Deep Learning Based Decision Support System

Author

Younas Khan

00000119548

MS-15(CSE)

Thesis Supervisor

Dr. Usman Qamar

Department of Computer and Software Engineering,

College of Electrical and Mechanical Engineering,

National University of Sciences and Technology, Islamabad

March 2019

بسم الله الرحمن الرحيم

In the name of God, most Gracious, most Compassionate

وَلَا يُحِيطُونَ بِشَيْءٍ مِّنْ عِلْمِهِۦٓ إِلَّا بِمَا شَآءَ ۚ

And they can’t encompass anything from His knowledge, but to extend He wills [2:255]

DEEP LEARNING BASED DECISION SUPPORT SYSTEM

Author
YOUNAS KHAN
00000119548

A thesis submitted in partial fulfillment of the requirements for the degree of
MS Computer Software Engineering

Thesis Supervisor:
DR. USMAN QAMAR

DEPARTMENT OF COMPUTER AND SOFTWARE ENGINEERING, COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING, NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,
ISLAMABAD
MARCH 2019

Language Correctness Certificate

This thesis has been read by an expert of the English language and is found free of typing, syntax, semantic, grammatical and spelling mistakes. The stated is also according to the format given by the university.

Signature of Student
Younas Khan
Registration Number
00000119548

Signature of Supervisor
Dr. Usman Qamar

Declaration

I certify that this research work titled "*Deep Learning Based Decision Support System"* is my own work. The work has not been presented elsewhere for assessment. Material derived from other sources has been properly acknowledged/referred.

Signature of Student
Younas Khan
MS-15(CSE)
00000119548

This page is intentionally left blank

Acknowledgements

I am extremely thankful to my ALLAH, The Almighty, for bestowing countless blessings upon me and vesting in me patience and ability to complete this manuscript. It is merely because of His kindness and assistance which enabled me to endure and succeed. I am very grateful to my supervisor **Dr. Usman Qamar** for his guidance and efforts. Throughout this journey, he has always been there to guide me and boost my confidence. He has been a source of great knowledge and his expertise made this possible. I want to thank my teachers **Dr. Saba Bashir** and **Dr. Farhan Hassan Khan** for their time, and ideas. Furthermore, I want to extend my heartiest gratitude to **Dr. Muhammad Abbas** and **Dr. Wasi Haider Butt** for their availability, guidance and support. I want to appreciate my ABBA Jan, who left no stone unturned to educate me. I want to thank my MOOR Jana for her prayers and love. I want to thank **Ma'am Mahvesh Khan, NNBS** and their sponsors for their financial support. I am grateful to my brothers, and friends **Muhammad Asad**, **Babar Zeb**, **Nazish Yousaf**, and **Syed Ikram Shah** for motivating and encouraging me.

Dedicated to my Beautiful Wife

Abstract

Heart diseases are amongst the leading causes of death worldwide. Death becomes certain if heart diseases are not diagnosed in the early stages. It has been causing death irrespective of age, gender or any other demographics. That is why diagnosing heart disease is extremely important. The most commonly used clinical method for diagnosing heart disease is angiography which is a very expensive procedure. Studies have also suggested that it has negative side effects. Lately, a lot of research has been conducted based on the use of machine learning for heart disease diagnosis. This thesis has also conducted a systematic literature review in order to thoroughly analyze the existing literature and look for gaps in it. The results of the review reveal that the most popular classification techniques are Support Vector Machine, Neural Networks, and ensemble classifiers. The main objective of this thesis is to present a technique that can be used to develop a decision support system or a computer-aided diagnosis system for heart disease particularly coronary artery disease. Studies reveal that the performance of neural networks can be increased by systematically initializing attribute weights instead of random weights initialization. In order to attain a reliable methodology, four feature selection i.e. weights by Support Vector Machine, weights by Principle Component Analysis, weights by Gini Index and weights by Information Gain and four weight optimization techniques i.e. Forward, Backward, Particle Swarm Optimization and Evolution Strategy have initially been used to provide optimized attribute weights in order to improve the performance of artificial neural network. The results of the initial experimentation are promising. Later, an ensemble is formed by getting an average attributes weights of three weight optimization techniques i.e. PSO, ES and backward weight assignment. The average weights are then provided to the input layer of the neural network. The accuracy attained by the proposed system is over 94%, which is promising. In future, the proposed technique can be used to form a reliable and assistive system which can be used as a diagnostic tool in order to add clinicians and physicians.

Table of Contents

List of Equations

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

Every passing day brings advancement in information technology, and the emphasis of automating systems has been increasing simultaneously. Irrespective of domain, manual systems have been automated. And when it comes to the medical domain, systems have rightly been automated. What is noteworthy is the fact that in the domain of healthcare, two factors are of the most importance i.e. accuracy and speed. Because in healthcare systems, a person's life actually relies on systems, therefore the room for error should tend to none. When it comes to manual systems, they are not hectic, and time-consuming but often experts whose time is extremely expensive are kept busy doing mundane tasks. If these tasks are automated, it will not only save time of the experts but will assist in improving accuracy as well. As a matter of fact, today an entire organization can be automated with the help of the internet, computers and modern-day technologies. For instance, manually storing a record of sales of a store is extremely difficult to manage, the same when done with a PoS or Point of Sale becomes a piece of cake. The PoS can be used to save/store and process loads of sales and inventory data without any difficulty. Likewise, in healthcare detecting patterns, identifying, extracting, and selecting features/attributes, and diagnosing diseases when done manually is an extremely difficult, time-consuming and costly.

As far as research related to the automation of medicine or healthcare is concerned, it has been emphasized particularly in the last 3 decades. Systems have been developed to effectively assist clinicians and physicians. However, there are so many challenges when we talk about automating medical systems. Medical records usually comprise sensitive details, and a mere small mistake can bring hazardous or even life-threatening effects.

Medical records more often than not are unstructured, but machine learning has been evolved enough to deal with this challenge with numerous data pre-processing techniques. Then there is the constraint of time, automated systems should be as fast as possible especially when they are used for healthcare due to the fact that valuable human life depends on them. This constraint is normally associated with redundant features and can be tackled with feature selection and feature reduction techniques.

1.1Overview

With the passage of time heart diseases have taken more lives than any other disease. The problem is not limited to a particular gender, age, or any other demographic characteristics. This is why the accurate prediction, preferably, at an early stage is extremely vital in order to save someone's life [36]. In the literature on machine learning, a stupendous amount of work has already been done to diagnose heart and other diseases. However, what is mind-boggling is that an agreement of which particular technique is best for which kind of data or problem is yet to be established. But what is agreed upon is that feature selection, noise reduction, and techniques of optimization are really helpful in enhancing the accuracy and efficiency of the existing techniques [37]. That is one of the reasons why this research study has been conducted.

1.2Background and Motivation

Heart diseases are one of the most fatal diseases and have been causing death all around the world. It affects people of every age, gender, and profession, and many factors are responsible to cause heart diseases. If not diagnosed timely it can cause death in a matter of months if not days. Angiography is used most commonly for diagnosing it, but studies show that it has side effects and is costly as well. Moreover, numerous studies and research works have been carried out in order to diagnose heart disease and to a great extent, scientists have achieved success. However, there is still room for improvement, and until 100% accuracy is achieved the struggle for diagnosing and treating diseases would continue.

The literature has a huge amount of work on decision support systems or computer-aided diagnosis systems for healthcare. There also exists a decent amount of work on heart disease diagnosis but very few studies have tried to enhance the performance of artificial neural networks by initializing the input attribute weights with optimized weights. This the reason why the need for exploring this factor has been felt. Moreover, the fact that computer-aided systems can be used to assist physicians, clinicians and other experts for saving a life is another motivating factor of this thesis.

1.3 Objective and Contributions

Heart failure has been causing deaths irrespective of age, gender or any other demographics, around the globe. When heart diseases are not diagnosed timely, heart failure and consequently death becomes certain. The existing clinical techniques for instance: angiography is used for detecting CAD, but it has reportedly side effects and it is a very expensive procedure. Modernized hospitals often contain patients' data, that data can be used by machine learning in order to predict or classify patients. The main objective of this thesis is to present a technique which can be used to assist physicians, clinicians and other medical experts to diagnose heart diseases, particularly, coronary artery disease in order to save valuable lives.

A heart disease dataset i.e. Z-Alizadeh Sani Heart Disease Dataset has been used in this study as the main dataset. Different feature selection and weight optimization techniques are applied to the stated dataset for experimentation purposes. The final algorithm has been established, where the main contribution of this thesis is initializing the weights of attributes with an ensemble that is based on Particle Swarm Optimization, Evolution Strategy, and Backward weight optimization, in the input layer of an artificial neural network. The performance of an artificial neural network has been enhanced with this technique. The research conducted in this thesis has been able to achieve a classification accuracy of over 94%.

1.4 Outline

Apart from this chapter, this manuscript has six more chapters. Chapter II is used to present the related terminology, it gives the reader some required background knowledge about computer-aided diagnosis systems, machine learning and the use of machine learning for healthcare and diagnosis systems. Chapter II also contains a systematically conducted literature review, which is used to analyze the existing similar techniques, and gaps in the existing literature. Since numerous feature selection and weight optimization techniques have first been trialed in order to thrive for the best possible system, the manuscript has presented a rough methodology in Chapter III, and different experiments have been performed in the same chapter. Chapter IV proposes the final methodology and its implementation, Chapter V has the results, comparisons, and analysis and lastly, Chapter VI discusses the conclusion and future work of this thesis.

1.5Summary

This is the first and introductory chapter of the thesis, it is used to provide the reader with an idea about the significance of information technology and particularly machine learning for decision support systems and computer-aided diagnosis systems. It explains how technology is used for assisting and aiding physicians, clinicians, and other medical experts. What challenges are faced and how are these challenges coped with. Moreover, an overview of the thesis, its background and motivation, its objective and contributions and its outline have also been presented in this introductory chapter.

CHAPTER 2

LITERATURE REVIEW

2.1Introduction

Getting an insight into previous and related literature is a compulsory feature of any project be that academic or industrial. What makes a literature review effective is the fact that it should create a solid base in order to advance knowledge. It plays a vital role in the facilitation of theory development, shuts areas where an excess of research presides and explores areas where there is need of research [1].

However, studies have proven that the conventional literature review has issues and more often than not the results attained might be misleading, and one of the foremost tools that are used for supporting evidence-based artifacts in most domains is the development of Systematic Literature Review. SLRs are made use of for aggregating the experiments from a defined range of various studies for answering a set of research questions. Reviews of the stated kinds employ some carefully defined criteria for determining which research papers are to be taken into consideration and analyzed [2].

In software engineering, the number of empirical studies has grown rapidly, and as a consequence of that systematic procedures should be placed in order to assess and aggregate the outcomes of research for providing an in-depth summary of a specific topic [3].

Keeping the importance of SLR in mind, this thesis has implemented its literature review in a systematic way. It has a predefined inclusion and exclusion criteria, i.e. existing studies that have been selected and analyzed must match the predefined criteria. These criteria are formed on the basis of publishers, and some quality assessment criteria.

Keeping the importance of SLR in mind, this thesis has implemented its literature review in a systematic way. It has a predefined inclusion and exclusion criteria, i.e. existing studies that have

been selected and analyzed must match the predefined criteria. These criteria are formed on the basis of publishers, and some quality assessment criteria.

Before we could begin the discussion of our main systematic literature review, there are certain factors or areas which need to be illustrated, for instance: decision support systems and their types, how are decision support systems used for healthcare and the benefit of its application, machine learning, and the application of machine learning in order to develop decision support systems or computer-aided diagnosis systems.

2.2Decision Support Systems

A computer decision support system or simply a DSS serves the purpose of assisting to determine and evaluate different alternative actions. It relies on data, normally on a huge chunk of data, which is provided to it, it then analyzes that data with the help of statistical procedures and equipment in order to derive meaningful data from it, it then tapers down the alternatives or the available choices of decisions by making use of decision theories. It does not replace a manager, or in some cases a manager or generally speaking a human being but assists them in decision making [4].

The evolution of DSS began in the time of distributed computing, these systems began around 1965. It is significant to keep a formal record of technologies, ideas, systems, and people involved in such a crucial area of applied IT. Literature is important because it provides an insight of history which not only serves the purpose of guiding future activities but also has the record of the actions and ideas of how ideas and practices are advanced. DSS can be termed as a diverse field and that is why its history is not linear and neat, and it has been perceived differently and from various vantage points [5].

2.2.1 Types of Decision Support Systems

DSS can be categorized in numerous groups, and amongst the most common ones, a few have been enlisted as follows.

2.2.1.1 Data-Driven Decision Support Systems

These kinds of DSS rely on systems of file drawing, data analyzing, the information analyzing, and data warehousing. These systems emphasize access to and on the manipulation of huge databases of structured data.

2.2.1.2 Model-Driven Decision Support Systems

As the name suggests model-driven DSS can be formed from different areas or disciplines of particularity and include models of accounting, finances, representations, and optimization etc. These kinds of systems emphasize more on accessing and manipulating models instead of data i.e. it makes of use data and parameters in order to assist decision making to analyze a particular case. What is noteworthy is that these systems do not rely on large databases as they are not consequently linked to the stated.

2.2.1.3 Knowledge-Driven Decision Support Systems

Knowledge-driven DSS are used for providing recommending or suggesting some schemes which assist users for the selection of a suitable alternative to the faced problem. These systems are often termed as intelligent DSS or management expert systems. These systems rely on knowledge and on the basis of a particular knowledge base they suggest a course of actions to the concerned personnel. Furthermore, these systems are equipped with special problem-solving tools and are often related to data mining, for instance, examining a great amount of data in order to present content relationships.

2.2.1.4 Document-Driven Decision Support Systems

With the help of document-driven DSS managers can retrieve and manage unstructured webpages and documents by merging various processing and storage technologies in order to get comprehensive retrieval and analysis of the document. These are also used for accessing documents, meeting minutes, corporate records and important correspondence etc. Such systems are usually run with the help of task-specific search engines.

2.2.1.5 Communication-Driven Decision Support Systems

They are known as GDSS or Grouped Decision Support Systems. They are different from the rest of the DSS as they are a hybrid DSS as they emphasize using communications and decision models for facilitating solutions of problems by grouping decision makers. Besides communication, Grouped Based Decision Support Systems also supports document sharing, scheduling, and other group productivities, in order to enhance decision making.

2.2.1.6 Combination or Hybrids of Decision Support Systems

- Combinations formed by making use of multiple types of DSS are termed as hybrid DSS. The most common amongst which is the Web-Based DSS, which is driven by formed by multiple models such as communication-driven, document-driven and knowledge-driven DSS. These are computerized systems which are used to deliver decision-assisting information or which work as decision assistive techniques for managers or analysts.
- OLAP or On-Line Analytical Processing is a group of software which is used to enable managers, analysts, and executives to acquire an insight of data via consistent, speedy and interactive access to a vast variety of possible perspectives of information. Initially, the systems are used to transform raw data into information in order to present the actual dimensionalities of an enterprise. Tools of OLAP are used to structure a hierarchy of data in such a way that a manager would perceive their enterprise. Moreover, these systems are also utilized to enable analysts to rotate data, which ultimately change the relationships and provide a detailed view of the corporate information [6].

The described hybrid systems are the most common and popular DSS as far as businesses and enterprises are concerned. While keeping a track of the different classifications and distinctions of DSS, a DSS ought to be illustrated in terms of the underlying model of the system, the targeted and/or potential users and a defined goal or purpose.

2.3Decision Support Systems for Healthcare

We now move towards the main theme of our study which is a decision support system for healthcare, which is also known as a clinical decision support system. It is used to provide clinicians, patients, staff or other people with people-specific information. The information is filtered intelligently and is presented at a suitable time in order to improve health or healthcare. Clinical Decision Systems are used to encompass various tools in order to improve decision-making in clinical environments. These techniques and tools include computerized reminders and alerts to both patients and clinicians; condition-particular order sets and guidelines; reports and summaries of patients, templates of documentation, support of diagnosis, and reference information that are contextually relevant, and many more services.

Furthermore, clinical decision support can be any sort of tool which is used to provide patients, administrative staff, clinicians, nurses, caregivers, or any other such personnel with filtered or targeted information. These systems are designed for improving the quality of care, preventing or avoiding errors, and ultimately enabling the care-team to improve their efficiency [7].

These systems can also be used to significantly increase the safety, quality, effectiveness, and efficiency of healthcare. The ONC or Office of the National Coordinator for Health Information Technology provisions determinations for developing, adopting, implementing, and evaluating the use of these systems in order to improve decision making concerned with healthcare [8].

Following are a few examples of Decision Support Systems for Healthcare

2.3.1 An Automatic Disease Diagnosis Method

The automation of medical diagnosis on the basis of enormous data can be seen as extremely important to doctor-patient contradiction regulation and the improvement of healthcare facilities. The study under discussion illustrates a disease diagnosis system that is both intelligent and automatic. On the foundation of an enormous number of established cases of diseases, the methodologies of probability and stats are made use of for finding relationships amongst symptoms and various diseases. The established sets of symptoms of various diseases have been acquired and later the diagnoses are provided stating the number of symptoms given by the patients and the established set of symptoms of various diseases. The system has been deployed beforehand and it proves that it can be of great value if applied for clinical purposes [9].

The technique has two basic parts, firstly the attainment of an established set of symptoms and secondly the particular method of diagnosis. The diseases successfully diagnosed by this system are nasopharyngitis and rhinitis [9].

2.3.2 Evidence-Based Health Care System

One area that now has been remaining in the limelight is healthcare analysis, but having said that it still postures consistent challenges, and that is the reason why a great amount of money has been spent on it by many countries. The Institute of Medicine illustrates many propositions in order to ameliorate healthcare quality. Various algorithms of data mining have been used on the stupendous amount of healthcare data and records in order to assist the process of decision making [10].

This particular system analyzes the health conditions of patients and diagnoses the exact disease by making use of the Evidence-Based Methodology and Big Data Analytics. Two things can be considered for disease diagnosis and suggesting drugs i.e. the history of patients and their current evidence. Whereas the report of patient analysis is produced for monitoring their recovery and the suggested drug's feedback is noted to stand-for the effectiveness of the process of diagnosis [10].

In this system, the sample dataset has a medical history which has been collected from 1,000 patients. The feedback that has been acquired from the patients is made use of for measuring the efficiency of the system. What is interesting is that machine learning has been utilized for analyzing and identifying the most suitable drug for patients having particular symptoms or clinical evidence [10].

With the availability of a dataset and the feedback acquired from the patients, the system can diagnose or predict disease and also suggest the drugs to the patients. Based on the provided symptoms, the system is able to diagnose Appendicitis and suggest different drugs to different patients. As far as the success rate of the recommender is concerned, it is around 85% which is an indication of the fact that this system is capable of processing an enormous amount of patients' medical records and also of suggesting the most appropriate drug based on the conditions of the patient. The system's results, however, should be cross-checked with the physicians' prescription [10].

2.3.3 CAD System for Automatic Diagnosis of Neurological Diseases

The diagnosis of neurological disease has been increasing over the course of time and now it has become a significant concern, as it can be counted amongst one the biggest and the toughest challenges in modern medicine and healthcare. According to the recent reports of WHO or World Health Organization, stroke to headache. Alzheimer disease and epilepsy influence around one billion people around the globe. A rough estimate states that around 7 million deaths occur because of neurological disorders, every year. Currently, magnetic resonance imaging and electroencephalogram are used as the diagnosis techniques, which produce a stupendous amount of data (both in dimension and in size) for detecting, monitoring and treating these diseases. Generally, the analysis of the stated amount of medical data is manually done by experts for identifying and understanding abnormalities. However, accumulating, managing, analyzing, and assimilating data of such great amount for humans is an extremely difficult and time-consuming task. That is the reason why decision support systems or more particularly computer-aided diagnosis plays an important role by automatically detecting abnormalities that neurological by making use of big data. Over the past few decades, a lot of research has been conducted for developing computer-aided diagnosis systems for managing big data in order to assess diagnosis. The system proposed in this particular study can improve diagnosis consistency and it can also increase the success rate of treatments, it can be used to save valuable lives and to reduce both the amount of time and money required [11].

This Computer Aided Diagnosis system has three major steps i.e. preprocessing, feature selection, and classification. In the first step, the acquired data that is a medical signal or image data are processed, and noise is removed from it. Noise removal comes in really handy for reducing the complexity and it also decreases the computation time of the algorithm. The second part i.e. feature selection/extraction is used to extract the biomarkers of disease recognition. Whereas, in the last step i.e. classification a vector based on the extracted features is made use of as an input, in the classification model, in order to assign a candidate to one of the two categories. In general cases, a computer-aided diagnosis system is of two types i.e. two-class classification system, where the candidates are either classified as normal or abnormal, more than two class classification system or multi-class classification, where the system is able to classify a candidate into numerous classes [11].

For assessment purposes, k-fold validation assessment can be used, there are other options such as leave one out and bootstrap methods as well. FROC or Free Response Receiver Operating Characteristic and ROC curves have been used for evaluating the general performance of the computer-aided diagnosis systems, on different points of operation. FROC is used to depict relationships amongst false positives and sensitivity. This is attained by placing thresholds on a particular parameter of the system or even on result or output of the classification model [11].

Lately, a decent amount of research has been conducted in order to develop computer-aided diagnosis systems to detect neurological issues for instance dementia, epileptic seizures, autism, Alzheimer disease, brain tumors, strokes, and sleeping disorders [11].

2.4Machine Learning

Learning is the procedure of developing a model after knowledge is covered from data, while machine learning is the complex computation procedure of automatically recognizing patterns and intelligent decision making on the basis of trained data samples. Machine learning falls under the umbrella of artificial intelligence; it is the ability of a machine to learn from a large set of data and predict, cluster or classify similar but unseen or new data based on its learning or training. Some famous machine learning techniques include Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree, self-organization map, and k-means clustering etc. There are ensemble approaches as well which integrate the outcomes of individual classification techniques and produce an overall better performance.

2.5Machine Learning for Healthcare Decision Support Systems

In this subsection, we discuss the different uses of machine learning for aiding healthcare decision support systems.

2.5.1 Data-Driven Based Approach to Aid Parkinson's Disease Diagnosis

The methodology proposed in this particular study is based on machine learning and is used to diagnose Parkinson's disease. It uses VGRFs or Vertical Ground Reaction Forces data acquired from the gait cycle. Just like any other computer-aided diagnosis system, this one too has a classification model which is used to classify patients either Parkinsonian or normal. The process of diagnosis has four major parts; preprocessing of data, extraction, and selection of features, classification of data and evaluation of performance. Feature selection is done with the help of a wrapper technique attained by utilizing a random forest algorithm. Those features which are extracted and then selected are made use of as inputs for every classifier. The methodology presented not only uses supervised classification techniques but also unsupervised classification techniques. Techniques derived from the former method include: decision trees, K-nearest neighbor, Naïve Bayes, random forest, and support vector machines while those derived from the latter include: GMM or Gaussian Mixture Model, and K-means. In order to assess the effectiveness of the presented technique, a dataset acquired from three different research papers has been used. As stated earlier this set of data has vGRF measurements attained from 8 force sensors kept under both feet of the patients. 93 patients who are Parkinsonian and 72 are normal, the stated subjects have been successfully categorized [12].

As far as the performance of the system is concerned, it is measured on the basis of many parameters for instance precision, accuracy, F-measure and recall. The evaluation of classification performance has been carried out with the help of cross-validation, the leave one out to be specific. It is clear from the results that the proposed system has the capability to differentiate between healthy and Parkinsonian disease patients/subjects with really good accuracy. In order to validate the system, the system has also been evaluated on another dataset which consists of patients neurodegenerative diseases such as HD or Huntington's disease and ALS or Amyotrophic Lateral Sclerosis. And the effectiveness of the system with the new dataset prove its effectiveness, as they are able to differentiate amongst PD patients from patients having other neurological disorders [12].

2.5.2 CAD of Alzheimer's type dementia combining support vector machines and discriminant set of features

Alzheimer's disease is one of the most common causes of dementia as far as older people are concerned. According to research, it influences about 30 million people worldwide. It has also been predicted that in the course of the coming 50 years this number will be tripled. At the same time diagnosing AD in the earliest stages seems to be an extremely difficult challenge [13].

SPECT or more commonly known as Single Photon Emission Computed Tomography and PET or Positron Emission Tomography are the most commonly used diagnostic procedures. Simultaneously, these conventional techniques do rely on reorientation that is manual, and also on semi-quantitative analysis of particular ROIs (Region of Interests) and on a visual reading of tomographic slices. The stated steps are extremely time-consuming, prone to error and subjective [13].

That is why the need for automatic systems can be felt, and an automatic CAD or computer-aided diagnosis has been proposed in this particular study in order to detect AD in the earliest stages. The technique presented has been founded on image parameter selection and SVM or support vector machines classification. For dealing with ROIs, a study has been conducted and image parameters that are the most discriminant ones are recognized in order to reduce input space dimensionality and to improve the system's accuracy. As stated earlier, features are evaluated and amongst all the evaluated features, and those which have actually reduced the input space dimensionality and improved the accuracy of diagnosis (when RBF SVM is used) are sagittal correlation parameters and coronal standard deviation. The system developed on the stated technique has been able to achieve a reliable accuracy of over 90% as far as the early diagnosis of AD is concerned. The stated system is also stated to have outperformed the previous techniques [13].

The system has made use of machine learning tools and techniques for developing an entirely automated CAD system in order to enhance the early diagnosis of AD. It has used support vector machine for classification and image parameter selection for feature extraction. It is developed by hunting for the most differentiating inputs set in the first and second sagittal statistics, transversal and coronal portions of the brain. The sagittal correlation and coronal standard deviation are stated to be the most differentiating AD image parameters. The FDR based feature selection is not only able to reduce input space dimensionality but has also led to a two coefficient vector which yielded a high discrimination accuracy particularly when it is used with an RBF kernel. This is the reason why the presented CAD system yields a phenomenal accuracy of over 90%. As far as the sensitivity and specificity of the system are concerned, they are 93.10% and 87%, respectively. Since these mindboggling figures have been achieved in the early detection of AD, it can be said that a reliable system has been developed because of machine learning [13].

2.5.3 Improve CAD with ML Techniques Using Undiagnosed Samples

Applying machine learning techniques in computer-aided diagnosis is a popular practice, where the objective more often than not is to learn a hypothesis from samples that are diagnosed, which at the end of the day are used to help medical personnel to make a diagnosis. Machine learning relies on a good amount of data and likewise, in order to learn a good hypothesis a huge amount of diagnosed samples are needed. Despite the fact that the stated samples can easily be acquired from mundane medical checkups, usually, it is not possible for the medical experts to come up with a diagnosis for every sample. Furthermore, this heavy burden can be released from the medical experts if a given hypothesis can be learned with the help of a great number of undiagnosed samples [14].

The system brought forth in this particular study is founded on a novel algorithm based on semi-supervised learning termed as Co-forest. It proposes to make use of an ensemble called RF or more commonly known as Random Forest in order to extend the co-training paradigm. The ensemble is used to enable Co-forest for estimating the labeling confidence as far as undiagnosed samples are concerned and also for easily producing the concluding hypothesis. Benchmark datasets have been used for verifying the effectiveness of the presented technique. A successful application on the diagnosis of breast cancer of microcalcification and scenarios and cases from three medical datasets reveal that undiagnosed samples are actually assistive when it comes to developing a computer-aided diagnosis system. Moreover, Co-forest is capable of enhancing the performance of the hypothesis which has been learned with the help of a few diagnosed samples [14].

In computer-aided diagnosis systems, the process of diagnosing samples by training it may rely on medical experts and it may also place a heavy burden of mundane tasks on them. This burden can be lifted if the algorithm used for training can use data that is unlabeled. In this particular system, a novel technique called Co-forest has been presented which can make use of undiagnosed samples for boosting the system's performance which got trained with diagnosed samples. What this technique does is that it extends the paradigm of co-training by exploiting the capabilities of Random Forest in order to tackle the issue of opting for confident undiagnosed samples for labeling and for producing the final hypothesis. The effectiveness of Co-forest has been verified as the system has been trialed on UCI datasets. A successful application on the diagnosis of breast cancer of microcalcification and scenarios and cases from three medical datasets reveal that undiagnosed samples are actually assistive when it comes to developing a computer-aided diagnosis system. Furthermore, Co-forest is capable of enhancing the performance of the hypothesis which has been learned with the help of a few diagnosed samples [14].

However, the system leaves behind some research gap, the Co-forest is stated to have been tending toward the underestimation of error-rates of the associated ensembles that is why the appropriate error-rates of the ensembles have not been estimated properly. Moreover, Co-forest's performance can be further enhanced if Query by Committee is incorporated with it. Incorporating the stated would help in the attainment of assistive information from the experts on particular samples [14].

2.5.4 A CAD System for Thyroid Disease Using ELM

In this particular case, ELM or Extreme Learning Machine and PCA or Principle Component Analysis have been used for developing a computer-aided diagnosis system that can be used for detecting thyroid disease. This computer-aided diagnosis system consists of three major portions. The focus of the first stage is to reduce dimensionality and for that, it uses PCA and constructs a derived set of features that is termed as the most discriminative one. As far as the second portion is concerned, it targets model construction. In which an ELM classifier is reconnoitered for training a predictive model that is optimal in other words which have optimized parameters. The performance of ELM mainly relies on the number of hidden neurons, and that is why this particular study has proposed a method of experimentally hunting for the best number. The last portion deals with the diagnosis of thyroid disease, for the stated purpose the attained model of ELM is used. It uses the new features set and the optimized parameters. This computer-aided diagnosis system has been named PCA-ELM, and its effectiveness has been estimated rigorously on UCI machine learning repository's thyroid disease dataset [15].

The system's accuracy has also been compared with other relevant systems. In order to validate or test the system, a 10 fold cross validation has been used. The results obtained from experimentation reveal that PCA-ELM has outperformed all the existing systems. As PCA-ELM has achieved a mindboggling accuracy of 98%. What is also noteworthy is that PCA-ELM works really fast, it has also outclassed computer-aided diagnosis systems based on support vector machine. So, it can be concluded that the proposed system should be trialed as a novel tool for detecting thyroid disease, as it not only works faster than the existing systems but also performs better [15].

2.5.5 CAD of ECG data on the least square support vector machine

The technique or methodology proposed in this particular research is based on LSSVM or Least Square Support Vector Machine and is successful in classifying arrhythmia, it has been applied on a dataset of ECG signals. The main objective of this system is to differentiate between diseased subjects (having arrhythmia) and healthy subjects, for the stated purpose it applies LSSVM to a dataset of ECG signals. In order to validate the system, this research has used three different split techniques, 50-50 split i.e. 50% for training the system and 50% for testing it, 70% for training and 30% for testing, and 80% for training and 20% for testing. In order to test the performance of the system, the study has not only considered classification accuracy but has also analyzed specificity, sensitivity and ROC curves. The system is said to have achieved the perfect accuracy of 100% on all the split sets. Based on the analysis of the results it is clear that this system has outclassed all the previous computer-aided diagnosis systems, as far as the classification of arrhythmia patients is concerned. Furthermore, it can also be inferred from the results that this system should be used as a novel intelligent system for assisting the process of diagnosis [16].

When it comes to classifying arrhythmia, the structure of LSSVM that has been built has delivered extremely promising figures. The study in no way claims that this system should be used to replace the existing devices but suggests that this technique should be used in complementation with the existing devices. So, that physician has absolutely no doubt when diagnosing arrhythmia patients [16].

2.6A Systematic Literature Review of Machine Learning Techniques Used for Heart Disease Diagnosis

One of the leading causes of death all around the world is heart failure, it is seen as a major disease in old, and middle ages. Coronary Artery Disease (CAD) particularly, is termed a widespread cardiovascular illness having high rates of mortality. Clinical diagnosis techniques, such as angiography, are recommended by physicians as they call it the best diagnosis for CAD. On the contrary, it is very costly and possesses side effects. In order to present alternatives to such clinical diagnosis, much research has been conducted in the field of machine learning [17]. Heart disease has gradually become a communal problem of public health, around the globe, it is because of unawareness, unhealthy consumption, and poor lifestyle. Today, its accurate prediction and diagnosis are great challenges for hospitals and practitioners. The development of computing technologies has assisted healthcare facilities in the collection and storage of data for founding clinical decision making. Hospitals, located in many developed countries, collect and store patients' data in the digital and manageable form [52]. Coronary Heart Disease (CHD) is a proven significant public health problem not only in some parts of the world but globally.

Machine Learning (ML) is used to interpret datasets by using computers that acquire knowledge from experiences. ML for health-informatics has emerged as an interdisciplinary science of dealing with healthcare data using complex computational techniques [53]. More often than not the healthcare datasets are colossal, that is where data mining comes into play. Data mining transforms the huge sets of data into information which are later used to make better predictions and decisions [54].

Numerous research studies have been carried out to discover which ML techniques have been used for diagnosing heart disease. For instance [55], which investigates and compares various data mining classification procedures, and ensembles. ML techniques can ameliorate the results of cardiac arrest forecast, but, there is a further need for research to improve the prediction and generalization of ML techniques [56].

In this part of the manuscript, we discuss a survey that presents research studies focusing on ML techniques applied to heart disease datasets and which have been published since 2012. The study intends to find gaps in the existing literature and suggests that a general solution for various healthcare problems is needed to be proposed. There are a few studies which seem to have done a similar job, however, there is no work done particularly on heart disease datasets. Hence, we feel the need of conducting thorough research to objectively and critically analyze past papers. This survey provides the current gaps in research and is helpful for researchers who want to apply machine learning and data sciences techniques for diagnosing heart failure.

2.6.1 Methodology

The research process for this SLR has been carried out in a systematic manner. The collection and selection processes of papers have been done in a way which meets predefined criteria. Four keywords have been used to narrow down the search process i.e. Heart failure, Heart Diseases, Risk Prediction, and Neural Network. The stated keywords enable us to select from a reduced number of researches.

2.6.1.1 Quality Assessment Criteria

In order to ensure that only quality research papers are considered for the survey, some quality assessment criteria have been defined. All the selected papers have met these defined criteria.

a) Publishers

The publishers that are taken into account for this survey are Elsevier, IEEE, and ACM. Figure 1 presents an overview of the papers selected from the stated publishers and the details of individual research studies for each publisher is given in Table 1.

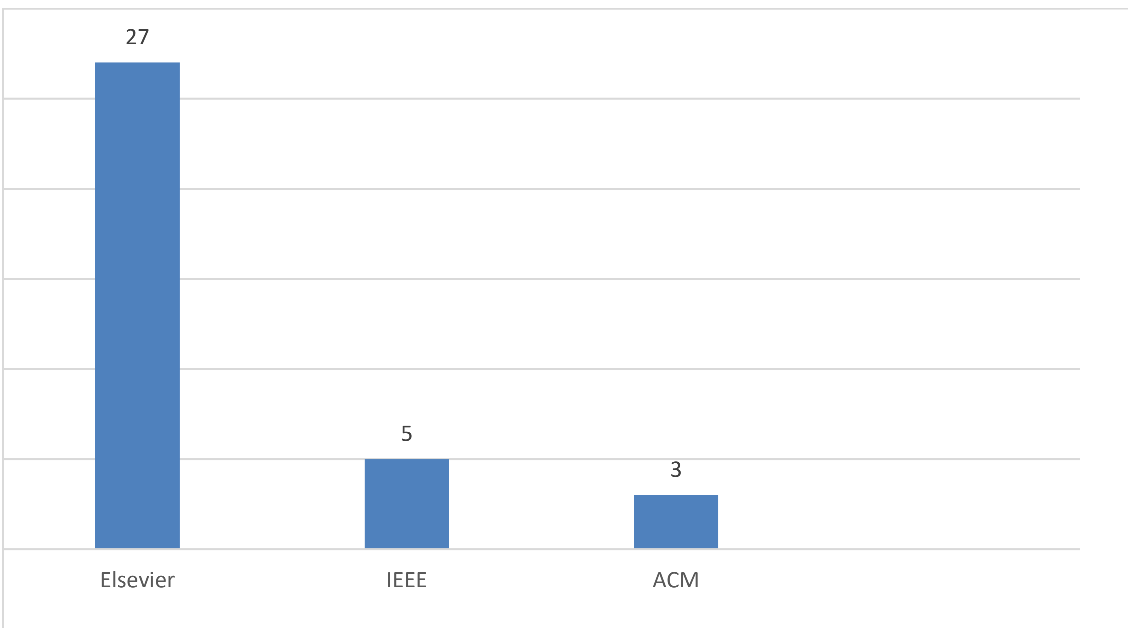

Figure 1: Selected researches per publisher

Table 1: Details of Research Studies per Publisher

Publisher	Selected Research Studies	No. of Researches
Elsevier	[17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [36] [37] [38] [39] [41] [42] [43] [44] [45]	27
IEEE	[40] [46] [47] [48] [51]	5
ACM	[35] [49] [50]	3
Total		*35*

b) Presentation of an Effective Technique

One of the main concerns about the selected papers is that should have proposed a model or a technique for detecting heart disease. Papers which do not meet this criterion have been excluded from the results of the search.

c) Validation of Results

An assessment of the results presented or claimed in the research papers should be depicted on a heart disease dataset.

d) Repetition

It is necessary for a paper to be select to have proposed and validated a novel methodology. In other words, research papers which presented similar models or methodologies have been excluded.

e) Recent Papers

In order to make sure that the latest papers are selected, only journal papers from 2012 to present have been selected for this survey as shown in Figure 2. At the end of the selection process, a total of 35 journal papers have been opted for.

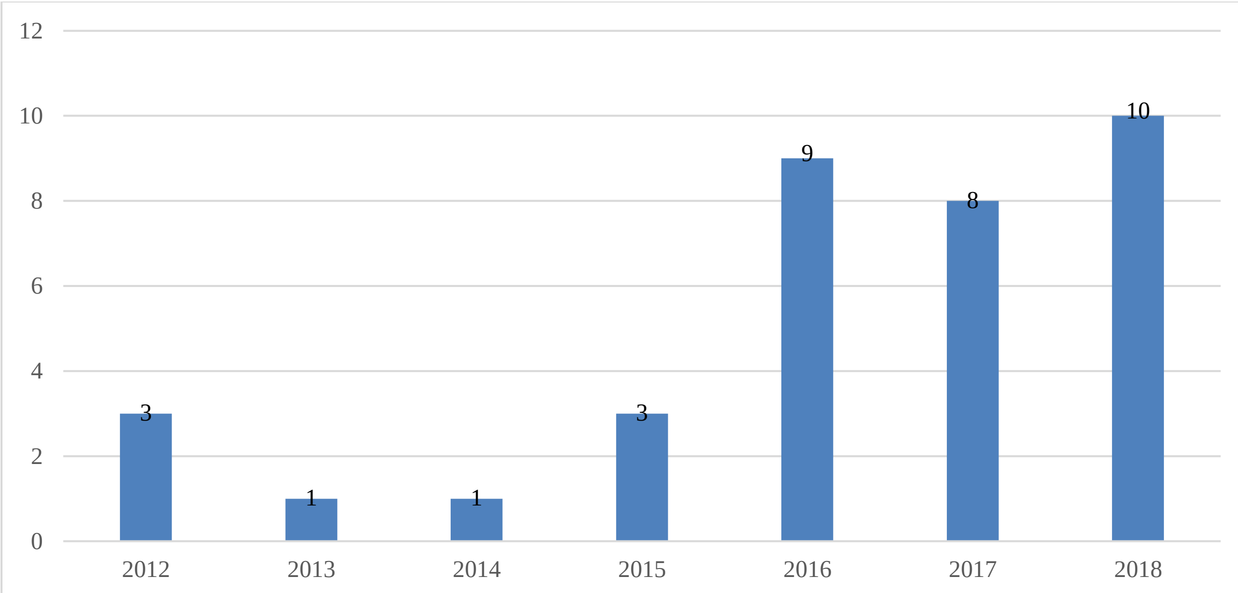

Figure 2: Selected researches per year

2.6.2 Related Work

In this section, the main findings of the selected papers have been discussed after analyzing them. It presents a brief of the methodologies brought forth in the selected papers. The research papers have been categorized on the basis of the classification techniques. Some of the selected papers have applied their techniques to various datasets, these datasets include, Diabetes datasets, Heart disease datasets, Breast Cancer datasets, Liver Disease, and Hepatitis datasets. However, in this survey those papers have been selected which implement their techniques on heart disease datasets, for instance: UCI, PHP, BIDMC CHF dataset, PTB Diagnostic ECG and others.

The following techniques are found to be the most popularly used for heart disease and HF classification.

2.6.2.1 Support Vector Machine (SVM)

Sung and Yuan Lee make use of Genetic Algorithm (GA) for feature selection and SVM for classification. The results of this technique validate its effectiveness. When SVM is applied without GA, it outclasses two other techniques in the literature by producing an accuracy of 96.38%, when GA is applied the accuracy is improved by 3.14% [17].

A.D. Dolatabadi et al., [18] proposes a technique for automatic diagnosis of Coronary Artery Disease (CAD) and normal conditions by making use of Heart Rate Variability (HRV) signals that are derived from an electrocardiogram (ECG). The technique makes use of Principal Component Analysis (PCA) for dimensionality reduction and then applies SVM for classification purpose.

Zeinab A. et al., [21] presents a system for optimizing parameters and settings of multiple algorithms of HRV feature extraction, it selects the best subset of features and tunes SVM parameters at the same time for maximizing prediction performance.

A. Mustaqeem et al., [28] collects enough data to form its own dataset; accuracy and kappa statistics are two evaluation measures used to assess the performance of the proposed algorithm.

H. Fujita et al., [31] ranks the features, which are clinically significant and are obtained by numerous means, using their t-value and feeds them to classifiers like k-nearest neighbor (kNN), decision tree and SVM. The technique predicts Sudden Cardiac Death (SCD) one, two, three and four minutes prior to the SCD with an accuracy of 97.3%, 89.4%, 89.4%, and 94.7%, respectively.

U. Rajendra et al., [35] feeds nonlinear features to these classifiers: Decision Tree (DT), kNN, and SVM, the mix of Discrete Wavelet Transform (DWT) and a nonlinear analysis of the ECG signals is empowered to predict SCD accuracies of 92.11% (kNN), 98.68% (SVM), 93.42% (kNN) and 92.11% (SVM) before one, two, three and four minutes before SCD occurrence, respectively.

Y. Zheng [38] employees Least Square-SVM (LS-SVM) in order to implement an intelligent diagnosis, an analysis of the results depict that the prediction accuracy of chronic heart failure is 95.39%.

The technique proposed in [41] aims to predict the severity of heart failure in patients and uses SVM with Gaussian Radial, and it solves an optimization issue by maximizing hyperplanes amongst the designed classes. Another study, [43] designs a system based on SVM and genetic algorithm, which is coupled with a 10 fold cross-validation technique for selecting features and optimizing parameters of the classifier.

2.6.2.2 Neural Networks

Zeinab A. et al., [19] proposes a highly accurate hybrid algorithm for diagnosing CAD. The proposed algorithm increases the performance of the neural network by around 10%; it assigns the initial weights via GA and then feeds it to an NN.

Roshan et al., [24] feeds the extracted features to two classification algorithms i.e. NN and LS-SVM, the higher accuracy is achieved by using the first approach with a blend of principal components of ECG beats with approximately 98.11% average accuracy.

Oluwarotimi et al., [26] consults a seasoned cardiac clinician to select 13 features and calculates their weights to train an ANN in order to predict the risks of HF. The technique is then applied to 297 HF patients. It predicts HF risks with an accuracy of 91.1%.

D. Tay et al., [33] introduces an algorithm that is inspired from neurons, state of the art testing is conducted on Honolulu Heart Program (HHP) dataset and its results are compared with SVM and Evolutionary Data-Conscious Artificial Immune Recognition System (EDC-AIRS). Results depict that the proposed algorithm outclasses both SVM and (EDC-AIRS).

Altan, G. et al., [34] focuses on diagnosing CHF and CAD with multilayer perceptron NN; its classification performance is state of the art. Leema, Khanna, and Kannan [39] proposes a Computer Aided Diagnostic system which makes use of ANN that is trained by Particle Swarm Optimization (PSO) and Gradient Descent Backpropagation in order to classify clinical datasets.

The techniques presented in [44] though is formed on Convolutional Neural Network, but can be distinguished from it as it proposes to use many options for segmentation, and a huge number of settings is tested.

A CNN based technique has been presented in [45], which has 2 convolution layers and 2 pooling layers, moreover, it makes use of LSTM (Long Short-Term Memory) i.e. an RNN but with a hidden memory layer and an output layer. The memory block in LSTM has 3 gates units which are called input, forget and output and a self-recurrent connecting neuron.

In [47], a neural network based technique for the prediction of heart failure, on patient's electronic medical data, has been presented. However, especially a one-hot encoding and word vectors have been employed for modeling and predicting HF with LSTM.

In [48] a novel technique has been presented which is based on CNN, the tool aims to classify healthy and pathological people by making use of an auditory sensor for FPGA (Field Programmable Gate Array, it is used to decompose audio in real time to frequency bands).

2.6.2.3 Decision Trees

Maryam T. et al., [20] uses a dataset of 1159 healthy, 405 negative angiography and 782 positive angiography participants, 10 variables out of 12 are entered into a Decision Tree and the model is able to identify the risk factors with an accuracy of 94%. Whereas, a Bagged Decision Tree ensemble is used by [27] for classifying two groups with a 5 fold cross-validation and a 20% holdout validation, yielding 98.1% and 99.5% respective accuracies.

K. Sudarshan [37] makes use of dual-tree complex wavelets transform or DTCWT in order to automatically differentiate ECG signals CHF from normal. The proposed methodology has been tested on ECG segments of 2 seconds. The study uses five different techniques for feature ranking, which are fed to kNN, and decision trees for classification. The presented method has achieved an accuracy of 99.86%. Since the method is based on merely 2 seconds of ECG signals, the clinicians would have sufficient time for further investigations.

2.6.2.4 Rules-Based, Rough Sets, and Fuzzy Logic

Purushottam, Kanak, and Richa [22] plans a framework for foreseeing the levels of risks of patients based on provided parameters. The major influence of this Rules Based algorithm is to assist a non-specialized physician in making the correct decision regarding the risk levels. While, Debabrata Pa et al., [23] helps in detecting CAD at an early stage, through rules formulation from doctors and a fuzzy expert system technique.

Nguyen, Phayung, and Herwig [25] makes use of fuzzy logic and presents a technique called interval type-2 fuzzy logic system (IT2FLS) which utilizes a blended learning procedure consisting of fuzzy c-mean clustering and parameters tuned in by Firefly and GA. Since there is high dimensionality involved the computation cost of this technique is high, however, an assessment of the experiments conducted by IT2FLS reveals that it outclasses other ML techniques, for instance, NB, SVM, and ANN.

A fuzzy clinical decision support system has been presented in [49], which has been founded on the C5.0 decision tree for classifying CAD and healthy conditions. The proposed algorithm is able to achieve an accuracy of 90.50%. One of the key features delivered by the system is the automatic detection of fuzzy rules without any aid from experts.

The algorithm presented in [50] makes use of a mining method to conduct the diagnosis. The algorithm consists of three parts i.e. firstly a fuzzy membership function is built by utilizing statistical methods and medical guidelines, secondly a decision tree creates the rules and lastly, fuzzy inferencing is used to predict heart disease.

2.6.2.5 Regression Models

A. Mustaqeem et al., [30] proposes the application of a classification algorithm called Contrast Pattern Aided LR (CPXR (Log)). It develops and validates prognostic risk models in order to forecast survival of one, two and five years in HF by implementing Electronic Health Records (EHRs).

Long, N. C., et al., [41] proposes MulSLR or Multilinear Sparse Logistic Regression, which can be seen as an extension of Sparse-LR. It differs from conventional LR in a sense that it solves for K classification vectors instead of one. The convergence problem is tackled by block proximal descent approach. Another research [51] implements 5 machine learning approaches i.e. NN, SVM, fuzzy rules, CART (Classification and Regression Tree) and random forest. Best results are produced by CART and random forests.

2.6.2.6 kNN

For the sake of experimentation, in [31] the entire set of clinically important features is first ranked and then fed to numerous classifiers including k-nearest neighbor (kNN), DT and SVM. Where for the given dataset SVM achieves the highest accuracy. While in [35], non-linear features are input to these two classifiers kNN, and SVM. The respective accuracies of the stated classifiers are 92.11% and 96.68%.

2.6.2.7 Random Forest

Zerina and Abdulhamit [36] examines five different classifiers namely C4.5, DT, kNN, SVM, ANN, and Random Forest, where results reveal that random forest achieves the highest accuracy. ECG signals are obtained from BIDMC CHF and PTB Diagnostic ECG databases.

A random forest inspired technique called improved random forest has been proposed by F, Miao et al., [46]. They introduce a split rule and stopping threshold in order to recognize more accurate predictors which are used to differentiate between survivors and non-survivors.

2.6.2.8 Ensembles

Saba, Usman, and Farhan [29] presents an ensemble technique with Multilayer classification by making use of improved bagging, and weighting. The proposed model which is called HM-BagMoov is empowered to overcome the cons of conventional performance bottlenecks by using a seven heterogeneous-classifiers ensemble.

C.-H. Weng et al., [32] investigates the performance of various classifiers, including classifiers involved in an ensemble and solo classifiers. They conclude that NN ensembles can improve the ability of generalization of learning systems significantly by training a limited number of NNs and then integrating their results.

Raid et al., makes use of three classifiers i.e. ANN, LS-SVM, and NB to develop an ensemble framework. The proposed ensemble is implemented on a real series telehealth data of chronic heart disease patients, and the results depict that the ensemble yields incredible accuracy and can be used to reduce the risk of incorrect recommendations [40].

A two-level neural attention model over an RNN (Recurrent Neural Network) has been formed in the shape of RETAIN model in [42]. The ensemble gets a patient's history as an input and produces a binary prediction on whether the patient is going to have an onset heart failure or not.

2.6.3 Analysis of the selected studies

In order to thoroughly analyze the research studies that have been selected, some parameters have been defined and all the selected papers are assessed against these parameters. These are the parameters that are used to evaluate and analyze the selected papers.

2.6.3.1 Research Problem

All the selected papers must implement its technique on a heart disease dataset, however, there is no restriction on the dataset's nature. The dataset can be signal images, numerical, or categorical.

2.6.3.2 Proposed Approach

The approaches proposed are of different natures, papers that implement individual classifiers are included and so are those that apply feature selection or reduction techniques before feeding it to the classifier. Moreover, ensemble frameworks are also included in this survey. Their proposed techniques have been depicted in Table 2.

2.6.3.3 Dataset

Heart disease datasets including UCI, PHP, BIDMC CHF dataset, and PTB Diagnostic ECG datasets, have been implemented in the selected papers. The proposed algorithms have been mainly applied to two types of datasets i.e. categorical or integer or real and ECG signals. The most common attributes in the former type are age, sex, chest pain, resting blood pressure, cholesterol, fasting blood sugar, resting ECG results, maximum heart rate, exercise-induced angina, and slope of peak exercise etc. As far as the ECG signals are concerned they consist of features of two major domains i.e. time and frequency. The former includes mean i.e. mean of all the RR intervals, standard deviation RR, square of the mean of the sum of the difference between adjacent intervals etc., and the latter includes the power of the low and high-frequency bands etc. Moreover, nonlinear dynamics measures are varied according to the applied or proposed techniques.

2.6.3.4 Validation Results

This is one of the most significant points in the analysis of the papers because the domain of the papers revolves around saving a life. Validation results i.e. accuracies have been given in Table 2, the accuracies given in the stated table are average accuracies and are given in percentages.

2.6.3.5 Future Work/Limitation

This section is going to play a vital role when it comes to assisting researchers in finding gaps in the existing literature. Due to the fact that HF is a severe life-threatening disease, the need for diagnosing it dire. This section of Table 2 contains adequate information about the possible future work and/or the limitation in each individual paper.

Table 2: ML techniques applied to Heart Disease Datasets

Sources	Research Problem	Proposed Approach	Dataset	Nature of Data	Accuracy in %	Future Work/Limitations
[17]	A method for congestive heart failure recognition.	Genetic Algorithm with Support Vector Machine.	CHF2DB and Normal Sinus Rhythm	ECG signals	98.79	-
[18]	A technique for automatically diagnosing normal and Coronary Artery Disease conditions.	It extracts features from Heart Rate Variability signals and applies PCA and SVM on it.	Long term ST DB	ECG signals	99.2	The study uses very little data for its training.
[19]	A hybrid method for diagnosing Coronary Artery Disease has been presented.	Genetic Algorithm and a neural network.	Z-Alizadeh Sani Dataset	Integer/Real	93.85	The number of attributes is way too many which may result in higher computation time.
[20]	A model for predicting coronary heart disease.	Decision trees.	A dataset of 2346 patients' records	Integer/Real	94	The algorithm has not been applied to any

			has been used in this study.			other relevant dataset.
[21]	An optimization algorithm for the prediction of paroxysmal atrial fibrillation.	SVM and Heart Rate Variability.	Atrial Fibrillation Prediction Database	ECG signals	87.7	Future works include stretching and implementing the presented algorithm on other HRV research gaps, for instance, detecting sleep apnea.
[22]	Assisting unspecialized doctors in making a decision regarding heart disease risk level.	Rules-based classifier.	Cleveland	Categorical, Integer, Real	86.7	Various combinations of confidence, MinItemSets, and the threshold for the algorithm can be attempted.
[23]	A screening system for the early detection of Coronary Artery Disease.	Rules and fuzzy experts approach.	Advanced Medical Research Institute	Categorical, Integer, Real	84.2	The current system has been designed for one disease, however, the rule organization supports for an extension of

						multiple disease expert system.
[24]	The study classifies five types of ECG signals for predicting abnormal cardiac activity.	PCA.	MIT-BIH	ECG signals	98.11	-
[25]	A system for diagnosing heart disease has been proposed in this study.	An integration of rough sets based on attribute reduction and interval type-2 fuzzy logic.	UCI and SPECTF	Categorical, Integer, Real	82.6	When it comes to training time and dealing with large records, the technique can be improved by utilizing Levy flights in moving strategy of Firefly.
[26]	This research works to diagnose the risk of heart failure.	Fuzzy analytic hierarchy process for computing global and an ANN classifier.	Cleveland Heart Disease Dataset	Categorical, Integer, Real	91.1	Automatic determination of the best number of hidden nodes and their links can still be seen as a challenge, it is evident from the fact that the ANN classifiers are trained numerously.
[27]	This paper aims to diagnose	A Probabilistic Symbol Pattern	PhysioNet database	ECG signals	98.8	Patient traits such as genetic risk

	Congestive Heart Failure for evading life-threatening events.	Recognition method has been deployed for detecting subjects in CHF, while an ensemble of bagged decision trees is used for classification.				factors, demographics, and co-morbidities may further increase the classification.
[28]	A hybrid technique which predicts disease and recommends medicines to cardiac patients.	It uses SVM, Random Forest (RF) and Multi-layer Perceptron for prediction and for the recommendation it uses a knowledge base.	POF Hospital	Categorical, Integer, Real	97.8	The model can be protracted to analyze the influence of features like age, gender and weight etc., moreover, smartphone or web-based interface can also be developed, for the recommender, in future.
[29]	An ensemble has been proposed for a medical decision support system.	HM-BagMoov: an ensemble of seven heterogeneous classifiers.	Cleveland	Categorical, Integer, Real	86.2	A publicly available is deemed to be developed for people who can make use of the application via the internet.

[30]	The research develops and validates prognostic risk models for predicting 1, 2, and 5-year survival in HF.	Contrast Pattern Aided Logistic Regression integrated with loss functions.	EHR data Mayo Clinic	Electronic Health Records	91.4	-
[31]	The paper automatically classifies HRV signals to predict sudden cardiac deaths.	The student's t-test has been used for feature ranking while DT, kNN, and SVM are used for classification.	MIT-BIH and NormalSinus Rhythm	ECG signals	94.7	It is intended that the HRV signal analysis is incorporated in a novel ECG and the clinicians are warned 24 hours prior to the SCD.
[32]	This study proposes a method for disease prediction.	Ensemble mainly consisting of ANN.	UCI	Categorical, Integer, Real	85.31	An issue of overfitting might occur if the training dataset is high.
[33]	A technique has been proposed that performs clinical risk prediction i.e. it estimates the possibility of disease-risk faced by a patient.	A novel, brain-inspired, an algorithm called Artificial Neural Cell System for classification or ANCSc.	Honolulu Heart Program	ECG signals	73.6	The first limitation is that the technique i.e. ANCSc has been evaluated using a single risk prediction task, whereas, the second one is that there is a huge room of

						improvement in its accuracy.
[34]	This research is focused on the prediction of congestive heart failure and coronary artery disease.	Multilayer perceptron neural network.	Normal Sinus Rhythm and Long-term ST.	ECG signals	97.83	Future works include a real-time model for the early prediction of CHF and algorithms based on a genetic search for enhancing accuracy.
[35]	A newly integrated index for the diagnosis of Sudden Cardiac Death using ECG signals has been proposed and evaluated in this paper.	Discrete Wavelet Transform with SVM has been deployed to predict SCD two minutes before death.	MIT-BIH SCD Holter and Normal Sinus Rhythm	ECG signals	98.68	The study calculates an index for SCD which can be constituted into a tool for clinicians and physicians to empower them in SCD diagnosis.
[36]	Long-term ECG time series has been classified as normal and congestive heart failure.	Autoregressive Burg for feature extraction and for classification five methods have been implemented; amongst which random forest	BIDMC CHF and PTB Diagnostic ECG databases	ECG signals	100	-

		classifies with maximum accuracy.				
[37]	Automatic prediction of Congestive Heart Failure.	Dual-tree complex wavelets transform.	MIT-BIH NSR, Fantasia, and BIDMC CHF	ECG signals	99.86	The method is planned to be evaluated on data with a bigger subject pool size.
[38]	Computer-aided chronic heart failure diagnosis.	Least-Square SVM has been used in this technique.	A dataset of 152 heart sound samples has been compiled for this study.	ECG signals	95.39	The dataset is very limited, moreover, it does not contains heart murmurs samples.
[39]	Computer-aided classification of clinical datasets.	An ensemble of ANN, PSO, and Differential Evolution.	UCI	Categorical, Integer, Real	86.66	-
[40]	Supporting recommendations for heart patients	Least-Squares SVM, ANN, and Naïve Bayes.	Tunstall	ECG signals	94.83	Future works include the application of this technique on a more reliable and appropriate dataset.
[41]	The study works on an adaptive and context-aware	The proposed algorithm is based	Compiled their own dataset.	ECG signals	87.9	The articles do not say much about their

	decision-making system for Intensive Health Care provision.	on RBF SVM and LKF SVM.				dataset, and it should be evaluated on a bigger and reliable dataset.
[42]	Heart risk prediction.	This study proposes and evaluates RETAIN on an enormous dataset for risk prediction.	Cerner Health Facts	Electronic Health Records	82	The study does not take into account the factors liable for the varying accuracies among the datasets, which is stated to be their future work.
[43]	This article aims to recognize cardiac health on the basis of ECG signals.	SVM with Genetic Algorithm and 10 fold cross-validation.	MIT-BIH Arrhythmi a	ECG signals	98.85	Future works include the development of a prototype for fetching ECG signals and this algorithm to diagnose heart problems.
[44]	The study aims to automatically classify heart sound for pathology detection.	Convolutional Neural Network.	UoC-murmur and PhysioNet -2016	PCG and ECG signals	84.6	This study lacks an insight analysis as far as the sub-band filtering processes are concerned.

[45]	Identification of coronary artery diseases ECG signals.	Long Short-Term Memory network and CNN.	PhysioNet database	ECG signals	99.85	Future works include the installation of the algorithm in portable devices so that healthcare personnel can diagnose CAD at the earliest.
[46]	Prediction of heart failure patients' mortality.	The algorithm is based on improved random survival forest.	Multi-parameter Intelligent Monitoring in Intensive Care	Categorical, Integer, Real	82.1	ECG, PPG, and BP variation may be merged with the input data to improve the performance of the algorithm.
[47]	Heart failure risk prediction.	Neural networks and Long Short-term Memory network.	-	Electronic Health Records	66.55	In the future, expert knowledge can be incorporated into the system.
[48]	Recognition and classification of heart murmurs.	A novel convolutional neural network.	PhysioNet /CinC Challenge	EEG signals	97	-
[49]	Diagnosis of coronary artery diseases.	The presented model is founded on Random Forest, C5.0, and fuzzy modeling.	UCI	Categorical, Integer, Real	90.50	Future works include containing more features and generalizing the algorithm for other datasets.

[50]	Heart disease prediction	Fuzzy rules-based method.	Personal Health Record	Categorical, Integer, Real	69.22	In the future, an analysis of the proposed technique on a large scale dataset can be performed.
[51]	Analysis of heart patients data for severity evaluation and type prediction	Regression models called: Classification And Regression Tree	Cardiolog y Departme nt at the St. Maria Nuova Hospital in Florence, Italy	Categorical, Integer, Real	84.7	The findings can be generalized with larger sample size.

2.6.4 Discussion

HF has been proven one of the leading causes of deaths worldwide, this is the main reason why accurate prediction of HF risks is extremely vital in order to prevent and treat it [26]. A timely diagnosis of CHF is critical for evading a life-endangering event [27]. Apart from timeliness, accuracy plays an extremely significant role in the medical domain as it is related to the life of a person. A great and extensive amount of research has already been carried out on disease classification and prediction using ML techniques. Conversely, an agreement is yet to be made about which technique or classifier is best suited for which datasets. However, it is proven that the feature selection and reduction technique increases the accuracy and reliability of classifiers. Moreover, it is also clear that classifier ensembles have been proven to have improved classification accuracy [29].

When it comes to health-care problems there are certain aspects which cannot be overlooked. These aspects include: time taken to execute the technique and computational complexity which is majorly dependent on the number of features, accuracy, and generalization. According to our analysis, some of the studies have been able to focus on avoiding the inclusion of redundant features. The use of state of the art techniques e.g. GA in [17], [33] and [43], weights by SVM in [18], information gain in [28], F-score in [29], for feature selection has helped in negating the problem of too many features and hence decreasing the time of execution and the computational complexity.

When we carry out an insightful analysis of the papers, we realize that the curse of dimensionality has been ameliorated by some studies using advanced techniques of feature reduction. For example, PCA in [18], [19], [24] chaos firefly in [25], and minimum Redundancy maximum Relevance in [37] are extremely efficient for dimensionality reduction. On the contrary, we also realize that the question that whether feature reduction causes information loss or not has not been answered in most of the studies.

Furthermore, when it is a question of life, accuracy is the most important aspect and since machines are being used to do that prediction, the margin of error ought to tend to zero. According to our analysis, some studies have focused on achieving the highest accuracies by proposing extremely reliable and efficient techniques. The most accurate and reliable techniques are proposed and evaluated in these papers, e.g. GA with SVM to achieve over 98% accuracy in [17], PCA and SVM to gain over 99% accuracy in [18]. PCA for getting over 98% accuracy in [24], and an ensemble of bagged decision trees for attaining around 99% accuracy in [27], more such techniques and their accuracies are depicted in Table 2. What is alarming is that most of the techniques have been applied to limited datasets and a dire need of evaluating them on a large dataset still exists.

The last aspect too is a significant one, sometimes researchers commit the mistake of over-training their algorithm on one particular dataset and not exposing or assessing their performance on unseen data. Which causes the problem of overfitting, where an algorithm performs outstandingly well on a particular dataset and fails miserably when tested on an unseen dataset. In this research we find some studies which have tackled this problem, these studies include [29] and [37] where the algorithms are evaluated on 5 datasets each. The accuracies in these techniques can still be improved because the mean accuracy that has been achieved by [29] is 85.75% and the same achieved by [19] is 88.28%.

Figure 3 shows the use of different algorithms and techniques in the most recent researches. This survey aims to explore, summarize, and critically analyze the most recent and state of the art research papers in order to find research gaps for future studies. This research has been conducted systematically, to help readers gain the knowledge of previous researches conducted in the domain of heart failure and risk detection. The limitations and future work provided in Table 2 can assist researchers in fulfilling the need for future research and gap in research. Whereas, the discussion section can help them direct their research.

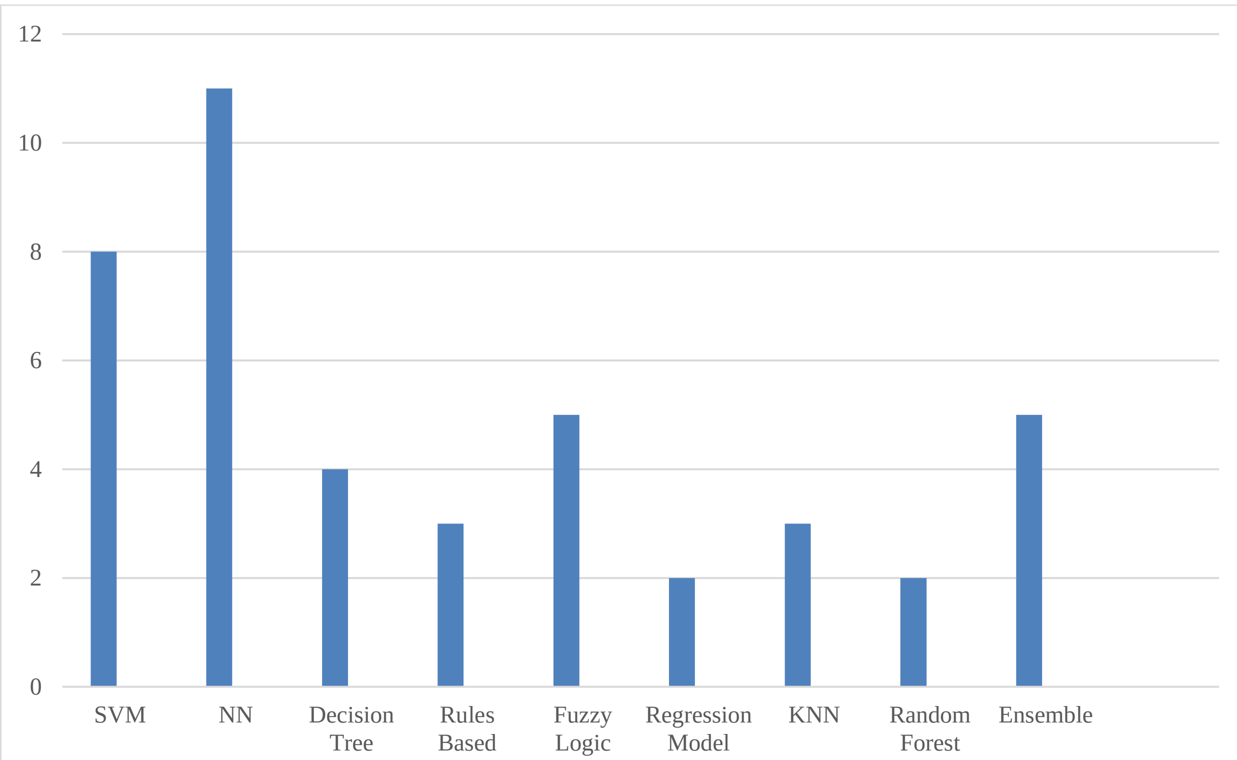

Figure 3: Techniques used in recent researches

2.7 Summary

As the title of the chapter suggests, this chapter of the manuscript deals with the literature review and background study of the thesis. It starts with an illustration of decision support systems, which serves the purpose of assisting to determine and evaluate different alternative actions. It relies on data, normally on a huge chunk of data, which is provided to it, it then analyzes that data with the help of statistical procedures and equipment in order to derive meaningful data from it, it then tapers down the alternatives or the available choices of decisions by making use of decision theories. Then it proceeds the types of decision support systems, a few of which are: Data-Driven Decision Support Systems, Model-Driven Decision Support Systems, Knowledge-Driven Decision Support Systems, Document-Driven Decision Support Systems, Communication Driven Decision Support Systems, and some hybrid decision support systems.

Then it describes decision support systems for health care, which can be used to provide clinicians, patients, staff or other people with people-specific information. The information is filtered intelligently and is presented at a suitable time in order to improve health or healthcare. Clinical Decision Systems are used to encompass various tools in order to improve decision-making in clinical environments. It also presents a few examples of clinical decision support systems.

Then a brief of Machine Learning is provided, which falls under the umbrella of artificial intelligence; it is the ability of a machine to learn from a large set of data and predict, cluster or classify similar but unseen or new data based on its learning or training. Some famous machine learning techniques include Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree, self-organization map, and k-means clustering etc. Then the use of Machine learning for developing decision support systems has been presented with examples.

After setting the tone for and providing the basic background of decision support systems and machine learning, the chapter proceeds to present a systematic literature review of machine learning techniques applied to heart disease dataset. The SLR aims to explore, summarize, and critically analyze the most recent and state of the art research papers in order to find research gaps for future studies. This research has been conducted systematically, to help readers gain the knowledge of previous researches conducted in the domain of heart failure and risk detection.

CHAPTER 3

EXPERIMENTATION

3.1Introduction

This chapter of the thesis discusses first discusses the initially proposed methodology which is used for the purpose of experimentation. Our main dataset is Z-Alizadeh Sani dataset, which contains 303 records of heart patients. The stated dataset has 54 features, covering symptomatic, demographic, checkup and electrocardiograph results. The given number of features is really high, so four feature selection techniques are used for narrowing the number of features down. For validation purpose 10 fold cross-validation technique is used. After splitting the dataset into two parts i.e. training and testing, four weight optimization techniques are used, and for learning purpose an artificial neural network is used.

3.2Initial Methodology

In this section, an initially proposed methodology has been illustrated, where four different techniques have been trialed for feature selection, 10 fold cross-validation technique has been used for assessing the performance of the algorithm, four different techniques have been used for weight optimization, and a multi-layer perceptron has been used for learning and classification purpose.

3.2.1 Data

Since this research aims to diagnose heart diseases our primary dataset is the Z-Alizadeh Sani, which contains 303 records in which 87 are healthy and 216 are heart patients. The dataset has 54 features which cover a wide range of demographic, symptomatic, electrocardiograph results, check-up results.

3.2.2 Feature Selection

When it comes to predictive modeling, the decision of which features should be used is of the highest significance. This question becomes particularly difficult and requires a great amount of knowledge in the domain of the problem. This is where automatic feature selection comes into play. In some cases, feature selection can also be termed as attribute or variable selection. However, its essence remains the same i.e. the automatic selection of features from data in hand or columns in a tabular dataset, which are deemed as the most relevant for predictive modeling. Feature selection is often confused with dimensionality reduction, due to the fact that their purpose is somewhat the same i.e. reducing the number of features. However, dimensionality reduction does this job by developing novel combinations of features, on the other hand, feature selection methods opt for including and excluding features from the dataset without altering them. Some examples of dimensionality reduction include: Singular Value Decomposition, Principal Component Analysis, and Mapping [57].

Feature selection methods mostly are used to get aid from in creating an accurate model for prediction. As they help us, in opting for features that are going to provide us with better accuracy whilst asking for lesser data. These methods are used for identifying and removing irrelevant, unneeded and redundant features from the dataset, which do not back the accuracy of the model or even worse, they may be decreasing the accuracy of our model. With fewer features, the complexity of the model is reduced, and with that comes the ease of understanding and explaining.

Numerous techniques, such as filter methods, wrapper methods, and embedded methods can be used for feature selection. Which work on the basis of different methodologies i.e. statistical measures (in case of filter methods), preparing, evaluating and comparing various combination as search problems (in case of wrapper methods), and regularization (in case of embedded methods) [62]. However, based on previous similar studies, this research work opts for the following feature selection techniques

3.2.2.1 Weight by SVM

Support Vector Machine is a powerful tool used for both classification and regression. What it does is that it makes use of kernel trick in order to change the data in order to find an optimized boundary in them [63]. Support vectors are formed which are points that are nearest to the hyperplane, and if these points were to be removed they would change the position of the boundary or hyperplane. The hyperplane can be thought of as a line linearly distinguishes between instances of a dataset [64].

However, Weight by SVM is a built-in technique, available in RapidMiner. It is used to calculate the relevance of features by calculating the input weight of each feature with respect to the class label. This technique is inspired by SVM as it makes use of the coefficients of the normal SVM vectors and considers them as feature weights [57] [65].

Table 3 provides a list of all the features selected by weight by SVM.

Table 3: Features Selected weight by SVM

1	HTN
2	FH
3	Non-Anginal Chest Pain
4	ST Elevation
5	Blood Pressure
6	Low Threshold Angina
7	Region with RWMA
8	Atypical
9	ESR
10	VHD
11	Fasting Blood Sugar
12	LVH
13	Q Wave
14	Diastolic Murmur
15	Function Class
16	Dyspnea
17	Ejection Fraction
18	PR
19	Age
20	Sex
21	DM
22	Current Smoker
23	Typical Chest Pain
24	ST Depression
25	T inversion
26	Cr
27	TG
28	K

3.2.2.2 Weight by Gini Index

Gini Index which is also termed as Gini Coefficient is a statistical measure for distribution, it is named after its developer Gini Corrado. This index ranges from 0 to 1, with the former representing perfect equality and the latter perfect inequality [66].

As far as Weight by Gini Index is concerned, it is also a built-in operator in RapidMiner, which is used for feature selection. It calculates the relevance of the feature of a given dataset on the basis of their Gini Index. It computes the weight of a feature with respect to their class label by calculating the Gini Index of that particular class distribution, given that the provided dataset had been split on the basis of that feature. In simple words, if the weight assigned to a particular feature is higher it is more relevant [59] [67].

Table 4 provides a list of the features selected by weight by Gini Index.

Table 4: Features Selected weight by Gini Index

1	Dyspnea
2	Non-Anginal Chest Pain
3	ST Depression
4	ESR
5	Lymph
6	VHD
7	PR
8	Ejection Fraction
9	DM
10	Neut
11	Q Wave
12	HTN
13	Atypical
14	TG
15	Cr
16	ST Elevation
17	Region with RWMA
18	BUN
19	Age
20	BMI
21	BP
22	Diastolic Murmur
23	Typical Chest Pain
24	T inversion
25	Fasting Blood Sugar
26	K
27	Na

3.2.2.3 Weight by PCA

Principal Component Analysis is used to figure out linearly independent dimensions which can be used to represent data points, and then use those new dimensions for predicting or reconstructing the original dimensions, whilst minimizing the reconstruction error. The process is completed in five steps, a) calculating covariance matrix “X” for data points, b) calculating Eigen Vectors and their corresponding Eigen Values, c) sorting Eigenvectors in accordance with their respective Eigenvalues in descending order, d) choosing the first k Eigenvectors and hence the new “k” dimensions, and lastly, e) transforming the original “n” dimensions into “k” dimensions [68].

Weight by PCA is a built-in operator of RapidMiner, which is used to create feature weights of a given dataset by making use of a component developed by PCA. Where the weights of the feature project its relevance with respect to the class label, where the higher the weight, the more relevant it is [58] [69].

Table 5 provides a list of all the features selected by weight by Principal Component Analysis.

Table 5: Features Selected by weight by PCA

1	Age
2	HTN
3	Lymph
4	PLT
5	Length
6	BMI
7	K
8	DLP
9	Na
10	Function Class
11	FBS
12	VHD
13	TG
14	Neut
15	Weight
16	WBC
17	Cr
18	HDL
19	LDL
20	Sex
21	DM
22	BP
23	PR
24	ST Depression
25	BUN
26	ESR
27	Ejection Fraction
28	Region with RWMA

3.2.2.4 Weight by Information Gain

Information in a particular dataset can be calculated by subtracting entropy from one, and if an observation regarding the data is formed, the latest information can also be calculated. Whereas the difference between the two values of information is termed as information gain [70].

Weight by Information Gain is also an operator provided by RapidMiner, which is used to calculate the relevance of the features on the basis of information gain and then gives them weights. It calculates the weights of features with respect to the class label by making use of information gain. The higher the weight the more relevant the feature [71].

Table 6 provides a list of all the features selected by weight by Information Gain.

Table 6: Features Selected by weight Information Gain

1	HTN
2	Diastolic Murmur
3	ST Elevation
4	ST Depression
5	Poor R Progression
6	FBS
7	TG
8	Lymph
9	Age
10	T inversion
11	PR
12	Region with RWMA
13	Non-Anginal CP
14	VHD
15	Q Wave
16	Dyspnea
17	Na
18	DM
19	Atypical
20	BMI
21	ESR
22	BUN
23	Ejection Fraction
24	BP
25	Typical Chest Pain
26	Cr
27	K
28	Neut

3.2.3 Validation

Validation is a significant step that has to be placed in order to assess the viability and accuracy of a technique. We use the validation process right after the feature selection takes place [60]. We have opted for 10 fold cross-validation technique which uses 90% of data for the purpose of training and 10% for testing.

3.2.4 Weight Optimization

Till the year 2015, machine learning developers have been working with randomly initialized weights as the beginning point for neural networks. They have recently realized that the initial weights of features or attributes actually played a significant role in figuring out the global minimum of the neural network's cost function [72].

Numerous initialization methods can be used for neural networks, and as a matter of fact, it is considered as a whole new field of study because careful initialization of neural networks can not only speed up the learning process but can also improve accuracy. Today, libraries are available for deep learning, one of such famous libraries is Keras. Keras offers numerous such initialization techniques, and an interesting fact about these methods is that all of them are some of the other kind of variation of random initializations. Currently, these methods are supported by Keras for initializing weights of neural networks: zeros, ones, constants, normal distribution, uniform distribution, truncated normal distribution, an orthogonal matrix, and identity matrix etc. [73].

3.2.4.1 Optimize weights-PSO

Particle Swarm Optimization, as the name suggests is a technique used for optimization. It works in numerous iterations, in which the values of a group of variables are adjusted on the basis of the member which is the closest to the target at a particular time. To understand this, let us imagine a flock of birds circulating over an area and smelling some hidden food. The bird that is the closest to the food gives a signal to the others by chirping and the other birds follow him. And if any other bird becomes close to the food than the previous one it chirps louder and the remaining then follow this one. This goes on until they find the food. Three global variables are involved in this variable: target condition, global best or gBest, and stopping criterion. Each particle has a presentation of a possible solution, velocity i.e. possibility of change in data, and personal best or pBest [74].

Optimize weights-PSO is a built-in operator of RapidMiner, which works on the very principles of Particles Swarm Optimization.

3.2.4.2 Optimize weights-Evolution Strategy

This is another operator provided by RapidMiner, it used to calculate the relevance of a feature of a provided dataset by making use of the evolutionary approach, where the weights are computed with a Genetic Algorithm. GA is termed as a search heuristic algorithm, mimicking natural evolution processes. It is used to routinely to produce viable solutions for optimizing problems. GA belongs to a larger category of evolutionary algorithms, which are used to solve optimization issues by making use of natural evolution inspired techniques for instance: inheritance, mutation, selection, and crossover [61] [75].

3.2.4.3 Optimize weights-Backward

This operator is provided by RapidMiner, it assumes that the attributes are independent and assigns 1's as initializing weights to them [76].

3.2.4.4 Optimize weights-Forward

RapidMiner provides this operator as well, it assigns weights based on the assumption that the attributes are independent. All the attributes are assigned 1's as initial weights [76].

3.2.5 Neural Network

It is known that the main idea behind the logic of artificial neural network is that of the human brain. ANN works on something termed as hidden states, which are similar to neurons. Every hidden state is a transient method which possesses probabilistic behaviors. A sequence of such hidden states exists between inputs and outputs [77]. Just like our brains, where there are 100s of millions of neurons, in which information is processed in the form of electric signals. External stimuli/information is acknowledged by the neurons and is processed in the cell body of the neuron, and then transformed into the output and send via the axon to another neuron [78].

$$\cdot f\left(b + \sum_{i=1}^{n} x_i w_i\right)$$

Equation 1: Artificial Neural Network

Where b is biased, x is the input to neurons, w is weight, n is the number of inputs from the previous layer, and i is a counter from 0 to n.

For training and classification purpose we have used an artificial neural network, which has three kinds of layers i.e. input, hidden and output layers. Every layer has neurons (processing units). The behavior of an ANN is dependent on the links amongst the variables of inputs and outputs. Our first layer is the input layer, which has weights of features. Our second layer is the hidden layer, we have experimented with it and have used 1, 2 and 3 hidden layers here and each one has 13 neurons. Our last layer is the output layer, which states that either a record is classified as a CAD patient or normal. We have made use of the feedforward training technique and Multi-Layer Perceptron to update weights. The learning rate of our ANN is 0.01 and the momentum is 0.9.

3.2.6 Classification

Data with the selected features have been input to our neural network, but before that the weights are optimized using four different techniques. In order to decide which techniques have achieved how much accuracy, let us have a look at the change in classification accuracy for each optimization technique.

The following table i.e. Table 7 is used to illustrate the results of an artificial neural network when its initial weights are assigned with different weight optimization techniques. The features used in the stated table have been selected with the help of weight by support vector machine. As the table reveals different blends of artificial neural networks have been used i.e. ANN having one, two and three hidden layers. In this particular case, ANN having three hidden layers is producing the best results.

Table 7: Applying Weights Optimization Techniques to Features selected by Weight by SVM

Weight Optimization Technique	Hidden Layers	Accuracy in %
Particle Swarm Optimization	1	85.51
	2	83.80
	3	**87.80**
Evolution Strategy	1	84.82
	2	84.17
	3	86.80
Backward	1	85.80
	2	86.77
	3	87.11
Forward	1	82.84
	2	82.84
	3	82.84

The following table i.e. Table 8 is used to describe the results of an artificial neural network when its weights are initialized with different weight optimization techniques. The features used in the following table have been selected with the help of Principal Component Analysis. It has been clearly depicted in Table 8 that ANN having a different number of hidden layers have been tried for the purpose of experimentation. Though the accuracies attained on the basis of features produced with Principal Component Analysis are lower as compared to those features that are selected by Support Vector Machine. However, just like the previous instance the ANN having three hidden layers is performing better than the rest.

Table 8: Applying Weights Optimization Techniques to Features selected by Weight by PCA

Weight Optimization Technique	Hidden Layers	Accuracy in %
Particle Swarm Optimization	1	76.31
	2	79.17
	3	**80.51**
Evolution Strategy	1	78.25
	2	78.58
	3	77.25
Backward	1	79.19
	2	78.22
	3	76.27
Forward	1	71.31
	2	71.31
	3	71.31

Table 9 is used to illustrate an artificial neural network having one, two and three hidden layers. Features that have been selected by weight by information gain have been taken into consideration for this particular scenario. Whereas four different weight optimization techniques have been used for the initializing attribute weights in the input layer of the ANN. It can be observed that when the backward weight optimization technique is used and the ANN has three hidden layers, the performance of the system is better than the rest.

Table 9: Applying Weights Optimization Techniques to Features selected by Weight by Information Gain

Weight Optimization Technique	Hidden Layers	Accuracy in %
Particle Swarm Optimization	1	87.13
	2	87.42
	3	87.45
Evolution Strategy	1	82.16
	2	82.84
	3	84.81
Backward	1	86.44
	2	87.44
	3	**87.44**
	1	76.23

Forward	2	76.19
	3	71.32

Table 10 illustrates the performance of an artificial neural network having different hidden layers. Features selected by weight by Gini Index have been considered in this case, whereas the weight initialization in the input layer has been trialed with different weight optimization techniques. It can be seen that ANN with three hidden layers performs better than the rest of the combinations.

Table 10: Applying Weights Optimization Techniques to Features selected by Weights by Gini Index

Weight Optimization Technique	**Hidden Layers**	**Accuracy in %**
Particle Swarm Optimization	1	86.81
	2	85.15
	3	85.18
Evolution Strategy	1	86.45
	2	86.11
	3	86.80
Backward	1	87.51
	2	87.16
	3	**88.49**
Forward	1	71.38
	2	71.38
	3	71.38

By analyzing the results produced, it can be realized that when the weights of features selected through Gini Index are optimized with Backward Weight Optimization, the accuracy is the highest. The next best weight optimization technique is PSO when features selected by SVM are used, and the remaining results can be seen in tables 7 to 10. These results are used to propose a novel ensemble technique.

3.2.7 Results on Other Datasets

In order to make sure that the algorithm has not led to overfitting, and it does support generalization, the technique has also been applied to two other datasets. These datasets are Wisconsin Breast Cancer and Cleveland's Heart Disease. The technique has also been assessed on a multi-class classification problem that is why, Cleveland's datasets has been used which has five possible outputs, while Wisconsin is a binary classification problem.

Furthermore, to have an idea of whether our technique has improved the performance of the simple NN or not, the results of the proposed technique have also been compared with those of a simple neural network.

Table 11: Comparing the performance of NN and the first technique

Dataset	Technique	Accuracy in %
Cleveland	**Proposed NN**	**95.01**
	NN	89.75
Wisconsin	**Proposed NN**	**96.67**
	NN	88.40

3.3Summary

In this chapter, experimentations have been performed in order to assess the methodology proposed in the previous chapter. The main dataset used in this manuscript is Z-Alizadeh Sani Dataset, which is a heart disease dataset. The stated dataset has over 300 records and 54 features. The number of features is high, so four feature selection techniques have been used to narrow the number of features down. The main objective of this thesis is to ameliorate the performance of an artificial neural network and for that purpose, four weight optimization techniques have been trialed to optimize the initial weights assigned to the input layer of the artificial neural network.

The experimental results reveal that when the weights of features selected through Gini Index are optimized with Backward Weight Optimization, the accuracy is the highest. These results have been taken into consideration in order to propose the final methodology.

CHAPTER 4

PROPOSED METHODOLOGY AND IMPLEMENTATION

4.1Introduction

In this chapter, the final methodology has been discussed. It is very clear from the results that the feature selected Gini Index are the best ones and hence the Gini Index has been chosen in the final methodology. For validation purpose, 90% of the data has been used to training and the remaining 20% data has been used for testing. While an ensemble has been presented based on the average weights assigned by three of the weight optimization techniques, i.e. Evolution Strategy, Particle Swarm Optimization and Backward Weight Optimization. The weights of the attributes in the input layer of the neural network have been initialized with the average of weights produced by these techniques.

4.2Proposed Methodology

A flow chart which depicts the proposed methodology has been presented in the following figure i.e. Figure 4.

4.2.1 Data

The same dataset i.e. Z-Alizadeh Sani dataset has been used, which contains 303 records in which 87 are healthy and 216 are heart patients. The dataset has 54 features which cover a wide range of demographic, symptomatic, electrocardiograph results, check-up results.

4.2.2 Feature Selection

As far as feature selection is concerned weights by Gini Index has been used for the stated purpose.

4.2.3 Validation

In order to validate the proposed technique, the dataset has been divided into two parts i.e. training and testing, 90% of the data has been used for the former purpose and 10% for the latter.

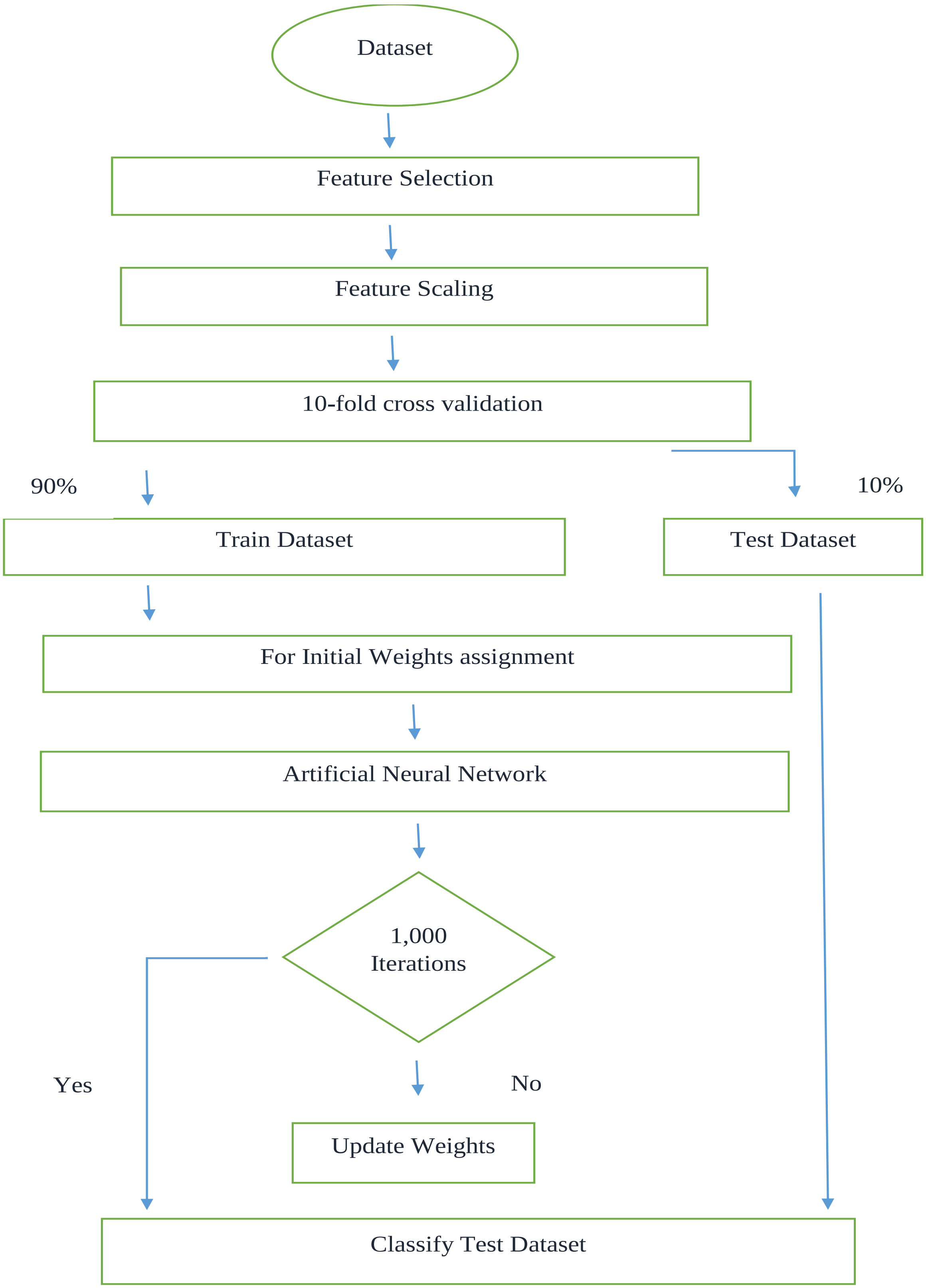

Figure 4: Proposed Methodology

4.2.4 Feature Scaling

Before we start feeding our data to the artificial neural network, we need to take care of the scaling problem. Feature scaling is said to have a significant impact when some algorithms are used in other cases it might have any influence. In order to grasp a good idea about scaling let us discuss it. More often than not our data has such features which vary extremely highly as far as magnitude, range and units are concerned. If not dealt with, some algorithms would ignore the units and only take feature's magnitude into consideration. This will have a huge impact on the results because e.g. 5kg and 5000g will not be considered the same [79].

In order to deal with this issue, we make use of Keras *StandardScaler* function, which is used to scale the data in such a way that the distribution is centered on zero, and one as standard deviation. What it does is that it calculates the standard deviation and mean for the features and then the features are scaled respectively [80].

$$SS = xi - mean(x)/stdev(x)$$

Equation 2: Standard Scaler

4.2.5 Weight Optimization

Our main contribution is in this particular step, in the previous methodology four different techniques have been trialed for weight optimization. Whereas, in this step of the final methodology, an ensemble has been developed on the basis of averaging the weights of three weight optimization techniques. These weight optimization techniques are: PSO, ES and Backward weight optimization. The average weights of these techniques have been used to initialize the weights of the features or attributes, in the input layer of the neural network.

4.2.6 Neural Network

For training and classification purpose we have used an artificial neural network, which has three kinds of layers i.e. input, hidden and output layers. Every layer has neurons (processing units). The behavior of an ANN is dependent on the links amongst the variables of inputs and outputs. The first layer is the input layer, which has weights of features, the second layer is the hidden layer and the last layer is the output layer, which states that either a record is classified as a CAD patient or normal. The stopping criterion for the artificial neural network is 1,000 iterations.

4.2.6.1 Activation Function

In order to understand what activation functions do, we need to understand the functionality of an artificial neuron. Neurons calculate the weighted sum of their inputs, adds a bias to it and make the decision of whether they should be fired or not. The value of Y can range from +inf to – inf, and the neuron cannot bound the value. Then how is the decision of whether to fire or not made? This is where activation functions come into play [81].

There are numerous kinds of activation functions, namely: step function, linear function, a sigmoid function, *tanh* function, and ReLU etc. However, the selection of these activation function varies from case to case. In our case, we have used two different activation functions i.e. ReLU and Sigmoid.

4.2.6.1 Rectified Linear Units or ReLU

ReLU makes sure that the output does not become a negative value. So, when *z* is greater than zero the output stays *z*, and if it does below zero the output stays zero. Moreover, it is also used when there are numerous output possibilities. In our case, ReLU is used for going from the input layer to the hidden layer [82].

$$f(z) = \max(0, z)$$

Equation 3: ReLU

What gives ReLU an edge is that it does not activate the entire set of neurons at once. E.g. when a negative input is received it will be converted into zero, and the neuron will not get activated. In other words, at a given point in time a few neurons will be active and hence ReLU helps in making the artificial neural network sparse and hence increases its efficiency [83].

4.2.6.1 Sigmoid Function

Sigmoid functions are used to bind the output in a range (0, 1). In our case, we use it in the output layer i.e. whether the subject has CAD or not [82].

$$\mathrm{f(z)} = 1 / (1 + \mathrm{e}\hat{}(-\mathrm{x}))$$

Equation 4: Sigmoid

The sigmoid function is used when we have a binary classification problem at hand, i.e. just like our example where the decision that whether a particular subject has CAD or not. However, sigmoid cannot be used for multi-class classification, but since our problem is binary classification, we opt for sigmoid at the output layer.

4.2.6.2 Optimization

Usually, forward propagation and backward propagation techniques are used as error functions or weight optimization. But in our case, we use the Adaptive moment estimation technique. However, we use the following technique for optimization.

4.2.6.2 Adam

Since 2014, a special optimization algorithm in the shape of Adam (Adaptive Moment Estimation) for deep neural networks is present [84]. Adam is one of the best methods that are used to calculate adaptive learning rates for every parameter. It computes the adaptive learning rates for all the parameters. Apart from storing the exponentially decaying averages of previously squared gradients, for instance, RMSporp and Adadelta, it also keeps something similar to momentum. If momentum is thought of like a ball going down a slope, Adam can be termed as a heavy ball having friction and hence providing us with flat minima [85].

4.2.6.3 Binary Class Entropy

When are working on a binary classifier, we more often than not opt for a binary class entropy or the log loss. Here is the equation of binary class entropy.

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Equation 5: Binary Class Entropy/Log Loss

Where y is the label that is its value is either 1 (CAD patients) or 0 (normal), p(y) can be termed as the probability predicted for the outcome being 1. The equation reads as: for each CAD, *log(p(y))* has to be added to the loss, that is the log probability of it being a CAD. And conversely, it sums *log(1-p(y))*, i.e. the log probability of it being normal, for every normal point y =0 [86].

4.3 Summary

This chapter of the manuscript presents the main idea and the main methodology proposed in the thesis. It builds on the experimentation that has been performed in the previous chapter i.e. trailing four different feature selection and the same number of weight optimization techniques for enhancing the performance of an artificial neural network in order to diagnose coronary artery disease in an improved and ameliorated way. Based on the results presented in the previous chapter, the feature selected by weight by Gini Index best describes our main dataset that is Z-Alizadeh Sani Dataset. And that is the reason why we have selected weight by Gini Index as the final feature selection technique. For optimizing weights an ensemble technique has been brought forth that takes an average of weights produced by three of the weight optimization techniques i.e. PSO, ES, and backward and feeds that as the initial weights to the artificial neural network.

CHAPTER 5

RESULTS, COMPARISON AND ANALYSIS

5.1Introduction

In this chapter, the results of the proposed methodology have been analyzed and discussed. The confusion matrix i.e. accuracy, sensitivity, and specificity have been presented. Moreover, a learning history i.e. epoch vs accuracy graph has also been illustrated.

5.2Results

The proposed technique has achieved over 94% of accuracy, whereas, its sensitivity is 100% and its specificity is 73%. Moreover, some other results have been depicted in the following graphs.

Figure 4 is used to present epoch vs accuracy graph, it is clear from the figure that the learnability and hence accuracy of the model have increased as the number of iterations increase. However, after the 200th iteration, the model has stopped learning.

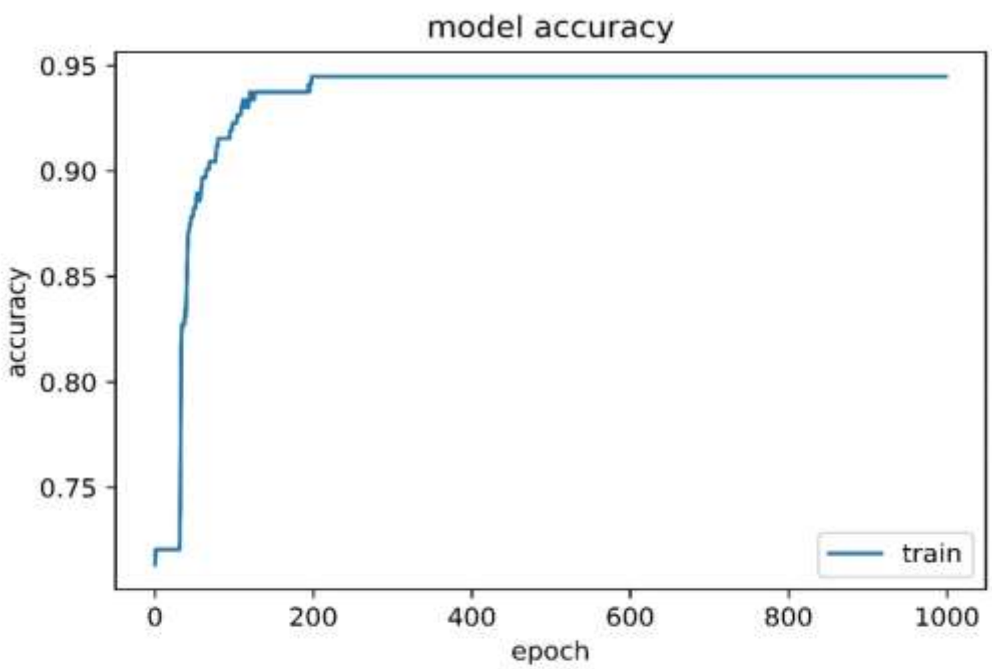

Figure 5 : Accuracy vs Epoch

The following figure i.e. Figure 5 is used to represent loss vs epoch graph, it is clear from the graph that loss has decreased as the number of iterations increase.

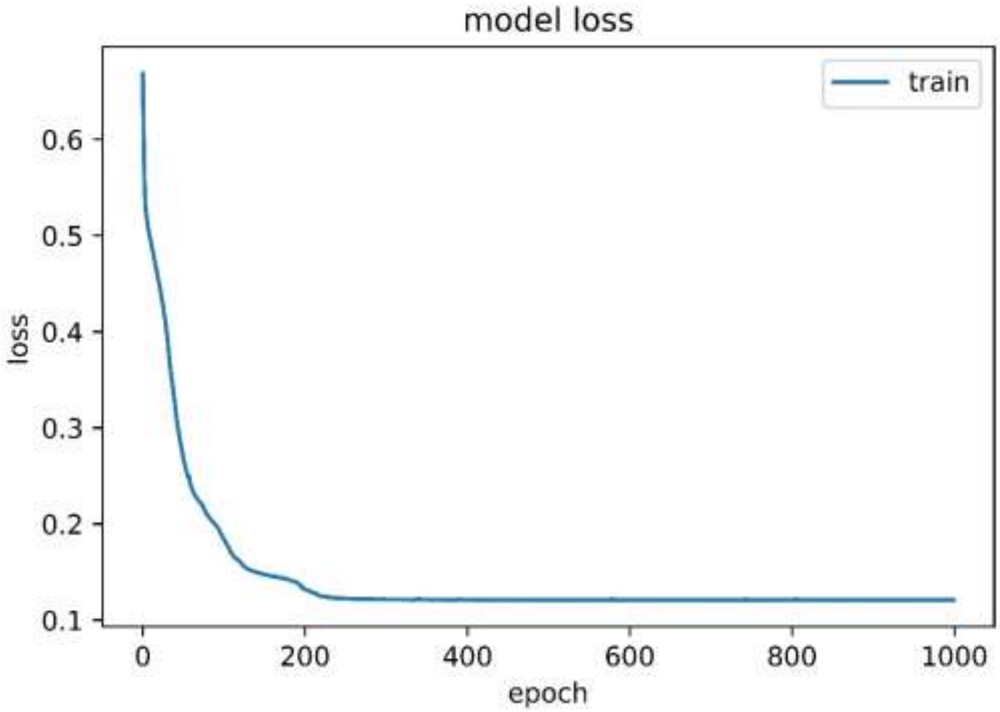

Figure 6: Loss vs Epoch

The following i.e. Figure 6 is used to present the Receiver Operating Characteristics graph.

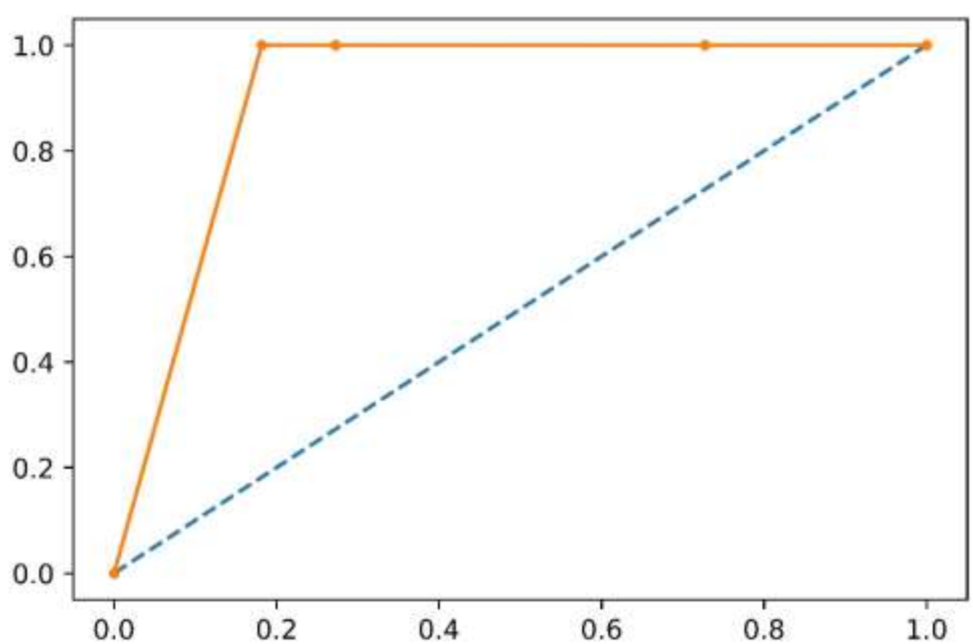

Figure 7: ROC Curve

5.3State of the Art

In this section, various research studies have been selected in quality journals and conferences in the most recent years i.e. 2016, 2017 and 2018 have been analyzed. The research problems, proposed techniques, and accuracies of these research studies have been illustrated in the following table i.e. table 12.

Table 12: Comparison with previous similar research works

Source	Research Problem	Proposed Approach	Dataset	Nature of Data	Accuracy in %
[19]	A hybrid method for diagnosing Coronary Artery Disease has been presented.	Genetic Algorithm and a neural network.	Z-Alizadeh Sani Dataset	Integer/Real	93.85
[21]	An optimization algorithm for the prediction of paroxysmal atrial fibrillation.	SVM and Heart Rate Variability.	Atrial Fibrillation Prediction Database	ECG signals	87.7
[22]	Assisting unspecialized doctors in making a decision regarding heart disease risk level.	Rules-based classifier.	Cleveland	Categorical, Integer, Real	86.7
[26]	This research works to diagnose the risk of heart failure.	Fuzzy analytic hierarchy process for computing global and an ANN classifier.	Cleveland Heart Disease Dataset	Categorical, Integer, Real	91.1
[30]	The research develops and validates prognostic risk models	Contrast Pattern Aided Logistic	EHR data Mayo Clinic	Electronic Health Records	91.4

	for predicting 1, 2, and 5-year survival in HF.	Regression integrated with loss functions.			
[39]	Computer-aided classification of clinical datasets.	The ensemble of ANN, PSO, and Differential Evolution.	UCI	Categorical, Integer, Real	86.66
[42]	Heart risk prediction.	This study proposes and evaluates RETAIN on an enormous dataset for risk prediction.	Cerner Health Facts	Electronic Health Records	82
[47]	Heart failure risk prediction.	Neural networks and Long Short-term Memory network.	-	Electronic Health Records	66.55
[49]	Diagnosis of coronary artery diseases.	The presented model is founded on Random Forest, C5.0, and fuzzy modeling.	UCI	Categorical, Integer, Real	90.50
[62]	Heart Disease Diagnosis	Decision Trees	Z-Alizadeh Sani	Integer, Real	85.38
		Naïve Bayes			83.33
		SVM			86.67
		NN			86.32
[63]	Automatic Detection of CAD	KNN, DT and SVM	Physionet database.	ECG	90

[64]	CAD Detection	SVM	Z-Alizadeh Sani	Integer, Real	84.27
[65]	CAD Classification	NB	Z-Alizadeh Sani	Integer, Real	85
		RF			85
		SVM			83
		KNN			86
		ANN			85
		Ensemble			85

The table above shows that the technique proposed in this manuscript can be placed amongst the best systems of the literature.

5.4Analysis

This thesis is actually an extension of [19] aims to present a hybrid method in order to diagnose coronary artery disease. It makes use of a genetic algorithm and an artificial neural network and implements their technique on Z-Alizadeh Sani Dataset and achieves 93.85% of accuracy. The main objective of this study is to enhance the performance of an artificial neural network by initializing the input weights of attributes with an optimization technique and not randomly assigning weights to it.

Our proposed technique has been able to outclass the previously presented techniques in terms of accuracy. It has been able to achieve the highest accuracy because of the different experimentations carried out in order to analyze their results and hence present a better technique. The fact that four different feature selection techniques i.e. weight by SVM, weight by Gini Index, weight by Information Gain, and weight by PCA have been used for selecting features, and then using four different techniques i.e. Particles Swarm Optimization, Evolution Strategy, Backward and Forward techniques, for weight optimization, has really helped us in analyzing the numerous combinations and their results. We analyze that the features selected by weight by Gini Index are the best ones. We also realize that three of the four weight optimization techniques i.e. PSO, ES, and backward perform well. And hence we have been able to present an ensemble that is based on the average of weights produced by PSO, ES, and backward.

As far as the question that why our technique has been able to produced better results is concerned, numerous studies have already been conducted on the fact that random weight initialization for an artificial neural network can actually influence the performance of the ANN negatively. That is the reason why we conduct different experiments with different feature selection and weight optimization techniques in order to present a better technique.

5.5Summary

In this chapter, the results produced by the proposed technique have been presented in terms of accuracy, sensitivity, and specificity. Different graphs have also been provided to provide the reader with a visual illustration of the performance of the algorithm. Moreover, similar studies carried out in the years 2015, 2016, and 2018 have also been summarized in order to compare their techniques, and performances with the proposed system.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1Introduction

After the literature review, initially proposed a methodology, experimentation, proposed methodology, and results and comparison, this manuscript has finally been concluded. In this chapter, an overview of the conducted research has been presented.

This thesis is dedication towards presenting a reliable computer-aided diagnosis system for coronary artery disease. It aims to efficiently diagnose the stated heart disease, in order to prevent the fatality. For the stated purpose different experimentations have been conducted on the main dataset i.e. Z-Alizadeh Sani Dataset.

6.2Conclusion

One of the leading causes of death all around the world is heart failure, it is seen as a major disease in old, and middle ages. Coronary Artery Disease (CAD) particularly, is termed a widespread cardiovascular illness having high rates of mortality. Clinical diagnosis techniques, such as angiography, are recommended by physicians as they call it the best diagnosis for CAD. On the contrary, it is very costly and possesses side effects. In order to present alternatives to such clinical diagnosis, much research has been conducted in the field of machine learning. Heart disease has gradually become a communal problem of public health, around the globe, it is because of unawareness, unhealthy consumption, and poor lifestyle. Today, its accurate prediction and diagnosis are great challenges for hospitals and practitioners. The development in computing technologies has assisted healthcare facilities in the collection and storage of data for founding clinical decision making. Hospitals, located in many developed countries, collect and store patients' data in the digital and manageable form. Coronary Heart Disease (CHD) is a proven significant public health problem not only in some parts of the world but globally.

Machine Learning (ML) is used to interpret datasets by using computers that acquire knowledge from experiences. ML for health-informatics has emerged as an interdisciplinary science of dealing with healthcare data using complex computational techniques. More often than not the healthcare datasets are colossal, that is where data mining comes into play. Data mining transforms the huge sets of data into information which are later used to make better predictions and decisions.

Numerous studies have been carried out to discover which ML techniques have been used for diagnosing heart disease. For instance, which investigates and compares various data mining classification procedures, and ensembles. ML techniques can ameliorate the results of cardiac arrest forecast, but, there is a further need for research to improve the prediction and generalization of ML techniques.

For the stated purpose Z-Alizadeh Sani Dataset has been used. Our proposed technique has been able to outclass the previously presented techniques in terms of accuracy. It has been able to achieve the highest accuracy because of the different experimentations carried out in order to analyze their results and hence present a better technique. The fact that four different feature selection techniques i.e. weight by SVM, weight by Gini Index, weight by Information Gain, and weight by PCA have been used for selecting features, and then using four different techniques i.e. Particles Swarm Optimization, Evolution Strategy, Backward and Forward techniques, for weight optimization, has really helped us in analyzing the numerous combinations and their results. We analyze that the features selected by weight by Gini Index are the best ones. We also realize that three of the four weight optimization techniques i.e. PSO, ES, and backward perform well. And hence after the experimentations, an ensemble based on the average weights of PSO, ES and Backward weight optimization has been developed in order to assign optimized initial weights to an artificial neural network. The proposed technique has been able to achieve an accuracy of over 94%, which suggests that it can be transformed into a reliable system.

6.3Future Work

The accuracy and sensitivity of the proposed technique are promising, however, there is still room for improvement as far as specificity is concerned. In this day and age where technology can be used to save valuable lives, why stop until profound perfection has been achieved. There is always a window available to ameliorate and this particular system can further be bettered. Its performance can be improved by applying more weight optimization techniques.

For future work, novel techniques of initial weight optimization such as normal distribution, uniform distribution, truncated normal distribution, an orthogonal matrix, and identity matrix can be applied in order to further improve the specificity of the technique. After improving the specificity, an automated system for the diagnosis of CAD can be developed.

6.4Summary

This is the finishing chapter of the manuscript and presents the conclusion and future work of this particular research. In the first section of this chapter i.e. conclusion, the fatality of heart diseases has been discussed. The significance of early detection or diagnosis of heart disease has been highlighted and emphasized, and the use of machine learning in order to assist physicians, clinicians, and other medical experts has been illustrated. Moreover, the future work of this thesis has been presented in the second section.

REFERENCES

[1] J. Webster and R. Watson, "Analyzing the past to prepare for the future: writing a literature review", *JSTOR*, 2002. [Online]. Available: http://www.jstor.org/stable/4132319. [Accessed: 07- Jan- 2019].

[2] D. Budgen and P. Brereton, "Performing systematic literature reviews in software engineering", *Proceeding of the 28th international conference on Software engineering - ICSE '06*, 2006. Available: https://dl.acm.org/citation.cfm?id=1134500. [Accessed 5 January 2019].

[3] P. Brereton, B. Kitchenham, D. Budgen, M. Turner and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain", *Journal of Systems and Software*, vol. 80, no. 4, pp. 571-583, 2007. Available: 10.1016/j.jss.2006.07.009 [Accessed 8 January 2019].

[4] BD, "What is decision support system (DSS)? definition and meaning", *BusinessDictionary.com*, 2019. [Online]. Available: http://www.businessdictionary.com/definition/decision-support-system-DSS.html. [Accessed: 09- Jan- 2019].

[5] D. Power, "A Brief History of Decision Support Systems", *Dssresources.com*, 2019. [Online]. Available: http://dssresources.com/history/dsshistoryv28.html. [Accessed: 10- Jan- 2019].

[6] V. Sauter, "Decision Support Systems (DSS)", *Umsl.edu*, 2019. [Online]. Available: http://www.umsl.edu/~sauterv/analysis/488_f02_papers/dss.html. [Accessed: 23- Jan- 2019].

[7] J. Bresnick, "Understanding the Basics of Clinical Decision Support Systems", *HealthITAnalytics*, 2019. [Online]. Available: https://healthitanalytics.com/features/understanding-the-basics-of-clinical-decision-support-systems. [Accessed: 05- Jan- 2019].

[8] Health. IT, "Clinical Decision Support | HealthIT.gov", *Healthit.gov*, 2019. [Online]. Available: https://www.healthit.gov/topic/safety/clinical-decision-support. [Accessed: 08- Jan- 2019].

[9] X. Luo, Y. Chang and J. Yang, "An Automatic Disease Diagnosis Method Based on Big Medical Data", *2015 2nd International Conference on Information Science and Security (ICISS)*, 2015. Available: 10.1109/icissec.2015.7370979 [Accessed 10 January 2019].

[10] C. Pasupathi and V. Kalavakonda, "Evidence Based health care system using Big Data for disease diagnosis", *2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, 2016. Available: https://ieeexplore.ieee.org/document/7538393. [Accessed 5 January 2019].

[11] S. Siuly and Y. Zhang, "Medical Big Data: Neurological Diseases Diagnosis Through Medical Data Analysis", *Data Science and Engineering*, vol. 1, no. 2, pp. 54-64, 2016.

Available: https://link.springer.com/article/10.1007/s41019-016-0011-3. [Accessed 9 January 2019].

[12] N. Khoury, F. Attal, Y. Amirat, L. Oukhellou and S. Mohammed, "Data-Driven Based Approach to Aid Parkinson's Disease Diagnosis", *Sensors*, vol. 19, no. 2, p. 242, 2019. Available: https://www.mdpi.com/1424-8220/19/2/242. [Accessed 10 January 2019].

[13] J. Ramírez et al., "Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features", *Information Sciences*, vol. 237, pp. 59-72, 2013. Available: https://www.sciencedirect.com/science/article/pii/S0020025509002291. [Accessed 4 January 2019].

[14] M. Li and Z. Zhou, "Improve Computer-Aided Diagnosis With Machine Learning Techniques Using Undiagnosed Samples", *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 37, no. 6, pp. 1088-1098, 2007. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4342802. [Accessed 3 January 2019].

[15] L. Li, J. Ouyang, H. Chen and D. Liu, "A Computer Aided Diagnosis System for Thyroid Disease Using Extreme Learning Machine", *Journal of Medical Systems*, vol. 36, no. 5, pp. 3327-3337, 2012. Available: https://link.springer.com/article/10.1007%2Fs10916-012-9825-3. [Accessed 11 January 2019].

[16] K. Polat, B. Akdemir and S. Güneş, "Computer aided diagnosis of ECG data on the least square support vector machine", Digital Signal Processing, vol. 18, no. 1, pp. 25-32, 2008. Available: https://www.sciencedirect.com/science/article/pii/S1051200407000929. [Accessed 13 January 2019].

[17] Yu, S. and Lee, M. 2012. Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability. Computers in Biology and Medicine. 42, 8 (2012), 816-825.

[18] Davari, D. A. et al. 2017. Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM. Computer Methods and Programs in Biomedicine. 138, (2017), 117-126.

[19] Arabasadi, Z. et al. 2017. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. Computer Methods and Programs in Biomedicine. 141, (2017), 19-26.

[20] Tayefi, M. et al. 2017. hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm. Computer Methods and Programs in Biomedicine. 141, (2017), 105-109.

[21] Boon, K. et al. 2018. Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm III. Computer Methods and Programs in Biomedicine. 153, (2018), 171-184.

[22] Purushottam et al. 2016. Efficient Heart Disease Prediction System. Procedia Computer Science. 85, (2016), 962-969.

[23] Pal, D. et al. 2012. Fuzzy expert system approach for coronary artery disease screening using clinical parameters. Knowledge-Based Systems. 36, (2012), 162-174.

[24] Martis, R. et al. 2012. Application of principal component analysis to ECG signals for automated diagnosis of cardiac health. Expert Systems with Applications. 39, 14 (2012), 11792-11800.

[25] Long, N. et al. 2015. A highly accurate firefly based algorithm for heart disease prediction. Expert Systems with Applications. 42, 21 (2015), 8221-8231.

[26] Samuel, O. et al. 2016. An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. (2016), 163-172.

[27] Mahajan, R. et al. 2017. Improved detection of congestive heart failure via probabilistic symbolic pattern recognition and heart rate variability metrics. International Journal of Medical Informatics. 108, (2017), 55-63.

[28] Mustaqeem, A. et al. 2017. A statistical analysis based recommender model for heart disease patients. International Journal of Medical Informatics. 108, (2017), 134-145.

[29] Bashir, S. et al. 2016. IntelliHealth: A medical decision support application using a novel weighted multi-layer classifier ensemble framework. Journal of Biomedical Informatics. 59, (2016), 185-200.

[30] Taslimitehrani, V. et al. 2016. Developing EHR-driven heart failure risk prediction models using CPXR(Log) with the probabilistic loss function. Journal of Biomedical Informatics. 60, (2016), 260-269.

[31] Fujita, H. et al. 2016. Sudden cardiac death (SCD) prediction based on nonlinear heart rate variability features and SCD index. Applied Soft Computing. 43, (2016), 510-519.

[32] Weng, C. et al. 2016. Disease prediction with different types of neural network classifiers. Telematics and Informatics. 33, 2 (2016), 277-292.

[33] Tay, D. et al. 2015. A novel neural-inspired learning algorithm with application to clinical risk prediction. Journal of Biomedical Informatics. 54, (2015), 305-314.

[34] Altan, G. et al. 2016. A new approach to early diagnosis of congestive heart failure disease by using Hilbert–Huang transform. Computer Methods and Programs in Biomedicine. 137, (2016), 23-34.

[35] Acharya, U. et al. 2015. An integrated index for detection of Sudden Cardiac Death using Discrete Wavelet Transform and nonlinear features. Knowledge-Based Systems. 83, (2015), 149-158.

[36] Masetic, Z. and Subasi, A. 2016. Congestive heart failure detection using random forest classifier. Computer Methods and Programs in Biomedicine. 130, (2016), 54-64.

[37] Sudarshan, V. et al. 2017. Automated diagnosis of congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2 s of ECG signals. Computers in Biology and Medicine. 83, (2017), 48-58.

[38] Y. Zheng, X. Guo, J. Qin and S. Xiao, "Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics", 2018.

[39] Leema, N. et al. 2016. Neural network classifier optimization using Differential Evolution with Global Information and Back Propagation algorithm for clinical datasets. Applied Soft Computing. 49, (2016), 834-844.

[40] J. Zhang, R. Lafta, X. Tao, Y. Li, F. Chen, Y. Luo and X. Zhu, "Coupling a Fast Fourier Transformation With a Machine Learning Ensemble Model to Support Recommendations for Heart Disease Patients in a Telehealth Environment", IEEE Access, vol. 5, pp. 10674-10685, 2017.

[41] M. Aborokbah, S. Al-Mutairi, A. Sangaiah and O. Samuel, "Adaptive context aware decision computing paradigm for intensive health care delivery in smart cities—A case analysis", 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2210670717305000?via%3Dihub. [Accessed: 01- Aug- 2018].

[42] L. Rasmy, Y. Wu, N. Wang, X. Geng, W. Zheng, F. Wang, H. Wu, H. Xu and D. Zhi, "A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set", Journal of Biomedical Informatics, vol. 84, pp. 11-16, 2018.

[43] P. Pławiak, "Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system", Expert Systems with Applications, vol. 92, pp. 334-349, 2018.

[44] B. Bozkurt, I. Germanakis and Y. Stylianou, "A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection", Computers in Biology and Medicine, vol. 100, pp. 132-143, 2018.

[45] J. Tan, Y. Hagiwara, W. Pang, I. Lim, S. Oh, M. Adam, R. Tan, M. Chen and U. Acharya, "Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals", Computers in Biology and Medicine, vol. 94, pp. 19-26, 2018.

[46] F. Miao, Y. Cai, Y. Zhang, X. Fan and Y. Li, "Predictive Modeling of Hospital Mortality for Patients With Heart Failure by Using an Improved Random Survival Forest", IEEE Access, vol. 6, pp. 7244-7253, 2018.

[47] B. Jin, C. Che, Z. Liu, S. Zhang, X. Yin and X. Wei, "Predicting the Risk of Heart Failure With EHR Sequential Data Modeling", IEEE Access, vol. 6, pp. 9256-9261, 2018.

[48] J. Dominguez-Morales, A. Jimenez-Fernandez, M. Dominguez-Morales and G. Jimenez-Moreno, "Deep Neural Networks for the Recognition and Classification of Heart Murmurs

Using Neuromorphic Auditory Sensors", IEEE Transactions on Biomedical Circuits and Systems, vol. 12, no. 1, pp. 24-34, 2018.

[49] S. Mokeddem, "A fuzzy classification model for myocardial infarction risk assessment", Applied Intelligence, 2017.

[50] J. Kim, J. Lee, D. Park, Y. Lim, Y. Lee and E. Jung, "Adaptive mining prediction model for content recommendation to coronary heart disease patients", Cluster Computing, vol. 17, no. 3, pp. 881-891, 2013.

[51] G. Guidi, M. Pettenati, P. Melillo and E. Iadanza, "A Machine Learning System to Improve Heart Failure Patient Assistance", IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 6, pp. 1750-1756, 2014.

[52] K. Raza, "Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule", U-Healthcare Monitoring Systems, vol. 1, pp. 179-196, 2018.

[53] J. Wassan, H. Wang and H. Zheng, "Machine Learning in Bioinformatics", Encyclopedia of Bioinformatics and Computational Biology, vol. 1, pp. 300-308, 2018.

[54] M. Amin, Y. Chiam and K. Varathan, "Identification of significant features and data mining techniques in predicting heart disease", Telematics and Informatics, 2018.

[55] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease", IEEE Computer Society, 2018. [Online]. Available: https://www.computer.org/csdl/proceedings/iscc/2017/1629/00/08024530-abs.html. [Accessed: 19- Dec- 2018].

[56] S. Layeghian Javan, M. Sepehri and H. Aghajani, "Toward analyzing and synthesizing previous research in early prediction of cardiac arrest using machine learning based on a multi-layered integrative framework", Journal of Biomedical Informatics, vol. 88, pp. 70-89, 2018.

[57] Y.-W. Chen , C.-J. Lin , Combining SVMs with various feature selection strategies, in: Feature Extraction: Foundations and Applications, 2006, pp. 315–324 .

[58] K. Fukunaga , Introduction to Statistical Pattern Recognition, Academic press, 2013 .

[59] T. Pang-Ning , M. Steinbach , V. Kumar ,Introduction to Data Mining, Pearson Addison Wesley, Boston, MA, 2006 .

[60] F. Wani, S. Rizvi, M. Sharma and M. Bhat, "A study on cross-validation for model selection and estimation", INTERNATIONAL JOURNAL OF AGRICULTURAL SCIENCES, vol. 14, no. 1, pp. 165-172, 2018.

[61] Russell C. Eberhart and Yuhui Shi, “Comparison between Genetic Algorithms and Particle Swarm Optimization”

[62] J. Brownlee, "An Introduction to Feature Selection", Machine Learning Mastery, 2019. [Online]. Available: https://machinelearningmastery.com/an-introduction-to-feature-selection/. [Accessed: 20- Jan- 2019].

[63] kdnuggets, "What is a Support Vector Machine, and Why Would I Use it?", Kdnuggets.com, 2019. [Online]. Available: https://www.kdnuggets.com/2017/02/yhat-support-vector-machine.html. [Accessed: 19- Jan- 2019].

[64] N. Bambrick, "Support Vector Machines: A Simple Explanation", Kdnuggets.com, 2019. [Online]. Available: https://www.kdnuggets.com/2016/07/support-vector-machines-simple-explanation.html. [Accessed: 19- Jan- 2019].

[65] R. GmbH, "Weight by SVM - RapidMiner Documentation", Docs.rapidminer.com, 2019. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/feature_weights/weight_by_svm.html. [Accessed: 20- Jan- 2019].

[66] W. KENTON, "Gini Index", Investopedia, 2018. [Online]. Available: https://www.investopedia.com/terms/g/gini-index.asp. [Accessed: 20- Jan- 2019].

[67] R. GmbH, "Weight by Gini Index - RapidMiner Documentation", Docs.rapidminer.com. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/feature_weights/weight_by_gini_index.html. [Accessed: 19- Jan- 2019].

[68] R. Kumar, "Understanding Principal Component Analysis – Rishav Kumar – Medium", Medium, 2018. [Online]. Available: https://medium.com/@aptrishu/understanding-principle-component-analysis-e32be0253ef0. [Accessed: 13- Jan- 2019].

[69] R. GmbH, "Weight by PCA - RapidMiner Documentation", Docs.rapidminer.com. [Online]. Available: https://docs.rapidminer.com/8.0/studio/operators/modeling/feature_weights/weight_by_pca.html. [Accessed: 18- Jan- 2019].

[70] IGI Dictionary, "What is Information Gain", www.igi-global.com. [Online]. Available: https://www.igi-global.com/dictionary/information-gain/14407. [Accessed: 18- Jan- 2019].

[71] R. GmbH, "Weight by Information Gain - RapidMiner Documentation", Docs.rapidminer.com. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/feature_weights/weight_by_information_gain.html. [Accessed: 20- Jan- 2019].

[72] A. Ananthram, "Random Initialization For Neural Networks : A Thing Of The Past", Towards Data Science, 2018. [Online]. Available: https://towardsdatascience.com/random-initialization-for-neural-networks-a-thing-of-the-past-bfcdd806bf9e. [Accessed: 20- Jan- 2019].

[73] J. Brownlee, "Why Initialize a Neural Network with Random Weights?", Machine Learning Mastery, 2018. [Online]. Available: https://machinelearningmastery.com/why-initialize-a-neural-network-with-random-weights/. [Accessed: 18- Jan- 2019].

[74] Mnemstudio, "Introduction to Particle Swarm Optimization", Mnemstudio.org. [Online]. Available: http://mnemstudio.org/particle-swarm-introduction.htm. [Accessed: 15- Jan- 2019].

[75] R. GmbH, "Optimize Weights (Evolutionary) - RapidMiner Documentation", Docs.rapidminer.com. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/optimization/feature_weighting/optimize_weights_evolutionary.html. [Accessed: 18- Jan- 2019].

[76] R. GmbH, "Optimize Weights (Forward) - RapidMiner Documentation", Docs.rapidminer.com. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/optimization/feature_weighting/optimize_weights_forward.html. [Accessed: 20- Jan- 2019].

[77] T. Srivastava, "Introduction to Artificial Neural Network : Simplified", Analytics Vidhya, 2014. [Online]. Available: https://www.analyticsvidhya.com/blog/2014/10/introduction-neural-network-simplified/. [Accessed: 20- Jan- 2019].

[78] J. Mahanta, "Introduction to Neural Networks, Advantages and Applications", Towards Data Science, 2017. [Online]. Available: https://towardsdatascience.com/introduction-to-neural-networks-advantages-and-applications-96851bd1a207. [Accessed: 19- Jan- 2019].

[79] S. Asaithambi, "Why, How and When to Scale your Features – GreyAtom – Medium", Medium, 2017. [Online]. Available: https://medium.com/greyatom/why-how-and-when-to-scale-your-features-4b30ab09db5e. [Accessed: 18- Jan- 2019].

[80] Benalexkeen, "Feature Scaling with scikit-learn – Ben Alex Keen", Benalexkeen.com. [Online]. Available: http://benalexkeen.com/feature-scaling-with-scikit-learn/. [Accessed: 17- Jan- 2019].

[81] A. Sharma V, "Understanding Activation Functions in Neural Networks", Medium, 2017. [Online]. Available: https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0. [Accessed: 17- Jan- 2019].

[82] T. Babs, "The Mathematics of Neural Networks – Coinmonks – Medium", Medium, 2018. [Online]. Available: https://medium.com/coinmonks/the-mathematics-of-neural-network-60a112dd3e05. [Accessed: 19- Jan- 2019].

[83] D. GUPTA, "Fundamentals of Deep Learning - Activation Functions and their use", Analytics Vidhya, 2017. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/10/fundamentals-deep-learning-activation-functions-when-to-use-them/. [Accessed: 19- Jan- 2019].

[84] V. Bushaev, "Adam—latest trends in deep learning optimization.", Towards Data Science, 2018. [Online]. Available: https://towardsdatascience.com/adam-latest-trends-in-deep-learning-optimization-6be9a291375c. [Accessed: 21- Jan- 2019].

[85] S. Ruder, "An overview of gradient descent optimization algorithms", Sebastian Ruder, 2016. [Online]. Available: http://ruder.io/optimizing-gradient-descent/index.html#adam. [Accessed: 21- Jan- 2019].

[86] D. Godoy, "Understanding binary cross-entropy / log loss: a visual explanation", Towards Data Science, 2018. [Online]. Available: https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a. [Accessed: 20- Jan- 2019].