limited datasets. Moreover, it is worth noting that feature selection approaches based on deep learning might impose a significant computational burden due to the necessity of training extensive neural networks. To address the challenges stemming from these limits, the academic literature offers an alternative category of algorithms that employ randomised movement as a fundamental approach to achieving optimal solutions within much-reduced timeframes. These algorithms are within the category of metaheuristic algorithms.

## 2.2 Meta-heuristic Algorithms

MHAs are a class of approximation optimization approaches that employ probabilistic techniques for solving problems. In essence, meta-heuristics provide domain-specific expertise through the utilization of a high-level strategy that facilitates the sharing of solutions among multiple nodes executed in parallel. The algorithm functions are divided into two distinct phases, namely the exploration phase and the exploitation phase. The

exploration phase is characterised by the search for a global solution within the search space through unpredictable movements. On the other hand, the exploitation phase aims to converge towards the global solution of the current iteration. The efficient search process is heavily influenced by the trade-off between the two phases, which is determined by the hyper-parameters specified for the algorithm. From a motivational perspective, metaheuristic algorithms can be categorised into four distinct groups: biology-based, mathematics-based, physics-based, and human-social interaction-based [52–55].

### 2.2.1 Physics Based Algorithms

Physics-based algorithms are derived from the fundamental rules of physics, encompassing principles from several domains such as physics, chemistry, music, dynamic systems, and metallurgical processes. The authors of the study [37] introduced a novel approach for dimensionality reduction in the context of speech emotion recognition, specifically, they recommended the utilization of the golden ratio-based equilibrium optimization method. Simulated Annealing (SA) and Gravitational Search Algorithm (GSA) are some well-known physics-based MHAs.

#### 2.2.1.1 Simulated Annealing (SA)

Simulated Annealing (SA) is an optimization algorithm that draws inspiration from the annealing process observed in metallurgy. The SA algorithm, first presented in [56], has gained significant popularity in addressing intricate optimization issues that pose challenges for conventional optimization methods. The algorithm emulates the process of gradual cooling and crystallization observed in materials to explore the global optimum within a specified search space.

SA operates on the notion of probabilistic acceptance, which enables it to successfully explore the search space and avoid being trapped in local optima. The algorithm commences by initialising a solution and afterwards proceeds to iteratively perturb it through the introduction of minor random modifications. Subsequently, the quality of the newly proposed solution is assessed using an objective function, and a decision is made about its acceptance or rejection. This decision is influenced by a probability that is defined by the temperature at the current state and the disparity between the objective function

values of the new solution and the current solution. As the algorithm advances, there is a steady drop in temperature, resulting in a heightened concentration of the search around potential solutions that show promise. The mathematical formulation of simulated annealing encompasses the utilization of a temperature schedule and a cooling function.

## 2.2.2 Human-Social Interaction Based Algorithms

### 2.2.2.1 Group Teaching Optimization Algorithm (GTOA)

GTOA is a metaheuristic optimization algorithm that draws inspiration from the instructional methods employed in classroom settings. It was first proposed by Zhang *et al.* [57] as a method to effectively explore and find the best solution for intricate optimization issues by simulating the collaborative interaction between students and professors. The algorithm employs a combination of exploration and exploitation tactics to attain a harmonious equilibrium between global and local search.

In the context of GTOA, the population of candidate solutions is metaphorically represented as a collective of pupils, while the optimal solution is analogously regarded as the most proficient instructor. The method commences by initialising the placements of the student group within the search space through a random generation process. Each student thereafter modifies their position by taking into account the impact of the highest-performing student within the group, the influence of the most accomplished teacher, and the historically optimal position. This procedure facilitates the examination of various regions within the search space, while simultaneously capitalising on the knowledge gained from the most optimal solutions discovered thus far. The efficacy of GTOA can be influenced by many parameter choices, including the group size (i.e., the number of students in the group) and the influence coefficients utilised in the position update equation. Precise parameter calibration is crucial to guarantee the attainment of optimal performance across various problem areas.

### 2.2.2.2 Tabu Search (TS)

TS algorithm, which was proposed by Fred W. Glover in his 1998 work [58], is a metaheuristic optimization technique. The inspiration for this notion stems from the idea

of utilising memory and leveraging past search experiences to enhance the effectiveness of the search process. Tabu search (TS) is well-suited for addressing combinatorial optimization issues characterised by the presence of a vast search space and a multitude of local optima, making the identification of the global optimum particularly problematic. Within the context of TS, the algorithm effectively manages a transient memory component referred to as the Tabu list. This list serves the purpose of monitoring and recording solutions that have been recently explored. The presence of this memory effectively inhibits the algorithm from re-examining previously explored solutions or executing actions that result in known unsatisfactory regions. The implementation of a Tabu list in TS restricts some moves, hence encouraging the investigation of unexplored locations and facilitating the diversity of the search process. The algorithm employs an iterative approach to systematically investigate the search space by executing successive neighbourhood moves from the current solution. The evaluation of each move is conducted through the utilization of an objective function, and thereafter, the optimal move is selected based on its ability to enhance the objective value. Nevertheless, TS also permits the consideration of moves that result in a deterioration of the objective value to avoid being stuck in local optima. This acceptance of sub-optimal motions is typically determined by aspiration criteria or strategic oscillation tactics.

### **2.2.3 Biology Based Algorithms**

Algorithms falling under the category of biology-based algorithms draw inspiration from the social behaviour exhibited by avian, pis-cine, and mammalian species, as well as the biological mechanisms governing evolutionary processes. The Particle Swarm Optimization (PSO) algorithm and Genetic Algorithm (GA) are the most notable examples within this category.

#### **2.2.3.1 Particle Swarm Optimization (PSO)**

The PSO method is widely recognised as a prominent metaheuristic approach that draws inspiration from the collective behaviour observed in bird flocking or fish schooling. It was initially proposed by Kennedy and Eberhart in 1995 [59], and has gained significant popularity in the field of optimization due to its straightforwardness and effectiveness. The method emulates the combined intellectual capacity of a group of organisms,

wherein particles symbolising various solutions traverse the search space to locate the most effective solution. Every particle updates its position by considering both its own best solution and the best solution discovered by the entire swarm. The aforementioned iterative procedure persists until an acceptable solution is attained. PSO algorithm can be formally expressed in the following manner:

Every individual particle  $i$  within the swarm possesses a location vector denoted as  $X_i$  and a velocity vector denoted as  $V_i$ . The position and velocity are iteratively updated using the formulae provided below:

$$X_n(k+1) = X_n(k) + V_n(k+1), \quad (2.1)$$

$$V_n(k) = w.V_n(k) + c_1.r_1(P_n - X_n(k)) + c_2.r_2(G - X_n(k)), \quad (2.2)$$

Here,  $X_n(k+1)$  represents the updated position of particle  $n$  at iteration  $k+1$ . Similarly,  $V_n(k+1)$  denotes the updated velocity of the particle. The variable  $w$  corresponds to the inertia weight, while  $c_1$  and  $c_2$  represent the acceleration coefficients. The values  $r_1$  and  $r_2$  are random numbers ranging from 0 to 1. The term  $P_n$  denotes the personal best position of particle  $n$ , and  $G$  represents the global best position among all particles in the swarm.

Numerous research studies have provided evidence of the efficacy of PSO in addressing diverse optimization challenges. For example, in [60] the author employed PSO to tune the parameters of Support Vector Machines (SVMs) for classification tasks. Their study showed that PSO outperformed other optimization approaches in terms of achieving greater performance.

In a separate inquiry conducted by Zhang *et al.* (2019) [61], the utilization of PSO was employed to enhance the arrangement of wind turbines within a wind farm, leading to notable enhancements in power generation efficiency. The aforementioned examples serve to demonstrate the adaptability and wide-ranging applicability of PSO in various fields. In [62] Luo *et al.* introduced a PSO-based approach for feature selection in the analysis of gene expression data with high dimensionality. The algorithm employed a process of gene selection, identifying a subset of genes that had the highest discriminatory power in discriminating between various types of cancer. This resulted in a notable enhancement in the accuracy of cancer classification.

Although PSO has demonstrated usefulness, it is important to acknowledge that it also possesses certain limits. One of the primary obstacles encountered in swarm optimization is the phenomenon known as premature convergence, wherein the collective fails to sufficiently explore the entirety of the search space and becomes trapped in a poor solution. Several strategies have been proposed to address this problem, including adaptive inertia weights and velocity clamping. Another constraint that should be considered is the susceptibility of PSO to the selection of control parameters, such as inertia weight and acceleration coefficients, which necessitate meticulous calibration for varying issue scenarios. Furthermore, the issue of scalability arises in the context of PSO when confronted with optimization problems of significant magnitude, mostly due to the substantial computational expenses involved.

### 2.2.3.2 Genetic Algorithm (GA)

Genetic Algorithm (GA) belongs to a category of optimization algorithms that draw inspiration from the ideas of natural selection and genetics. GA was first proposed by Holland in the 1970s [63]. Since then, they have become widely recognised and utilised for their remarkable ability to efficiently solve intricate optimization problems. The algorithm emulates the phenomenon of evolution by upholding a population of prospective solutions and iteratively implementing genetic operators, such as selection, crossover, and mutation, to produce new progeny. By use of an iterative procedure, GA aims to locate an ideal solution by progressively enhancing the population across successive generations.

GA can be formally stated in the following manner: The representation of a population of individuals is in the form of a collection of chromosomes. These chromosomes consist of strings of binary or real-valued genes that encode potential solutions. The algorithm progresses iteratively over a sequence of generations. Within each successive cohort, individuals undergo assessment and scrutiny depending on their level of fitness, which serves as a measure of the excellence and efficacy of their proposed solutions. The utilization of selection operators, such as roulette wheel selection or tournament selection, is a common practice in evolutionary algorithms to determine the individuals with superior fitness levels who will be chosen as parents for the subsequent generation. Crossover operators are utilised to merge the genetic material of two parents, resulting



in the creation of kids. Conversely, mutation operators introduce random alterations to the genetic makeup of the offspring's genes. The aforementioned procedure is iterated over multiple generations until an acceptable resolution is achieved.

A plethora of studies have provided evidence about the efficacy of GA across several disciplines. Li *et al.* (2019) in [64] employed a GA to optimise the deployment of wireless sensor networks, resulting in enhanced coverage and energy efficiency when compared to alternative approaches. In a separate investigation conducted by Babatunde, *et al.* [65], GA was utilised to perform feature selection in the context of picture recognition. This approach led to a reduction in the dimensionality of the feature space and an improvement in the accuracy of classification. The aforementioned examples underscore the multifaceted nature of GA in effectively addressing optimization challenges within several domains.

One of the primary obstacles faced by GA is computing complexity, particularly when dealing with large-scale optimization problems. This arises from the necessity of evaluating the fitness of an entire population in each generation as a fundamental step of the algorithm. Several strategies, including parallelization and fitness approximation, have been suggested as potential solutions to address this problem. GA suffers from premature convergence, a phenomenon in which the algorithm becomes ensnared in a sub-optimal solution, impeding its ability to thoroughly explore the entirety of the search space. To tackle this issue, scholars have put up sophisticated operators and procedures for maintaining diversity. Furthermore, the selection of parameter values, such as the size of the population and the rate of mutation, can have a substantial influence on the effectiveness of GA and frequently necessitate meticulous calibration.

### 2.2.3.3 Grey Wolf Optimization (GWO)

GWO algorithm is a distinctive and robust metaheuristic algorithm that draws inspiration from the hunting behaviour of grey wolves in the natural world. The GWO algorithm, as presented by Mirjalili *et al.* [66], has garnered significant interest within the optimization community owing to its straightforwardness and effectiveness in addressing intricate optimization challenges. The algorithm simulates the social structure and cooperative hunting tactics observed in grey wolves, employing an iterative approach to systematically explore and identify the most efficient option. A collective group of

wolves in the context of GWO symbolically represents a population of potential solutions. The process commences by randomly initialising the positions of the alpha, beta, and delta wolves, which serve as representations of the most optimal solutions discovered thus far. Subsequently, every individual wolf within the pack proceeds to modify its position by the positions held by the alpha, beta, and delta wolves, while also taking into consideration its capacities for exploration and exploitation. The programme iteratively advances, enabling the wolves to adjust and optimise their placements until a desirable result is achieved.

The usefulness of GWO has been proven in a variety of optimization challenges. Liu *et al.* (2021) in [67] utilised the GWO to optimise the parameters of an SVM classifier. Their study demonstrated that the application of GWO resulted in enhanced classification accuracy when compared to alternative optimization algorithms. In a separate investigation conducted by Guezgouz *et al.*, [68], GWO was employed to optimise the allocation of renewable energy resources within a power system. This optimization approach resulted in improved overall system efficiency and cost reduction. The aforementioned examples demonstrate the multifaceted nature and wide-ranging utility of GWO in several fields.

Although GWO has demonstrated encouraging outcomes, it is important to acknowledge its inherent limits. One constraint that should be considered is the model's susceptibility to parameter adjustment, namely the hunting intensity coefficients and the beginning placements of the alpha, beta, and delta wolves. The efficacy of GWO is contingent upon the values assigned to its parameters, necessitating meticulous calibration to cater to diverse optimization challenges. One additional constraint is the possibility of premature convergence, wherein the search procedure becomes stagnant prematurely, thereby impeding any further investigation of the search domain. To address this concern, scholars have put forth potential approaches, including the implementation of dynamic population size and the utilization of adaptive parameter control.

#### **2.2.3.4 Whale Optimization Algorithm (WOA)**

WOA is a novel optimization method inspired by nature that was proposed by Mirjalili *et al.* in their work, work, [69]. The exploration and exploitation technique employed by WOA, which draws inspiration from the hunting behaviour of humpback whales,

has demonstrated significant potential for effectively addressing a range of optimization challenges. The algorithm emulates the bubble-net hunting behaviour observed in whales, wherein a group of potential solutions, represented as whale positions, undergoes repeated searching to find the most effective solution.

Within the framework of WOA, the algorithm effectively manages a population of cetaceans and systematically adjusts their spatial coordinates through the use of two primary methodologies: exploration and exploitation. The process of exploration involves the use of a random search equation to update the position of each whale, thereby facilitating a wide exploration of the search space. In contrast, exploitation is accomplished through the manipulation of the optimal solution's position by using a spiral equation, enhancing its effectiveness. The integration of these two tactics in the WOA methodology facilitates efficient exploration of the search space while simultaneously converging towards the most ideal solution.

WOA has also been used for the purpose of FS in [70], the study covers a wide array of benchmark datasets and compares the performance of this algorithm with GA, PSO, GWO and Ant Lion Optimizer (ALO). The results indicated the performance WOA is better than conventional algorithms, however, the algorithm suffers greatly when dealing with high dimensional datasets and there is still room for improvement.

### **2.2.3.5 Chimp Optimization Algorithm (ChOA)**

ChOA is a novel technique proposed by Khishe *et al.* in [71], this NIA is based on the hunting and sexual behaviours of chimpanzees (sometimes referred to as chimps) as shown in Figure 2.1. Chimps live in the form of fission-fusion groups merging and collapsing while covering a wide range of areas. Among these groups, the chimps are further divided into different groups namely the divers, chasers, barriers and attackers defending their territory against other groups. By modelling and simulating the hunting behaviour and learning strategies of chimps in the wild, it was proposed as a metaheuristic algorithm for optimization tasks in engineering, science, and other fields. ChOA proposed by Khishe *et al.* in [71] gains inspiration from the hunting preferences of chimps in a natural setting. The hunting behaviour of chimps can be categorised into two distinct phases: exploration and exploitation. The process of exploration involves

the act of attacking, blocking, and chasing the prey, which in turn leads to the discovery of a broader region inside the search space on a global scale. On the other hand, exploitation involves the act of attacking the prey, which provides possibilities for local search within the promising areas that were identified during the exploration phase.

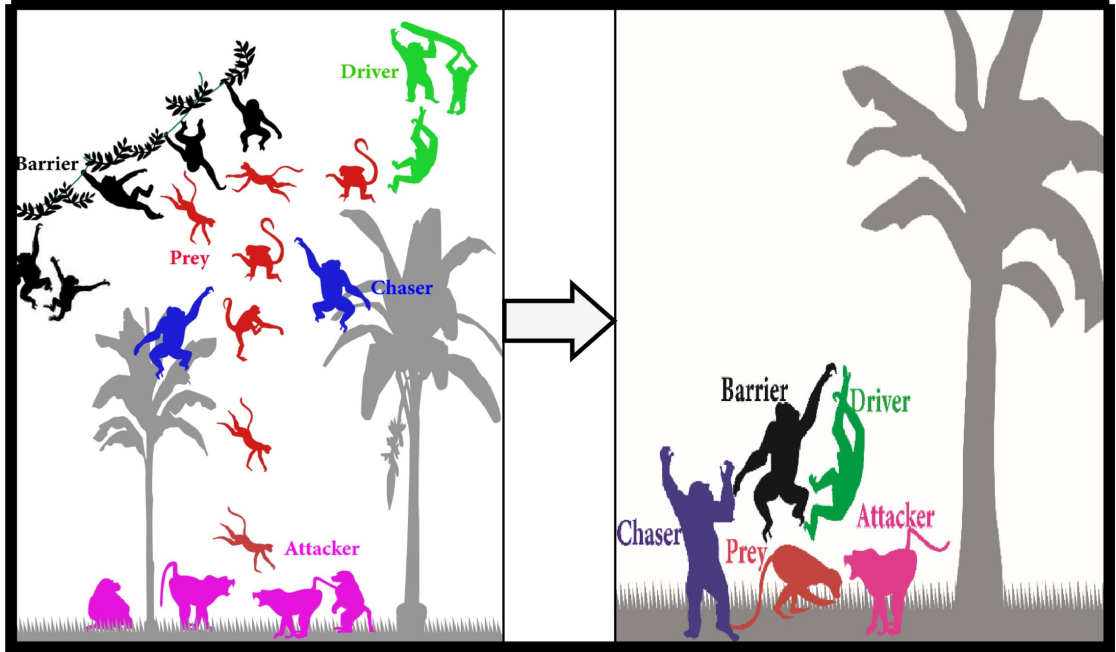


FIGURE 2.1: Attacking Strategy of Chimps[71]

To execute the sequential actions involved in hunting, a division of labour is observed among chimpanzees, consisting of four distinct roles: the driver, chaser, barrier, and attacker. Every chimp within the population serves as a candidate solution within the search space. The chimpanzees categorised as attacker, barrier, chaser, and driver represent the most optimal, second most optimal, third most optimal, and fourth most optimal solutions, respectively. At each iteration, the rest of the chimps are forced to update their position based on the location of the attacker ( $X_A$ ), barrier ( $X_B$ ), chaser ( $X_C$ ) and diver ( $X_D$ ) chimps. The updated position of chimps can be calculated as.

$$\begin{aligned}
 X_A(k+1) &= X_A(k) - A_1 \cdot D_A^n, \\
 X_B(k+1) &= X_B(k) - A_2 \cdot D_B^n, \\
 X_C(k+1) &= X_C(k) - A_3 \cdot D_C^n, \\
 X_D(k+1) &= X_D(k) - A_4 \cdot D_D^n,
 \end{aligned} \tag{2.3}$$

$$X_n(k+1) = \frac{X_A(k+1) + X_B(k+1) + X_C(k+1) + X_D(k+1)}{4}, \tag{2.4}$$

Here  $X_A(k), X_B(k), X_C(k), X_D(k)$  show the current position of the chimps in the search space, while  $X_A(k+1), X_B(k+1), X_C(k+1), X_D(k+1)$  show their next approximate move, the next position of the  $n^{th}$  chimp  $X_n(k)$  for  $k^{th}$  iteration is calculated using (2.4).  $D_A^n, D_B^n, D_C^n$  and  $D_D^n$  are the distance of the attacker, barrier, chaser and diver from the  $n^{th}$  chimp for  $k^{th}$  iteration and  $A_1, A_2, A_3$  and  $A_4$  are random vectors. The values of  $D_A^n, D_B^n, D_C^n$  and  $D_D^n$  can be calculated as follows.

$$\begin{aligned} D_A^n &= |C_1 \cdot X_n - M \cdot X_A(k)|, \\ D_B^n &= |C_2 \cdot X_n - M \cdot X_B(k)|, \\ D_C^n &= |C_3 \cdot X_n - M \cdot X_C(k)|, \\ D_D^n &= |C_4 \cdot X_n - M \cdot X_D(k)|, \end{aligned} \tag{2.5}$$

$A_1, A_2, A_3, A_4$ , and  $C_1, C_2, C_3, C_4$  are the random coefficient vectors calculated by using the following equations.

$$\begin{aligned} A_1 &= (2r_{11} - 1)f, \\ A_2 &= (2r_{12} - 1)f, \\ A_3 &= (2r_{13} - 1)f, \\ A_4 &= (2r_{14} - 1)f, \end{aligned} \tag{2.6}$$

$$\begin{aligned} C_1 &= 2r_{21}, \\ C_2 &= 2r_{22}, \\ C_3 &= 2r_{23}, \\ C_4 &= 2r_{24}, \end{aligned} \tag{2.7}$$

$$M = \text{chaotic\_value}, \tag{2.8}$$

Here,  $r_{ij}$  ( $i, j = [1, 2, 3, 4]$ ) are random values in the range  $[0, 1]$  based on uniform distribution and  $f$  decreases non-linearly from 2.5 to 0 throughout the process.  $M$  is a random chaotic value and could be generated from Gaussian or Bernoulli distribution. The complete position update rule for each chimp  $n$  for the  $k^{th}$  iteration is given by (2.9). The variable  $M$  represents the impact of sexual incentives on the behaviour of chimpanzees during the concluding phases of hunting. This motivation leads to unpredictable behaviour, wherein the chimps relinquish their hunting duties and make desperate efforts to get prey. To update the position of the chimpanzees during the optimization process, a probability of 50% is employed to randomly choose between the chaotic model and the conventional part updating procedure. The subsequent equation represents the

theoretical framework:

$$X_n(k) = \begin{cases} \frac{X_A + X_B + X_C + X_D}{4}, & \text{if } \phi < 0.5 \\ M, & \text{if } \phi > 0.5 \end{cases} \quad (2.9)$$

The procedure is initialised by randomly generating a population, then calculating the fitness value for these populations and equating the top four chimps as the attacker ( $X_A$ ), barrier ( $X_B$ ), chaser ( $X_C$ ), and driver ( $X_D$ ). The algorithm iterates through the following steps until it satisfies a specified termination requirement. Initially, the fitness function is employed to assess each solution inside the population. Afterwards, the system proceeds to update the positions and scores of the attacker, barrier, chaser, and driver. Next, the method proceeds to update the values of the  $F$ ,  $r_{ij}$ , and  $M$  coefficients. Thirdly, the determination of the primary coefficients of  $A$  and  $C$  is achieved by employing the values of the  $r_{ij}$ , and  $F$  parameters, as specified in equations (2.6) and (2.7). The chimps' positions are updated using equations (2.3), (2.7), and (2.9) in the last step. Consequently, the optimal resolution, specifically the location of the prey, is retrieved. Algorithm 1 describes the pseudo-code for the ChOA. ChOA has been used for FS process; In [72] a binary chimp optimization algorithm (BChOA) is developed, which uses the ChOA to address the task of feature selection. This work examines the use of two transfer functions, namely the S-shaped and V-shaped functions, to convert the continuous form of ChOA into a binary representation. The technique that has been created is subsequently subjected to a comparative analysis with established algorithms such as GA, PSO, BA, and ACO.

The algorithm, however, is accompanied by some obstacles and restrictions that impede its effectiveness when used for large datasets. The agent's ability to explore the search space is constrained by the mobility restrictions imposed by the Gaussian randomization process. Furthermore, as a result of the intricate nature of the algorithm, the agents become trapped in local optima and are unable to adequately traverse the search space. Consequently, this leads to inferior solutions when dealing with diverse situations. In addition to these problems, the chimp update rule as shown by (2.4) shows that the new chimp location is updated using a simple average that could lead to divergent behaviour as  $X_A$  is the chimp with the best fitness while the others have a lower fitness value, hence there is room for improvement in this regard. Considering these problems, the ChOA exhibits increased vulnerability to premature convergence and the local minima problem, which leaves room for improvement.

---

**Algorithm 1** Steps for ChOA

---

Initialise population size  $N$ , dimension  $d$ , max iterations  $k_{\max}$ .  
Set  $k = 0$ , and generate random population  $X_n (n = 1, 2 \dots N)$ .  
Initialise fitness values for population.  
**while**  $k \leq k_{\max}$  **do**  
  **for**  $n = 1 : N$  **do**  
    Define the group strategy.  
    Update agents using group strategy.  
  **end for**  
  **for**  $j = 1 : d$  **do**  
    **if**  $j < 1$  **then**  
      Update current position.  
    **else if**  $j > 1$  **then**  
      Select random chimp.  
    **end if**  
  **end for**  
  Update Attacker, Barrier, Driver and Chaser.  
   $k = k + 1$ .  
**end while**  
**return** *Global Best*

---

## 2.2.4 Mathematics Based Algorithms

MHAs that are based on mathematics employ mathematical models and methodologies to address intricate optimization problems, which encompass the task of feature selection. The Arithmetic Optimization Algorithm (AOA) and the Sine Cosine Algorithm (SCA) are two often employed metaheuristic algorithms in the field of mathematics [73]. The article by Sindhu [74] assessed the efficacy of SCA as a method for feature selection in face recognition and handwritten digit recognition data sets.

### 2.2.4.1 Sine-Cosine Algorithm (SCA)

SCA is a stochastic optimization technique that utilises sine and cosine mathematical functions. It is designed to solve optimization issues in a population-based manner. The search process in SCA is initiated by employing a random population. In each subsequent generation, this agent set is modified using mathematical expressions that rely on sine and cosine functions. The position of each individual is evaluated by the

formulae shown below [75].

$$\zeta_i^{k+1} = \begin{cases} \zeta_n^k + (r_1 * \sin r_2 * |x_n^k - r_3 \zeta_n^k|), & \text{if } r_4 < 0.5 \\ \zeta_n^k + (r_1 * \cos r_2 * |x_n^k - r_3 \zeta_n^k|), & \text{if } r_4 \geq 0.5 \end{cases} \quad (2.10)$$

The variable  $\zeta_n^k$  represents the  $n^{\text{th}}$  agent's position for  $k^{\text{th}}$  iteration while  $\zeta_n^{k+1}$  shows the next position of this agent. Additionally,  $r_1$  is a deterministic parameter, while  $r_2$ ,  $r_3$ ,  $r_4$  are evenly distributed random numbers that influence the exploitation and exploration of the search space. Lastly,  $x_n^k$  denotes the position of the best agent in the  $n^{\text{th}}$  dimension after  $k^{\text{th}}$  iterations.

The deterministic variable  $r_1$  determines the magnitude of each step and can be in the space between the current and best solution or outside this space as shown in Figure 2.2. The value of  $r_1$  is determined by the following equation. The variable  $r_2$  is a random number on the range  $[0, 2\pi]$  and determines the direction and margin of movement.  $r_3$  is a weight determined randomly to find the best solution.  $r_3$  is defined by  $r_3 = a * w$  where  $a$  is a constant number and  $w$  is a random number in the range  $[0, 1]$ .

$$r_1 = f - f \left( \frac{k}{k_{\max}} \right), \quad (2.11)$$

where  $f$  is a constant  $k$  is the current iteration and  $k_{\max}$  is the total number of iterations.

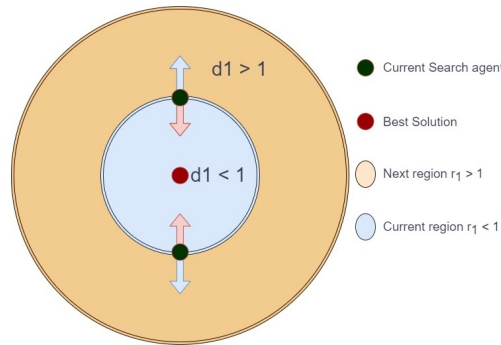


FIGURE 2.2: Particle Movement in SCA [75]

Finally, the variable  $r_4$  is a Gaussian distributed random number which switches the algorithm between sine and cosine functions. The variables  $r_1$ ,  $r_2$ ,  $r_3$  and  $r_4$  are modified at each iteration to enhance the diversity of the solutions. By modifying the ranges of the sine and cosine functions, the agents can explore regions outside the interval between



their respective targets. Furthermore, this prompts an investigation into the search space. Figure 2.2 illustrates how the aforementioned equation delineates the spatial relationship between two individuals inside the search space. Algorithm 2 describes the pseudo-code of SCA.

---

**Algorithm 2** Steps for SCA

---

```

Initiate population  $N$ , max iterations  $k_{\max}$ .
Set  $k = 0, \rho = 0$ .
while  $k \leq k_{\max}$  do
    Evaluate parameter  $\rho$ .
    for each agent do
        Update and search using Sine and Cosine functions.
    end for
    Update position of current agents.
     $k = k + 1$ .
end while

```

---

### 2.2.5 Variants for Metaheuristic Algorithms

This part of the literature review demonstrates a wide range of applications in which hybrid, enhanced, and multi-objective variations of metaheuristic algorithms have been implemented.

#### 2.2.5.1 Hybrid MHAs

The FSAs proposed by Yu *et al.* utilise a hybrid approach that combines PSO and GA in [34]. The PSO technique is employed to build an initial population of solutions, whereas the GA is utilised to execute the selection and crossover procedures. The technique underwent testing on multiple benchmark datasets and showed superior performance compared to various other feature selection methods. A comparable dataset was also introduced in Stephan *et al.* in [35], where a hybrid method combining the Artificial Bee Colony (ABC) Optimization algorithm and the WOA was employed. In the study conducted by Wan *et al.* [76], two models were introduced that combined the techniques of Ant Colony Optimization (ACO) and Genetic Algorithm (GA). These models were then evaluated against the current state-of-the-art methods using various datasets derived from the UCI repository. A hybridised approach combining the GA and Cuckoo search techniques was proposed in a study conducted by Jona *et al.* [77].

This approach utilised a hybridised wrapper-based method for addressing the feature selection task. The authors Moradi *et al.* [78] introduced a local exploration model in their study, which aimed to assist PSO in the selection of marginal deducts based on their correlation data. The authors of the study investigated the utilization of robotic manipulators. They enhanced a modified PSO algorithm by using Differential Evolution (DE) techniques. The objective of this research was to present a comprehensive analysis of various types of manipulators, including serial and parallel manipulators, as well as the 10-DOF hybrid redundant serial-parallel robots. The work was documented by Mao *et al.* [36] in 2017. The reactive power problem was addressed by Lenin *et al.* [38] by the utilization of a hybrid approach that included TS and SA algorithms. The approach was also utilised to solve a symmetrical instance of the travelling salesman issue. Mafarja *et al.* [79] introduced a hybrid feature selection strategy that combines the use of SA and GA in their study. The algorithm has shown favourable performance when evaluated on the UCI dataset, specifically in terms of the number of selected attributes, when compared to existing state-of-the-art methodologies.

#### 2.2.5.2 Improved MHAs

The literature shows various methods for enhancing MHAs, to mitigate the issue of premature convergence. One potential approach is to dynamically modify the random step size or search range in response to the algorithm's progress. The utilization of this adaptive mechanism guarantees the preservation of a balanced exploration-exploitation equilibrium during the entirety of the optimization process. The paper by Stützle and Hoos [80] presents the MAX-MIN Ant System (MMAS), an ACO algorithm that is derived from the Ant System. This algorithm demonstrates improved performance compared to its predecessor on well-established benchmark problems, including the Traveling Salesman Problem and the Quadratic Assignment Problem. In their study, Cai *et al.* [81] introduced an Adaptive Particle Swarm Optimization (APSO) algorithm. This algorithm incorporates a stability criterion to dynamically adjust the parameters of the PSO algorithm. The aim is to achieve optimal convergence while avoiding premature convergence. Through simulations conducted on established problems, the authors demonstrate the superior performance of APSO compared to the conventional PSO algorithm. The techniques for addressing challenging problems using Fuzzy Adaptive Simulated Annealing (Fuzzy ASA) and regular ASA are presented by the author in [82]. These

algorithms are advanced global optimization methods that are rooted in the simulated annealing paradigm. The author provides comprehensive illustrations of their practical implementation and emphasises their minimal mathematical prerequisites. The research conducted by Cheung *et al.* [83] introduced an adaptive firefly algorithm (AdaFa) that integrates three methodologies for parameter selection and adaptation. The paper showcases the enhanced performance of AdaFa on benchmark functions and a real-world scenario. The article by Mahdavi *et al.* [84] introduces an enhanced harmony search algorithm that incorporates a novel approach for generating solution vectors and a strategy for modifying its parameters. The authors demonstrate the algorithm's efficacy by successfully addressing a range of benchmarking and conventional engineering optimization tasks. Deterministic chaos maps, such as the Logistic map, Chebyshev map, and Tent map, are mathematical functions that generate a succession of seemingly random integers. The generation of unique and unpredictable patterns in these maps is contingent upon a starting state. The utilization of non-periodic, non-converging, and random-like outcomes makes them a viable alternative to random sequence generators within MHAs. The rising popularity of chaos-based meta-heuristics can be attributed to its notable convergence rate and heightened level of randomness. The methodology presented in the work by Hefny *et al.* [85] introduces a novel approach that integrates chaotic agents with PSO to improve the effectiveness and accuracy of search algorithms. This approach specifically aims to overcome the limitations of premature convergence and parameter sensitivity that are inherent in PSO. In the study [86] a GA is proposed that incorporates a chaotic optimization operation. This approach employs the logistic function as a crossover operator, resulting in enhanced computational efficiency when compared to conventional genetic algorithms for the optimization of multimodal functions. [87] puts forth six distinct algorithms for chaotic league championship (LCA) and conducts an analysis of their efficacy in intricate benchmark functions. The findings of this investigation demonstrate encouraging outcomes, indicating that the integration of LCA with chaotic maps has the potential to enhance optimization quality in select problem domains.

The Gaussian distribution is commonly employed as a refining tool in a variety of algorithms. In addition, certain methodologies involve the utilization of sampling from a Gaussian distribution as a substitute for conventional solution updating techniques.

In the study conducted by [88], an investigation is carried out to analyse the underlying characteristics that contribute to the effectiveness of the PSO. This is achieved by removing certain conventional components and modifying the algorithm through the elimination of the velocity formula. Consequently, the study reveals the algorithm's resemblance to other problem-solving methods based on stochastic population dynamics, thereby proposing novel directions for further research [88]. In their study, Wang *et al.* [89] introduces Gaussian bare-bones differential evolution (GBDE), an optimization algorithm that requires minimal parameter tuning. They also present a modified version of GBDE, referred to as MGBDE. The authors demonstrate that both GBDE and MGBDE exhibit superior or comparable performance to existing state-of-the-art differential evolution (DE) variants when applied to benchmark functions and real-world problems. In a scholarly publication by Zou *et al.* a novel iteration of the teaching-learning-based optimization algorithm named Bare Bones Teaching Learning Based Optimization (BBTLBO) was introduced [90]. This variant integrates the learning approach of a teacher, Gaussian sampling, and neighbourhood search. The study evaluates the efficacy of BBTLBO in comparison to TLBO and other rudimentary algorithms using both benchmark functions and real-world problems.

### 2.2.5.3 Multi-objective Algorithms

Single-objective meta-heuristics are not well-suited for addressing optimization situations characterised by numerous conflicting objectives. Hence, the emergence of the notion of multi-objective optimization has occurred, to identify a collection of potential solutions that do not exhibit dominance over one another. The methodology commonly known as "Pareto optimization" or "vector optimization" is frequently employed to optimise many criteria concurrently while adhering to a predetermined set of restrictions. Numerous academics have put forth novel solution representations to tackle multi-objective optimization issues by employing selection processes based on Pareto dominance. In their study, Guo *et al.* [91] proposed a chromosome-based approach to address a multi-objective order scheduling problem. In the proposed methodology, it is established that every gene present on the chromosome corresponds to a specific order group. The value assigned to each gene indicates the plant that is responsible for processing the respective group. In their study, Xiao *et al.* [92] proposed a hybrid encoding technique that integrates both indirect and direct representations to enhance the

optimization of air traffic network flow. In their methodology, the authors utilised a direct encoding strategy for representing departure time slots and an indirect encoding approach for representing aircraft routes. Wang and Ma [93] proposed a chromosomal representation based on sets to tackle a multi-objective problem related to placement and allocation. The provided depiction successfully encompasses pertinent details about both the decision-making process for facility location and the various types of facilities involved.

Particle updating and population reproduction are essential elements of search algorithms based on PSO. Several researchers, such as Zhang *et al.* [94], have put forward multiple modifications aimed at improving the performance of PSO in a range of optimization tasks. An instance of a modified PSO algorithm was introduced, specifically designed to effectively address complex nonlinear resource allocation problems (RAPs). The proposed modification integrates a particle updating mechanism that does not require any parameters with an adaptive grid-based approach to efficiently update the global particle leaders. In a separate study conducted by Zhang *et al.* [95], a sorting method for scheduling problems was suggested. This mechanism involves the selection of solutions from personal optimal solutions to construct the global optimal solution set during each iteration. In addition, the researchers devised a dual population evolution mechanism and implemented a crowding entropy-based hierarchical elitism preservation approach within the PSO algorithm. These modifications were aimed at attaining a diverse and evenly distributed collection of Pareto optimal solutions. These novel adaptations have demonstrated the potential for enhancing the overall efficacy of PSO in addressing diverse optimization issues.

The research conducted by Qian *et al.* [96] introduces HDE, a hybrid algorithm that combines differential evolution with a largest-order-value rule to transform continuous values into job permutations. Additionally, the algorithm incorporates problem-specific local search techniques to prioritise exploitation. The objective of this algorithm is to address the multi-objective permutation flow shop scheduling problem, taking into account the presence of limited buffers. In their study, Burke *et al.* introduce a memetic approach to enhance the robustness objectives of airline schedules through the concurrent optimization of flight re-timing and aircraft rerouting. This approach is applied to actual schedules obtained from KLM Royal Dutch Airlines, leading to notable enhancements in the targeted objectives [97]. In the study conducted by Khalili *et al.* [98] a

novel approach is presented that combines a multi-objective electromagnetism algorithm (MOEM) with simulated annealing. The purpose of this approach is to address a challenging bi-objective flow-shop scheduling problem that involves transportation times. The authors demonstrate the efficacy of their suggested method through the use of experimental examples and performance measurements.

### 2.2.6 Limitations of MHAs

MHAs have become increasingly prominent in the realm of optimization, particularly in the context of feature selection within the field of machine learning. Nevertheless, despite their achievements, these algorithms possess certain constraints that can impact their efficacy. One limitation is that the efficacy of metaheuristic algorithms is contingent upon the initial solution and stopping condition, hence impeding the comparability of outcomes across various metaheuristic algorithms. In addition, it is worth noting that the efficacy of metaheuristic algorithms can be influenced by the calibre of the initial solution, potentially resulting in outcomes that are less than optimal.

The selection of a stopping criterion holds significant importance in metaheuristic algorithms due to its impact on the efficiency and convergence of the algorithm [99]. Nevertheless, it is important to note that there is no universally optimal stopping criterion applicable to all situations. The selection of a stopping criterion can have a substantial impact on the algorithm's performance. One commonly employed strategy involves setting a maximum number of iterations or a threshold for the fitness function. Nevertheless, this approach may not always yield satisfactory results, especially in intricate issues characterised by an expansive search area. One further constraint of metaheuristic algorithms pertains to the challenge of ascertaining the most optimal parameters for the algorithm [100]. In the context of the GA, pertinent parameters encompass the population size, mutation rate, crossover rate, and selection operator. The selection of these parameters has a substantial impact on the algorithm's performance, and determining the most suitable values for these parameters can pose a considerable challenge. Hence, it is crucial to optimise the parameters to attain optimal performance of the algorithm. Nevertheless, the selection of suitable parameter values for metaheuristic algorithms poses a formidable challenge, necessitating a profound understanding of the domain and a comprehensive process of trial-and-error testing.

The efficacy of metaheuristic algorithms can also be influenced by the intricacy of the task at hand. As the complexity of the problem rises, the search space expands, hence posing a greater challenge for the algorithm to identify the most effective solution [101]. This phenomenon can result in increased execution duration and an elevated likelihood of becoming trapped in local optima. In the context of feature selection problems, it is worth noting that the number of potential feature subsets might exhibit exponential growth, resulting in a search space that is both extensive and intricate. In addition, the efficacy of metaheuristic algorithms might be influenced by the existence of noise within the dataset [102]. In the context of feature selection challenges, the inclusion of noisy features has the potential to diminish the accuracy of the algorithm and compromise the overall quality of the selected features. This phenomenon occurs due to the algorithm's tendency to perceive noisy features as significant and include them in the selection process, hence resulting in less-than-ideal outcomes.

Premature convergence represents an additional obstacle encountered in the realm of metaheuristic algorithms. Premature convergence is a phenomenon that arises when an algorithm reaches a sub-optimal solution prematurely, hence impeding its ability to explore alternative regions within the search space. This phenomenon may arise in situations where the population size is quite small or when the search operator lacks sufficient diversity. Numerous scholarly investigations have put forth various approaches to address this constraint, including the integration of a diversity measure within the algorithm or the use of multiple beginning points. One further constraint is the scalability of metaheuristic algorithms. According to Liu *et al.* [103], when the problem's dimensionality increases, the search space experiences exponential growth, hence posing greater challenges for the algorithm in locating the ideal solution. This phenomenon can result in increased execution duration, perhaps rendering the algorithm incapable of identifying an optimal solution within a feasible time frame. Hence, the development of efficient algorithms capable of effectively addressing high-dimensional issues holds significant importance. The selection of the goal function can have an impact on the efficacy of metaheuristic algorithms. The quality of the solution and the performance of the algorithm can be considerably influenced by the objective function. Nevertheless, the task of choosing a suitable objective function might pose difficulties, especially when dealing with intricate situations [104]. Hence, it is crucial to meticulously select the objective function and assess its influence on the algorithm's performance.

## 2.3 Research Gap

This literature survey indicates several key areas where the conventional and hybrid MHAs performance have room for improvement particularly when dealing with unbalanced high dimensional datasets. Typically bio-medical datasets are of such nature and considering the importance of these datasets this area is of particular importance. The algorithms discussed in the literature survey namely the ChOA, GWO and WOA have been used for FS, however there are some aspects that have not been examined. GWO has not been tested for bio-medical datasets, while WOA suffers from pre-mature convergence and local minimal problems. PSO on the other hand has a sluggish response when used for high-dimensional datasets. ChOA has improved performance when compared to PSO, GWO and WOA, however, it suffers from premature convergence. There is room for improvement in terms of accuracy and reduction ratio of selected features.

## 2.4 Problem Statement

The significance of feature selection in the biomedical domain becomes apparent due to its crucial role in improving the efficacy and comprehensibility of predictive models, decreasing computational intricacy, and facilitating the identification of significant biomarkers. Nevertheless, the primary obstacle involves developing an optimization methodology that proficiently directs the choice of pertinent characteristics from biomedical datasets with a large number of dimensions. This methodology must guarantee that the chosen characteristics substantially enhance model accuracy, comprehensibility, and the detection of crucial biological understandings. Additionally, it should tackle challenges such as overfitting, computational load, and ethical concerns about data privacy and patient confidentiality. This research investigates the efficacy of employing hybrid variants of MHAs for the purpose of feature selection, aiming to decrease the dimensionality of large-scale datasets by eliminating unnecessary data. The project aims to review several FSAs' for dimensionality reduction, investigate the capabilities of metaheuristic algorithms, and evaluate their applicability in solving the feature selection problem across datasets with different sizes and multiple classes. The objective of the proposed solution is to surpass the performance of traditional, conventional, and enhanced versions of metaheuristic techniques. This will be accomplished by employing a



novel hybrid approach and a technique for minimising dataset variance. The ultimate goal is to achieve a globally optimal solution within a shorter time frame, while also preserving the relationship between the attributes and the output classes in the dataset.

## 2.5 Thesis Objectives

The key objectives of this thesis are listed below:

1. The optimization of storage and processing of large data by the reduction and elimination of redundant information.
2. Develop a hybrid metaheuristic strategy as a means to overcome the limitations associated with conventional feature selection methods.
3. Develop a mathematical model for the hybrid algorithm that successfully incorporates the exploration and exploitation phases of metaheuristic algorithms.
4. Conduct an in-depth study of feature selection techniques to identify suitable benchmarks and do comparative analysis.

## 2.6 Research Methodology

As discussed in the section 2.4, the objective of this work is to develop a novel hybrid meta-heuristic algorithm that combines the strengths of the SCA and BChOA methods. To achieve the intended objective, a mathematical model is formulated to represent the novel technique. The proposed algorithm, referred to as the Binary Sine-Cosine Chimp Optimization Algorithm (BSChOA), is then evaluated using high dimensional and unbalanced bio-medical datasets, as well as other benchmark datasets from the UCI repository. The study evaluates the efficacy of the suggested technique in improving the performance of Feature Selection (FS) algorithms. The comparison is conducted against several existing algorithms, namely BChOA, BGWO, WOA, GA, PSO, and SCA. The evaluation is based on various metrics such as accuracy, precision, F1-score, train time, and the number of selected features to determine the effectiveness of the proposed algorithm for the purpose of FS.

## Chapter 3

# Binary Sine-Cosine Chimp Optimization Algorithm (BS-ChOA)

In this section, we discuss the mathematical model for the proposed hybrid MHA. The proposed BS-ChOA is developed from the ChOA hybridized with the mathematics-based SCA to make use of the best abilities of both algorithms during the exploration and exploitation phases. The base model is derived from the ChOA, as discussed in Section 2.2.3.5, during the exploration phase, we make use of the SCA. The algorithm uses the sine and cosine functions to circle its way towards an optimum solution.

To begin, an initial chimp population matrix ( $Ch$ ) is established, consisting of randomly assigned values inside a defined search space. This matrix represents the initial positional vectors of all chimps and the prey.

$$Ch = \begin{bmatrix} Ch_1 \\ Ch_2 \\ \vdots \\ Ch_N \end{bmatrix} = \begin{bmatrix} Ch_1^1 & Ch_1^2 & \dots & Ch_1^d \\ Ch_2^1 & Ch_2^2 & \dots & Ch_2^d \\ \vdots & \vdots & \dots & \vdots \\ Ch_N^1 & Ch_N^2 & \dots & Ch_N^d \end{bmatrix} \quad (3.1)$$

Here  $Ch_n$  ( $n = 1, 2, 3 \dots N$ ) represents  $N$  candidate solutions in the search space with  $d$  dimensions. The chimps  $Ch_n$  are initialised randomly and then evaluated using a defined cost function. They are then categorised with the best four solutions categorised

as *attacker*, *barrier*, *chaser* and *diver* and the rest of the chimp population is updated during the iterative process. After each iteration, the chimps are re-categorised and the best four are selected accordingly. The graphical representation of the initial search area is depicted in Figure 3.1. The chimps initiate their hunting behaviour by strategically placing the barriers and the chasers to ensnare their prey, Figure 3.2 shows the movement matrix of each chimp within the search space. The algorithm is implemented for a predetermined number of iterations, as specified by the user at the beginning. During each iteration, the chimp position update technique is performed about the goal fitness function. To augment the exploitation capabilities of the ChOA, a more targeted strategy is employed by integrating a revised SCA into the ChOA framework. The following steps are used for the mathematical formulation of the proposed algorithm.

1. **Initialization:** A random population  $X_i$  with  $n$  samples and  $i$  dimensions is initialised.
2. **Evaluation:** Each vector of the population is evaluated using a predefined fitness function and the results are arranged from best to least, with the top four vectors being categorised as the attacker, barrier, chaser and diver.
3. **Exploration:** In this scenario, chimps modify their location in order to locate the closest possible position of their next prey. With the assistance of this update, chimps may potentially become trapped in optimal solutions inside the complex space, at least over multiple iterations.

$$X_A(k+1) = \begin{cases} X_A(k) - \cos(r_2) * A_1 * D_A^n, & \text{if } r_3 < 0.5 \\ X_A(k) - \sin(r_2) * A_1 * D_A^n, & \text{if } r_3 \geq 0.5 \end{cases} \quad (3.2)$$

$$X_B(k+1) = \begin{cases} X_B(k) - \cos(r_2) * A_1 * D_B^n, & \text{if } r_3 < 0.5 \\ X_B(k) - \sin(r_2) * A_1 * D_B^n, & \text{if } r_3 \geq 0.5 \end{cases} \quad (3.3)$$

$$X_C(k+1) = \begin{cases} X_C(k) - \cos(r_2) * A_1 * D_C^n, & \text{if } r_3 < 0.5 \\ X_C(k) - \sin(r_2) * A_1 * D_C^n, & \text{if } r_3 \geq 0.5 \end{cases} \quad (3.4)$$

$$X_D(k+1) = \begin{cases} X_D(k) - \cos(r_2) * A_1 * D_D^n, & \text{if } r_3 < 0.5 \\ X_D(k) - \sin(r_2) * A_1 * D_D^n, & \text{if } r_3 \geq 0.5 \end{cases} \quad (3.5)$$

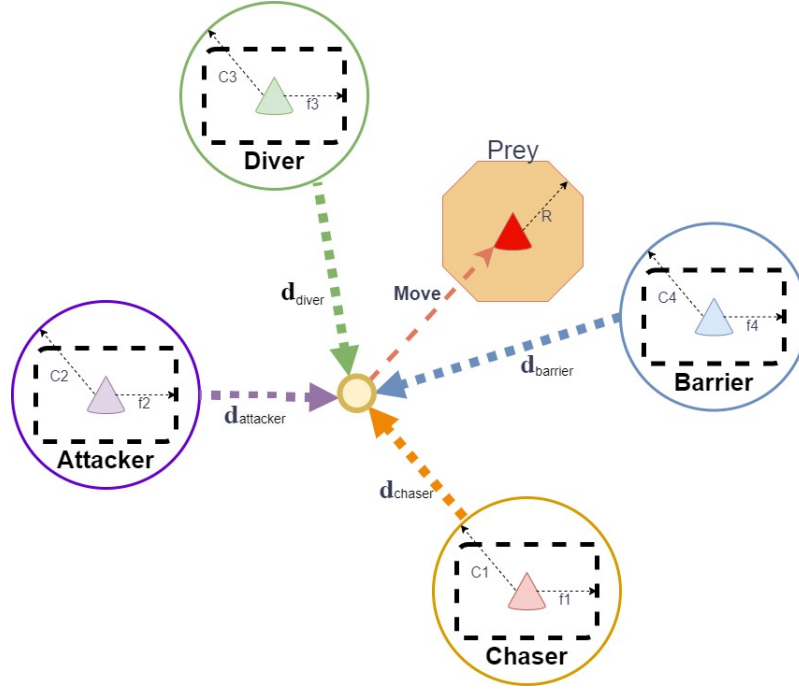


FIGURE 3.1: Chimp Movement in the search space

Here  $r_2 = (2\pi) * rand()$  and lies in the range  $[0, 2\pi]$ , the variables  $X_A(k)$ ,  $X_B(k)$ ,  $X_C(k)$  and  $X_D(k)$  shows the current position of the top four chimps.  $D_A^n$ ,  $D_B^n$ ,  $D_C^n$  and  $D_D^n$  show the distance of the  $n^{th}$  chimp from the *attacker*, *barrier*, *chaser* and *diver* calculated using (2.5).  $X_A(k+1)$ ,  $X_B(k+1)$ ,  $X_C(k+1)$  and  $X_D(k+1)$  shows the updated position of these chimps with respect to the  $n^{th}$  chimp.  $A_1, A_2, A_3$  and  $A_4$ , and  $C_1, C_2, C_3$  and  $C_4$  are the random coefficient vectors calculated by using (2.6) and (2.7). Traditionally the update rule of chimp position involves a simple average as shown in (2.4) in Section 2.2.3.5, however, this would mean that all four top-ranking chimps have an equal pull on the chimp to be updated.

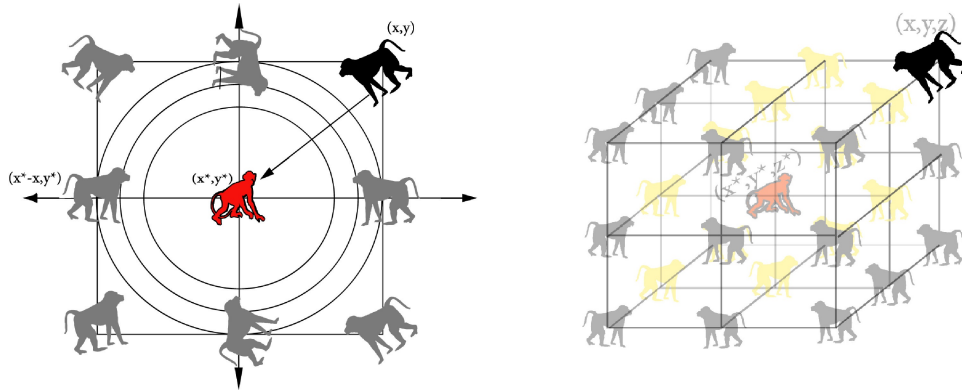


FIGURE 3.2: a) 2D Movement matrix of Chimps b) 3D Movement matrix of Chimps [71].

To solve this issue a weighted average is suggested to update the position and leverage the update in favour of the chimp with the best fitness namely the attacker in this case as shown by (3.6).

$$X_n(k+1) = b_1 * X_A(k+1) + b_2 * X_B(k+1) + b_3 * X_C(k+1) + b_4 * X_D(k+1), \quad (3.6)$$

$$\sum_{i=1}^4 b_i = 1, \quad (3.7)$$

The variable  $X_n (i = 1, 2, 3...N)$  shows the chimp population and  $b_i (i = 1, 2, 3, 4)$  are coefficients for the weighted average and are determined during the experiment keeping in mind (3.7).

4. **Stopping Condition:** Subsequently, stop criteria were set up to govern the search process inside the search domain, with the objective of attaining the most ideal value. The aforementioned criteria were employed to ascertain the selection procedure for all community search agents and thereafter assign them positions based on the performance of the most efficient search agent.

The process is iterative and is repeated until it satisfies the specified stopping condition, such as reaching the maximum allowable number of iterations or achieving an expedited solution. The pseudo-code for the BS-ChOA is shown in 3.

---

**Algorithm 3** Steps for ChOA

---

Initialise population size  $N$ , dimension  $d$ ,  $k_{\max}$ .  
Set  $k = 0$ , and generate random population  $X_n (n = 1, 2...N)$ .  
Initialise fitness values for population.  
**while**  $k <= k_{\max}$  **do**  
  **for**  $n = 1$  to  $N$  **do**  
    Define the group strategy.  
    **for**  $j = 1$  to  $d$  **do**  
      Calculate distance from  $X_A(k)$ ,  $X_B(k)$ ,  $X_C(k)$ ,  $X_D(k)$ .  
      Evaluate the next position of chimps w.r.t *attacker*, *barrier*, *chaser*, *diver*.  
      Update chimp position.  
    **end for**  
    Update Attacker, Barrier, Driver and Chaser.  
  **end for**  
   $k = k + 1$ .  
**end while**  
**return** Global best

---

### 3.1 Feature Selection Using BS-ChOA

Feature selection is a complex optimization problem. In order to determine an optimally reduced feature set a binary version of the proposed MHA is obtained by using complex transformation functions. For the purpose of this research, two functions have been used, namely the s- and v-functions [71]. Equations (3.8) and (3.9) show the mathematical representation of the transformation function.

$$TF_s(n, j) = \frac{2}{1 + e^{-2 * X[n, j]}} - 1, \quad (3.8)$$

$$TF_v(n, j) = \frac{X[n, j]}{\sqrt{1 + X[n, j]^2}} - 1, \quad (3.9)$$

Here  $X[n, j]$  represents the distance of a  $j^{th}$  feature point for a  $n^{th}$  chimp from the prey location as determined by (3.6). At each iteration, this distance is evaluated, and a transformation rule is applied to determine whether to select a particular feature or mask its value. At any point, a random mutation can also occur based on a random value that can alter the state of the feature point. Figure 3.3 shows a graphical representation of the BS-ChOA for the purpose of feature selection. The following steps are followed to execute the FS process using BS-ChOA:

- **Step 1:** The initial dataset is employed to generate a population of chimps. Each individual chimpanzee inside the collective is considered a potential candidate for a feature subset.
- **Step 2:** The BS-ChOA is executed and the population is assessed by the utilization of a fitness function by utilising a classifier.
- **Step 3:** Following the assignment of fitness values to each solution, the population undergoes a selection process where the four solutions with the highest fitness values are chosen.
- **Step 4:** The algorithm will update the primary coefficients f, m, C, A, and D. Subsequently, the other chimps adjust their respective positions based on the optimal position of the leading chimp.
- **Step 5:** Transformation Functions (TFs) are employed to assess the likelihood of modifying the components of position vectors and constraining their range of motion within the interval of 0 to 1.
- **Step 6:** The algorithm is repeated until a set stopping condition is met and a reduced FV is returned.

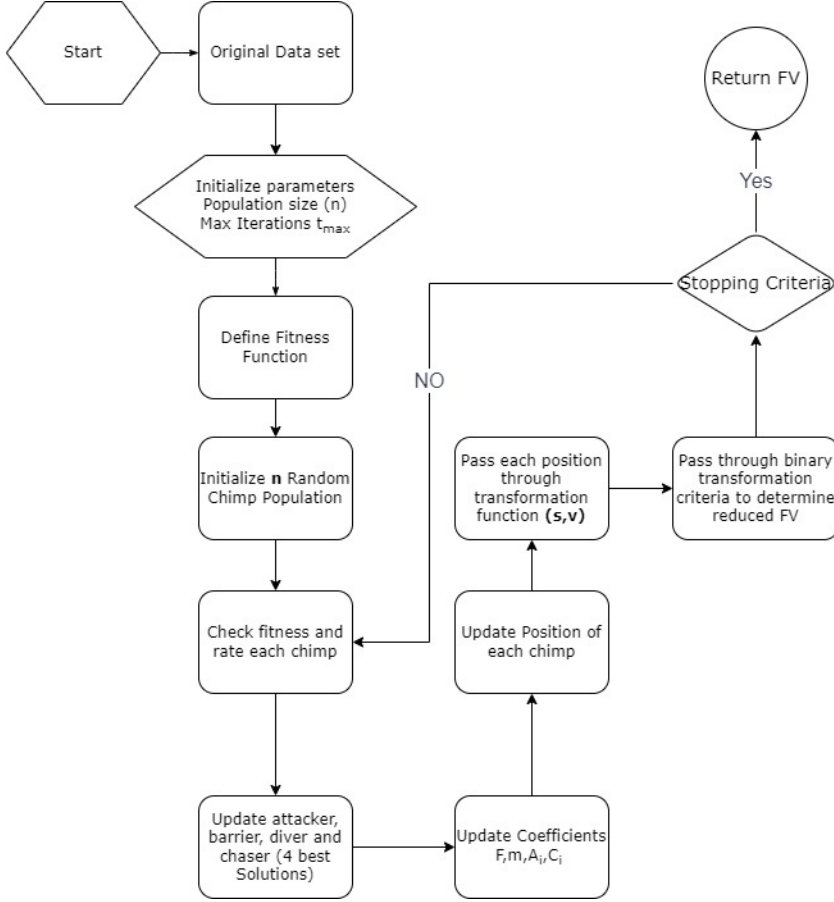


FIGURE 3.3: Flow Diagram: Feature Selection Using BS-ChOA

Traditionally the fitness function  $J$  for variance-based feature selection is evaluated using (3.10).

$$J = \alpha * \mathbf{acc}(X_n) + (1 - \alpha) * \frac{D - M}{D}, \quad (3.10)$$

where  $X_n$  is the reduced feature set and  $\alpha$  is a set coefficient in the range  $[0, 1]$   $D$  is the total number of features, and  $M$  is the number of selected features. A trade-off is set between the selection of hyper-parameter  $\alpha$  between the accuracy of the desired result and the reduction ratio of the FV. When dealing with high-dimensional datasets the value  $\frac{D-M}{D}$  does not have much impact on the fitness value as this fraction would be very small. For this purpose, a new objective function is also proposed for such large datasets.

$$J = \alpha * \mathbf{Acc}(P_i) + (1 - \alpha) * \frac{\ln D - \ln M}{\ln D}, \quad (3.11)$$

Using the  $\ln$  function, the size of the reduced dataset can have a useful impact on the fitness value calculated for each chimp thereby effectively improving the performance of the proposed algorithm for large datasets.

## 3.2 Complexity Analysis

The temporal complexity of the BS-ChOA initialization is  $O(N * d)$ , where  $N$  represents the population size and  $d$  represents the dimension of the test function used to modify the solutions within the boundary. In the subsequent phase, the fitness evaluation of each search agent requires a temporal complexity of  $O(k_{\max} * N * d)$ , where  $k_{\max}$  is the upper limit on the number of iterations required to replicate the proposed BS-ChOA method. The overall time complexity of the BS-ChOA algorithm can be expressed as  $O(t_{\max} * N * d)$ .



## Chapter 4

# Results and Discussions

In this section, we present the experimental evaluation of the technique proposed in our study. The comprehensive assessment of existing literature indicates that MHAs demonstrate appropriate and well-balanced characteristics in addressing situations that do not have exact solutions. Due to this rationale, our study employed the utilization of distinctive algorithms. In order to accomplish this goal, it is imperative to provide a comprehensive and accurate depiction of the situation. While MHAs do not offer a guarantee of attaining optimal solutions, their objective is to generate solutions that closely approximate the optimal answer, while simultaneously minimising the computer resources needed. Every metaheuristic technique that has been suggested in the existing body of literature possesses its own distinct set of merits and drawbacks. The objective of this study is to enhance the strategy by using a wider array of parameters to achieve a more comprehensive and precise approach. The incorporation of a substantial quantity of datasets facilitates a thorough examination, whereas the use of multidimensional datasets and problems involving several classes showcases the algorithm's efficacy and exactitude over a wide range of scenarios.

The findings presented in this section demonstrate that our suggested approach, which prioritises a broader range of parameters, successfully and efficiently addresses the multi-objective optimization problem across different dataset sizes. The findings indicate that our method exhibited superior performance compared to the baseline techniques, as it successfully identified options that were more closely aligned with the best solution.

## 4.1 Dataset Preprocessing

In this study, the proposed technique for feature selection is applied to a total of eight high-dimensional biomedical datasets. These datasets consist of publicly available datasets from the University of California Irvine (UCI) database and the Mendeley dataset for Gene Selection and Classification. The specifics of each dataset, including attributes, instances, and the number of classes, are presented in Table 4.1. The datasets consist of a combination of category and numerical data, and certain features within them do not contribute to the overall classification of the dataset. In addition, the number of instances within the dataset has an impact on the computing time required for performing feature selection. Each dataset is preprocessed using min-max normalization to scale input features to a uniform range, ranging from 0 to 1. This is done to ensure that each feature contributes equally to the training of the model, thereby preventing the dominance of features with high scaling. Data was oversampled in order to overcome problems due to unbalanced datasets.

## 4.2 Experimental Setup

The proposed algorithm is developed and tested on Python 3.7 on a Windows system using a core i3-8th Gen Laptop with 16GB RAM and datasets taken from University of California, Irvine Machine Learning Repository (UCI), <https://jundong1.github.io/scikit-feature/datasets.html> and Mendeley Data for Gene Selection, <https://data.mendeley.com/datasets/fhx5zgx2zj/1>.

TABLE 4.1: Dataset Details

S.No.	Datasets [105]	# of Instances	# of Attributes	Classes
1	DLBCL	76	5469	2
2	Prostate Tumour	102	10509	2
3	Arcene	100	10000	2
4	Leukaemia	72	6114	5
5	Brain Tumour 1	90	5920	5
6	Brain Tumour 2	50	10367	4
7	BLDCL	60	5723	9
8	11 Tumours	202	1035	11

The algorithm is compared with its modern counterparts namely the **ChOA**, **GWO**, **WOA**, **PSO**, **GA** and **SCA**. The experimental setup is as follows.

- Each dataset is run using the KNN and RF-based FS methods, with a maximum iteration size of 100 and a population size of 30 to attain a balanced result on metrics. Each algorithm is run 20 times, each with a different train test split to provide a well-rounded average and the best results for all datasets.
- Each dataset is run with kNN and RF classifiers with both s and v functions.
- Convergence curves with a ln based update rule and simple average are also recorded.
- Selected features and time complexity for each algorithm are recorded for each iteration.
- Hyper-parameters are kept consistent as suggested by the original authors in their respective papers.

### 4.3 Results

The study employs two commonly used classifiers to test the performance of BS-ChOA with other algorithms. The results recorded exhibit that the BS-ChOA performs better than its counterparts for most datasets in terms of average accuracy, precision, f1-score and number of selected features. The results also indicate that there is an obvious trade-off in terms of train time for the BS-ChOA when compared with GA and in a number of selected features when compared with BWOA, however it is evident from the recorded average metrics that the trade-off rules in favour of the proposed BS-ChOA when overall performance is considered.

#### 4.3.1 Average and Best Metrics

The average and best metrics (accuracy, precision, f1-score, number of selected features) for each dataset were recorded during the experiment and are discussed in this section. Table 4.2 shows the average and best accuracy with v-function and RF classifier. Using the hybrid BS-ChOA a 2-8% increase in average accuracy is observed with the Brain

Tumour 1 dataset recording the highest increase of 8%. Figure 4.1 and 4.2 show the average and best accuracy for all data sets. It is observed that the BS-ChOA performs better than other candidate algorithms for most datasets when comparing average accuracy, a similar pattern is observed when comparing the average and best precision and f1-score.

Table 4.3 shows the mean and best precision recorded for 20 iterations and Figure 4.3 and 4.4 show bar graphs for the achieved precision in the experiment. Table 4.4 shows the mean and best F1-score, while Figure 4.8 and 4.9 show the mean and best F1-score. As the data is unbalanced, the F1-score proves to be a better metric when testing the performance, particularly for bio-medical datasets where the datasets are more likely to have multi-class unbalanced high-dimension datasets. The hybridization of ChOA with SCA has proven to be effective when comparing the average metrics as this enhances both the exploration and exploitation characteristics resulting in an overall increase in average metrics for the selected datasets.

Figure 4.10 and 4.11 show bar graphs of the recorded data of the number of selected features for each dataset, here the number of selected features, shown on the vertical axis are taken on a logarithmic scale to better represent the output. Figure 4.12 shows the average, metrics recorded for each individual dataset, these graphs indicate that BS-ChOA has shown the best average metrics throughout the iteration process.

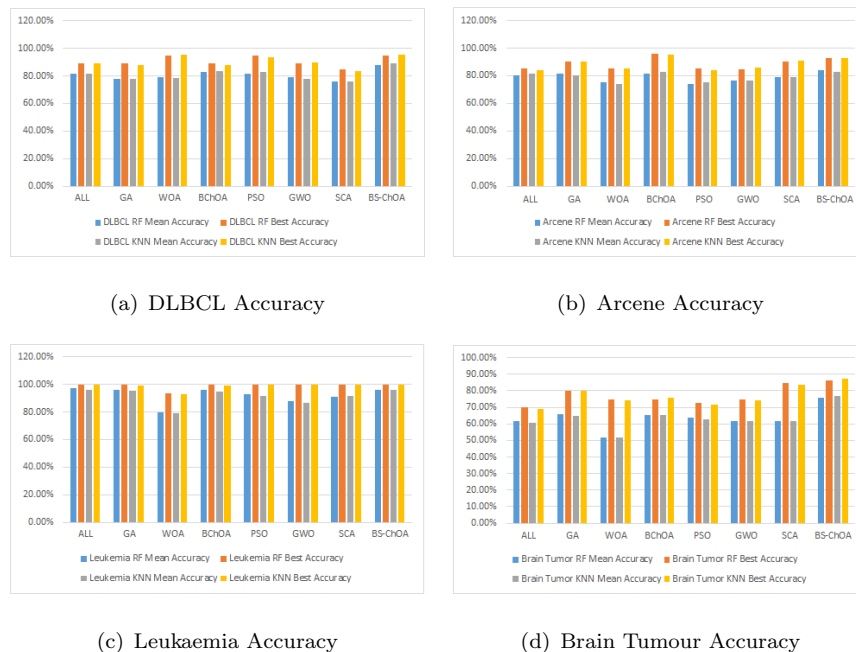
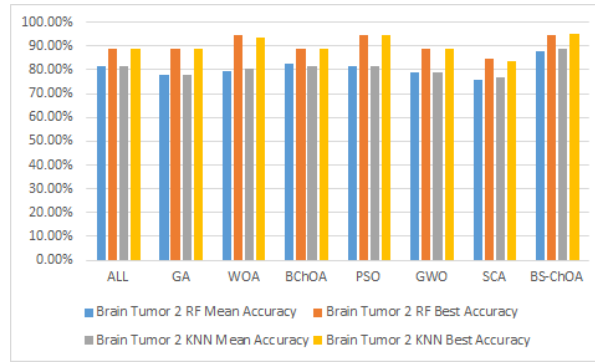
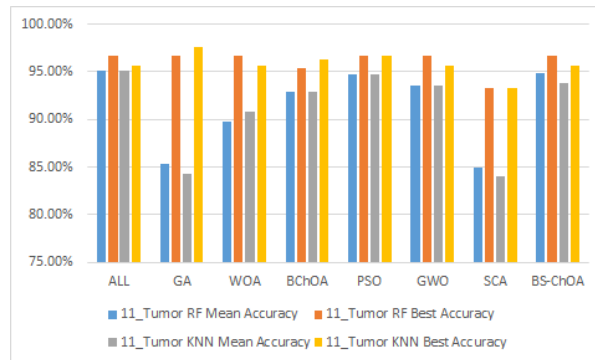


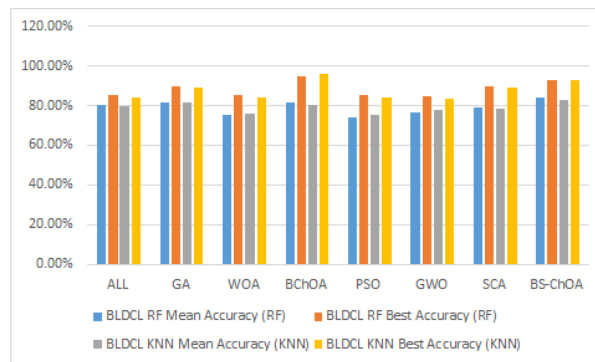
FIGURE 4.1: Average and Best Accuracy



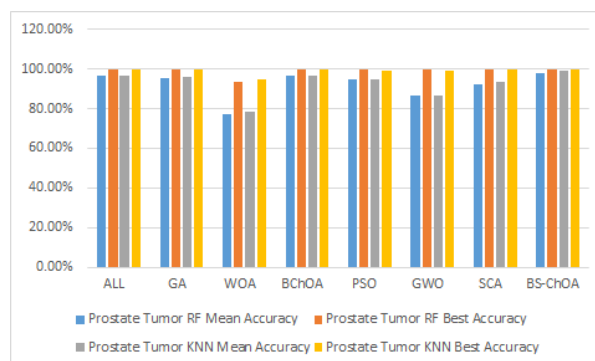
(a) Brain Tumour 2 Accuracy



(b) 11 Tumour Accuracy



(c) BLDCL Accuracy

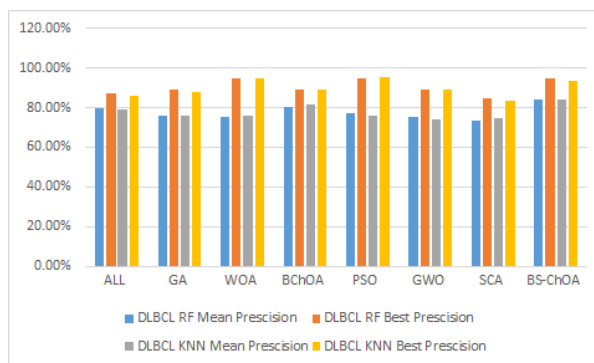


(d) Prostate Tumour Accuracy

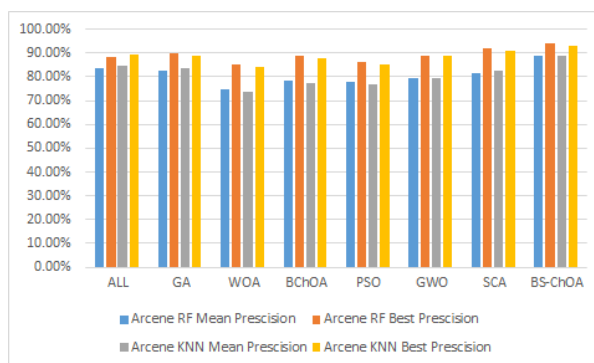
FIGURE 4.2: Average and Best Accuracy

TABLE 4.2: Mean and Best Accuracy

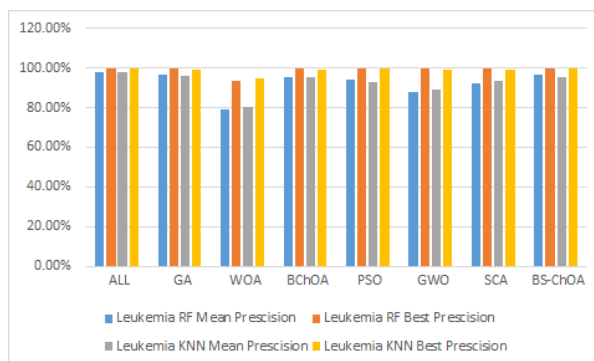
Dataset	Classifier		ALL	GA	WOA	BChOA	PSO	GWO	SCA	BS-ChOA
Brain Tumour	RF	Mean Accuracy	61.74%	65.84%	51.69%	65.22%	63.56%	61.87%	61.64%	75.65%
		Best Accuracy	70.00%	80.00%	75.00%	75.00%	72.73%	75.00%	84.55%	86.20%
	KNN	Mean Accuracy	60.74%	64.84%	51.69%	65.22%	62.56%	61.87%	61.64%	76.65%
		Best Accuracy	69.00%	80.00%	74.00%	76.00%	71.73%	74.00%	83.55%	87.20%
BLDCL	RF	Mean Accuracy	80.37%	81.53%	74.97%	81.49%	74.27%	76.79%	79.30%	83.87%
		Best Accuracy	85.00%	90.00%	85.00%	95.00%	85.19%	84.62%	90.00%	92.59%
	KNN	Mean Accuracy	79.37%	81.53%	75.97%	80.49%	75.27%	77.79%	78.30%	82.87%
		Best Accuracy	84.00%	89.00%	84.00%	96.00%	84.19%	83.62%	89.00%	92.59%
Prostate Tumour	RF	Mean Accuracy	96.52%	95.09%	77.31%	96.57%	94.98%	86.40%	92.41%	98.03%
		Best Accuracy	100.00%	100.00%	93.75%	100.00%	100.00%	100.00%	100.00%	100.00%
	KNN	Mean Accuracy	96.52%	96.09%	78.31%	96.57%	94.98%	86.40%	93.41%	99.03%
		Best Accuracy	100.00%	100.00%	94.75%	100.00%	99.00%	99.00%	100.00%	100.00%
11_Tumour	RF	Mean Accuracy	95.15%	85.33%	89.80%	92.96%	94.67%	93.52%	85.00%	94.86%
		Best Accuracy	96.67%	96.67%	96.67%	95.35%	96.67%	96.67%	93.33%	96.67%
	KNN	Mean Accuracy	95.15%	84.33%	90.80%	92.96%	94.67%	93.52%	84.00%	93.86%
		Best Accuracy	95.67%	97.67%	95.67%	96.35%	96.67%	95.67%	93.33%	95.67%
Leukaemia	RF	Mean Accuracy	97.14%	96.26%	79.84%	95.80%	92.75%	87.79%	90.69%	96.25%
		Best Accuracy	100.00%	100.00%	93.75%	100.00%	100.00%	100.00%	100.00%	100.00%
	KNN	Mean Accuracy	96.14%	95.26%	78.84%	94.80%	91.75%	86.79%	91.69%	96.25%
		Best Accuracy	100.00%	99.00%	92.75%	99.00%	100.00%	100.00%	100.00%	100.00%
DLBCL	RF	Mean Accuracy	81.38%	77.70%	79.35%	82.49%	81.61%	79.06%	75.97%	87.89%
		Best Accuracy	88.89%	88.89%	94.44%	88.89%	94.44%	88.89%	84.44%	94.44%
	KNN	Mean Accuracy	81.38%	77.70%	78.35%	83.49%	82.61%	78.06%	75.97%	88.89%
		Best Accuracy	88.89%	87.89%	95.44%	87.89%	93.44%	89.89%	83.44%	95.23%
Brain Tumour	RF	Mean Accuracy	81.38%	77.70%	79.35%	82.49%	81.61%	79.06%	75.97%	87.89%
		Best Accuracy	88.89%	88.89%	94.44%	88.89%	94.44%	88.89%	84.44%	94.44%
	KNN	Mean Accuracy	81.38%	77.70%	80.35%	81.49%	81.61%	79.06%	76.97%	88.89%
		Best Accuracy	88.89%	88.89%	93.44%	88.89%	94.44%	88.89%	83.44%	94.95%
Arcene	RF	Mean Accuracy	80.37%	81.53%	74.97%	81.49%	74.27%	76.79%	79.30%	83.87%
		Best Accuracy	85.00%	90.00%	85.00%	96.25%	85.19%	84.62%	90.00%	92.59%
	KNN	Mean Accuracy	81.37%	80.53%	73.97%	82.49%	75.27%	76.79%	79.30%	82.87%
		Best Accuracy	84.00%	90.00%	85.00%	95.50%	84.19%	85.62%	91.00%	92.59%



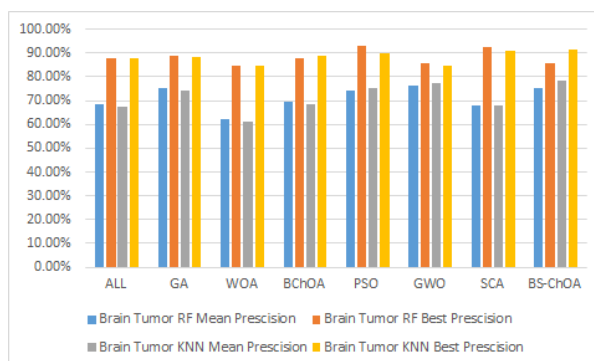
(a) DLBCL Precision



(b) Arcene Precision

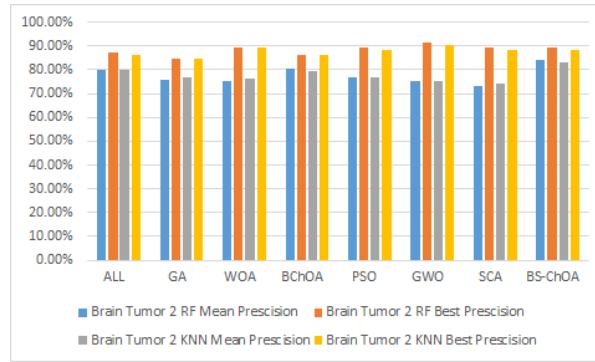


(c) Leukaemia Precision

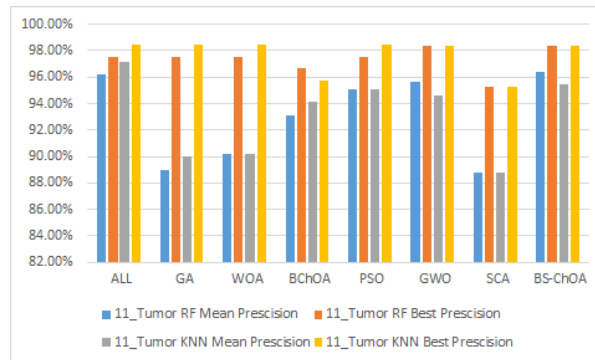


(d) Brain Tumour Precision

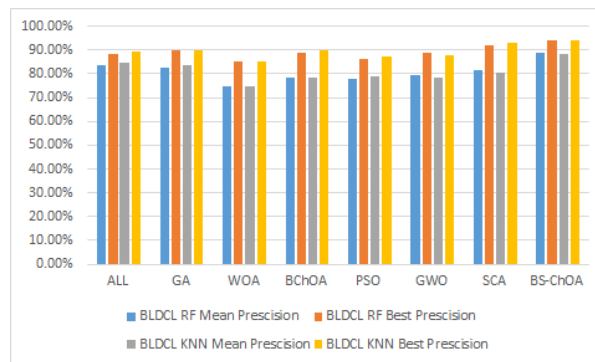
FIGURE 4.3: Average and Best Precision



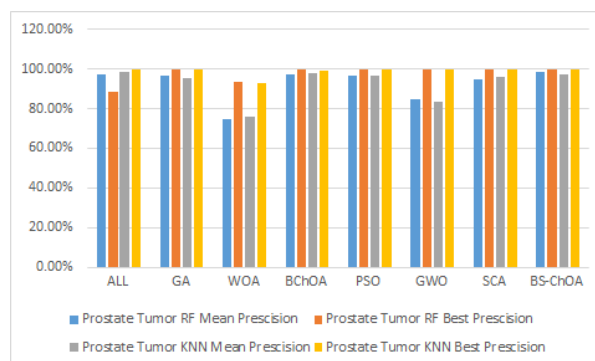
(a) Brain Tumour 2 Precision



(b) 11 Tumour Precision



(c) BLDCL Precision



(d) Prostate Tumour Precision

FIGURE 4.4: Average and Best Precision



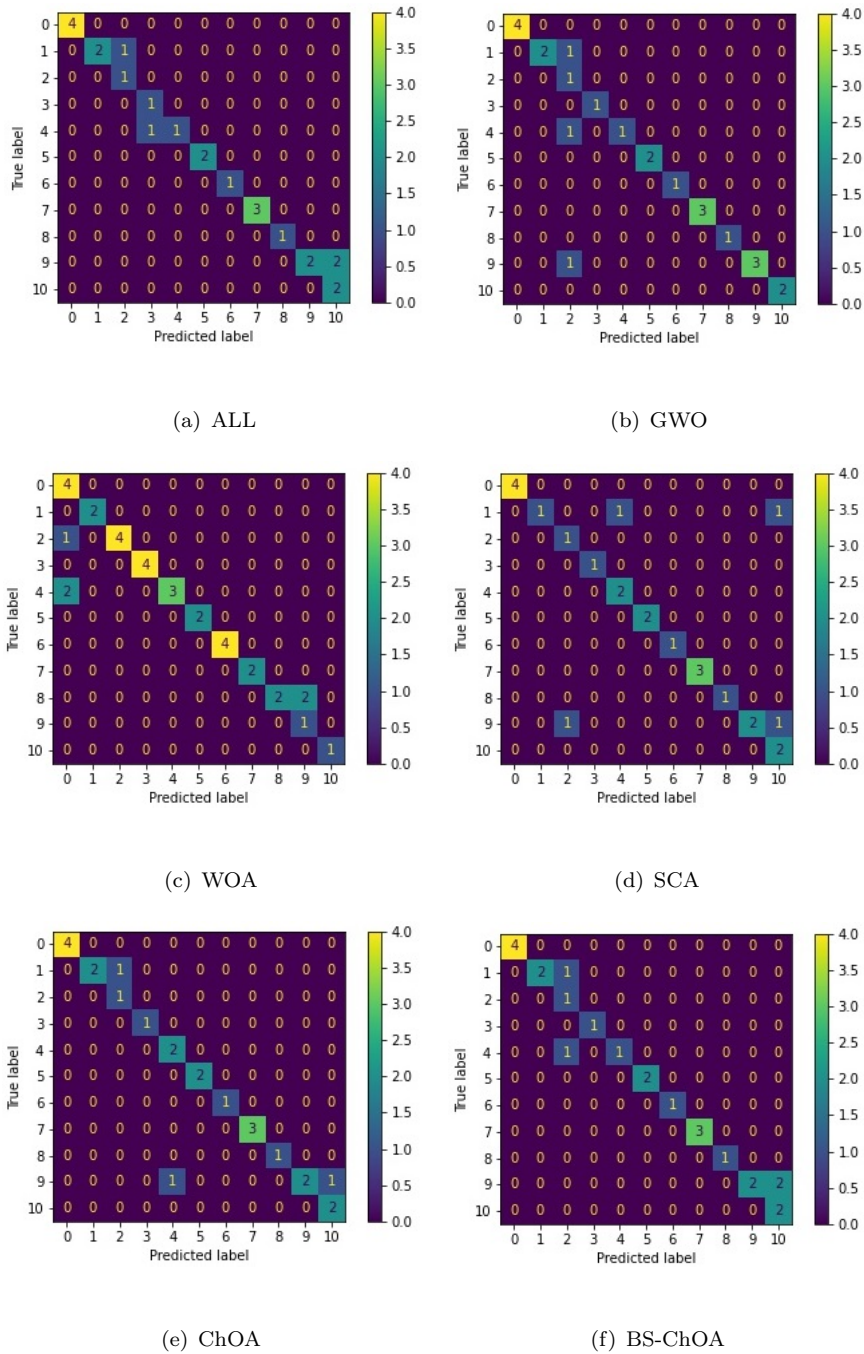


FIGURE 4.5: Confusion Matrix 11 Tumour Dataset

In the case of selected features, we observe that the BS-ChOA is better than most of its contemporaries. WOA is an exception in this case however, the poor accuracy, precision and F1-score of WOA indicate that the algorithm suffers from pre-mature convergence and over-fitting during the training process. Table 4.5 shows the number of selected genes for each algorithm, taking this information in conjunction with other metrics, it is evident that the BS-ChOA has better convergence characteristics than other algorithms.

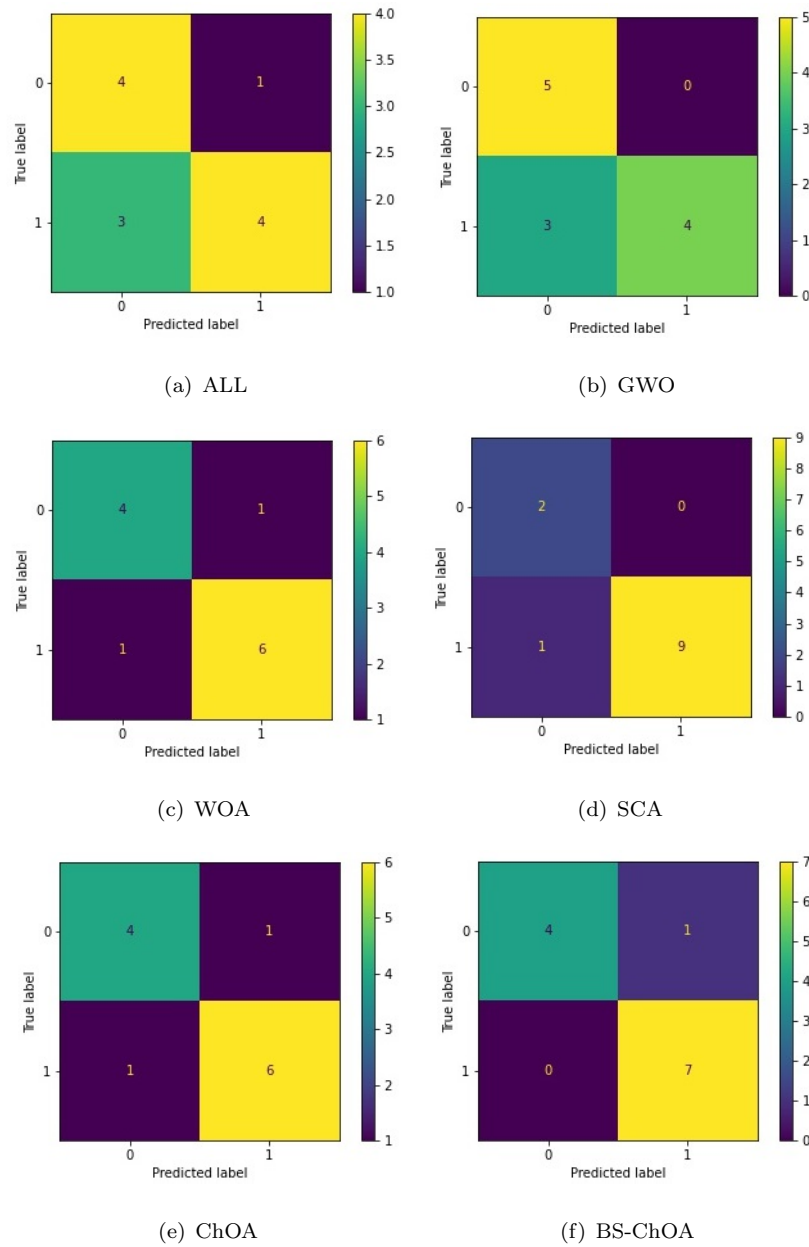
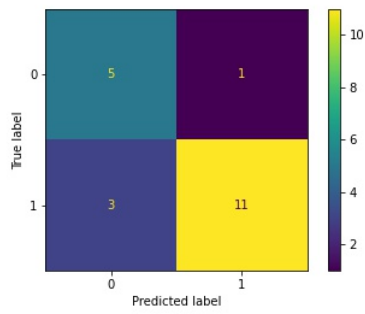
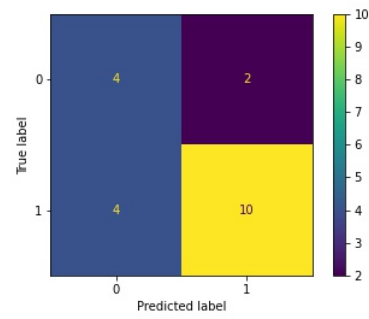


FIGURE 4.6: Confusion Matrix Prostate Tumour Dataset

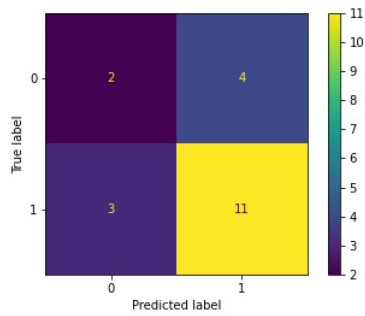
The recorded data shows that for Brain\_Tumour, Brain\_Tumour2, Arcene and DLBCL datasets there is a 5-7% gain in the mean precision and 6-9% increase in the mean F1-score. The increase in the F1-score is an important factor in this regard as biomedical datasets are heavily unbalanced and accuracy and precision are not termed as good metrics to measure the efficacy of optimization algorithms. To further strengthen the results Figure 4.5 - 4.7 provide a graphical representation of the confusion matrix obtained during this experiment. These results show that the BS-ChOA has better metrics when compared with other algorithms for multi-modal datasets.



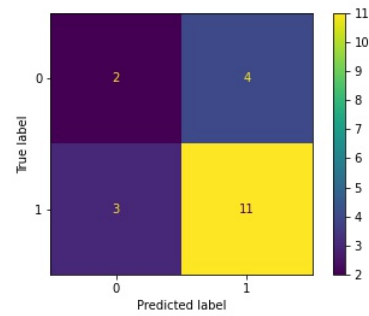
(a) ALL



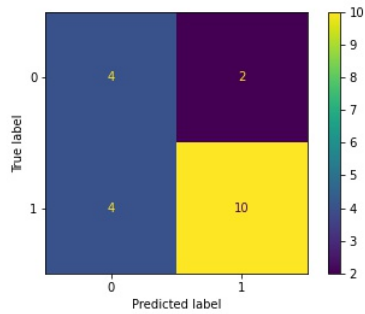
(b) GWO



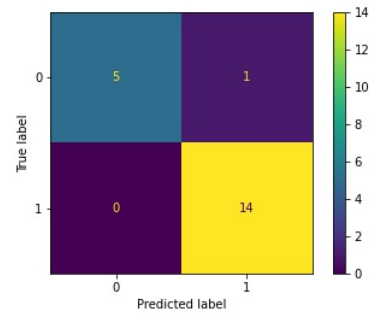
(c) WOA



(d) SCA



(e) ChOA



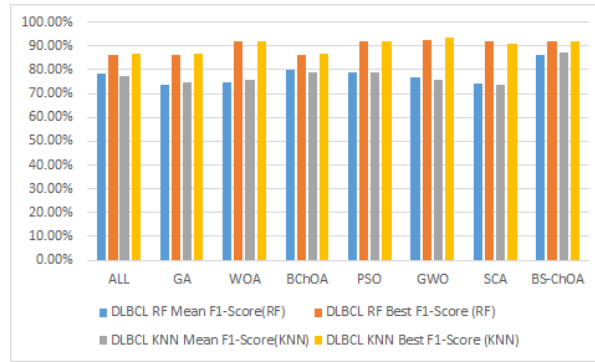
(f) BS-ChOA

FIGURE 4.7: Confusion Matrix Prostate Cancer Dataset

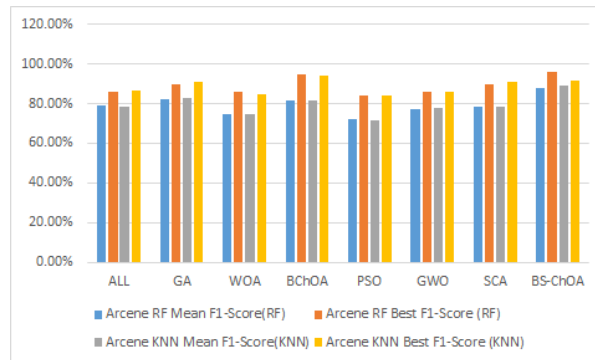
Figure 4.12 and 4.13 show the average metrics for all datasets, we can see that the proposed BS-ChOA algorithm provides a much-balanced output in terms of high accuracy, precision, f1-score and low no. of selected features when compared with other algorithms.

TABLE 4.3: Average and Best Precision

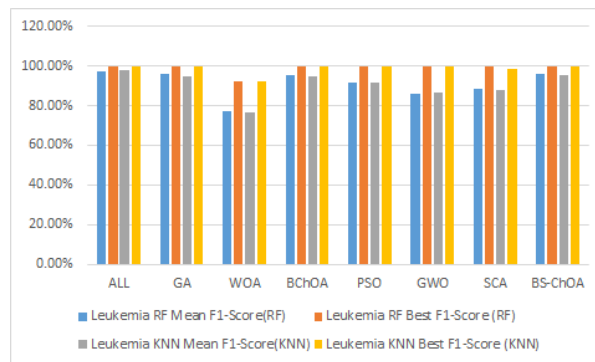
Dataset	Classifier		ALL	GA	WOA	BChOA	PSO	GWO	SCA	BS-ChOA
Brain Tumour	RF	Mean Precision	68.59%	75.47%	62.29%	69.71%	74.23%	76.49%	67.76%	75.42%
		Best Precision	87.88%	89.09%	84.55%	87.88%	93.18%	85.91%	92.50%	85.91%
	KNN	Mean Precision	67.59%	74.47%	61.29%	68.71%	75.23%	77.49%	67.76%	78.55%
		Best Precision	87.88%	88.09%	84.55%	88.88%	89.98%	84.91%	90.92%	91.25%
BLDCL	RF	Mean Precision	83.58%	82.47%	74.50%	78.47%	77.78%	79.27%	81.73%	89.10%
		Best Precision	88.50%	90.00%	85.00%	88.85%	86.20%	88.85%	92.00%	94.00%
	KNN	Mean Precision	84.58%	83.47%	74.50%	78.47%	78.78%	78.27%	80.73%	88.10%
		Best Precision	89.50%	90.00%	85.00%	89.85%	87.20%	87.85%	93.00%	94.00%
Prostate Tumour	RF	Mean Precision	97.42%	96.33%	74.95%	97.08%	96.51%	84.59%	94.67%	98.52%
		Best Precision	88.50%	100.00%	93.75%	100.00%	100.00%	100.00%	100.00%	100.00%
	KNN	Mean Precision	98.42%	95.33%	75.95%	98.08%	96.51%	83.59%	95.67%	97.52%
		Best Precision	100.00%	100.00%	92.75%	99.00%	100.00%	100.00%	100.00%	100.00%
11_Tumour	RF	Mean Precision	96.18%	88.98%	90.15%	93.12%	95.10%	95.61%	88.76%	96.42%
		Best Precision	97.50%	97.50%	97.50%	96.71%	97.50%	98.33%	95.28%	98.33%
	KNN	Mean Precision	97.18%	89.98%	90.15%	94.12%	95.10%	94.61%	88.76%	95.42%
		Best Precision	98.50%	98.50%	98.50%	95.71%	98.50%	98.33%	95.28%	98.33%
Leukaemia	RF	Mean Precision	97.78%	96.68%	79.10%	95.10%	94.06%	87.91%	92.30%	96.35%
		Best Precision	100.00%	100.00%	93.75%	100.00%	100.00%	100.00%	100.00%	100.00%
	KNN	Mean Precision	97.78%	95.68%	80.10%	95.10%	93.06%	88.91%	93.30%	95.35%
		Best Precision	100.00%	99.00%	94.75%	99.00%	100.00%	99.00%	99.00%	100.00%
DLBCL	RF	Mean Precision	79.76%	75.91%	75.18%	80.50%	77.01%	75.30%	73.37%	84.06%
		Best Precision	87.18%	88.89%	94.44%	88.89%	94.44%	88.89%	84.44%	94.44%
	KNN	Mean Precision	78.76%	75.91%	76.18%	81.50%	76.01%	74.30%	74.37%	84.06%
		Best Precision	86.18%	87.89%	94.44%	88.89%	95.44%	88.89%	83.44%	93.44%
Brain Tumour 2	RF	Mean Precision	79.76%	75.91%	75.18%	80.50%	77.01%	75.30%	73.37%	84.06%
		Best Precision	87.18%	84.92%	89.29%	86.30%	89.29%	91.67%	89.26%	89.32%
	KNN	Mean Precision	79.76%	76.91%	76.18%	79.50%	77.01%	75.30%	74.37%	83.06%
		Best Precision	86.18%	84.92%	89.29%	86.30%	88.29%	90.67%	88.26%	88.32%
Arcene	RF	Mean Precision	83.58%	82.47%	74.50%	78.47%	77.78%	79.27%	81.73%	89.10%
		Best Precision	88.50%	90.00%	85.00%	88.85%	86.20%	88.85%	92.00%	94.00%
	KNN	Mean Precision	84.58%	83.47%	73.50%	77.47%	76.78%	79.27%	82.73%	89.10%
		Best Precision	89.50%	89.00%	84.00%	87.85%	85.20%	88.85%	91.00%	93.00%



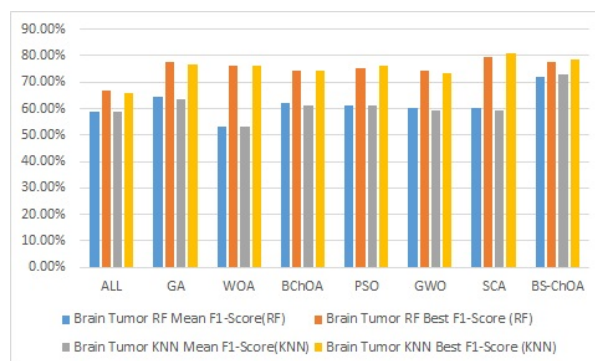
(a) DLBCL F1-Score



(b) Arcene F1-Score

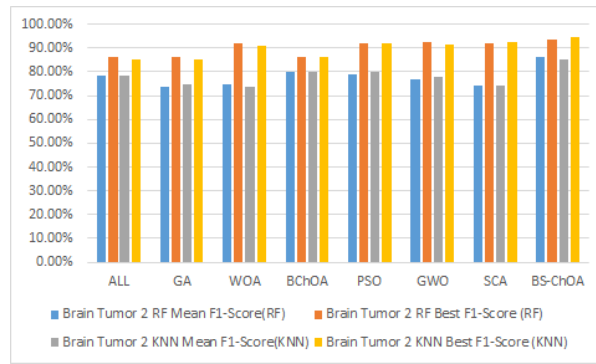


(c) Leukaemia F1-Score

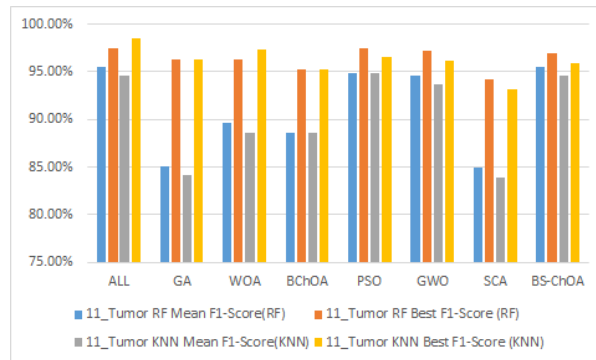


(d) Brain Tumour F1-Score

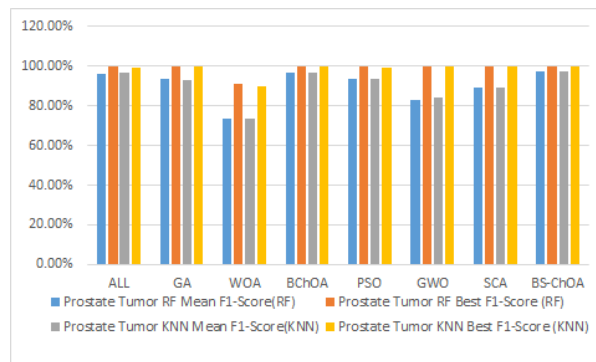
FIGURE 4.8: Average and Best F1-Score



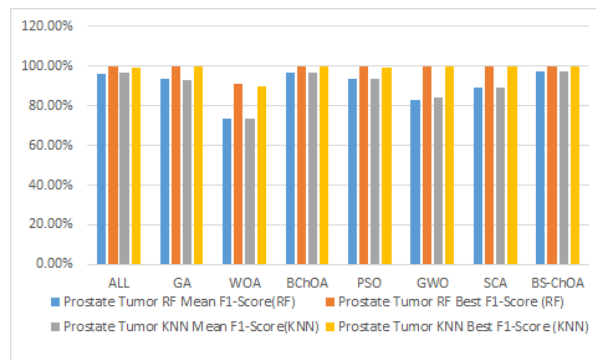
(a) Brain Tumour 2 F1-Score



(b) 11 Tumour F1-Score



(c) BLDCL F1-Score

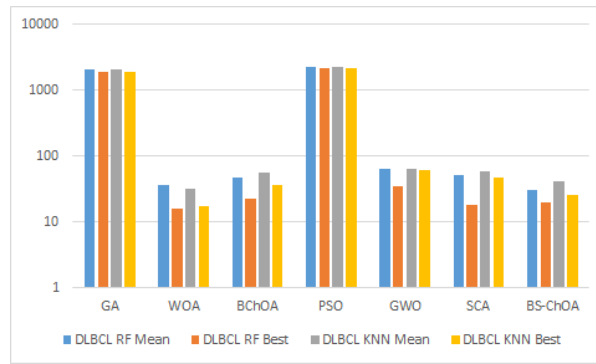


(d) Prostate Tumour F1-Score

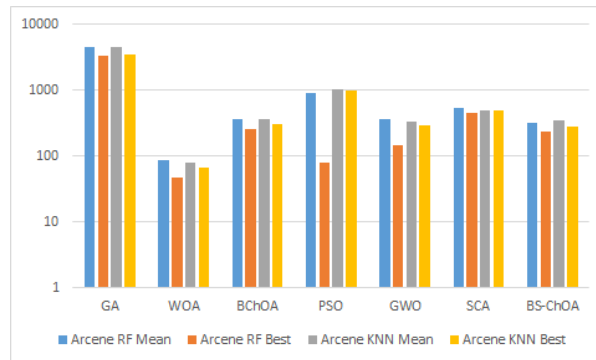
FIGURE 4.9: Average and Best F1-Score

TABLE 4.4: Average and Best F1-Score

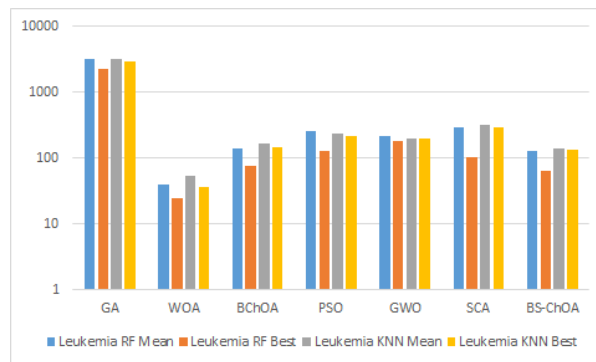
Dataset	Classifier		ALL	GA	WOA	BChOA	PSO	GWO	SCA	BS-ChOA
Brain Tumour	RF	Mean F1-Score	58.74%	64.66%	53.00%	61.99%	61.20%	60.25%	60.33%	71.74%
		Best F1-Score	66.98%	77.50%	76.11%	74.21%	75.06%	74.21%	79.71%	77.50%
	KNN	Mean F1-Score	58.74%	63.66%	53.00%	60.99%	61.20%	59.45%	59.33%	72.74%
		Best F1-Score	65.98%	76.50%	76.11%	74.21%	76.06%	73.21%	80.71%	78.50%
BLDCL	RF	Mean F1-Score	83.58%	82.47%	74.50%	78.47%	77.78%	79.27%	81.73%	89.10%
		Best F1-Score	85.70%	90.00%	85.70%	95.00%	83.80%	85.70%	90.00%	92.30%
	KNN	Mean F1-Score	82.58%	82.47%	74.50%	77.47%	76.78%	78.27%	81.73%	89.10%
		Best F1-Score	84.70%	91.00%	84.70%	95.00%	82.80%	85.70%	91.00%	91.30%
Prostate Tumour	RF	Mean F1-Score	95.90%	93.79%	73.45%	96.54%	93.32%	82.92%	89.32%	97.46%
		Best F1-Score	100.00%	100.00%	90.86%	100.00%	100.00%	100.00%	100.00%	100.00%
	KNN	Mean F1-Score	96.90%	92.79%	73.45%	96.54%	93.32%	83.92%	89.32%	97.46%
		Best F1-Score	99.00%	100.00%	89.86%	100.00%	99.00%	100.00%	100.00%	100.00%
11_Tumour	RF	Mean F1-Score	95.58%	85.09%	89.58%	88.55%	94.92%	94.62%	84.87%	95.55%
		Best F1-Score	97.50%	96.35%	96.35%	95.28%	97.50%	97.22%	94.17%	96.89%
	KNN	Mean F1-Score	94.58%	84.09%	88.58%	88.55%	94.92%	93.62%	83.87%	94.55%
		Best F1-Score	98.50%	96.35%	97.35%	95.28%	96.50%	96.22%	93.17%	95.89%
Leukaemia	RF	Mean F1-Score	97.08%	95.75%	77.32%	95.63%	91.64%	85.69%	88.62%	96.07%
		Best F1-Score	100.00%	100.00%	92.10%	100.00%	100.00%	100.00%	100.00%	100.00%
	KNN	Mean F1-Score	98.08%	94.50%	76.32%	94.63%	91.64%	86.69%	87.62%	95.07%
		Best F1-Score	100.00%	99.50%	92.10%	100.00%	100.00%	100.00%	98.75%	100.00%
DLBCL	RF	Mean F1-Score	78.52%	73.83%	74.95%	80.07%	79.10%	76.64%	74.47%	86.22%
		Best F1-Score	85.98%	85.98%	91.77%	85.98%	91.77%	92.59%	91.76%	91.78%
	KNN	Mean F1-Score	77.52%	74.83%	75.95%	79.07%	79.10%	75.64%	73.47%	87.22%
		Best F1-Score	86.98%	86.98%	91.77%	86.98%	91.77%	93.59%	90.76%	91.78%
Brain Tumour 2	RF	Mean F1-Score	78.52%	73.83%	74.95%	80.07%	79.10%	76.64%	74.47%	86.22%
		Best F1-Score	85.98%	85.98%	91.77%	85.98%	91.77%	92.59%	91.76%	93.50%
	KNN	Mean F1-Score	78.52%	74.83%	73.95%	80.07%	80.10%	77.64%	74.47%	85.22%
		Best F1-Score	84.98%	84.98%	90.77%	85.98%	91.77%	91.59%	92.76%	94.50%
Arcene	RF	Mean F1-Score	79.28%	81.96%	74.37%	81.38%	72.22%	76.98%	78.10%	87.88%
		Best F1-Score	85.70%	90.00%	85.70%	95.00%	83.80%	85.70%	90.00%	96.00%
	KNN	Mean F1-Score	78.28%	82.96%	74.37%	81.38%	71.22%	77.98%	78.10%	88.88%
		Best F1-Score	86.70%	91.00%	84.70%	94.00%	83.80%	85.70%	91.00%	91.30%



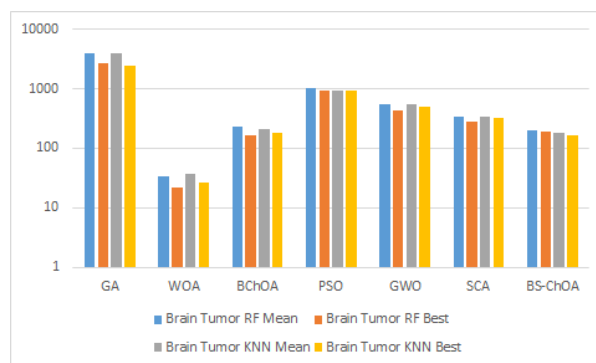
(a) DLBCL



(b) Arcene



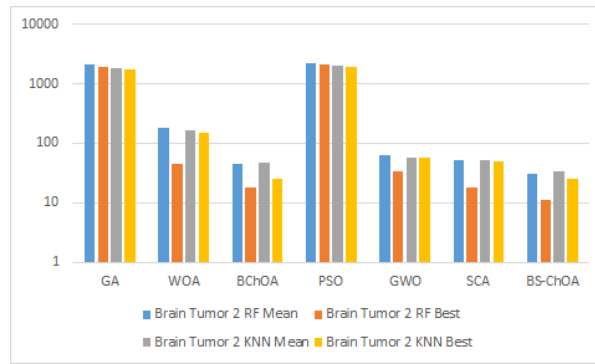
(c) Leukaemia



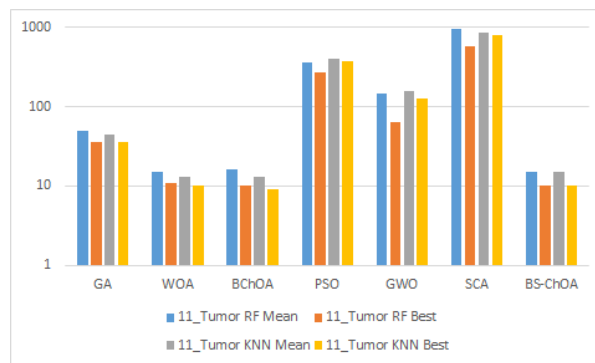
(d) Brain Tumour

FIGURE 4.10: Average and Best Number of Selected features

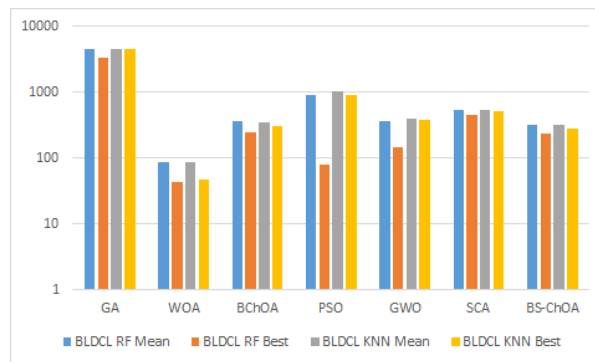




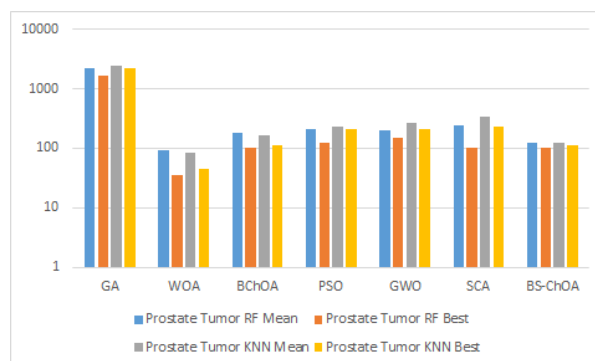
(a) Brain Tumour 2



(b) 11 Tumour



(c) BLDC L

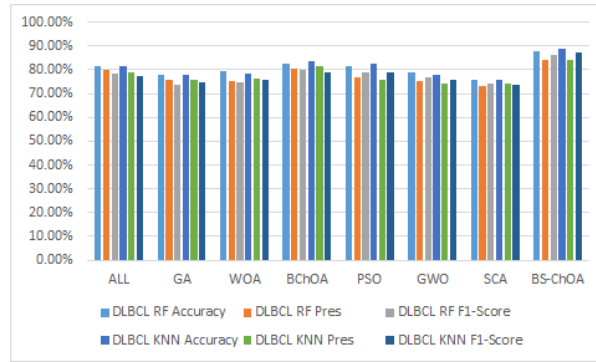


(d) Prostate Tumour

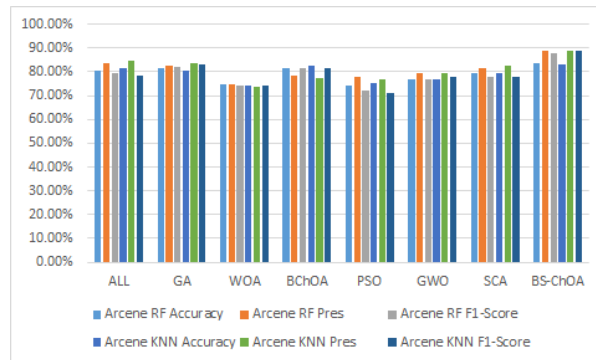
FIGURE 4.11: Average and Best Number of selected features

TABLE 4.5: Average and Best Selected Features

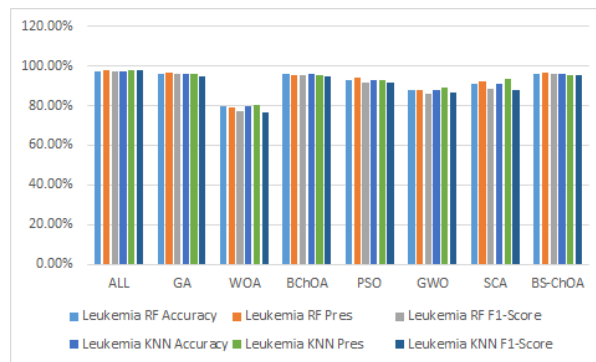
Dataset	Classifier		ALL	GA	WOA	BChOA	PSO	GWO	SCA	BS-ChOA
Brain Tumor	RF	Mean	10368	3869	33	226	1048	538	340	204
		Best		2730	22	165	910	436	283	188
	KNN	Mean		3869	37	213	943	538	340	183
		Best		2438	26	185	927	488	319	165
BLDCL	RF	Mean	10000	4569	88	357	918	363	546	313
		Best		3359	44	245	78	148	441	238
	KNN	Mean		4569	86	353	1010	399	546	313
		Best		4448	46	299	916	372	516	282
Prostate Tumor	RF	Mean	6448	2198	93	182	205	197	237	125
		Best		1680	36	100	125	152	103	100
	KNN	Mean		2417	84	163	226	265	336	125
		Best		2210	46	112	213	211	236	112
11.Tumor	RF	Mean	12534	49	15	16	359	145	943	15
		Best		36	11	10	273	64	570	10
	KNN	Mean		44	13	13	395	159	848	15
		Best		36	10	9	368	125	795	10
Leukemia	RF	Mean	7129	3171	40	140	259	218	286	125
		Best		2297	25	75	125	177	103	65
	KNN	Mean		3171	54	168	233	196	315	138
		Best		2883	36	144	218	195	290	133
DLBCL	RF	Mean	5921	2070	36	46	2215	64	52	30
		Best		1882	16	22	2115	34	18	20
	KNN	Mean		2070	32	57	2215	64	58	42
		Best		1911	17	36	2109	62	46	26
Brain Tumor 2	RF	Mean	5921	2070	180	44	2215	64	52	30
		Best		1882	45	18	2115	34	18	11
	KNN	Mean		1863	162	47	1994	57	52	33
		Best		1727	148	25	1881	57	50	25
Arcene	RF	Mean	10000	4569	88	355	918	363	546	313
		Best		3359	46	260	78	148	441	238
	KNN	Mean		4569	79	358	1010	327	491	344
		Best		3460	66	300	968	297	489	275



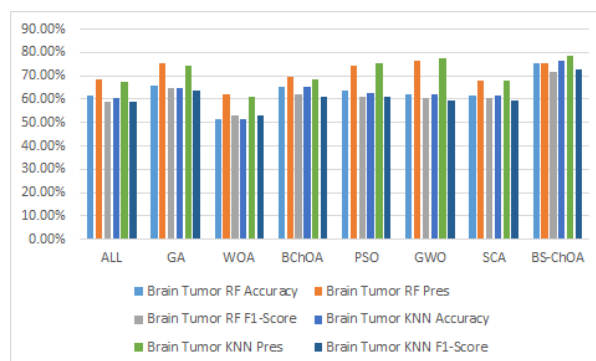
(a) DLBCL



(b) Arcene

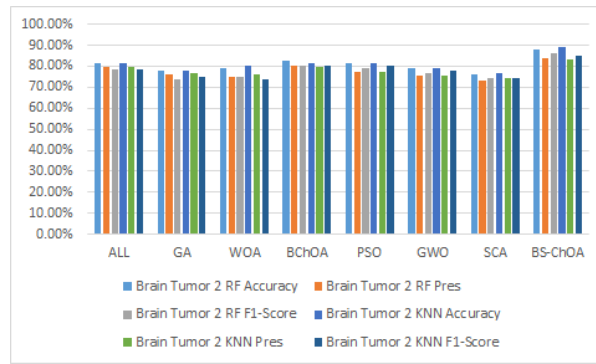


(c) Leukaemia

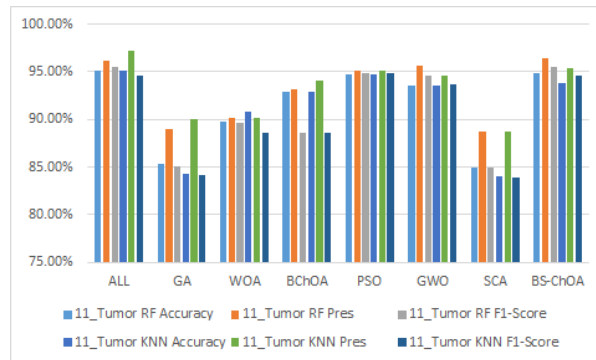


(d) Brain Tumour

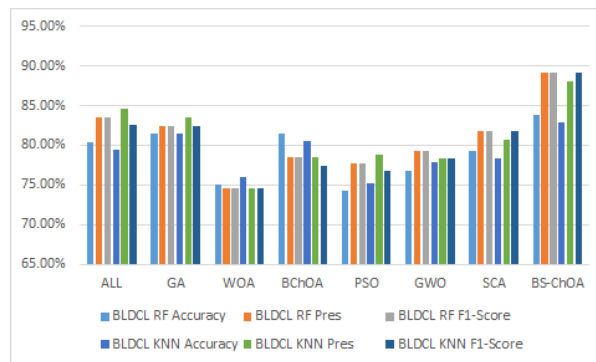
FIGURE 4.12: Average Metrics for All datasets



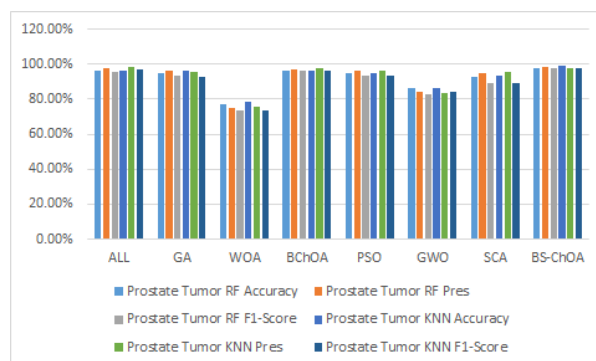
(a) Brain Tumour 2



(b) 11 Tumour



(c) BLDCL



(d) Prostate Tumour

FIGURE 4.13: Average Metrics for All datasets

### 4.3.2 Transformation Functions

For the purpose of this experiment two transformation functions namely the s- and v-function were used. Figure 4.15 shows the average metrics comparison chart between the s and v functions. When comparing the average metrics both s- and v- functions show good results. Figure 4.14 shows the average number of features selected when using each TF, from the comparison chart we can see that the s-function tends to give a smaller dataset when compared to the v-function.

This is owed to the fact that the v-function equation (3.9) depends on the size of datasets, since we are dealing with very large datasets the v-function output will not vary much with small changes. This is not the case with the s-function hence it performs better when comparing the number of features selected for high dimensional datasets.

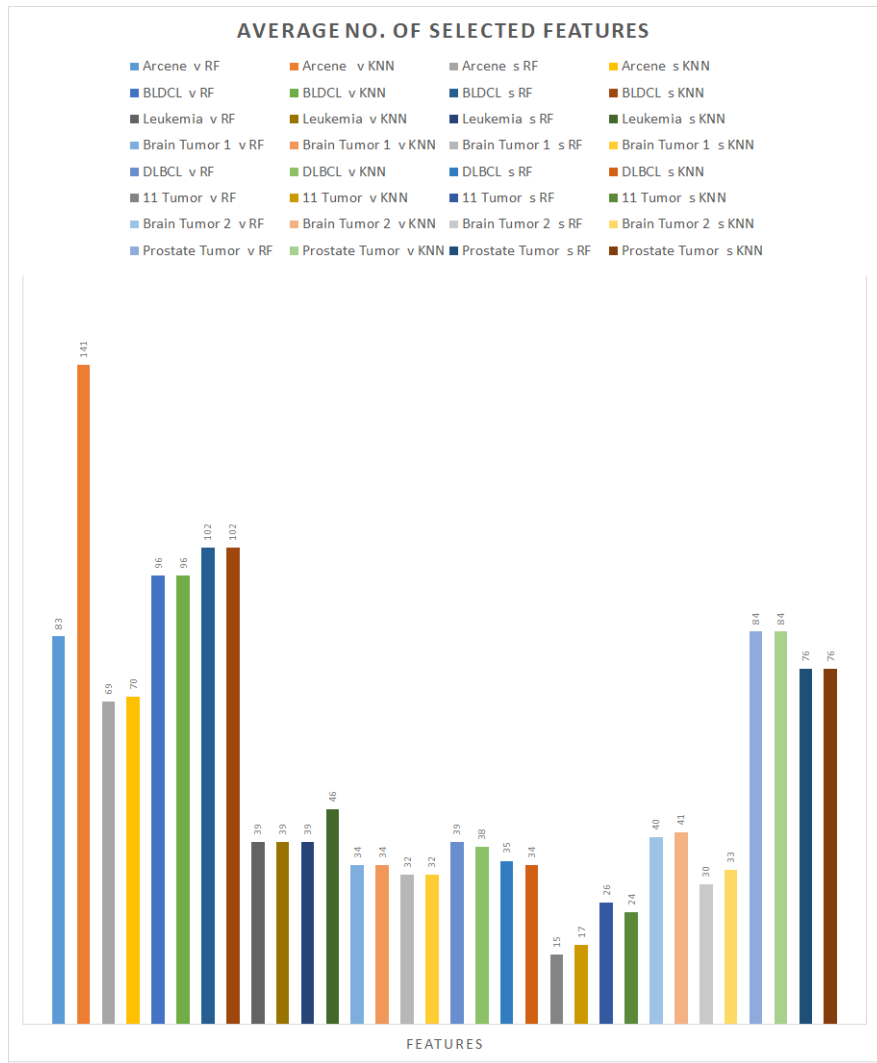


FIGURE 4.14: Average number of features selected using s and v TF

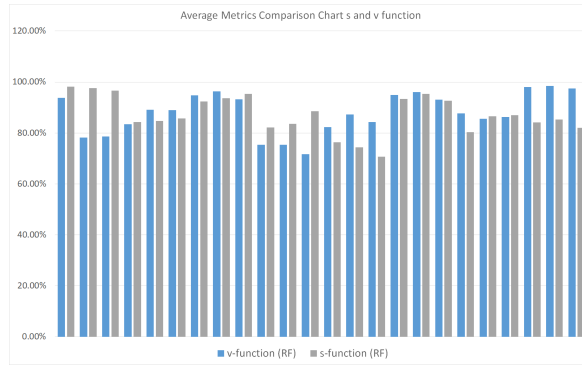


FIGURE 4.15: Average Metrics for s and v TF

### 4.3.3 Convergence Curve

As discussed in the previous chapter the update rule for the ChOA used simple averages to update the position of each chimp. As we can see in Figure 4.16 this method leads to non-ideal convergence characteristics as a population other than the best candidate also pulls the chimps with the same weight as the best solution. To mitigate this problem a weighted average update rule is introduced.

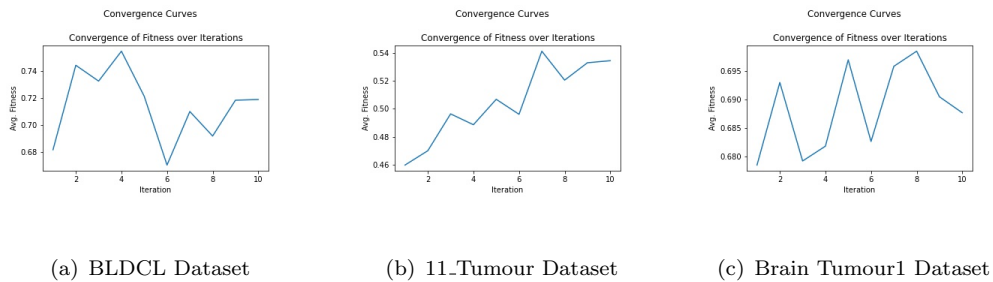


FIGURE 4.16: Convergence curve for BS-ChOA with simple average for 10 iterations

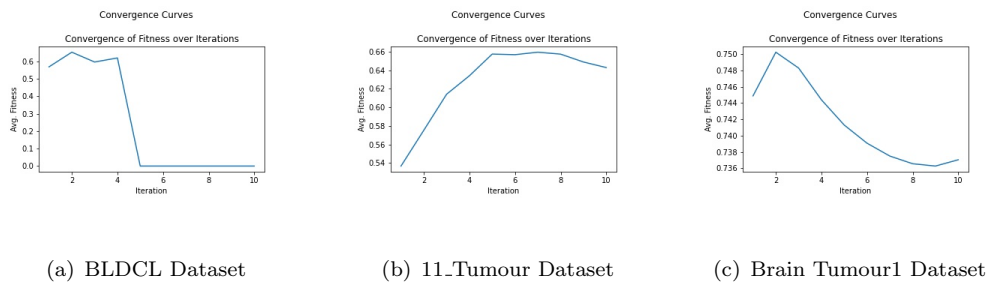


FIGURE 4.17: Convergence curve for BS-ChOA with weight average for 10 iterations

For the purpose of this experiment the coefficients  $b_i$  ( $i = 1, 2, 3, 4$ ) as discussed in (3.6) are set as  $b_1 = 0.5$ ,  $b_2 = 0.3$ ,  $b_3 = 0.15$  and  $b_4 = 0.05$ . As shown in Figure 4.17

by changing the update rule the convergence curve has improved and hence a better solution is achieved.

## 4.4 Discussion

The datasets used in the experiment provide a fair comparison for analysing the proposed method for high-dimensional datasets, the selected 8 datasets have the necessary variation in the number of classes and data unbalance required to carefully estimate the efficacy of BS-ChOA. Section 4.3 provides thorough details of the experiments conducted to validate the performance of the proposed algorithm and the necessary changes made to improve the performance of the base algorithms. The hybridization of ChOA with SCA, the change in update rule for standard ChOA and the use of ln based fitness function evaluation have proven to improve the performance and convergence characteristics to effectively solve FS problems for complex datasets. The results shown in Section 4.3.1 show a considerable improvement in the average and best metrics of the proposed algorithm when compared with its contemporaries. Section 4.3.3 provides valuable insights regarding the impact of using a weighted average-based update rule in comparison with simple average and shows that the convergence curves with weighted average are much improved than the conventional algorithm. Furthermore, Table 4.5 shows that the BS-ChOA has improved the FS process as well and decreased the number of features by a margin of 3-5% more than the other algorithms whilst maintaining good precision, accuracy and F1-score. In light of these results, the BS-ChOA has proved to be better than the other algorithms when comparing the accuracy, precision and f1-score, however, this algorithm has its limitations. The use of sin and cos functions during the exploration phase causes an increase in the computational expense which increases the train time. The algorithm performs well for unbalanced datasets providing a considerable increase in average accuracy, however, this increased computational compromises its effectiveness when used for simple balanced classes.

## Chapter 5

# Conclusion

Given the essential nature and profound importance of feature selection, along with the intrinsic intricacy of the problem, there is an increasing emphasis on the investigation and implementation of feature selection approaches based on metaheuristics. The efficacy of machine learning algorithms is influenced by the existence of superfluous or inconsequential data. The findings of our research demonstrate that metaheuristics exhibit significant efficacy as optimization techniques, enabling them to attain optimal performance across feature selection challenges of varying dimensions, including low-dimensional, medium-dimensional, and high-dimensional scenarios. The objective of this study was to perform a comparative analysis of different recent MHAs for high dimensional bio-medical datasets and present an innovative solution to overcome the problems faced by these algorithms. The proposed algorithm has proven to be effective in dealing with multi-modal datasets and showed improved and consistent results when compared with its counterparts. Consequently, a new methodology named the Binary Sine Cosine-Chimp Optimization Algorithm (BS-ChOA) was developed. The performance of the BS-ChOA was enhanced through the integration of the SCA, which leveraged its expertise in determining ideal feasible regions, resulting in benefits for the ChOA. The efficacy of the proposed BS-ChOA technique was assessed by employing benchmark datasets sourced from the University of California Irvine (UCI) and Mendeley Data for Gene Selection and juxtaposing it against alternative competitive algorithms. The experimental findings provided evidence of the superior efficacy of the Hybrid BS-ChOA method in comparison to various established metaheuristic algorithms, such as the original ChOA and SCA, as well as GA, BGWO, BPSO, and BWOA. Our findings



indicate that in lower-dimensional problems, the impact of feature selection on the challenges associated with big data and its storage is minimal. As the complexity of issues grows, the importance of metaheuristics in the process of feature selection becomes increasingly evident, and the variations in performance between different metaheuristics become significant. The importance of carefully choosing a suitable metaheuristic approach for feature selection in practical scenarios is underscored, as the selection of a metaheuristic method can have a substantial influence on the efficacy and efficiency of the feature selection procedure.

## 5.1 Future Work and Limitations

In future works, it is recommended to evaluate the efficacy of the proposed BSChOA in numerous problem domains. These domains encompass solving a wide range of multi-objective problems, optimising engineering constraints, optimising hyperparameters in machine learning, and performing multilevel threshold segmentation. The assessment of BSChOA's performance in various scenarios will offer valuable insights into its adaptability and efficacy in tackling diverse optimization difficulties. Furthermore, as the proposed method has shown its effectiveness when dealing with bio-medical datasets, a use case for the development of ANN/DNN techniques using FS can be made for the purpose of federated learning, as this would not only optimise the use of transmission bandwidth but also save storage and training time for the ANN models to be used. Although the BS-ChOA technique has shown promising results in feature selection, it is important to recognise certain limitations. Initially, it should be noted that the proposed methodology may not universally align with all datasets and problems, since distinct datasets may necessitate diverse feature selection methodologies. Additionally, due to the iterative nature of the approach, datasets containing millions of entries may require a longer processing time to identify pertinent features.

# Bibliography

- [1] C. D. STRUCTURING, “End2end unstructured data processing, confidential data structuring & storage using image processing, nlp, machine learning, and blockchain,” *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 13, 2022.
- [2] M. Naeem, T. Jamal, J. Diaz-Martinez, S. A. Butt, N. Montesano, M. I. Tariq, E. De-la Hoz-Franco, and E. De-La-Hoz-Valdiris, “Trends and future perspective challenges in big data,” in *Advances in Intelligent Data Analysis and Applications: Proceeding of the Sixth Euro-China Conference on Intelligent Data Analysis and Applications, 15–18 October 2019, Arad, Romania*, pp. 309–325, Springer, 2022.
- [3] D. Singh and C. K. Reddy, “A survey on platforms for big data analytics,” *Journal of Big Data*, vol. 2, p. 8, Oct 2014.
- [4] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, “Trends in big data analytics,” *Journal of Parallel and Distributed Computing*, vol. 74, no. 7, pp. 2561–2573, 2014. Special Issue on Perspectives on Parallel and Distributed Processing.
- [5] C. Samitsch, *Introduction*, pp. 1–3. Wiesbaden: Springer Fachmedien Wiesbaden, 2015.
- [6] Z. Obermeyer and E. J. Emanuel, “Predicting the future-big data, machine learning, and clinical medicine,” *New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016.
- [7] Y. Sun, Y. Shi, and Z. Zhang, “Finance big data: Management, analysis, and applications,” *International Journal of Electronic Commerce*, vol. 23, no. 1, pp. 9–11, 2019.

- [8] G. Zhao, D. X. Jiang, K. Li, and J. H. Diao, "Data mining for fault diagnosis and machine learning for rotating machinery," in *Key Engineering Materials*, vol. 293, pp. 175–182, Trans Tech Publ, 2005.
- [9] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131–156, 1997.
- [10] I. T. Jolliffe, *Principal component analysis*. John Wiley & Sons, 2002.
- [11] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [12] A. E. Eiben and J. E. Smith, *Introduction to evolutionary computing*. Springer, 2015.
- [13] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," *bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.
- [14] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proceedings of the 20th international conference on machine learning (ICML-03)*, pp. 856–863, 2003.
- [15] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009.
- [16] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.
- [17] Y. Masoudi-Sobhanzadeh, H. Motieghader, and A. Masoudi-Nejad, "Featureselect: a software for feature selection based on machine learning approaches," *BMC bioinformatics*, vol. 20, no. 1, pp. 1–17, 2019.
- [18] Y. Li, C.-Y. Chen, and W. W. Wasserman, "Deep feature selection: theory and application to identify enhancers and promoters," *Journal of Computational Biology*, vol. 23, no. 5, pp. 322–336, 2016.
- [19] 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.

- [20] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [21] L. Breiman, “Random forests,” *Machine learning*, vol. 45, pp. 5–32, 2001.
- [22] V. Bolón-Canedo, N. Sánchez-Marroño, and A. Alonso-Betanzos, “Feature selection for high-dimensional data,” *Progress in Artificial Intelligence*, vol. 5, pp. 65–75, 2016.
- [23] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.
- [24] T. Hasanin, T. M. Khoshgoftaar, J. Leevy, and N. Seliya, “Investigating random undersampling and feature selection on bioinformatics big data,” in *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, pp. 346–356, IEEE, 2019.
- [25] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [26] L. Gao, J. Song, X. Liu, J. Shao, J. Liu, and J. Shao, “Learning in high-dimensional multimedia data: the state of the art,” *Multimedia Systems*, vol. 23, pp. 303–313, 2017.
- [27] R. E. Neapolitan *et al.*, *Learning bayesian networks*, vol. 38. Pearson Prentice Hall Upper Saddle River, 2004.
- [28] Y. Fan, J. Liu, P. Liu, Y. Du, W. Lan, and S. Wu, “Manifold learning with structured subspace for multi-label feature selection,” *Pattern Recognition*, vol. 120, p. 108169, 2021.
- [29] F. Moslehi and A. Haeri, “A novel hybrid wrapper–filter approach based on genetic algorithm, particle swarm optimization for feature subset selection,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, pp. 1105–1127, 2020.
- [30] D. Arthur and S. Vassilvitskii, “k-means++: The advantages of careful seeding,” tech. rep., Stanford, 2006.

- [31] D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in neural information processing systems*, vol. 13, 2000.
- [32] J. Sánchez-Monedero, M. Pérez-Ortiz, A. Saez, P. A. Gutiérrez, and C. Hervás-Martínez, "Partial order label decomposition approaches for melanoma diagnosis," *Applied Soft Computing*, vol. 64, pp. 341–355, 2018.
- [33] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on knowledge and data engineering*, vol. 17, no. 4, pp. 491–502, 2005.
- [34] X. Yu, A. Aouari, R. F. Mansour, and S. Su, "A hybrid algorithm based on pso and ga for feature selection," *Journal of Cybersecurity*, vol. 3, no. 2, p. 117, 2021.
- [35] P. Stephan, T. Stephan, R. Kannan, and A. Abraham, "A hybrid artificial bee colony with whale optimization algorithm for improved breast cancer diagnosis," *Neural Computing and Applications*, vol. 33, no. 20, pp. 13667–13691, 2021.
- [36] B. Mao, Z. Xie, Y. Wang, H. Handroos, H. Wu, and S. Shi, "A hybrid differential evolution and particle swarm optimization algorithm for numerical kinematics solution of remote maintenance manipulators," *Fusion Engineering and Design*, vol. 124, pp. 587–590, 2017.
- [37] A. Dey, S. Chattopadhyay, P. K. Singh, A. Ahmadian, M. Ferrara, and R. Sarkar, "A hybrid meta-heuristic feature selection method using golden ratio and equilibrium optimization algorithms for speech emotion recognition," *IEEE Access*, vol. 8, pp. 200953–200970, 2020.
- [38] K. Lenin, B. R. Reddy, and M. Suryakalavathi, "Hybrid tabu search-simulated annealing method to solve optimal reactive power problem," *International Journal of Electrical Power & Energy Systems*, vol. 82, pp. 87–91, 2016.
- [39] J. Seo, S. Kim, S. Lee, H. Jeong, T. Kim, and J. Kim, "Data-driven approach to predicting the energy performance of residential buildings using minimal input data," *Building and Environment*, vol. 214, p. 108911, 2022.
- [40] S. A. Ebiaredoh-Mienye, E. Esenogho, and T. G. Swart, "Integrating enhanced sparse autoencoder-based artificial neural network technique and softmax regression for medical diagnosis," *Electronics*, vol. 9, no. 11, p. 1963, 2020.

- [41] F.-Z. El-Alami, A. El Mahdaouy, S. O. El Alaoui, and N. En-Nahnahi, "A deep autoencoder-based representation for arabic text categorization," *Journal of Information and Communication Technology*, vol. 19, no. 3, pp. 381–398, 2020.
- [42] P. M. Tuan, T.-L. Phan, M. Adel, E. Guedj, and N. L. Trung, "Autoencoder-based feature ranking for alzheimer disease classification using pet image," *Machine Learning with Applications*, vol. 6, p. 100184, 2021.
- [43] X. Xu, H. Gu, Y. Wang, J. Wang, and P. Qin, "Autoencoder based feature selection method for classification of anticancer drug response," *Frontiers in genetics*, vol. 10, p. 233, 2019.
- [44] K. Manohar and E. Logashanmugam, "Hybrid deep learning with optimal feature selection for speech emotion recognition using improved meta-heuristic algorithm," *Knowledge-Based Systems*, vol. 246, p. 108659, 2022.
- [45] M. Nasser, N. Salim, H. Hamza, and F. Saeed, "Deep belief network for molecular feature selection in ligand-based virtual screening," in *Recent Trends in Data Science and Soft Computing: Proceedings of the 3rd International Conference of Reliable Information and Communication Technology (IRICT 2018)*, pp. 3–14, Springer, 2019.
- [46] J. Smolander, M. Dehmer, and F. Emmert-Streib, "Comparing deep belief networks with support vector machines for classifying gene expression data from complex disorders," *FEBS Open Bio*, vol. 9, no. 7, pp. 1232–1248, 2019.
- [47] Z. He and W. Yu, "Stable feature selection for biomarker discovery," *Computational biology and chemistry*, vol. 34, no. 4, pp. 215–225, 2010.
- [48] A. Aliper, S. Plis, A. Artemov, A. Ulloa, P. Mamoshina, and A. Zhavoronkov, "Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data," *Molecular pharmaceuticals*, vol. 13, no. 7, pp. 2524–2530, 2016.
- [49] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, "Spotfake: A multi-modal framework for fake news detection," in *2019 IEEE fifth international conference on multimedia big data (BigMM)*, pp. 39–47, IEEE, 2019.

- [50] X. Chen, J. Ji, T. Ji, and P. Li, "Cost-sensitive deep active learning for epileptic seizure detection," in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 226–235, 2018.
- [51] I. Ahmad, X. Wang, M. Zhu, C. Wang, Y. Pi, J. A. Khan, S. Khan, O. W. Samuel, S. Chen, and G. Li, "Eeg-based epileptic seizure detection via machine/deep learning approaches: A systematic review," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [52] M. Gendreau and J.-Y. Potvin, "Metaheuristics in combinatorial optimization," *Annals of Operations Research*, vol. 140, pp. 189–213, 2005.
- [53] I. Fister Jr, X.-S. Yang, I. Fister, J. Brest, and D. Fister, "A brief review of nature-inspired algorithms for optimization," *arXiv preprint arXiv:1307.4186*, 2013.
- [54] A. K. Sangaiah, Z. Zhiyong, M. Sheng, *et al.*, *Computational intelligence for multimedia big data on the cloud with engineering applications*. Academic Press, 2018.
- [55] L. Wen, Q. Yin, and P. Guo, "Ant colony optimization algorithm for feature selection and classification of multispectral remote sensing image," in *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, vol. 2, pp. II–923, IEEE, 2008.
- [56] D. Bertsimas and J. Tsitsiklis, "Simulated annealing," *Statistical science*, vol. 8, no. 1, pp. 10–15, 1993.
- [57] Y. Zhang and Z. Jin, "Group teaching optimization algorithm: A novel meta-heuristic method for solving global optimization problems," *Expert Systems with Applications*, vol. 148, p. 113246, 2020.
- [58] F. Glover and M. Laguna, *Tabu search*. Springer, 1998.
- [59] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, vol. 4, pp. 1942–1948, IEEE, 1995.
- [60] V. P. Kour and S. Arora, "Particle swarm optimization based support vector machine (p-svm) for the segmentation and classification of plants," *IEEE Access*, vol. 7, pp. 29374–29385, 2019.

- [61] Y. Ma, A. Zhang, L. Yang, C. Hu, and Y. Bai, "Investigation on optimization design of offshore wind turbine blades based on particle swarm optimization," *Energies*, vol. 12, no. 10, p. 1972, 2019.
- [62] J. Luo, D. Zhou, L. Jiang, and H. Ma, "A particle swarm optimization based multiobjective memetic algorithm for high-dimensional feature selection," *Memetic Computing*, vol. 14, no. 1, pp. 77–93, 2022.
- [63] H. Jh, "Adaptation in natural and artificial systems," *Ann Arbor*, 1975.
- [64] Z. Wei, M. Li, Q. Zhao, Z. Lyu, S. Zhu, and Z. Wei, "Multi-mc charging schedule algorithm with time windows in wireless rechargeable sensor networks," *IEEE Access*, vol. 7, pp. 156217–156227, 2019.
- [65] O. H. Babatunde, L. Armstrong, J. Leng, and D. Diepeveen, "A genetic algorithm-based feature selection," 2014.
- [66] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in engineering software*, vol. 69, pp. 46–61, 2014.
- [67] D. Liu, M. Li, Y. Ji, Q. Fu, M. Li, M. A. Faiz, S. Ali, T. Li, S. Cui, and M. I. Khan, "Spatial-temporal characteristics analysis of water resource system resilience in irrigation areas based on a support vector machine model optimized by the modified gray wolf algorithm," *Journal of Hydrology*, vol. 597, p. 125758, 2021.
- [68] M. Guezgouz, J. Jurasz, B. Bekkouche, T. Ma, M. S. Javed, and A. Kies, "Optimal hybrid pumped hydro-battery storage scheme for off-grid renewable energy systems," *Energy Conversion and Management*, vol. 199, p. 112046, 2019.
- [69] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in engineering software*, vol. 95, pp. 51–67, 2016.
- [70] M. Mafarja and S. Mirjalili, "Whale optimization approaches for wrapper feature selection," *Applied Soft Computing*, vol. 62, pp. 441–453, 2018.
- [71] M. Khishe and M. Mosavi, "Chimp optimization algorithm," *Expert Systems with Applications*, vol. 149, p. 113338, 2020.
- [72] J. Wang, M. Khishe, M. Kaveh, and H. Mohammadi, "Binary chimp optimization algorithm (bchoa): a new binary meta-heuristic for solving optimization problems," *Cognitive Computation*, vol. 13, pp. 1297–1316, 2021.



- [73] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, and A. H. Gandomi, "The arithmetic optimization algorithm," *Computer methods in applied mechanics and engineering*, vol. 376, p. 113609, 2021.
- [74] R. Sindhu, R. Ngadiran, Y. M. Yacob, N. A. H. Zahri, and M. Hariharan, "Sine-cosine algorithm for feature selection with elitism strategy and new updating mechanism," *Neural Computing and Applications*, vol. 28, pp. 2947–2958, 2017.
- [75] S. Mirjalili, "Sca: a sine cosine algorithm for solving optimization problems," *Knowledge-based systems*, vol. 96, pp. 120–133, 2016.
- [76] Y. Wan, M. Wang, Z. Ye, and X. Lai, "A feature selection method based on modified binary coded ant colony optimization algorithm," *Applied Soft Computing*, vol. 49, pp. 248–258, 2016.
- [77] J. Jona and N. Nagaveni, "Ant-cuckoo colony optimization for feature selection in digital mammogram.," *Pakistan journal of biological sciences: PJBS*, vol. 17, no. 2, pp. 266–271, 2014.
- [78] P. Moradi and M. Gholampour, "A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy," *Applied Soft Computing*, vol. 43, pp. 117–130, 2016.
- [79] M. Mafarja and S. Abdullah, "Investigating memetic algorithm in solving rough set attribute reduction," *International Journal of Computer Applications in Technology*, vol. 48, no. 3, pp. 195–202, 2013.
- [80] T. Stützle and H. H. Hoos, "Max-min ant system," *Future generation computer systems*, vol. 16, no. 8, pp. 889–914, 2000.
- [81] T. Cai, F. Pan, and J. Chen, "Adaptive particle swarm optimization algorithm," in *Fifth World Congress on Intelligent Control and Automation (IEEE Cat. No. 04EX788)*, vol. 3, pp. 2245–2247, IEEE, 2004.
- [82] H. A. e Oliveira, L. Ingber, A. Petraglia, M. R. Petraglia, and M. A. S. Machado, "Stochastic global optimization and its applications with fuzzy adaptive simulated annealing-volume 35," 2014.
- [83] N. J. Cheung, X.-M. Ding, and H.-B. Shen, "Adaptive firefly algorithm: parameter analysis and its application," *PloS one*, vol. 9, no. 11, p. e112634, 2014.

- [84] M. Mahdavi, M. Fesanghary, and E. Damangir, "An improved harmony search algorithm for solving optimization problems," *Applied mathematics and computation*, vol. 188, no. 2, pp. 1567–1579, 2007.
- [85] H. A. Hefny and S. S. Azab, "Chaotic particle swarm optimization," in *2010 The 7th International Conference on Informatics and Systems (INFOS)*, pp. 1–8, IEEE, 2010.
- [86] P. Snaselova and F. Zboril, "Genetic algorithm using theory of chaos," *Procedia computer science*, vol. 51, pp. 316–325, 2015.
- [87] H. Bingol and B. Alatas, "Chaotic league championship algorithms," *Arabian journal for science and engineering*, vol. 41, pp. 5123–5147, 2016.
- [88] J. Kennedy, "Bare bones particle swarms," in *Proceedings of the 2003 IEEE Swarm Intelligence Symposium. SIS'03 (Cat. No. 03EX706)*, pp. 80–87, IEEE, 2003.
- [89] H. Wang, S. Rahnamayan, H. Sun, and M. G. Omran, "Gaussian bare-bones differential evolution," *IEEE Transactions on Cybernetics*, vol. 43, no. 2, pp. 634–647, 2013.
- [90] F. Zou, L. Wang, X. Hei, D. Chen, Q. Jiang, and H. Li, "Bare-bones teaching-learning-based optimization," *The Scientific World Journal*, vol. 2014, 2014.
- [91] Z. Guo, W. K. Wong, Z. Li, and P. Ren, "Modeling and pareto optimization of multi-objective order scheduling problems in production planning," *Computers & Industrial Engineering*, vol. 64, no. 4, pp. 972–986, 2013.
- [92] M. Xiao, K. Cai, and H. A. Abbass, "Hybridized encoding for evolutionary multi-objective optimization of air traffic network flow: A case study on china," *Transportation Research Part E: Logistics and Transportation Review*, vol. 115, pp. 35–55, 2018.
- [93] S. Wang and S. Ma, "Efficient methods for a bi-objective nursing home location and allocation problem: A case study," *Applied Soft Computing*, vol. 65, pp. 280–291, 2018.
- [94] E. Zhang, Y. Wu, and Q. Chen, "A practical approach for solving multi-objective reliability redundancy allocation problems using extended bare-bones particle

- swarm optimization,” *Reliability Engineering & System Safety*, vol. 127, pp. 65–76, 2014.
- [95] R. Zhang, P.-C. Chang, S. Song, and C. Wu, “Local search enhanced multi-objective pso algorithm for scheduling textile production processes with environmental considerations,” *Applied Soft Computing*, vol. 61, pp. 447–467, 2017.
- [96] B. Qian, L. Wang, D.-x. Huang, W.-l. Wang, and X. Wang, “An effective hybrid de-based algorithm for multi-objective flow shop scheduling with limited buffers,” *Computers & Operations Research*, vol. 36, no. 1, pp. 209–233, 2009.
- [97] E. K. Burke, P. De Causmaecker, G. De Maere, J. Mulder, M. Paelinck, and G. V. Berghe, “A multi-objective approach for robust airline scheduling,” *Computers & Operations Research*, vol. 37, no. 5, pp. 822–832, 2010.
- [98] M. Khalili, “A multi-objective electromagnetism algorithm for a bi-objective hybrid no-wait flowshop scheduling problem,” *The International Journal of Advanced Manufacturing Technology*, vol. 70, no. 9-12, pp. 1591–1601, 2014.
- [99] L. Rere, M. I. Fanany, and A. M. Arymurthy, “Metaheuristic algorithms for convolution neural network,” *Computational intelligence and neuroscience*, vol. 2016, 2016.
- [100] A. H. Gandomi, X.-S. Yang, and A. H. Alavi, “Mixed variable structural optimization using firefly algorithm,” *Computers & Structures*, vol. 89, no. 23-24, pp. 2325–2336, 2011.
- [101] V. Kesavan, R. Kamalakannan, R. Sudhakarapandian, and P. Sivakumar, “Heuristic and meta-heuristic algorithms for solving medium and large scale sized cellular manufacturing system np-hard problems: A comprehensive review,” *Materials today: proceedings*, vol. 21, pp. 66–72, 2020.
- [102] R. Guha, A. H. Khan, P. K. Singh, R. Sarkar, and D. Bhattacharjee, “Cga: A new feature selection model for visual human action recognition,” *Neural Computing and Applications*, vol. 33, pp. 5267–5286, 2021.
- [103] S. Liu, Q. Lin, J. Li, and K. C. Tan, “A survey on learnable evolutionary algorithms for scalable multiobjective optimization,” *IEEE Transactions on Evolutionary Computation*, 2023.

- 
- [104] P. Agrawal, T. Ganesh, and A. W. Mohamed, “Chaotic gaining sharing knowledge-based optimization algorithm: an improved metaheuristic algorithm for feature selection,” *Soft Computing*, vol. 25, no. 14, pp. 9505–9528, 2021.
- [105] “Uci archives.” <https://archive-beta.ics.uci.edu/>. [Accessed March 29, 2023].

## Turnitin Originality Report

A Hybrid Sine-Cosine Binary Chimp Optimization Algorithm for Feature Selection Applications by Hassan Mohyuddin



From Ms Theses (CUST Library)

- Processed on 23-Nov-2023 17:46 PKT
- ID: 2236896617
- Word Count: 18174

Similarity Index

8%

Similarity by Source

Internet Sources:

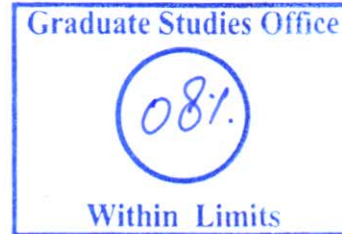
4%

Publications:

5%

Student Papers:

1%



## sources:

- 1 1% match (Elnaz Pashaei, Elham Pashaei. "An efficient binary chimp optimization algorithm for feature selection in biomedical data classification", Neural Computing and Applications, 2022)  
[Elnaz Pashaei, Elham Pashaei. "An efficient binary chimp optimization algorithm for feature selection in biomedical data classification". Neural Computing and Applications, 2022](#)
  - 2 < 1% match (student papers from 31-May-2016)  
[Submitted to Skolkovo Institute of Science and Technology \(Skoltech\) on 2016-05-31](#)
  - 3 < 1% match (student papers from 08-Jan-2013)  
[Submitted to University of Ulster on 2013-01-08](#)
  - 4 < 1% match (Internet from 04-Oct-2022)  
[https://www.researchgate.net/publication/259440613\\_Intriguing\\_properties\\_of\\_neural\\_networks](https://www.researchgate.net/publication/259440613_Intriguing_properties_of_neural_networks)
  - 5 < 1% match (Internet from 03-Feb-2023)  
[https://www.researchgate.net/publication/330759145\\_Task\\_scheduling\\_in\\_cloud\\_computing\\_based\\_on\\_hybrid\\_moth\\_search\\_algorithm\\_a](https://www.researchgate.net/publication/330759145_Task_scheduling_in_cloud_computing_based_on_hybrid_moth_search_algorithm_a)
  - 6 < 1% match (Internet from 30-Jan-2023)  
[https://www.researchgate.net/publication/332951385\\_Feature\\_Selection\\_and\\_Feature\\_Extraction\\_in\\_Pattern\\_Analysis\\_A\\_Literature\\_Rev](https://www.researchgate.net/publication/332951385_Feature_Selection_and_Feature_Extraction_in_Pattern_Analysis_A_Literature_Rev)
  - 7 < 1% match (Internet from 06-Feb-2023)  
[https://www.researchgate.net/publication/286834616\\_Moving\\_to\\_outsourcing\\_in\\_production\\_scheduling\\_Quantitative\\_approaches](https://www.researchgate.net/publication/286834616_Moving_to_outsourcing_in_production_scheduling_Quantitative_approaches)
  - 8 < 1% match (Internet from 24-Feb-2023)  
[https://www.researchgate.net/publication/302948808\\_A\\_Computational\\_Intelligence\\_Method\\_for\\_Effective\\_Diagnosis\\_of\\_Heart\\_Disease](https://www.researchgate.net/publication/302948808_A_Computational_Intelligence_Method_for_Effective_Diagnosis_of_Heart_Disease)
  - 9 < 1% match (Internet from 30-Jan-2023)  
[https://www.researchgate.net/publication/322729054\\_Adaptive\\_scheduling\\_of\\_cloud\\_tasks\\_using\\_ant\\_colony\\_optimization](https://www.researchgate.net/publication/322729054_Adaptive_scheduling_of_cloud_tasks_using_ant_colony_optimization)
  - 10 < 1% match ("Advances in Intelligent Manufacturing and Mechatronics", Springer Science and Business Media LLC, 2023)  
["Advances in Intelligent Manufacturing and Mechatronics". Springer Science and Business Media LLC, 2023](#)
  - 11 < 1% match (Qi Liu, Xiaofeng Li, Haitao Liu, Zhaoxia Guo. "Multi-objective metaheuristics for discrete optimization problems: A review of the state-of-the-art", Applied Soft Computing, 2020)  
[Qi Liu, Xiaofeng Li, Haitao Liu, Zhaoxia Guo. "Multi-objective metaheuristics for discrete optimization problems: A review of the state-of-the-art". Applied Soft Computing, 2020](#)
  - 12 < 1% match (Internet from 19-May-2022)  
[https://link.springer.com/article/10.1007/s00500-021-06229-8?code=18a7b0e3-b6da-4084-8db1-a702f7019896&error=cookies\\_not\\_supported](https://link.springer.com/article/10.1007/s00500-021-06229-8?code=18a7b0e3-b6da-4084-8db1-a702f7019896&error=cookies_not_supported)
- < 1% match (Internet from 13-Aug-2020)