OPTIMIZING ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS) WITH DRAGONFLY ALGORITHM FOR CARDIOVASCULAR DISEASE

WADA MOHAMMED JINJIRI

UNIVERSITI SAINS MALAYSIA

2023

OPTIMIZING ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS) WITH DRAGONFLY ALGORITHM FOR CARDIOVASCULAR DISEASE

by

WADA MOHAMMED JINJIRI

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

June 2023

ACKNOWLEDGEMENT

First and foremost, I am thankful to the Almighty ALLAH for giving me the opportunity, strength, and determination for the completion of this thesis. I am very grateful to express my heartiest gratitude to my supervisor Dr Pantea Keikhosrokiani for her skilful guidance, technical approach and inspiring attitude which helped me in the right direction for accomplishing this research.

I also like to express my gratitude to my Co-supervisor Dr Nasuha Lee Abdullah for her beneficial assistance and art of making useful suggestions for the success of this thesis. I also extend my gratitude to the Examiners of my research Prof. Madya Dr. Umi Kalsom Yusof and Prof. Madya Dr. Nasriah Zakaria for their necessary suggestions towards the successful completion of this research. Also, I extend my appreciation to the entire academics and staff of the School of Computer Science Universiti Sains Malaysia (USM) for their guidance and support throughout this period of study.

I would also like to thank my parents for their love, support, assistance, and prayers throughout my entire life. I also like to express my appreciation to my beloved brothers and sisters for being supportive and encouragement. I would like to thank my friends, especially Abdulmajid Babangida Umar, Abdulkadir Ahmad and Abdulrauf Garba for their support and for helping out each other.

Finally, special thanks to the Tertiary Education Trust Fund (TETFUND) for granting me the scholarship.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ABC	Artificial Bee Colony (Optimizer)
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural network
CSO	Chicken Swamp Optimization
CVD	Cardiovascular Disease
DA	Dragonfly Algorithm
DT	Decision Tree
FIS	Fuzzy Inference System
GD	Gradient Descent
GA	Genetic Algorithm
GENFIS	Generated Fuzzy Inference System
GWO	Grey Wolf Optimization
KNN	K-Nearest Neighbour
LR	Logistic Regression
LR	Linear Regression
MAE	Mean Absolute Error
MSE	Mean Square Error
NB	Naïve Bayes
NFS	Neuro-Fuzzy System
PSO	Particle Swarm Optimization
RMSE	Root Mean Squared Error
SVM	Support Vector Machine
ML	Machine Learning

MENGOPTIMUMKAN SISTEM INFERENS KABUR NEURO ADAPTIF (ANFIS) DENGAN ALGORITMA PEPATUNG UNTUK PENYAKIT KARDIOVASKULAR

ABSTRAK

Penyakit kardiovaskular (CVD) merupakan satu kebimbangan besar dalam bidang penjagaan kesihatan. CVD adalah punca utama kadar kematian yang tinggi di seluruh dunia. Penyelidikan ini menggunakan sistem inferens neuro-fuzzy Adaptive (ANFIS) dan menangani masalah topologi dan konfigurasi parametrik yang membawa kepada ralat ramalan untuk CVD. Oleh itu, pendekatan yang dicadangkan akan mengoptimumkan premis dan parameter akibat ANFIS menggunakan algoritma pepatung dengan mengemas kini dua lapisan ANFIS dan meminimumkan ralat ramalan untuk ramalan penyakit kardiovaskular yang cekap. Dataset CVD yang digunakan untuk menguji model ramalan ini diperolehi dari repositori Kaggle. Dapatan dari kajian ini juga diukur menggunakan metrik seperti min ralat mutlak (MAE), min kuasa dua ralat (MSE), dan punca min ralat kuasa dua (RMSE). Antara model pembelajaran mesin untuk regresi yang digunapakai dalam kajian ini termasuk DT, LR, LR, KNN, RF dan SVR. Hasil analisis menunjukkan SVR dan LR memberi keputusan yang terbaik iaitu dengan nilai ralat ramalan yang terendah dengan catatan masing-masing: - MAE = 0.42693, MSE = 0.24656, RMSE = 0.48753 (SVR) dan MAE = 0.41114, MSE = 0.23602, RMSE = 0.48638 (LR). Semetara itu, model ANFIS juga dilatih dengan pengelompokan subtraktif untuk model regresi dan kemudian dioptimumkan menggunakan algoritma pepatung, ANFIS-DA. Keputusan eksperimen ini memberi nilai MEA = 0.05921, MSE = 0.09009, RMSE = 0.10066 untuk data latihan, dan MEA = 0.03088, MSE = 0.22197, RMSE = 0.10127 untuk data ujian. Ia

membuktikan Model ANFIS-DA yang dicadangkan menghasilkan ramalan yang lebih tepat berbanding dengan pendekatan pengoptimuman ANFIS-PSO, ANFIS-GA dan ANFIS-GWO.

OPTIMIZING ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (ANFIS) WITH DRAGONFLY ALGORITHM FOR CARDIOVASCULAR DISEASE

ABSTRACT

Cardiovascular disease (CVD) remains a great concern in the field of healthcare. It is responsible for the highest mortality rate leading cause of death worldwide. This research utilizes an Adaptive neuro-fuzzy inference system (ANFIS) and addresses the problem of topology and parametric configurations that lead to the prediction error for CVD. Therefore, the proposed approach will optimize the premise and consequent parameters of ANFIS using the dragonfly algorithm by updating the two ANFIS layers and minimizing prediction error for efficient cardiovascular disease prediction. A cardiovascular disease dataset obtained from the Kaggle repository is used to validate the models. The results are also evaluated using metrics such as mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). Several machine learning models for regression such as DT, LR, LR, KNN, RF and SVR are also utilized. The results from analysing these algorithms observed the SVR with MAE = 0.42693, MSE = 0.24656, RMSE = 0.48753 and LR with MAE = 0.41114, MSE = 0.23602, RMSE = 0.48638 performs the best the lowest prediction error. Moreover, the ANFIS model is trained by subtractive clustering for the regression model and then optimized using the dragonfly algorithm. The experimental results show the estimation of error loss after optimization with MEA = 0.05921, MSE = 0.09009, RMSE = 0.10066 for training data, and MEA = 0.03088, MSE = 0.22197, RMSE = 0.10127 for testing data respectively. The proposed ANFIS-DA model compared with other optimization approaches proved a lower prediction error than ANFIS-PSO, ANFIS-GA, and ANFIS-GWO.

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Cardiovascular disease (CVD) is involving the heart or blood vessels. It is a series of diseases relating to the circulatory system that includes myocardial infarction, coronary heart disease, and heart failure. Consequently, it also covers a wide range of disorders such as cardiac muscle disease and the vascular supply system for the heart, brain, and other vital organs (Ortega, Lavie, and Blair 2016). CVD is the biggest cause of death in the world, according to the World Health Organization (WHO). In a recent study, it was estimated that 17.7 million deaths had occurred worldwide as a result of CVD (WHO 2021) presenting 31% of all world losses. However, the cause of CVD is due to some of the risk factors, such as being overweight, physical inactivity, age, tobacco use, stress, harmful use of alcohol, depression, and inheritance unhealthy diet. Thus, over 80% of cardiovascular disease can be prevented if it can be detected early and increase the chance of a successful treatment (Aribarg, Supratid, and Lursinsap 2009). Several studies at this time are trying to apply machine learning and statistical algorithms in predicting the risk of getting the disease and preventing patients from suffering a heart attack.

However, it is necessary to understand the risk factors to improve diagnosis. In the traditional approach, the patient's symptoms and medical history are analysed, for example, ECG, blood sugar levels, blood pressure, and cholesterol levels. This is a time-consuming and expensive procedure. Therefore, it is necessary to detect people with a risk of cardiovascular diseases. The process becomes simpler through the use

of soft computing techniques, which saves time and therefore improves diagnosis efficiency (Kibria and Matin 2022).

Various health symptoms and habits that contribute to cardiovascular disease are documented electronically in health records. Healthcare databases have collected a vast amount of cardiovascular disease datasets that comprise many features that may be irrelevant and or/redundant, unbalanced data, or the volume of data because of the disease complexity (Indhumathi and Kumar 2021), these negatively affect the performance of the model and is difficult to diagnose. Yet, human expertise in healthcare may also sometimes result in incorrect predictions of the disease (Chatterjee, Cymberknop, and Armentano 2017). As a result, to conduct a successful analysis, different methods are utilized for predicting cardiovascular disease. Researchers have implemented numerous models to diagnose CVD precisely for a long time to prevent patients from suffering from the disease.

Today, with the advancement of technology especially for machine learning and computational intelligence techniques, these technologies assist a diagnosis of a medical judgment (Sikchi, Sikchi, and Ali 2013). Thus, artificial intelligence offers a computer able to learn and accomplish a task from a set of given data. As an intelligent system, a machine learning approach is used to know the meaning of data reliably and suggest an appropriate output from raw data for several resolutions (Injadat et al. 2021). This makes an available means of analysing a vast amount of data to discover patterns and relationships among various entities that are undetectable without means of advanced analysing techniques (Abeykoon et al. 2016).

Moreover, machine learning has become a widespread method for analysing cardiovascular disease predictions (Krittanawong et al. 2020). Despite the existence of numerous machine learning techniques, determining the best suitable approach that is feasible for cardiovascular disease datasets remains a challenge (Nadakinamani et al. 2022). Therefore, two machine learning methods (fuzzy logic and neural network) combined into a single technique known as an Adaptive neuro-fuzzy inference system (ANFIS) that is developed by (Jang, Sun, and Mizutani 1993) is utilized.

In general, the adaptive neuro-fuzzy inference system (ANFIS) is among the conventional neuro-fuzzy inference expert technique that utilizes the Takagi-Sugeno fuzzy inference framework, As reported by (Opeyemi and Justice 2012) ANFIS provides a method for acquiring data information through a fuzzy modelling process to build membership function parameters that allow the associated fuzzy inference system to track the given input/output data. The membership function parameters of ANFIS to be optimized are the premise parameters. These parameters define the shape of the membership functions (Walia, Singh, and Sharma 2015). The ANFIS uses the features of the dataset to learn and modifies the system parameters based on an error criterion (Ewees and Abd Elaziz 2020).

Though, it was disclosed that ANFIS learns from neural networks and uses fuzzy logic reasoning towards handling a variety of non-linear and complicated issues with high accuracies (Agrawal and Ashtankar 2013; Salleh, Hussain, and Talpur 2019; Wang et al. 2015). Thus, achieving an improved accuracy for CVD prediction is quite challenging considering that many of the ANFIS parameters are fuzzy variables and that many times the problem of ANFIS topology and parametric configuration

becomes a complicated task with the existing ANFIS model that make the predictions unfeasible (Salleh, Talpur, and Hussain 2017).

To overcome these problems, hybrid strategies are employed to enhance the prediction performance of medical diagnosis. Amongst the hybrid approaches, the evolutionary algorithm (EA) is one of the soft-computing techniques integrated to generate these hybrid models, they are powerful search and optimization paradigms influenced by a biological mechanism. Evolutionary algorithms (EA) are population-based and inspired by animals, insects, or birds' biological behaviours that can minimize the occurrence of prediction errors by finding an optimal solution to complex problems (Shahid and Singh 2020). Quite a lot of researchers have utilized some of the EA such as Genetic algorithm (GA), Particle swarm optimization (PSO), Artificial bee colony (ABC), and Greywolf optimization (GWO) to improve ANFIS for various disease predictions with an improved performance result.

Therefore, this research utilized machine learning and computational intelligence techniques by introducing a hybrid ANFIS technique for cardiovascular disease prediction with the optimization power of swarm intelligence as a significant concern in the process of the results generated by the healthcare information system. thus, the ANFIS model is integrated with a metaheuristic approach to optimize the process parameters for more optimal performance with improved cardiovascular disease prediction.

1.2 Motivation

Cardiovascular disease (CVD) is a disease with about 80% of survival chances of prevention if identified in its early stage. With the recent increase and exploitation of computational intelligence methods, they can assist medical organizations to deploy the adaptive neuro-fuzzy inference system (ANFIS) to perform regression and prediction of such diseases. Even though several works that use ANFIS for diagnosis has presented promising result, but still, there's a need for effective parameter training. With advancements in computational intelligence techniques, these technologies can be employed to assist in medical diagnosis.

Furthermore, ANFIS can be regarded as a local linearization model for estimation, it has broad applicability in system modelling. The efficient implementation of ANFIS-based models necessitates effective parameter training for increased accuracy. It has a significant chance of falling into local minima, which uses derivative-based learning. In this case, the derivative-free methods using metaheuristic algorithms are more potent. Therefore, to achieve maximum results with improved prediction accuracy, this gives the motivation to propose an optimized ANFIS using an evolutionary algorithm to tunned its learning parameters. Thus, the work of this research performed an empirical evaluation of computational intelligence approaches and discussed the conclusions from these findings.

1.3 Problem Statement

Cardiovascular disease (CVD) is one such ailment that has recently gained a lot of attention from researchers as it accounts for 33% of deaths worldwide annually. Physicians have estimated prognosis by subjectively integrating the patient's looks,

medical indicators, and laboratory investigations as said by (Rossello et al. 2019). Precise predictions of cardiovascular disease are critical for diagnosis. Researchers day by day remains self-motivated to develop an effective decision system as cardiovascular disease remains a challenge (Nadakinamani et al. 2022).

This research aims to improve the ANFIS model by minimizing prediction errors using a swarm intelligence technique. Predicting cardiovascular disease may be considered an optimization problem, where various computational intelligence models have been implemented to improve prediction accuracies (Richard and Mala 2018).

ANFIS is a combination of two machine learning "fuzzy logic and neural network" techniques, and its parameters are fuzzy variables. The problem of topology and parametric configurations in ANFIS turns out to be a complicated task (Salleh et al. 2017). The tunning of ANFIS and its configuration is generally thru trial and error and most of the researcher's concern is efficient parameter training that achieves efficient results. Thus, researchers agreed that tunning membership functions (MF) parameters are more complex (Liu, Leng, and Fang 2013; Petković et al. 2014).

Researchers (Sagir and Sathasivam 2017) disclosed that ANFIS uses gradient descent to learn its parameters which as a result affected by the initial point. The problem of obtaining a suitable parameter for training the memberships function, and weights between the ANFIS layers are eliminated, these resulted in getting stuck into the local point (Song, Chen, and Antoniou 2021). However, these problems may lead to premature convergence or computational complexity. To achieve these several researchers have proposed techniques that tuned ANFIS training parameters using

different optimization techniques to solve the prediction problems and achieve better accuracy (Hussain, Salleh, and Leman 2019).

To overcome these issues, population-based computation techniques are employed due to their capability of finding global optimum solutions with a less computational cost for optimization problems. These methods are population-based and inspired by animals, insects, or birds' biological behaviours that can minimize the occurrence of prediction errors by finding an optimal solution to complex problems (Shahid and Singh 2020). Quite a lot of researchers have utilized some of these techniques such as genetic algorithm (GA) (Feng, Wang, and Li 2021), beetle swarm optimization (BSO) (Singh et al. 2021), moth-flame optimization (MFO) (Canayaz 2019), and particle swarm optimization (PSO) (Goldar, Ghiasi, and Badamchizadeh 2020) are utilized to improve ANFIS for various disease predictions.

Accordingly, this study aims to address the above-mentioned problems to propose an enhanced technique that efficiently predicts cardiovascular disease.

1.4 Research Questions

Several research questions are highlighted based on the problem statement. The question this study attempts to answer are:

1. How to minimize cardiovascular disease prediction errors?
2. How to enhance the parameters of the adaptive neuro-fuzzy inference system (ANFIS)?

1.5 Objectives of the Research

The key objective of this research is to perform comprehensive machine learning analysis, prediction, and optimization by using computational intelligence methods. Therefore, to achieve this, the research objectives are as follows:

1. To propose a model to minimize cardiovascular disease (CVD) prediction errors using the adaptive neuro-fuzzy inference system (ANFIS).
2. To optimize the parameters of ANFIS by using a dragonfly algorithm (DA).

1.6 Scope of the Research

This research depends on the benefit of some computational intelligence methods from machine learning (fuzzy logic and neural networks) and swarm-intelligence techniques that focus on developing a model that efficiently enhances the predictive performance of diagnosing cardiovascular disease (CVD). These approaches are utilized and analysed to observe a method that can efficiently diagnose cardiovascular disease. In general, the metaheuristic dragonfly algorithm (DA) is used and enhance learning parameters (Consequent and Premise) of the adaptive neuro-fuzzy inference system (ANFIS) model, these were amongst the popular techniques used in the medical domain that has been chosen to achieve the objectives of this research.

1.7 Significance of the Study

Enhanced prediction of cardiovascular disease is regarded as one of the most significant focuses of this research for healthcare data analysis. The findings are beneficial in healthcare applications due to the greater demand for the diagnosis of CVD which justifies the need for practical life-changing approaches. Today, Soft-computing technologies that can perform data analysis and modelling help clinicians

achieve the best and most timely clinical judgments required in healthcare applications.

To ensure health significance has promptly assessed, the diagnosis of the cardiovascular disease reached prime importance that vows to the use of computational-based technological models for the assessment of health risk. Thus, this research came up with an integrated computational intelligence ANFIS-DA technique that can improve the effectiveness and increases prediction accuracies. With this advancement, physicians can quickly predict and diagnose their patients, saving costs and time with more effective results.

1.8 Expected Contributions

This research contributes to the area of computational intelligence for the medical healthcare domain, by developing machine learning models that can predict cardiovascular disease from input data to output results. Thereby helping physicians and medical health centres to quickly detect the disease in a minimal amount of time for early diagnosis.

The proposed work contributed by introducing a new approach for predicting cardiovascular disease from a large set of clinical data. Several machine learning methods are employed to test their reliability and analyse their efficiency towards predicting cardiovascular disease. Moreover, this research is expected to contribute to the current understanding of how optimization process stages can be effectively managed and achieved. Therefore, an optimized ANFIS-DA technique is expected to

overcome the prediction error with the idea of fine-tuning parameters to achieve maximum results for cardiovascular disease prediction.

1.9 Thesis Organization

This thesis consists of five chapters. It begins with this introductory chapter that presents the background, motivation, problem statement, research questions, research objectives, scope, significance, and contributions that are expected from this research.

Chapter 2: Discuss relevant computational approaches in the literature, guiding the reader through the resources of this research topic, particularly those connected to the components of this work. Specifically, it gives an overview of the ideas of relevant theories behind the classification and prediction processes and neuro-fuzzy systems as well as evolutionary approaches and hyperparameter optimization processes that are proven towards enhancement in this research.

Chapter 3: Introduced the proposed methodological framework that is used for this research. The technique introduces the components and their relationships to explain how the proposed solution is built using a set of processes and flows at each particular stage.

Chapter 4: Discussed and presents the implemented results in a series of charts, tables, and figures with proper and detailed explanations and processes that are involved for each stage that is initiated. Also, the results are then compared and discussed the outcome in a detailed manner.

Chapter 5: This chapter concludes this research by summing the concise comments on the findings and, also discussed the contributions and recommendations for future work to explore other techniques that may increase the model performance.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter provides an overview of the relevant literature for this research. It gives an overview of data mining and soft computing technology that have been carried out by various researchers. The chapter discusses some machine learning methods in the medical healthcare domain. Also, the chapter discusses the predictive modelling of an adaptive neuro-fuzzy inference system (ANFIS) and its applications from previous works by researchers. On the other hand, this chapter confers the metaheuristic approaches for several optimization problems that had been and provide some comparative performance of these works done. Finally, the chapter concluded with a discussion and summary. Also, shown in figure 2.1 is the structural organization of the chapter.

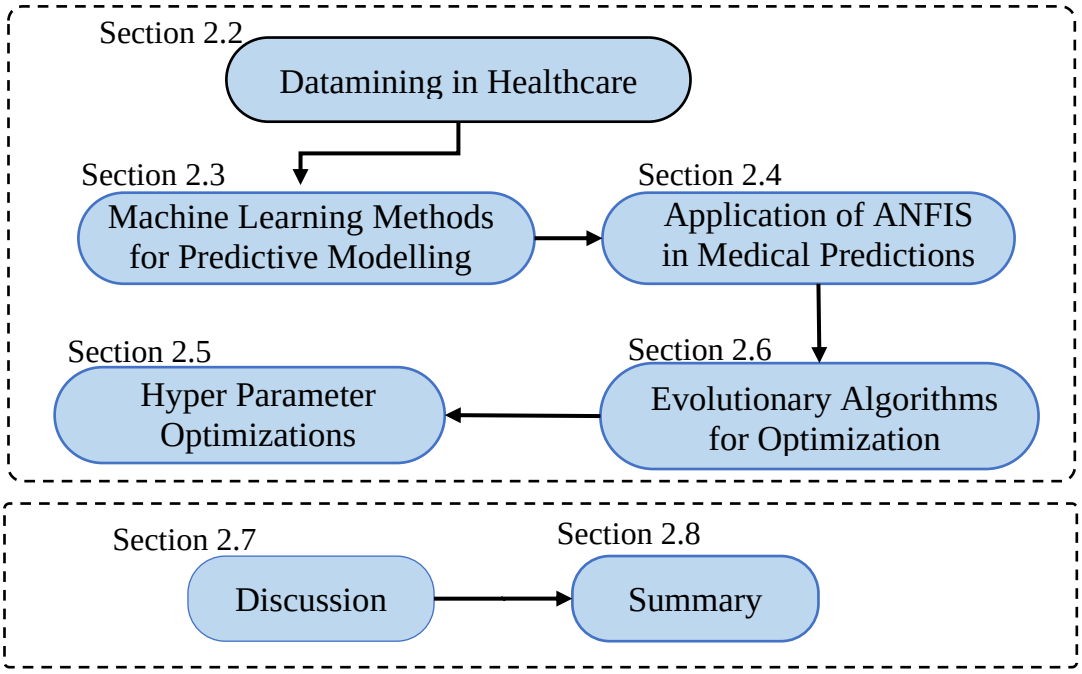

Figure 2.1 Structural organization of the chapter

2.2 Data Mining in Healthcare

Healthcare data mining has become one of the most prominent research areas nowadays. It involves finding meaningful information and identifying patterns in data. It is possible to predict data trends using various data mining approaches, which empower specialists to make judgments and predictions with certainty. Many algorithms are applied in data mining techniques to solve real-life problems (Madni, Anwar, and Shah 2017). Healthcare data mining is an important element of extracting patterns and information from previously unknown clinical data, by identifying patterns and trends from large complex data, data mining can improve decisions.

The need for clinical analysis has become increasingly essential to make informed decisions (Nega and Kumlachew 2017). Doctors often make clinical decisions based on their intuition and experience, not on databases containing knowledge-rich information. Thus, patients' quality of care is compromised by bias, errors, and excessive medical costs associated with this practice (Lashari et al. 2018). Therefore, data mining can help with the prediction and determination of diseases in this area.

There are numerous data mining techniques and applications with a vast potentiality of solutions that can assess the efficacy of medical therapies by comparing and contrasting causes and symptoms, as well as identifying effective interventions for a certain disease (Shouman, Turner, and Stocker 2012). Thus, cardiovascular disease is one of the prevalent diseases caused by many factors, and with the rapid growth of digital technologies, health centres store a huge amount of data that is very complex and challenging to analyse (Beyene and Kamat 2018).

With the application of data mining, cardiovascular disease prediction can be taken to a new level by identifying and extracting useful information from a huge clinical dataset with minimal user input and effort. An accurate prediction of cardiovascular diseases is possible through effective data mining implementation in healthcare (Amin, Chiam, and Varathan 2019). Moreover, there have been reports of successful data mining applications in the health sector that have vast potential. For example, data mining applications can evaluate the effectiveness of medical treatments by comparing and contrasting the causes and symptoms, and also identify successful standardized treatments for specific diseases. These can also help healthcare management by identifying and tracking the conditions and risks of the disease in patients (Shafique et al. 2015).

A survey by Pavithra & Jayalakshmi, (2019) aimed to assess the current state of data mining techniques in health care that can be used to predict cardiovascular diseases with the means of increasing accuracy for effective diagnosis. Similarly, several data mining techniques are utilized for comparative study by observing an efficient method that predicts and diagnose heart disease effectively (Raju et al. 2018). Also, two data mining techniques were used to predict the presence or absence of chronic heart disease in patients at its early stage based on various health parameters to help reduce the risk of the disease (Nalluri et al. 2020).

Therefore, healthcare data mining is important, but a challenging task that must be accomplished precisely, developing a solution for predicting the presence and absence of CVDs through data mining is urgently needed to extract useful knowledge from clinical data (Martins et al. 2021). With the use of data mining tools and machine

learning approaches, healthcare organizations can identify significant patterns and detect correlations and relationships between multiple variables within huge databases (Sivakami and Prabhu 2019). A wide range of medical problems can be solved using data mining, by employing machine learning methods that can perform analysis, classification, and regression (Shafique et al. 2015).

2.3 Machine Learning Methods for Predictive Modelling

Today, we are living in the age of algorithms, in which machine learning has become global and essential for solving complex problems in the medical domain. A large amount of heterogeneous data and information is generated daily by healthcare providers, making it impossible to analyse and process it using traditional methods. To gain actionable insights from this data, machine learning methods are used (Qayyum et al. 2020). In its most basic form, machine learning is a study of computer algorithms that employ several statistical, probabilistic, and optimization methods to learn from experience and detect useful patterns from unstructured and complex data (Uddin et al. 2019). Thus, these methods had become more practical in the area of medical predictions by using data fed into them for more accurate analysis.

The use of machine learning involves a variety of learning approaches including supervised as well as unsupervised learning which are differentiated according to the presence of feedback. These approaches had contributed to solving diverse medical problems ranging from disease identification, diagnosis, and analysis (Smiti 2020). Numerous researchers have worked on various machine learning methods and agreed these algorithms work well in solving various medical problems (Al-Janabi, Qutqut, and Hijjawi 2018).

To constrain the discussion, develop concrete examples, and represent the vast majority of clinical machine learning research, this research focuses on the supervised learning approach. Supervised learning is a learning mechanism that is trained by a machine using a labelled dataset in the form of a compressed input/output pair, these datasets are available in discrete or continuous form (Dai et al. 2015). Supervised learning has its importance in the medical field concerning heart disease scenarios where inputs can be symptoms of heart disease such as cholesterol, blood pressure, chest pain, etc. and the output can be suffering from the disease or not, since there were clearly defined variables in this study, supervised learning was employed.

Several researchers have implemented different machine-learning methods for the prediction and diagnosis of heart-related diseases. For example, Support Vector Machine (SVM), Naïve Bayes (NB), Logistic Regression (LR), Random Forest (RF), and Gradient Boosting (GR) are used for anticipating cardiovascular disease illness, the outputs and the accuracies generated are reviewed for each method (Dinesh et al. 2018). Researchers Basharat et al., (2016) applied a supervised learning approach to classify unstructured data of cardiac patients, using logistic regression, they classified unstructured data into multinomial class levels to make intelligence decisions for predictive analysis.

In a study performed by (Maini, Venkateswarlu, and Gupta 2018) where they proposed a Decision Tree (DT), Naïve Bayes (NB), Artificial Neural Network (ANN), and Boosting Ensemble for cardiovascular disease prediction. Cleveland heart disease dataset is used in evaluating their system performance for each of the techniques that are used, though the performance of these techniques where judge based on precision,

recall, and f1_score parameters. The best performance method from their result is the ensemble and recommended for cardiovascular disease prediction.

Similarly, a study by N. K. Kumar et al., (2020) analyses and predicts cardiovascular disease using machine learning trees classifier techniques such as decision tree, random forest, k nearest neighbour, logistic regression, and support vector machine which were broken down depending on their precision and ROC AUS score. Their investigation of foreseeing cardiovascular disease has achieved a higher precision with a random tree classifier. Also, researchers Arumugam et al., (2021) in their work performed multiple disease predictions amongst a cardiovascular disease by employing a Support Vector Machine, Naïve Bayes, and a Decision Tree approach towards making a significant effort to improve the diagnostic process efficiency. The outcome of their work suggests the decision tree technique for optimum performance in predicting cardiovascular disease in diabetic patients over the other two techniques.

Table 2.1 summarises some of the machine learning methods that have been utilized by various researchers to forecast and diagnose heart-related diseases.

Table 2.1 Applications of Machine Learning Techniques in Healthcare

Application	Title	Author/s	ML technique/(s)	Best technique/(s)	Dataset
Heart disease detection	Application of machine learning for the detection of heart disease.	(Yadav et al. 2020)	K-means, KNN, NB, LR, Fuzzy, NN	NN	Cleveland heart disease
CVD Prediction	A Cardiovascular disease prediction using machine learning algorithms.	(Rubini et al. 2021)	LR, RF SVM, NB,	RF	Framingham & UCI heart disease
Heart disease prediction	Heart disease prediction using machine learning techniques.	(Shah, Patel, and Bharti 2020)	DT, SVM, KNN, NB	KNN	UCI heart disease
Heart disease	Using machine learning for heart disease prediction.	(Salhi, Tari, and Kechadi 2020)	NN, SVM, KNN	NN	EHS Hospital Algeria
CVD prediction	Clinical data analysis for prediction of CVD using ML techniques.	(Nadakinamani et al. 2022)	NB, J48, RT, JRIP	RT	Hungarian & Statlog data
Coronary HD prediction	Prediction of coronary heart disease using machine learning.	(Gonsalves et al. 2019)	SVM, NB, DT	NB	South African heart disease
Coronary artery diagnosis	Machine learning-based coronary artery disease diagnosis.	(Alizadehsani et al. 2019)	ANN, NB, SVM, KNN, C4.5,	ANN, SVM	67 datasets used
Heart disease prediction	Heart disease prediction using machine learning techniques	(Shah et al. 2020)	DT, LR, NB, RF	RF	UCI heart disease
Heart diseases prediction	Heart disease prediction using machine learning algorithms	(Jindal et al. 2021)	LR, RF, KNN	KNN	UCI heart disease
Cardiac Prediction	Prediction of cardiac disease using supervised Machine Learning algorithms.	(Princy et al. 2020)	NB, DT, LR, RF, SVM, KNN	DT	Kaggle CVD data

However, technological advancements have helped to generate an overwhelming amount of clinical data, which is now difficult for standard machine learning algorithms to manage (Dash et al. 2019). Most of these clinical data are fuzzy by their nature, and conventional algorithms are less likely to forecast the problem than hybrid models for the stated task. Existing studies have shown that machine learning algorithms in medical predictions are needs an in-depth investigation necessary to identify techniques that will improve prediction performance (Princy et al. 2020). Thus, one of the hybrid neuro-fuzzy inference expert systems that can capture the benefits of both neural networks and fuzzy inference systems is the adaptive neuro-fuzzy inference system (ANFIS). ANFIS has been utilized in most medical applications because of its efficiency in forecasting and optimization capabilities which makes it the most preferred architecture (Kour, Manhas, and Sharma 2020).

Therefore, the next Section 2.4, discusses predictive procedures and applications that are been utilized by ANFIS for heart-related and other diseases in the medical domain.

2.4 Application of ANFIS for Prediction

The development of an expert system, especially an adaptive neuro-fuzzy inference system (ANFIS) method over the past few decades, has made a crucial role in complex and uncertain medical tasks such as diagnosis and prediction of disease. (Feng et al. 2021; Rivera, Rodriguez, and Yu 2019). The most important advantages of these systems are the expression of human knowledge using special linguistic concepts and fuzzy rules, nonlinearity and compatibility, and better accuracy of these methods in terms of data constraints compared to other methods (Karaboga and Kaya

2019; Yadollahpour et al. 2018). The combination of fuzzy systems, which are based on logical rules and artificial neural networks with the ability to knowledge acquisition from numerical information, enables to use of human knowledge in the construction of the prediction model (Korzhakin and Sugiharti 2021).

ANFIS is one of the hybrid neuro-fuzzy expert systems that was developed by Jang et al., (1993). ANFIS provides a mechanism for learning data knowledge to compute membership function parameters, which allows the connected fuzzy inference system to monitor the input/output data (Ashadevi, Selvi, and Revathi 2017). There are several ANFIS systems based on the neuro-fuzzy system that are being implemented to attain desired outcomes in modelling nonlinear functions. Thus, the core feature of ANFIS is the fuzzy inference system (FIS), which uses 'if-then' rules to predict the behaviour of any uncertain system (Tiwari et al. 2020).

To begin with the modelling process, ANFIS initiates by obtaining a set of pre-processed data. The dataset contains desired input/output data pairs for the target system and is split into training and testing sets, the best and most commonly used split proportion is 70/30 or 80/20 training (Song et al. 2021). Next is to define the ANFIS structure and set the initial parameters for learning by generating the initial fuzzy inference system (FIS) configuration from functions (Genfis1, Genfis2 or Genfis3) that is initially referring to grid-partitioning, subtractive clustering, and fuzzy c-means methods. Genfis1 generates the FIS structure from a training dataset using a grid-partitioning, the number of membership function (MFs) types, and input and output membership functions can also be specified.

On the other hand, Genfis2 generates an initial ANFIS structure for training by implementing the subtractive clustering on the dataset, this function does this by extracting a set of rules that modelled the behaviour of the system represented by the dataset. The rule extraction method determines the number of rules and antecedent membership functions and then uses linear least-squares estimation to determine each rule's consequent equations. As shown in Figure 2.2, is a progression of ANFIS modelling for most predictive applications in a medical domain.

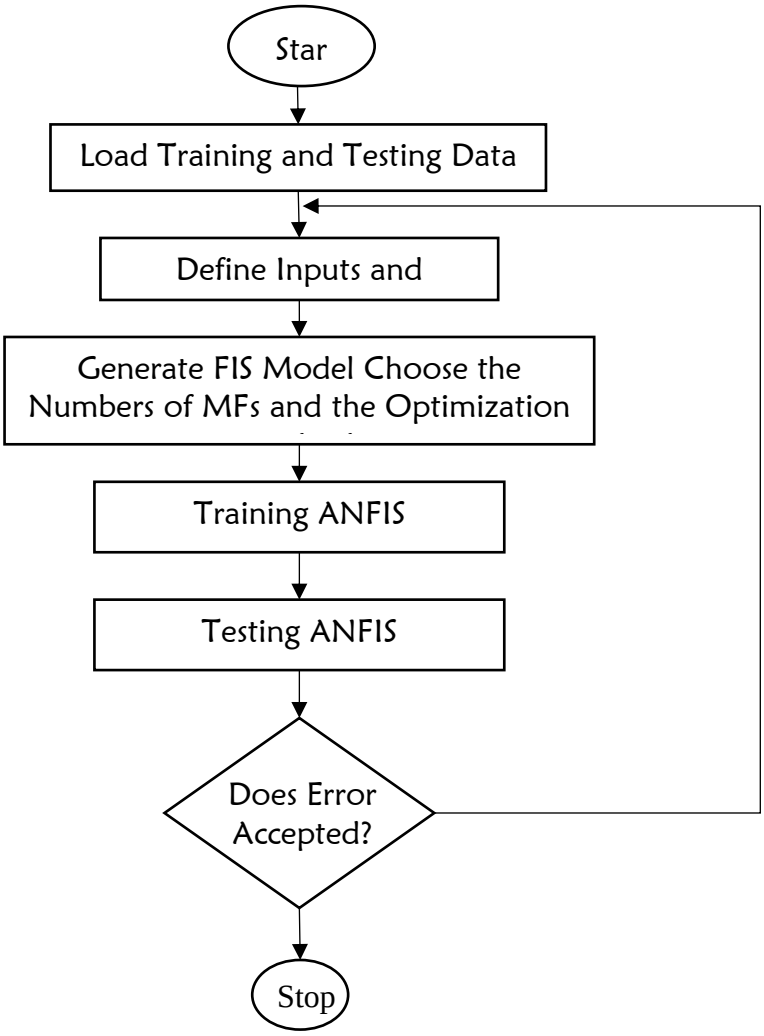

Figure 2.2 Descriptive progression of the ANFIS model (Noureldeen and Hamdan 2018)

Next is training the ANFIS to make predictions. Once the FIS is generated ANFIS is then trained using the function “anfis()”. This is a major training function that uses hybrid learning to identify the parameters of the Sugeno-type fuzzy inference system. It applies a combination of the backpropagation gradient descent method for training FIS membership function parameters to emulate a given training data set. The training process stops whenever the maximum epoch number is reached, or the training error goal is achieved.

However, an adaptive neuro-fuzzy inference system is suitable for the solutions of medical disease detection and diagnosis. Several researchers have implemented applications of ANFIS that can predict and diagnose diverse medical diseases. An E-healthcare system using ANFIS proposed by Sarangi et al., (2016) as illustrated in Figure 2.2 aims to detect cardiovascular disease smartly and communicate the detection with a physician as well as preferred medication to the patients.

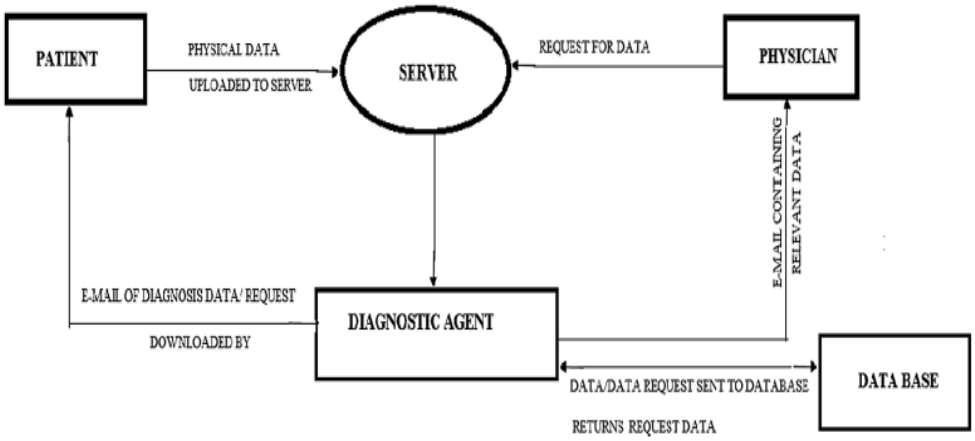

Figure 2.3 Intelligent E-health system (Sarangi et al. 2016)

In another work by Subbulakshmi et al., (2019) implemented a Sugeno-type method ANFIS application for coronary artery disease which analysis various

parameters to identify the patient's well-being condition. Another similar work by Shuriyaa & Rajendranb, (2018) utilized the ANFIS approach which aimed to predict cardiovascular disease, to obtain the highest performance possible, different parameters were investigated with various values.

Also, another work by Song et al., (2021) in their paper, examine the application of ANFIS in heart disease prediction, as well as some innovative combinations of ANFIS and other techniques to support the diagnosis of heart disease aiming to improve the model performance.

Similarly, Paul & Karn, (2021) implemented an ANFIS-based application for the prediction of diabetes mellitus by subtractive clustering method, ANFIS is trained using Pima Indian diabetes dataset with 8 inputs which provides a maximum percentage accuracy of about 99% in estimating the presence or absence of the disease. Another work by Damodara & Thakur, (2021) developed a chronic kidney disease ANFIS prediction system based on a Sugeno inference system where the network parameter is updated continuously using the learning method.

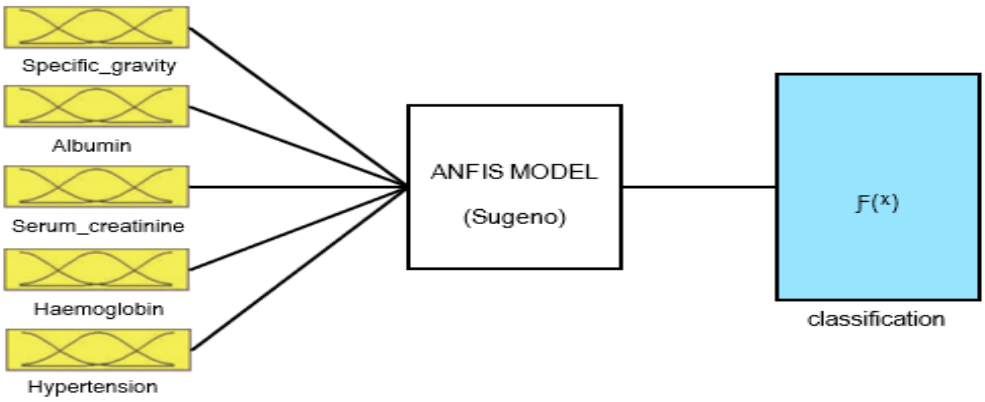

Figure 2.4 Neuro-fuzzy inference CKD stage prediction system (Damodara and Thakur 2021)

Figure 2.3 depicts the ANFIS system for the prediction of chronic kidney disease, hybrid algorithm is used which involves the combination of least squares and back propagation algorithms. As part of the system, input parameters are mapped to output parameters using a membership function and a set of if-then rules. The performance of their model can predict CKD with an accuracy of 94%.

Meanwhile, a breast cancer prediction is proposed by HomaieFasih & Ahsani (2020) to implement a fuzzy model for Iranian breast cancer patients to predict the disease in its early stage to help with treatment and improve the lives of the patients. In their work, Sugeno fuzzy inference, Mamdani fuzzy inference, and ANFIS are implemented and achieved the best prediction error by ANFIS using the fuzzy c-means approach. Also, a work by Santoso & Amrullah, (2020) proposed an ANFIS application that can predict lung disease, a patient's symptoms are used to determine the effectiveness of the system and their application gives the best performance using a subtractive clustering process.

In another work by Ukaoha et al., (2020) proposes an ANFIS model to diagnose the coronavirus disease as an alternative due to limited test kits in the hospitals as the number of patients keeps increasing. The performance of their model has successfully predicted the disease by 96% over other models compared. As a result of the ANFIS model, COVID-19 can be diagnosed promptly and differentially, which in return helps doctors make more informed clinical decisions and reduces diagnostic errors.

Therefore, Table 2.2 summarizes ANFIS-based applications that have been discussed in this section.