PREDICTING DIAGNOSTIC TESTS AND PROBABLE DISEASE BY UTILIZING META HEURISTIC OPTIMIZATION

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Computer Science and Engineering

By

Priyanka Shivaprasad More

Registration Number: 41900136

Supervised By

DR Baljit Singh Saini (22078)

Computer Science and Engineering

(Professor)

Lovely Professional University

Co-Supervised by

Dr Rakesh Kumar Sharma

Obstetrics and Gynecology

(Professor)

D. Y. Patil Medical College

LOVELY PROFESSIONAL UNIVERSITY

2024

DECLARATION

I, hereby declared that the presented work in the thesis entitled "**Predicting Diagnostic Tests and Probable Disease by Utilizing Meta Heuristic Optimizations**" in fulfilment of degree of **Doctor of Philosophy (Ph. D.)** is outcome of research work carried out by me under the supervision of **Dr Baljit Singh Saini**, working as a Professor, in the School of Computer Science and Engineering of Lovely Professional University, Punjab, India, and **Dr Rakesh Kumar Sharma**, working as Dean and Professor in the department of Obstetrics and Gynecology of D Y Patil medical college, Kolhapur, Maharashtra. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of another investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

Name of the scholar: **Priyanka Shivaprasad More**

Registration No.: **41900136**

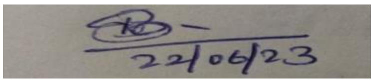
Department: School of Computer Science and Engineering

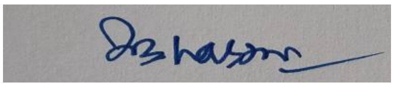
Lovely Professional University,

Punjab, India.

CERTIFICATE

This is to certify that the work reported in the Ph. D. thesis entitled "**Predicting Diagnostic Tests and Probable Disease by Utilizing Meta Heuristic Optimizations**" submitted in fulfillment of the requirement for the reward of degree of **Doctor of Philosophy (Ph.D.)** in the School of Computer Science and Engineering, is a research work carried out by **Priyanka Shivaprasad More (41900136)**, is bonafide record of her original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.





Supervisor: Dr Baljit Singh Saini

Designation: Professor

School: Computer Science and Engineering

University: Lovely professional University, Punjab.

Co-Supervisor: Dr Rakesh Kumar Sharma

Designation: Professor

Department: Obstetrics and Gynecology

University: D. Y. Patil Medical College, Kolhapur, Maharashtra.

ABSTRACT

The World Health Organization (WHO) proposes a doctor-to-population correlation to be 1:1000, meaning one doctor should be available for every 1000 individuals in the population. However, India, with a population of 1.38 billion, has a doctor-to-population ratio of 1:850, which is below the recommended ratio. The COVID-19 pandemic has further weakened India's already fragile healthcare system, causing high levels of stress among healthcare workers due to the overwhelming demand and insufficient resources. Research has shown that doctors experienced high levels of psychological stress during the pandemic phase, primarily due to an insufficient medical workforce. Additionally, patients in remote areas do not have access to quality healthcare facilities, and have to make multiple visits to complete their treatment. Consequently, there is a need for feasible and cost-effective solutions to address these issues.

There are several ways in which Artificial intelligence (AI) can be used in the healthcare system, including risk stratification of patient populations, medical analysis/prediction, treatment effectiveness, infection rate prediction, and intelligent medicine dispensing. By automating many mundane tasks, AI can assist health professionals in performing their roles better, leading to improved patient outcomes. Hospitals and medical systems that incorporate AI capabilities are essential, challenging, but achievable.

Healthcare records are electronically stored in large datasets due to technological advancements, which can be analyzed to improve the healthcare system. With the help of these datasets, trends in the data can be found, and artificial intelligence (AI) and machine learning (ML) approaches help to create systems that automatically recommend laboratory tests to support current clinical practices. The most critical research question at this stage is how to efficiently store, retrieve, and maintain patient data. Retrieving useful information from the enormous size of medical datasets is challenging. Lately, a

significant research trend is using data analytics systems to validate assumptions and identify intriguing patterns in a significant volume of data from national or regional medical centers. Additionally, metaheuristic-based data mining algorithms can link symptoms and specific diseases, as well as the cause and effect of a person's behavior on their health. Incorrect laboratory tests can result in incorrect, overlooked, or delayed patient diagnoses. However, policymakers and clinicians often overlook the importance of ordering lab tests correctly.

The objectives of this research are to: 1) implement a system that reduces the workload in the hospital management system, 2) develop a system that reduces the workload of doctors, 3) reduce the number of patient visits to the hospital to avoid the risk of infection, 4) help patients in remote areas access quality healthcare in a cost-effective manner, and 5) develop a hybrid metaheuristic optimization algorithm that predicts probable diseases efficiently.

ACKNOWLEDGEMENT

I extend my profound gratitude to all those who have accompanied me on my journey towards completing my doctorate in Computer Science and Engineering. This thesis owes its existence to the steadfast support, guidance, and encouragement from various individuals and institutions.

Foremost, I express deep appreciation to my supervisors, Dr Baljit Singh Saini and Dr Rakesh Kumar Sharma, for their exceptional mentorship, valuable feedback, and unwavering dedication. Their expertise has been pivotal in shaping my research's trajectory.

My heartfelt thanks go to Dr D Y Patil Medical College Hospital and Research Institute, Kolhapur, Maharashtra, for their invaluable contribution in providing the dataset that underpinned my research in artificial intelligence within the healthcare domain. Their commitment to collaboration enriched the depth and scope of my study.

The high-quality dataset from Dr D Y Patil Medical College Hospital and Research Institute played a crucial role in validating hypotheses in this thesis. It not only facilitated model testing but also opened avenues for novel exploration. I appreciate Dr Shimpa Sharma, Dr Rajesh Khyalappa, and Dr Jasmin Nath at Dr D Y Patil Hospital for their support, discussions, and proactive assistance during data acquisition. Their expertise ensured dataset relevance and accuracy, reflecting true interdisciplinary cooperation.

Lastly, my family deserves heartfelt acknowledgment for their unwavering love, encouragement, and sacrifices. Their belief in me provided the strength to overcome challenges. In conclusion, this thesis signifies not only my personal achievement but a collective endeavour involving numerous contributors who propelled my growth and success. Your integral roles in this remarkable journey are deeply appreciated.

Place: LPU, Phagwara.

Date: 10/09/2024 (Priyanka Shivaprasad More)

CONTENTS

List of Tables

List of Figures

Abbreviations

ACO	Ant Colony Optimization
ABC	Artificial Bee Colony
AI	Artificial Intelligence
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
CS	Cuckoo Search
DL	Deep Learning
DE	Differential Evaluation
DALY	Disability-Adjusted Life Years
EHR	Electronic Health Records
ELM-BA	Extreme Learning Machine with Bat Algorithm
FFNN	Feed Forward Neural Network
FA	Firefly Algorithm
FCM	Fuzzy C-Means
GOBL	Generalized Opposition-Based Learning
GSO	Glow-worm Swarm Optimization
GSA	Gravitational Search Algorithm
GWO	Grey Wolf Optimizer
HCAI	Healthcare Associated Infections
IPC	Infection Prevention and Control
ICU	Intensive Care Units
LMIC	Lower- and Middle-Income Counties
ML	Machine Learning
MRI	Magnetic resonance imaging
CT	Magnetic resonance imaging
MFO	Moth Flame Optimization
NLP	Natural Language Processing
PSO	Particle Swarm Optimization
PID	Pima Indians Diabetes

PET	Positron Emission Tomography
PCA	Principle Component Analysis
SVM	Support Vector Machine
SI	Swarm Intelligence
UCL	University College London
WOA	Whale Optimization Algorithm
WHO	World Health Organization

Chapter 1

Introduction

1.1 Overview

The healthcare industry faces unprecedented challenges compounded by the scarcity of healthcare professionals worldwide. The shortage of healthcare workers has profound implications, not only for the professionals themselves but also for patient care outcomes and the overall functionality of healthcare systems. This scarcity places immense pressure on healthcare professionals, resulting in increased workload burdens, heightened stress levels, and a higher risk of burnout. Consequently, these factors can compromise the quality of patient care and adversely impact the health.

Moreover, the risk of healthcare-associated infections (HAIs) further complicates the healthcare landscape. HAIs pose significant threats to patient safety and public health, leading to increased morbidity, mortality, and healthcare costs. The shortage of healthcare professionals exacerbates the risk of HAIs by straining resources, compromising infection control measures, and increasing the likelihood of lapses in hygiene protocols.

In response to these challenges, there is huge passion in using artificial intelligence (AI) to alleviate the strain on healthcare professionals and optimize healthcare delivery processes. AI holds promise in revolutionizing various aspects of healthcare, offering opportunities to streamline workflows, improve diagnostic accuracy, enhance treatment planning, and ultimately, enhance patient outcomes. By automating routine tasks and augmenting decision-making processes, AI technologies have the potential to alleviate the workload burden on healthcare professionals and mitigate the risk of burnout.

Furthermore, the incorporation of High-Performance Computing (HPC) in AI applications enables the processing of vast amounts of healthcare data at unprecedented speeds, facilitating real-time decision-making and enabling personalized medicine approaches. HPC infrastructure empowers AI algorithms to analyze complex datasets, ranging from electronic health records to medical imaging scans, with greater efficiency and accuracy.

In addition to leveraging AI and HPC technologies, the integration of metaheuristic optimization algorithms presents a novel approach to addressing healthcare-related challenges. Metaheuristic optimization techniques, offer versatile tools for solving complex optimization problems across various domains, including healthcare. The present era of medicine is deeply intertwined with advancements in computing infrastructure and state-of-the-art technologies, which play a pivotal role in supporting diverse applications to aid human activities. These applications span a wide spectrum, encompassing surgical assistance, medicinal testing and composition, and utilization of variety of instruments for education purposes at different universities [1]. Across the medical domain, diverse computing tools find application in conducting examinations of the human body. Employing computer-aided automated processes for evaluating various diseases not only enhances performance but also facilitates superior treatment outcomes. [2]. The healthcare domain generates and accumulates immense volumes of data, necessitating robust mechanisms for efficient storage in clinic-associated databases. Consequently, there is a critical imperative to devise methodologies capable of effectively accessing and extracting pertinent information from this expansive dataset. This extracted data holds significant potential in facilitating accurate diagnoses and targeted treatment strategies for a diverse range of patient conditions. [3]

In the context of healthcare, metaheuristic optimization algorithms can be applied to optimize resource allocation, scheduling procedures, and treatment planning, among other tasks. By efficiently exploring solution spaces and identifying optimal solutions, these algorithms have the potential to enhance operational efficiency, improve resource utilization, and ultimately, enhance patient care delivery. A multitude of researchers are dedicating their efforts to automating disease detection and prediction with the help of different ML techniques. This concerted endeavour aims to harness the potential of these algorithms to effectively enhance treatment strategies across a broad range of medical conditions. [4]. Researchers are presently employing a diverse array of machine learning and data mining algorithms within the domain of disease diagnosis [5]. It is clear that a variety of methods such as “data mining, AI, and meta Heuristics” are used to build an intelligent and computerized system for disease diagnosis [6].

This thesis explores the intersection of AI, HPC, and metaheuristic optimization algorithms to address the challenges caused by shortage of healthcare professionals and the potential risks of infections linked to healthcare. After a thorough analysis of current research articles, case studies, also empirical analyses, this research aims to elucidate the potential applications of these technologies in optimizing healthcare delivery processes, mitigating the workload burden on healthcare professionals, reducing the risk of HAIs, and improving patient care outcomes.

By examining the synergistic integration of AI, HPC, and metaheuristic optimization algorithms in healthcare settings, this thesis seeks to contribute to the advancement of innovative solutions that address the evolving needs of healthcare systems and improve the quality and efficiency of patient care delivery. Through interdisciplinary collaboration and technological innovation, we endeavor to pave the way towards a more sustainable and resilient healthcare ecosystem that prioritizes the well-being of both healthcare professionals and patients alike.

1.2 Scarcity of healthcare professionals

Healthcare workers represent a critical component of the global workforce, playing an indispensable role while providing medical care and the preservation and enhancement of community health through various interventions such as health promotion and attentive patient care (WHO, 2021). Despite their pivotal contribution, healthcare professionals often face challenges stemming from the lack of recognition and the oversight of factors affecting their occupational well-being.

The scarcity of healthcare workers is a pressing issue observed not only in developing nations but also in affluent and developed nations, as per documented of the World Health Organization's 2021 report on the health workforce. The pandemic caused by Corona virus accentuated this shortage, as healthcare personnel grappled with illness due to the surge in patient numbers and exposure to the virus. The dwindling numbers of healthcare workers have profound implications for healthcare systems, impacting the delivery of quality care, disease prevention efforts, and health promotion initiatives.

Multiple factors contribute to the persistent shortage of healthcare workers. Notable among these are the migration patterns of nurses from low- to high-income countries in pursuit of better career prospects, an aging health workforce characterized by a higher number of retirees than incoming graduates, and the escalating healthcare demands of an aging population. Furthermore, disillusionment with the profession prompts some healthcare workers, particularly nurses, to exit the workforce prematurely. Additionally, challenges such as the conflict between career and family obligations faced by female healthcare workers, coupled with instances of violence in healthcare settings leading to emotional or physical trauma, exacerbate the already stressful work environment.

Projections indicate that the shortage of healthcare personnel will escalate, potentially reaching 18 million till 2030, especially in nations with poor and lower middle incomes. Healthcare workers are indispensable for the systematic functioning of healthcare systems and the provision of essential care to populations in need. Contributing factors to the persistent shortage include inadequate investments in healthcare education and training programs, coupled with a mismatch between educational curricula and workforce demands within healthcare systems. Furthermore, the phenomenon of international migration of healthcare workers from low- and middle-income countries exacerbates the shortage crisis.

The shortage of healthcare personnel places immense strain on those actively engaged in healthcare delivery, often resulting in extended and strenuous working hours. The mental and physical wellness of healthcare professionals must come first in mitigating burnout and ensuring sustained workforce productivity. However, amidst the backdrop of workforce shortages and additional stressors such as pandemics and labor strikes, there is an urgent need to explore innovative strategies that enable healthcare workers to deliver quality care while safeguarding their overall well-being [7].

In total, 90% of low-income nations do not have an acceptable number of healthcare workers, which is reported as less than 4.4 competent workers per 1000 people, according to the World Health Organisation (WHO). This shortage is believed to contribute to an excessive workload for healthcare providers, potentially compromising the quality of care. The WHO suggests that insufficient staffing may force providers to cut corners, reducing the quality of services. UNICEF has expressed similar worries, saying that excessive workloads could result into compromised medical care. High workloads are thought to decrease provider motivation and ability to deliver appropriate care, potentially explaining the gap between theoretical knowledge and practical application [8].

The shortage of healthcare workers contributes to burnout among healthcare professionals in several ways. When there are not enough nurses to meet patient needs, the existing nursing staff may be required to work longer hours, take on heavier workloads, and deal with increased stress levels. This can lead to physical and emotional exhaustion, feelings of depersonalization towards patients, and a reduced sense of personal accomplishment - all key components of burnout. Additionally, the constant pressure caused by staffing shortages can create a challenging work environment that further exacerbates burnout among healthcare workers. The combination of high demand for care with limited resources can increase job dissatisfaction and contribute to feelings of being overwhelmed or unsupported. In summary, the shortage of nurses intensifies workload pressures on existing staff members which in turn increases their risk for experiencing burnout as highlighted in this article's findings about factors affecting nursing workforce shortages [9].

1.3 Burnout and workloads among healthcare workers

The pandemic caused by Corona virus has affected healthcare workers' mental health and general wellbeing, specifically doctors, in terms of stress, burnout, and workload [10]. The different surveys include data and findings indicating that the pandemic has significantly worsened mental health issues, increasing burnout, exhaustion, and trauma among healthcare workers, including doctors. Factors contributing to this include the overwhelming number of COVID-19 patients, increased work hours, stress, and witnessing colleagues getting sick or dying. Healthcare personnel' high levels of stress, anxiety, tiredness, and burnout were discovered by surveys, indicating the pandemic's negative effects on their mental health. There is a concern about increased rates of depression, anxiety, PTSD, and other mental health conditions among healthcare professionals, underscoring the need for support and intervention to address these challenges [11]. There is a need to address the issues of pre-existing high rates of burnout and mental health concerns in the healthcare workforce for building a more resilient healthcare system for the future [12].

Another study look into how job control affected the relationship between healthcare workers' workload and burnout. The study's main objectives were job control, exhaustion, cynicism, and workload. The study encompassed 352 hospital employees from public hospitals in Italy. Findings underscored the crucial role of job control in moderating the impact of workload on exhaustion, thereby reinforcing the sequential connection from exhaustion to cynicism. The study underscores the importance of organizational management strategies that advocate for job control to alleviate burnout risks among healthcare workers. The results underscore the necessity for implementing preventative measures against burnout and enhancing working conditions in healthcare environments. The model examined in the study addressed the variation in exhaustion and cynicism, emphasizing the significance of reducing stressors and fostering job control to mitigate burnout risks among healthcare professionals [13].

The workforce shortage in healthcare can have a significant impact on patient care. With fewer healthcare workers available, the existing staff may be required to take on heavier workloads, leading to fatigue and increased stress levels. This can result in decreased quality of care, longer wait times for patients, and an increased risk of medical errors. Additionally, the shortage may also lead to difficulties in accessing timely appointments or necessary treatments for patients. Overall, the workforce shortage can strain the healthcare system and compromise patient safety and satisfaction [14].

1.4 Healthcare associated Infections

A second highest cause of death due to "Healthcare-associated infections" (HAIs), put significant burden on rates of morbidity and death. Interestingly, various economies have varying rates of HAIs. For example, 7 patients belongs to developed economies and 10 patients belongs to emerging economies, respectively, get at least one form of HAI for every 100 admissions. Alarmingly, the majority of pathogenic microorganisms have acquired antimicrobial resistance, while the development of novel antimicrobial agents remains limited. Contagious and infectious diseases are the second greatest cause of death for both plants and animals worldwide and provide serious problems. The emergence of epidemics caused by epizootic diseases has been well-documented throughout history, underscoring the persistent threats posed by communicable infectious diseases.

Despite the previous notion of achieving complete control over infectious diseases, they persist as a significant public health concern worldwide. As the foremost cause of both mortality and morbidity, the financial burden associated with their treatment continues to rise. Within the past 35 years, no less than 30 newly emergent contagious and communicable diseases have been identified. In recent times, the profound impact of infectious diseases on the vast and mobile global population has become increasingly evident in socioeconomic, environmental, and ecological domains. This population faces the combined difficulty of facing new infectious agents, as demonstrated by the Coronavirus pandemic, and managing the massive epidemic of existing infectious diseases that are resistant to multiple medications [15].

Healthcare-associated infections (HAIs) contribute significantly to morbidity and mortality rates, representing the second highest cause of death on a global scale [16] [17]. According to reports by researchers, there is an incidence of at least one type of healthcare-associated infection (HCAI) in 8% of patients in affluent economies and 11% of patients in developing nations. Among these affected patients, the mortality rate is reported to be 10% [18]. For instance, within the United States, an assessed 1.75 million individuals contract healthcare-associated infections (HAIs) on an annual basis, yielding a prevalence rate of 4.5%. This concerning statistic further translates to a mortality range of 90,000 to 99,000 individuals impacted by these infections [19]. A separate study conducted in the European Economic Area disclosed an annual incidence of 2,609,911 newly diagnosed cases of healthcare-associated infections (HAIs). This epidemiological burden resulted in a total loss of 2,506,091 Disability-Adjusted Life Years (DALYs) every year, equating to an average of 501 DALYs per 100,000 people in the population [20].

It has been estimated that the rate of HAIs varies between 5.7% and 19.1% in "Lower and Middle Income Countries" (LMIC). Nevertheless, the data pertaining to HAIs in LMIC is characterized by its fragmented nature, primarily attributed to inadequate infrastructures for comprehensive data record-keeping and limited availability of resources[21][22]. The World Health Organization (WHO) conducted a multicentre study to assess the prevalence of healthcare-associated infections (HAIs) specifically in Intensive Care Units (ICUs). According to the results, a significant 51% of individuals hospitalized in intensive care units (ICUs) experienced a health-related infection (HAI), which led to extended hospital stays and heightened susceptibility to further infections and related complications. Infectious diseases account for a staggering 15 million annual deaths globally, with emerging economic nations bearing the brunt, contributing to 95% of these fatalities. The principal causes of these deaths include acute respiratory infections, diarrheal diseases, measles, AIDS, malaria, and tuberculosis. Moreover, the global estimate suggests that over 1.4 million patients concurrently suffer from HAIs in both advanced and emerging countries, imposing significant financial burdens on individuals, communities, and public healthcare systems at large [23]. A significant portion of healthcare-associated infections (HAIs), however, can be mitigated through the effective use of infection prevention and control policies and strategic planning [24].

1.5 AI in healthcare

The AI emerged as promising technology in the healthcare sector, holding the capacity to bring about revolutionary transformations across multiple domains, including enhanced patient outcomes and cost reduction. AI can be broadly classified into two distinct categories: rule-based AI and machine learning (ML)-based AI. Rule-based AI systems operate by following predetermined sets of rules, whereas ML-based AI systems leverage algorithms capable of learning from data.

AI has a wide range of applications in healthcare. Outlined below are several pivotal domains where AI is currently being leveraged:

i. **Medical Imaging**:
   AI is revolutionizing medical imaging by enhancing diagnostic accuracy, improving workflow efficiency, and enabling new applications. According to a study published in The Lancet Digital Health [25], AI algorithms achieved human-level performance in detecting breast cancer from mammograms, with an 'area-under-the-curve' (AUC) of 0.95. Another paper in Nature Medicine [26] reported that an AI system outperformed radiologists in detecting lung cancer from CT scans, with a sensitivity of 94.4% compared to 72.8% for the radiologists. Furthermore, a study in Radiology demonstrated that AI-assisted triage of head CT scans reduced the time to interpret critical findings by 63%, improving patient care. AI also enables novel applications like virtual biopsy, where algorithms can non-invasively characterize tissue properties, as shown in a Science Translational Medicine paper [27]. As AI continues to advance, it holds immense potential to enhance medical imaging, leading to earlier diagnoses, personalized treatments, and improved patient outcomes.

ii. **Disease Diagnosis and Prognosis**:

AI is making significant strides in improving disease diagnosis and prognosis. A study published demonstrated that an AI system could predict cardiovascular risk factors from retinal fundus images with remarkable accuracy, performing better than conventional risk calculation models [28]. Another research found that an AI algorithm could detect Alzheimer's disease an average of 6 years before the final diagnosis, by analyzing language patterns in speech transcripts [29]. Recent research has demonstrated the profound impact of artificial intelligence (AI) on disease diagnosis and prognosis, showcasing its potential to revolutionize healthcare practices. A study published in the Journal of the American Medical Association (JAMA) in 2021 highlighted AI's pivotal role in enhancing diagnostic accuracy and prognostic capabilities across various medical disciplines [30].

iii. **Drug Discovery and Development**:

The integration of AI tools in pharmaceutical companies aims to address challenges in drug development and product lifecycle management, contributing to the rise of startups in the sector. The healthcare industry faces complex issues such as rising drug costs, necessitating significant changes. AI-enabled manufacturing allows for personalized medications tailored to individual patient needs, including dose and release parameters. Adoption of AI technologies accelerates product development timelines, enhances product quality, improves production safety, and optimizes resource utilization, making automation increasingly crucial. Overall, AI's role in pharmaceuticals promises cost-effective solutions that enhance efficiency and address key industry challenges [31]. AI is being used to help speed up the process of drug discovery and development. AI algorithms can analyze large datasets to identify potential drug targets and predict how different drugs will interact with the body [32].

iv. **Personalized Medicine**:

AI plays a pivotal role in the development of personalized medicine by facilitating data analysis and treatment optimization for individual patients. This approach addresses the variability in treatment responses observed in clinical trials and enhances decision-making regarding drug combinations and dosing strategies. Through phenotypic personalized medicine (PPM), AI-driven platforms like the quadratic phenotypic optimization platform (QPOP) and CURATE.AI have shown superior outcomes by minimizing side effects and maximizing efficacy, thereby improving individuals' quality and length of life[33]. Ongoing studies in AI and precision medicine are working towards providing personalized medical information for doctors and patients. Early disease detection and prevention can reduce the overall burden of sickness and healthcare costs for all parties involved in this collaboration. The main objective is to improve health outcomes and encourage early intervention to keep people healthier [34].

v. **Patient Monitoring**:
RPM helps doctors keep an eye on patients with chronic or acute illnesses from afar, like elderly people at home or those in hospitals. Instead of relying on staff time, RPM uses new technologies for monitoring that don't need invasive methods. The study reviews how AI is changing RPM, its challenges, and what might happen in the future[35]. The AI facilitates “remote patient monitoring” (RPM), optimizing hospitalization and complication avoidance, thereby reducing costs. The FDA defines the regulations for AI in medical devices, ensuring the safety, effectiveness, and transparency of AI solutions [36]. RPM and telehealth facilitate quick accessibility to medical data and help provide patients with affordable, excellent treatment [37].

However, there are also some challenges and ethical considerations associated with the use of AI in healthcare. These include privacy concerns, prejudice in AI systems, and the possibility of healthcare workers losing their jobs.

AI has the power to completely transform healthcare by expediting medical research, cutting expenses, and increasing patient outcomes. In addition to the applications mentioned in the previous answer, AI is also being used in healthcare for natural language processing (NLP), predictive analytics, and virtual assistants.

NLP involves the analysis of natural language to extract insights and meaning from unstructured data, such as medical notes and patient records. AI-powered NLP can help healthcare providers to identify patient needs and recommend treatment options based on the information contained in their medical records.

Predictive analytics involves the use of AI algorithms to detect the patterns and correlation in data to predict future health issues. For example, AI has the capability to examine patient data and detect people who have a high possibility of acquiring a certain medical condition. This enables healthcare practitioners to take early action and avoid the occurrence of the illness.

Virtual assistants fall under the AI domain which is useful in healthcare. It offered personalized support and assistance to patients , such as reminders to take medication, information about their condition, and access to healthcare providers.

There are also several challenges associated with the use of AI in healthcare. It involves concerns related to data confidentiality and protection and the likelihood of healthcare workers losing their jobs. To ensure the ethical and successful application of AI in healthcare, it is necessary to address these difficulties.
Recent research in AI and healthcare has focused on a range of applications, including medical imaging, drug discovery, and personalized medicine. For example, a study published in National Library of Medicine in 2023 by Qing Gao et. al. used AI to predict the likelihood of patients with non-small cell lung cancer responding to immunotherapy, a type of cancer treatment. The study found that the AI algorithm was able to accurately predict treatment response in 75% of patients [38].

Another study published by Khanna et al in-Network Modelling Analysis in Health Informatics and Bioinformatics used AI to analyze medical images from patients with COVID-19, identifying patterns and features that could be used to predict disease severity and patient outcomes. The study found that AI was able to accurately predict patient outcomes, such as the need for mechanical ventilation, with a high degree of accuracy [39].

AI is also being used to predict the results of diagnostic tests based on a patient's symptoms and other clinical data. One example of this is the use of ML approaches to predict the early symptoms in patients. In a study published in the journal PLOS ONE, Hasan et al. used machine learning algorithms to predict the likelihood of a patient having type 2 diabetes based on their symptoms and other clinical data [40]. The researchers collected data from electronic health records of patients who had been diagnosed with type 2 diabetes and those who had not. They then used machine learning algorithms to analyze the data and identify patterns that were associated with the disease. The researchers found that the machine learning algorithm was able to accurately predict the likelihood of type 2 diabetes in patients based on their symptoms and other clinical data. This information can help doctors make more informed decisions about diagnostic testing and treatment for patients with suspected type 2 diabetes.

In another study published in the journal Computational Intelligence and Neuroscience, Taher et al. [41] used machine learning algorithms to predict the likelihood of a patient having a particular genetic disorder based on their symptoms and other clinical data. The researchers trained the machine learning algorithm using data from patients who had been diagnosed with the genetic disorder and those who had not. The researchers found that the machine learning algorithm was able to accurately predict the likelihood of the genetic disorder in patients based on their symptoms and other clinical data. This information can help doctors make more informed decisions about genetic testing and treatment for patients with suspected genetic disorders.

Overall, AI has the potential to improve the accuracy and efficiency of diagnostic testing by helpin**g** doctors make more informed decisions based on the analysis of large datasets. However, it's important to note that AI should be used in conjunction with clinical expertise and judgment, rather than as a replacement for it.

1.6 HPC in healthcare

High-performance computing (HPC) technology is improving cardiovascular (CV) science in two main ways: by helping us learn from data and by helping us build new models. Clinical trials, which compare different treatments, are one example of learning from data. In HPC, we use artificial intelligence (AI) models to predict future CV events using large datasets. These models can also help us understand how different factors affect individual risk. Sometimes, AI shows us unexpected connections between data and outcomes, which can help us find new ways to treat CV diseases. By using HPC, we can explore and understand CV diseases better, which could lead to new treatments [42].

The integration of high-performance computing (HPC), high-performance data analytics (HPDA), and AI, known as HPC+, has revolutionized industries like healthcare and pharmaceuticals. This special section of Technology and Health Care examines how HPC+ drives innovation, improves patient outcomes, and accelerates drug discovery. Articles in this issue focus on HPC+'s potential in medical imaging, personalized medicine, drug discovery, and decision support [43].

1.7 Metaheuristic optimization algorithms in healthcare domain

The research community is presently engaged in the integration of a multitude of sensor types for comprehensive data collection. However, this acquisition technique poses intricate challenges in terms of forecasting, decoding, and recognition. Furthermore, diverse metaheuristic algorithms are employed to analyse the data's validity for optimal utilization with data mining algorithms, thereby augmenting the performance of various healthcare algorithms in handling and processing the data [44]. The utilization of metaheuristics in medical services exhibits a wide range of applications, encompassing enhanced classification systems, proficient detection systems, and elevated disease detection rates [45]. The implementation of these techniques has effectively contributed to the optimization of treatments, resulting in a significant reduction in complications experienced by patients enduring protracted diseases. It is worth noting that certain diseases carry the inherent risk of widespread transmission if left undiagnosed for extended periods, potentially endangering a substantial number of individuals. Early detection of such diseases plays a crucial role in pre-emptively mitigating the occurrence of such outbreaks.

Today, many optimization approaches are used to tackle various problems in healthcare. Some algorithms ensure optimal solutions and metaheuristic approaches ensure near optimal solutions. Figure 1.1 shows types of optimization techniques and it is further divided into heuristics and metaheuristic algorithms. The commonly used EA (Evolutionary Algorithms) are Genetic Algorithm (GA), Differential Evolution (DE), flower pollination algorithm (FPA), and Evolutionary strategy and Physical algorithm are Black hole Algorithm (BHA), Gravitational search algorithm (GSA) and Central force algorithm (CFA). Then the commonly used SI (Swarm Intelligence) algorithms are Particle Swarm Optimization (PSO), Krill herd algorithm (KH), Whale optimization Algorithm (WOA), cuckoo Search Algorithm (CSA) and Salp Swarm Algorithm (SSA).

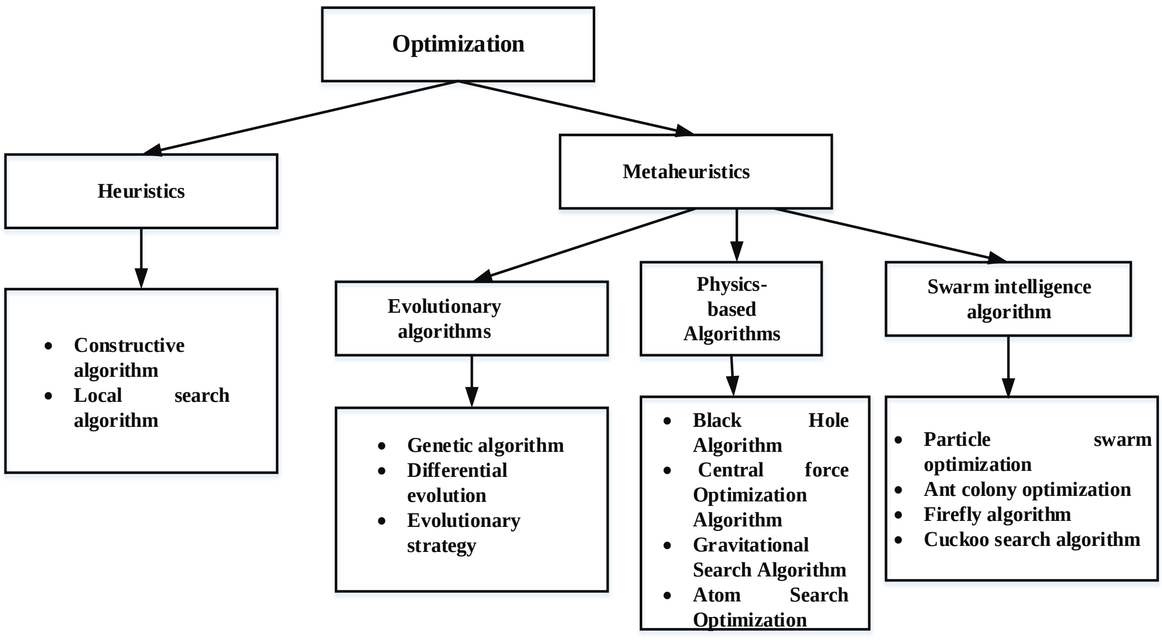


Figure 1.1: Types of optimization techniques

Metaheuristic techniques can be classified into two primary categories: algorithms based on single solutions and algorithms based on population solutions [46]. Based on the approach used to find solutions, the two groups can be identified. Algorithms that rely solely on solutions produce solutions at random until the ideal solution is reached. Population-based algorithms, on the other hand, produce a random set of solutions and iteratively update each solution's value. Iterations lead to increased refinement of the optimal solution [47]. A plethora of innovative metaheuristic algorithms, including but not limited to “Artificial Immune Systems”(AIS), “Lion Optimization” (LO), “Moth Flame Optimization” (MFO), “Cat Swarm Optimization” (CSO), “Particle Swarm Optimization” (PSO) etc., prove to be immensely valuable in facilitating feature selection and extraction for diverse disease diagnosis and early detection applications [48][49].

The application of metaheuristic optimisation algorithms in healthcare to tackle challenging issues has been growing, ranging from predicting clinical outcomes, optimizing treatment plans, and reducing hospital costs. Metaheuristic optimization algorithms are computational techniques that mimic natural phenomena such as genetic evolution, swarm intelligence, or natural selection to find the optimal solution for a given problem. These algorithms can efficiently solve complex optimization problems that cannot be solved with traditional mathematical techniques.

One of the main applications of metaheuristic optimization algorithms in healthcare is in predicting clinical outcomes. In a study by Tao Xie et al., they proposed a hybrid metaheuristic optimization algorithm that used PSO algorithm and utilised Support Vector Machines (SVM) to forecast colorectal cancer patients' chances of survival. The study reported that the proposed algorithm had higher accuracy and lower prediction error compared to traditional prediction models [50].

Another application of metaheuristic optimization algorithms is in optimizing treatment plans. In a study by Badra et al., they proposed a metaheuristic optimization algorithm based on the Grey Wolf Optimizer (GWO) to optimize the treatment plan of patients with prostate cancer [51]. The proposed algorithm aimed to minimize the total radiation dose to healthy organs and tissues while maintaining the radiation dose to the tumor. The outcomes demonstrated that the suggested algorithm performed better than conventional optimisation methods in terms of reducing the radiation dose to healthy organs and tissues.

Metaheuristic optimization algorithms have also been used to reduce hospital costs. In a study by Walter et al., they proposed a hybrid metaheuristic optimization algorithm based on PSO and the ACO algorithm to optimize the layout of hospital wards and reduce the average walking distance of hospital staff [52]. The study reported that the proposed algorithm reduced the average walking distance of hospital staff by 28% and improved the utilization of hospital space by 14%.

Overall, the use of metaheuristic optimization algorithms in healthcare has shown promising results in improving clinical outcomes, optimizing treatment plans, and reducing hospital costs. These algorithms provide a powerful computational tool for solving complex problems in healthcare, which can ultimately improve the quality of patient care.

Hybrid optimization algorithms have become an attractive option for solving complex problems in healthcare due to their ability to combine multiple optimization techniques and overcome the limitations of individual methods. In this section, we will discuss the applications of hybrid optimization algorithms in healthcare for feature selection, disease detection, classification, and segmentation, along with examples from published research articles.

The researchers and scholars tried implementing the metaheuristic algorithms in various key technologies of healthcare applications. Every method presented in this field tried to address a particular problem with its own particular set of parameters and declared to prove better simulation outcomes with respect to previous traditional methods. Moreover, some researchers and scholars used a hybrid model for solving a single problem. Figure 1.2 summarizes various metaheuristic optimization algorithms that are used in healthcare based on the domain like segmentation, feature selection and classification. Some of the optimization techniques used for segmentation are Particle Swarm Optimization (PSO), Ant lion optimizer (ALO) and Glow worm Swarm optimization(GSO) . The optimization techniques utilized in feature selection are GA , ABC, Cuckoo Search Algorithm (CSA) and GWO.

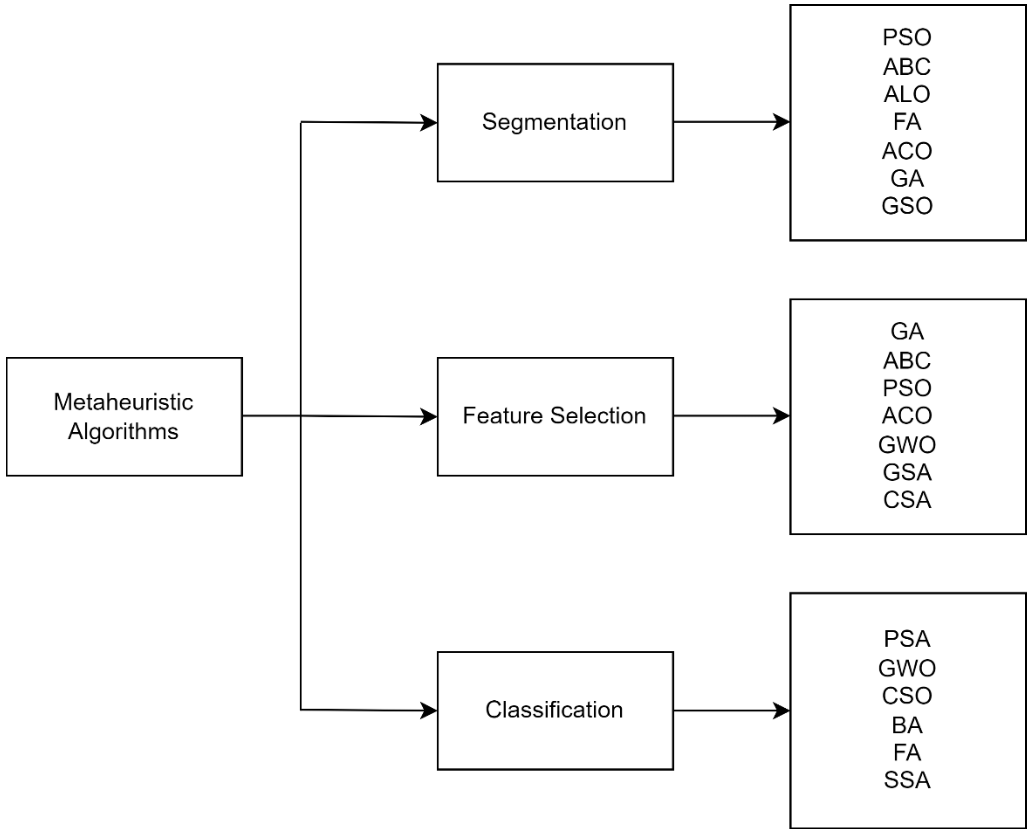


Figure 1.2: Summary of optimization algorithms in healthcare

Feature selection is an essential step in the analysis of high-dimensional data in healthcare. It involves selecting a subset of relevant features from a large set of variables, to improve the accuracy of prediction models and reduce computational complexity. Hybrid optimization algorithms have shown great potential for feature selection in healthcare applications. For example, a hybrid algorithm based on PSO and “differential evolution” (DE) was proposed for feature selection in breast cancer diagnosis using gene expression data [53]. The algorithm aims to select the most informative genes from a large set of gene expression data to improve the accuracy of breast cancer diagnosis. The proposed hybrid algorithm outperformed other feature selection algorithms in terms of accuracy and the number of selected genes.

Disease detection is another important application of hybrid optimization algorithms in healthcare. It involves identifying the presence of a particular disease or condition in patients using various diagnostic tests. Hybrid algorithms have been used for disease detection in various medical imaging modalities such as “computed tomography” (CT), “magnetic resonance imaging” (MRI), and “positron emission tomography” (PET). For example, a hybrid algorithm based on “artificial bee colony” (ABC) and “gravitational search algorithm” (GSA) was proposed for the discovery of lung cancer using CT pictures [54]. The algorithm aims to identify lung nodules in CT images, which is a challenging task due to the variability in nodule size, shape, and texture. When compared to other algorithms, the suggested hybrid algorithm performed better in terms of sensitivity and specificity.

Classification is another important application of hybrid optimization algorithms in healthcare, which involves classifying patients into different groups based on their medical data. Hybrid algorithms have been used for classification in various healthcare domains such as diabetes, cancer, and Alzheimer's disease. For example, a hybrid algorithm based on the cuckoo search (CS) and firefly algorithm (FA) was proposed for the classification of diabetes using medical data [55]. The algorithm aims to classify patients into diabetic and non-diabetic groups based on their medical data. In terms of sensitivity and accuracy, the suggested hybrid algorithm performed better than the other classification algorithms.

Finally, segmentation is another important application of hybrid optimization algorithms in healthcare. It involves dividing medical images into different regions based on their intensity and texture features, to assist in diagnosis and treatment planning. Hybrid algorithms have been used for segmentation in various medical imaging modalities such as MRI, CT, and ultrasound. For example, a hybrid algorithm based on PSO and fuzzy C-means (FCM) was proposed for brain tumor segmentation in MRI images [56]. The algorithm aims to segment brain tumors into different regions, such as tumor and non-tumor regions, to assist in diagnosis and treatment planning. The proposed hybrid algorithm showed improved performance compared to other segmentation algorithms in terms of accuracy and Dice similarity coefficient.

In summary, hybrid optimization algorithms have shown great potential in healthcare applications for various tasks such as feature selection, disease detection, classification, and segmentation. These algorithms have the potential to improve the accuracy and efficiency of medical data analysis, which can ultimately lead to better patient outcomes.

1.8 Research Motivation:

The medical field relies extensively on the analysis of disease images acquired through advanced digital technologies for a wide spectrum of diagnoses. The incorporation of artificial intelligence (AI) in the interpretation of medical images has ushered in a new era of automatic and precise evaluations. This integration has not only lightened the burden on healthcare professionals but also reduced diagnostic errors, accelerated diagnosis times, and enhanced the accuracy of detecting and predicting various diseases. Achieving optimal diagnoses through AI techniques necessitates a diverse array of medical data sources, encompassing ultrasound, magnetic resonance imaging, brain scans, blood tests, genetic history, speech and gait analysis, and specialized cognitive tests. Moreover, artificial intelligence has brought about significant improvements in the overall patient experience within medical facilities and streamlined the process of transitioning patients to home-based rehabilitation programs. This transformative technology continues to shape the future of healthcare by optimizing diagnostic processes and enhancing patient care. As it evolves and matures through the accumulation of gathered information, it becomes vulnerable to misuse and unauthorized access to collected data. With the advancement of artificial intelligence (AI), the realms of technology and medicine are forging stronger collaborations aimed at enhancing technological capabilities. Several existing obstacles in the integration of AI algorithms into clinical applications stem from the limited availability of healthcare data suitable for machine learning. The adoption of standardized data formats, such as Fast Healthcare Interoperability Resources (FHIR), holds promise for more effective data aggregation. However, it's important to note that improved interoperability alone may not completely resolve the challenge of inconsistent semantic coding in healthcare data.

Although risk prediction, early detection, diagnosis, and precise prognosis are ongoing areas of development for certain types of cancer, they continue to present challenges in various cancer cases. Nevertheless, AI-driven tools hold the potential to enhance human endeavors, ultimately striving to enhance personalized care and improve outcomes for oncology patients. AI encompasses a range of subfields, including ML, DL, ANN, fuzzy logic, and speech recognition,possessing unique capabilities and functionalities that can enhance the performance of modern medical sciences. These intelligent systems streamline human involvement in clinical diagnosis, medical imaging, and decision-making processes. Artificial intelligence's application in the field of medicine is currently a topic of significant interest, particularly in the context of diagnosing and predicting medical conditions based on image analysis. Integrating an AI tool into clinical practice necessitates a thorough validation process to ensure its clinical effectiveness, especially in the realm of digital pathology. Crucial considerations encompass dataset variations, the inadvertent inclusion of confounding factors rather than genuine signals, inadvertent perpetuation of biases in clinical procedures, ensuring AI algorithms are interpretable, establishing dependable metrics for model confidence, and addressing the challenge of applying the AI model to diverse populations. It's particularly vital to conduct an analysis to comprehend the implications of a new algorithm. For instance, if the AI system identifies a different spectrum of diseases compared to current clinical practices, it becomes essential to assess the advantages and disadvantages of detecting this distinct range of conditions. In the case of mammography, this might involve identifying less severe cases of ductal carcinoma in situ, which could potentially lead to increased treatment with minimal improvements in patient outcomes.

1.9 Research Problem

We can condense the major issues in healthcare as follows:

- **Difficulty in accessing quality care**: Patients face problem to access online medical advice due to various problems. Sometimes these problems include their remote location or some patients cannot leave their homes.
- **Requirement for work absence:** For visiting doctor and do diagnostic tests patients has to take time off
- **Elevates the risk of infection**: Since patients must spend time in a waiting area, they face an increased likelihood of acquiring a new illness from another patient or potentially transmitting their own condition to someone else.
- **Handling high flow of patients:** It becomes difficult for doctors to handle large number of patients with efficiency, which affects their performance in hospital and also their health.

Apart from the major issues discussed, healthcare systems across the globe are grappling with various other challenges. One such significant obstacle is the surging cost of healthcare. This high cost makes healthcare unaffordable for many, particularly those residing in low-income areas. Additionally, the shortage of medical professionals, especially in remote areas, leads to prolonged wait times and insufficient care, adding to the already-existing problems.

To overcome these challenges, there is an urgent need for a feasible and cost-effective solution that can provide quality healthcare to all, regardless of their socioeconomic status or location. The primary objective of this research is to implement a system that can address some of the most significant challenges faced by healthcare systems today, including reducing the workload of hospital management systems and doctors, minimizing patient visits to hospitals, and preventing healthcare-associated infections.

To achieve these objectives, the research aims to tackle key issues such as predicting primary diagnostic tests before the first visit to the doctor by analyzing the patient's symptoms, designing a questionnaire that is user-friendly and provides sufficient information for test recommendations, and utilizing metaheuristic algorithms to optimize the prediction results for generating the minimum but adequate number of diagnostic tests. It is also crucial to verify whether the predicted tests help in correctly diagnosing the patient, as this will significantly improve patient outcomes.

The objective of this research is to establish a system that will 1) reduce workload in the hospital management system 2) reduce the workload of doctors and 3) reduce number of visits to the hospital to avoid risk of infection. This will help patients in the remote areas for getting quality healthcare in cost effective way.

The following key issues will be resolved by this research:

(1) How to reduce the healthcare associated infections.
(2) How to predict primary diagnostic tests before the first visit to the doctor using the patient's symptoms.
(3) How to design questionnaire that is easily understandable by the patients and provide sufficient information for test recommendation.
(4) How the metaheuristic algorithms can help to optimize the prediction result to generate minimum but sufficient diagnostic test.
(5) How to verify whether the predicted tests help in correctly diagnosing the patient.

In conclusion, implementing such a system would have a significant impact on healthcare outcomes, providing a viable solution to the challenges faced by healthcare systems worldwide. By improving access to quality care, reducing costs, and optimizing patient outcomes, this system has the potential to revolutionize healthcare delivery.

1.10 Research Objectives

This research come up with a futuristic approach to predict diagnostic tests and probable disease based on symptoms. The application of metaheuristic algorithms in the Decision Support System for diagnostic test prediction represents an innovative approach. In this research patients' symptoms are collected through a doctor verified questionnaire. A model is implemented to predict diagnostic tests. An empirical analysis of different metaheuristic algorithms is performed. A novel competitive verse water wave optimization algorithm is applied on the proposed model to optimize the prediction results. Based on the problem statement, the following three objectives are defined:

(1) To analyze existing optimization methods.
(2) To design and develop the prediction methodology for diagnostic tests and disease using Meta heuristic approach.
(3) To Validate the outcomes of the proposed methodology.
(4) To comparing and evaluate the obtained results using standard performance assessment metrics.

1.11 Thesis outline

This dissertation is organised in six different chapters.

Chapter 1 provides background and motivation behind predicting diagnostic tests and probable disease using metaheuristic algorithms. It covers the comprehensive explanation of current state of healthcare issues with respect to patient management during communicable disease outbreak, health workers mental health and hospital management system. The debate and the flaws in existing system are discussed to further explain the topic. This chapter paves the foundation for the problem statement. The chapter also lists the objectives of research undertaken. This is followed by a description of the dataset, an explanation of the issue, and a set of goals.

Chapter 2 reviews the metaheuristic algorithms used in healthcare domain for various purposes like feature selection, image segmentation, image classifications and so on.
Chapter 3 is devoted to discussing how to implement the model that will predict diagnostic test and probable disease using metaheuristic algorithm. Predicting diagnostic tests is an important component proposed based on the symptoms of the patient. A novel meta-heuristic Competitive verse Water Wave optimization algorithm to detect disease is also presented in this chapter.

Chapter 4 describes the process used to validate the outcome of proposed methodology.

Chapter 5 describes the process used to compare and evaluate the obtained results using standard performance assessment metrics. The results of the proposed model are discussed along with performance measurement metrics.

In Chapter 6, the major contributions of the finished work are highlighted to offer the research's concluding thoughts. Towards the end of the chapter, instructions for further practical research work are given.

Chapter 2
Analysis of Existing Optimization Methods

2.1 Introduction

Computing facilities and advanced technologies have made a significant contribution to the current era of medicine, enabling humans to use various tools for surgery, testing, and developing multiple medicines. Feature selection and segmentation are crucial aspects of the healthcare industry, with a rapidly growing pace of storing diverse healthcare data in clinic-related databases. The challenge of reducing the dimensionality of large datasets while maintaining performance accuracy is a major real-world problem in numerous scientific fields, including medicine Henceforth, feature reduction is categorized into two primary divisions: feature extraction or generation and feature selection. Metaheuristic methods have gained increasing prominence in addressing the complexities of optimizing design problems that encompass numerous variables subject to intricate constraints.

Due to the rapid increase in health data accessibility and computing power, healthcare applications now incorporate Artificial Intelligence (AI), transforming them into Smart Healthcare [57]. This strategy fosters sustainable development. Healthcare sustainability seeks to enhance both the social and financial effects of healthcare services concurrently, with the potential to shape the future of healthcare. Today, a variety of optimization methods are employed to address diverse healthcare challenges. Certain algorithms ensure optimal solutions, whereas metaheuristic approaches guarantee solutions that are close to optimality. In Figure 2.1, you can observe various optimization techniques divided into heuristics and metaheuristic algorithms. Commonly utilized Evolutionary Algorithms (EA) encompass “Genetic Algorithm” (GA), “Flower Pollination Algorithm” (FPA), “Differential Evolution” (DE), and Evolutionary Strategy. On the other hand, physical algorithms include “Black Hole Algorithm” (BHA), Central Force Algorithm” (CFA). Furthermore, in the context of healthcare optimization techniques, frequently employed Swarm Intelligence (SI) algorithms comprise PSO, “Krill Herd Algorithm” (KHL), “Whale Optimization Algorithm” (WOA), “Cuckoo Search Algorithm” (CSA), and “Salp Swarm Algorithm” (SSA).

Researchers are working on automating the diagnosis and prediction of several diseases using multiple ML techniques to effectively treat patients. Metaheuristic algorithms are efficient techniques that provide practical solutions to problems in the computing world, especially in healthcare data diagnosis. These algorithms possess two crucial attributes: intensification, which involves searching around the present optimal solutions and selecting the most exceptional alternatives, and diversification, which effectively explores the search space. Contemporary metaheuristic algorithms have been devised to expedite problem-solving, tackle sizable difficulties, and acquire resilient algorithms.

This review studies different metaheuristic, optimization-based disease diagnosis method, to detect and classify various health-related diseases based on their optimized results. The approach combines different SI optimization methods, including KHO, ABC, PSO, GWO, and MFO method, to perform segmentation, feature selection, detection, and classification. Metaheuristic optimization techniques are valuable tools for extracting and selecting features in various illness diagnostic scenarios, enabling early disease identification.

2.2 Disease Detection Through Swarm Intelligence Techniques

Choubey and colleagues [58] introduced a two-stage approach for diabetes detection. In the initial stage, they gathered data from two distinct sources, namely the Pima Indian Diabetes and Localized Diabetes datasets. Subsequently, in the second stage, these collected datasets underwent processing and analysis through two separate methods. The first method involved classification using various classifiers, while the second method utilized “Principal Component Analysis” (PCA) and PSO to reduce the number of features. They conducted a comparative analysis of several models, demonstrating the potential applicability of this model to other diseases as well.

Hassan and Hassanien [59] employed Whale Optimization Algorithm (WOA) for the segmentation of retinal fundus vasculature, which consisted of two main phases. In the initial phase, they enhanced the fundus images through pre-processing techniques. Then, in the subsequent phase, they employed multilevel thresholding in combination with WOA to segment the blood vessels. The resulting model demonstrated greater accuracy in comparison to existing segmentation models. Moreover, it was shown that this algorithm exhibited enhanced resilience in the face of sickness symptoms, disturbances, and brightness-related problems.

Majhi [60] introduced a diabetes classification approach employing a FFNN combined with MFO. The MFO played a role in adjusting the weights of the FFNN. This experimental study was conducted using the UCL repository PID dataset. Ultimately, the hybrid model demonstrated the capability to predict outputs and achieve accurate classification results.

Si *et al.* [61] presented a clustering technique utilizing GSO for the segmentation of MRI lesions. The process began with noise removal and intensity correction in MRI images. Subsequently, they assessed clustering using GOBL and GSO. In a post-processing stage, the lesions were distinguished from normal tissue, and the results were compared with existing technologies.

Lu and colleagues [62] employed an ELM in combination with the BA for the diagnosis of pathological brain conditions. Initially, ELM-BA was trained to distinguish between pathological brains and healthy tissues. Subsequently, a 10-fold cross-validation was conducted, resulting in a sensitivity rate of 99.4%. The experimental results demonstrated the robustness and accuracy of this method in diagnosing pathological brain disorders, potentially aiding medical experts in their diagnostic processes.

Gunavathi and Premalatha [63] applied CSA algorithm to feature selection for cancer classification. Initially, genes were ranked using T and F-statistics as well as SNR. This method is straightforward and can be seamlessly integrated with statistical feature selection models. CSA was then utilized to identify informative genes from the top-ranked ones. Following this, the KNN algorithm was optimized using CSA, and the resulting model was tested on 10 distinct datasets.

Kumar and Singh [64] presented a model for breast cancer detection that utilized EGWO in conjunction with SVM. This hybrid approach was designed to identify the most effective subset of tumor features for distinguishing between benign and malignant cases. The model was evaluated using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, and the results demonstrated a substantial enhancement in classification accuracy.

Ha *et al.*[65] introduced an automated model for breast cancer detection by combining CNN with an Improved DHO algorithm. The process began with pre-processing to eliminate noise and enhance classification. Subsequently, features were extracted using Haralick features. The model was further enhanced by integrating ICA to reduce the number of features. Ultimately, the classification performance was improved through the utilization of CNN with Improved DHO, and their accomplishments were substantiated.

Sudha *et al.*[66] employed an Enhanced Local Attraction (ILA) technique for selecting features in the classification of breast cancer. The initial steps involved pre-processing mammogram images and segmenting them using a region-growing algorithm. Subsequently, features were extracted and chosen using ILA in combination with three different classifiers. The results were assessed based on the number of features and the resulting classification accuracy.

Taherian Dehkordi *et al.* [67] employed WWO to predict and diagnose diabetes mellitus, leading to improved precision in diabetes prediction. In this study, the parameters of a NN were simultaneously determined while training was conducted using WWO to optimize these values. The parameters identified by WWO effectively reduced classification errors in diagnosis. Ultimately, the system's performance was validated using various metrics and compared with baseline approaches.

Tabrizchi *et al.* [68] introduced a method for breast cancer detection using the MVO-GBDT. Their approach focused on enhancing the performance of GBDT by fine-tuning its parameters through optimization by MVO. To achieve greater accuracy, they employed k-fold cross-validation during the GBDT optimization process. The entire implementation was conducted using the WDBC dataset, resulting in reduced variance in breast cancer diagnosis.

Rani *et al.* [69] introduced KHO-RF as a method for feature selection and classification. This approach consists of two stages: In the initial stage, redundant and irrelevant genes were filtered out, and subsets of genes were selected from a vast pool using KHO. Then, in the second stage, classification was performed using the genes selected in the first stage. To validate their performance, experiments were conducted on ten different datasets.

Kuo *et al* [70] Conducted a meta-analysis to evaluate the diagnostic accuracy of machine learning models in predicting COVID-19. Research results indicate that non-image data may be used to accurately predict COVID-19 with satisfactory performance..

Abayomi-Alli *et al* [71] conducted a study where they explored the utilization of ensemble learning techniques to create predictive models for the efficient identification of COVID-19 using regular laboratory blood test outcomes. The results indicate that an ensemble learning model combining DNN and ExtraTrees achieved an average accuracy of 99.20% and an area under the curve (AUC) of 99.40%. Additionally, AdaBoost yielded an average accuracy of 99.27% and an AUC of 98.80% when applied to the San Raffaele Hospital dataset.

Fathi Kazerooni *et al* [72] investigated the additive value and independent reproducibility of integrated diagnostics in the prediction of overall survival (OS) in isocitrate dehydrogenase (IDH)-wildtype GBM patients, by combining conventional and deep learning methods. Cox-PH modelling showed a concordance index of 0.65

(95% CI 0.6–0.7) for clinical data improving to 0.75 (95% CI 0.72–0.79) for the combination of all omics. Li *et al* [73] Designed a noninvasive AI system to detect SRCC of GC and determine which patients with SRCC will benefit from postoperative chemotherapy using preoperative contrast-enhanced CT. The AI model based on CT scans showed excellent performance in detecting SRCC, accurately stratifying patient prognosis, and forecasting responses to treatment.

Combalia *et al* [74] performed simulations of similar real-world situations to evaluate the influence of elements such as the institutions that provide the photos, diagnoses that are not included in the training dataset, and other image artifacts on the accuracy of classification. The objective of this research was to provide significant information to doctors and regulatory authorities about safety issues and real-world accuracy. The most successful algorithm attained a balanced accuracy of 58·8% on the BCN20000 dataset, which was specifically developed to simulate real-world clinical circumstances. In comparison, the algorithm earned a balanced accuracy of 82·0% on the HAM10000 dataset, which was earlier used as a benchmark in a previous study.

Rashid *et al* [75], anticipated the occurrence of five common chronic illnesses, including cancer, diabetes, myocardial infarction, hepatitis, and renal disease. The results are tested with other datasets of chronic diseases and shown to surpass existing benchmark methods. Rashid et al assessed the precision of artificial intelligence (AI) software in a practical, three-level network comprising several hospitals dedicated to stroke care. The accuracy measures provide evidence that Viz LVO is a valuable supplementary tool in stroke diagnosis. Rapid and precise diagnosis, with a strong negative predictive value, reduces the risk of overlooking individuals who may possibly be saved.

Soerensen *et al* [76] Assessed the efficacy of an AI framework for forecasting the likelihood of cancer in individuals referred from primary care using regular blood tests. The AI risk score has the potential to be a significant tool in clinical decision-making. Additional validation of the score is necessary to ascertain its suitability in different populations.

Moreira *et al* [77] illustrated the architecture's appropriateness using a case study of a low-cost device used for COVID-19 diagnostics. This work highlights two important findings: the assessment of low-cost devices in effectively and correctly handling COVID-19 prediction tasks, and the quantitative evaluation of CNN models implemented on these low-cost devices.

Almotairi *et al* [78] Explored many applications and implementations of contemporary technology to battle the COVID-19 pandemic across several domains, including image processing, illness monitoring, outcome prediction, and computational medicine. The findings demonstrate that CT scans are effective in diagnosing individuals with COVID-19 infection. This research work makes a valuable contribution to the field of AI-based diagnostic test prediction, which is a viable approach for managing the healthcare challenges of the new normal age.

Table 2.1 represents disease detection through swarm intelligence techniques

Table 2.1: Disease detection through swarm intelligence techniques

Authors and citation	**Algorithm**	**Data**	**Merits**	**Demerits**
Choubey *et al.* [58]	PCA- PSO	Pima Indian Diabetes (PID) and Localized Diabetes datasets	Enhance accuracy and less time of computation	There was no guarantee that the classification models achieving higher accuracy on the given dataset would necessarily yield the same level of accuracy on other datasets.
Hassan and	WOA	DRIVE	Attained high accuracy with abnormal images and obtained	Suffered from slow convergence

Hassanien [59]			precise segmentation results for small blood vessels.	
Majhi [60]	FFNN-MFO	PID	Achieved better accuracy and error rates	Trapped by local optima
Si *et al.* [61]	GSO	MRI slices of patient human brain	Successfully detected small and large lesions	Takes more time for processing
Lu *et al.* [62]	ELM-BA	Website of Harvard Medical School	Resilient and precise in diagnosing pathological brain disorders	Understanding the entropy values was difficult
Gunavathi and Premalatha [63]	KNN-CSA	Kent Ridge Biomedical Data Repository	Achieved 100% accuracy	This approach relied on the initial condition
Kumar and Singh [79]	EGWO-SVM	WDBC	Effectively identified breast cancer	Performance will be reduced for large datasets
Ha *et al.* [65]	CNN-IDHO	DCE-MRI	Attained improved accuracy and precision.	However, there is still some residual noise in the system, which diminishes the accuracy

Sudha *et al.* [66]	ILA-3 classifiers	Digital database for screening mammography (DDSM)	Achieved better accuracy with less features	Has a high computational complexity
Taherian Dehkordi et al. [67]	NN-WWO	PID	Simple and cost-effective in diabetic diagnosing	The execution time exceeded that of the existing models.
Tabrizchi *et al.* [68]	MVO-GBDT	WDBC	Attained superior accuracy without encountering overfitting problems.	The runtime was not substantial due to the limited number of instances.
Rani *et al* [69].	RF-KHO	UCI machine learning	Statistical and biological significance were assessed.	Suffered from convergence speed

2.3 disease detection using hybrid optimization techniques

The hybrid algorithm combines both local and global search methods to enhance disease prediction performance. Furthermore, various approaches, including chaotic concepts, population initialization, and self-adaptive strategies, are integrated with swarm intelligence (SI) algorithms. The following section discusses literature that focuses on hybrid optimization techniques

Sahu *et al.* [80] applied a combination of PSO and Reverse FA for cancer classification. They initially employed FCBF to identify redundant and irrelevant features. Subsequently, they utilized SVM as the classifier, implementing a 10-fold cross-validation approach, and optimized SVM using PSO and Reverse FA. This model successfully struck a better balance between exploitation and exploration.

Alsarori *et al* [81] examined the detection of cancer cells in histological nuclei images through the application of ABC and PSO. Initially, they isolated the red component of the histological images. During the smoothing process, two different filters were applied. ABC was employed to ascertain the optimal threshold value, while PSO was utilized to determine the positions and velocities of the bees.

Saleh and Rüştü [82] adapted MTLBO for melanoma classification, incorporating aspects of ABC. They commenced with preprocessing steps to eliminate noise and enhance image quality. Subsequently, segmentation was performed using Otsu's thresholding, and feature extraction was carried out using GLCM. Finally, they trained a neural network classifier using the MTLBO-ABC optimization approach. The experimental results demonstrated the superior performance of their proposed model compared to baseline models.

Dubey [3] introduced a hybrid optimization approach that combined the LA with the BOA for addressing multiple diseases. They began by collecting a dataset from UCI ML, which included various disease types. The hybrid LA-BOA method was applied to determine the optimal feature selection for all attributes within the dataset. These selected features were subsequently employed by the classifier, resulting in enhanced outcomes.

Kora et al. [83] conducted research on the detection of heart disease employing a hybrid approach that combined BFO and PSO. The primary objective of this hybrid method was to create a straightforward and cost-effective model. The process began with preprocessing the raw signal and extracting features using DWT. Subsequently, they introduced a novel modification to the BFO and integrated it with DWT and SVM for evaluating their detection performance.

Balasubramanian and Ananthamoorthy [84] employed an Improved ANFIS combined with a modified GSO and DE for medical diagnosis. The DE-modified GSO technique was used to diagnose neurodegenerative and ocular illnesses. The findings demonstrate that this model attained decreased error measures and the utmost R2 value. Moreover, this technique has the capability to be expanded to other categories of illnesses, such as cardiovascular disease, cancer, and Covid-19, as well as for predicting commodity-related data.

Thawkar [85] showcased a mammography feature selection and classification approach employing the Salp Swarm Algorithm (SSA) combined with Teaching-Learning-Based Optimization (TLBO). The selected features obtained through this hybrid optimization technique were subsequently evaluated using an Adaptive Neuro-Fuzzy Inference System (ANFIS). The experimental results clearly indicate that this hybrid optimization approach outperformed the basic algorithm. Additionally, the model's robustness was put to the test using benchmark datasets, further confirming its effectiveness.

Gurav *et al.* [86] proposed a model for diagnosing prostate cancer through histopathology images, which involved the integration of Fuzzy logic with the Rider Optimization Algorithm (ROA) and the Salp Swarm Algorithm (SSA). Initially, the input image underwent preprocessing and segmentation using color space transformation thresholding. Subsequently, features were extracted based on the local optimal oriented pattern and classified using a neural network. Notably, the neural network was optimized using the SSA and ROA techniques. The results were rigorously validated using various performance metrics. Table 2.2 provides an overview of disease detection using hybrid optimization techniques.

Table 2.2: Diagnosing Diseases with Hybrid Optimization

Authors and citation	**Algorithm**	**Data**	**Merits**	**Demerits**

Sahu *et al.*[80]	SVM-PSO-RFA	Twelve datasets	Reduced Noise and Robust Approach	This model was inadequate for extensive datasets.
Alsarori *et al.* [81]	SVM-PSO-ABC	PSB-2015 crowd-sourced nuclei dataset	Achieves better efficiency and accuracy	Takes more time for computation
Saleh and Rüştü [82]	MTLBO-ABC	dermoscopy dataset	This model was more accurate	In certain instances, the accuracy fell short of what the proposed model achieved.
Dubey [3]	LA-BOA	UCI machine learning	Better prediction outcomes in predicting several diseases in the HC sector	Long computation time
Kora et al. [83]	PSO-BFO	MIT-BIH AF	Economical model	The training process takes a considerable amount of time

Balasubramanian and Ananthamoorthy [84]	modified GSO and DE	UCI machine learning	This model can be applied to various types of diseases	Long computation time
Thawkar [85]	SSA-TLBO	UCI machine learning	Achieves better convergence with less computation	As the number of samples and features increased, the performance deteriorated.
Gurav *et al.* [86]	SSA- ROA	National Cancer Institute GDC Data portal	This model efficiently detected the prostate cancer	Performance was reduced when evaluated on large database

2.4 A Review of Various Case Study

Meta-heuristic optimization approaches have been gaining attention in recent years due to their effectiveness in solving complex problems that are difficult to solve through traditional methods. These approaches draw inspiration from natural systems and employ techniques such as randomization and local search to explore the solution space efficiently.

Nature-inspired algorithms can be broadly categorized into population-based genetic algorithms and trajectory-based algorithms. Population-based genetic algorithms work by evolving a population of potential solutions through genetic operators such as crossover and mutation, whereas trajectory-based algorithms optimize a single solution by iteratively adjusting its trajectory towards a better solution. Despite the growing popularity of meta-heuristic approaches, selecting the best algorithm for a given problem can be challenging due to the sheer number of nature-inspired algorithms available. Therefore, researchers must carefully consider the problem's characteristics and the available resources to determine the most suitable algorithm.

In computer vision applications, image segmentation is considered a critical operation that partitions an image into meaningful regions for further analysis and interpretation. It plays a vital role in a wide range of applications, including scene understanding, object recognition, medical image analysis, etc.. The segmentation process can be challenging due to the inherent variability and complexity of images, making it a prime target for meta-heuristic optimization techniques.

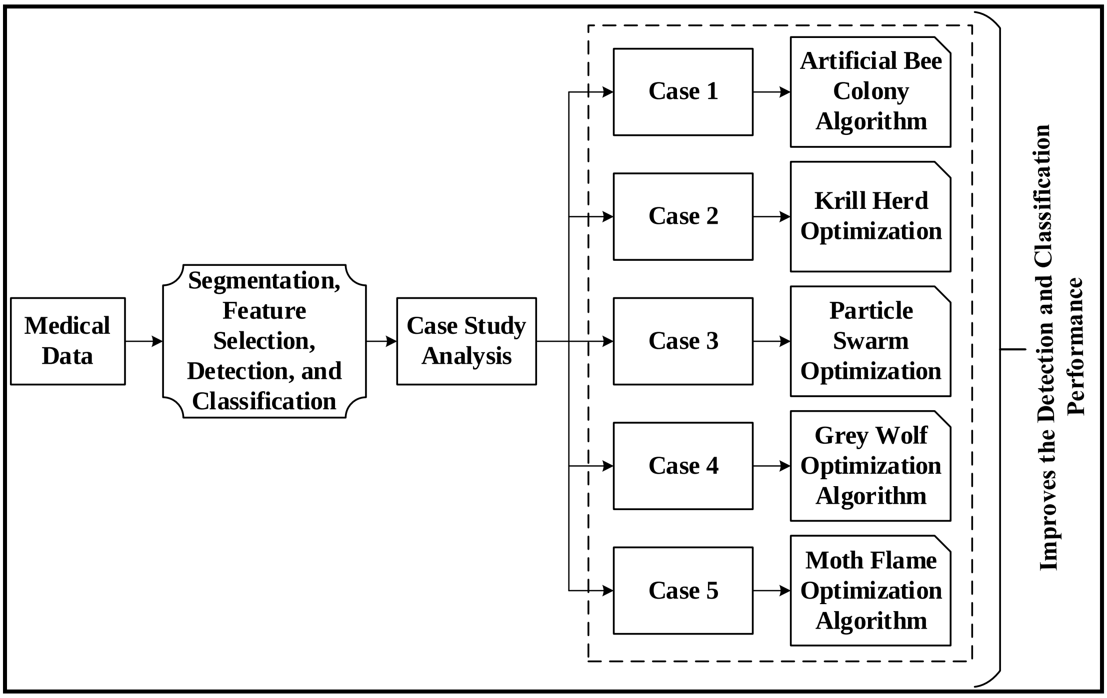


Figure 2.1: Diagrammatic representation of reviewed case studies

The diagrammatic representation of reviewed case studies is provided in figure 2.1. Feature selection is a preliminary phase in data mining where a subset of pertinent features is chosen to construct models. The quest for an ideal feature subset within a high-dimensional feature space presents a challenging NP-complete problem. Consequently, conventional optimization algorithms prove inefficient in tackling extensive feature selection tasks. As a result, metaheuristic algorithms find widespread use in effectively addressing feature selection challenges. In this study, five case analyses are presented like analysis of the ABC approach, KH, PSO algorithm, analysis of GWO algorithm, and MFO algorithm-based segmentation and feature selection method for improving the classification performance.

2.4.1 Case 1: Analysis of Artificial Bee Colony Approach

Image segmentation is a crucial technology for image processing, and several methods have been proposed to address this task. Among these methods, thresholding is a simple yet essential technique. To enhance the segmentation performance, Li et al. [87] presented ABC based on the honey-collecting behavior of bees in nature. The act of bees seeking for high-quality food sources might be seen as an endeavor to find the most advantageous solutions to an optimization issue.

The intelligent swarm paradigm has four primary components: food supply, employed bee, spectator bee, and scout bee. Each answer corresponds to a certain food source, and the hired bee visits each food source one by one. After the food source is thrown away, the bee that was previously working becomes a bee that searches for other sources. The abundance of nectar in the food supply is governed by the level of suitability of the matching solution. Li et al. conducted an analysis of the ABC algorithm's effectiveness in selecting features for diagnosing Parkinson's disease. while Ye et al. [88] studied this algorithm for image segmentation. Nonetheless, a detailed description of this method is presented below.

The search for food sources in ABCs is performed by bees in three distinct steps:

1. Employed bees explore potential food sources and store pertinent information.
2. Onlooker bees select employed bees with superior performance and trail them.
3. If an employed bee has not been updated for a specific duration, it becomes a scout bee and explores a new food source at random.

Every food source corresponds to a potential solution for the 2D function problem, and the quantity of nectar collected by a food source signifies the quality of the corresponding solution, as indicated by its fitness value. To begin, the ABC creates a starting population of SN solutions, where SN represents the number of food sources and is identical to the number of employed bees. Each solution, denoted $x_i(i = 1,2,\dots,SN)$, is an n-dimensional vector, where n corresponds to the number of optimization parameters, which in this case is 2. The initial population is randomly generated and uniformly distributed.

The fitness function utilized in ABC for 2D thresholding is defined as follows:

$$fit_i = J_F(s,t) \tag{1}$$

Here, $J_F(s,t)$ represents the objective function value of solution i, which can be computed by mapping the position of the food source into a threshold vector. The fitness value of solution i is denoted as fit_i. Initially, a population of solutions is developed and then undergoes multiple cycles of search procedures performed by employed bees, observer bees, and scout bees. The maximum cycle number (MCN) is one of the four control parameters used in the ABC algorithm. [89].

During each iteration of the algorithm, the employed bees seek out a food source within the vicinity of their current location and assess its fitness level. The i^{th} food source is denoted as $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$. The amount of nectar in the food source located at Xi is represented by $F(X_i)$. Following the observations of the employed bees, onlooker bees move towards the region of the food source at Xi with a probability defined by:

$$p_i = fit_i / \sum_{j=1}^{NS} fit_j \tag{2}$$

The following expression is used by the ABC to generate a candidate food position (threshold pair) from the old one in memory:

$$v_{ij} = x_{ij} + \emptyset_{ij}\left(x_{ij} - x_{kj}\right) \tag{3}$$

where, $k \in \{1, 2, \dots, SN\}$ and $j \in \{1, 2, \dots, n\}$ are indices chosen at random; The value of k should be distinct from i and $\emptyset_{ij}$ represents a random number ranging from -1 to 1. After generating each candidate food source position v_{ij}, the ABC evaluates its performance using an artificial bee. Next, the performance of the new candidate is evaluated by comparing it to that of the previous candidate using a greedy selection method. If the new food source has a quality that is equivalent to or superior to the old one, the old food source stored in memory is substituted with the new one. In the ABC algorithm, if a location reaches its maximum potential after a certain number of cycles, the related food source is deemed abandoned. The "limit for abandonment" is a critical control parameter of the ABC algorithm, which determines the preset number of cycles. When an abandoned source, represented by $X_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ is identified, the scout finds a new food source to replace it. This process can be described as:

$$x_{ij} = l_i + \delta(u_i - l_i) \tag{4}$$

where, $j \in \{1,2,\ldots,n\}$, l_i and u_i are the lower and upper bounds of the parameter x_{ij} and δ is a random number in the range $[0, 1)$, Where, l_i and u_i are lower and upper bounds of the variable X_i, in this paper, l_i is 0 and u_i is 255.

For $j \in \{1,2,\ldots,n\}$, l_i and u_i represent the lower and upper limits of the parameter x_{ij}, respectively. δ is a random number in the range of [0,1), where l_i and u_i are the lower and upper limits of the variable X_i. In this study, l_i is set to 0 and u_i to 255.

Algorithm: ABC Algorithm

Initialize the population solutions x_{ij}, $x = 1,2,\ldots,SN$, $j = 1,2,\ldots,n$ by (3)

Evaluate the fitness value of the population

cycle = 1

repeat

Produce new solutions v_{ij} for the employed bees by using (4) and evaluating them

Apply the greedy selection process

Calculate the probability values p_{ij} for the solutions x_{ij} by (2)

Produce the new solutions v_{ij} for the onlookers from the solutions x_{ij} selected depending on p_{ij} and evaluate their fitness value

Apply the greedy selection process

Determine the abandoned solution for the scout, if exists, and replace it with a new randomly produced solution x_{ij} by (3)

Memorize the best solution achieved so far

cycle = cycle + 1

until cycle = MCN

Figure 2.2: Pseudo Code for ABC Algorithm

The pseudo-code of the ABC algorithm is shown in figure 2.2. Based on the above description, it can be inferred that the ABC algorithm utilizes four control parameters: the number of food sources, which is equivalent to the number of employed or onlooker bees (SN), the limit value, and the MCN.

2.4.2 Case 2: Segmentation and Feature Selection using Krill Herd (KH)

In Resma, *et al* [90], presented meta-heuristic KHO algorithm has been suggested for tackling the picture segmentation issue. This algorithm employs multilayer thresholding. Nevertheless, the effectiveness may be enhanced by using the KH algorithm, since it relies on the behavior of the krill herd, thus making it a valuable swarm intelligence optimization method as presented by Sumanth et al [91]. The KH algorithm involves the exploration of krill food sources in a search space with several dimensions, followed by the generation of a variety of choices. However, the focus will be on the distance between the krill individuals and the surplus food linked to the expenses. Subsequently, a krill individual's time-dependent position can be assessed by the three operational procedures: (i) motion, which was induced by the presence of other individuals, (ii) foraging motion, and (iii) random physical diffusion. The following are the KH algorithm's key benefits: all agents have a role in the procedure; it is not essential to have derivative information; both the crossover operator and the mutation operator are exploited. However, the KH algorithm necessitates an optimal approach for determining the primary krill distribution as well as parameters and more comprehensive basic motions within the algorithm.

Three actions can be used to calculate the time-varying position of an individual krill: random diffusion (DK_i), hunting activity (FK_i) and Krill's motion initiation (NK_i), The Krill Herd Optimization algorithm has utilised a d-dimensional search space Lagrangian model (Damaševičius, *et al* [31]).

$$\frac{dX_i}{dt} = NK_i + FK_i + DK_i \tag{5}$$

The Krill Herd Optimization algorithm looks like this:

Step 1: Set the initial value for the variables position of Krill (X), maximum foraging speed (V_f), maximum number of iterations (I_{max}), maximum diffusion speed (D^{max}), maximum induced speed (N^{max}) and number of Krill (n)

Step 2: for $i = 1$ to I_{max} Repeat Steps 3 to Step 7

Step 3: Determine the movement of each Krill

Step 4: Motion as an effect of other Krill

The movement is caused by the krill individuals' collective actions as they constantly work to maintain a high density. The nearby target and a repulsive swarm density are used to determine the motion induction direction, represented by α_i .

$$NK_i^{new} = N^{max}\phi_i + \omega_n NK_i^{old} \tag{6}$$

where, ω_n which has a value in the interval [0,1], is the inertia weight connected to the produced motion. To indicate the latest motion induced NK_i^{old} is used.

$$\phi_i = \phi_i^{local} + \phi_i^{target} \tag{7}$$

The local influence that the neighbours offer is represented by ϕ_i^{local} and has the following definition:

$$\phi_i^{local} = \sum_{j=1}^{Nn} \widehat{K}_{i,j}\widehat{X}_{i,j} \tag{8}$$

where,

$$\widehat{X}_{i,j} = \frac{X_j - X_i}{\|X_j - X_i\| + z} \tag{9}$$

$$\widehat{K}_{i,j} = \frac{K_i - K_j}{K^{worst} - K^{best}} \tag{10}$$

K_i is the fitness of i^{th} Krill, K_j is the fitness of j^{th} neighbour, z is a tiny positive number to prevent singularities, and N_n represents the number of neighbours and K^{worst} and K^{best} are the best and worst fitness values for Krill individuals. The optimal Krill individual provides the effect of target direction, indicated by ϕ_i^{target} and it can be calculated as follows:

$$\phi_i^{target} = C^{best}\widehat{K}_{i,best}\widehat{X}_{i,best} \tag{11}$$

The effective coefficient of a krill individual with the highest fitness to an i^{th} krill individual represented by C^{best} is as follows:

$$C^{best} = 2\left(rand + \frac{1}{I_{max}}\right) \tag{12}$$

I stands for the iteration number, $rand$ shows a random number between 0 and 1. Determine the sensing distance using

$$d_{s,i} = \frac{1}{5N}\sum_{j=1}^{N} X_i - X_j \tag{13}$$

Where, X_i denotes the associated position of i^{th} Krill and N is the total number of krill individuals. If X_i and X_j distance is less than the designated sensing distance then X_j and X_i are neighbours.

Step 5: Motion induced by foraging activities

The location of the food and past experiences with that area are used to define the motion induced by foraging actions. Fk_i (Foraging motion) is given by,

$$Fk_i = V_f \gamma_i + \omega_f FK_i^{old} \tag{14}$$

With γ_i^{best} denoting i^{th} Krill's current best fitness and ω_f representing the inertia weight of foraging motion (a random value in the interval [0,1]) respectively.

$$\gamma_i = \gamma_i^{food} + \gamma_i^{best} \tag{15}$$

Food attractive β_i^{food} is given by,

$$\gamma_i^{food} = C^{food} \widehat{K}_{i,food} \hat{X}_{i,food} \tag{16}$$

The food coefficient C^{food} is computed as:

$$C^{food} = 2\left(1 - \frac{1}{I_{max}}\right) \tag{17}$$

Step 6: Diffusion through physical means

The following is a definition of physical diffusion motion in terms of a stochastic directional vector and the diffusion speed at peak:

$$Dk_i = D^{max}\left(1 - \frac{1}{I_{max}}\right)\Psi \tag{18}$$

In the range [1,1], Ψ stands for the random directional vector.

Step 7: Genetic operators

Genetic reproduction methods, such as crossover and mutation, are combined with the KHO algorithm to enhance its performance. The m^{th} component of X_i denoted as $x_{i,m}$ can be derived by the subsequent formula:

$$x_{i,m} = \begin{cases} x_{r,m}, rand_{i,m} < Co \\ x_{i,m}, else \end{cases} \tag{19}$$

$$Co\ (crossover\ probability)\ =\ 0.2\widehat{K}_{i,ibest} \tag{20}$$

$$x_{i,m} = \begin{cases} x_{gbest,m} + \mu\left(x_{p,m} - x_{q,m}\right), rand_{i,m} < Mt \\ x_{i,m}, else \end{cases} \tag{21}$$

Mutation probability (M_t) is denoted as:

$$Mt = 0.05/\widehat{K}_{i,ibest}$$

Step 8: The vector of position of a krill individual through the interval $[t, t + \Delta t]$ can be obtained various movement parameters as follows:

$$X_i(t + \Delta t) = X_i(t) + \Delta t \frac{dX_i}{dt} \tag{22}$$

It maximises entropy criteria or traditional between-class variance approaches' goal functions. When segmenting the given image, a multilevel thresholding method based on KHO finds the appropriate threshold values.

Parameter Setting

An empirical analysis is used to estimate the parameters of KHO algorithms for multilevel thresholding. Table 2.3 displays the values for the parameters in the suggested KHO algorithm.

Table 2.3: Parameter Values for KHO

Parameter		**Value**
	Number of Iterations (I_{max})	20
Motion Parameters	Induced Motion (N_{max})	10
	Foraging Motion (V_f)	5
	Physical Diffusion (D_{max})	20

2.4.3 Case 3: Study of PSO Algorithm for Feature Selection

PSO is a population-based computational intelligence technique, proposed to solve optimization problems which are presented by Chen, *et al* [92] for feature selection and were applied to an actual case on the diagnosis of obstructive sleep aponia. The

PSO technique employs a population, referred to as a swarm, consisting of individuals known as particles, to identify the optimal solutions. A particle serves as a potential solution to the situation at hand. Assume that the search space is D-dimensional and the i^{th} a particle of a swarm can be a D-dimensional position vector $X_i = [x_{i1}, x_{i2}, \dots, x_{iD}]$. The particle velocity of particle i is denoted as $V_i = [v_{i1}, v_{i2}, \dots, v_{iD}]$. The PSO algorithm also takes into account the best position visited by the particle that provides the highest fitness value, which is represented by $PB_i = [pb_{i1}, pb_{i2}, \dots, pb_{iD}]$, and the best position explored so far, which is represented by $GB = [gb_1, gb_2, \dots, gb_D]$. Each particle is updated according to equation (23).

Swarm intelligence is a distributed solution to complex problems which intends to solve complicated problems through interactions between simple agents and their environment.

$$v_{id}^{new} = w \times v_{id}^{old} + c_1 r_1 \left(pb_{id}^{old} - x_{id}^{old}\right) + c_2 r_2 \left(gb_d^{old} - x_{id}^{old}\right), d = 1,2,\dots,D \tag{23}$$

where, c_1 denotes the cognitive learning element and c_2 represents the social learning factor. These two elements are positive constants with typical values ranging from 0 to 4. Inertia weight w steadily decreases particle velocity, keeping the swarm within restraint. The value of the attribute w is typically between 0.4 and 0.9. The Random variables r_1 and r_2 have a uniform distribution from 0 to 1. Upper and lower bound limit velocity vectors provide support to prevent particles from moving too quickly in the search space [93]. As a result, particle velocities are constrained to the range $[v_{min}, v_{max}]$. Each particle goes to a new solution according to equation 24.

$$x_{id}^{new} = x_{id}^{new} + v_{id}^{new}, d = 1,2,\dots,D; i = 1,2,\dots,N \tag{24}$$

Where, N is the size of the swarms.

Solution Representation

To solve the feature selection problem using PSO, we represented each feature with a binary digit. A binary value of 0 indicates that a feature is not selected, while a value of 1 means that a feature is selected. We encoded each particle using a binary alphabet string, where each character in the string represented a feature. For instance, the particle 101000 with six features means that the first and third features are selected. Finally, we updated the dimension d of each particle i using equation (4).

$$x_{id}^{new} = \begin{cases} 1, & if\ sigmoid\ (v_{id}^{new}) > U(0,1) \\ 0, & otherwise \end{cases} \tag{25}$$

Where, the sigmoid (v_{id}^{new}) is $1/1 + e^{-v_{id}^{new}}$.

Population Initialization

In PSO, a population of particles can be initialized randomly or using a well-adapted PSO. Nevertheless, it is believed that initializing PSO with a good starting point can lead to improved outcomes. This research employed a complete random selection process for bit values of 0 and 1 with a probability of 0.5. Specifically, if $U(0,1) > 0.5$, x_{id}^{0} is set to 1, otherwise, x_{id}^{0} is set to 0.

Assessing the Fitness of Every Particle within a Group

The fitness function plays a crucial role in the PSO algorithm as it determines the quality of each particle's solution. The ultimate goal of PSO is to find the global optimum solution through a cooperative search. The particles with the best fitness values represent the most promising solutions to the problem, and they are tracked and recorded to guide the swarm towards convergence. By continuously evaluating the fitness of each particle, the PSO algorithm can iteratively refine and improve the swarm's collective performance until an optimal solution is found.

2.4.4 Case 4: Analysis of Grey Wolf Optimization Algorithm for Feature Selection

The GWO is a metaheuristic algorithm inspired by the social hierarchy and hunting behaviour of grey wolves. The GWO algorithm has been successfully applied to various optimization problems, including feature selection in disease detection.

The social structure of grey wolves is based on four adjustable parameters:

- Alpha wolves (α) are the leaders of the pack who make the hunting decisions. The alpha wolves are the most dominating members of the pack, and their acts are imitated by the rest of the pack. The alpha wolf does not always have to possess the most physical strength among the pack members, but they must excel in effectively overseeing and leading the whole pack.
- Beta wolves (β) occupy the second position in the hierarchy. They advise the alpha and assist in making decisions. If the alpha wolf dies or becomes old, the beta wolf takes over their position. Their primary duty is to strengthen the alpha's orders and maintain consistency among the subordinate tiers of the hierarchy.
- Omega wolves (ω) reside at the lowest rung of the hierarchy and function as individuals who are blamed for the mistakes or problems of others. They submit to the top wolves within the hierarchy and consume their meal after everyone else.Delta wolves (δ) are subordinate wolves in the pack who are neither alpha, beta, nor omega. Delta wolves report to alphas or betas but have authority over omega wolves [94].

The feature selection problem in disease detection involves identifying a subset of relevant features from a large pool of potential features, which can be used to improve the accuracy of disease diagnosis. GWO has been shown to effectively address this problem by selecting an optimal subset of features. The GWO algorithm has been widely used for feature selection in disease detection.

Several studies have explored and extended the potential of GWO in this area, demonstrating its superior performance compared to other feature selection methods. For example, in a recent study by Zheng et al. [95], GWO was applied for feature selection in the detection of breast cancer using mammography images. The study used a dataset consisting of 1,000 mammography images with various features such as texture, density, and size. GWO was used to select the most relevant features, which were then fed into a SVM to diagnose the presence of breast cancer. The results showed that GWO outperformed other feature selection methods such as genetic algorithm and particle swarm optimization in terms of classification accuracy.

Another study by Jin et al. [96]applied a hybrid version of GWO and a CNN for feature selection in the diagnosis of COVID-19 using chest X-ray images. The study used a dataset consisting of 2,040 chest X-ray images with various features such as ground-glass opacities, consolidation, and pleural effusion. GWO was used to select the most relevant features, which were then fed into a CNN to diagnose the presence of COVID-19. The hybrid GWO-CNN approach outperformed other feature selection methods such as GA and PSO in terms of accuracy, sensitivity, and specificity.

In another study by Jawarneh et al. [97], a modified GWO algorithm was proposed that incorporates a fitness scaling factor to improve the search process and avoid premature convergence. The modified GWO algorithm was applied for feature selection in the diagnosis of liver disease using ultrasound images, and the results showed improved performance compared to the original GWO algorithm.

One recent study published in the Journal of Medical Systems used GWO for selecting relevant features for the ultrasound-based detection of thyroid cancer. The study used a dataset consisting of 300 ultrasound images with various features such as echogenicity, shape, and margin. GWO was used to select the most relevant features, which were then fed into a SVM to diagnose the presence of thyroid cancer. The results showed that GWO outperformed other feature selection methods such as genetic algorithm and PSO in terms of classification accuracy. The study concluded that GWO can be a useful tool for the early detection of thyroid cancer.

Another recent study published in the Multimedia tools and application applied a hybrid version of GWO and an artificial neural network (ANN) for feature selection in the diagnosis of diabetes using retinal images. The study used a dataset consisting of 1,000 retinal images with various features such as microaneurysms, haemorrhages, and exudates. GWO was used to select the most relevant features, which were then fed into an ANN to diagnose the presence of diabetes. The hybrid GWO-ANN approach outperformed other feature selection methods such as GA and PSO in terms of accuracy, sensitivity, and specificity. The study concluded that the hybrid GWO-ANN approach can be a promising tool for the early detection of diabetes [98].

Furthermore, some studies have proposed modifications to the original GWO algorithm to improve its performance in feature selection for disease detection. One recent study published in the International Journal of Biological Macromolecules proposed a modified GWO algorithm that incorporates a fitness scaling factor to improve the search process and avoid premature convergence. The modified GWO algorithm was applied for feature selection in the diagnosis of heart disease using electrocardiogram (ECG) signals, and the results showed improved performance compared to the original GWO algorithm. The study concluded that the modified GWO algorithm can be a useful tool for the early detection of heart disease [99].

Since the initial application of GWO in feature selection for disease detection, many studies have been conducted to investigate its potential in different medical diagnosis problems. For instance, a study applied the GWO algorithm to select features in the detection of cervical cancer using Pap smear images. The study used a dataset of 918 images and compared the performance of GWO with other feature selection methods. The results showed that GWO outperformed other methods in terms of accuracy and F-measure [100].

Another study applied the GWO algorithm for feature selection in the diagnosis of liver diseases using ultrasound images. The study used a dataset consisting of 200 images with various features such as echogenicity, shape, and texture. GWO was used to select the most relevant features, which were then fed into a k-nearest neighbor (k-NN) classifier to diagnose the presence of liver diseases. The results showed that GWO outperformed other feature selection methods in terms of classification accuracy and sensitivity [101].

In another study, the GWO algorithm was applied for feature selection in the diagnosis of prostate cancer using MRI images. The study used a dataset consisting of 100 images with various features such as volume, shape, and texture. GWO was used to select the most relevant features, which were then fed into a SVM classifier to diagnose the presence of prostate cancer. The results showed that GWO outperformed other feature selection methods such as GA and PSO in terms of classification accuracy [102].

Furthermore, some studies have proposed modifications to the original GWO algorithm to improve its performance in feature selection for disease detection. For example, one recent study proposed a modified GWO algorithm that incorporates an adaptive inertia weight to balance exploration and exploitation. The adapted GWO algorithm was utilized to select features for breast cancer diagnosis based on mammogram images and the results showed improved performance compared to the original GWO algorithm [103].

Another study proposed a hybrid GWO and particle swarm optimization (PSO) algorithm for feature selection in the diagnosis of brain tumor using MRI images [104]. The study used a dataset consisting of 100 images with various features such as volume, shape, and texture. The hybrid GWO-PSO algorithm was used to select the most relevant features, which were then fed into a random forest classifier to diagnose the presence of brain tumor. The results showed that the hybrid algorithm outperformed other feature selection methods in terms of classification accuracy and sensitivity.

In conclusion, the GWO algorithm has shown great promise for feature selection in disease detection, with ongoing research exploring its potential in various medical diagnosis problems. The development of modified versions of GWO and hybrid approaches with other machine learning methods further expand its potential and open up new opportunities for optimizing disease diagnosis accuracy.

2.4.5 Case 5: Analysis of Moth Flame Optimization Algorithm-based Feature Selection Method for Improving the Classification Performance

Abu Khurmaa, and colleagues [105] introduced the idea of integrating the MFO algorithm into a wrapper FS framework to improve classification tasks within medical contexts. We've recognized the beneficial attributes of MFO in this domain and are eager to explore its effectiveness in tackling the FS problem. It's important to highlight that the original MFO algorithm was initially tailored for optimizing continuous search spaces, whereas FS deals with a binary problem. Consequently, certain adjustments are required to enable the native optimizer to operate effectively in a high-dimensional binary search space. Therefore, it is necessary to adapt the MFO algorithm to suit the specific requirements of the FS problem to ensure effective and efficient optimization.

The MFO algorithm represents a newly devised approach within the realm of SI optimization. Diverging from conventional optimizers that draw inspiration from the survival tactics of various creatures, MFO stands out due to its distinctive optimization philosophy. Moths employ a transverse orientation mechanism, allowing them to navigate in a straight trajectory while maintaining a consistent angle relative to distant light sources, such as moonlight. However, this orientation mechanism becomes ineffective when the moths fly closer to the light sources. The behaviour of moths was converted into a mathematical model using the MFO algorithm [106].

The MFO algorithm is crafted using a population-centered approach, where moths are treated as individuals or potential solutions within the problem's search space. These moths have the capability to navigate within both one-dimensional and multidimensional spaces by configuring their position vectors. The development of the MFO methodology is guided by two distinct design viewpoints: structural and functional. From a structural perspective, the representation of a single moth and a population of moths involves utilizing a vector (a one-dimensional array) and a matrix (a two-dimensional array), respectively. To illustrate, the M matrix, corresponding to a population comprising 'n' moths in 'd' dimensions, is exemplified in equation 32.

$$M = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,d} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ m_{n,1} & m_{n,2} & \cdots & m_{n,d} \end{bmatrix} \tag{32}$$

Within the MFO optimizer, each moth serves as a potential solution, occupying a distinct position within the search space. To gauge the quality and proximity of each moth to the optimal solution, an assessment is conducted using predefined evaluation criteria. This assessment process employs a custom fitness function that is tailored to the specific problem being studied. The resultant fitness values are then stored in a one-dimensional array, as illustrated in equation 33. In this equation, "n" denotes the number of moths, while "OM1" signifies the fitness value associated with the first moth, which corresponds to the initial row in the M matrix.

$$M = \begin{bmatrix} OM_1 \\ OM_2 \\ \vdots \\ OM_n \end{bmatrix} \tag{33}$$

Equations (34) and (35) represent other important structural components of the MFO optimizer, which include the matrix of flames and their respective fitness values. In equation (34), "n" represents the number of moths, while "d" represents the number of dimensions. The moths and flames are both solutions, but they differ in their update strategy during the optimization process. Moths represent candidate solutions that explore the search space, while flames represent the best solutions discovered so far. While moths persistently seek improved solutions, each flame acts as a marker that is hoisted when a moth uncovers a superior solution. This mechanism ensures that the optimizer never loses the best solution, making it a highly effective approach for optimization problems.

$$F = \begin{bmatrix} F_{1,1} & F_{1,2} & \cdots & F_{1,d} \\ F_{2,1} & F_{2,2} & \cdots & F_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ F_{n,1} & F_{n,2} & \cdots & F_{n,d} \end{bmatrix} \tag{34}$$

$$OF = \begin{bmatrix} OF_1 \\ OF_2 \\ \vdots \\ OF_n \end{bmatrix} \tag{35}$$

The MFO optimizer was developed using two design perspectives: structural and functional. The functional design of MFO involves three tuples, as presented in equation 36. The initialization function, I, generates random moths with their corresponding fitness values, using methods such as the random distribution function, as shown in equation 37. The update function, P, is responsible for changing the positions of moths in the search space. This is accomplished by receiving a matrix of moths as input and delivering the revised moth matrix as output, as shown in equation 38. Finally, the termination condition is checked using the T function, as shown in equation 39. This function checks whether the termination condition is met or not. By following this functional design perspective, MFO can efficiently explore the search space and converge to optimal solutions.

$$MOF = (I, P, T) \tag{36}$$

$$I: \Theta \longrightarrow \{M, OM\} \tag{37}$$

$$P: M \longrightarrow M \tag{38}$$

$$T: M \longrightarrow \{true, false\} \tag{39}$$

After the initialization step is finished, the P function is iteratively carried out until a predefined stopping criterion is satisfied. The role of the P function is to emulate the spiral movement of moths circling around the flames, as expressed in equation 40. In this equation, M_i refers to the i^{th} moth, F_j denotes the j^{th} flame, and S represents the spiral function. The logarithmic spiral is the commonly used function for modelling the spiral motion, as depicted in equation 41. Here, D_i represents the distance between the i^{th} moth and the j^{th} flame, as given in equation 42. The parameter "b" dictates the configuration of the logarithmic spiral, while "t" represents a random number within the interval [-1,1]. A value of t=1 indicates that the moth is in its closest position to the flame, whereas a value of t= -1 indicates that the moth is farthest away from the flame. To extend the exploration of the search space, the parameter "t" is examined within the interval [r,1], with "r" decreasing linearly from -1 to -2 as the iterations advance, thus promoting exploitation.

$$Mi = S(Mi, Fj) \tag{40}$$

$$S(Mi, Fj) = Di.e^{bt}.cosS(2\Pi) + Fj \tag{41}$$

$$Di = |Mi - Fj| \tag{42}$$

The MFO optimizer also includes an adaptive mechanism for the number of flames, which decreases during the iteration process. This mechanism is depicted in Equation 43, where "l" represents the current iteration count, "N" denotes the maximum number of flames, and "T" signifies the maximum number of iterations. This adaptive mechanism was introduced by Al-Tashi in 2022. The steps involved in the P function are summarized in Figure 2.3, which presents the pseudocode for the basic MFO algorithm.

Algorithm: Pseudocode MFO Algorithm

Input: $Max\_iteration$, n (number of moths), d (number of dimensions)

Output: Approximated global solution

Initialize the position of moths

while $l \leq Max\_iteration$ **do**

Update flame no using equation (43)

$OM = Fitness\ Function\ (M)$;

if $l == 1$ **then**

$F = sort(M)$;

$OF = sort(OM)$;

else

$F = sort(M_{l-1}, M_l)$;

$OF = sort(OM_{l-1}, OM_l)$;

end if

for $i = 1\!:n$ **do**

for $j = 1\!:d$ **do**

Update r and t;

Calculate D using equation (42) with respect to the corresponding month;

Update $M(i, j)$ using equation (40) and equation (41) with respect to the corresponding month;

```
        end for
    end for
    l = l + 1
end while
```

Figure 2.3: Pseudocode for MFO Algorithm

$$FlameNo = round(N - l * (N - 1)/T) \tag{43}$$

Maintaining a balance between exploration and exploitation is achieved by the gradual reduction in the number of flames.

2.4.6 Analysed Results for the Applied Metaheuristic Optimization Algorithm

The creation of this new dataset is a significant development in disease detection and diagnosis prediction systems in the medical field. The use of images collected from various sources, distinct from the melanoma images, ensures that the dataset is diverse and representative of a broad range of medical conditions.

The validation of the dataset using different medical datasets from UCI, Keel, and Kaggle data repositories further reinforces the credibility and reliability of the dataset. The use of a proprietary medical knowledge base and a commercial rule-based AD system to generate patients' characteristics is a novel approach that adds value to the dataset.

The dataset contains detailed information on the patients, including their socio-demographic data, the specific disease they are experiencing, a collection of symptoms and previous occurrences associated to this pathology, and a differential diagnosis. This comprehensive dataset offers a thorough understanding of the patients' ailments. The inclusion of non-binary symptoms and antecedents is particularly noteworthy, as it can lead to more efficient and natural interactions between ASD/AD systems and patients.

The dataset's potential for disease detection and diagnosis prediction systems in the medical domain is immense. It can enable researchers to develop more accurate and effective systems, leading to improved patient outcomes. Moreover, the dataset can be a valuable resource for medical professionals, students, and researchers interested in exploring various medical conditions and their characteristics. Overall, this new dataset is a significant contribution to the medical field and has the potential to facilitate advancements in disease detection and diagnosis prediction systems.

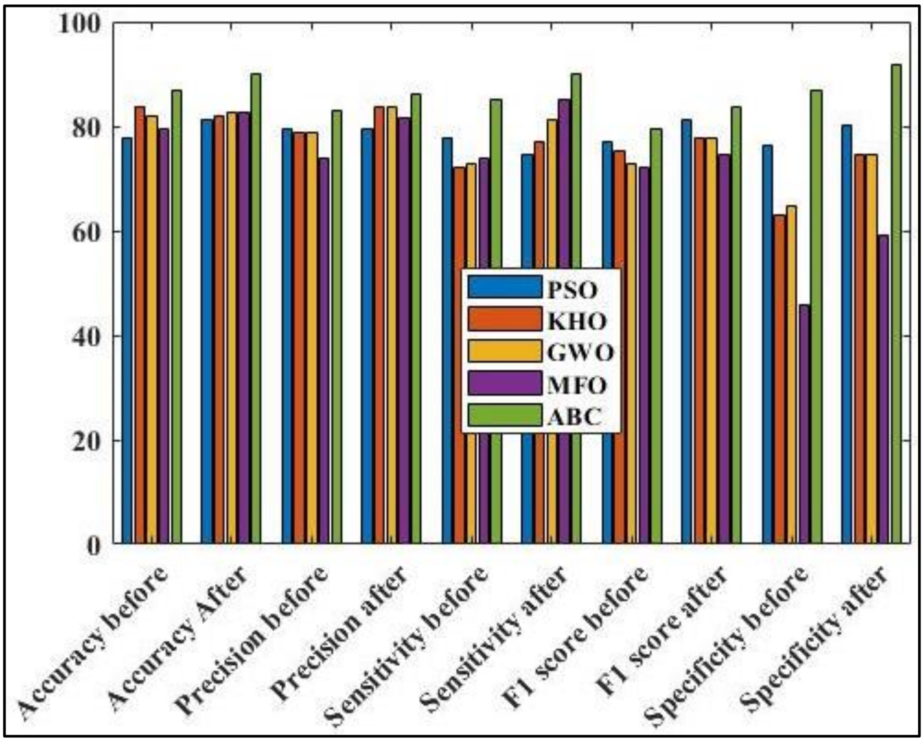


Figure 2.4: Performance Analysis Before and after using Metaheuristic Optimization Algorithm

The graph in Figure 2.4 illustrates the performance of all classes both before and after applying optimization techniques. This study utilized several swarm intelligence optimization methods, including PSO, KHO, GWO, MFO, and ABC algorithms. After the optimization process, all disease detection and classification techniques achieved a higher accuracy of 80%, a sensitivity of 74%, a specificity of 81%, an F1-score of 84%, and a precision of 80% for the Particle Swarm Intelligence technique. The KHO algorithm produced average accuracy, precision, sensitivity, specificity, and F1-score values of 81%, 87%, 76%, 75%, and 78%, respectively. Meanwhile, the GWO algorithm achieved an average sensitivity of 80.5%, a specificity of 85%, an accuracy of 83%, an F1-score of 77%, and a precision of 85%. On the other hand, the MFO algorithm obtained an average sensitivity of 87%, a specificity of 59%, an accuracy of 83%, an F1-score of 76%, and a precision of 82%. The ABC technique reached an average sensitivity of 90%, a specificity of 94%, an accuracy of 91%, an F1-score of 86%, and a precision of 84%. By incorporating these optimization techniques, the performance of the study was enhanced, resulting in improved performance values as demonstrated in Figure 2.4.

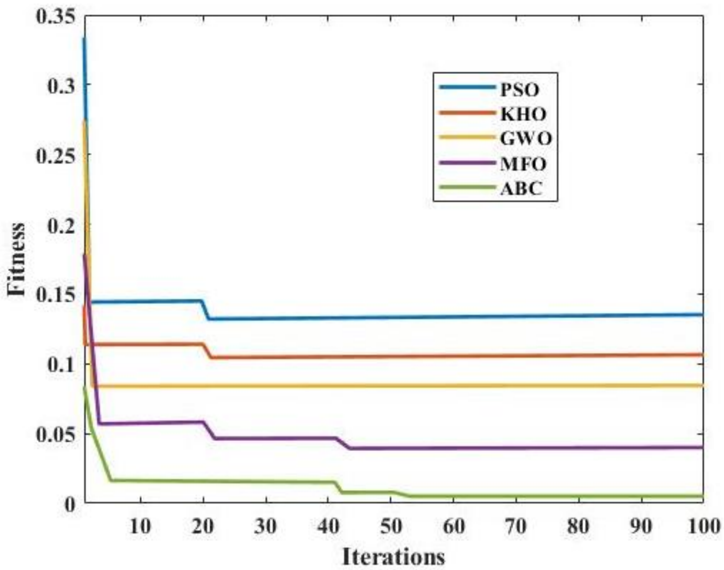


Figure 2.5: Convergence Curve for Optimization Techniques

The convergence curve of this optimization technique is presented in figure 2.5. It analyses the convergence curve for SO, KHO, GWO, MFO, and ABC algorithms. The algorithm has demonstrated improved management of premature convergence across a majority of datasets, in contrast to the observed behaviour of the BGWO, BBA, and BCS algorithms which exhibited premature convergence on most of the datasets.

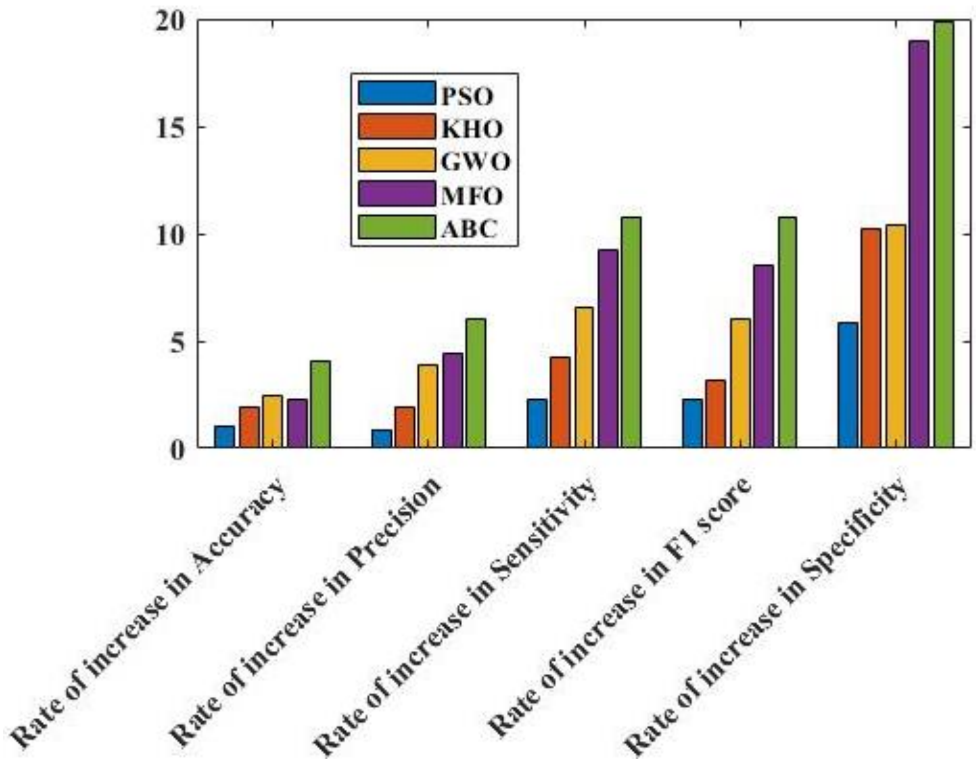


Figure 2.6: Rate of Increase in Performance Values Using Optimization Techniques

Figure 2.6 shows the effect of using optimization techniques for image segmentation and feature selection. The figure depicted that the rate of increase in accuracy became 0.9%, 2%, 2.3%, 2.1%, and 4.2% using PSO, KHO, GWO, MFO, and ABC algorithms. The rate of increase in precision became 0.6%, 1.8%, 4%, 4.3% and 5.8% using PSO, KHO, GWO, MFO, and ABC algorithms, respectively. Sensitivity also increased after using the optimization techniques became 2%, 3.9%, 5.15%, 9.2, and 11% of PSO, KHO, GWO, MFO, and ABC algorithms. However, the rate of F1 score increased by 2.3%, 3%, 6.1%, 8.8%, and 11% for the PSO, KHO, GWO, MFO, and ABC-based optimization techniques. Accordingly, the rate of increase in specificity became 5.8%, 10.2%, 10.3%, 19.12%, and 19.95% for PSO, KHO, GWO, MFO, and ABC algorithms, respectively.

2.5 Summary

In the healthcare optimization sector, the primary aim is to elevate critical performance metrics such as accuracy, sensitivity, F-measure, precision, and specificity. Various practical approaches rooted in optimization techniques present a spectrum of optimization challenges. Notably, a significant hurdle in healthcare optimization involves grappling with NP-hard problems through intricate mathematical functions. Examining the current landscape of healthcare optimization reveals notable progress, especially in the domains of disease prediction and classification through the employment of metaheuristic optimization methodologies. The study reveals that SI models effectively tackle several challenges in illness prediction. Nevertheless, there are still some constraints. Moreover, several current models are focused on optimizing healthcare performance from a singular perspective. However, it is important to take into account some constraints and drawbacks while using the SI algorithm in the healthcare sector. Some of the limitations are:

1) ***The optimization issues of healthcare integrated with recent SI models****:* The experimental findings demonstrated that using SI (Statistical Inference) yielded superior categorization in illness prediction and obtained more favorable outcomes. With the advancement of SI, namely the introduction of BES [107], Sparrow Search Algorithm (SaSA) [108], Aquila Optimizer [109] and Manta Ray Foraging Optimization [110].
2) ***Integration with practical implication****:* Based on the aforementioned study on metaheuristics, it has been shown that some SI techniques outperform the current models in certain circumstances. However, several conventional SI techniques rely on simulation and only a limited number of methods are actually used in real-life healthcare. Conducting an assessment of these models in an actual healthcare setting will be a valuable area of future study.
3) ***Improvement in algorithm****:* Metaheuristic techniques are population-based and include factors such as population size, algorithm variables, and maximum iteration for initiating the optimization process. The effectiveness of the model is highly influenced by the selection of these parameters. This demonstrates the need of appropriately adjusting parameters before applying them to a problem. In order to address this issue, several studies have lately focused on analyzing self-adaptive models.

4) ***Increasing computational efficiency:*** to improve computation efficiency, the researchers are analyzing parallelization models (which is well applicable for EA and SI methods) and multi algorithm that integrate two or more search algorithms.

The process of diagnosing different diseases after examining a patient's symptoms and indicators is known as disease detection. Various diagnostic tests are employed to obtain signs and symptoms in order to diagnose various illnesses. Asymptotic cases—those in which the symptoms are not observed—make this process more difficult. Since the patients in all these circumstances are asymptomatic, various tests are necessary to diagnose multiple disorders. The significance of metaheuristic algorithms for various illness detection methods is discussed in this analysis. The method is unique in that it applies metaheuristic algorithms to Decision Support Systems for the purpose of identifying different diseases either entirely on their own or in conjunction with larger, hybrid systems. This analysis looks at new metaheuristic algorithms that conduct image segmentation, feature selection, and classification and how they might be used to diagnose and predict diseases. The different metaheuristic algorithms like KHO,ABC,PSO,GRO and MFO are considered. When dealing with classification difficulties, it is crucial to meticulously choose a limited set of characteristics for the purpose of prediction. By choosing this specific subset of characteristics, the processing time may be significantly reduced, while still achieving robust and better outcomes compared to employing the whole collection of features. Selecting the most effective subset for classification, which optimizes performance, poses a significant challenge. In response to this issue, a metaheuristic optimization algorithm is introduced. This algorithm is designed to enhance classification performance by addressing the inherent difficulty associated with selecting the optimal subset. The performance of disease detection is improved by utilizing the metaheuristic optimization technique. The accuracy of the optimization algorithms produces 80%, 81%, 83%, 83%, and 91% for PSO, KHO, GWO, MFO, and ABC algorithms. The rate of increase in accuracy became 0.9%, 2%, 2.3%, 2.1%, and 4.2% for PSO, KHO, GWO, MFO, and ABC algorithms. This analysis portrayed that the classification and detection performance is improved by using a metaheuristic algorithm. In future studies, a prediction methodology is designed and developed using a metaheuristic approach for diagnostic tests and early detection of disease, which improves the system performance, respectively.

Chapter 3

Design and develop the prediction methodology for diagnostic tests and disease using the Metaheuristic approach

3.1 Introduction

AI-powered predictive systems can be used to predict diagnostic tests based on symptoms in real-time. Davenport et al [57] discussed the potential applications of AI in healthcare along with ethical implications in their research. The applications of AI in healthcare are vast and continue to expand as AI and healthcare technologies advance.

- **i.** Early Disease Detection: AI-based predictive systems can analyze symptoms reported by patients and provide early detection of diseases. For instance, predicting the likelihood of developing conditions such as diabetes, heart disease, or certain types of cancer based on symptoms and risk factors.
- **ii.** Triage and Resource Allocation: AI systems can help in triaging patients and efficiently allocating healthcare resources. By analyzing symptoms, severity, and demographic data, the system can prioritize patients based on the urgency of their condition, ensuring that critical cases receive immediate attention.
- **iii.** Differential Diagnosis Support: AI models can assist healthcare professionals in differential diagnosis, where multiple diseases may have similar symptoms. By considering a patient's symptoms, medical history, and other relevant factors, the system can generate a ranked list of potential diagnoses to aid clinicians in making accurate and timely decisions.
- **iv.** Personalized Medicine: AI-powered systems can analyse a patient's symptoms and medical history to predict how they may respond to different treatment options. This can enable personalized medicine, where treatment plans are tailored to individual patients, optimizing efficacy and minimizing adverse effects.
- **v.** Telemedicine Triage: In telemedicine settings, AI can play a vital role in remotely assessing and triaging patients. By analyzing symptoms reported

during virtual consultations, the system can provide recommendations for further tests, referrals, or appropriate levels of care.

vi. Public Health Surveillance: Artificial intelligence (AI) powered prediction systems have the ability to observe and examine symptom data from several sources, including social media and health surveys, in order to detect possible illness outbreaks or monitor the transmission of transmissible illnesses in real-time. This information has significant value for public health initiatives and resource planning.

3.2 Competitive Multiverse Optimization Algorithm

The Competitive Multi-Verse Optimizer (CMVO) is a metaheuristic optimization algorithm inspired by the concept of multiple universes competing with each other. It was proposed by Benmessahel et al [111] as an optimization technique for solving complex optimization problems. Figure 3.1 depicts the comprehensive procedure of the MVO algorithm. The chart clearly demonstrates that a universe with a lower inflammation rate has a tendency to accept more objects from superior universes, even the best universe generated so far, in order to enhance its inflammation rate.

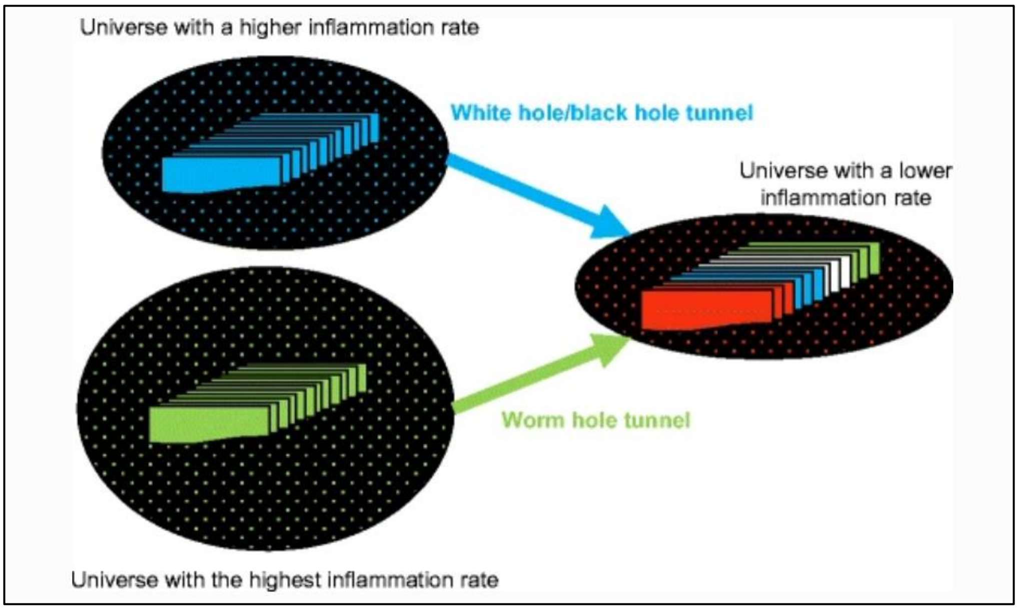


Figure 3.1: The overall process of MVO algorithm

The flowchart of basic functionality of CMVO algorithm is provided in figure 3.2

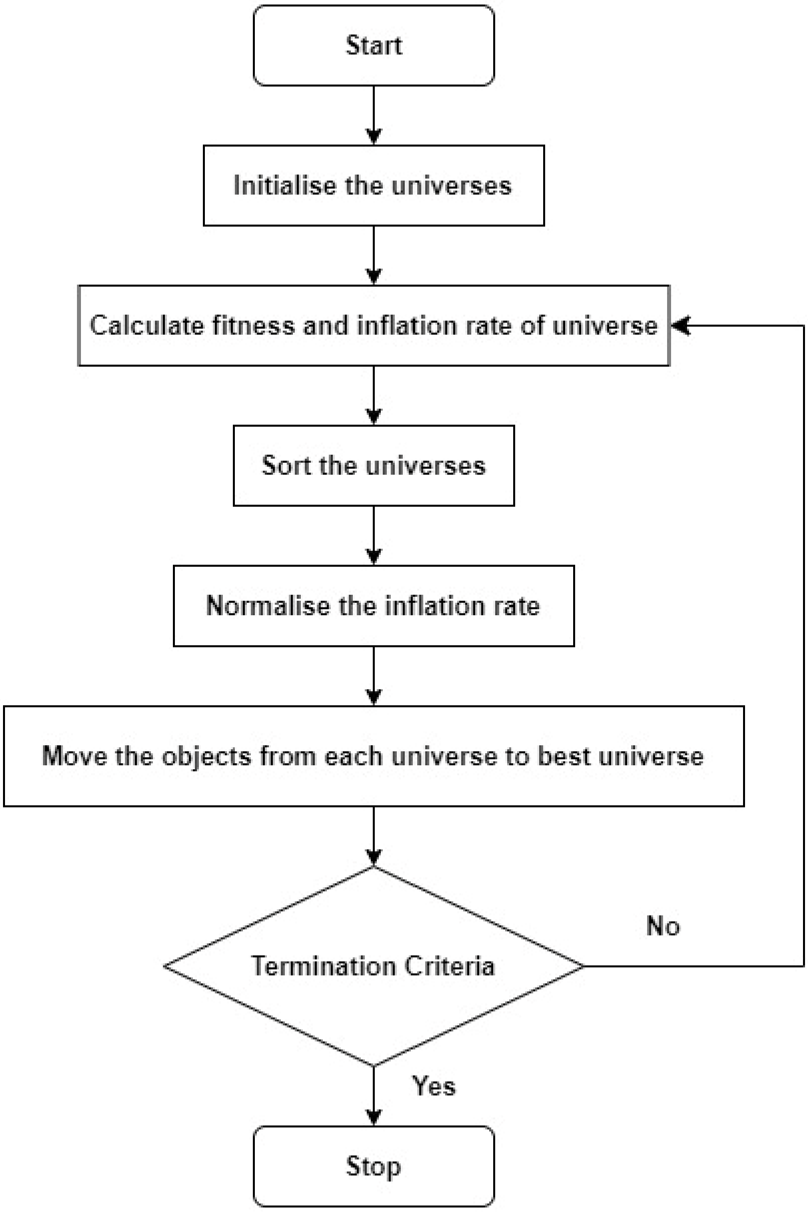


Figure 3.2 Flowchart of CMVO algorithm

Here are the key features and steps involved in CMVO:

i. Initialization:

- Initialize a population of potential solutions (individuals) randomly or using some heuristic approach.
- Each individual represents a point in the search space.

ii. Multiverse Representation:

- CMVO introduces the concept of multiple universes, each representing a different subpopulation of individuals.
- The number of universes is typically predefined and fixed.

iii. Universe Competition:

- In each iteration, the universes compete with each other to improve their fitness and evolve towards better solutions.
- Competition is driven by the fitness values of individuals in each universe.

iv. Universe Evolution:

- During competition, some universes may perform better and dominate others.
- The dominating universes undergo evolution, where individuals with higher fitness values are selected and undergo certain operations such as crossover, mutation, or local search to generate new solutions.
- This process allows the dominating universes to explore and exploit the search space effectively.

v. Universe Migration:

- To preserve variety and prevent abrupt convergence, CMVO incorporates a migration mechanism among universes.
- Some individuals from less dominant universes are migrated to more dominant universes to inject diversity and explore new regions of the search space.

vi. Termination:

- The optimization process continues for a predefined number of iterations or until a termination criterion is met.
- Termination criteria can be a maximum number of iterations, reaching a specific fitness value, or no significant improvement over a certain number of iterations.

vii. Output:

- The output of CMVO is the best solution found, which corresponds to an individual with the highest fitness value obtained during the optimization process.

CMVO aims to balance exploration and exploitation by allowing universes to compete and evolve. The competition encourages exploration by allowing less dominant universes to migrate and explore new areas, while evolution within dominant universes promotes exploitation by refining solutions in promising regions of the search space. The Competitive Multi-Verse Optimizer (CMVO) can be used in various optimization scenarios, particularly when dealing with complex problems that require balancing exploration and exploitation.

The Competitive Multi-Verse Optimizer (CMVO) algorithm can be applied in healthcare solutions to address various optimization problems. Here are a few examples of how CMVO can be used in healthcare:

i. Feature Selection: In healthcare, it is crucial to identify the most relevant features or variables that contribute to a particular medical condition or outcome. CMVO can be used to optimize the selection of features by evaluating different combinations of features and maximizing a performance metric (e.g., accuracy, sensitivity, specificity) in classification or regression tasks.

ii. Treatment Optimization: CMVO can be applied to optimize treatment plans for patients based on various factors such as patient demographics, medical history, and available resources. By formulating the treatment plan as an optimization problem, CMVO can explore different combinations and schedules of treatments to maximize patient outcomes while considering constraints such as drug interactions, dosage limits, and cost.

iii. Resource Allocation: Healthcare organizations often face challenges in efficiently allocating limited resources such as hospital beds, medical equipment, and healthcare staff. CMVO can be utilized to optimize resource allocation by considering factors like patient demand, urgency, and resource availability. The algorithm can explore different allocation strategies and prioritize resources to maximize healthcare service efficiency and patient satisfaction.

iv. Clinical Trial Design: Designing effective clinical trials is essential to evaluate the safety and efficacy of new drugs or treatment interventions. CMVO can aid in optimizing the design of clinical trials by determining the optimal sample size, treatment allocation ratio, patient recruitment strategies, and other trial parameters. The algorithm can search for the best combination of trial design factors to enhance statistical power and reduce cost and time.

v. Disease Diagnosis and Prognosis: CMVO can be employed to optimize the diagnostic and prognostic models in healthcare. By optimizing the model parameters or feature selection, CMVO can enhance the accuracy and reliability of disease diagnosis and prognosis. This can contribute to improved patient outcomes, personalized medicine, and early detection of diseases.

These are just a few examples of how CMVO can be applied in healthcare solutions. The algorithm's ability to explore multiple universes and compete for the best solutions makes it suitable for various optimization problems in the healthcare domain. It offers the potential to enhance decision-making, resource utilization, and patient care in the healthcare industry.

3.3 Water Wave Optimization Algorithm

Water Wave Optimization (WWO) is a metaheuristic optimization algorithm inspired by the behaviours of water waves. To efficiently search in solution space, the WWO primarily mimics the processes of propagation, reflection, and shattering activities. Figure 3.3 depicts the various wave forms in shallow and deep sea. A water wave moves from deep to shallow seas, increasing in height, decreasing in wavelength, and increasing in fitness value.

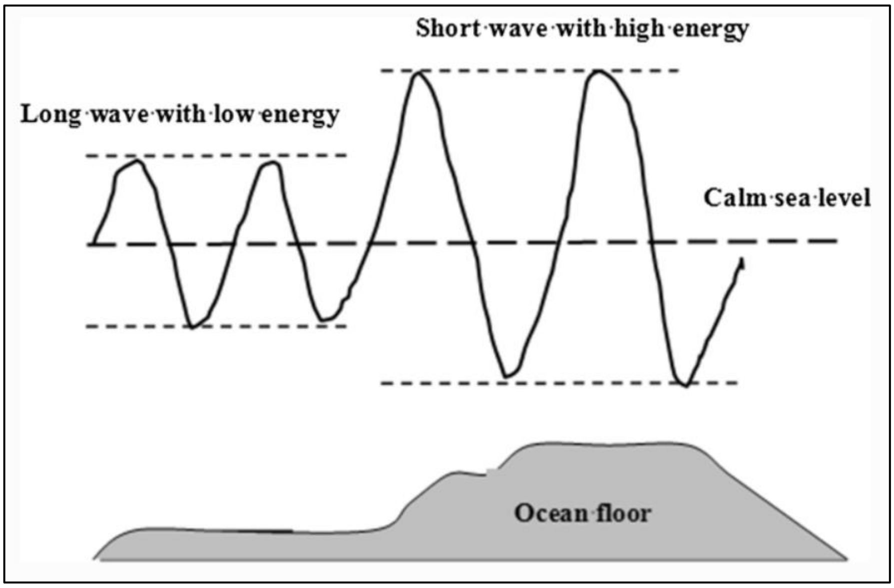


Figure 3.3: various forms of waves

It was proposed by Zheng [112]. It simulates the natural phenomenon of waves in bodies of water to solve optimization problems. WWO has been applied to various optimization tasks, including engineering design, data clustering, image segmentation, and healthcare management.

The basic idea of WWO is to simulate the behaviour of water waves, which involves three main components:

i. Wave Initialization: WWO starts by initializing a population of water waves, which represent potential solutions to the optimization problem. Each wave represents a candidate solution and is characterized by its position in the search space.

ii. Wave Propagation: The waves in WWO undergo several wave propagation operations, mimicking the behaviour of real waves:

- Reflection: Waves reflect from boundaries or obstacles in the search space, which helps explore new regions. The reflection operation modifies the positions of the waves based on the reflection equation.
- Refraction: Waves refract towards regions with higher fitness values, similar to how waves bend towards shallow areas in water bodies.

Refraction operation adjusts the positions of waves based on the refraction equation.

- Dispersion: Waves disperse to promote diversification and exploration in the search space. The dispersion operation adds a random perturbation to the positions of the waves.

These wave propagation operations are performed iteratively to update the positions of the waves and explore the search space efficiently.

iii. Wave Selection: After each iteration of wave propagation, the fitness of each wave is evaluated using an objective function specific to the optimization problem. The waves with higher fitness values are more likely to be selected for the next iteration, influencing the exploration-exploitation trade-off.

The process of wave propagation and selection continues for a specified number of iterations or until a termination condition is met. The algorithm aims to converge towards an optimal solution by iteratively adjusting the positions of the waves based on their fitness values. WWO offers advantages such as simplicity, robustness, and flexibility in handling different optimization problems. It provides a balance between exploration and exploitation, allowing the algorithm to search effectively across the solution space.

In healthcare solutions, the Water Wave Optimization (WWO) algorithm can be applied to perform exploitation by focusing on refining and improving promising solutions within the search space. Exploitation in WWO involves the following aspects:

i. Treatment Optimization: WWO can be utilized to optimize treatment plans for patients. By considering various parameters such as medication dosages, therapy durations, and treatment schedules, WWO can search for the optimal combination of treatment factors that maximize patient outcomes while minimizing side effects or costs.

ii. Resource Allocation: WWO can aid in optimizing resource allocation in healthcare facilities. This can involve determining the optimal allocation of medical staff, equipment, and hospital beds to ensure efficient and effective healthcare delivery. The algorithm can consider factors like patient demand, availability of resources, and operational constraints to find the best allocation strategy.

iii. Patient Scheduling: WWO can assist in optimizing patient scheduling, particularly in scenarios where multiple patients need to be scheduled for medical procedures or appointments. By considering factors such as patient preferences, urgency, and available resources, WWO can optimize the scheduling process to minimize waiting times, improve resource utilization, and enhance overall patient satisfaction.

iv. Healthcare Facility Layout: WWO can be applied to optimize the layout and design of healthcare facilities. By considering factors like patient flow, accessibility, and efficiency, the algorithm can determine the optimal arrangement of departments, equipment, and amenities within a healthcare facility to improve patient experience, reduce waiting times, and enhance the overall workflow.

3.4 Competitive verse Water Wave Optimization

1. Load and Pre-process the Dataset: Load the symptom database into a suitable data structure, such as a Pandas DataFrame. Pre-process the data by handling missing values, encoding categorical variables, and performing any necessary feature scaling or normalization.

2. Define the Objective Function: Define an objective function that measures the effectiveness of a diagnostic test recommendation based on the symptoms provided. This objective function can consider factors like accuracy, sensitivity, specificity, or a combination of these metrics.

3. Implement CMVO Algorithm:

- Initialize a population of universes (solutions) with random values within appropriate ranges.
- Evaluate the fitness of each universe using the defined objective function.
- Select the best universe(s) based on their fitness.
- Update the positions of universes using the CMVO update equation, which involves competition and migration between universes.
- Repeat the selection and update steps for a specified number of iterations.

4. Implement WWO Algorithm:
    - Initialize a population of water waves (solutions) with random values within appropriate ranges.
    - Evaluate the fitness of each wave using the defined objective function.
    - Sort the waves based on their fitness values.
    - Apply wave behaviour such as reflection, refraction, and dispersion to update the positions of waves.
    - Repeat the evaluation and update steps for a specified number of iterations.
5. Execute CMVO and WWO Optimization:
    - Run the CMVO algorithm to search for the optimal test recommendation based on symptoms. Keep track of the best solution found during the iterations.
    - Run the WWO algorithm as an alternative optimization approach to find an optimal solution for the test recommendation.
6. Compare and Evaluate Results:
    - Compare the results obtained from CMVO and WWO to determine the best diagnostic test recommendation based on their fitness values.
    - Evaluate the performance of the recommended test by validating it against known diagnoses from the dataset or domain experts.

7. Iterate and Fine-tune:

    - Iterate and fine-tune the optimization process by adjusting the algorithm parameters, such as population size, iteration count, or specific update rules, to improve the performance and convergence of the algorithms.

3.5 Dataset

The dataset used in this research is a rich collection of data gathered through a meticulously designed questionnaire. The questionnaire is verified and validated. This dataset aims to provide a comprehensive collection of participant responses which helped in analyzing and exploring the relationships between symptoms and potential diseases. The questionnaire covers a broad spectrum of symptoms, allowing for the identification of patterns and correlations that has assisted in predicting diagnosis.

3.5.1 Data collection and study population

The questionnaire data are collected from the Kolhapur district of Maharashtra, India for diagnosis test prediction. The dataset is formed by collecting the data from patients visiting the OPD (Out Patient Department) and those who are admitted to the department of Medicine in Dr D. Y. Patil Hospital, Kolhapur, Maharashtra. The data is collected from patients through the verified and validated questionnaire. The hospital staff has also helped in collecting data. Patients' demographic information is also recorded in it. The Institutional Ethics committee at D Y Patil Medical College, Kolhapur, Maharashtra has granted approval for this study. The following figure 3.4 shows the approval letter from the Ethics committee. The data is collected between April 2022 and December 2022. All the patients belong to the Kolhapur district in Maharashtra.

D. Y. PATIL
MEDICAL COLLEGE
KOLHAPUR

Constituent Unit of D. Y. Patil Education Society (Deemed to be University), Kolhapur.
Re-accredited by NAAC with 'A' Grade

Dr. Rakesh Kumar Sharma
Dean & Professor (Obst. & Gyn)

Padmashree Dr. D. Y. Patil
Founder President

Dr. Sanjay D. Patil
President

Ref No : DYPMCK/01/2022/IEC

Date : 06 / 01 /2022

INSTITUTIONAL ETHICS COMMITTEE, D. Y. PATIL MEDICAL COLLEGE, KOLHAPUR.

This is to certify that the research project titled,

Predicting diagnostic tests and probable disease by utilizing Meta heuristic optimization

Submitted by : **Mrs. Priyanka Shivaprasad More, Lovely Professional University, Punjab**

Under the supervision of appointed Guide (if any) : **Dr. Baljit Singh Saini**

Has been studied as expedited review and granted approval by the Institutional Ethics Committee (IEC) for the study with effect from 06th January 2022 with the following caveats:

1. If you desire any change in the protocol or standard recording document at any time, please submit the same to the IEC for information and approval before the change is implemented.
2. As per recommendations of ICMR, you must register your study with the Central Trials Registry- India (CTRI), hosted at the ICMR's National Institute of Medical Statistics (http://icmr-nims.nic.in). The registration details as provided by the website are to be submitted to the Institutional Ethics Committee within a period of 3 months from issue of this letter.
3. All serious and/or unexpected adverse events due to the drug/procedures tested in the study must be informed to the IEC within 24 hours and steps for appropriate treatment must be immediately instituted.
4. In case of injury/disability/death of any participant attributable to the drug/procedure under study, all compensation is to be made by the sponsor of the study.
5. The Chief investigator/Researcher must inform the IEC immediately if the study is terminated earlier than planned with the reasons for the same.
6. The final results of the study must be communicated to the IEC within 3 months of the completion of data collection.
7. The researcher must take all precautions to safeguard the rights, safety, dignity and wellbeing of the participants in the study.
8. The researcher must be up to date about all information regarding the risk/benefit ratio of any drug/procedure being used and any new information must be conveyed to the IEC immediately. The IEC reserves the right to change a decision on the project in the light of any new knowledge.
9. Before publishing the results of the study, the researcher must take permission from the Dean of the Institution.
10. Annual progress report should be submitted for all sponsored projects to the committee.
11. Unethical conduct of research in non-sponsored projects will result in withdrawal of the ethics approval and negation of all data collected till that date.

Shimpa

Dr. Mrs. Shimpa R. Sharma
(Member Secretary, IEC)

Address : 869, 'E Ward, D. Y. Patil Vidyanagar, Kasaba Bawada, Kolhapur- 416 006 (MS) India | phone no. : (0231) 2601235-36
fax : (0231) 2601238 | email : dypatilmedicalcollege@gmail.com | website : www.dypatilmedicalkop.org

Figure 3.4: Approval Letter of Institutional Ethics Committee, D Y Patil Medical College, Kolhapur.

3.5.2 Questionnaire preparation

When preparing a questionnaire to collect symptoms for research purposes, it's important to design it carefully to ensure the collection of accurate and relevant information. Following steps have been followed to develop a well-designed questionnaire to collect symptoms for the research. This enabled to gather reliable and relevant data to support the research objectives.

i. **Define the Objective**: Clearly define the objective of the research. Determine what specific symptoms or information you are interested in collecting and how they relate to the research goals.

ii. **Review Existing Literature**: Conduct a thorough review of existing literature and research studies related to the topic. This helps to identify commonly reported symptoms and any validated scales or questionnaires that have been used in similar studies.

iii. **Select Appropriate Questions**: Based on the study objective and the literature review, select the most appropriate questions to capture the symptoms of interest. Questions can be in the form of checkboxes, rating scales, or open-ended responses.

iv. **Keep it Clear and Concise**: Make sure the questionnaire is simple to comprehend and respond to. Utilize straight forward and succinct language, refraining from using jargon or technical terminologies that participant may not be acquainted with. Provide clear instructions and guidelines for answering each question.

v. **Use Validated Measures**: When researchers encounter validated measures or scales for the specific symptoms they are investigating, they should consider incorporating them into their questionnaire. The utilization of established measures ensures the reliability and validity of the collected data.

vi. **Consider Response Options**: Provide response options that accurately capture the severity, frequency, or duration of symptoms, if applicable. This can include Likert scales, visual analog scales, or numerical rating scales.

vii. **Include Demographic Questions:** Along with symptom-related questions, include demographic questions to collect basic information about participants, such as age, gender, and any relevant medical history. This information can help provide context to the symptom data.

viii. **Pilot Testing:** Perform a pilot test on the questionnaire using a small group of participants to identify any probable concerns, like unclear or ambiguous questions. Make necessary alterations to the questionnaire based on the response received during the pilot testing stage.

ix. **Obtain Ethical Approval**: If your research involves human participants, ensure that you obtain the necessary ethical approval from the appropriate review board or institution before distributing the questionnaire.

x. **Data Analysis Plan:** Plan how you will analyze the collected data. Determine the statistical methods or qualitative analysis techniques that will be used to interpret and draw conclusions from the symptom data.

The goal of the questionnaire was to collect following information from patient:

- demographic information
- Symptoms
- Tests recommended by the doctor
- Actual diagnosis

Accordingly, the questionnaire was designed. There are validated questionnaires available through different sources. These questionnaires are disease specific only. For few diseases the guidelines are provided for symptoms by WHO (World Health Organization). To generate the questionnaire following sources have been referred along with Doctor’s input.

3.5.3 Validating a Questionnaire

Once the new or translated questionnaire items have successfully undergone preliminary pilot testing and successive revisions, it is appropriate to proceed with conducting a pilot test among the intended respondents to initiate the validation process. In this pilot test, the final version of the questionnaire is administered to a substantial and representative sample of respondents who are the target audience for the questionnaire. It is important to note that if the pilot test is conducted with small sample sizes, the relatively large sampling errors may diminish the statistical power required for validating the questionnaire.

i. Pilot testing:

The pilot test was organized in D Y Patil Hospital, Kolhapur, Maharashtra from June to August 2022. They were provided with the “paper-based questionnaire”. The response was recorded and data collection progressed smoothly. The questionnaire was primarily researcher administered questionnaire. The final questionnaire will be self-administered questionnaire.

Researcher-administered questionnaires involve a researcher or interviewer who presents the questionnaire items to participants and records their responses. The researcher or interviewer reads the questions aloud or presents them visually, and participants provide responses orally or in writing. This approach might entail conducting interviews in person, over the phone, or through computer-assisted means. Researcher administered questionnaires allow for clarification of questions, ensuring participants understand the items correctly. Researchers can also provide guidance and support throughout the process, potentially increasing response rates and minimizing missing data. The presence of a researcher may influence participant responses, potentially leading to social desirability bias or altered responses due to the social interaction.

Self-administered questionnaires are questionnaires that participants complete on their own without the presence of a researcher or interviewer. Participants are provided with the questionnaire and asked to read and respond to the items independently. They may complete the questionnaire in writing (paper-based) or through electronic means (online surveys, email questionnaires). Self-administered questionnaires offer privacy and anonymity to participants, allowing them to answer questions more honestly and openly. This method is often convenient and time-efficient for participants. Since there is no immediate guidance from a researcher, the quality of responses can be influenced by factors such as participant understanding of the questions, motivation to respond accurately, and potential difficulties in interpreting certain items.

Considering all these things, the questionnaire is researcher administered for pilot testing. Patients filled the questionnaire with the assistance from several doctors and nurses. Ensuring the accurate alignment of questionnaire items with the research questions was a crucial factor. The pilot test additionally assessed the clarity and appropriateness of the questionnaire, verifying if the questions were properly defined, easily understood, and consistently presented.

The pilot study of the questionnaire revealed various issues observed among patients, including:

- The participants' capacity to understand the instructions provided in the questionnaire.
- The participants' comprehension of questionnaire items, the terminology employed, the order of questions, and the logical progression of statements.
- The presentation, together with the font and layout
- The extent or duration to complete the questionnaire.
- Additional remarks provided by patients.

Upon careful consideration of all the comments and errors, revisions were made and the questionnaire was re-piloted repeatedly until no further modifications were deemed necessary.

ii. Result analysis of pilot testing

The response from 50 patients is recorded in this pilot testing. Out of these 50 patients 28 were females and 22 were male. Based on the response from these patients, the questionnaire is upgraded. The patients in the IPD of department of Medicine are primarily considered for this research study. It typically took around 3-5 minutes to elucidate the research project and acquire consent from the participants. Every patient who received an invitation willingly agreed to take part in the study. They completed the questionnaires while waiting for their medications in the waiting room.

On average, it took the respondents approximately 10-15 minutes to finish the questionnaires. Although they made efforts to answer all the questions, there were certain items that they overlooked. Notably, there was significant inconsistency in the responses to certain questions, either due to vagueness or language barriers, leading to participants not fully comprehending the questions.

The pilot test has shown that the questionnaire is viable. The project did not seem to cause excessive disruptions in the hospital or significantly affect the staff's time. Moreover, it was well-received by patients in the waiting room. This test has highlighted the efficacy of conducting a pilot test to identify shortcomings in a questionnaire that can be addressed through appropriate revisions before employing it in a comprehensive study. Furthermore, it has provided valuable insights into the implementation of the survey. In this particular instance, doctors and nurses were occasionally required to assist patients with the questionnaire and verify response completion.

iii. Questionnaire Reliability

The reliability of a questionnaire pertains to the coherence of the survey results. Internal consistency examines the degree to which the questionnaire items are interconnected or demonstrate consistency in measuring the same construct. To assess internal consistency, the coefficient alpha, commonly referred to as Cronbach's alpha, is often utilized. Cronbach's alpha is a well-established statistical measure employed to evaluate the internal consistency reliability of a test or questionnaire. It indicates the degree to which the items within the instrument consistently measure the same underlying construct.

The Kuder-Richardson (KR) formulas are a group of statistical measures utilized for evaluating the reliability of internal consistency in tests or questionnaires. They are commonly employed in educational and psychological research to evaluate the consistency and reliability of scores obtained from various types of items.

Both the Kuder-Richardson Formula and Cronbach's alpha are statistical measures employed to evaluate the internal consistency reliability of tests or questionnaires. While they serve a similar purpose, there are differences in their computation and application.

a. **Kuder-Richardson (KR) Formulas:**
    - **Applicability**: The KR formulas, such as KR-20 and KR-21, are applicable when the test or questionnaire contains dichotomous (yes/no) items.
    - **Assumptions**: KR formulas assume that the items are of equal difficulty and that the underlying trait being measured is unidimensional.
    - **Computation**: KR formulas estimate the reliability of internal consistency by assessing the extent to which the true score variance contributes to the total score variance. Here, it takes into account the proportion of correct and incorrect responses for each item.
    - **Interpretation**: KR coefficients range between 0 to 1, where higher values indicate better internal consistency reliability.

b. **Cronbach's Alpha:**
    - **Applicability**: Cronbach's alpha is applicable when the test or questionnaire includes items with multiple response options, such as Likert scale items.

- **Assumptions**: Cronbach's alpha assumes that the items measure a single latent construct and that the inter-item correlations are approximately equal.
- **Computation**: Cronbach's alpha assesses the reliability of internal consistency by analyzing the average correlation between items within a questionnaire. It quantifies the proportion of the total variance attributable to the true score variance compared to the overall variance, accounting for both the true score variance and measurement error.
- **Interpretation**: Cronbach's alpha coefficients span from 0 to 1, where higher values indicate stronger internal consistency reliability.

Both the KR formulas and Cronbach's alpha offer valuable insights into the internal consistency of a test or questionnaire. However, the selection between the two depends on factors such as the item response format and specific needs of the study.

The questionnaire used in this research is contains dichotomous responses. So here, **Kuder–Richardson formula** is used to calculate reliability of the test.

iv. Kuder–Richardson method

The Kuder-Richardson coefficient (KR) [113]is a specific instance of Cronbach's alpha, calculated for dichotomous scores. It is commonly suggested that a high KR coefficient suggests test homogeneity. However, similar to Cronbach's alpha, homogeneity (or unidimensionality) is an assumption rather than a definitive conclusion drawn from reliability coefficients. It is important to note that a high KR coefficient can still be obtained even with a multidimensional scale, particularly when there is a large number of items.

The KR coefficient spans from 0 to 1, where higher values signify increased internal consistency reliability. The formula relies on determining the portion of the variance in the overall test scores that can be credited to the variance in the true score, rather than the variance caused by errors.

The formula to calculate the questionnaire reliability is as follows:

$$\boldsymbol{r_{KR} = \left(\frac{k}{k-1}\right) \times \left(1 - \frac{\sum p\,q}{\sigma^2}\right)}$$

where,

r_{KR} is the reliability using Kuder-Richardson equation

k : total number of test items

p : the proportion of the test takers who pass an item

q is the proportion of test takers who fail an item

σ^2 is the variation of the entire test

To calculate KR coefficient, you would need the responses of a group of participants on a multiple-choice test, where each item can be answered with either "yes" or "no." You would calculate the proportion of participants who answered each item yes and no and then apply the formula. The following table shows the reliability and its interpretation.

Table 3.1: Interpreting a correlation coefficient

Range Correlation Coefficient	**Reliability Interpretation**
r_{KR}= 1	Perfect
0.90 ≤ r_{KR}	Excellent
0.80 ≤ r_{KR}< 0.90	Good
0.7 0 ≤ r_{KR}< 0.80	Acceptable
0.60 ≤ r_{KR}< 0.70	Questionable
0.5 ≤ r_{KR}< 0.6	Poor Reliability
r_{KR}< 0.5	Unacceptable reliability
r_{KR}= 0	No Reliability

In the provided table (Table 3.1), the coefficient of correlation can be observed ranging from 0.0, indicating the weakest correlation, to 1.0, indicating the strongest correlation. The table reveals that values closer to 1 are considered the most desirable for obtaining a good correlation. For instance, a strong correlation is indicated by Kuder-Richardson reliability test results ranging from 0.7 to 0.8.

The results of the test were used to determine the reliability of the questionnaire designed to collect symptoms data from patients. In this research, there is a phase called data collection, where the patient's data was obtained through the application of a questionnaire.

v. Calculating questionnaire reliability using Kuder–Richardson method

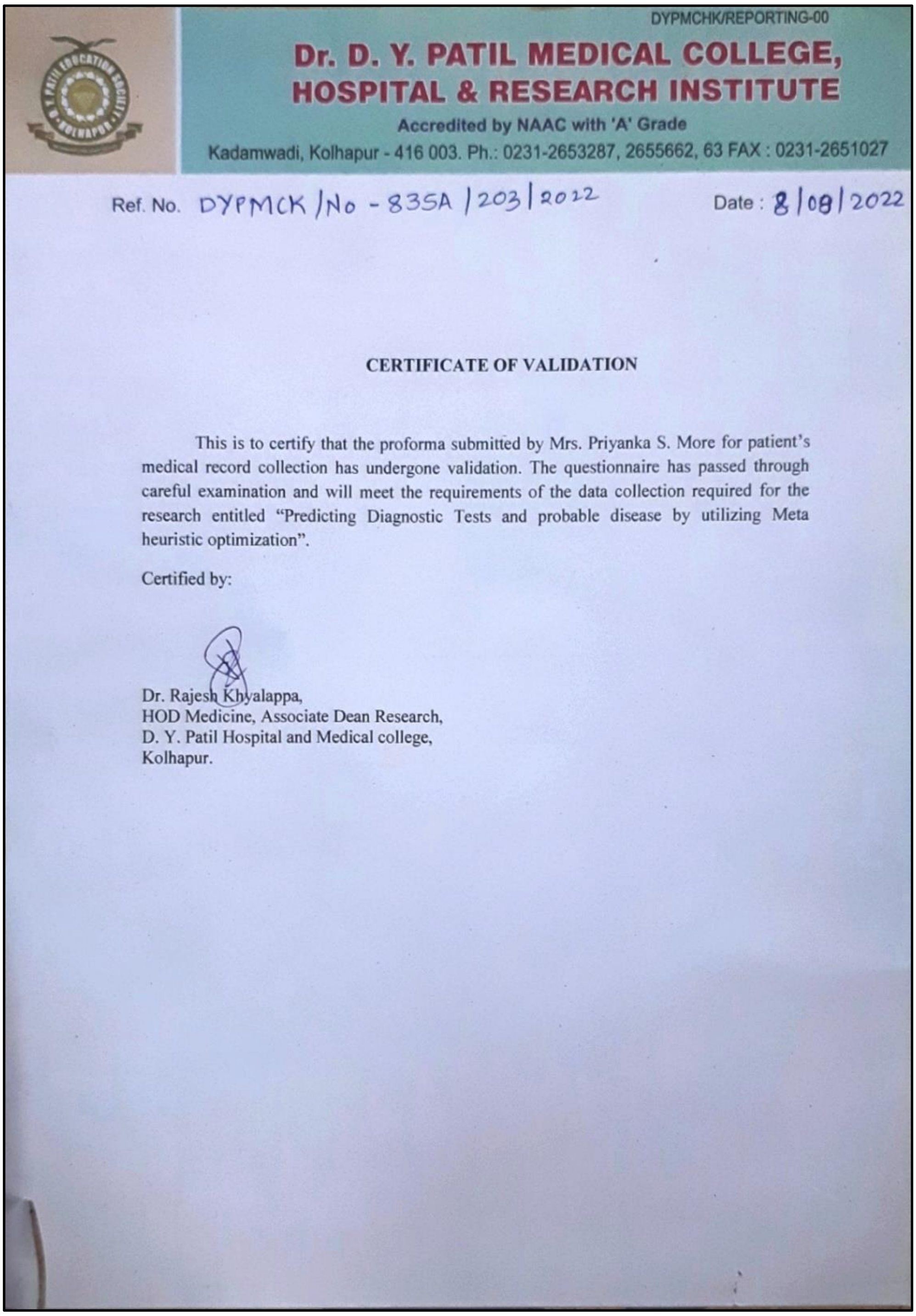
The following table 3.2 shows the calculation of the questionnaire reliability based on the dataset. This dataset contains symptoms for communicable diseases like malaria, Flu, covid-19, Chikungunya, etc.

Table 3.2: Questionnaire reliability calculation

Question No.	p	q	pq	σ^2	$\sum pq$	n	k	r_{KR}
Q1	0.886493	0.11425	0.101282	10.26	3.21688	4035	20	0.722593
Q2	0.213135	0.787608	0.167867					
Q3	0.193061	0.807683	0.155932					
Q4	0.569021	0.431722	0.245659					
Q5	0.53259	0.468154	0.249334					
Q6	0.534077	0.466667	0.249236					
Q7	0.152912	0.847831	0.129644					
Q8	0.320942	0.679802	0.218177					
Q9	0.625527	0.375217	0.234708					
Q10	0.638662	0.362082	0.231248					
Q11	0.09715	0.903594	0.087784					
Q12	0.028996	0.971747	0.028177					
Q13	0.391078	0.609665	0.238427					
Q14	0.002478	0.998265	0.002474					

Q15	0.538042	0.462701	0.248953					
Q16	0.165924	0.834076	0.138393					
Q17	0.009658	0.990342	0.009565					
Q18	0.15057	0.84943	0.127898					
Q19	0.143883	0.856117	0.123181					
Q20	0.354879	0.645121	0.22894					

From the results of the reliability testing above it can be seen that the reliability of the questionnaire is 0.722593. This is an **acceptable reliability** as shown in table 3.1. With this result the questionnaire is validated. The following figure 3.5 shows the doctor's approval on validation of questionnaire. The figure 3.6 shows the Dean's permission letter to collect patients' data from hospital.



DYPMCHK/REPORTING-00

Dr. D. Y. PATIL MEDICAL COLLEGE, HOSPITAL & RESEARCH INSTITUTE

Accredited by NAAC with 'A' Grade

Kadamwadi, Kolhapur - 416 003. Ph.: 0231-2653287, 2655662, 63 FAX : 0231-2651027

Ref. No. DYPMCK/No - 835A/203/2022 Date : 8/08/2022

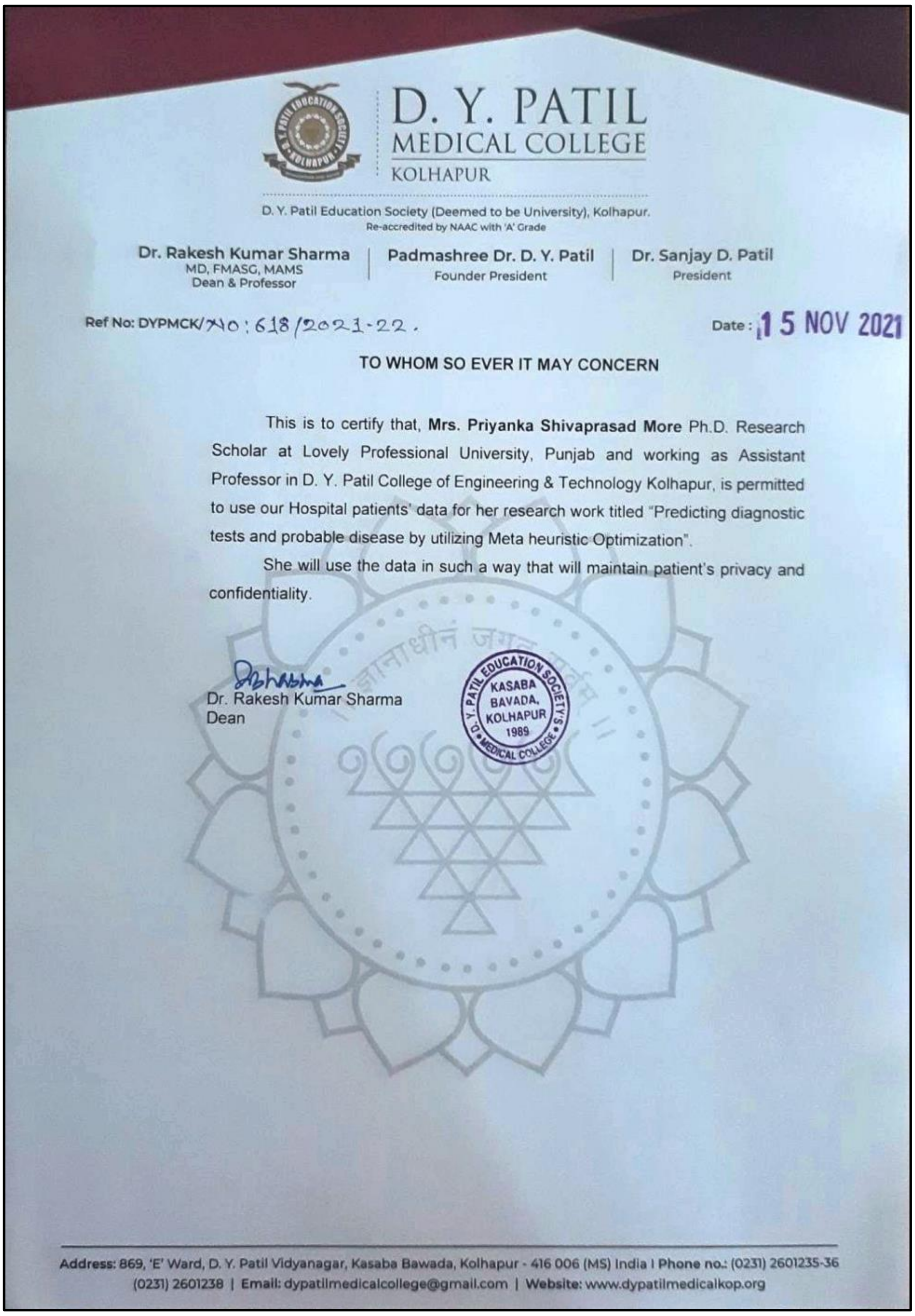
CERTIFICATE OF VALIDATION

This is to certify that the proforma submitted by Mrs. Priyanka S. More for patient's medical record collection has undergone validation. The questionnaire has passed through careful examination and will meet the requirements of the data collection required for the research entitled "Predicting Diagnostic Tests and probable disease by utilizing Meta heuristic optimization".

Certified by:

Dr. Rajesh Khyalappa,
HOD Medicine, Associate Dean Research,
D. Y. Patil Hospital and Medical college,
Kolhapur.

Figure 3.5: Questionnaire Validation Certificate by Doctor



D. Y. PATIL
MEDICAL COLLEGE
KOLHAPUR

D. Y. Patil Education Society (Deemed to be University), Kolhapur.
Re-accredited by NAAC with 'A' Grade

Dr. Rakesh Kumar Sharma, MD, FMASG, MAMS, Dean & Professor | Padmashree Dr. D. Y. Patil, Founder President | Dr. Sanjay D. Patil, President

Ref No: DYPMCK/740 : 618/2021-22. Date : 15 NOV 2021

TO WHOM SO EVER IT MAY CONCERN

This is to certify that, **Mrs. Priyanka Shivaprasad More** Ph.D. Research Scholar at Lovely Professional University, Punjab and working as Assistant Professor in D. Y. Patil College of Engineering & Technology Kolhapur, is permitted to use our Hospital patients' data for her research work titled "Predicting diagnostic tests and probable disease by utilizing Meta heuristic Optimization".

She will use the data in such a way that will maintain patient's privacy and confidentiality.

Dr. Rakesh Kumar Sharma
Dean

D. Y. PATIL EDUCATION SOCIETY'S MEDICAL COLLEGE KASABA BAVADA, KOLHAPUR 1989

Address: 869, 'E' Ward, D. Y. Patil Vidyanagar, Kasaba Bawada, Kolhapur - 416 006 (MS) India | Phone no.: (0231) 2601235-36 (0231) 2601238 | Email: dypatilmedicalcollege@gmail.com | Website: www.dypatilmedicalkop.org

Figure 3.6: Permission letter for data collection from Dr D. Y. Patil hospital, Kolhapur

3.5.4 Dataset population information:

The population under consideration for this research is from the Kolhapur district which is in Maharashtra state of India. The deidentified patient's data is extracted from electronic health record of hospitalization at the D. Y. Patil hospital, Kolhapur via clinical data warehouse. Also, some records are generated via a questionnaire through a personal interview at the OPD of the same hospital. The structured data covers patients' symptoms, demographic information, tests recommended by doctors and final diagnosis. Table 3.3 shows the population information for the 4035 sampled patients.

Table 3.3 Dataset population information

Total number of patients	4035
Age – Mean (Years)	40.6
Age – Standard Deviation (Years)	21.24
Percentage Female / Male	65:45

The data features extracted include

1) age: in years
2) sex: 1 = male; 0 = female
3) fever: 1 = Yes, 0 = No
4) cold: 1 = Yes, 0 = No
5) rigor: 1 = Yes, 0 = No
6) body pain: 1 = Yes, 0 = No
7) fatigue: 1 = Yes, 0 = No
8) headache: 1 = Yes, 0 = No
9) skin rash: 1 = Yes, 0 = No
10) runny nose: 1 = Yes, 0 = No
11) sore throat: 1 = Yes, 0 = No
12) cough: 1 = Yes, 0 = No
13) diarrhoea: 1 = Yes, 0 = No
14) chest pain: 1 = Yes, 0 = No
15) vomiting: 1 = Yes, 0 = No
16) abdominal pain: 1 = Yes, 0 = No
17) joint swelling: 1 = Yes, 0 = No
18) conjunctivitis: 1 = Yes, 0 = No
19) COVID-19 vaccination: 1 = Yes, 0 = No

20) diagnosis tests recommended: diagnostic tests

21) final diagnosis

3.5.5 Sample Size

There are so many methods to decide the minimum sample size for the research. After exploring different methods and after consulting the statistics department of the Dr D. Y. Patil Medical College and Research Insititure, Kolhapur, the sample size is decided as follows.

Here, **Slovin's formula** is used to decide the sample size.

$$\textbf{Sample size} = n = \frac{N}{1+N\,e^2}$$

Where,

N = Population size

e = Margin of error

For this research the population (N) under consideration is population of Kolhapur district. So, N = 4147506.

By considering margin of error = e = 2 % = 0.02 for N = 4147506

$$n = \frac{4147506}{1 + 4147506 \times\ (0.02)^2}$$

$$n = \frac{4147506}{1660.0024}$$

n = 2499

Minimum Sample size required is **2499**.

3.6 Proposed Metaheuristic Optimization Based Diagnostic Tests and Probable Disease Prediction Model

This section explains the adopted approach followed in this research to achieve the proposed objectives. Figure 3.7 shows the proposed model for the novel approach to predict diagnostic tests and probable disease using symptoms. The focus is on identifying the most relevant features for recommending tests based on symptoms. The problem of diagnostic test prediction falls under classification rather than clustering.

In clustering, the goal is to group similar data points together based on their intrinsic properties or similarity measures, without any predefined labels or classes. Clustering algorithms aim to discover hidden patterns or structures in the data and assign data points to different clusters.

On the other hand, in diagnostic test prediction, the objective is to classify or predict the presence or absence of a specific disease and accordingly predict the diagnostic tests to be done. The problem involves assigning data points (patients) to predefined classes (positive or negative for a particular disease). It is a supervised learning task where labelled training data is available to train a classification model. The model is then used to predict the class labels of unseen data points (patients) based on their test results.

3.6.1 Data collection through questionnaire for model designing

The data is collected through the verified and validated questionnaire from patients. This data contains record of patients' symptoms, demographic information, Diagnostic tests recommended by doctors and final diagnostic. This dataset is used for the model development and training purpose.

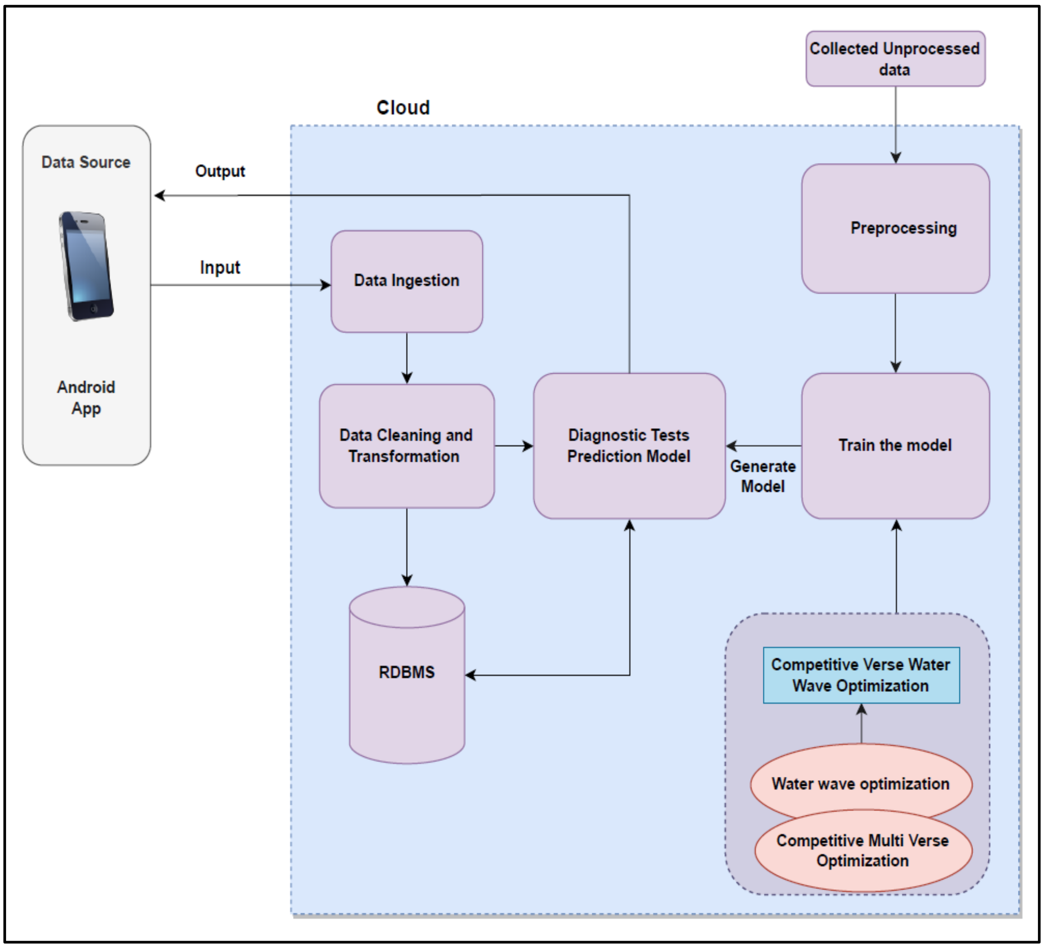


Figure 3.7: Methodology for metaheuristic optimization based diagnostic tests prediction model.

3.6.2 Preprocessing

Data pre-processing in the diagnostic test prediction using symptoms involves preparing and cleaning the data to make it suitable for analysis and modelling. Data Cleaning includes variety of tasks which includes identifying and handling missing values, outliers, and errors in the dataset. This can involve techniques such as imputation (replacing missing values with estimated values), removing outliers that are significantly different from other data points, and correcting errors in the data.

By applying these techniques, model can identify the symptoms that have the strongest associations or influence on the recommended tests. These features can then be used to optimize the test recommendation process based on the symptom database.

Data cleaning is performed to address inconsistencies and errors in the questionnaire responses. The responses in the questionnaire are mostly in dichotomous form. Here, the symptom columns that contain categorical data, such as "Yes" or "No" responses, are converted into numerical format. For example, "Yes" is converted to 1 and "No" is converted to 0. This conversion allows to apply feature selection techniques that require numerical input.

3.6.3 Training procedure for the proposed prediction model

The proposed prediction model is trained using a novel CVWWO algorithm, created by merging CMVO [111] and WWO [112]. CMVO is a population-based algorithm designed for addressing global optimization problems. In this strategy, when the universe is outperformed, it learns from the winning solution in each generation. This approach is particularly effective due to its rapid convergence and simplified structure.

On the other hand, WWO is a metaheuristic approach inspired by shallow water wave models, employed for solving optimization challenges. WWO offers the advantages of simplicity and operates with a small population size. By combining WWO with CMVO, the CVWWO algorithm achieves global optimal solutions while mitigating computational complexity concerns. The algorithmic phases of the CVWWO algorithm are detailed below.

Step 1: Initialization

The initial step involves creating random universes according to the population size and the dimension of the search space, as indicated by the following expression

$$Z = \{Z_1, Z_2, \ldots, Z_a, \ldots, Z_d;\ 1 \le a \le d\} \tag{3.1}$$

In this context, d represents the entire set of solutions and Z_a denotes the a^{th} indivisual solution.

Step 2: Fitness computation

The fitness function is employed to assess the most favourable solution, considering the fitness function with the lowest error value as the optimal solution. The equation for the fitness function is represented as follows:

$$\gamma = \frac{1}{\ell}\sum_{\vartheta=1}^{\ell}\left[O_{\tau} - D_{o}\right]^{2} \tag{3.2}$$

In this context, γ represents the fitness metric, D_o represents the classifier's output, and O_τ corresponds to the desired target output.

Step 3: Bi-competition

Here, the universes within the population are comprised of randomly two sets. The two different sets of population at all the iteration are allowed to connect in bi-competition. Hence, the population set exhibiting the highest fitness is termed the "winner," while the others are deemed "losers.".

Step 4: Update a solution

The victor promptly progresses to the next iteration, while the defeated party undergoes an update. The universe's position is adjusted using the updated equation of the proposed CVWWO algorithm. The conventional adjustment equation of the CMVO[111] algorithm is as follows:

$$Z_{u,v}^{i} = g_1 * TDR + g_2\left(Z_{w,v} - Z_{u,v}\right) + g_3 * \left(Z_v - Z_{u,v}\right) \tag{3.3}$$

$$Z_{u,v}^{i} = g_1 * TDR + g_2 * Z_{w,v} - g_2 * Z_{u,v} + g_3 * Z_v - g_3 * Z_{u,v} \tag{3.4}$$

$$Z_{u,v}^{i} = -Z_{u,v}\left(g_2 + g_3\right) + g_1 * TDR + g_2 * Z_{w,v} + g_3 * Z_v \tag{3.5}$$

Through the integration of CMVO with WWO, it becomes possible to enhance global convergence. Consequently, the standard equation of WWO [112] is provided as follows:

$$Z_{u,v}^{i} = Z_{u,v} + rand\left(-1,1\right).\lambda Q_m \tag{3.6}$$

$$Z_{u,v} = Z_{u,v}^{i} - rand(-1,1).\lambda Q_m \tag{3.7}$$

By substituting equation (3.7) in equation (3.5), the following equation becomes,

$$Z_{u,v}^{i} = -(Z_{u,v}^{i} - rand(-1,1).\lambda Q_m)(g_2 + g_3) + g_1 * TDR + g_2 * Z_{w,v} + g_3 * Z_v \tag{3.8}$$

$$Z_{u,v}^{i} = -Z_{u,v}^{i}(g_2 + g_3) + rand(-1,1).\lambda Q_m(g_2 + g_3) + g_1 * TDR + g_2 * Z_{w,v} + g_3 * Z_v \tag{3.9}$$

$$Z_{u,v}^{i} + Z_{u,v}^{i}(g_2 + g_3) = rand(-1,1).\lambda Q_m(g_2 + g_3) + g_1 * TDR + g_2 * Z_{w,v} + g_3 * Z_v \tag{3.10}$$

$$Z_{u,v}^{i}(1 + g_2 + g_3) = rand(-1,1).\lambda Q_m(g_2 + g_3) + g_1 * TDR + g_2 * Z_{w,v} + g_3 * Z_v \tag{3.11}$$

$$Z_{u,v}^{i} = \frac{1}{1 + g_2 + g_3}\left[rand(-1,1).\lambda Q_m(g_2 + g_3) + g_1 * TDR + g_2 * Z_{w,v} + g_3 * Z_v\right] \tag{3.12}$$

$$TDR = 1 - \frac{a^{1/t}}{Q^{1/t}} \tag{3.13}$$

In this context,

g_1, g_2 and g_3 → random variable which ranges from $[0,1]$,

Q_m → the length of the m^{th} dimension of search space,

$Z_{w,v}$ → the victorious universe in the v^{th} round of competition,

$Z_{u,v}$ → universe that did not succeed in v^{th} round,

Z_v → the mean position value of the relevant universe,

Q → maximum iteration,

$m \rightarrow$ the ongoing iteration,

$t \rightarrow$ the constant value of 6.

The TDR (Travelling Distance Rate) is variation coefficient. It is a determining factor for measuring the rate of change in distance. The TDR value is gradually reduced throughout each iteration to provide more accurate exploitation and local search in the vicinity of the best universe produced. The ordinary world revolves around the present optimal universe with TDR. Therefore, TDR is responsible for determining both the velocity and extent of the investigation and use of a universe around the optimal universe. A higher initial TDR value facilitates the algorithm's exploration of the global space, while the subsequent decrease in TDR value enables the program to utilize the vicinity of the present optimal solution.

Step 5: Re-computation of fitness

In this case, the fitness is recalculated for every solution in order to uncover the optimal one.

Step 6: Termination

All the above illustrated processes are repeated until optimal solution is achieved. Table 3.1 portrays the pseudo code of proposed CVWWO algorithm.

Table 3.4: Algorithm for innovative CVWWO algorithm

Sl. No	Algorithm for innovative CVWWO algorithm
1	**Start**
2	Set the population size and define the search space dimension
3	For each iteration check if m <=maximum iteration, if yes
4	Calculate variable TDR
5	Calculate the fitness value for individual universe
6	For each universe **do**

7	Update the objects among the universes
8	Transfer objects within each universe to defeated universes.
9	Update the solution using equation (3.12)
10	end for
11	Re-compute the fitness
12	end for
13	Achieve optimal Solution
14	**Stop**

The proposed prediction model effectively predicts the primary diagnostic tests to be done.

3.7 Advantages of Hybrid Metaheuristic Optimization Algorithms in Symptom-Based Diagnostic Test Prediction

In the context of designing an AI-based system that predicts diagnostic tests based on a patient's symptoms through a questionnaire, the use of **hybrid metaheuristic optimization algorithms** offers significant advantages. Such a system, aimed at reducing hospital footfall by recommending diagnostic tests before a patient’s first visit to a doctor, presents unique challenges in terms of symptom interpretation, test selection, and disease prediction. The inherent complexity of these tasks requires a sophisticated approach, and hybrid metaheuristic algorithms are particularly well-suited to this problem domain due to their capacity to optimize multiple aspects of the decision-making process simultaneously. This section outlines the primary advantages of hybrid metaheuristics in the context of this problem.

i. Optimizing Test Selection and Improving Accuracy

The primary function of the proposed system is to predict the most relevant diagnostic tests based on patient symptoms. This requires accurate mapping of symptoms to diagnostic procedures. Hybrid metaheuristic algorithms, by combining multiple search strategies such as genetic algorithms (GA), particle swarm optimization (PSO), and ant colony optimization (ACO), ensure that the selection process is optimized. The combination of these algorithms allows for a balance between **exploration of the solution space** (discovering a wide range of test possibilities) and **exploitation** (refining and optimizing test recommendations). This results in **improved accuracy** in test selection, ensuring that patients are recommended the most appropriate and comprehensive set of tests, thereby enhancing clinical outcomes and aiding in early disease detection.

ii. Handling Complex, Nonlinear Relationships Between Symptoms and Diagnostic Tests

The relationship between symptoms, diseases, and the corresponding diagnostic tests is often nonlinear and highly complex. A single symptom can be indicative of multiple conditions, while a combination of symptoms may point to a particular disease or set of diseases. Hybrid metaheuristic optimization algorithms are well-suited to handle such complexity. By incorporating multiple optimization techniques, these algorithms are able to efficiently navigate large and complex solution spaces, **identifying patterns and relationships** that may not be immediately obvious. This enables the system to generate more accurate and relevant test recommendations, even in cases where symptoms do not have a straightforward correlation with specific diagnostic tests.

iii. Robustness Against Noisy and Incomplete Symptom Data

Medical datasets, particularly those derived from patient questionnaires, are often subject to **noisy, incomplete, or imprecise data**. Patients may misreport symptoms, or their symptoms may vary in intensity, leading to ambiguity in the input data. Hybrid metaheuristic algorithms are inherently robust against such noise and missing data. By evaluating multiple candidate solutions in parallel and using adaptive mechanisms to refine these solutions over iterations, hybrid algorithms can tolerate variability in input while still producing reliable outputs. This ensures that the system can still provide accurate and relevant test recommendations even when faced with uncertain or incomplete symptom information.

iv. Efficient Prediction of Multiple Diagnostic Tests

The system is tasked with predicting a set of diagnostic tests that are likely to yield the most useful information for the clinician before the patient's first visit. Given the need for **timely and accurate** test recommendations, hybrid metaheuristics offer a significant advantage due to their ability to **converge faster** toward optimal or near-optimal solutions. The combination of different metaheuristic techniques ensures that the system does not waste time evaluating unpromising solutions, thereby accelerating the process of generating recommendations. This increased convergence speed ensures that the system provides **timely, reliable predictions**, which is critical for reducing delays in diagnosis and treatment.

v. Feature Selection for Symptom-Test Mapping

An essential challenge in this system is determining which **symptoms are most relevant** to the diagnostic test selection process. Hybrid metaheuristics excel at optimizing **feature selection**, allowing the system to prioritize the most impactful symptoms when making recommendations. By optimizing this process, the system can avoid recommending unnecessary tests, while ensuring that all relevant tests are included in the recommendation. This enhances the **efficiency and accuracy** of the system, as it ensures that only the most relevant diagnostic tests are suggested based on the patient's symptoms.

vi. Avoiding Local Optima in Diagnostic Test Recommendations

One of the common challenges with single optimization techniques is the tendency to become trapped in **local optima**, where suboptimal solutions are incorrectly identified as the best solution. In the context of test recommendation, this could lead to the system consistently recommending a limited set of generic tests, without considering more specialized tests that may be necessary based on the patient's unique symptoms. Hybrid metaheuristic algorithms overcome this issue by **combining exploration and exploitation** strategies from different algorithms, enabling the system to escape local optima and search for more **global solutions**. This ensures that the system can identify a more diverse and relevant set of diagnostic tests, particularly in cases where rare or complex diseases are involved.

vii. Adaptability to Dynamic Symptom and Test Data

The medical field is constantly evolving, with new diagnostic tests being introduced and new symptoms or diseases being identified. Hybrid metaheuristic algorithms are inherently adaptable, allowing the system to incorporate **new data and adjust** its recommendations as medical knowledge evolves. This adaptability ensures that the system remains relevant over time, as it can seamlessly integrate new tests or changes in symptom-disease relationships into its recommendation process. The ability to **dynamically update** its internal models based on new information enhances the long-term utility of the system.

viii. Multi-Disease Prediction Flexibility

In cases where patients present with symptoms that could be indicative of multiple diseases, the system needs to account for **multi-objective optimization**. Hybrid metaheuristics, with their ability to simultaneously optimize multiple objectives, allow the system to consider the possibility of multiple co-occurring or related diseases when making diagnostic test recommendations. This ensures that the patient is tested for all relevant conditions, reducing the risk of misdiagnosis and improving overall patient outcomes.

ix. Balancing Test Coverage and Patient Convenience

A key design goal of the system is to reduce hospital footfall by ensuring that patients undergo necessary tests before visiting the doctor. However, it is also important to avoid overburdening patients with too many tests. Hybrid metaheuristic optimization algorithms can help the system balance **test effectiveness with patient convenience** by minimizing the number of unnecessary tests while still ensuring adequate diagnostic coverage. This balance reduces the number of diagnostic tests that need to be performed without sacrificing accuracy, thus improving patient satisfaction and reducing overall healthcare costs.

x. Minimizing Redundant Testing and Improving Efficiency

Finally, hybrid metaheuristic algorithms can optimize the test recommendation process by **minimizing redundant testing**. This reduces both the cost burden on patients and the strain on healthcare facilities. By optimizing test selection, the system can suggest the minimum number of tests required to accurately diagnose a potential condition, improving the overall **efficiency** of the healthcare system. This leads to a more streamlined process, in which patients are only recommended tests that are highly likely to contribute to a more accurate diagnosis.

In conclusion, the use of hybrid metaheuristic optimization algorithms in a symptom-based diagnostic test prediction system offers numerous advantages. These include improved accuracy in test recommendations, the ability to handle complex and nonlinear symptom-disease-test relationships, robustness to noisy and incomplete data, and greater efficiency in generating test recommendations. Moreover, hybrid metaheuristics provide adaptability to dynamic data, multi-objective optimization for disease prediction, and improved balance between test coverage and patient convenience. Together, these advantages make hybrid metaheuristic algorithms a powerful tool in the development of AI-based systems for healthcare diagnostics.

Chapter No. 4

Validate the outcomes of the proposed methodology

4.1 Introduction

This chapter delves into the validation of a methodology designed to predict diagnostic tests and probable diseases based on patients' symptoms, focusing on communicable diseases including dengue, malaria, COVID-19, chikungunya, and viral fever. The methodology aims to streamline the diagnostic process by suggesting preliminary tests before the patient's initial visit to a healthcare professional, thereby reducing the number of hospital visits and saving time for both patients and doctors. An Android application has been developed to implement the methodology in a real-time environment, where patients' symptoms are collected through a questionnaire. The collected symptoms serve as inputs to a cloud-hosted model, which predicts the preliminary tests to be conducted. However, the validation process emphasizes the necessity of medical professionals' intervention in validating and approving the suggested tests, ensuring the ethical use of AI in healthcare.

4.2 Methodology implementation details:

The proposed methodology involves the development of an Android application designed to collect patients' symptoms through a structured questionnaire. The questionnaire serves as a screening tool to identify potential communicable diseases based on reported symptoms. The collected data is then processed using machine learning algorithms to predict the most probable diseases and recommend preliminary diagnostic tests.

The methodology involves a multi-step process:

i. **Symptom Collection through Questionnaire:**

Patients input their symptoms using a structured questionnaire via an Android application. The Android application provides a user-friendly interface for patients to input their symptoms through a structured questionnaire. The questionnaire covers a wide range of symptoms associated with communicable diseases under consideration, ensuring comprehensive data collection.

ii. **Cloud-hosted AI Model:**

The symptom data is transmitted to a cloud-hosted proposed AI model, trained to predict preliminary diagnostic tests for communicable diseases. This model utilizes advanced algorithms to analyze the symptom data and predict the preliminary tests to be conducted based on the reported symptoms.

iii. **Doctor's Intervention:**

Preliminary test recommendations are sent to a medical professional for validation before reaching the patient. The doctor has the authority to approve or disapprove the suggested tests based on their expertise and judgment.

iv. **Feedback Mechanism:**

If the suggested tests are approved by the doctor, they are then provided to the patient for further action. If not approved, the doctor selects alternative diagnostic tests for the patient, ensuring patient safety and accurate diagnosis. This feedback loop helps refine the model.

v. **Model Learning Process:**

Not approved recommendations are fed back into the AI model for continuous learning, enhancing accuracy and precision. This feedback loop aims to improve the accuracy and precision of the model's predictions over time, enhancing its performance in predicting diagnostic tests and probable diseases.

4.3 Results and Analysis

Initially, during training and testing, the model exhibited excellent performance with an accuracy of 94.86%, precision of 93.72%, recall of 96%, and F1 score of 94.14%. However, when deployed in real-time, the model's performance deteriorated significantly, with accuracy dropping to 74.20%, precision to 68.50%, recall to 79.30%, and F1 score to 73.60%.

The disparity between the model's performance during training-testing and real-time deployment highlights the challenges of adapting AI models to dynamic and real-world healthcare environments. The poor performance metrics indicate the need for refinement and optimization to enhance the model's predictive capabilities in real-time scenarios. The following table 4.1 shows the performance evaluation of proposed model before improvements.

Table 4.1: Performance evaluation of proposed model before improvements

Metric	Before Improvement
Accuracy	74.20%
Precision	68.50%
Recall	79.30%
F1 Score	73.60%

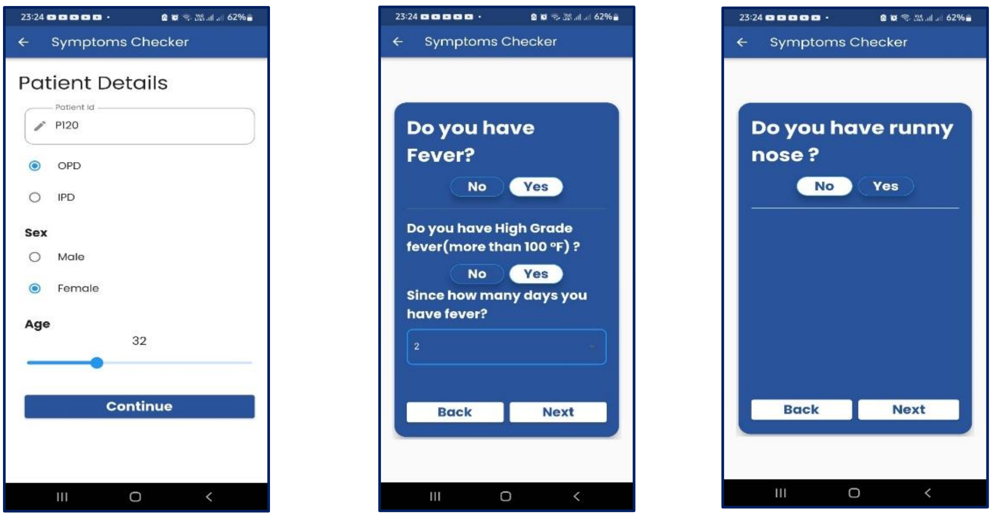
Following the refinement of the model with "Not Approved" cases, there is a notable improvement in performance metrics. The accuracy increased to 86.40%, precision to 81.20%, recall to 89.50%, and F1 score to 85.10%. The incorporation of "Not Approved" cases enabled the model to adapt and learn from real-world scenarios, resulting in enhanced predictive accuracy and reliability. The improved performance metrics underscore the significance of iterative refinement and continuous learning in optimizing AI-driven healthcare systems for real-time deployment. The following table 4.2 shows the performance evaluation of proposed model before improvements

Table 4.2: Performance evaluation of proposed model after improvements

Metric	After Improvement
Accuracy	86.40%
Precision	81.20%
Recall	89.50%
F1 Score	85.10%

The model's predictions are compared with the doctor's validation to determine its reliability and effectiveness. Discrepancies between the model's predictions and the doctor's validation are analyzed to identify areas for improvement and refinement. Feedback from both doctors and patients is integrated into the analysis to gain insights into the usability and effectiveness of the proposed methodology. This feedback helps in identifying strengths, weaknesses, and opportunities for enhancement.

Figure 4.1 shows screenshots of the questionnaire interface within the Android app. It shows the user-friendly design and the ease of inputting symptoms.



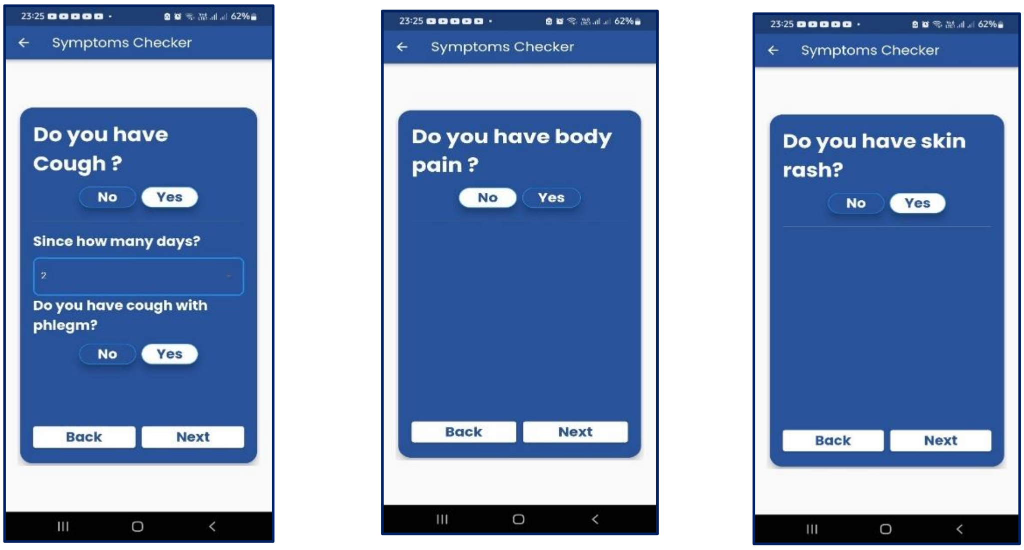


Figure 4.1: Screenshot from Android app showing patient's data collection through questionnaire

Figure 4.2 capture screenshots of the interface where patients receive the preliminary test suggestions based on their symptoms. It Show how the model predicts the probable diseases and suggests appropriate tests.

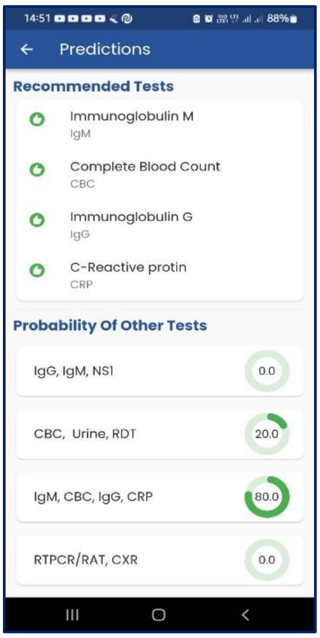


Figure 4.2: Screenshot from Android app showing model prediction interface

The figure 4.3 illustrates the interface where doctors validate the suggested tests. This showcases how doctors review the preliminary test recommendations and either approve or suggest alternative tests. This highlights the feature that facilitate the doctor's decision-making process.

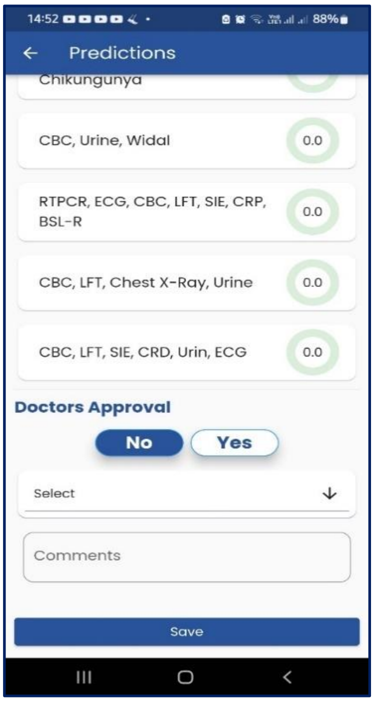


Figure 4.3: Screenshot from Android app showing doctor's validation process

The figure 4.4 illustrates the interface where the doctors are provided with the facility to select other diagnostic tests, if they don't approve of the diagnostic tests generated by model. This will work as a feedback mechanism to gather doctors' insights and refine the model based on real-world validation.

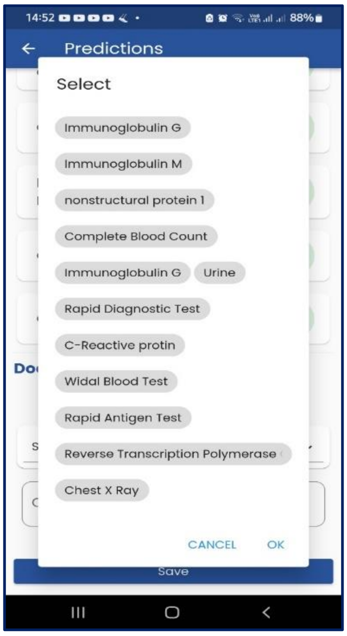


Figure 4.4: Screenshot from Android app showing doctor's intervention process

4.4 Ethical Considerations

- **Doctor's Intervention**: The validation process emphasizes the importance of medical professionals' intervention in validating and approving the suggested tests. This ensures patient safety and mitigates the risks associated with relying solely on AI-driven predictions for medical diagnosis.
- **Patient Privacy and Confidentiality**: Measures are implemented to ensure the privacy and confidentiality of patients' health information collected through the Android application. Data encryption and secure transmission protocols are employed to safeguard sensitive patient data.
- **Informed Consent**: Patients are provided with clear information about the purpose and use of their symptom data collected through the Android application. Informed consent is obtained from patients prior to collecting their data, ensuring transparency and respect for patient autonomy.
- **Continuous Improvement**: The ethical use of AI in healthcare necessitates a commitment to continuous improvement and refinement of the methodology. Regular monitoring, feedback integration, and updates to the machine learning

model are essential to ensure the accuracy, fairness, and transparency of the diagnostic process.

By addressing these implementation details, conducting thorough results and analysis, and adhering to ethical considerations, the proposed methodology aims to enhance diagnostic efficiency while upholding principles of patient safety, privacy, and informed consent. The validation process provides insights into the effectiveness of the proposed methodology for predicting diagnostic tests and probable diseases based on patients' symptoms. The integration of an Android application for symptom collection streamlines the diagnostic process and enhances the efficiency of healthcare delivery. The validation results serve to validate the feasibility and practicality of the proposed system in real-world healthcare settings.

Chapter 5

Compare and evaluate the obtained results using standard performance assessment metrics

5.1 Experimental Setup

The Microsoft excel is used for the storage of the dataset. The PyCharm 22.1.3 is used for the analysis and implementation of algorithms on computer with 4GB of RAM, Microsoft Windows 10 operating system and an Intel Core i3 processor.

5.2 Evaluation Metrics

There are different performance parameters used in this research. On the basis of these parameters the research has obtained the most accurate classifier and optimization technique. List of parameters is as follows:

i. **Accuracy:**

Accuracy evaluates the ratio of correctly predicted instances, encompassing both positive and negative cases, to the total number of instances. This metric offers a comprehensive evaluation of the model's predictive performance. The highest accuracy is 1.0, while the worst is 0.0. The formula for accuracy is:

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

ii. **Precision (Positive Predictive Value):**

Precision quantifies the ratio of correctly predicted positive instances relative to all instances predicted as positive. It emphasizes the accuracy of positive predictions. The optimal precision, or Positive Predictive Value (PPV), attains a value of 1.0, while the poorest precision is denoted by 0.0. The precision is calculated as:

$$Precision = \frac{TP}{TP + FP}$$

iii. Recall (True positive rate/Sensitivity):

Recall, also referred to as sensitivity or true positive rate, evaluates the proportion of correctly predicted positive instances among all actual positive instances. It highlights the model's capability to accurately identify positive cases. The highest possible sensitivity score is 1.0, while the lowest is 0.0. The recall is calculated as:

$$Recall = \frac{TP}{TP + FN}$$

iv. F1-score:

The F1 score is a statistic that quantifies the balance between accuracy and recall by calculating their harmonic mean. It functions as a unified metric that achieves a harmonious equilibrium between accuracy and recall, thereby merging both features into a single value. The F1 score is calculated as:

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

The F1 score proves particularly beneficial in scenarios involving imbalanced datasets or when both false positives and false negatives hold equal importance.

5.3 Machine learning techniques used for comparative evaluation

There are several machine learning techniques available for pattern recognition. The techniques that have been used for comparative analysis are RF, KNN, Adaptive-Boosting (AdaBoost) and GBM.

The AdaBoost was implemented using the AdaBoostClassifier7, and Two hyperparameters that were examined are ‘learningrate’ and ‘n-estimators’. The n-estimatos parameter shows the highest number of steps that the model generate in training phase, and ‘learningrate’ showa how much weight is assigned to each stump per iteration. A higher learningrate magnifies the influence of every classifier, resulting in increased contributions from stumps during training. Lower learningrate values may lead to decreased accuracy in classification, while higher values can introduce instability in the model [114].

The KNN algorithm implemented using the KNeighborsClassifier11, and during the execution, three hyperparameters were subjected to testing: metric, weights and n-neighbors. Specifically, n-neighbors represents how many neighbors were employed during the training. The weights parameter is responsible for determining the function that assigns weights to each neighbor during training, while metric governs the function used to compute the distance to each neighbor.

The RF algorithm was performed using the RandomForestClassifier8, and during the execution criterion and n-estimators these hyperparameters were examined. Similar to Adaboost, n_estimators denote the maximum number of Decision Trees (DTs) generated by the model. On the other hand, criterion represents the function responsible for choosing the optimal splits for each Decision Tree node.

Using the GradientBoostingClassifier9, GB M algorithm was implemented and the hyperparameters explored were max-depth and n-estimators. Each Decision Tree (DT) in the model has a depth that is represented by its max-depth. DTs with more nodes are correlated with a higher max-depth. Like RF and Adaboost, n-estimators indicates the highest number of DTs that the model can produce.

5.4 Performance Analysis of models

This section assesses the performance of the proposed methodology with the help of evaluation parameters: accuracy, precision, recall and F1- score using training data and

K-fold method. The different techniques used for comparing and evaluating are RF, KNN, GBM and adaptive-boosting(AdaBoost) with the proposed metaheuristic optimization based diagnostic tests and probable disease prediction model.

5.4.1 Assessment utilizing training data:

In Figure 5.1, the performance of the proposed prediction model is compared using evaluation metrics using training data.

i. Accuracy:

Table 5.1(a): Accuracy evaluation for training data

Training percentage	RF	KNN	AdaBoost	GBM	Proposed novel approach
50	77.81601	80.35625	80.94897	86.26807	91.06336
60	79.1708	81.62637	83.48921	88.55428	91.773
70	80.19334	82.89649	86.45553	89.1255	93.258
80	80.44092	83.50953	88.44353	89.26199	94.743
90	81.71104	84.60014	89.29483	89.85641	96.000

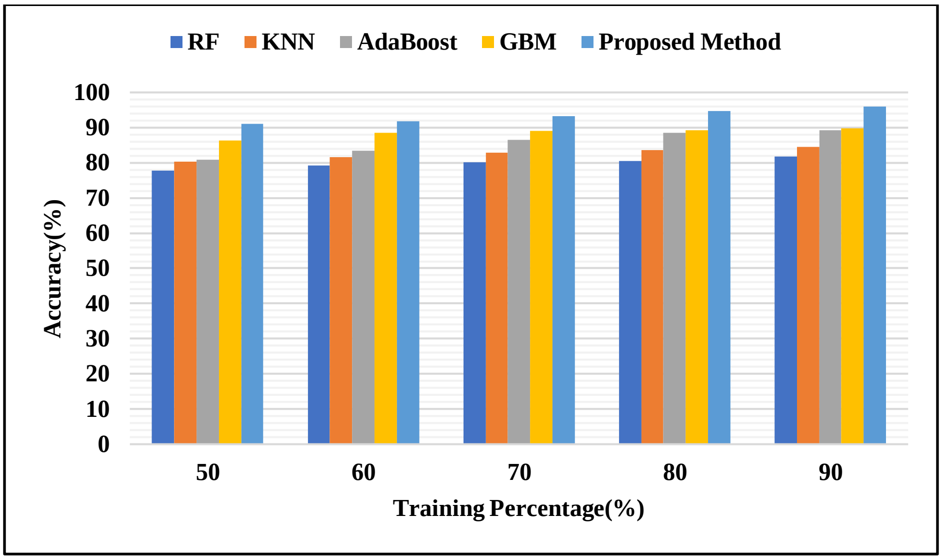


Figure 5.1(a): Evaluating the effectiveness of the developed method using training data percentage: Accuracy

ii. Precision:

Table 5.1(b): Precision evaluation for training data

Training percentage	RF	KNN	AdaBoost	GBM	Proposed novel approach
50	71.82024	79.90178	80.37677	82.67422	88.59286
60	73.57195	80.0619	81.0814	83.19235	89.10000
70	73.70849	81.07248	82.06211	83.52811	89.38929
80	78.88452	81.76911	82.6436	83.6901	90.11429
90	81.96186	82.03755	83.25546	86.0771	91.14286

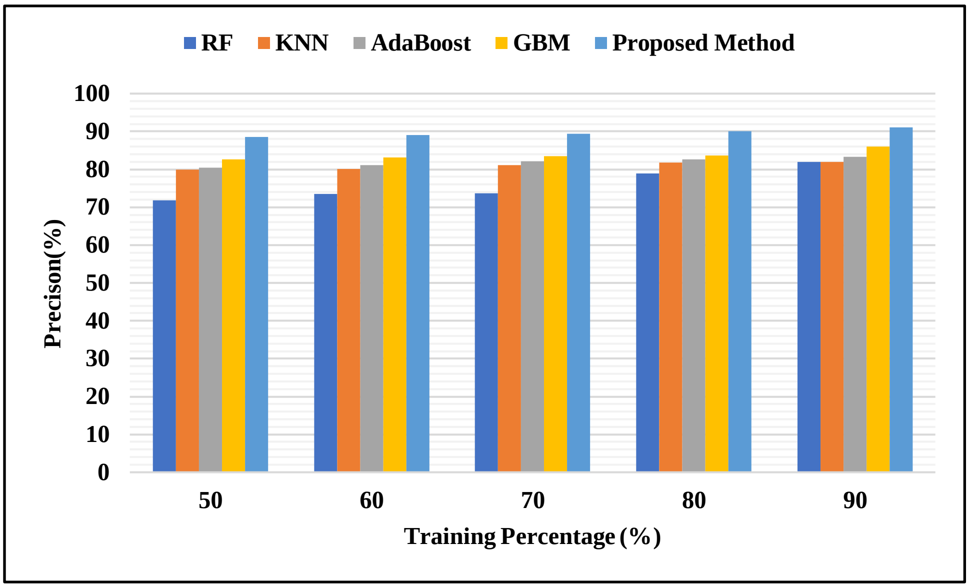


Figure 5.1(b): Evaluating the effectiveness of the developed method using training data percentage: Precision

iii.Recall:

Table 5.1(c): Recall evaluation for training data

Training percentage	RF	KNN	AdaBoost	GBM	Proposed novel approach
50	79.67211	84.64209	86.87808	87.17331	90.34179
60	82.80839	84.99968	87.2387	88.02409	90.51674
70	82.91136	86.2575	87.56036	88.65383	90.85935
80	83.10849	86.71776	88.23545	89.26129	91.1622
90	85.5086	87.91211	89.5716	90.46068	91.8814

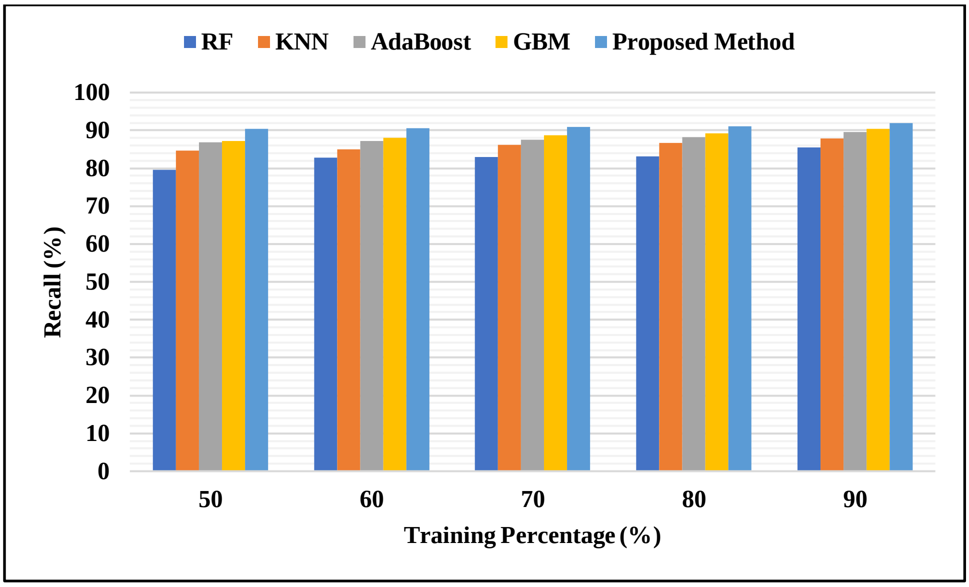


Figure 5.1(c): Evaluating the effectiveness of the developed method using training data percentage: Recall

iv. F1-score:

Table 5.1(d): F1-score evaluation for training data

Training percentage	RF	KNN	AdaBoost	GBM	Proposed novel approach
50	84.11049	86.21251	88.14827	89.46778	91.65298
60	85.43995	87.5956	89.29483	90.66559	92.367

70	86.76941	88.76628	89.5519	90.87157	93.852
80	86.89351	88.88026	90.7635	91.88859	94.9638
90	87.7461	88.97869	90.95554	92.27535	96.000

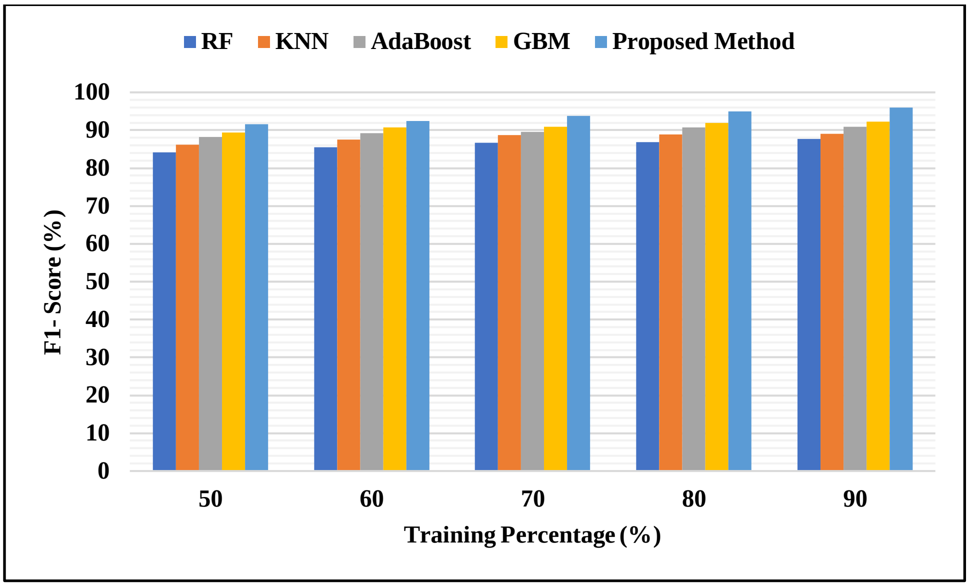


Figure 5.1(d): Evaluating the effectiveness of the developed method using training data percentage: F1- Score

Here, the performance comparison of proposed novel approach is performed using evaluation metrics. Figure 5.1(a), 5.1(b), 5.1(c), 5.1(d) shows results for accuracy, precision, recall, and F1-score are presented in sequence. The novel approach utilizing GBM showcased superior performance, outperforming all other models. Notably, the RF model exhibited subpar performance in comparison. This discrepancy raises the possibility that the RF model underfitted because it had difficulty generalising the data during training. As a consequence, its performance suffered when evaluated on the test set. Figure 5.1(a) shows the evaluation using accuracy metric. In comparison to RF, KNN, AdaBoost, GBM, which had accuracy values of 80.44%, 83.50%, 88.44% and 89.26% respectively, the proposed prediction model calculated a accuracy of 94.74% utilizing the training data 80%. Figure 5.1(b) shows the evaluation using precision metric. In comparison to RF, KNN, AdaBoost, GBM, which had accuracy values of 78.88%, 81.76%, 82.64% and 83.69% respectively, the proposed prediction model calculated a precision of 90.11% utilizing the training data 80%. Figure 5.1(c) shows the evaluation using recall metric. In comparison to RF, KNN, AdaBoost, GBM, which had recall values of 83.10%, 86.72%, 88.23% and 89.26% respectively, the proposed prediction model calculated a recall of 91.16% utilizing the training data 80%. Figure 5.1(d) shows the evaluation using F1-score metric. In comparison to RF, KNN, AdaBoost, GBM, which had F1-score values of 86.89%, 88.88%, 90.76% and 91.88% respectively, the proposed prediction model measured a F1-score value of 94.96% for the training data 80%.

5.4.2 Analysis using k- fold

i. Accuracy:

Table 5.2(a): Accuracy evaluation for k-fold

K-Fold	RF	KNN	AdaBoost	GBM	Proposed novel approach
1	79.67885	80.77302	82.26312	88.33424	90.10479
2	80.35625	80.94897	82.81266	89.10145	90.47483
3	80.76862	81.77506	86.82336	90.67437	92.0968
4	81.62637	84.46297	88.28823	92.25236	93.7238
5	82.89649	87.10397	91.98564	93.17396	94.86682

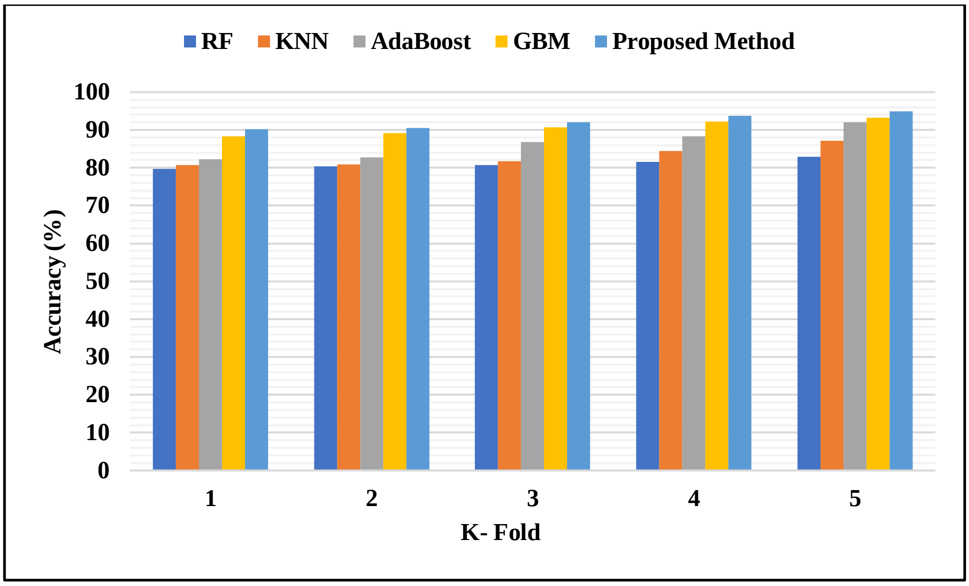


Figure 5.2 (a): Evaluating the effectiveness of the developed method with K-fold consideration - Accuracy

ii. Precision:

Table 5.2(b): Precision evaluation for k-fold

K-Fold	**RF**	**KNN**	**AdaBoost**	**GBM**	**Proposed novel approach**
1	74.44086	80.72877	81.23929	83.36289	88.62418
2	79.09174	80.94965	81.91726	83.81141	89.31672
3	80.00899	81.23929	82.76941	83.90419	91.16942
4	80.0619	82.29656	83.13518	85.08697	92.92632
5	81.1613	82.41667	83.95886	85.42001	93.7238

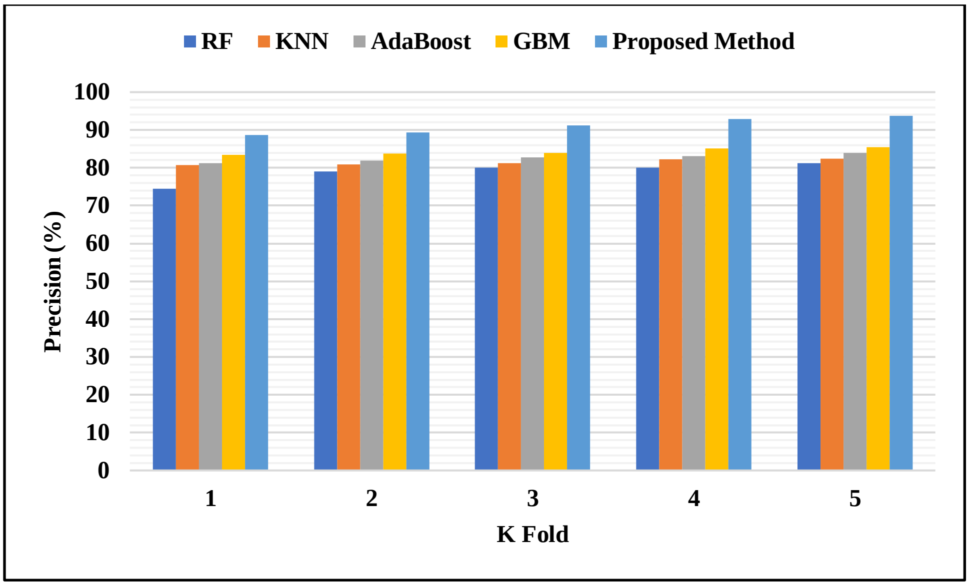


Figure 5.2 (b): Evaluating the effectiveness of the developed method with K-fold consideration: Precision

iii. Recall:

Table 5.2(c): Recall evaluation for k-fold

K-Fold	RF	KNN	AdaBoost	GBM	Proposed novel approach
1	88.13353	88.32607	88.33424	90.29926	91.8072
2	89.02705	89.7871	89.94706	91.4843	93.159
3	89.83067	90.26502	91.23997	92.76444	94.644
4	92.69283	92.71056	92.91197	93.96225	95.6484
5	93.59921	94.1457	94.944	94.944	96

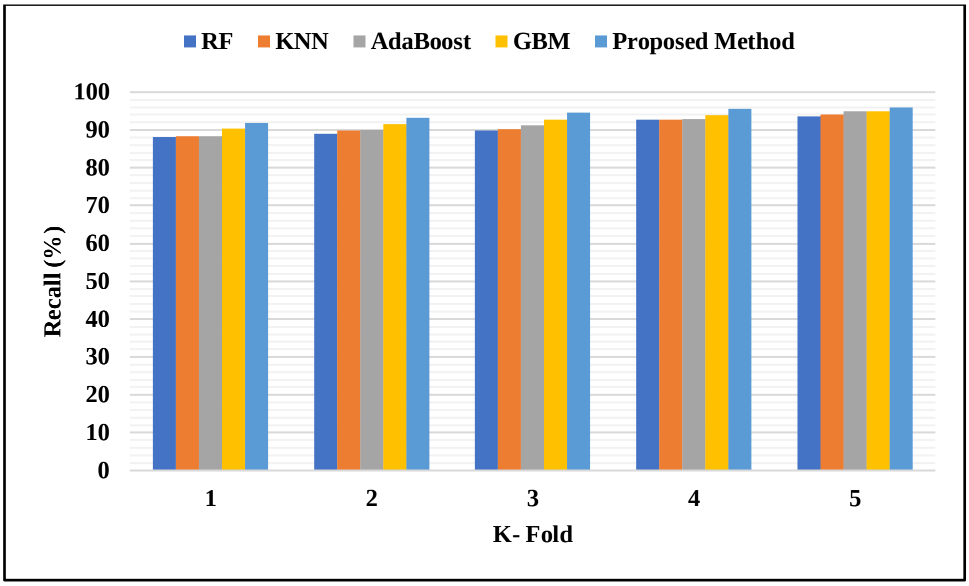


Figure 5.2 (c): Evaluating the effectiveness of the developed method with K-fold consideration: Recall

iv. F1-Score:

Table 5.2(d): F1-score evaluation for k-fold

K-Fold	RF	KNN	AdaBoost	GBM	Proposed novel approach
1	79.84888	82.24958	84.13571	85.6791	89.17211
2	82.1041	83.5744	86.29538	87.0992	90.24741
3	84.21335	84.54934	87.25256	88.48047	90.70677
4	84.7612	86.19745	88.55325	90.10027	93.75914
5	86.91278	87.20711	91.09007	91.68633	94.19838

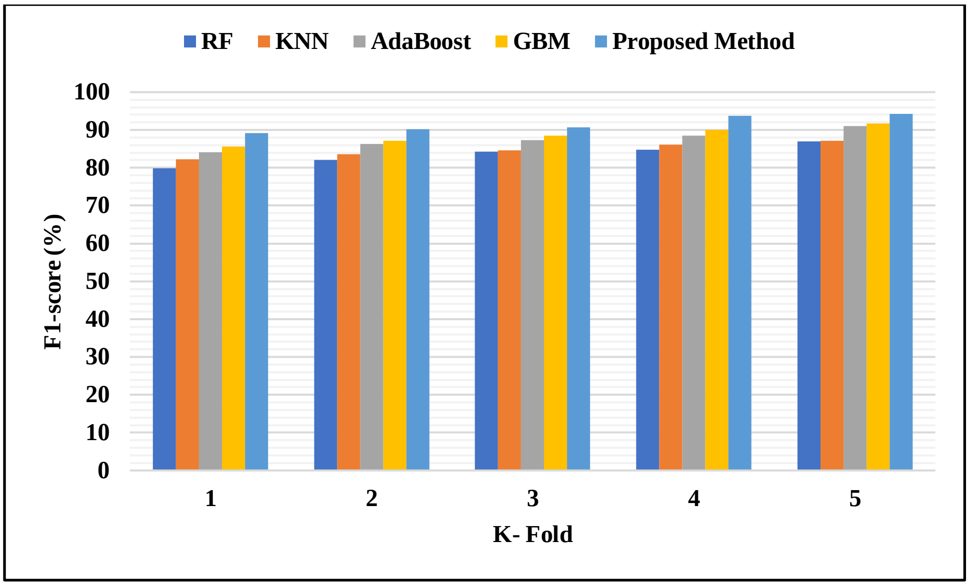


Figure 5.2 (d): Evaluating the effectiveness of the developed method with K-fold consideration: F1- Score

The comparison of the new approach with regard to performance measures is illustrated in Figure 5.2. The analysis based on accuracy is depicted in Figure 5.2(a). The accuracy accessed by RF is 82.90%, KNN is 87.10%, AdaBoost is 91.98%, GBM is 93.17%, and the proposed prediction model is 94.86% when the k-fold number is 5. Figure 5.2(b) shows the evaluation using precision metric. The proposed prediction model measures the precision value of 93.72% in comparison to RF, KNN, AdaBoost, GBM, which had precision values of 81.16%, 82.41%, 83.95% and 85.42% respectively, for the k-fold number is 5. The analysis based on recall metric is depicted in figure 5.2(c). With the k-fold number set to 5, the recall computed by the proposed prediction model is 96%, compared to the other approaches' recall value of 93.60% for RF, 94.14% for KNN, 94.95% for AdaBoost and 94.95% for GBM. Figure 5.1(d) shows the evaluation using F1-score metric at K fold value of 5. In comparison to RF, KNN, AdaBoost, GBM, which had F1-score values of 86.91%, 87.20%, 91.09% and 91.68% respectively, the proposed prediction model measured a F1-score value of 94.19% for the k fold number 5.

The tree-based models showed the most promise. Generally, tree-based models excel in solving problems involving tabular data. Surprisingly, the RF model showed the weakest results, suggesting potential underfitting. In contrast, the GBM model achieved the best results, prompting specialists to thoroughly analyse its attributes and develop a new GBM model. Despite yielding slightly inferior results, this newly proposed GBM-based model offered better interpretability for physicians. These findings indicate that the GBM-based approach stands out as the most accurate model. It is important to mention that while multi-class classification is more representative of real-world situations, the majority of classification approaches still prioritize binary classification [115] . This preference stems, in part, from the challenges associated with training models to effectively generalize over more than two classes. This increased complexity contributes to multi-class models exhibiting lower performance when compared to binary models.

5.5 Deep learning algorithms for model building:

In the current study, machine learning algorithms were utilized for model building, as the dataset primarily comprised text or numeric attribute values. The selection of machine learning methods was suitable for the structured nature of the data. However, with advancements in technology, there exists significant potential for incorporating deep learning models to enhance the prediction accuracy and explore a wider range of diagnostic possibilities. This could be considered for future work.

One of the key future directions of this research lies in expanding the model to detect a broader spectrum of diseases by including patient symptoms in image format. Many symptoms, such as skin rashes, X-rays, and medical scans, are often represented visually. Incorporating such image-based data into the model would allow for more comprehensive disease prediction capabilities. Deep learning methods, specifically Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs), are well-suited for handling such visual data and could be explored as viable alternatives.

Furthermore, in the context of predicting diagnostic tests based on patient symptoms, the choice of algorithm should be carefully tailored to the format and nature of the input data. While machine learning algorithms are effective for handling structured data (such as text and numeric values), CNNs are highly effective for image-based symptoms. Similarly, Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) are particularly useful for sequential or time-series symptom data, as they can capture temporal dependencies in the input. Additionally, Transformer models can be employed for text-based symptom descriptions, leveraging their ability to process long and complex sequences of textual information.

In the context of predicting diagnostic tests based on patient symptoms prior to a medical consultation, the selection of algorithms is contingent upon the structure and nature of the input data. This section provides a detailed technical overview of the suitability of various deep learning and artificial neural network (ANN) algorithms for different data modalities.

5.5.1 Algorithm Selection Based on Data Modality:

i. **Symptoms Represented as Structured/Tabular Data:** For structured data, such as numerical or categorical attributes (e.g., age, symptom severity, or gender), **Multilayer Perceptron (MLP)** or **Autoencoders** are optimal choices. These models can effectively capture non-linear dependencies between input features (symptoms) and target outputs (diagnostic tests). MLP, a fully connected feedforward network, is particularly adept at learning complex mappings through multiple layers of neurons, making it well-suited for tabular data where feature interaction is critical.

ii. **Symptoms Represented as Sequential or Time-Dependent Data:** For time-series or sequential data, such as symptom progression or historical medical records, models designed to capture temporal dependencies are essential. **Recurrent Neural Networks (RNN)**, along with their advanced variants like **Long Short-Term Memory (LSTM)** and **Gated Recurrent Units (GRU)**, are well-suited for this task. RNNs can capture the evolving nature of symptoms over time by maintaining hidden states, while LSTMs and GRUs further enhance the network’s ability to model long-term dependencies by addressing issues like vanishing gradients. These models are ideal for situations where symptom development is influenced by previous time steps or historical medical data, making them appropriate for predicting diagnostic tests based on symptom trajectories over time.

iii. **Symptoms Represented as Textual Data:** When symptoms are provided in textual format, such as free-text responses from patient questionnaires or unstructured medical notes, **Attention-based models** and **Transformers** (e.g., BERT, GPT) are highly effective. These models leverage self-attention mechanisms to capture contextual and sequential dependencies in natural language data. Transformers, in particular, excel at handling long-range dependencies and complex syntactic structures, making them well-suited for processing and extracting relevant information from natural language descriptions of symptoms. **LSTM** models can also be applied in this domain, though Transformers are generally preferred due to their superior performance in large-scale text processing tasks.

iv. **Symptoms Represented as Image Data:** For symptoms that manifest visually, such as medical imaging (e.g., X-rays, CT scans, or dermatological images), **Convolutional Neural Networks (CNNs)** are the most appropriate models. CNNs are designed to process spatially structured data and are highly effective in feature extraction from images, leveraging convolutional layers to detect patterns and hierarchies in the input data. CNNs can link features from medical images to specific diagnostic tests, making them ideal for detecting visual patterns indicative of diseases or conditions.

5.5.2 Deep Learning and ANN Algorithms Applicable to the Problem:

i. **Multilayer Perceptron (MLP):** MLP is a class of feedforward neural networks capable of modeling complex non-linear relationships between inputs and outputs. It is particularly effective for structured/tabular data where symptoms are encoded as numeric or categorical values. MLP can capture the intricate mappings between symptom features and the recommended diagnostic tests through multiple hidden layers.

ii. **Convolutional Neural Networks (CNN):** CNNs, primarily used for spatially structured data such as images, are proficient in feature extraction from visual inputs. Their convolutional layers scan the input data to identify patterns and features, which can then be used to predict the necessary diagnostic tests. CNNs are particularly useful in medical applications where images, such as X-rays or MRI scans, are linked to specific test recommendations.

iii. **Recurrent Neural Networks (RNN):** RNNs are designed for sequential data, making them suitable for tasks where the temporal progression of symptoms is essential. They maintain hidden states that propagate information across time steps, allowing the model to capture dependencies between past and current symptoms. RNNs are well-suited for predicting diagnostic tests in cases where symptom data is collected over time or has a sequential structure.

iv. **Long Short-Term Memory (LSTM):** LSTMs are a specialized form of RNN that address the problem of vanishing gradients, allowing for better retention of long-term dependencies. They are ideal for scenarios where symptoms or medical history span long time periods. By utilizing memory cells and gating mechanisms, LSTMs can selectively retain or forget information over extended sequences, making them particularly useful for predicting diagnostic tests for chronic conditions or when past symptoms influence current diagnosis.

v. **Gated Recurrent Units (GRU):** GRUs are a simplified variant of LSTMs, designed to provide similar functionality with reduced computational overhead. They are more efficient than LSTMs in terms of both memory usage and training time while retaining the ability to capture dependencies in sequential data. GRUs are particularly suited for real-time applications or scenarios where rapid diagnostic test prediction is required based on short-term symptom histories.

vi. **Autoencoders:** Autoencoders can be leveraged for dimensionality reduction or anomaly detection in datasets with complex or high-dimensional symptom sets. By learning a compressed representation of the input data, autoencoders can focus on the most critical features of the symptom data. Additionally, they can identify anomalous symptom patterns that may require specialized diagnostic tests.

vii. **Generative Adversarial Networks (GANs):** While primarily used for data generation, GANs can be adapted to augment training datasets by generating synthetic examples of patient symptoms. This can be particularly useful when training data is scarce or incomplete, improving model generalization by simulating diverse symptom patterns. GANs can help enhance the prediction of diagnostic tests by creating synthetic datasets that reflect real-world variability in symptom presentation.

viii. **Deep Belief Networks (DBN):** DBNs are effective for unsupervised feature learning and can be employed for pretraining networks on large, unlabeled datasets. They are useful for capturing hierarchical relationships in the data and can be fine-tuned for predicting diagnostic tests once sufficient labeled data is available. DBNs excel in scenarios where limited labeled data is available but large amounts of unlabeled symptom data can be utilized.

ix. **Attention-based Models and Transformers:** Transformers, which employ self-attention mechanisms, are the current state-of-the-art models in natural language processing. These models are particularly useful for handling textual data, such as patient symptom descriptions or medical notes, by capturing long-range dependencies and understanding contextual relationships. Their ability to process large sequences of text makes them highly effective in predicting diagnostic tests from natural language symptom descriptions.

The selection of the appropriate algorithm for predicting diagnostic tests based on patient symptoms depends heavily on the format and complexity of the input data. Structured and tabular data are best handled by models like MLP, while sequential and time-dependent data benefit from RNNs, LSTMs, and GRUs. Textual data is most effectively processed by attention-based models and Transformers, while CNNs are the optimal choice for image-based data. Future research can further expand the scope of this work by integrating multiple data modalities, leveraging the strengths of different deep learning architectures, and improving the model's predictive performance across diverse symptom representations.

Chapter 6

Conclusion and Future Work

Communicable diseases can significantly impact hospital management systems due to their potential to overwhelm limited healthcare staff. When an outbreak occurs, hospitals might experience an influx of patients needing care, testing, and isolation. With a shortage of healthcare workers, it becomes challenging to provide prompt and effective care to all patients, potentially leading to delays, increased wait times, and compromised patient safety. Furthermore, the growing count of infected individuals within hospital settings can heighten the likelihood of transmission among patients and healthcare staff. Close proximity and contact can facilitate the spread of the disease, even if rigorous infection control measures are in place. This puts both patients and healthcare workers at a higher risk of infection. Patients are increasingly seeking strategies to minimize hospital visits as a proactive measure to mitigate the risk of infection. This trend is substantiated by several published articles and statistics. The patients' desire to reduce hospital visits to minimize infection risk is supported by studies citing concerns about infections, vulnerability of certain demographics, time constraints, and the convenience of virtual consultations. These factors collectively underscore the need for healthcare systems to adapt and provide patient-centric alternatives to in-person care.

Instead of solely concentrating on the pandemic, today's trailblazers in the healthcare field are shifting their primary focus towards adapting to new realities in medical management. Notably, emerging healthcare leaders have identified fresh key priorities for the year 2022 and beyond. These include enhancing the staff's professional experience and bridging the gap between the potential of predictive analytics and its current implementation, among other objectives. Given the projection of a substantial shortage in the healthcare workforce by 2030, enhancing the staff's overall experience has risen to the forefront of the present healthcare agenda. The latest report indicates that providing enhanced training in digital health technologies will be pivotal for advancement. This training will help alleviate the sense of being overwhelmed by data-driven processes and enable the workforce to readily embrace new workflows.

However, amplifying training is just one component of the larger solution. Addressing the long-term labor crisis will fundamentally hinge on effective collaboration among governments, regulators, and the industry as a whole. This collaboration is essential to enhance working conditions comprehensively. An array of factors, including data isolation, interoperability concerns, and technological infrastructure limitations, has contributed to the uneven adoption of predictive analytics thus far. Encouragingly, a growing number of healthcare trailblazers are now introducing predictive analytics technology. Their efforts are inspiring others to foster its adoption within their own facilities. As organizations use computer insights in healthcare to make better choices and reduce paperwork, more people will learn from each other. Early users will help late users, and tech companies will make healthcare better.

This research provides data driven, systematic method to predict primary diagnostic tests to be done prior to the first visit to the hospital or doctor based on the symptoms. This will help the patients to reduce hospital visits. This proposed novel approach will also help the hospitals to optimize the patient screening process. It will provide the hospital the feasibility to attend more patient and reduce workload of the healthcare system. The patient's symptoms and demographic information is provided to the proposed model through a validated and verified questionnaire using API. The primary data obtained from the questionnaire is used to train and tests the model. This research proposes a novel approach to predict diagnostic tests and probable disease with the aid of a Competitive verse Water Wave Optimization method. The model predicts diagnostic tests to be done for the input symptoms with the help of proposed CWWO method. An innovative machine learning framework is put forth for the multi-class categorization of diseases such as Dengue, Chikungunya, and other communicable diseases within the Kolhapur district of Maharashtra, India. This model exclusively relies on the patients' clinical details and socio-demographic information. The study encompasses an assessment of four distinct machine learning models: Random Forest

(RF), K-Nearest Neighbors (KNN), AdaBoost, Gradient Boosting Machine (GBM) and the proposed GBM with Competitive verse Water Wave Optimization method. These models are evaluated with two methods as varying train-test split and K-fold method against the metrics- accuracy, precision, recall, F1-score.

Random Forest (RF) is an ensemble learning technique based on constructing a multitude of decision trees during training and aggregating their predictions. The algorithm excels in handling structured data, mitigating overfitting through its ensemble approach. In the evaluation, RF demonstrated gradual improvement in performance metrics as the training percentage increased. Specifically, accuracy improved from 77.82% to 81.71% with 50% to 90% training data. Precision and recall also showed consistent enhancement, with precision reaching 81.96% and recall 85.51% at the highest training ratio. The F1-score exhibited a similar upward trend, culminating at 87.75%. Despite these improvements, RF was outperformed by GBM and the proposed hybrid approach in final accuracy and overall predictive performance

K-Nearest Neighbors (KNN) operates by classifying data based on the majority vote of the nearest neighbors. Its performance is notably sensitive to the dimensionality and size of the dataset. In this study, KNN displayed solid performance with accuracy ranging from 80.36% to 84.60% across different training percentages. Precision peaked at 82.04% and recall at 87.91%, while the F1-score reached 88.98%. KNN's performance improved with the increase in training data, but it still lagged behind GBM and the hybrid approach in terms of accuracy and F1-score, highlighting its limitations in handling complex patterns compared to more sophisticated algorithms.

AdaBoost, or Adaptive Boosting, enhances weak classifiers by focusing on errors made in previous rounds of training. It is effective in improving the performance of base learners and is less prone to overfitting compared to single classifiers. The algorithm showed progressive improvements across all metrics as the training percentage increased. Accuracy ranged from 80.95% to 89.29%, with precision reaching 83.26% and recall 89.57%. The F1-score increased steadily, achieving 90.96% at the highest training ratio. AdaBoost outperformed RF and KNN in terms of accuracy and recall but was outpaced by GBM and the proposed hybrid approach, especially in F1-score and overall accuracy.

Gradient Boosting Machine (GBM) constructs an ensemble of trees sequentially, where each tree corrects the errors of its predecessors. This approach effectively minimizes prediction errors by focusing on residuals. GBM demonstrated strong performance metrics across training percentages, achieving an accuracy of up to 93.17%. Precision and recall also showed substantial improvements, with values peaking at 85.42% and 94.94%, respectively. The F1-score reached 91.69% in the most extensive training scenario. GBM consistently outperformed RF, KNN, and AdaBoost, showcasing its superior capability in handling complex data and improving predictive accuracy.

The proposed novel approach integrates Competitive Multiverse Optimization (CMVO) and Water Wave Optimization (WWO) with GBM to refine hyperparameters and enhance model performance. This hybrid methodology optimizes GBM's parameters through metaheuristic techniques, leading to significant improvements in predictive accuracy. The proposed model achieved an accuracy of 96.00%, surpassing all other algorithms. Precision and recall were also elevated, with the precision reaching 93.72% and recall 96.00%. The F1-score peaked at 96.00%, indicating exceptional balance between precision and recall. This approach not only outperformed RF, KNN, AdaBoost, and GBM but also demonstrated enhanced generalization and accuracy, validating the efficacy of metaheuristic optimization in advancing predictive modeling for diagnostic tests.

While RF, KNN, and AdaBoost provide competitive results, GBM and the proposed hybrid approach deliver superior performance. The proposed hybrid method, leveraging metaheuristic optimization with GBM, significantly enhances model accuracy and predictive power, demonstrating its effectiveness in the context of diagnostic test prediction based on patient symptoms. This comparative analysis underscores the advanced capabilities of hybrid optimization techniques in improving model performance in complex healthcare datasets.

The predicted results will be visible to the patients only after the doctor's authorization. The outcomes of the proposed approach application show that proposed method with 94.86% of accuracy, 93.72% precision, 96% recall and 94.19% of F1 score predicts the probable disease and diagnostic tests to be done as accurate and with a less error than methods such as RF, KNN, AdaBoost and GBM. Among the models suggested, the GBM-based approach demonstrated the most favourable outcomes overall. On the other hand, the RF model yielded the least favourable results, exhibiting a decline in performance compared to training with grid search, hinting at a potential underfitting issue. Notably, the GBM-based model showcased superior performance compared to all alternative models.

The findings of this investigation highlight the viability of achieving robust classification outcomes using exclusively medical and demographic information. The novel model in this context serves as a valuable, economically efficient, and time-effective alternative. This holds particular significance within resource-constrained scenarios, wherein solely the data gathered from the patient's health facility visit remains accessible, making the model a potentially invaluable asset

Currently, the research scope is limited to the following communicable diseases: Dengue, Malaria, Chikungunya, Typhoid, Covid-19, Viral fever and Flu. In the future, this proposed approach can be applied to other diseases. The hyperparameters employed for both feature selection and classification algorithms are established standards commonly drawn from existing literature. However, the exploration of modifications and optimizations to these hyperparameters could pave the way for future research opportunities within this domain.

Implementing changes in clinical practice demands a multifaceted approach. It's essential to educate patients, helping them recognize situations where adopting a "less is more" strategy holds value in healthcare [116]. Equally vital is the education of clinicians, enabling them to quantitatively decipher diagnostic performance traits with the aid of effective decision-support tools. While these challenges are substantial, achieving a proficient equilibrium between the advantages and drawbacks of diagnostic testing is paramount. This equilibrium plays a pivotal role in optimizing cost, quality, and accessibility within our healthcare system concurrently.

The ensembled machine learning models has the potential to assist in detecting and categorizing communicable diseases [117]. It offers clinicians a diagnostic instrument grounded in actual data that can complement their clinical expertise. Additionally, these models could serve as a robust surveillance tool in the period preceding an epidemic outbreak.

Patient's Screening Data Collection

Proforma

Date:

Name	
Age	
Gender	
OPD/IPD	
Address	
Contact Info	
Occupation	

Complaints:

Cold	Do you have running nose? (yes/no)	
	Do you have sore throat? (yes/no)	
Cough	Do you have cough? Yes/No	
	Duration – < 3 weeks/3- 8 weeks/>8 weeks	
	Do you have expectoration? Yes/No	
	Amount – scanty/copious	
	Color of sputum – yellowish/greenish/blood tinged	
	Aggravated on change of posture? Yes/No	
	Associated with stridor? Yes/No	
	Associated with breathlessness/ chest pain	
	History of beedi/ cigarette smoking? Yes/ No	
	Past History of Tuberculosis/bronchial asthma/HIV	
Breathlessness	Do you have breathlessness? Yes/No	

	Do you have breathing difficulty while performing routine activities? Yes/No	
	Complaints of breathing difficulty in the night? Yes/No	
	Aggravated on exertional activities? Yes/No	
	Associated with chest pain/cough	
	Associated with loss of weight/ loss of appetite	
Fever	Do you have fever? Yes/No	
	Is it high grade/low grade? Low grade: High Grade:	
	Duration(days)	
	Associated with chills and rigor? Yes/No	
	Diurnal variation? Yes/No	
	Continuous/ intermittent	
	Relieved with medications Yes/No	
	Associated with chest pain/breathlessness/cough	
Loose stool	Do you have loose stools? Yes/No	
	Duration(days)	
	No of episodes per day	
	Associated with blood- Yes/No	
	Associated with abdominal pain/ fever/vomiting	
	Fowl smelling? Yes/No	
	Relieved with medications? Yes/No	
Vomiting	Do you have vomiting? Yes/No	
	Number of episodes	

	Color of vomitus	
	Associated with abdominal pain/loose stools/fever	
	Relieved with medications- Yes/No	
Abdominal Pain	Do you have abdominal pain? Yes/No	
	Duration	
	Localized / Diffused	
	Relieved with medications? Yes/No	
	Associated with fever/loose stools/vomiting	
Headache	Do you have headache? Yes/No	
	Duration	
	Localized / Diffused	
	Associated with fever/giddiness/ blurring of vision/vomiting	
	Relieved with medications – Yes/No	
Past Complaints	Similar Complaints in the past? Yes/ No	
Covid-19 Vaccination	First dose/second dose/not vaccinated	
Any other symptoms		

Recommended Tests:

Sr. No.	**Tests**

Diagnosis

Publication Details

Published :

1. Priyanka Shivaprasad More, Baljit Singh Saini & Rakesh Kumar Sharma (2023) Optimisation algorithm in health care: review on the State-of-the-Art models, Journal of Experimental & Theoretical Artificial Intelligence, DOI: 10.1080/0952813X.2023.2217813
2. Priyanka Shivaprasad More & Baljit Singh Saini (2023) Competitive verse water wave optimisation enabled COVID-Net for COVID-19 detection from chest X-ray images, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 11:4, 1418-1430, DOI: 10.1080/21681163.2022.2140074
3. Priyanka Shivaprasad More & Baljit Singh Saini (2023) KH-FC: krill herd-based fractional calculus algorithm for text document clustering using MapReduce structure, International Journal of Computational Science and Engineering, Vol 25, No. 6, 668-684, DOI: 10.1504/IJCSE.2022.127188
4. P. S. More, B. S. Saini, and K. S. Bhatia, (2020), "Krill Herd (KH) algorithm for text document clustering using TF–IDF features", International Conference on Smart Machine Intelligence and Real-Time Computing (SmartCom 2020), Pauri, Garhwal, Uttarakhand, India, 26-27 June 2020. Smart Computing: CRC Press, 2021, pp. 502-512, (https://doi.org/10.1201/9781003167488)
5. Priyanka Shivaprasad More, Baljit Singh Saini & Rakesh Kumar Sharma(2023), "AI-Based Diagnostic Test Prediction- An Effective Way To Handle The New Normal Era In Healthcare", 6th International Conference on Contemporary Computing and Informatics (IC3I-2023), Amity University, Greater Noida, 14-16 Sept, 2023.

Communicated:

1. Case study on different metaheuristic optimization techniques for disease detection and classification.

2. A Novel Approach for Predicting Diagnostic Tests and Probable Disease by Utilizing Meta-Heuristic Optimization.

Copyright:

1. Forecasting clinical tests using metaheuristic optimization technique. Registration number: **L-150909/2024**

Bibliography:

[1] K. Kaur and Y. Kumar, "Swarm Intelligence and its applications towards Various Computing: A Systematic Review," *Proc. Int. Conf. Intell. Eng. Manag. ICIEM 2020*, pp. 57–62, 2020, doi: 10.1109/ICIEM48762.2020.9160177.

[2] M. A. Khan and F. Algarni, "A Healthcare Monitoring System for the Diagnosis of Heart Disease in the IoMT Cloud Environment Using MSSO-ANFIS," *IEEE Access*, vol. 8, pp. 122259–122269, 2020, doi: 10.1109/ACCESS.2020.3006424.

[3] A. K. Dubey, "Optimized hybrid learning for multi disease prediction enabled by lion with butterfly optimization algorithm," *Sadhana - Acad. Proc. Eng. Sci.*, vol. 46, no. 2, 2021, doi: 10.1007/s12046-021-01574-8.

[4] X. Li, J. Zhang, and F. Safara, "Improving the Accuracy of Diabetes Diagnosis Applications through a Hybrid Feature Selection Algorithm," *Neural Process. Lett.*, vol. 55, no. 1, pp. 153–169, 2023, doi: 10.1007/s11063-021-10491-0.

[5] W. Qiao and Z. Yang, "Solving large-scale function optimization problem by using a new metaheuristic algorithm based on quantum dolphin swarm algorithm," *IEEE Access*, vol. 7, pp. 138972–138989, 2019, doi: 10.1109/ACCESS.2019.2942169.

[6] K. Surabhi and K. Yogesh, *Nature-Inspired Metaheuristic Algorithms for Constraint Handling: Challenges, Issues, and Research Perspective*. 2021.

[7] L. E. Søvold *et al.*, "Prioritizing the Mental Health and Well-Being of Healthcare Workers: An Urgent Global Public Health Priority," *Front. Public Heal.*, vol. 9, no. May, pp. 1–12, 2021, doi: 10.3389/fpubh.2021.679397.

[8] R. Kovacs and M. Lagarde, "Does high workload reduce the quality of healthcare? Evidence from rural Senegal," *J. Health Econ.*, vol. 82, no. February, p. 102600, 2022, doi: 10.1016/j.jhealeco.2022.102600.

[9] A. T. Tamata and M. Mohammadnezhad, "A systematic review study on the factors affecting shortage of nursing workforce in the hospitals," *Nurs. Open*, vol. 10, no. 3, pp. 1247–1257, 2023, doi: 10.1002/nop2.1434.

[10] I. Torjesen, "Doctors are under more work pressure than during height of

covid-19 pandemic in 2020," *BMJ*, vol. 375, p. n3088, 2021, doi: 10.1136/bmj.n3088.

[11] L. A. Morgantini *et al.*, "Factors contributing to healthcare professional burnout during the COVID-19 pandemic: A rapid turnaround global survey," *PLoS One*, vol. 15, no. 9 September, pp. 1–11, 2020, doi: 10.1371/journal.pone.0238217.

[12] U. S. D. of H. and H. S. Assistant Secretary for Planning and Evaluation, "Impact of the COVID-19 Pandemic on the Hospital and Outpatient Clinician Workforce: Challenges and Policy Responses," *ASPE - Off. Heal. Policy*, vol. 2022, no. 13, pp. 1–27, 2022, [Online]. Available: https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//196851/COVIDNH.pdf

[13] I. Portoghese, M. Galletta, R. C. Coppola, G. Finco, and M. Campagna, "Burnout and workload among health care workers: The moderating role of job control," *Saf. Health Work*, vol. 5, no. 3, pp. 152–157, 2014, doi: 10.1016/j.shaw.2014.05.004.

[14] A. S. Alonazi *et al.*, "The Effect of Increasing the Health Practitioner's Workload on Patients and Their Safety," *Saudi J. Nurs. Heal. Care*, vol. 5, no. 12, pp. 301–307, 2022, doi: 10.36348/sjnhc.2022.v05i12.001.

[15] M. Haque *et al.*, "Strategies to prevent healthcare-associated infections: A narrative overview," *Risk Management and Healthcare Policy*, vol. 13. Dove Medical Press Ltd, pp. 1765–1780, 2020. doi: 10.2147/RMHP.S269315.

[16] H. A. Khan, F. K. Baig, and R. Mehboob, "Nosocomial infections: Epidemiology, prevention, control and surveillance," *Asian Pac. J. Trop. Biomed.*, vol. 7, no. 5, pp. 478–482, 2017, doi: 10.1016/j.apjtb.2017.01.019.

[17] G. Ripabelli, A. Salzo, A. Mariano, M. L. Sammarco, and M. Tamburro, "Healthcare-associated infections point prevalence survey and antimicrobials use in acute care hospitals (PPS 2016–2017) and long-term care facilities (HALT-3): a comprehensive report of the first experience in Molise Region, Central Italy, and targeted in," *J. Infect. Public Health*, vol. 12, no. 4, pp. 509–515, 2019, doi: 10.1016/j.jiph.2019.01.060.

[18] R. Danasekaran, G. Mani, and K. Annadurai, "Prevention of healthcare-

associated infections: protecting patients, saving lives," *Int. J. Community Med. Public Heal.*, vol. 1, no. 1, p. 67, 2014, doi: 10.5455/2394-6040.ijcmph20141114.

[19] R. M. Klevens *et al.*, "Estimating health care-associated infections and deaths in U.S. Hospitals, 2002," *Public Health Rep.*, vol. 122, no. 2, pp. 160–166, 2007, doi: 10.1177/003335490712200205.

[20] M. Sartelli, J. Mckimm, and M. A. Bakar, "Health care-associated infections – an overview," pp. 2321–2333, 2018.

[21] A. Cassini *et al.*, "Burden of Six Healthcare-Associated Infections on European Population Health: Estimating Incidence-Based Disability-Adjusted Life Years through a Population Prevalence-Based Modelling Study," *PLoS Med.*, vol. 13, no. 10, pp. 1–16, 2016, doi: 10.1371/journal.pmed.1002150.

[22] M. Haque, J. McKimm, B. Godman, M. Abu Bakar, and M. Sartelli, "Initiatives to reduce postoperative surgical site infections of the head and neck cancer surgery with a special emphasis on developing countries," *Expert Rev. Anticancer Ther.*, vol. 19, no. 1, pp. 81–92, 2019, doi: 10.1080/14737140.2019.1544497.

[23] D. Pittet, B. Allegranzi, J. Storr, and L. Donaldson, "'Clean Care is Safer Care': the Global Patient Safety Challenge 2005-2006," *Int. J. Infect. Dis.*, vol. 10, no. 6, pp. 419–424, 2006, doi: 10.1016/j.ijid.2006.06.001.

[24] I. H. S. Guidelines Review Committee, *Guidelines on core components of infection prevention and control programmes at the national and acute health care facility level*. 2016. [Online]. Available: http://apps.who.int/bookorders.

[25] K. Dembrower *et al.*, "Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study," *Lancet Digit. Heal.*, vol. 2, no. 9, pp. e468–e474, 2020, doi: 10.1016/S2589-7500(20)30185-0.

[26] D. Ardila *et al.*, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nat. Med.*, vol. 25, no. 6, pp. 954–961, 2019, doi: 10.1038/s41591-019-0447-x.

[27] J. M. Murray, B. Wiegand, B. Hadaschik, K. Herrmann, and J. Kleesiek, "Virtual Biopsy: Just an AI Software or a Medical Procedure?," *J. Nucl. Med.*,

vol. 63, no. 4, pp. 511–513, 2022, doi: 10.2967/jnumed.121.263749.

[28] J. De Fauw *et al.*, “Clinically applicable deep learning for diagnosis and referral in retinal disease,” *Nat. Med.*, vol. 24, no. 9, pp. 1342–1350, 2018, doi: 10.1038/s41591-018-0107-6.

[29] E. Fristed *et al.*, “A remote speech-based AI system to screen for early Alzheimer’s disease via smartphones,” *Alzheimer’s \& Dement. Diagnosis, Assess. \& Dis. Monit.*, vol. 14, no. 1, p. e12366, 2022, doi: https://doi.org/10.1002/dad2.12366.

[30] A. Rodman, T. A. Buckley, A. K. Manrai, and D. J. Morgan, “Artificial Intelligence vs Clinician Performance in Estimating Probabilities of Diagnoses before and after Testing,” *JAMA Netw. Open*, vol. 6, no. 12, p. E2347075, 2023, doi: 10.1001/jamanetworkopen.2023.47075.

[31] D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, and R. K. Tekade, “Artificial intelligence in drug discovery and development,” *Drug Discovery Today*, vol. 26, no. 1. 2021. doi: 10.1016/j.drudis.2020.10.010.

[32] H. Askr, E. Elgeldawi, H. Aboul Ella, Y. A. M. M. Elshaier, M. M. Gomaa, and A. E. Hassanien, “Deep learning in drug discovery: an integrative review and future challenges,” *Artif. Intell. Rev.*, vol. 56, no. 7, 2023, doi: 10.1007/s10462-022-10306-1.

[33] A. Blasiak, J. Khong, and T. Kee, “CURATE.AI: Optimizing Personalized Medicine with Artificial Intelligence,” *SLAS Technology*, vol. 25, no. 2. 2020. doi: 10.1177/2472630319890316.

[34] K. B. Johnson *et al.*, “Precision Medicine, AI, and the Future of Personalized Health Care,” *Clinical and Translational Science*, vol. 14, no. 1. 2021. doi: 10.1111/cts.12884.

[35] T. Shaik *et al.*, “Remote patient monitoring using artificial intelligence: Current state, applications, and challenges,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 13, no. 2. 2023. doi: 10.1002/widm.1485.

[36] A. Dubey and A. Tiwari, “Artificial intelligence and remote patient monitoring in US healthcare market: a literature review,” *J. Mark. Access Heal. Policy*, vol. 11, no. 1, 2023, doi: 10.1080/20016689.2023.2205618.

[37] T. Malche *et al.*, "Artificial Intelligence of Things- (AIoT-) Based Patient Activity Tracking System for Remote Patient Monitoring," *J. Healthc. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/8732213.

[38] Q. Gao, L. Yang, M. Lu, R. Jin, H. Ye, and T. Ma, "The artificial intelligence and machine learning in lung cancer immunotherapy," *Journal of Hematology and Oncology*, vol. 16, no. 1. 2023. doi: 10.1186/s13045-023-01456-y.

[39] V. V. Khanna, K. Chadaga, N. Sampathila, S. Prabhu, R. Chadaga, and S. Umakanth, "Diagnosing COVID-19 using artificial intelligence: a comprehensive review," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 11, no. 1. 2022. doi: 10.1007/s13721-022-00367-1.

[40] H. T. Abbas *et al.*, "Predicting long-term type 2 diabetes with support vector machine using oral glucose tolerance test," *PLoS One*, vol. 14, no. 12, 2019, doi: 10.1371/journal.pone.0219636.

[41] T. M. Ghazal *et al.*, "Supervised Machine Learning Empowered Multifactorial Genetic Inheritance Disorder Prediction," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/1051388.

[42] S. Goto, D. K. McGuire, and S. Goto, "The Future Role of High-Performance Computing in Cardiovascular Medicine and Science -Impact of Multi-Dimensional Data Analysis," *J. Atheroscler. Thromb.*, vol. 29, no. 5, pp. 559–562, 2022, doi: 10.5551/jat.RV17062.

[43] N. Lewandowski and B. Koller, "Transforming medical sciences with high-performance computing, high-performance data analytics and AI," *Technol. Heal. Care*, vol. 31, no. 4, pp. 1505–1507, 2023, doi: 10.3233/THC-237000.

[44] R. S. Bennett *et al.*, "Kikwit Ebola virus disease progression in the rhesus monkey animal model," *Viruses*, vol. 12, no. 7, pp. 1–14, 2020, doi: 10.3390/v12070753.

[45] M. Kumar, S. K. Khatri, and M. Mohammadian, "Review on Breast Cancer Disease Predictive Modelling using Swarm Intelligence," *2020 Int. Conf. Comput. Perform. Eval. ComPE 2020*, pp. 523–530, 2020, doi: 10.1109/ComPE49325.2020.9200117.

[46] F. Wang *et al.*, "Generation of a Hutchinson–Gilford progeria syndrome monkey model by base editing," *Protein Cell*, vol. 11, no. 11, pp. 809–824,

2020, doi: 10.1007/s13238-020-00740-8.

[47] J. R. Lambert and E. Perumal, “Oppositional firefly optimization based optimal feature selection in chronic kidney disease classification using deep neural network,” *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 4, pp. 1799–1810, 2022, doi: 10.1007/s12652-021-03477-2.

[48] V. Hajihashemi, Z. Hassani, I. Dehmajnoonie, and K. Borna, “Hybrid Algorithms of Whale optimization algorithm and k-nearest neighbor to Predict the liver disease,” *EAI Endorsed Trans. Context. Syst. Appl.*, vol. 6, no. 16, p. 156838, 2019, doi: 10.4108/eai.13-7-2018.156838.

[49] S. Kaur, Y. Kumar, A. Koul, and S. Kumar Kamboj, “A Systematic Review on Metaheuristic Optimization Techniques for Feature Selections in Disease Diagnosis: Open Issues and Challenges,” *Archives of Computational Methods in Engineering*. Springer Science and Business Media B.V., Apr. 01, 2022. doi: 10.1007/s11831-022-09853-1.

[50] T. Xie, J. Yao, and Z. Zhou, “DA-based parameter optimization of combined kernel support vector machine for cancer diagnosis,” *Processes*, vol. 7, no. 5, 2019, doi: 10.3390/pr7050263.

[51] B. K. Kihel, S. K. Kihel, and S. Chouraqui, “Immune Grey Wolf Optimizer for Attribute Reduction: Application for Medical Systems,” *Int. J. Intell. Eng. Syst.*, vol. 14, no. 4, 2021, doi: 10.22266/ijies2021.0831.04.

[52] W. J. Gutjahr and M. S. Rauner, “An ACO algorithm for a dynamic regional nurse-scheduling problem in Austria,” *Comput. Oper. Res.*, vol. 34, no. 3, 2007, doi: 10.1016/j.cor.2005.03.018.

[53] J. Wang, X. Wang, X. Li, and J. Yi, “A Hybrid Particle Swarm Optimization Algorithm with Dynamic Adjustment of Inertia Weight Based on a New Feature Selection Method to Optimize SVM Parameters,” *Entropy*, vol. 25, no. 3, 2023, doi: 10.3390/e25030531.

[54] S. Perumal and T. Velmurugan, “Lung cancer detection and classification on CT scan images using Enhanced Artificial Bee Colony Optimization,” *Int. J. Eng. Technol.*, vol. 7, no. 2.26 Special Issue 26, 2018, doi: 10.14419/ijet.v7i2.26.12538.

[55] J. Nayak, B. Naik, P. Dinesh, K. Vakula, and P. B. Dash, “Firefly Algorithm in

Biomedical and Health Care: Advances, Issues and Challenges," *SN Computer Science*, vol. 1, no. 6. 2020. doi: 10.1007/s42979-020-00320-x.

[56] H. Verma, D. Verma, and P. K. Tiwari, "A population based hybrid FCM-PSO algorithm for clustering analysis and segmentation of brain image," *Expert Syst. Appl.*, vol. 167, 2021, doi: 10.1016/j.eswa.2020.114121.

[57] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Futur. Healthc. J.*, vol. 6, no. 2, 2019, doi: 10.7861/futurehosp.6-2-94.

[58] D. K. Choubey, P. Kumar, S. Tripathi, and S. Kumar, "Performance evaluation of classification methods with PCA and PSO for diabetes," *Netw. Model. Anal. Heal. Informatics Bioinforma.*, vol. 9, no. 1, 2020, doi: 10.1007/s13721-019-0210-8.

[59] G. Hassan and A. E. Hassanien, "Retinal fundus vasculature multilevel segmentation using whale optimization algorithm," *Signal, Image Video Process.*, vol. 12, no. 2, 2018, doi: 10.1007/s11760-017-1154-z.

[60] S. K. Majhi, "How Effective Is the Moth-Flame Optimization in Diabetes Data Classification," in *Advances in Intelligent Systems and Computing*, vol. 740, 2019. doi: 10.1007/978-981-13-1280-9_7.

[61] T. Si, A. De, and A. K. Bhattacharjee, "MRI brain lesion segmentation using generalized opposition-based glowworm swarm optimization," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 14, no. 5, 2016, doi: 10.1142/S0219691316500417.

[62] S. Lu, S. Liu, S. H. Wang, and Y. D. Zhang, "Cerebral Microbleed Detection via Convolutional Neural Network and Extreme Learning Machine," *Front. Comput. Neurosci.*, vol. 15, 2021, doi: 10.3389/fncom.2021.738885.

[63] C. Gunavathi and K. Premalatha, "Cuckoo search optimisation for feature selection in cancer classification: A new approach," *Int. J. Data Min. Bioinform.*, vol. 13, no. 3, 2015, doi: 10.1504/IJDMB.2015.072092.

[64] S. Kumar and M. Singh, "No Title," *Vietnam J. Comput. Sci.*, vol. 8, no. 2, p. 177, 2021.

[65] W. Ha and Z. Vahedi, *Automatic Breast Tumor Diagnosis in MRI Based on a Hybrid CNN and Feature-Based Method Using Improved Deer Hunting*

Optimization Algorithm, vol. null. 2021.

[66] M. N. Sudha, S. Selvarajan, and M. Suganthi, “Feature selection using improved lion optimisation algorithm for breast cancer classification,” *Int. J. Bio-Inspired Comput.*, vol. 14, no. 4, 2019, doi: 10.1504/IJBIC.2019.103963.

[67] S. T. Dehkordi, A. Khatibi Bardsiri, and M. H. Zahedi, “Prediction and Diagnosis of Diabetes Mellitus using a Water Wave Optimization Algorithm,” *J. AI Data Min.*, vol. 7, no. 4, 2019.

[68] H. Tabrizchi, M. Tabrizchi, and H. Tabrizchi, “Breast cancer diagnosis using a multi-verse optimizer-based gradient boosting decision tree,” *SN Appl. Sci.*, vol. 2, no. 4, 2020, doi: 10.1007/s42452-020-2575-9.

[69] R. R. Rani and D. Ramyachitra, “Krill Herd Optimization algorithm for cancer feature selection and random forest technique for classification,” in *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, 2018, vol. 2017-November. doi: 10.1109/ICSESS.2017.8342875.

[70] K. M. Kuo, P. C. Talley, and C. S. Chang, “The accuracy of machine learning approaches using non-image data for the prediction of COVID-19: A meta-analysis,” *International Journal of Medical Informatics*, vol. 164. 2022. doi: 10.1016/j.ijmedinf.2022.104791.

[71] O. Abayomi-Alli, R. Damaševičius, R. Maskeliunas, and S. Misra, “An Ensemble Learning Model for COVID-19 Detection from Blood Test Samples,” *Sensors*, vol. 22, p. 2224, May 2022, doi: 10.3390/s22062224.

[72] A. Fathi Kazerooni *et al.*, “Clinical measures, radiomics, and genomics offer synergistic value in AI-based prediction of overall survival in patients with glioblastoma,” *Sci. Rep.*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-12699-z.

[73] C. Li *et al.*, “Deep learning-based AI model for signet-ring cell carcinoma diagnosis and chemotherapy response prediction in gastric cancer,” *Med. Phys.*, vol. 49, no. 3, 2022, doi: 10.1002/mp.15437.

[74] M. Combalia *et al.*, “Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: the 2019 International Skin Imaging Collaboration Grand Challenge,” *Lancet Digit. Heal.*, vol. 4, no. 5,

2022, doi: 10.1016/S2589-7500(22)00021-8.

[75] J. Rashid *et al.*, “An Augmented Artificial Intelligence Approach for Chronic Diseases Prediction,” *Front. Public Heal.*, vol. 10, 2022, doi: 10.3389/fpubh.2022.860396.

[76] P. D. Soerensen, H. Christensen, S. Gray Worsoe Laursen, C. Hardahl, I. Brandslund, and J. S. Madsen, “Using artificial intelligence in a primary care setting to identify patients at risk for cancer: a risk prediction model based on routine laboratory tests,” *Clin. Chem. Lab. Med.*, vol. 60, no. 12, 2022, doi: 10.1515/cclm-2021-1015.

[77] L. F. Rodrigues Moreira, R. Moreira, B. A. N. Travençolo, and A. R. Backes, “An Artificial Intelligence-as-a-Service Architecture for deep learning model embodiment on low-cost devices: A case study of COVID-19 diagnosis,” *Appl. Soft Comput.*, vol. 134, 2023, doi: 10.1016/j.asoc.2023.110014.

[78] K. H. Almotairi *et al.*, “Impact of Artificial Intelligence on COVID-19 Pandemic: A Survey of Image Processing, Tracking of Disease, Prediction of Outcomes, and Computational Medicine,” *Big Data and Cognitive Computing*, vol. 7, no. 1. 2023. doi: 10.3390/bdcc7010011.

[79] S. Kumar and M. Singh, “Breast Cancer Detection Based on Feature Selection Using Enhanced Grey Wolf Optimizer and Support Vector Machine Algorithms,” *Vietnam J. Comput. Sci.*, vol. 8, no. 2, 2021, doi: 10.1142/S219688882150007X.

[80] B. Sahu, A. Panigrahi, S. Mohanty, and S. S. Panigrahi, “A hybrid Cancer Classification Based on SVM Optimized by PSO and Reverse Firefly Algorithm,” *Int. J. Control Autom.*, vol. 13, no. 4, 2020.

[81] F. A. Alsarori, H. Kaya, J. Rahebi, D. E. Popescu, and D. J. Hemanth, “Cancer cell detection through histological nuclei images applying the hybrid combination of artificial bee colony and particle swarm optimization algorithms,” *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, 2020, doi: 10.2991/ijcis.d.200915.003.

[82] R. A. A. Saleh and R. Akay, “Classification of Melanoma Images Using Modified Teaching Learning Based Artificial Bee Colony,” *Eur. J. Sci. Technol.*, 2019, doi: 10.31590/ejosat.637846.

[83] P. Kora, A. Abraham, and K. Meenakshi, "Heart disease detection using hybrid of bacterial foraging and particle swarm optimization," *Evol. Syst.*, vol. 11, no. 1, 2020, doi: 10.1007/s12530-019-09312-6.

[84] K. Balasubramanian and N. P. Ananthamoorthy, "Improved adaptive neuro-fuzzy inference system based on modified glowworm swarm and differential evolution optimization algorithm for medical diagnosis," *Neural Comput. Appl.*, vol. 33, no. 13, 2021, doi: 10.1007/s00521-020-05507-0.

[85] S. Thawkar, "A hybrid model using teaching–learning-based optimization and Salp swarm algorithm for feature selection and classification in digital mammography," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 9, 2021, doi: 10.1007/s12652-020-02662-z.

[86] S. B. Gurav, K. V. Kulhalli, and V. V. Desai, "Fuzzy integrated salp swarm algorithm-based RideNN for prostate cancer detection using histopathology images," *Evol. Intell.*, vol. 15, no. 2, 2022, doi: 10.1007/s12065-020-00402-y.

[87] H. Li *et al.*, "A Hybrid Feature Selection Algorithm Based on a Discrete Artificial Bee Colony for Parkinson's Diagnosis," *ACM Trans. Internet Technol.*, vol. 21, no. 3, 2021, doi: 10.1145/3397161.

[88] Z. Ye, Z. Hu, X. Lai, and H. Chen, "Image segmentation using thresholding and swarm intelligence," *J. Softw.*, vol. 7, no. 5, 2012, doi: 10.4304/jsw.7.5.1074-1082.

[89] M. H. Horng, Y. L. Hong, Y. N. Sun, Z. Y. Zhan, and C. Y. Hong, "Learning one-class support vector machine by using artificial bee colony algorithm and its application for disease classification," in *ACM International Conference Proceeding Series*, 2019. doi: 10.1145/3343147.3343152.

[90] K. P. Baby Resma and M. S. Nair, "Multilevel thresholding for image segmentation using Krill Herd Optimization algorithm," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 33, no. 5, pp. 528–541, 2021, doi: 10.1016/j.jksuci.2018.04.007.

[91] S. Sumanth and A. Suresh, "Krill Herd Optimized Feature Selection for Classification of Alzheimer ' s Disease from MRI Images," vol. 7, pp. 729–734, 2018.

[92] L. F. Chen, C. T. Su, K. H. Chen, and P. C. Wang, "Particle swarm

optimization for feature selection with application in obstructive sleep apnea diagnosis," *Neural Comput. Appl.*, vol. 21, no. 8, 2012, doi: 10.1007/s00521-011-0632-4.

[93] T. Y. Tan, L. Zhang, and C. P. Lim, "Intelligent skin cancer diagnosis using improved particle swarm optimization and deep learning models," *Appl. Soft Comput. J.*, vol. 84, 2019, doi: 10.1016/j.asoc.2019.105725.

[94] M. Devendran and R. Sathya, *An approach for cancer classification using optimization driven deep learning*, vol. null. 2021.

[95] H. Ye *et al.*, "Diagnosing Coronavirus Disease 2019 (COVID-19): Efficient Harris Hawks-Inspired Fuzzy K-Nearest Neighbor Prediction Methods," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3052835.

[96] Y. Jin, G. Zhang, and J. Li, "Enhancing COVID-19 Diagnosis Through a Hybrid CNN and Gray Wolf Optimizer Framework," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, 2023, doi: 10.14569/IJACSA.2023.0140650.

[97] M. Jawarneh *et al.*, "Influence of grey wolf optimization feature selection on gradient boosting machine learning techniques for accurate detection of liver tumor," *SN Appl. Sci.*, vol. 5, no. 7, 2023, doi: 10.1007/s42452-023-05405-9.

[98] S. Dubey and M. Dixit, "Recent developments on computer aided systems for diagnosis of diabetic retinopathy: a review," *Multimed. Tools Appl.*, vol. 82, no. 10, 2023, doi: 10.1007/s11042-022-13841-9.

[99] "A Modified Grey Wolf Optimizer algorithm for feature selection to predict heart diseases," *Int. J. food Nutr. Sci.*, vol. 11, no. 12, 2023, doi: 10.48047/ijfans/v11/i12/180.

[100] H. Basak, R. Kundu, S. Chakraborty, and N. Das, "Cervical Cytology Classification Using PCA and GWO Enhanced Deep Features Selection," *SN Comput. Sci.*, vol. 2, no. 5, 2021, doi: 10.1007/s42979-021-00741-2.

[101] R. M. Ghoniem, "A Novel Bio-Inspired Deep Learning Approach for Liver Cancer Diagnosis," *Inf.*, vol. 11, no. 2, 2020, doi: 10.3390/info11020080.

[102] M. Siddhartha, V. Kumar, and R. Nath, "Early-stage diagnosis of chronic kidney disease using majority vote – Grey Wolf optimization (MV-GWO)," *Health Technol. (Berl).*, vol. 12, no. 1, 2022, doi: 10.1007/s12553-021-00617-8.

[103] Preeti and K. Deep, "A Modified Lévy Flight Grey Wolf Optimizer Feature Selection Approach to Breast Cancer Dataset," in *Lecture Notes in Networks and Systems*, 2023, vol. 547. doi: 10.1007/978-981-19-6525-8_31.

[104] S. M. Vijithananda *et al.*, "Feature extraction from MRI ADC images for brain tumor classification using machine learning techniques," *Biomed. Eng. Online*, vol. 21, no. 1, 2022, doi: 10.1186/s12938-022-01022-6.

[105] R. Abu Khurmaa, I. Aljarah, and A. Sharieh, "An intelligent feature selection approach based on moth flame optimization for medical diagnosis," *Neural Comput. Appl.*, vol. 33, no. 12, 2021, doi: 10.1007/s00521-020-05483-5.

[106] A. Majee, S. Biswas, S. Chatterjee, S. Sen, S. Mirjalili, and R. Sarkar, "Moth-Flame Optimization-Based Deep Feature Selection for Cardiovascular Disease Detection Using ECG Signal," in *Handbook of Moth-Flame Optimization Algorithm: Variants, Hybrids, Improvements, and Applications*, 2022. doi: 10.1201/9781003205326-10.

[107] H. A. Alsattar, A. A. Zaidan, and B. B. Zaidan, "Novel meta-heuristic bald eagle search optimisation algorithm," *Artif. Intell. Rev.*, vol. 53, no. 3, 2020, doi: 10.1007/s10462-019-09732-5.

[108] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," *Syst. Sci. Control Eng.*, vol. 8, no. 1, 2020, doi: 10.1080/21642583.2019.1708830.

[109] L. Abualigah, D. Yousri, M. Abd Elaziz, A. A. Ewees, M. A. A. Al-qaness, and A. H. Gandomi, "Aquila Optimizer: A novel meta-heuristic optimization algorithm," *Comput. Ind. Eng.*, vol. 157, 2021, doi: 10.1016/j.cie.2021.107250.

[110] W. Zhao, Z. Zhang, and L. Wang, "Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications," *Eng. Appl. Artif. Intell.*, vol. 87, 2020, doi: 10.1016/j.engappai.2019.103300.

[111] I. Benmessahel, K. Xie, and M. Chellal, "A new competitive multiverse optimization technique for solving single-objective and multiobjective problems," *Eng. Reports*, vol. 2, no. 3, Mar. 2020, doi: 10.1002/eng2.12124.

[112] Y. J. Zheng, "Water wave optimization: A new nature-inspired metaheuristic," *Comput. Oper. Res.*, vol. 55, pp. 1–11, 2015, doi: 10.1016/j.cor.2014.10.008.

[113] D. McGartland Rubio, "Alpha Reliability," *Encycl. Soc. Meas.*, pp. 59–63, Jan.

2004, doi: 10.1016/B0-12-369398-5/00395-9.

[114] Y. Wu *et al.*, "Demystifying Learning Rate Policies for High Accuracy Training of Deep Neural Networks," in *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 2019. doi: 10.1109/BigData47090.2019.9006104.

[115] F. C. Cruz, EAS, "Challenges of multivariable and multiclass classification problems," *I+D Rev. Investig.*, vol. 1, no. 1, 2013.

[116] O. Kherad, E. Bottequin, D. Steiner, A. Alibert, R. Eurin, and H. Bothorel, "Implementing a Multifaceted Intervention among Internal Medicine Residents with Audit and Educative Data Feedback Significantly Reduces Low-Value Care in Hospitalized Patients," *J. Clin. Med.*, vol. 11, no. 9, 2022, doi: 10.3390/jcm11092435.

[117] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble Learning for Disease Prediction: A Review," *Healthcare*, vol. 11, no. 12, 2023, doi: 10.3390/healthcare11121808.