NON- INVASIVE SYSTEM FOR HEART DISEASE PREDICTION

Thesis submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

VAISHALI BAVISKAR
(E18SOE827)

School of Computer Science Engineering and Technology

BENNETT UNIVERSITY

(Established under UP Act No 24, 2016)

Plot Nos 8-11, Tech Zone II,

Greater Noida-201310, Uttar Pradesh, India

August, 2023

NON- INVASIVE SYSTEM FOR HEART DISEASE PREDICTION

Thesis submitted in fulfilment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY

By

VAISHALI BAVISKAR

(Registration No. E18SOE827)

Under the Supervision of

Dr. MADHUSHI VERMA and **Dr. PRADEEP CHATTERJEE**

BENNETT
UNIVERSITY
TIMES OF INDIA GROUP

School of Computer Science Engineering and Technology

BENNETT UNIVERSITY

Plot Nos 8-11, Tech Zone II

Greater Noida-201310, Uttar Pradesh, India

August, 2023

Dedicated to my beloved husband, kids and family

DECLARATION BY THE SCHOLAR

I hereby declare that the work reported in the Ph.D. thesis entitled **"Non Invasive System for Heart Disease Prediction"** submitted at **Bennett University, Greater Noida, India** is an authentic record of my work carried out under the supervision of **Dr. Madhushi Verma** and **Dr. Pradeep Chatterjee**. I have not submitted this work elsewhere for any other degree or diploma. I am fully responsible for the contents of my Ph.D. thesis.

Vaishali Baviskar
Enrollment No. E18SOE827
School of Computer Science Engineering and Technology
Bennett University, Greater Noida, India
Date:

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled **"Non Invasive System for Heart Disease Prediction"** submitted by **Vaishali Baviskar** at **Bennett University, Greater Noida, India** is a bonafide record of her original work carried out under my supervision. This work has not been submitted elewhere for any other degree or diploma.

Dr. Madhushi Verma
Associate Professor
School of Computer Science Engineering and Technology
Bennett University, Greater Noida, India
Date:

Dr. Pradeep Chatterjee
General Manager
Tata Motors
Pimpri, Pune
Date:

ACKNOWLEDGEMENTS

I want to sincerely thank to my supervisors **Dr. Madhushi Verma** and **Dr. Pradeep Chatterjee**, for their invaluable advice and ongoing guidance all through my Ph.D. study and related research. They have inspired by all depth of experience in my research work with their deep knowledge and I also want to say big thank you to Dr. Gaurav Singal for his technical support with my research. I would like to show appreciation to the dean, Dr. Deepak Garg, and Ph.D. coordinator, Dr. Akansha Singh. I relate my success in making the most of my time studying at Bennett University to their plentiful assistance and support. Finally, I want to thank my husband, my kids, and my parents. I never would have been able to complete my study without their incredible help and understanding over the past few years.

Contents

Abstract

Heart disease (HD) is considered as a potentially fatal disease. Because of multiple contributory risk factors such as abnormal pulse rate, diabetes, high blood pressure, excessive cholesterol, and so on, identifying the condition is difficult. Accurate and timely diagnosis is critical for HD therapy and prevention. As HD is an important reason of deaths in developing nations, the intentions of detecting the risk factors and early sign detection of this disease is important for which the present study has been undertaken. In spite of the various endeavors of conventional works, there is a scope for improvement with regard to accuracy. To attain high prediction accuracy, as per defined objective 1, in phase-I, the proposed work uses genetic algorithm (GA), particle swarm optimisation (PSO), african buffalo optimisation (ABO) and genetic sine algorithm (GSA) on the benchmark UCI dataset to remove the unnecessary and redundant features and prevent the features from becoming trapped in local minima. For the classification of specific features, the system employs deep learning (DL) based recurrent neural network (RNN), long short term memory (LSTM), and deep progressive attention (DPA-RNN+LSTM) to boost the model's classification rate. In phase-II, the study proposes penguin optimisation algorithm (POA) for feature optimisation and stacked sparse convolutional based auto-encoder for the classification of HD. As per defined objective 2, further in phase-III, the Raspberry-Pi using the arduino uno and AD8232 sensor framework is used. This made it possible to receive the patient's body parameters in real time. Since IoT-based data is compiled from a variety of sources, it may be chaotic and noisy. IoT-data mining is employed to help with tasks like defining typical links between data components and using them to address prognostication concerns . Also, validate the data for binary and multiclass. To achieve objective 3, in phase –IV, the vascular age, cardiac index, and cardiac risk score is estimated to predict regression values. Finally, the classification of heart disease and regression is deployed in the form of flask web page model. Overall, the proposed model is assessed with regard to significant metrics for ensuring its effectiveness than conventional algorithms in HD prognosis.

List of Tables

List of Figures

LIST OF ACRONYMS & ABBREVIATIONS

ABC	Artificial Bee Colony
ABO	African Buffalo Optimization
AE	Auto Encoder
AGAFL	Adaptive Genetic Algorithm with Fuzzy Logic
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area under the Curve
BD	Big Data
BF	Buffalo Reference
BP	Blood Pressure
BRF	Balanced Random Forest
CAD	Coronary Artery Disease
CART	Classification and Regression Tree
CCNN	Cascaded Convolutional Neural Network
CHAID	Chi-Squared Automatic Interaction Detection
CHD	Coronary Heart Disease
CNN	Convolutional Neural Networks
CVD	Cardio Vascular Disease
DAE	Deep Auto-Encoders

DL	Deep Learning
DM	Data Mining
DNN	Deep Neural Network
DPA	Deep Progressive Attention
DT	Decision Trees
ECG	Electrocardiogram
EDCNN	Enhanced DL aided Convolutional Neural Network
EMRs	Electronic Medical Records
EPO	Emperor Penguin Optimizer
ESC	European Society of Cardiology
FHS	Framingham Heart Study
FP	False Positive
GA	Genetic Algorithm
GAN	Generative Adversarial Network
GPIO	General Purpose I/O
GSA	Genetic Sine Algorithm
GSO	Galactic Swarm Optimization
HB	Haemoglobin
HD	Heart Disease
HDL	High Density Lipoprotein
HF	Heart Failure
IACPSO	Improved Auto Categorical PSO
IFA	Intelligent Firefly Algorithm
IHA	Indian Heart Association
IoMT	Internet of Medical Things
IoT	Internet of Things

K-NN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
LDL	Low Density Lipoprotein
LR	Logistic Regression
LSTM	Long Short Term Memory
LTP	Local Transform Pattern
MadeGAN	Memory-Augmented Deep autoEncoder with GAN
MC	Microcontrollers
MCC	Matthews Correlation Coefficient
MDT	Multidisciplinary Team
ML	Machine Learning
MLP	Multi-Layer Perceptron
MRM	Multivariate Regression Model
NLP	Natural Language Processing
NYHA	New York Heart Association
PCA	Principal Component Analysis
PCG	Phonocardiograms
PSO	Particle Swarm Optimization
PUC	Pooled Area Curve
RBFN	Radial Basis Function Network
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SA	Sine Algorithm
SRE	Square Root Error
SSC-AE	Stacked Sparse Convolutional Neural Network Based Auto Encoder

SVM	Support Vector Machine
WHO	World Health Organization
XG-Boost	eXtreme Gradient Boosting

LIST OF SYMBOLS

$\alpha F(D1)$	Classifier error rate
sf	Length of selected features subset
Af	Entire count
$\beta,(1-\alpha)$	Constraints which is used to control and classify accurateness weights combined with feature minimization
$f(y_i)$	Calculation of fitness value is utilized by the function representation
$x_{value} = \frac{N_{Top}}{2}(fit - sort)$	Breeding sector population
$Prob_c$	Probability function
B_s	Search -Space capability
N	Swarm size value
x_m	Search of buffalo's waa signals
lf_1 **and** lf_2	Learning constraints
c	Crossover rate
m	Mutation rate
δ	Activation functionality
A_u	Hidden layer vector
c_u	Process gate
j_u	Memory vector
e_u	Result gate
ωp	Weight-matrix
b_p	Bias-vector

α_t	Used for the computing the connection among context vector
v_p **and** v_{ot}	Context vector
ϕ	Velocity of wind
Ψ	Gradient
$T^{'}$	Calculated temperature profile in the vicinity of the huddle
T	Time
S	Radius
N	Movement parameter for collision avoidance
P	Accuracy of polygon grid
R	Random function
x_i	Input variable
h_i	Hidden variable
W_{i1}	Weight matrix
a_{i1}	i^{th} auto encoder one-sided encoding procedure
a_{i2}	i^{th} auto encoder one-sided decoding procedure
Y	Output
W	Convolutional kernel
X	Input
$J(\theta)$	Loss function of softmax regression
I	Indicator function
l_{p1} **and** l_{p2}	Learning parameters
$Sig(a)$	Sigmoid function
y	Feature representation
β	Beta
P	Cost function
σ_j	Average activation

VV^c_{min}	Minimum velocity
VV^c_{max}	Maximum velocity
PP_j	pbest of j^{th} particle
IP_g	gbest of particles
H	Whole particles from two random sequences
A_j^{u+1}	Position of present solution in the j^{th} dimension at the u^{th} iteration
Qj	Location of place point present in the u^{th} dimension
θt	Temperature profile
d	Present iteration
N	Parameter for movements
$input_{emp}$	Position of other emperor penguins
$\otimes$	Concatenation function

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

The cardio vascular disease (CVD) is one of the fatal disease in human community, which is rapidly increasing globally with high mortality rates. Usually, the heart supplies sufficient and adequate levels of blood to rest of the body parts, for accomplishing the normal functionalities. Among the other congenital invasive- based techniques, angiography is considered to be the well-known technique for the diagnosis of the issues but has some limitations such as age, gender and probability of success rates [67].

On the other hand, non-invasive based methods used in the diagnosis of heart diseases such as intelligent learning based computational techniques are found more prominent and effective in diagnosing the heart abnormality. Heart disease is considered to be the most perilous and life snatching disease globally as it is a chronic type disease. Heart failure occurs due to blockage and narrowing of the blood vessels also known as coronary arteries. According to the report generated by WHO, an average of 17.90 million humans mortality rates have been increased due to CVD in the year of 2016, which constituted 30% of overall death rates. Followed by another statistical report which stated that, about 0.2 million people expire due to heart ailment in pakistan annually. Every year, levels of victimizing people are increasing rapidly. The european society of cardiology known as (ESC) has circulated a report that about 26.5 million adults are identified with heart disease and 3.8 million were recognized each year. About 50–55% of patients with heart disease perish within 1–3 years, where the cost of treatment for heart disease is about 4% constituting the overall healthcare annual budget rate levels. Several types of heart abnormalities are depicted in figure 1.1 [7].

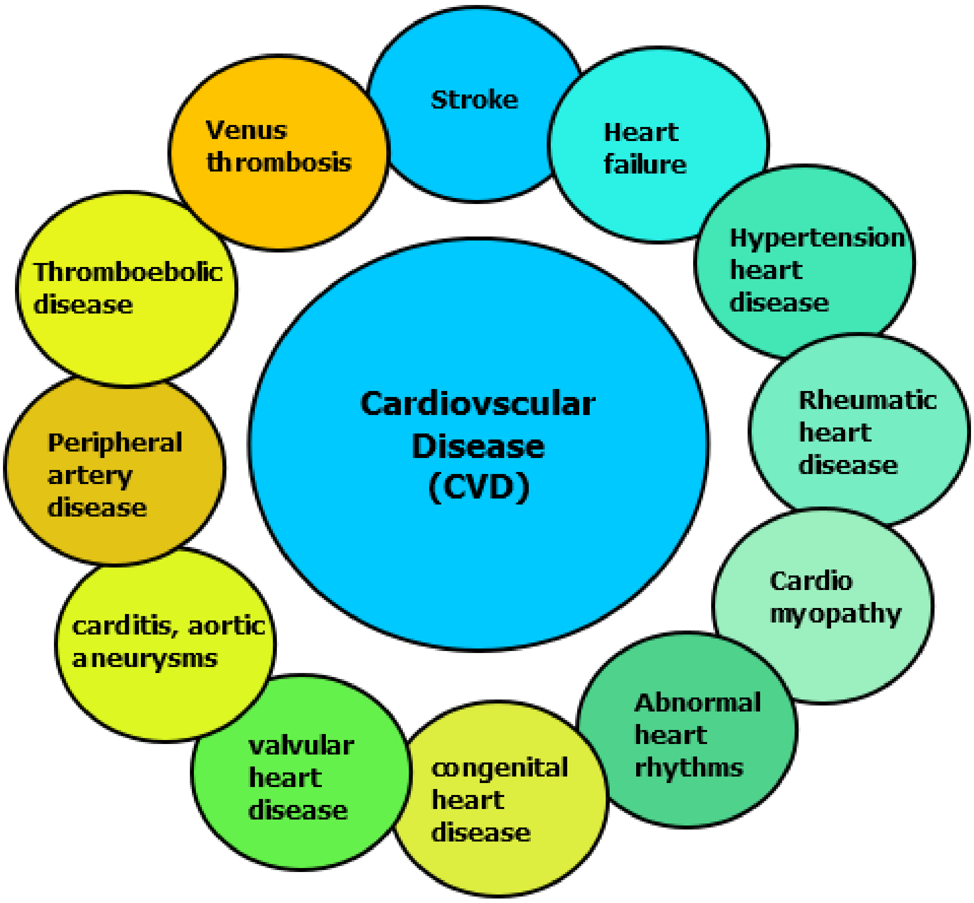

Figure 1.1: Several types of CVD in heart abnormality

A recent survey conducted at US depicts that most affected country by the heart disease and abnormality patients are high in number. The most common symptoms of the heart disease are the body weakness, swollen feet, weariness and some associated signs. The risk of heart disease may increase with the individual lifestyle such as smoking, fitness levels and unhealthy diet plans. This heart disease are of several types as CAD known to be coronary artery disease , leading to common defects in the heart such as chest pain, attack and stroke rates. The other type is the congenital heart disease which is found to be genetically inherited in an individual.

At the initial times, the traditional techniques used in the investigation and in the detection of the heart diseases were found complex due to simplified and less defined architectural design of the model [56].

Diagnosis and detection of cardiac abnormalities is complicated by the availability of medical diagnostic tools and specialists in developing countries. However, detailed and accurate diagnosis of cardiac abnormalities is very important in protecting patients from the development of unwanted abnormalities [70].

1.2 An Overview of Heart Disease Prediction and Classification

Prediction of CVD is regarded to be the prominent subjects in the sections of heart abnormality analysis. The load of CVD is rapidly increasing in all over the world. In recent times, many such researches have been conducted for addressing the diagnosis and classification of the heart abnormality rates. The major challenge is the heart disease detection. The early diagnosis of the heart disease plays a vital role in making the decision regarding the changes in the lifestyle of high-risk rate of patients and in reducing the complication. Machine learning (ML) and deep learning (DL) techniques have been implemented in classifying the presence of disease rates. These methods prove to be effective in decision making and prediction from a large quantity of data which are produced by health care sectors. The basic approaches and methods in the heart disease prediction and classification of the diseases are depicted in figure 1.2 [35].

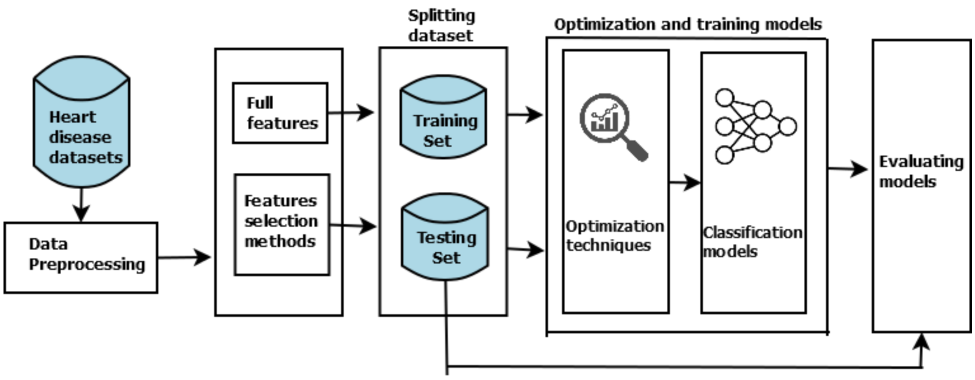

Figure 1.2: Basic methodology involved in disease prediction and classification

Though abnormality in heart occurs in different forms, the common set of core risk factors that influences the mortality rates in heart are analyzed and are detected at earlier rates to reduce the adversity of the abnormalities. However, monitoring the patients for a continuous period of 24 hours is not possible as it requires more of time and expertise with sapience requirements. Since technology and science is evolving at a great pace, detection and classification of the disease are easier by just recognizing the hidden patterns which are used for an easier detection of the diseases in case of health diagnosis in sectors of medicinal data. Hence, continuous efforts have been made in predicting the possibility at a prior stage [11].

Several approaches have been made by implementing both the ML and the DL techniques for analyzing and extracting the desired data that can be used in the appropriate analysis of the heart abnormality rates [6]. Timely and accurate detection of the CVD are critical in reducing the risk factors of myocardial infarction (MI). Relation among the factors of CVD are complex, ill-defied and are non-liner used in justifying the case of AI tools. These AI tools aid in the process of predicting and in classifying the CVD. Several risk prediction algorithms are recommended by integrating the data with the common risk factors. AI, ML and DL methods have been playing a vital role in the medical science for diagnosing numerous diseases more effectively in the patients [59].

1.3 Significance of Heart Disease Prediction and Classification using ML and DL Algorithms

The result of CVD is fatal as it causes both, circulatory system and heart to be dysfunctional and leads to death or physical paralysis. Thus, primary and instant detection of CVD can save many lives from earnest risk factors. Numerous researches have been carried out for attaining this particular objective, but there still exists room for perfection in case of both, performance and dependability. Currently, no large-scale approaches have been implemented for a routine medical information and in deploying ML in prognostic assessment of CVD. ML helps a cardiologist to predict diseases at an early stage and treat the patient accordingly. There are many ML techniques such as support vector machines (SVM), decision trees (DT) and k-nearest neighbor (K-NN), each with its strengths and weaknesses. These methods have been applied in broader areas; like in predicting human heart (echocardiogram signals) and skin diseases. With the ensemble and DL algorithms such as random forest (RF), xg-boost, and convolutional neural network (CNN) are used in the CVD prediction and classification in cases of picture categorization. These methods are used in the categories of multi-process strategy for developing a complete system that can forecast the probability of CVD [22]. Accurate and timely detection of the heart abnormality by using these approaches can result in protecting the patient form the rest of the adverse risk factors.

Figure 1.3: Data handling approaches using AI

Figure 1.3 represents the diagrammatic approach for the process of handling the medical examination data using the AI approaches [65]. These approaches are used in selecting only the preferred attributes which are simple and can be used in the medical examination of a heart disease patient. These systems are also cost effective leading to accurate detection and prediction with minimal requirements of tests [48].

1.4 Recent Trends in Heart Disease Prediction and Classification

Disease diagnosis is a crucial aspect of health care. If a disease is diagnosed at the right time then appropriate medical help can save lives and reduce the mortality rate [37]. Classification of diseases using ML and DL approaches have evolved at a great pace and supporting the medical sectors by providing instant, reliable and accurate predictions [54]. Data mining techniques and image fusion techniques can be deployed along with the ML and DL algorithms to resolve these issues, aid in managing the specific forms of data in the medical centers and in allying them with high specificity rates. Some of the recent positive attributes towards the heart disease diagnosis are given by [18, 27, 55].

- Image-fusion with DL approaches- used in fusing the couple of images to a single image with higher resolution integrated with clear set of data.
- Naïve bayes technique, from ML are easy to deploy and to develop, generally performing well in case of complex disease diagnosis.
- ANN are used in case of numerous and complex heart disease datasets as it has several layers used in receiving several inputs, and delivers only one prominent value as output.
- ML methods are also more useful in making the data to be free from noise features, by marking only the key traits for a better accuracy rates in prediction.
- ML are often used in case of predicting the likelihood of the illness occurrence.
- Hybrid models of DL are used in the heart disease prediction which are used in the prediction of CAD which are more accurate than using single models in predictions [55].

1.5 Major Pitfalls and Applications of Heart Disease Prediction and Classification

Some of the major pitfalls associated with the heart abnormality prediction and classification are discussed in this section. The recent report from US states that the ever-increasing risk factor for CVD are the minor pitfalls connected to the recognition and in the delayed diagnosis of the disease adversity rates. Due to these issues, the risk of heart failure for a general part of the population is expected to claim at the rate of 33.4%, in the year of 2060, where the risk of attack is more than 14%, since the year of 2020. It is one of the vital factors to enhance the detection and the prevention capabilities for the cardiac patients in saving lives and in improving the overall health factors. Rather than deploying only ECG signal analysis using the ML and DL techniques, EKG can also be analyzed for the severity prediction levels. Continuous monitoring of the heart abnormalities can result in missing up the 50% of arrhythmias occurring in the elderly heart patients. Some of the equipments such as pace makers and aneurysm clips can be unsafe in making the detection [60]. SVM methods are not sensitive to the redundant features, where the omission of significant features are more hazardous than the redundant features [45]. Moreover, the rhythm created by the atrial flutter creates a dynamic range of inaccurate detection when arte is done using the ECG signal analysis. This creates a difficult phase for them to distinguish among the PAC waves, leading to increased false-positive rates [32].

1.6 Problem Identification

In recent times, heart disease has become a crucial medical condition leading to severity of the problem such as stroke, angio and complicated surgical tasks. Higher mortality rates have been reached due to significant absence of prognostication of these diseases, leading to a high severity rate and resulting in increased number of mortality levels [62]. A statistical report from WHO, states that 8.9% of population is facing a delayed diagnosis of CHD on a daily basis. Among these, young patients or patients below the age of 15 with CHD are diagnosed in a period of 10 years. Moreover, several approaches related to the domains of DM, constituting both the ML and the DL concepts help in detecting the heart abnormality levels by implementing both, the feature extraction mechanism and the classification of both the presence or absence of the diseases. This faced several consequences especially related to heart abnormalities. Some of them are listed as,

- This particular approach at [14] has used a method of hybrid GA-ANN and are compared with several other DL approaches like ANN, DT, SVM, KNN, LR and RF. In these comparisons, the rate of accuracy has differed based on the algorithms as in rates of ANN in rates of 73.4%, DT in levels of 72.3%, and finally in case of in levels of SVM-68.9
- Subsequently, this approach using [23] adaptive genetic algorithm with fuzzy logic (AGAFL) showing higher rates of accuracy levels in case of using 90% for the datasets from hungary 91%, switzerland 89%, and cleveland 89%.

- One more approach for the heart abnormality detection in [19] has employed improved auto categorical PSO (IACPSO), showing, 98% as accuracy rate. In spite of better performance of conventional works in accordance with accuracy rate, there is a scope for improvement in this area in accurate levels of classification and in domains of feature extraction methods.

Moreover, lack of physicians, technical involvement rates, late diagnosis and non-appropriate diagnostic rate are some of the crucial reasons resulting in higher severity rates and levels in case of heart abnormality detections. Thus, making an appropriate and a timely classification and detection of these diseases can aid in reducing the severity and can help in timely delivery of the appropriate medications for an early cure. On consideration to these issues, some of the other approaches have been carried out in case of determining the complexities, and severity of the heart disease. But, Some setbacks are inevitable in this process. These are:

- ML based approaches such as PSO, ACO are associated with lower power processing systems where large dataset cannot be loaded for accurate prediction levels [34].

- ANN architecture which is DL based used in heart abnormality prediction, but the model adapted for training resulted in over-fitting and resulted in decreased accuracy rates and are less appropriate in making effective classification in case of large data set usage [10].

- Real-time dataset used in the DL methods constituting CCNN and GSO methods lacked in producing appropriate prediction rates and effective working on large-prototype dataset [51].

Based to these drawbacks, an appropriate and accurate disease abnormality detection and classification of abnormality in view of differentiating them into the case of presence or absence of abnormality in heart has to be done on the basis of implementing the DM mechanisms and approaches. This can aid in making an early and timely accurate detection, in view of providing medication at early stage resulting in less complications and in reducing higher rates of mortalities.

1.7 Motivation

Daily lifestyle of an individual indulge in affecting working of human heart. Many such issues initiating on a rapid pace and newly involving heart diseases are rapidly identified. Currently, more of stress levels involving in daily lives of individuals, heart noted as an initial and major organ in circulating the blood levels around the complete body. This indicates a proper functioning of heart that should be conserved for a restored living. There are several genetic factors involved in passing a heart disease or a case of abnormality down form generations. According to WHO, more than 12 million of the deaths are occurring worldwide, due to various abnormalities, affecting the entire heart and arteries. Every young-aged person around 20-30 years of age is affected by some kind of heart disease. The diagnosis of heart disease is important and considered as a complicated task in case of medical field. The main motive of carrying out the specific study lies in making a best classification of the heart disease via approximate feature extraction and data processing approaches which are more obvious in implementing AI approaches constituting both ML and DL. Early detection can make the healing process much more easier than detecting them at a later stage which leads to complicated and complex stage of recovery. Also, early prediction can reduce the likelihood of disease occurrence especially patients getting various heart diseases. The other main motive of early detection is intricate study of the individual's body pattern and behaviour related to heart risk parameters. The proposed study uses DL and ML techniques. This process undergoes feature extraction and classification of the absence or the presence of the abnormality especially at the heart's regions.

1.8 Aim and Objective

The primary aim of the study relates in employing various DM techniques comprising the ML and the DL approaches in executing an appropriate prediction and classification of the heart abnormalities. In this research, the aim relies on the objectives as mentioned,

1. **To improve the accuracy of hybrid deep learning algorithm using novel features selection technique for heart disease prediction** - to implement feature selection ap-

proach using genetic algorithm, PSO and ABO algorithm and classification using the DPA-RNN+LSTM methods for identifying whether a heart problem is present or absent. Use penguin optimization algorithm for the feature selection approach and implementing stacked sparse convolutional neural network based auto encoder (SSC-AE) for classification phase for classifying either being present or not abnormality in heart.

2. **Validate the proposed model on synthetic and sensing environment dataset for binary and multiclass classification** - to predict the efficiency of the binary and multiclass classification model, using different optimization algorithms with RNN, LSTM, deep progressive attention RNN and LSTM and SSC-AE using IoT in cases of retrieving data from patients.

3. **To calculate the heart severity with cardiac risk score, cardiac index (CI) and vascular age of heart (VAH)** - to perform feature selection using meta-heuristic algorithms comprising namely GA, PSO, GSA, ABO and POA and DPA-RNN+LSTM for classification and regression is employed based on cardiac risk score, vascular age of heart and cardiac index on real-time datasets by implanting web-deployment systems. Use performance metrics that include different attributes to assess the model's efficiency in predicting the prognosis of heart disease.

1.9 Significance of the Study

The heart disease diagnosis is made according to the signs, symptoms as well as by using the physical examination of the patient. The diagnosis of heart disease is considered to be a great challenge and an essential task and can be accomplished by evaluating the complete heart functioning and its working mechanism. The automated form of prediction is superior to a manual diagnosis and can be more effective compared to a manual one. Using an automated form of disease prediction will reduce the case of misdiagnosis and will result in appropriate diagnostic measures giving complete information of the heart system. Assessing the real-time data obtained using sensors for the appropriate predictions is more viable to provide timely and accurate medications to resolve abnormality levels as compared to implementing artificial data. It is observed that correct rate of prediction of heart diseases can prevent the cases of mortality, while incorrect rates of prediction can prove to be fatal. The models emphasized in the respective study are used both in case of appropriate feature extraction and in the classification of existence or absence of a heart problem. Additionally, IoT can be utilized in making a daily analysis over the patients such as data entry as well as making a periodical analysis over the patients. Thus, the significant part of study lies in accomplishing the faster and the accurate form of prognostics over the patients by effective classification and reducing the adverse risk factors. Also, by deploying the IoT and web-deployment factors in case of medical care and ailment, it is easy to keep a timely track on the illness as well as in the unusual status of patients. Apart from this; early detections, heart risk parameters detection etc are some of the vital features carried out in this approach.

1.10 Scope of the Study

The proposed work scope begins with care and concern for an individual who seeks medical practitioners' attention during the diagnostic process and requires continuous supervision in view of avoiding further future complications and complex forms of medical prognostic procedures. The data collected from UCI cleveland dataset, heart failure prediction dataset and IoT device are analyzed under the procedures of feature extraction used in obtaining only the exact and relevant features and the rest are removed from the redundant data for accurate range of predictions and are classified based on the disease's presence or absence using the ML and DL methods. Calculated the risk parameters of heart i.e. risk score, heart age and cardiac index. Further, the advancement is done by deploying the real-time data which obliges in analyzing the timely conditions of the patients. Moreover, the proposed study has an advantage in making use of the web-deployment in case of disease-prediction and in classification of presence or absence of disease by using the real-time data analysation from patients and predicting the heart risk parameters. These methods are more prominent in making personalized and successful prognostic measures in delivering effective forms of treatments.

1.11 Thesis Organisation

The entire study comprises of 8 different chapters constituting various aspects in wrapping the entire study. Chapter 1- this chapter pacts with the complete and a brief introduction comprising entire study, with some illustrative depictions and their prioritized explanations are provided for a clear understanding. Chapter 2- this chapter entirely deals with cutting-edge procedures and approaches utilized in heart disease and abnormality detection and classification phases which accomplishes various domains such as ML, DL based on various datasets such as real-time, artificial existing datasets. This chapter aids in providing a clear understanding about the proposed study using the existing study concepts. Chapter 3- this chapter handles the entire proposed technique and the methodology part which is projected in this specific study. This chapter additionally provides a complete overview on the proposed technique comprising the data collection methods, which are segregated to four different sub-sections. Chapter 4- this chapter deals with the overall proposed methodology, with the methods and approaches comprising both the classification and feature selection used in the study. Additionally, each section emphasizes the results, for the particular phase obtained from the proposed methods and approaches, and their appropriate discussions for a clear understanding. Chapter 5- this chapter deals with the overall proposed methodology, with the methods involved in both feature selection and the classification mechanism, with the appropriate results and summary sections. Chapter 6- this chapter deals with the entire methodology used, with the proposed design representation, approaches involved in a detailed manner and the results of the implemented model in the respective phase. Chapter 7- this chapter deals with the entire methodology used, with the proposed design representation, approaches involved in a detailed manner comprising the DL model in both classification and in regression, with the exhibition of results of the implemented model in the respective phase. Chapter 8- the final part of the thesis ending up with the appropriate conclusion and the possible future recommendation in aspects of ML and DL in heart disease diagnosis. The complete reference of the entire study is also presented in this paper.

1.12 Summary

One of the major causes of heart disease is mortality and morbidity among the world's population. Early and timely detection as well as prevention of heart disease is extremely important in serious and various health complications. Thus the entire chapter provides a complete and an elaborated introduction for a better understanding of the concepts including various sections such as the background, significance, scope, and the chapter also presents the main attribute, aim and objective to carry out the study in making the descriptive understanding to be more precise. The rest of the chapters will cover the entire study under various aspects and with their detailed explanations for a better understanding in cases.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In recent years, heart disease (HD) is deliberated to be among the foremost hazardous and life grabbing longterm diseases throughout the world. The reason for this is the heart declines to provide an adequate blood supply to further organs of the body for the accomplishment of the regular functionality. So, heart failure (HF) occurs because of the obstruction and also contraction of coronary arteries. Thus, the coronary arteries are accountable in support of the blood distribution into the heart. However, the current survey exposes that the united states is leading in heart diseases. The symptoms of HD comprises of physical weakness, swollen feet, shortness of breath and fatigue. Therefore, the risk of HD can be further augmented with the way a person lives, including their unhealthy diet, high blood pressure, lack of exercise, level of fitness, smoking and so on.

In accordance with the record [67] produced by the world health organization (WHO), 17.90 million humans on average passed away with CVD in 2016. Hence, this number signifies almost 30% of entire global bereavements.

There are various inadequacies of the presently obtainable risk extrapolation methods of HF. Initially, most preceding models are established by using conventional numerical methods like regression modelling and ML–based prediction models. As a result, [16] these methods have inadequate effectiveness to update population densities which cannot have the capability to attain further information on top of regularly collected health care statistics by insurance claims or by electronic medical records (EMRs) for registered patients. Furthermore, data established through the medicinal zone or else in hospitals is so enormous which occasionally converts into problematic when it is analyzed. Subsequently, by utilizing the ML approaches for prognostication and also data handling can be more effective for the physicians. Later, this study [33] has deliberated about the HD risk issues and described ML methods. By the usage of ML methods, HD can be prophesied and provided with the proportional study of the procedures for ML used in the prediction experiments. Consequently, deep learning (DL) can be compatible towards cardiovascular medication in which haemodynamic along with electrophysiological guides are progressively taken by an uninterrupted basis through adequate strategies and image separation in cardiac imaging. Additionally, in this study [36], DL has an important softness which includes complications in understanding its methods like the 'black-box' criticism, it's essential for widespread arbitrated is 'labelled' data in training, deficiency of regulation in scheme, nonexistence of data effectiveness in training, inadequate relevance towards medical studies, and other reasons. So, the optimum interventional presentation for DL needs cautious preparation of soluble issues, choice with most suitable DL procedures as well as data, and composed clarification of outcomes. The current research is entirely connected with the HD prediction through ML and DL models. As the existing works have focussed on different aspects of heart disease prediction through ML and DL, these studies are reviewed in this chapter.

2.2 Review on Conventional ML models in Heart Disease Prediction and Classification

Artificial intelligence (AI) and machine learning (ML) have recently generating significant responsiveness in the technical community and in broadcasting. Correspondingly, systems have pronounced prospects in medication field for providing personal patient care, which includes diagnosis of HD combined with heart failure management. Though, numerous physicians are acquainted with these details, but several are unacquainted with the algorithms basics and in what way it is implemented in the clinical field. Through this heart failure research, the recent methods of ML includes formation of innovative methods for diagnosis and categorizes the patients into unique phenotypical groups and enhances the exploration competences. Indeed, this paper [47] has dealt with the outline of ML directed for practicing physicians, it also calculates the recent ML applications in analysing, categorising and extrapolation of heart failure (HF).

Subsequently, in ML algorithm, the term ‘learn’ is referred to as the information that are received directly from the data and there is an enhancement in the performance consistently. Besides, this study [8] has physically examined cochrane and also medline databases with the references included in the applicable analytical studies along with comprised studies. Eventually, in this study nearly 122 significant applications have been taken. These studies have predominantly mentioned which are as follows (a) the contribution of ML to the HF patient classification of patients into different groups which has been involved in various treatment approaches, (b) HF patients can be differentiated against the hearty people or from other syndromes, (c) Evaluation of HF consequences, (d) recognition of HF patients against the automated history and patients with same features whomever have related treatment approaches,(e) concerning the fetched data from the medical records (f) evaluation of HF patients results with implanted gadgets like a left ventricular assist device and cardiac resynchronization therapy. Thus, ML method plays a significant part for the effective approaches for analysing, managing and for the evaluation of the resultants of HF people.

In accordance with disease diagnosis, ML methods can improve the rapidity of management, which can reduce the rate of incorrect positives. Further, in the recent trends in ML procedures have prepared a considerable effect on the identification and analysing of various syndromes. Here, several ML algorithms likes SVM, decision tree, KNN and naive bayes have been considered. However, the real-time employment of these algorithms have deliberated with the usage of python. These algorithms have been acquainted to diagnose numerous diseases like diabetes, cancer, heart attack, epilepsy and other diseases. Therefore, the study [49] has suggested a theoretical and scientific background for precision, recall, accuracy, and f1 score standards of an ML algorithm so as to identify disease. Certainly, HD premises a substantial death rate around the world, and it has turned out to be a health threat for a lot of people. Earlier, HD used to be predicted to save more lives by identifying the cardiovascular conditions like coronary artery diseases, heart attacks and so on. ML can fetch a current resolution for consequential and exact calculations. As a means to medicinal field, numerous upgrading occurred due to usage of ML methods. This study [42] has suggested that, innovative ML approach is in order to predict heart disease. Accordingly, the recommended study has manipulated the dataset of cleveland heart disease, with the data mining methods has been used for classification as well as regression. Hence, in ML methods, decision tree and the random forest have been used and has attained satisfactory results. Generally, HD is among the furthermost prevalent common diseases in recent days. A primary diagnosis of this disease is a critical work for several health care professionals to preclude the patients from the disease. Besides, this paper [1] has dealt with a reasonable study of various classifiers for the HD dataset's classification so that an appropriate classification and prediction of HD cases with minimum characteristics can be attained. As in consequence, the set has encompassed of Class attribute is one of 76 attributes that have been collected from 1025 patients from places like switzerland, cleveland, long beach and hungary. But, a subset of only 14 attributes have been taken, whereas each feature has a certain position value. Additionally, those algorithms have been utilized include naive bayes, k-nearest neighbor (K-NN), SVM, stochastic gradient decent (SGD), decision tree J48, adaboost, along with DT classifiers to express the efficiency of the selected classifications algorithms for efficient classification, or to predict the HD cases. Despite all this, It has been noted that within five years of diagnosis, heart failure ends up killing nearly half of the population. In the recent years, there is a development of various ML models for the initial prognosis of heart failure as it supports cardiologists in promoting the diagnostic procedure. In mentioned paper [3], the introduction of a skilled system has loaded two SVM models for an efficient prognosis of heart failure. Further, the second SVM model has been standardised and it can be accomplished as an analytical model. In order to improve the two models, a hybrid grid search algorithm (HGSA) has been used which can optimize two models concurrently. Hence, the efficacy of the recommended method has been calculated by using six dissimilar evaluation standards such as sensitivity, accuracy, specificity, the charts of receiver operating characteristic (ROC), area under the curve (AUC), and matthews correlation coefficient (MCC). Additionally, the suggested method has also demonstrated finer efficiency in comparison with the other advanced ML ensemble models. Certainly, DM and ML methods towards detection and prediction of HD can be of excessive clinical productiveness, then it is extremely thought-provoking towards the development. Additionally, there is a non - availability of cardiovascular expertise and a high rate of cases that are misdiagnosed in the world's poorest nations. This problem could be solved by creating an accurate and effective early-stage heart disease prediction system using electronic patient records to support clinical decision-making. In detail, specified study [3] has meant to recognise ML classifiers with the maximum accuracy. Various supervised ML procedures have utilized and correlated with the efficiency and accurateness in HD valuation. Henceforth, the feature significant scores can be predictable for the entire feasible procedures excluding MLP and also KNN. Entirely, the features have been classified based on the significance score to search extraordinary HD predictions. Besides, the paper has established that, dataset of HD has been collected against Kaggle, three-classification built on KNN, DT and RF procedures where the RF technique has attained accuracy besides specificity and sensitivity. Accordingly, a comparison with the explicit supervised ML process can be acquainted with HD predictions with greater correctness and tremendous probable efficacy. Certainly, cardiovascular disease prediction is a life-threatening task in the region of medical data exploration. Because of this, ML has been demonstrated to be active in assisting in decisions making and in predictions against the huge data quantity generated through the health services system. Accordingly, ML techniques acquainted in latest improvements in various areas on the internet of things (IoT). So that, several analysis has afforded an indication in valuating HD with ML methods. Besides, this paper [44] is aimed at retrieving the important feature by implementing ML methods which results in the increment of efficiency. As in consequence the extrapolation method is presented with dissimilar features groupings and numerous identified classification procedures. Accordingly, this paper [2] has represented a concurrent scheme for predicting HD of medical data categories which define a patient's present health position. The predominant aim of this suggested method has been to detect the maximal ML procedures which attains greater accuracy of HD prognostication. There are two kinds of features selections, they are uni-variate feature selection and relief which has been utilized to choose the significant features of the dataset. Subsequently, by applying modified hyperparameter and cross-validation, accuracy can be improved. Further, by the incorporation of apache Kafka against apache spark as the basic structure of the model, the RF classifier has outperformed the other models to maximum accuracy. As in consequence, the concurrent progresses in the arithmetical power, processes and internet-centered memory sources, ML-based AI has rapidly been an increased domain that exists as the resolution for numerous technical and social trials. Moreover, AI intellectual has emerged as a highly successful and over subtle in universities of dutch. Whereas, the foremost reserves have made in 2018 for the development and to form the leading utilizing AI in hospitals to progress patient supervision and to decrease healthcare expenditures. Subsequently, AI has the prospective for the great enhancement of conventional arithmetical analysis in several fields and have been established to permit the detection of ‘hidden’ data in exceedingly multifaceted datasets. This paper [9] has aimed to bring together the extensive cardiovascular community along with the fundamentals of modern ML-centered AI and has described about numerous algorithms. Correspondingly, diagnosis of the disease is the great crucial health-care concern. As if the disease is diagnosed previously, the usual or by scheduled period, the patient’s life can be saved. Further, the classification technique of ML can be valuable to support the medical field through conveying dependable and immediately diagnosing the disease. To add to this, paper [18] has included the classification approaches for ML and image fusion which can be established to support healthcare specialists in identification of HD. However, ML has been briefly encapsulated with the usage of classification procedures towards the HD diagnosis. Formerly, the demonstration of certain work was achievable with the usage of classification methods for ML and in image synthesis in this region. This paper [17] has dealt with the specific ML-models that have been formed on preliminary data to classify diagnostic community, complexity of disease, and newyork heart association (NYHA) class. Additionally, models have been established to evaluate necessity for conversation at multidisciplinary team (MDT) conferences and to measure prognostication of single patients. In accordance with the automated dependent disease difficulty-score resulting against the clinical data has been related to existence on cox analysis self-sufficiently for workout, demographic, research laboratory along with the parameters of ECG. By the usage of ML in medical research is developing progressively specified the cumulative obtainability of multifaceted medical datasets. As ML represents significant benefits with regard to prognostic performance and also classifying unknown subsections of patients with exact structure and predictions, this paper [58] has represented that, the approval of several physicians and scholars are not exactly acquainted with the estimation and deducing of ML considers. However, ML specialists deprived of medical involvement can provide the particulars of the exploration which is inferior for a clinical distribution to measure. Though, the vast indication has been described, the deprived reproducible and so ML method has been treated in clinical research by suggesting the necessity of ML analyses that can be obtained in an understandable method. ML study results are explicitly focussed towards the clinical scholars. So, ML has played a vital job in the clinical field. In the indicated paper [24], ensemble learning methodologies have been acquainted with the enriched performance of HD prediction. In connection to that, two feature extraction methods have been used namely principal component analysis (PCA) and linear discriminant analysis (LDA) to choose the necessary features against the dataset. Machine learning algorithms and ensemble learning methods have been used with certain features. Subsequently, the comparison has been done among the ML algorithm along with ensemble learning methods. Hence, various methodologies have been accomplished to evaluate the existing models. As a result, ensemble learning model with the DT has obtained the finest efficiency. Coronary artery disease (CAD) extrapolation is an inflexible and thought-provoking analysis in the medical field. Besides, the initial prediction on the medical industry particularly in the field of cardiovascular is the one among the professional. Eventually, the preclusion of cardiovascular can be achieved by a diet chart arranged by the apprehensive surgeon during primary probability. Though, this paper [12] has comprised of the CAD prediction through the suggested algorithm by making of pooled area curve (PUC) in the ML model. However, the knowledge based identification is a significant feature for precise calculation. Then, this significant method has supported a good influence to regulate difference in medical images with the presence of poor pixels contiguous it. Subsequently, this pooled region structure in ML algorithm is getting lessening veins and tissues with the support of congestion and plaque in the blood vessels. Further research has been sustained the current adaptive image-based classification methods and also by relating with the prevailing classification approaches and to predict CAD prior for a greater accuracy rate. According to the indian heart association (IHA), out of four persons expire due to HD every moment in India. The age group spans between 50 and 30. Hence, one-fourth of HF morality happens to individuals who are lesser than 40. Furthermore nine hundred people deceased below the age of 30 because of various HD per day in India. This paper [31] has aimed to observe and relate the correctness of four dissimilar with ROC curve of ML algorithms for forecasting along with diagnosing HD through 14 features against UCI cardiac datasets. Among the wide range of HD, CAD is deliberated as a typical cardiovascular condition with an extraordinary demise rate. The best known instrument used for recognizing CAD is angiography. It is recognised as cost effective and has minimum complicacy. Henceforth, the objective of this paper [29] deals with the increase in the accuracy of coronary heart disease analysis by selecting significant prognostic features in order of their position. Moreover, this study has suggested an incorporated model using ML. Thus, the ML of RTs, DT, SVM and DT Of chi-squared automatic interaction detection (CHAID) are used. The suggested method illustrates favourable results and also the analysis proves that the RTs methods surpasses other models. CVD is the foremost base of universal death and it is a main community health anxiety. Hence CVD prediction is one among the current methods for controlling CVD. Indeed, in this study [68], 29930 cases of CVD has been selected from 101056 people in the year 2014. The logistic regression analysis has exhibited that, approximately 30 gauges have been connected to CVD, comprising male, family revenue, old age, smoking, drinking, obesity, an abnormally large waist circumference, low levels of lipoprotein and cholesterol, and irregular fasting blood sugar and so on. Various approaches have been acquainted to construct prediction model which includes multivariate regression model (MRM), classification and regression tree (CART), ada boost, naive bayes, bagged trees and RF. MRM has been taken as a standard for performance assessment. However, CVD prediction model has been provided for a 3-year CVD risk assessment. Subsequently, it is formed on a huge numbers with precariousness of CVD by using RF algorithm, whichever can be provided with the reference on CVD prediction and treatment work.

2.3 Review of Conventional DL Models in Heart Disease Prediction and Classification

Predicting heart ailments in the initial stages can avert possible number of deaths because of heart attacks. The research aims to use actual-life databases as it helps in predicting the possible heart attacks in a timely manner. Research [40] concentrates on potential heart disease prediction by integrating state-of-the-art dataset accessible at UCI, and convolutional neural networks (CNN). These datasets consist of cardiac test parameters and common human habits. The research outcome displays that the presented model overtakes the prevailing methods concerning enactment evaluation metrics. The presented model has achieved better precision. Multi-stage approach has been presented in [20] to anticipate heart disease using imbalanced data comprising both quantitative and qualitative features. The presented model is initiated with LASSO regression application to detect the impact of essential variables in data variation. Utilizing several instances of unsystematically subsampled datasets, check the stability of the attribute impact, LASSO is implemented recurrently. This is an important step in the algorithm to regulate true-negatives of attribute selection. The accuracy obtained in the proposed method has achieved higher precision and it is better when compared to single accurateness of RF classifier or SVM. The algorithm results in greater test accuracy and specificity with greater values of area under the curve (AUC). The paper aims to predict the accuracy level, if each patient is at threat of HD [63]. The prediction have been done using DL algorithms with training data. CNN algorithm is a way to define the risk of heart disease in initial stages with the help of structured data. As soon as the individual enters the data that is required, the algorithm is then used with the data and the outcomes are generated. The accuracy will decrease when the clinical data is incomplete. Here, the prediction model is implemented over actual hospital data. CNN algorithm has been presented to forecast the threat of the disease by using both organized and unorganized data of patients. In the presented model, the accuracy obtained have been better than other approaches. The heart related disease can be possibly controlled by predicting at an earlier stage. The paper [37] aims to study the previously done research papers in the area of DL on ECG related coronary heart disease (CHD) prediction. Paper addresses the DL applications for CHD analysis through ECG. The outcomes clearly indicate that CNN is a method having highest accuracy rate on large dataset. In implementing DL, python plays a crucial role. This helps in minimizing the risk of late analysis by prior prediction and enhances the health care quality. DL methods in clinical diagnosis does not replace the specialists, the advancement of deep learning methods will assist the doctors in early decision making. In the paper, patients having high-risk of CHD are marked as the positive samples, and the remainder is labelled as negative [5]. DL approach is used to gain embedded clinical vectors. Research has been done on actual medical datasets. The results displays that the DL can greatly enhance the prediction accuracy of high-risk CHD when compared with the state-of-the-art method. So, the presented method aids doctors for precise prediction. The paper has presented CNN model to predict the ailment in initial stage [30]. The presented model involves two conventional layers in that one is an output layer and the other one is two dropout layers. This results in higher accuracy. Additional benefits of this method are the model itself performs feature extraction, prediction and pre-processing, while the prevailing algorithm uses various methods for every task. Paper [41] have presented two stage model to efficiently predict heart ailment. The primary level comprises training an enhanced sparse auto encoder (SAE) and the next stage is artificial neural network (ANN). SAE is an unsupervised network, to understand about the training data. ANN is utilized to estimate the health status on the basis of learning records. SAE model have been enhanced by adam algorithm. The accuracy of the presented model on test data has produced better accuracy in comparison with other traditional learning methods. Heart disease is the root of mortality and premature disability in the world, predicting it has become a core problem in the area of healthcare system. The presented work offers a contribution to understand and to create an AI based long short term memory (LSTM) method for predicting heart disease [43]. In order to improve accuracy and additional prediction variables for heart ailment, a comparative approach has been presented in this paper among the LSTM and multi-layer perceptron (MLP) techniques. The paper aims in developing an AI based LSTM method for heart ailment prediction to make a revised decision to monitor and prevent heart ailment and stroke. LSTM has been demonstrated to be an essential method for resolving the issues and other additional issues than MLP technique. Implementing this method will display a significant impact on the architecture and in applying heart ailment prediction system in both, in terms of economy and healthcare. A two-level hierarchical DL framework is presented in the study with generative adversarial network (GAN) for automatic analysis of ECG signs. In the two-level, the first model is comprised of memory-augmented deep autoencoder with GAN (MadeGAN), that targets in distinguishing the abnormal signs from the normal ECG for abnormality detection [64]. The next level aims in classifying robust multi-class for diverse arrhythmias detection. The enactment valuation of the presented method is done by actual clinical data from the database of MIT-BIH arrhythmia. From the results, it is evident that, the presented framework successfully has extracted the disease-altered feature structure from the ECG signals, gaining improved enactment in HD prediction with greater enactment scores when compared with the prevailing approaches. Heart ailment is the foremost cause of mortality globally, that accounts for around one-third of the mortality. The aim of the paper is to predict HD in its early stages [61]. For this purpose three different DL algorithms are presented, namely ANN, radial basis function network (RBFN) and CNN to validate test and train with normalized, separated and selected data features. RBFN, CNN and ANN are used to create trained neural network model. Diverse evaluation methods are created to completely understand the enactment of classification. The obtained accuracy from the RBFN, ANN, and CNN show better precision. The objective of the study have been to present a reliable IoHT system that act as a medical system for decision support along with the opinion of heart ailment [15]. Based on that, it is highlighted that, the precision level can be enhanced through DL algorithm, by not requiring hybrid-complex methods, and secure data processing is attained with tangle based and multi-authentication based method. Using AEN, heart sound has been classified and IoHT system have been built to support medicos in real-time. To apprise the complete IoHT system and diagnosis infrastructure through AEN, an inclusive evaluation has been done. Initially, the AEN model is tested with the two diverse heart sound dataset. Heart ailment categorized in Pascal dataset are extrasystoles, murmur, and normal. Whereas, the physio dataset is categorized as abnormal and normal datasets. In order to develop the diagnosis infrastructure using physiobank-physionet A-training, PASCAL B-training and AEN heart set datasets have been used consequently. The presented model has gained higher precision in sensitivity, specificity and accuracy. Heart sound comprises of several essential qualities that aids in detecting heart disease in the initial stages. Several models have been presented till now and many signal-processing methods have been utilized on heart sounds for detecting HD. A methodology is presented in this paper to detect heart disease with the help of heart sounds [69]. The presented method uses three stages, namely deep feature extraction, classification and spectrogram generation stage. In spectrogram generation step, the heart sounds are transformed into spectrogram images through time–frequency transformation. From the three diverse pre-trained CNN models such as VGG16, AlexNet and VGG19, the deep features have been extracted. In the third phase, support vector classifier is utilized. The presented technique is appraised on two datasets that are obtained from classifying the heart sounds. The results gained from the present study is compared with the prevailing methods. From the comparison, it is found that the presented method overtakes other existing methods. It has been found that the huge progress in the area of DL seeks to generate AI systems that assist medicos in predicting as well as determining the aliment using IoT system. Thus, enhanced DL aided convolutional neural network (EDCNN) has been presented to aid and enhance patient diagnosis of heart ailment [50]. This model concentrates on deeper design that involves multi-layer perceptron's model by regularization of learning methods. The EDCNN has been applied on internet of medical things platform (IoMT) mainly for decision support system that aids medicos to efficiently analyse data of heart patients in cloud. The results found are compared with other conventional methods namely DNN, CNN, RNN, NNE on the basis of the analysis. The generated diagnostic system can be used efficiently to determine heart ailment risk level. Experimentation results display that a subsequent tuning and flexible design of EDCNN hyperparameters have achieved a higher precision.

Currently, many deaths occur mainly due to heart ailments. To handle this condition, analysing heartbeat sound is a suitable way to predict heart ailment. Classifying heartbeat sound is a complex problem in heart sound feature extraction and segmentation. This study presented a DL approach for classifying heartbeat sound on the basis of down-sampling, RNN for dataset-B and data framing [53]. This method is used to effectively detect the signals of heartbeat and provide data in order to identify whether additional treatment is required or not. The dataset-B is categorized into 3 types namely extra-systole, murmur and normal. By filtering the unnecessary noise, signal of heartbeat is cleaned. Data framing transfers every audio file in sampling frame to fixed-size frame. Down-sampling reduces the dimension signal of heartbeat sound wave in order to abstract additional discriminative features. The accuracy gained in the presented study is displayed as highly competitive. CVD is a life-threatening ailment and it has to be explored in the early stages. AI based detection tool can aid in identifying heart diseases. In this paper, CNN model is presented mainly due to its greater precision as well as robustness to diagnose heart diseases automatically from the sound of the heart [53]. In order to enhance precision in noisy surrounding, the presented method has employed data augmentation for multi-classification and training of several heart diseases. In this method, the heart beat sound is recorded for a certain time length. To remove the noise in the background, gaussian filter has been utilized and then pure sound of heartbeat is extracted from the recorded phonocardiograms (PCG) signals. In order to rise the flexibility to noise, collected datasets have been conquered to data augmentation method. This model comprises seven convolutional layers with several other layers. The presented model have attained the better precision to detect various heart disease. For predicting CHD, deep learning based multilayer perceptron (MLP) have been presented in this paper [39]. Specifically framingham heart study (FHS) have detected several varying values that are being enhanced by using data pre-processing methods. Additionally, seven critical features were selected to diagnose heart ailment and are coded them consequently. This study present how the use of data pre-processing through the MLP following a DL method will enhance the quality of the data while evaluating the probability of an individual having CHD. After that, by using data of 70%, the model has been trained and the rest 30% data has been utilized for testing purpose. Following which, the model has been designed by MLP DL neural network that achieved a higher precision. To predict CVD mechanized screening methods can be implemented. A non-invasive and efficient method, electrocardiogram (ECG) based methods are extensively utilized to predict CVD [13]. Therefore, this paper presents a deep

CNN to categorize five CVD by a normal 12-lead electrocardiogram pulse rates. Initially, ECG signals have been segmented into diverse intervals namely one-second, two-seconds and three-seconds devoid of wave detection, then 3 different datasets have been attained. Secondly, as a substitute to several composite pre-processing, time length of original signals of ECG have been deliberated as input along modest normalization of min-max. Finally, a 10-fold cross validation approach used for ECG signals of one-second also additional two datasets are tested. By a sequence of tests, the outcome shows that the presented model have achieved high enactment with the specificity, sensitivity and accuracy. Predicting CHD in the initial stages averts the progress of HD and leads for appropriate medication. In the paper, the capability of DL method have been shown by BiLSTM and CNN method, in detecting several CHD via recordings of PCG [4]. The algorithm attained a highest precision by CNNBiLSTM system. The progressed model only needs a minimal time period to complete both the classification and training process. This has been the limitation in former studies. Moreover, the enactment of the model has been appraised and attained higher precision. Therefore this study directs in applying DL approach in CHD detection under medical setting in order to aid clinicians in making decisions and to avoid cases from heart anomaly development. A lightweight CRNN system, cardioXNet, has been presented for detecting diverse CHD automatically without conducting several steps of pre-processing on signals of PCG [57]. This model comprises both sequence residual and representation learning by using bi-LSTM layers and CNN to extract time-invariant and temporal features. The framework displayed state-of-the-art enactment with higher precision overtaking other former studies. Besides displaying greater results, the employment from beginning to end system with considerably minimum quantity of variables make this method a highly qualified for use in several system applications. Healthcare analysis is becoming dominant with the progresses in computing, collection of data, analysing data and its classification. For classification of heartbeat, DL methods are growing important due to the accessible data namely MIT-BIH arrhythmia. The paper has focused on ECG based heartbeat classification by DL LSTM approach [28]. The diverse LSTM-DL model variants have been applied for classification, of that bi-directional LSTM Deep learning method have provided the greater precision. Prominently, bi-directional LSTM deep learning approach's comparative analysis displayed the maximum values in specificity and sensitivity with the present works. Hence, it is obvious that the mentioned models offers a precise heartbeat classification. The results from classification can be utilized to detect heart ailment and additional treatment by practitioners. Commonly used methods for predicting CHD is automatic heartbeat sound auscultation. In this paper, cardiac sound classification technique based on DL for CHD is presented [66]. This chiefly includes three parts, they are pre-processing, deep CNN with attention appliance for 1-D waveform cardiac sound area classification. To improve the data flow of CNN, an architecture of block-stacked sort by clique block is used and in every clique block, a structure of bidirectional structure have been presented in CNN. Using transition block and stacked clique, the presented CNN attains both channel and spatial attention directing to a promising classification enactment. Test conducted based on the dataset that is publically accessible have displayed that, the presented approach has attained state-of-the-art classification enactment in a substantial variables-saving way. A new method is developed for classification of phonocardiograms in terms of segmenting the complete recording of the heartbeats that are then fed to an LSTM system [25]. The last prediction has been done by taking average of the output from LSTM for every heartbeat and the result values have been compared to threshold. In comparison, a completely automated segmentation have been presented that could aid to conquer the lack of databases. This method is based on the concept of inferring PCG recording like an audio track of multi-instrument. The results attained show that using algorithms from the computational sound analysis can stand the latent to address trails in processing PCG. Like-wise approaches have been also implemented to ECG signals. A DL method is presented in the paper to recognize automatically and to categorize the 10 class of heartbeats from ECG that is vital for the analysis of cardiac arrhythmia. However, managing with a great variable function that needs a huge quantity of labelled samples, the DL model displays its complete potential. A DL based method classification system for CD detection and monitoring is presented in [46] paper. Presented DL design is categorized into deep auto-encoders (DAE) and deep neural network (DNN). DAE is an unsupervised system of feature learning and DNN as classifier. The purpose of the study is to enhance on earlier ML method that comprises of various data processing phases namely feature selection, feature reduction and feature extraction. The presented integrated model has been given better enactment when compared with the several DL models. The results obtained higher precision in specificity, accuracy and sensitivity. An algorithm for classifying heartbeat automatically by ECG signal is presented. Classifying the heartbeat automatically by electrocardiogram is vital in supporting medicos and experts in diagnosing heart disease. Paper [38], local transform pattern (LTP) using a hybrid neural fuzzy-logic system with SOM is presented. The heartbeat classification is done by five heartbeat classes namely bundle or normal branch block, ventricular ectopic, supraventricular, fusion of ventricular and unknown beat, normal. Accordingly, the presented algorithm for classification of heartbeat has displayed robust enactment of accuracy and sensitivity. This method can be utilized to predict heartbeat types effectively and expanded to medical system and healthcare, and a highly enriched diagnosis of heart ailment in medical field. The subject's ECG samples have been used as a consideration as one of the required data for HD detection in the study [52]. The classification of HD using various DL and ML have been reported in a numbers in previous work. It has been found that, the detection precision is lower when HD data is imbalanced. The goal of the paper was to identify suitable DL and ML models and to develop and test the necessary classification models in order to advance HD detection. The GAN model is selected with an aim to manage imbalanced data by creating and utilizing supplementary false data for prediction. Besides, an ensemble approach by GAN and LSTM have been established in this research, which displays higher enactment when associated to single DL approach. The enactment ranking of the presented model is on the basis of two datasets. Future work can be done by selecting other diverse ensemble models and utilizing additional diverse datasets and enactment can obtained in the same manner and compared.

2.4 Research gap

From the review of above conventional works, certain significant drawbacks are emphasized in this section,

- Though the complications have been faced, ML methods can be refined to deal with heart failure management. In future, this method may lead to an accurate diagnoses and more accurate treatment [47].
- Additionally, greater the dataset's level of lost values can have an adversarial effect. Therefore, the demonstration to tackle the issues has been suggested but the dataset used in this method, can miss some significant values, which may create issues [26].
- Moreover, as an alternative of GA and PSO, other evolutionary approaches, such as artificial bee colony (ABC) and african buffalo optimization (ABO) where parameter optimization and also feature selection can be functional. In accordance with this, the DL

methods can be utilized as if additional inclusive CAD data can be available. Correspondingly, a huge CAD dataset with more features can be examined by using various ML models [1].

- Besides, the future sequence of this research can be effective with various combinations of ML methods towards finer prediction methods. Innovative use of the feature selection techniques established to acquire a comprehensive sensitivity of the important features to escalate the effectiveness of HD prediction [44].

- However, by using effectual open source big data methods and also data mining approaches can effortlessly ensure the identical work. Subsequently, with small modification, the identical system can prognosis others diseases, and also it can be protracted to other area. As of future work, the research is aimed to incorporate actual data sources, like sensor data, mobile devices, and societal media data towards system for interrelating user's requests [21].

- In the future, both the dietary and nutrition data records must be taken into consideration and it acts as a supplementary predictor factor for detecting CHD. Dietary features play a vital part in the occurrence of CHD and prediction precision of CHD through involving supplementary dietary features can be discovered. But, with suitable endogeneity treatment, dietary data addition is likely to offer additional insights and enhanced precision in diagnosing CHD [20].

- Deep risk involves several limitations. In first case, the enhancement of deep risk relies on high-grade temporal data. In second case, its efficiency on prediction work for any diseases have to be validated further. In upcoming years, improvement has to be made in high-risk detecting task by using additional latest approaches for demonstrating learning and integrating additional sorts of medical data, namely, medical records and notes. Also, generalization and scalability have to be improved, thus, it can be implemented to resolve additional medical problems [5].

- In secured IoHT system, supplementary enhancements for creating an integrated highly progressed medical decision support mechanism, that is also capable to detect different kinds of ailment and other supplementary works have to be developed in the upcoming years. The future work includes adding substitute analysis and methods of treatment, carrying out supplementary works in order to test the security level. Additionally, developments for the data reliability have to be frequently analysed as upcoming years appears to be growing a digital world with additional sensitive data [15].

- In future, the focus need to be on enhancing the enactment in terms of sensitivity, specificity and accuracy and also improvement in relation to efficiency. Additionally, in the event of data resource supplement, the focus must be to categorize more classes of ML. Furthermore, the pre-trained models must go through transfer learning including several datasets to identify CVDs and risk parameters [13].

2.5 Summary

The chapter discussed about the heart disease prediction and classification with the aid of artificial intelligence (AI) based algorithms namely deep learning (DL) and machine learning (ML). From the existing studies, in terms of ML approach, SVM model was extensively used and in terms of DL approach, CNN model was widely used. This chapter has elaborated on prediction and classification of heart disease with existing research studies

CHAPTER 3

METHODOLOGY

3.1 Introduction

All four methodologies given here are related to heart disease prediction which are compared with other existing methodologies. The methodology one is given in chapter 4, hybrid deep progressive model is implemented on 2 datasets with 4 optimization algorithms Genetic Algorithm (GA), Particle Swarm Optimization (PSO), African Buffalo Algorithm (ABO) and Genetic Sine algorithm (GSA). All models are compared with existing methods and GSA with deep progressive attention model has shown better accuracy on dataset 1 compared to other methods. The methodology 2 is given in chapter 5. In chapter 5, stacked sparse auto encoder model with penguin optimization algorithm (POA) is implemented. The stacked sparse auto encoder model in this chapter has compared with other existing methodologies and shown the better accuracy. The methodology 3 is implemented in chapter 6. IoT based system data is collected with various sensors and tested with novel classifiers implemented as per methodology 1 and methodology 2. In this chapter, binary and multi class classification is performed on the collected dataset. The classifiers used for comparison are RNN, LSTM, Stacked Sparse Auto encdoer (SSAE) deep progressive hybrid models. These all models are tested with five optimization algorithms i.e. GA, PSO, ABO, GSA and POA. Methodology 4 is given in chapter 7. To calculate vascular age, cardiac index and risk score, fourth methodology is implemented where regression models are used to predict the values of risk parameters.

3.2 Efficient Heart Disease Prediction Using Hybrid Deep Learning Classification Models

In the study, feature selection practices such as particle swarm optimization (PSO), genetic algorithm (GA) and african buffalo optimisation algorithm (ABO) are employed. For the further enhancement of the procedure, an innovatively established genetic sine algorithm (GSA) is suggested because of the capability in the selection of optimum features. Therefore, the deep learning (DL) has various benefits across conventional ML approaches which includes programmed feature learning, managing huge and multifaceted data, enhances performance, controlling nonlinear interconnections, managing unstructured and also structured data, prognostic modelling, managing lost data and successive data, scalability and overall performance. Hence, the features which are selected remain subjected towards the classification methods through recurrent neural network (RNN) that is interconnected with long short term memory (LSTM) procedures. Using these algorithms, filtration is done for removing invalid data and give weightage to the acute data, such that deep progressive attention-RNN+LSTM (DPA-RNN combined with LSTM) can be advanced by way of enhancing the classification level.

The suggested outcomes are efficient and proportional analysis can be implemented on two standard datasets i.e. HD diagnosis UCI dataset, heart failure clinical dataset. Furthermore, arithmetical analysis with regards to pearson correlation co-efficient, mann-whitney U-test, friedman rank and iman-davenport significant values are estimated. Hence, the acquired outcomes illustrates that the recommended scheme remains relatively more effective for HD diagnosis rather than other predictable methods.

3.3 Optimization Using Stacked Sparse Autoencoder Model for Heart Disease Prediction

Unsupervised feature learning approaches have established enormous consideration, meanwhile they do not completely depend on labelled dataset and are appropriate for training methods for excessive data. In recent times, ML methods are effectively utilized for the cardiovascular disease prediction. However, in initial analysis an extrapolation is essential for providing actual treatment and to evade greater death rates. Additionally, numerous classification algorithms are established in recent times which gratify the necessity, but indicates restricted accurateness during prediction of HD.

Therefore, the motivation of the study remains on premature prediction of HD and to progress the accuracy in prediction by using standard HD datasets like UCI cleveland dataset, HD clinical dataset through employing active classification and optimization procedures. Hence, the optimization procedures commonly demonstrates the advantage of allocating with multi-faceted non-linear problems with enhanced flexibility. So, emperor penguin optimization algorithm, which specifically select the finest features for classification is used in the study to develop the efficacy, minimization of restoration errors, and to escalate the quality of HD classification. Furthermore, the recently established stacked sparse convolutional neural network based auto encoder (SSC-AE) classification procedure is applied for substantial feature classification with advanced robustness and effectiveness. Then, the area under curve (AUC), accuracy, F1 score are certain measures used to relate the consequences of numerous ML procedures towards the recommended model. As the results express that the suggested method, SSC-AE, is higher than the further classification models.

3.4 IoT based Heart Disease Prognosis Multi Class Classification using Deep Learning

One of the important technologies for everyday life is IoT. As IoT devices are capable of collecting and analysing large datasets, and therefore they have a greater potential for clinical research purposes namely disease prediction. This study helps in forecasting the absence and presence of HD with its brutality level for that it presents a system in order to monitor a person's health status from real-time IoT on the basis of sensor data. This study attempts to prognosticate HD from real-time IoT dataset with appropriate usage of data mining procedures. Although, traditional researchers have attempted executing this, they lacked in precision rate and many of the studies did not consider real-time datasets. In the early stage, real-time information is gathered by IoT sensors. Once the data is collected, pre-processing is ended by rejecting unwanted values. After this, label-encoding is done by changing categorical-columns to numerical system because ML model only considers numerical data system.

This data is then fed into test and train split where 20% is tested after training the 80% of data. Subsequently, the feature ranges inside the dataset are then normalized as part of the feature scaling procedure. Real-time data normally include elements that appear to vary in terms of magnitude, units and range. Scaling of the features are done in order to comprehend them on a comparable scale. Additionally, this method enables quicker training and prevents the optimization procedure from being stuck at the local optimum. The feature selection procedure that comes next is essential for getting better results. By selecting only the important variables and avoiding redundant and irrelevant variables, this technique helps to improve the algorithmic predictability. After collecting data, feature selection is done by employing appropriate ML based meta-heuristic. Lastly, the data is categorized by DL based methods.

This study takes into account five diverse ML based meta-heuristics for selecting features that includes GSA, GA, POA, ABO and PSO. In final stage, classification is achieved using DL based classifiers i.e. SSA, RNN, DPA-RNN+LSTM and LSTM. The study collected 312 data from persons by IoT sensors. As the collected data was insufficient, the present study has employed SMOTE method for enhancing the data. By this approach, the data within datasets are enriched in balanced way and a total of 1591 data have been conquered. In this case, around

11 features have been considered namely systolic BP, heart rate, total cholesterol level, oxygen saturation, PR interval, haemoglobin (HB), QT interval, diastolic BP, total cholesterol, high density lipoprotein (HDL) cholesterol, and low density lipoprotein (LDL) cholesterol level. In the study, diverse sensors have been employed for sensing 11 features. Enactment of the presented method has been evaluated for multi and binary-classification. Here in this condition, analysis is done by building confusion matrix to identify incorrect and correct detection by which efficiency of the presented system can be established. This study has taken into account the real-time data, when compared with traditional works are not feasible in this case. On the other hand, multi-classification aided in determining normal, mild, moderate and severe cases. During the implementation phase, condition-based execution have been employed to find multi and binary classification. However, a confusion matrix has been obtained for the suggested system to confirm its effectiveness. Overall presented system is validated in terms of performance metrics for monitoring its original performance. The performance of the proposed methodology in disease prognostication has been confirmed through the results. This procedure is conducted for gaining enhanced grade calculation in order to achieve ideal multi and binary-classification. In this research, both binary and multi-classification have been performed where binary classification is achieved to identify if the disease is present or not.

3.5 IoT Based Heart Disease Prediction Using Regression Methods

Several learning approaches have been employed by researchers for understanding the HD, but they lacked in precise detection and classification. Therefore, ML and DL algorithms with meta-heuristic attribute architectures have been applied in this study for classifying HD and regression on the basis of IoT extracted sensor dataset. These algorithms automatically learn the features. The presented model has employed five diverse optimisation approaches namely POA, PSO, ABO, GSA and GA methods. DL methods namely LSTM, RNN, stacked sparse auto-encoder and deep progressive attention of LSTM and RNN have been used for HD classification and regression detection to identify the heart condition of the individual, on the basis of IoT retrieved dataset. Regression is detected by assessing the variables namely vascular age, cardiac index value and risk score by framingham equation. At last, classification of HD and regression have been performed by flask web page model.

This study has intended to implement HD classification and regression of IoT centered dataset. It concentrates on forecasting minimum error value, by assessing the MSE, MAE, and RMSE on basis of diverse optimisation techniques with LSTM, RNN, deep progressive attention LSTM-RNN and SSAE to predict values of heart regression.

In the study, data have been gathered from sensor dataset that performs as terminal node for collecting signals on IoT. Numerous real-time data, about patients have been collected by wearable IoT based sensors. The collected sensors are ECG sensor, BP sensor, glucose sensor and SPO2 sensor. To gather diastolic BP, pulse rate and systolic BP, the BP sensor have been used. ECG sensor have been employed to identify RR, QT intervals then SPO2 estimates the value of oxygen saturation. The loaded and fed data have been used to pre-process the acquired dataset. The crucial step in the ML cycle is data pre-processing, that carries out data analysis and boosts the system's efficiency and precision. Label encoding has been performed to the pre-processed data. The category characteristics are label encoded since supervised learning algorithms can only handle numerical input. Data have been grouped for the purpose of data splitting after the targeted qualities have been labelled. Data have been splitted into testing and training phase, wherein 20% of the data is tested and 80% of data from dataset is trained. Data testing is done to evaluate the model execution, based on testing and training data, the finest model has been selected. On trained data, feature scaling is done that have been encoded with the label. This method is employed to regulate the self-determining attributes of datasets in a particular range. In data modelling, feature selection is the vital and a beneficial step that aids in removing unrelated noisy and useless data.

The model having high accuracy is employed to make detections after training process, testing, and validating for precision and low rate of error of the models employed in model training. This trained model is imported into a site and stored. In this case, the flask web application system is used to apply and build online users to input personal information about heart illness. These inputs are evaluated by a trained model to determine whether the patient has heart disease or not, and to make a forecast. On the basis of cardiac risk score, vascular age, and cardiac index readings, regression is also anticipated. Later, the web page displays the existence of HD also denote, the cardiac risk level. Additionally, cardiac index value and vascular heart age is shown in the web page displays the heart condition according to the selected variables.

With the help of the web page, the users can gain information about the vascular age, cardiac index and heart risk score. Prediction of HD is detected by updating the input variables like age and ID of the patients. By a click, the model detects the input value and the outcome will be presented based on attributes entered. The obtained results will help in identifying the normal and abnormal heart. Model having less error rate is predicted and employed for HD categorization and regression. To reveal the effectiveness, presented model is calculated based on certain parameters namely accuracy, sensitivity, recall and F1-score.

CHAPTER 4

EFFICIENT HEART DISEASE PREDICTION USING HYBRID DEEP LEARNING CLASSIFICATION MODELS

4.1 Introduction

The most important goal of the study is the implementation of HD prognosis in humans. The outcomes of traditional processes have a high false positive (FP) rate. Hence, to resolve this problem and to predict HD with extreme accuracy, the recent study projected an enhanced optimization approaches, and hybrid classification methods.

4.2 Proposed System

In this proposed method, the flow of the process and the selection of datasets are shown in the figure 4.1. The process workflow originates with the dataset collection. Consequently, data pre-processing is achieved to create data at ease for utilization and explanation. The already stated procedure furthermore contributes in excluding the undesirable effect of the prediction rate of the model. Subsequently, it prevents the replicates or variations in the data. The processed data is fetched to train and test separately. Thus, scaling of feature can be done towards the optimization and shorten training times. Later on, feature selection is performed by using the suggested methods which comprises of ABO, GA, and PSO along with the recently established GSA. As a result, important features are selected such that the classification process can be enhanced. Following this, classification efficiency is increased of RNN and LSTM combined with DPA-RNN+LSTM. The consideration model, specifically focused on the important input consequences during this process, and as a result, there is a learning relationship between them. Further, the suggested method is examined with regards to key metrics like recall, accuracy, and the f-measure. Then, this section demonstrates the work flow on the overall structure. The figure 4.1 signifies the consecutive stages in the execution technique of the classification and feature selection procedures.

4.2.1 Data pre-processing

The pre-processing is done using the two stage process. Hence, the lost data values are relieved by using the mean-value features and the components of the string values are reformed towards the numeric classifications.

4.3 Meta-Heuristic Algorithms in Feature Selection

Several feature selection, feature extraction methods are there. They are GA, ABO and PSO together with DL classifiers.

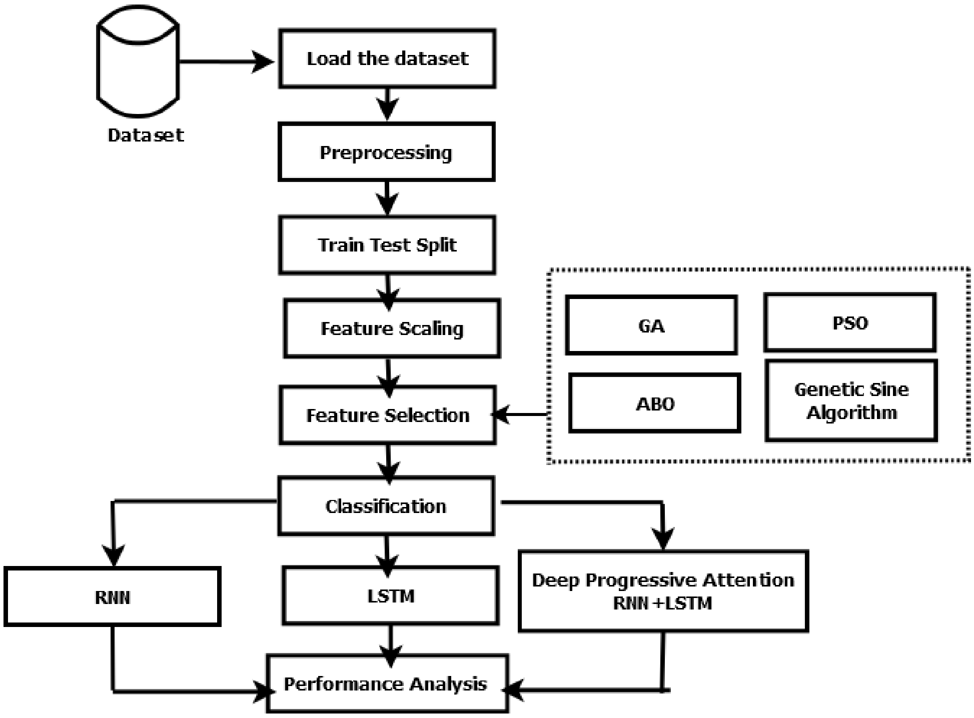

Figure 4.1: Recommended Flow of HD Prediction

4.3.1 Genetic algorithm (GA)

The GA is taken with the probabilistic process that represents the procedure of natural-evolution by utilizing these constraints. Hence, the GA uses the cross-over operators and the mutation operators. As the cross-over operator is used for matching entities prevailing in the parent population. Herein, the cross-over operator, each features are transformed at random, and various off-spring outcomes for the mutation-operators. In the specified algorithm 1, the created parent's representative is created by the off-spring. Naturally, the unique-point symmetric cross-over along with mutation can be obtained and attained by the bit-flipping procedure. Further, the minimization issues is inferred to the representation which is as follows:

$$fit = \alpha F(D1) + \beta \frac{|sf|}{|Af|} \tag{4.1}$$

Where, Classifier error rate is αF(D1), the subset of selected features length is sf , entire count is Af. β and 1- α are parameters which is used in controlling and classifying accuracy weights and feature reduction.

Selection - Process :

The selection of optimized population section is processed for the following generation breed. Based on the estimated fitness value is used in the equation (4.1) and hence the selection process is calculated.

Crossover - Process :

For breeding process, there is a selection of two parents randomly to obtain the normally selected pool. Hence, the similar process is continued without the attainment of population size towards the appropriate level. At one point, the cross-over procedure can occur, and this can be measured as the midpoint solution of the parent.

Mutation - Process :

The random resolutions are selected from the candidates selection for the determination of breeding and the bit-flipping process is performed. Hence, the differentiated solution set is progressed. This can represent the characteristics of the parents. Hence, the probability measures of the mutation process can retain with the mutation for the frequency feature.

In this regard, the GA algorithm is shown in the table 4.1. which is represented as, the initial stage. In the initiating the population size as M, $m-site$ value, $Prob_c$, max_{it} as function values. Hence, for each solution, the population is initialized at random and it is represented as $y_i = (y_{i1}, y_{i2} \ldots . y_{iD})$. Subsequently, ensuring the estimation is recursive if the standards are distinguished.

1. The calculation of fitness value is utilized by the function representation $f(y_i)$.
2. The breeding sector population is selected and it is expressed as $x_{value} = N_{\frac{Top}{2}}(fit-sort)$.
3. The selection of randomized value is greater than the probability function $Prob_c$.
4. Concurrently, the randomized mutation instances are chosen from y_{value}.
5. Then, the spontaneous innovative solution approaches are recognized with the recent prevailing solution.
6. The randomized value is made out, and the value appears to have greater value than the $Prob_m$ and the randomized mutation of the instance is chosen against the y_{value}.

Table 4.1: GA Algorithm

step 1. Set the initiator population size values M, m-site, $Prob_c, Prob_m, max_{it}$
step 2. At random population is initialized as $y_i = (y_{i1}, y_{i2} \ldots y_{iD})$
step 3. By using eqn 1, to assess the fitness of every solution step 4. previousfit ← fitness
step 5. Initially c as 0
step 6. To calculate fitness-sort=sort(fitness(y))
step 7. $x_{val} = \frac{M_{Tep}}{2}(fitnesssort)$
step 8. for k in range(0,M/2) do
step 9. $j\_k = f(\frac{M}{2} * rand) + 1$
step 10. if $(Prob_c > rand)$ then
step 11. $[x_{newval}(jk), x_{newval}(jl)] = crossover(y_{val}(jk), y_{val}(jl))$
step 12. fitness functions of each new solutions using $f(y_j)$ eqn (1)
step 13. if (fitness < previousfitness) then
step 14. $y = y_{newval}$
step 15. previousfitness=fitness
step 16. loopend
step 17. loopend
step 18. if $(Prob_m > rand)$ then
step 19. selection of m-site
step 20. $il = f(\frac{M}{2} * rand) + 1$
step 21. $y_{newval}(jm) = mutatex_{val}(jm)$
step 22. calculate fitness for every solution
step 23. if (fitness < previousfitness) then
step 24. $y = y_{newval}$

step 25. previousfitness=fitness
step 26. loopend
step 27. loop end
step 28. $y = mixy_{val} \& y_{newval}$
step 29. loop end of for
step 30. $[fit_{best}, i] = min(fit)$
step 31. $Sol_{best} = y(i)$
step 32. c= c+1
step 33. Until(c$<$ max_{it})
step 34. value return Sol_{best}

7. Updating the new spontaneous resolution with the further prevailing method outcomes.

8. The incorporation of x_{value} and $x_{newvalue}$ is generated. This is employed as the new creative solution. Henceforth, the results are enhanced, then effective solutions are obtained, and also it is measured as the best approach.

4.3.2 Particle Swarm Optimization (PSO)

The PSO optimization is a searching method based on population. Formerly, this method is denoted from the information sharing among the birds. As in the initial stage of PSO, the randomized population particles are set and the set is based on the further exchange particles. Then, the exchanged particles are progressed forward by the population absolute velocity range. Therefore, in each repetitive process the modified outcomes and the finest performance resultant values are obtained. Moreover, the velocity particles are reorganized based on the information. Hence, the weights are denoted as variables. Subsequently the algorithm which is represented below represents the transmission functionality type however, this procedure uses the modification of endless-value towards the binary-value. Hence, this is the relieved procedure towards the common tangent functionality. Therefore, each and every particle is represented as depicted and it is expressed by the D-dimensional vector. And so each particle is initialized at a random.

$$y_i = (y_{i1}, y_{i2} \dots y_{iD}) \in B_s \tag{4.2}$$

Where, the availability of the search space is denoted as B_s.

The velocity is represented as below and it is primarily initialized as 0 by the dimensional D-vector.

$$w_i = (w_{i1}, w_{i2} \dots w_{iD}) \in B_s \tag{4.3}$$

The highest portion set down for each and every swarm particle is represented as:

$$q_i = (q_{i1}, q_{i2} \dots q_{iD}) \tag{4.4}$$

The above algorithm 2 represents the swarm value size as N, and A_{c1}, A_{c2}, $w_{maximum}$, $w_{minimum}$, v_{max}, max_{it} values of the acceleration constant are initialized. As per the equations 4.2 and 4.3, the population and the vector velocity are initialized. Moreover, the estimation is performed recursively until the final conclusions are acquired.

The following are the stages of the process:

1. The updating weight of the inertia-value-w recently.
2. Updating the fitness functions of each solution by using the function $f(y_i)$.
3. Generally, the p_{best} and g_{best} are given as the test solutions.
4. The velocity of each and every particle is expressed accordingly with each repetitive procedure.
5. By utilizing the transfer functionality, the uninterrupted sequential values are mapped

Table 4.2: PSO Algorithm

step 1. Intialize size of population of swarm as M, Accl. const. B_{c1}, B_{c2},
step 2. $x_{maximum}, x_{minimum}, w_{maximum}, max_{it}$
step 3. Randomly usage of y value by using the eqn 2 and initiates population of every sol. and w is vector velocity where dimensional D as 0 vectors
step 4. c$\leftarrow$ cn
step 5. initializing cn as 0
step 6. $x = x_{max} - cn(\frac{x_{max}-x_{min}}{max_{it}})$
step 7. Every solution of fitness function is calculated using $f(y_i)$ eqn (1)
step 8. To assign value of p_{best} as well as $gbest$
step 9. for i range(1,M) do
step 10. $w_i^{c+1} = xw_i^{c+1} + B_{c1}rand_1(p_{best}(c) - y_i^c) + B_{c2}rand_2(gbest - x_i^c)$
step 11. loopend
step 12. Updation of practical velocity as w
step 13. for j range(1,M) do
step 14. for k range(1,D) do
step 15. If $w(Ij) > w_{maximum}$ then
step 16. $w(j,k) = w_{maximum}$
step 17. loopend
step 18. If $(w(I,j) < -w_{max})$
step 19. $w(i,j) = -w_{maximum}$
step 20. loopend
step 21. $l = w_{maximum}\frac{1}{1+e^(-w(j,k)}$
step 22. If $(random < l)$ then
step 23. y (j,k)=1
step 24. loop else

step 25.$y(j,k) = 0$
step 26. loopend
step 27. loopend
step 28. loopend
step 29. c=c+1 & y_{newval}
step 30. Until $c < max_{it}$
step 31. return gbest

towards the consistent binary-values. Henceforth, the resulting solution is created.

6. At last, the generation of the finest resultant is considered as the effective detected solution.

4.3.3 African-Buffalo Algorithm

To prevent the optimization issues of the data, the african buffalo algorithm (ABO) is executed. The table 4.3 represents the ABO algorithm is as follows,

In ABO algorithm, where the process and search of buffalo's waa signals is denoted by x_m and the buffalo reference is indicated as m. The buffalo's maa signals are demonstrating the stay to move the signals which is denoted as E_m. Hence x_m labels the application of further investigation. Another time $E_{m'}$ represents the request features of utilization process additionally. Let lf_1 and lf_2 are represented as the learning parameters. Then, the random number r is selected to attain the value is between 0 and 1. Hence the value have a dependence on the exploration issues. As if it shows the greater r value, it can show extra utilisation and less investigation is shown. Thus ABO algorithm has the ability to rectify the optimization problems which includes the estimated management issues, travelling salesman problem, and numerical call-function optimization procedure and parameter regulation mechanism of PID controller for the determination of automated voltage regulators and so on. In this segment, the procedure of ABO is achieved on the 16 metrics test-issues for optimization and hence the 16 optimization test-functionalities are expanded. The functionality selected, represents the differentiated set-

Table 4.3: ABO Algorithm

step 1. Initialization process:Buffalos are initialized at random place to nodes at the sol-space
step 2. Utilization of buffalos' explotation are updated from the eqn:
step 3. E'_m=E_m+l_f1 $(f_b - x_m)$ + $l_f2(i_b.m - x_m)$
step 4. Whereas E_m and x_m indicates the moves of exploitation and exploration, of the m_{th} buffalo $(m = 1,2...N)$; l_{f1} and l_{f2} is the learning factors
step 5. f_b is the best herd fitness and i_b is the herd's best location of the single buffalo
step 6. Updation of buffalo location with the eqn x'_m = x_m+ E_mr
step 7. Updation of $f_b$max ? Yes, updating, then go to step 5 else go to step 2
step 8. If the stopping rule does not satisfied or meet returned to update buffalos, else then go to next step

Table 4.4: ABO Solutions Steps

step 1. Intialization of buffalos at random in the solution-space
step 2. The parameters of ABO are initialized as l_{f1} and l_{f2}
step 3. Updation of buffalo's exploitation.The i_b for every buffalo and the overall herd f_b is determined in the eqn
step 4. Updation of the buffalos location in the second eqn step 5. Verify to assure the updation process of f_b, if yes then proceed
step 6. If the verification is not done, go to the second step
step 7. Validate the stopping-rules. If validation is completed then go to step 7 Else go to step 3
step 8. The effective outcomes are resulted

tings of global optimization. Which also comprise the encounters from the constrained issues to the un-constrained issues, single model problems to multi-model and independent issues to dependent difficulties.

In accordance with the algorithm, the search space is considered comprehensively, then two-equations can be applied to control the movement of buffalo on the solution space. Further, the democratic equations are depicted as:

$$E'_m = E_m + 1(f_b - x_m) + l_f 2(i_b.m - x_m) \tag{4.5}$$

The decision equation is expressed as:

$$X'_m = \frac{x_m + E_m}{\lambda} \tag{4.6}$$

The solution steps are specified as above. Hence the table 4.4 denotes the detailed methodology for generating the solution stages against the ABO algorithm. Further, the second equation of the decision equation produce the particular herd movement exploiting the signals of maa-move.

As in consequence, the initial equation represents the interface probability among the determined buffalos, which moves to the two challenging forces which are represented as f_b and i_b. Then accurate movement equation is the decision equation (4.6) which is engaged for perception, and also the results are obtained from the democratic equation (4.5). In this respect, algorithms such as PSO, genetic algorithm (GA), and ABO are precisely selected, because it is advantageous. However, PSO is the intelligible concept and rapid in conjunction which give optimal solution. Hence, PSO and GA depend on comparable values regarding cost-effective along with random element. Furthermore, PSO transmits the global and local searches spontaneously, when GA predominantly emphasis on global search. But ABO is normally have the capacity to ensure outstanding exploration and exploitation among the search space through consistent cooperation, communication and ethical memory of the previous individual exploits and also trapping against the herds collective exploits.

4.3.4 Genetic Sine Algorithm (GSA)

In this respect, GSA is suggested for choosing the appropriate features to obtain efficient classification. Usually, an optimization method is a genetic algorithm (GA) and experimental technique stimulated through natural evolution. Hence, this procedure moves the natural evolution concepts to the computerised domain which repeats the natural evolution. By nature the newly organisms acquire reformed with the atmosphere by evolution. Hence, the solution brings the different issues in indistinguishable method. Further, this method is a subset of the evolutionary techniques used for calculation and this method results with early solutions. Formerly, it is relocated to global and optimum resolution, and it breaks the searching process while it achieved the terminating circumstances. On conflict sine algorithm (SA) depend on population so that results with random search sets or else solution agent which are stable at random search space favourable towards the concern with optimization. In such instance, the search agents acquire support in optimum solution. This technique also includes the search agents compliable for precise population and retains the subsequent finest solution by accomplishing the population search agent with various repetitions. Hence, this can also be used for choosing the optimum feature set for improving the classification performance by the decrease in inappropriate features. In general GA is adaptable to several issues with simple large and distribution ability of search space towards the exploration of global optimization for evading the trapping of local optimization. The SA is an efficient optimization method. Moreover, it displays the various problems because of local optimal and there is an unstable exploitation or exploration inclinations in some situations. For solving this issue, the recent study suggested an innovative development in GSA for efficient feature section accompany with GA, ABO and PSO.

As illustrated in table 4.5, the dataset, feature sets, data size, size of population, cross-over rate and the mutation rate are utilized as input. Then initialization of population size, data size and population of individuals are done and the fitness of the individuals are calculated. Subsequently, the generation index is set. As in consequence, selection, cross-over and mutation operations are established which leads to innovatory population. Later, computation of the fitness value is done by using sine cosine algorithm and hence there is an enhancement of the previous individual's fitness value. Consequently, the fitness of the current individuals are estimated by recent dataset and so a unique population is created. Hence this procedure replicate until the population includes only a distinct individual.

4.4 Deep Learning Approaches in Classification

In this respect, LSTM, RNN and DPA-RNN+LSTM are suggested for classification procedure in order to distinguish the occurrence or non-appearance of HD.

4.4.1 RNN Technique and LSTM-Algorithm for Classification Process

A specific type of artificial neural network is a recurrent neural network (RNN). which utilizes the progressive data or else timing sequence data. Hence, these DL algorithms are normally used for temporal or else ordinal issues, namely natural language processing (NLP), recognition of speech, translation of language and also image captioning; which are integrated in popular applications. Hence, it is a prominent by the "memory" because the information is taken from

Table 4.5: Algorithm of GSA

Input: Feature-sets FS = f1,f2,...,fm dataset DS, initial data size iniD, initial population size N, crossover rate c, mutation rate m Output: Optimum feature subset
step 1. set size of population cur PopSize = cPs
step 2. Set data size curDataSize = iniD. To get data subset with the size cur-datasize from the dataset DS as the recent dataset
step 3. initialization of population P of cur-popsize individuals.
step 4. estimate the fitness of every individual in P, using the recent dataset

step 5. To set generation index GI = 1
do
a. carry-out selection, crossover, mutation operations, resulting in a new population P1
b.If GI = N/2
i. cur-popsize = N/2; N= cPs;
ii. cur-datasize =
2 * curdataSize. Get a data subset with the size of
cur-datasize from the dataset as the recent dataset
The subset doesnot contains the old dataset in the previous step
To check the search agent whether the value is 0, if 0 Genetic Algorithm
will be appliable.
else
sine cosine algorithm will exploit as fitness estimation
iii. Updation of fitness of earlier individuals in the recent population by the formula:
$Fitness(i) = (oldFit(i) + \sqrt{2} * newFit(i))/(1+\sqrt{2})$ in the earlier dataset
newfit(i) is the fitness weight of newfit(i) is increased by sqrt(2)
to reflex the new size of dataset.
The individual fitness i in the earlier dataset is retrieved from the previous runs.
the dataset
c. To estimate the new individual fitness in P1, using the recent dataset.

d. To create a new population P from P1, with population size cur-popsize. The selection of algorithm cur-popsize individuals with the maximum fitness in P1. e. To generate GI = GI + 1
step 6. While(cur-popsize>1)
step 7. return P

the preceding inputs to impact on recent input and also output. When the classical DNN accept the inputs and the results which are independent of each other, the outcome of the RNN be contingent on the preceding elements with the system. Whereas the imminent events can assist in defining the outcomes of the certain sequence, therefore unidirectional RNN cannot comprise these events in the predictions.

LSTMs is a kind of RNNs which can impede a long term dependences in consecutive data. Hence, LSTMs has the ability to procedure and investigate progressive data, like time sequences, manuscript, and language. Moreover, it uses the memory cell and gates are powered by information flow, permitting it particularly to preserve or reject data as required and therefore to avoid the vanishing gradient issues which outbreaks the classical RNNs. Further the LSTMs are extensively utilized in several applications like NLP, speech recognition, and time series prediction. In this method, the LSTM classification and RNN techniques are used in the HD prediction. Thus, the LSTM model is engaged as an effective RNN method. As a result, to create the LSTM procedure, the RNN model is used against the current hidden layer and then against the previous n levels of hidden layers. Based on the RNN model, the LSTM layers, in addition to the value nodes, exceeds the evaluation issues of long-term memory RNN. Generally the LSTM framework contains three gates pointed towards the typical RNN which includes the forget gate, the input gate, and the output gate. Hence, the aim of the LSTM technique is to integrate by controlling the dependent data along with non-linear controls towards the RNN cell. However, this cell is trained and confirms that the gradient functionality is removed based on the state signals. Therefore the LSTM and RNN model specification is shown in the table 4.6 which is as follows:

Table 4.6: RNN-LSTM model specification design

RNN-simple layer(No.of units-100)
LSTM-specification(layer-no of units-100)
Activation of softmax function
Usage of optimization-Adam
No of epochs-100
Batch size-4

4.4.2 Hybridized Technique of both RNN-LSTM Algorithm

In this respect, RNN and LSTM are incorporated because of its definite advantages. As in consequence each of the LSTM layer retains linear-recurrent prognosis layer which can improvise the prediction rate. Hence this framework has the capability to make effective parameter usage in the model compared with others and it congregates rapidly. Further it links to the LSTM layer with the input layer. However, the recurrent connection with the LSTM layer have a connection from the cell output units to the cell input units directly. Besides the cell output units are also connected with the output layer of such a network. As in consequence, LSTM-RNN uses the perception of gate-functions which regulates the activation value of the neuron's transforming or passing. However, LSTM-RNNs also uses optimally by providing the space through many layers. Because of that the function also results the input and layers are of same size. Then, LSTM-RNN also efficiently models dependencies in long term. As the tag sequence is denoted as $b = b_1 \ldots b_u$ according to the token sequence which is represented as $Z = z_1 \ldots z_u$. Further, the input layer (z1) and the hidden-state earlier step as $m_(u-1)$, the calculation of the hidden-layers and also it experience the hidden-layers computation.

$$m_u = \delta(A_{zm}[m_{u-1}]) \tag{4.7}$$

$$q_u = softmax(A_{mb}m_u) \tag{4.8}$$

$$\widetilde{b_u} = argmax\ q_u \tag{4.9}$$

At this notation denotes the matrices, and it positions the weights measures between the hidden and the input-layers and also among the output and hidden layers. Hence, the symbol δ denotes the activation functionality that is softmax or sigmoid.

$$\pi_u q(b_u|z_1 \cdots z_u) \tag{4.10}$$

The LSTM-gates formulation is expressed and the evaluation technique is represented as follows.

$$\begin{bmatrix} c_u \\ d_u \\ e_u \\ f_u \end{bmatrix} = \begin{pmatrix} sigmoid \\ sigmoid \\ sigmoid \\ softmax \end{pmatrix} A_u \tag{4.11}$$

At this stage, the notation of softmax and sigmoid are applied element-wise. Hence, A_u denotes the weight matrix Here in the notation, softmax and sigmoid were implemented element wise. A_u represents the weight-matrix.

$$j_u = d_u \odot j_u + c_u \odot f_u \tag{4.12}$$

$$m_u = e \odot soft - max(j_u) \tag{4.13}$$

In addition to the hidden layer vector A_u, the LSTM technique also accomplishes the memory vector j_u. However memory vector can be selected to write and read or else to reset the utilization of the gating mechanism and the sigmoid functionalities. As the input gate c_u can decrease the memory vector j_u. Hence the output gate e_u is denoted for scaling down the output level to obtain the resultant m_u.

4.4.3 Deep Progressive Attention RNN+LSTM

As in consequence, the attention mechanism generally has the capability to emphasise the diagnostic information and also filter the entire invalid data consistent towards the task by utilizing the attention mechanism which improve the efficient classification information. Later, it inspires the recent study which improves the innovative procedure with attention mechanism so called DPA-RNN+LSTM. The study deliberates various advantages. In general, RNN has the capability to process the inputs which has indefinite size and is able to recollect the individual data within a time setting which assists to predict the time series. On the other hand, LSTM give many parameters such as input biases, learning rates and the output biases which creates it to exclude the prerequisites for fine-adjustments. The current study recommends the recent technologically advanced algorithm. Hence the entire procedure of DPA-RNN+LSTM is illustrated in figure 4.2 and the structure of DPARNN+LSTM is exposed in figure 4.3.

Here, the attention mechanism formulae is expressed as

$$v_{ot} = \tanh \omega_p Y_t + b_p \tag{4.14}$$

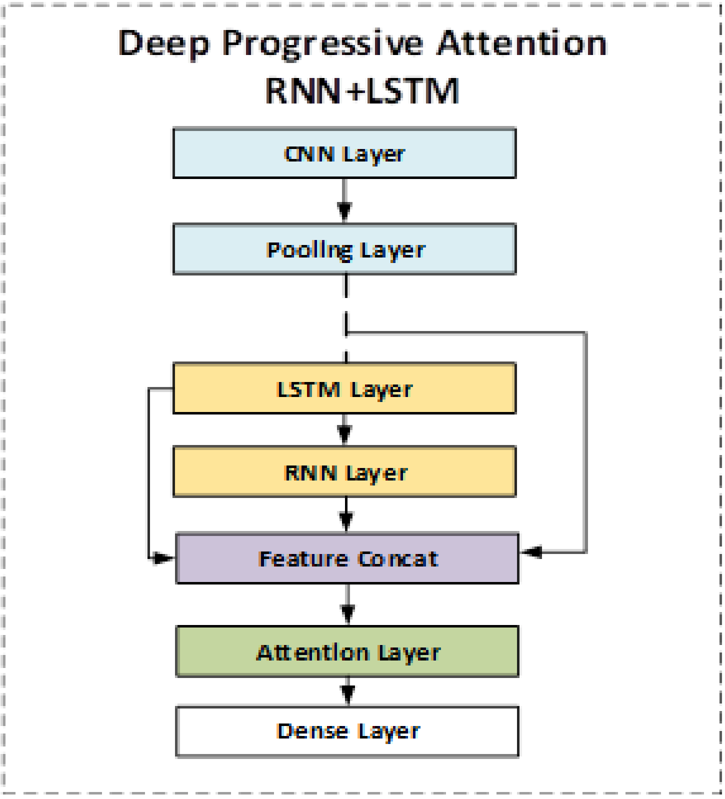

Figure 4.2: Process of DPA+RNN+LSTM

$$\alpha_t = \frac{exp((v_{\mathrm{ot}})^T \, v_P)}{\sum_t exp((v_{\mathrm{ot}})^T \, v_P)} \tag{4.15}$$

$$z_{os} = \sum_t \alpha_t Y_t \tag{4.16}$$

Where $\omega p represents weight - matrix$,b$_p$ denotes bias $-$ vector, α_t is used for the computing the connection among context vector v_p and v_{ot}. Using equation (4.14) DPA-RNN+LSTM is specified. Where ωp indicates weight-matrix, bp indicates bias-vector, αt is utilized for evaluating similarity between context vector (P) and vot. Finally, using equation 4.17, DPA-RNN+LSTM is given by,

$$j_u = d_u \odot j_u - 1 + c_u \odot f_u + z_{os} \tag{4.17}$$

Hence the deep progressive (DPA) hybrid model is illustrated as follows,

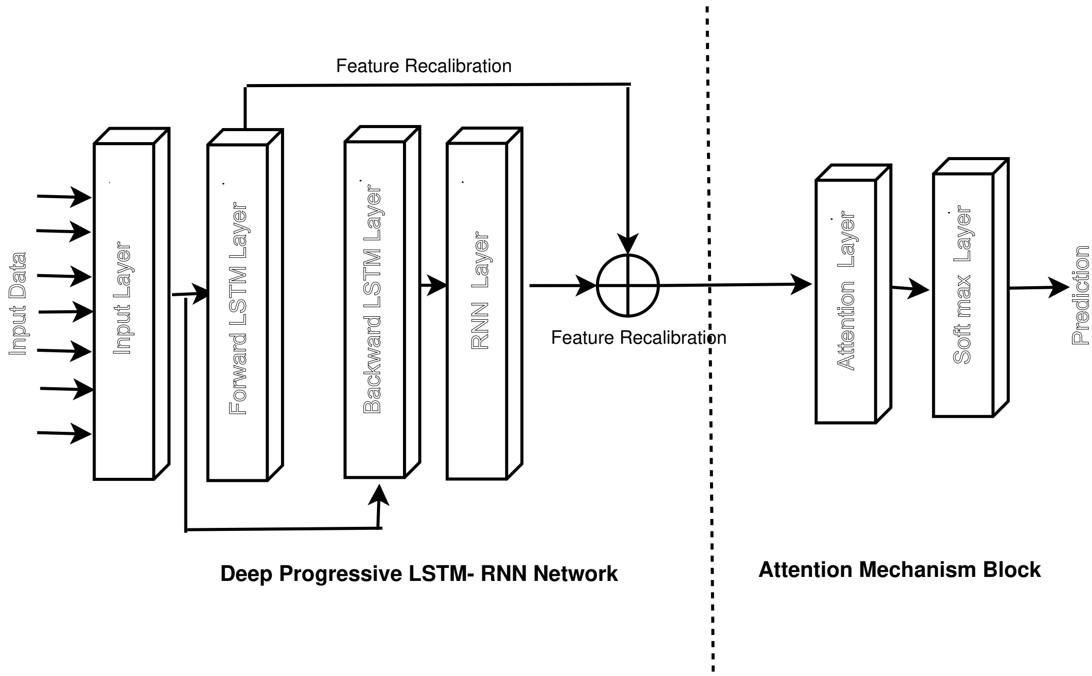

Figure 4.3: DPA-RNN+LSTM Architecture

Besides the figure 4.3 demonstrates the inclusive model for HD prediction. The framework includes two functional blocks where the initial block has forward LSTM layer, backward LSTM layer along with RNN layer. As in subsequence the attention mechanism block is utilized to incorporate of attention layer in addition with softmax layer. Hence the attention layer is used to improve the performance of the model. Finally, softmax layer is used as an activation function for classifying the occurrence or else non-occurrence of HD.

4.4.4 Performance Measures and Dataset

As in consequence, the details of the used dataset are specified for both the datasets which are utilized for training and testing. Correspondingly, performance measures are given for deep learning classifiers used.

4.4.4.1 Dataset Description

There are two types of datasets used in the application stage like HD diagnosis UCI dataset and heart failure (HF) clinical dataset.

Dataset 1: HD diagnosis using UCI-Dataset

In this kind of dataset involves of 76 attributes, on the other hand it retains a total of 14 sub-sets. Further, this Cleveland dataset is the distinct database which are used by ML scholars to be knowledgeable. Hence, the target attributes denoted the HD indention of patient data. Furthermore this value ranges from 0 to 4. In this suggested structure, Cleveland data-set HD are signified as the online dataset for prediction of HD. Hence, the dataset is obtainable the UCI-source of ML platform in virtual mode and the dataset attributes which are as follows

- Age
- Sex
- Resting Blood Pressure (BP)
- Fasting blood sugar value - 120 mg/dl
- Previous peak value
- The grade of the peak value
- 3 $\rightarrow$ normal one, 6$\rightarrow$fixed-defect, 7$\rightarrow$reversable defect
- Types of chest pain (four values)
- resting of electro cardiac resting
- Attainment of determined heart rate
- Main vessels count
- Serum cholesterol in mg/dl
- Trained persuaded angina

Further, the fields as the representatives of the social security statistics and then the names of the patient data are excluded against the used dataset. Hence the excluded values are replaced with the false values. Besides the above attributes gradients indicates the total amount of attribute fields utilized in the dataset.

Dataset 2: HF clinical dataset

As in consequence, the dataset is exposed to the feature selection processes, classification procedures and the performance measure is examined on the two datasets which lead to predictive results. Hence, the dataset against the author Davide Chicci, Giuseppe utilizes ML algorithms, and also it forecasted the endurance rate of the patient pretentious with HF problems. So, this prediction is resultant from the data of serum-creatinine and also from election fraction alone. However, the dataset of HF clinical records are utilized as the succeeding dataset and the dataset attributes which are included are as follows.

- Age
- Creatinine phosphate
- Anaemia
- Ejection fraction
- Diabetes
- Platelets
- High BP
- Serum sodium
- Serum creatinine
- Sex factor

4.4.4.2 Performance Measures

The results of prediction model are measured through the performance factors where the performance metrics comprises of accuracy, precision, recall and f-measure value. Hence, the metrics which are specified as follows.

Accuracy

Effective classified data against testing datasets are exposed in the percentage values and are symbolised as the accuracy value. Hence, the evaluation of the accuracy factor is specified below and the accuracy formula can be attained as specified below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.18}$$

Precision

Classified data of hale and hearty people are properly valued are characterised through the precision value and these values are calculated by the specified formula which is given below

$$Precision for prediction = \frac{TN}{FP+TN} \tag{4.19}$$

Recall Recall value metric describes the associated occurrences of extents which are improved. Later both recall and accuracy are dependent on the significance results basis and results which are measured. Hence, these values are projected by the formula which is expressed below:

$$Recall for prediction = \frac{TP}{TP+FN} \tag{4.20}$$

F-measure

F-measure is well-defined as a mean value of accuracy and recall of the results. Hence, the value can be calculated as below:

$$F1Scoreforprediction = \frac{2 * Precision * Reall}{Precision + Recall} \tag{4.21}$$

4.5 Results and Discussions

The performance of the suggested system is evaluated by comparing the predictable methods according to the datasets. Hence, the performance assessment of the suggested structure is accepted by including RNN, LSTM and DPA classifier as deliberated below

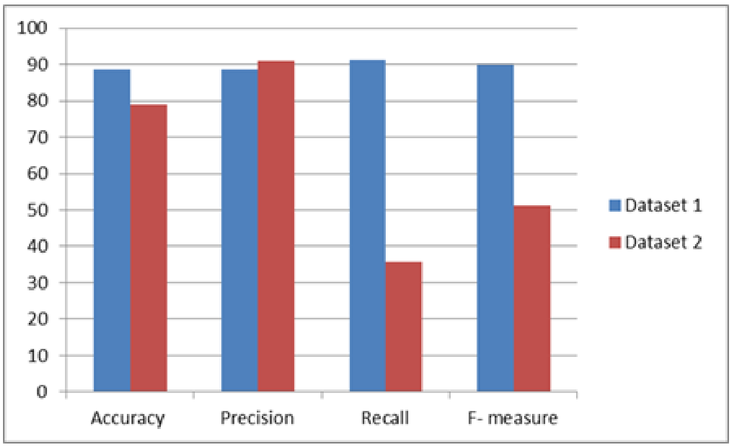

Figure 4.4: Performance Evaluation of RNN classifier without evolving GA-algorithm along with PSO-algorithm

Hence, figure 4.4 demonstrates the current results attained without applying GA, PSO and hybrid classifier methods and it illustrates the efficiency of the recent study which is based on RNN classification which express a greater accuracy for dataset :1 when compared with dataset:2.

As in consequence, figure 4.5 demonstrates the current results acquired devoid of employing GA, PSO and hybrid classifier methods and it express that the performance of the current study which is based upon LSTM classification which illustrates greater accuracy for dataset 1 against the dataset 2.

Henceforth, figure 4.6 demonstrates the current results attained with the application of the GA algorithm method and it expresses by implementing GA for feature selection which is based

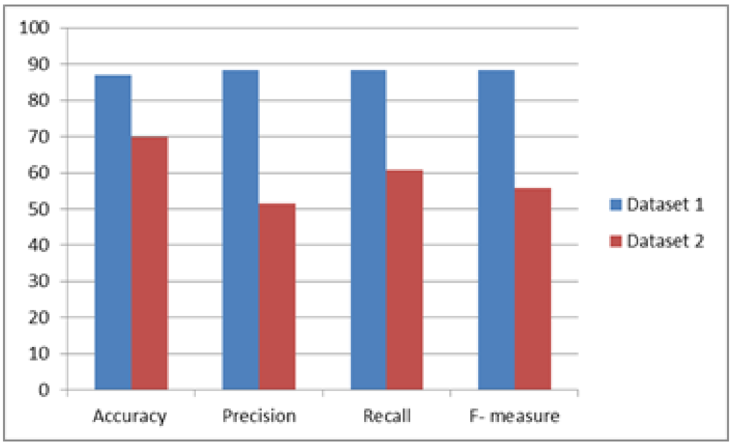

Figure 4.5: Performance Evaluation LSTM classifier without evolving GA-algorithm along with PSO-algorithm

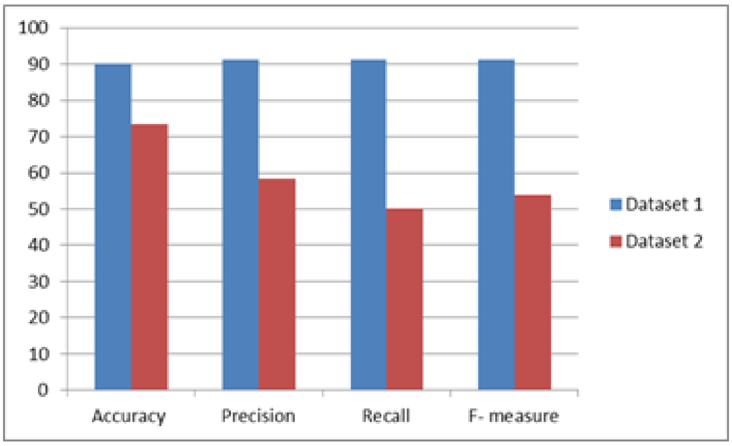

Figure 4.6: Performance Evaluation of RNN classifier with GA algorithm with feature selection

RNN classification and it illustrates a finer performance for dataset 1 against dataset 2.

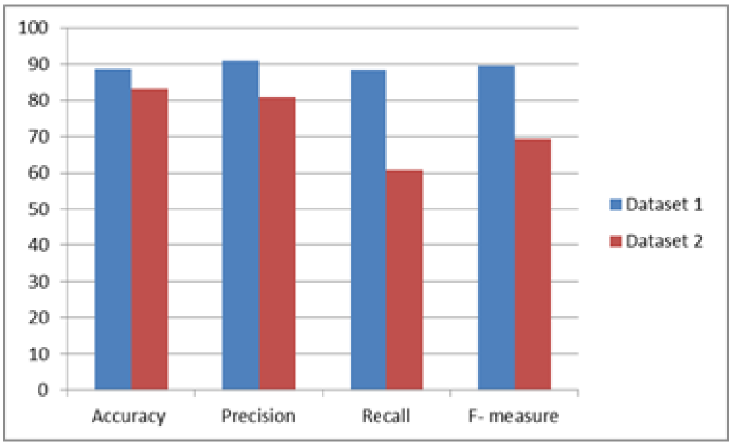

Figure 4.7: Performance Evaluation of RNN classifier with GA algorithm as feature selection

Hence, figure 4.7 demonstrates the existing results achieved with the application of the GA algorithm method and it illustrates implementation of GA for feature selection which is based on LSTM classification and it demonstrates enhanced performance for dataset 1 in comparison with dataset 2.

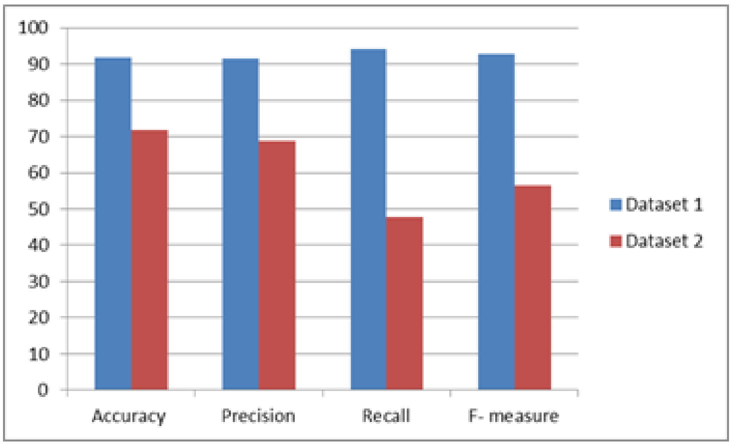

Figure 4.8: Performance Evaluation of the RNN classifier with PSO algorithm with feature selection

Hence, figure 4.8 demonstrates the current results acquired with the application of the PSO algorithm method and is displays that the performing PSO for feature selection based on RNN classification and it shows a greater performance for dataset 1 in comparison with dataset 2.

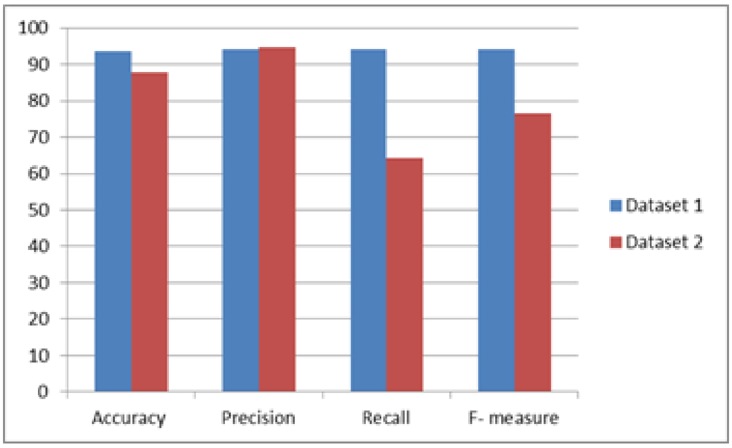

Figure 4.9: Performance Evaluation of the LSTM classifier with PSO-algorithm as feature selection

As in consequence, figure 4.9 demonstrates the results achieved with the application of the PSO algorithm method and it displays that implementing PSO for feature selection based on LSTM classification demonstrates an enhanced performance for dataset 1 from dataset 2.

Table 4.7: Features accuracy measures of features for dataset 1

Methods	Selected feature	Feature count	Accuracy	Time
GA-model	1111111111010	11	0.9016390	47.312710
PSO-model	1110101011001	8	0.9180330	92.512820
ABO-model	1001100101011	7	0.9836065	3.1326892
GSA	100111011011	8	0.9921	3.012548

Table 4.8: Feature count, selected feature, accuracy and time taken for dataset 2

Methods	Selected feature	Feature count	Accuracy	Time
GA-model	1111110011010	9	83.33	40.2598
PSO-model	1110101011111	10	80.0	59.3178
ABO-model	1111100001011	8	85.0	4.89513
GSA	10111010101	7	93.56	4.15691

Moreover, performance of recommended models for cleveland dataset and HF dataset are evaluated and attained results are exposed in table 4.7 and table 4.8. Since, the table 4.7 demonstrates current features selected against the GA, PSO, ABO and GSA. Hence, for the GA application, the sum of the features is 11 and achieves an accurateness rate of 90.16 % which is less when compared with PSO. So, PSO algorithm selective features number remains with 8 counts, and it represents the accurateness value as 91 %.

This can illustrate additional analytical time when compared with GA algorithm. Consequently, the correctness rate as of PSO algorithm displays the pre-eminent performance as an alternative to the GA-algorithm. But, ABO-algorithm selected features count of 7 and achieves an accurateness of 98% whereas GSA has exposed concentrated accurateness as 99.21%. By considering 8 counts of features selection of Cleveland dataset is find to be the best accuracy amongst the further three algorithms and involves less calculating time when compared with ABO, PSO and GA process. Correspondingly, table 4.8 analyses the performance of suggested models for HF dataset. Hence, GA model has illustrated 83.33% accurateness by taking 40.2598 seconds, PSO-model has determined 80% accurateness by 59.3178 seconds, and ABO model has exposed 85% accurateness with time taken of 4.89513 seconds, whereas GSA has realized 93.56% accuracy of 4.15691 seconds and as 7 set of features for HF dataset.

In general, GSA has the capability of choosing optimal feature sets for improving the performance of classifier by reducing the inappropriate features by this means of avoiding the trapping of local optimal. Because of this benefit, it has exposed operational performance compared with other three algorithms. Consequently, by internal evaluation of the suggested algorithms, GSA has revealed great accuracy compared to further three methods which evaluates its efficacy. The above mentioned tables 4.7, 4.8 has represented the associations of the recommended structure with the current approaches for HD-UCI and HF clinical dataset correspondingly. From the combination of the hybrid LSTM- RNN method with the ABO-model has exposed maximum accurateness in the HD predictions for cleveland dataset.

4.6 Comparative results of dataset

The recommended system is evaluated by predictable techniques connected to precision, accuracy, recall and also f-measure and the acquired results are deliberated in this segment. Primarily, comparison is achieved with cleveland dataset (dataset 1) and HF dataset (dataset 2). Further, results achieved for performance valuation for dataset 1 is revealed in figure 4.10, and results accomplished for performance valuation for dataset 2 is exposed in figure 4.11 and consequences achieved for performance estimation by comparing the dataset 1 and dataset 2 is specified in figure 4.12.

Since figure 4.10, DPA-RNN+LSTM has exposed accurateness of 94.7%, DPARNN+LSTM+GA has identified 96% as accurateness rate, DPA-RNN+LSTM+PSO has perceived an accuracy of 97.7%, DPA-RNN+LSTM+ABO has exposed an accuracy of 98.2%, whereas, DPA-RNN+LSTM+GSA has exposed 99.21% accurateness. Correspondingly, accuracy, f-measure and recall rate of the entire algorithms are established to be above 90%. Hence, an accuracy is the important measure in medicinal domain, in accordance with DPA-RNN+LSTM+genetic sine has exposed determined accuracy rate of 99.21% in contrast to further methods for dataset 1 which displays its efficacy.

Correspondingly, figure 4.11 has established that DPA-RNN+LSTM has exposed accurateness of 79%, DPARNN+LSTM+GA has derived accuracy rate as 86%, DPA-RNN+LSTM+PSO has realized the accurateness of 83%, DPA-RNN+LSTM+ABO has obtained 90% accuracy,

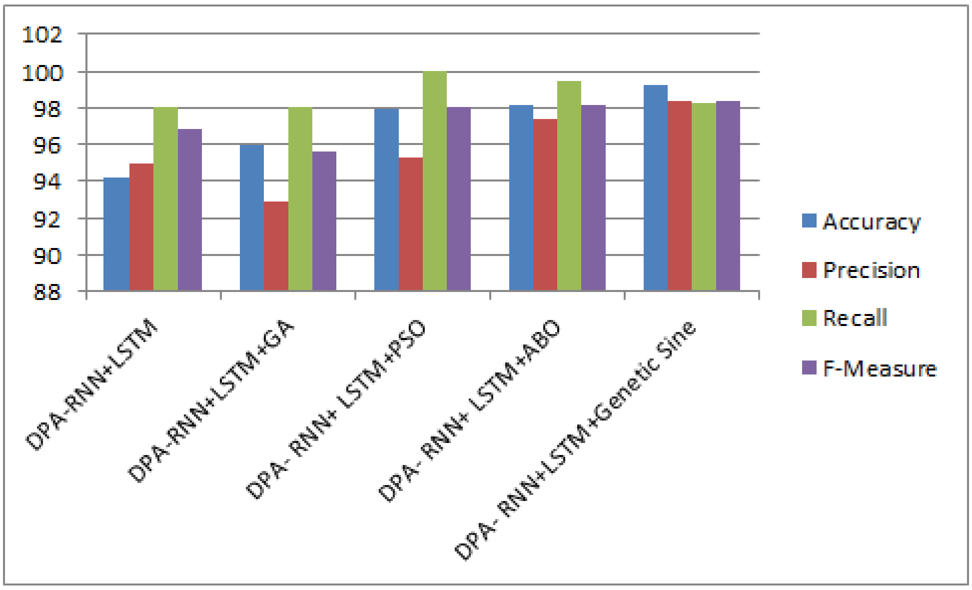

Figure 4.10: Performance valuation for dataset 1

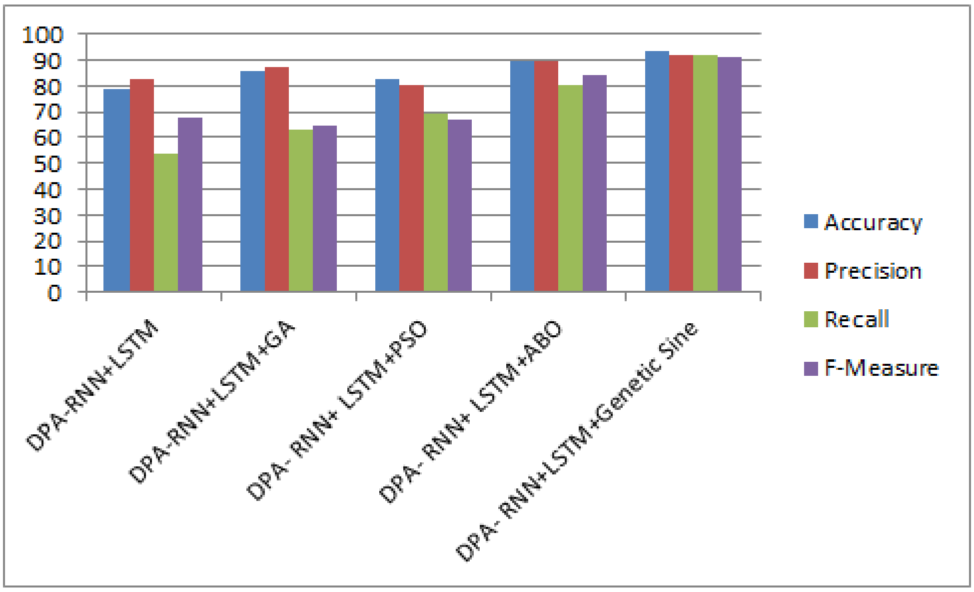

Figure 4.11: Performance valuation for dataset 2

whereas, DPA-RNN+LSTM+genetic sine has presented 93.56% accuracy. In the same way, precision, recall along with f-measure rates of entire algorithms are initiated to be improved. Then an accuracy is the important metric in medicinal domain. In accordance with DPA-RNN+LSTM+genetic sine has acquired maximum accurateness rate as 93.56% compared to other methods for dataset 2 which illustrates its efficiency.

Consequently, performance valuation for both the datasets are exposed in figure 4.12 has exposed the extreme accurateness of 99.21% which is achieved for cleveland dataset and

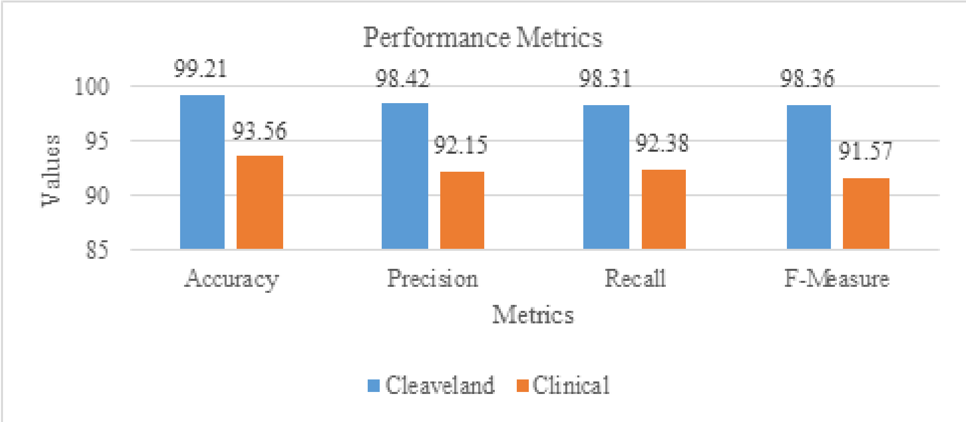

Figure 4.12: Performance valuation comparison for dataset 1 and dataset 2

93.56% for HF clinical dataset. Hence recall, precision, and also rate of F-measure of suggested system is established a better performance in both the datasets with maximum accuracy. Additionally, accuracy of the suggested and traditional methods are observed and the attained results have been shown in figure 4.13.

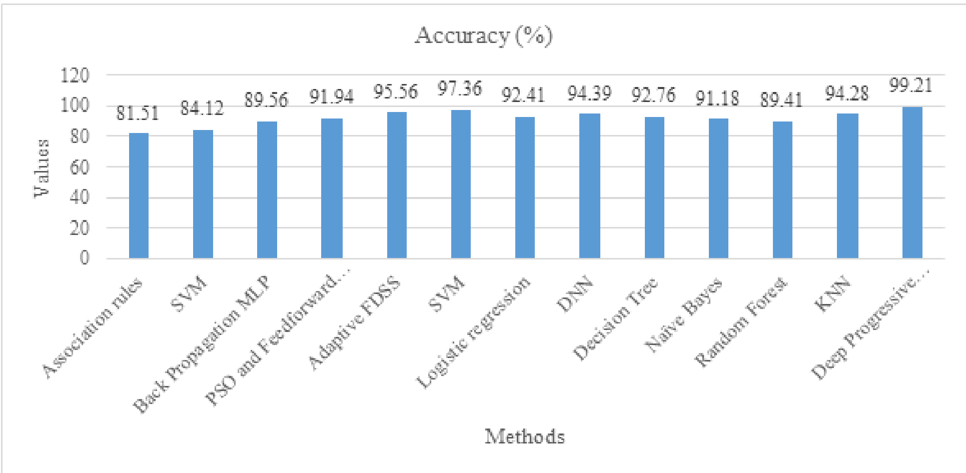

Figure 4.13: Performance valuation comparison for dataset 1 and dataset 2

Consequently figure 4.13 and figure 4.14 has exposed the suggested method implemented on cleveland dataset overthrows the prevailing methodology with regard to several performance measures such as precision, NPV, MCC, and f1-score values, and accuracy, sensitivity, specificity, and other metrics. In addition to that, the suggested system is relatively examined with prevailing methods like logistic regression (LR), Support vector machines (SVM), random forests (RF), decision trees (DT), naive bayes, and K nearest neighbor (KNN) are a few examples of machine learning techniques and attained results are presented in table 4.9.

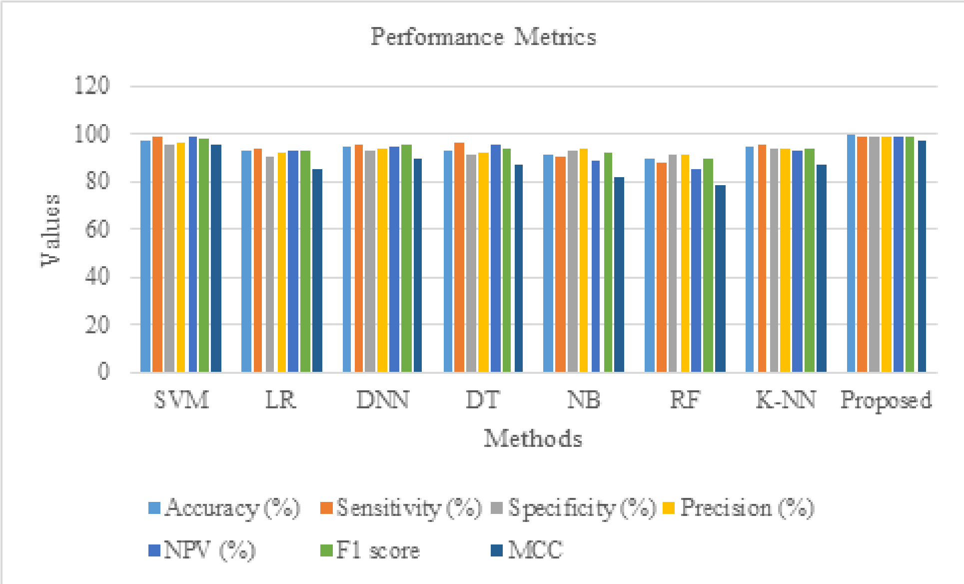

Figure 4.14: Performance evaluation comparison for dataset 1 and dataset 2

As in consequence, table 4.9 has exhibits the prevailing method such as bi-LSTM has illustrated maximum accuracy rate of 90.45%. But, when compared to the traditional methods, the suggested method has exposed great accurateness of about 98.42% through testing and 97.36% in training. Whereas, the precision rate of prevailing technique such as Bi-LSTM has exhibited 89%, and recommended system has displayed 98.42%. Similarly, recall as well as f1 score of the suggested method is constructed to be greater than established approaches through evaluating recall rate as 98.78% and f1 score rate as 98.36%.

Subsequent to the results, it is clear that suggested system accomplishes finer than other prevailing approaches such as bi-LSTM with maximum accurateness. Furthermore, current balanced random forest (BRF) technique is related to suggested process with reference to accurateness, sensitivity, ROC-AUC, MCC, G-mean and specificity and the acquired. Figure 4.15 presents the results.

Correspondingly, figure 4.15 illustrates the prevailing BRF classifier with 72.93% accuracy, whereas recommended technique has acquired 93.56%. Hence, the performance of BRF established lesser value compared with suggested methods with regards to further deliberated measures as well. Consequently, it is clearly stated that, recommended method is better than the prevailing BRF. In addition to that, arithmetical test analysis is achieved and the consistent results are revealed in table 4.10.

An essential tool for interpreting machine learning experiment outcomes are statistical sig-

Table 4.9: Comparative analysis of conventional and proposed methods

Models	Training accuracy	Test accuracy	Precision	Recall	F1 score
Logistic regression	85.38	80.22	80	80	80
SVM	89.62	79.12	78	79	77
Naive Bayes	85.38	81.32	81	82	82
KNN	88.68	84.32	84	84	84
Decision trees	83.6	79	80	79	79
RF	84.5	81	81	81	81
Existing- Bi-LSTM	90.45	89.01	89	88	89
Proposed Method	99.21	97.36	98.42	98.78	98.36

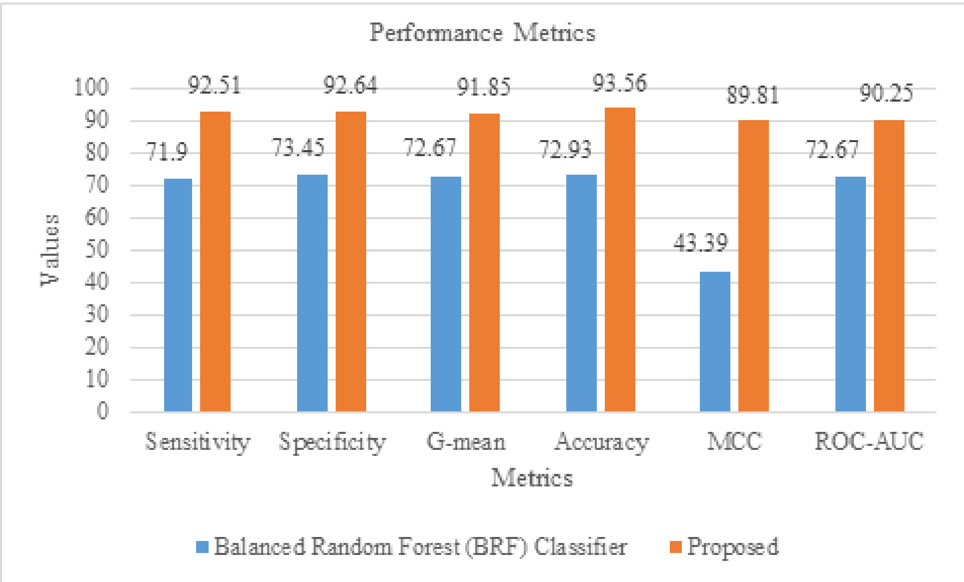

Figure 4.15: Comparative Analysis of prevailing and suggested method

nificance tests. The information gathered from these tools can assist in selecting the appropriate algorithms and setups for predictive modelling problem, as well as in presenting experimental results more effectively and confidently. As in consequence from table 4.10, the classifiers results use a two-step arithmetical test. Hence, based on the friedman rank, the iman-davenport is performed. Assume the variances in performance of classifier is recognised, friedman rank test is accompanied. Moreover, the test observes the grading of the advanced while classifiers iman-davenport recognises whether minimum classifier displays a significant modification contrary to others. Besides, the important variance is calculated by p value and it can be reduced with

Table 4.10: Arithmetical test analysis

Method	Friedman rank	Iman-Davenport p value
Proposed Algorithm	1.456	3.69E-10

Table 4.11: Mann-Whitney U-Test

Selected Feature	abs(PCC)	p-value
Age	0.000167	0.254
Anaemia	0.25297	0.066
Creatinine phosphokinase	0.68404	0.049
Diabetes	0.973913	0.002
Ejection fraction	0.000001	0.269
High blood pressure	0.171016	0.079
Platelets	0.425559	0.063
Serum creatinine	0	0.294
Serum sodium	0.000293	0.195
Sex	0.941292	0.004
Smoking	0.82819	0.013

0.05 threshold value. Whenever, the classifier rank is low, formerly it demonstrates better classifier. Furthermore, results of arithmetical analysis by using mann-whitney u-test are depicted in table 4.11.

From the above mentioned table 4.11, it is exposed that, the bio-statistical test (MannWhitney U test) which recognised the serum creatinine, ejection fraction as most important features and also the p-value is adjacent to 0 in uni-variate biostatistics analysis tests. Similarly, the pearson correlation co-efficient (PCC) results showed age in 3rd place as a maximum feature among the ejection fraction and serum creatinine. As a final point, recommended system is relatively examined with current approaches for cleveland and HF clinical dataset and acquired results are illustrated in table 4.12 and 4.13.

Consequently, from table 4.12, it is observed that the present method such as DNN has revealed 93.33%, whereas LSTM+PSO shows 93.5%, whereas, the suggested DPA-RNN+LSTM

Table 4.12: Recommended methods comparison with prevailing methods for HD diagnosis UCI dataset (dataset 1)

Methods	Accuracy
DNN + χ^2 Statistical model	(K-fold)91.57
DNN + χ^2 Statistical model	(holdout) 93.33
RNN+GA	90
RNN+ PSO	92
LSTM+GA	90
LSTM+ PSO	93.5
RNN+LSTM (hybrid)	95.08
RNN+LSTM +PSO	96.72
RNN+ LSTM + ABO (Proposed method)	98.36
DPA	96
RNN+LSTM (hybrid)+ DPA (Proposed method)	93.44
DPA-RNN +LSTM	94.7
DPA-RNN+LSTM+PSO	97.7
DPA-RNN+LSTM+ABO	98.2
DPA-RNN + LSTM with GSA	99.21

Table 4.13: comparison of current methods with recommended method for HF clinical dataset (dataset 2)

Methods	Accuracy
LR with all features	83.3
LR with 2 features	83.8
DPA-RNN+LSTM (Proposed method)	79
DPA-RNN+LSTM+ PSO (Proposed method)	83
DPA-RNN+LSTM + GA (Proposed method)	86
DPA-RNN+ LSTM + ABO (Proposed method)	90
DPA-RNN+ LSTM with GSA (Proposed method)	93.56

with GSA has acquired maximum accuracy of about 99.21% for cleveland dataset.

Since table 4.13, current method such as LR has obtained 83.8% of accuracy, whereas, recommended DPA-RNN+LSTM with GSA has resulted accuracy of 93.56% which is greater than prevailing methods. In general analytical results of suggested method found to be functioned finer than the prevailing techniques. Hence, the suggested technique excludes the prerequisite for fine modifications, highlights the critical data, individual/ optimization standards, over-all optimization and removes inacceptable information. Recommended system is operational because of its beneficial nature compared to current methods for HD prediction.

4.7 Summary

In this chapter, efficient HD prediction by using hybrid DL classification models are discussed. The algorithms like RNN, LSTM, ABO, and PSO are analysed individually as well as in a hybrid manner. Hence, the result analysis shows the occurrence of HD and the HF clinical datasets are compared and illustrated that the hybrid DL classification method can give better results than further models.

CHAPTER 5

OPTIMIZATION USING STACKED SPARSE AUTOENCODER MODEL FOR HEART DISEASE PREDICTION

5.1 Introduction

In recent years, ML methods are employed for the effectual HD prediction. Initially, prediction and diagnosis is essential for providing operational treatment to evade greater rates of death. Hence, some algorithms for classification were established to predict HD. Therefore, the application of research is based on early forecast of HD and also to progress the prediction precision by standard HD datasets namely HD clinical dataset and UCI cleveland dataset by applying effectual optimization and classification algorithms to display their utility in early detection of heart disease prediction. Later the study emphasizes on the initial prediction of HD and to enhance the prediction accuracy by using standard HD datasets like UCI cleveland dataset and HD clinical dataset by employing efficient optimization and classification procedures to illustrate the usage of HD prediction in premature period.

In general, the optimization algorithm demonstrates the advantages of producing the multifaceted non-linear problems with the finer adaptability and tractability. In addition to that, emperor penguin optimization procedure can choose the finest features towards the classification is used in this respect to progress the efficacy, reduction of resetting error, and increment in the HD classification excellence. Furthermore, the recently technologically advanced SSC-AE classification procedure is engaged for an important feature classification with greater advantage and effectiveness. As a result, this estimation compares the various ML techniques and their results in relation to a number of standards, including precision, f1 score, AUC score, sensitivity, and precision values. Moreover, the recommended technique SSC-AE demonstrates the greater result than the other classification methods where the discriminant features are extracted and its performance is compared with further advanced processes.

5.2 Proposed System

The study deals with the innovative methods which aims at classifying the significant features by inducing ML method with the appropriate optimization approaches which results in prediction of cardiovascular disease effectually. The suggested study involves pre-processing data, using the emperor penguin optimizer (EPO) for effective feature selection, and categorizing the key attributes using the SSC-AE technique. As a result, the suggested framework is shown in figure 5.1. Further, the presented approach is estimated with regard to several performance measures like recall, sensitivity, precision, and accuracy and specificity rates. In this section, the entire workflow is illustrated in figure 5.1 and it is denoted as the consecutive steps in the application procedure for the feature selection and classifying procedures.

5.2.1 Feature Selection at EPO Algorithm

As in consequence, in antarctic winter an emperor penguins(EP) is measured based on single classes which endure in cluster. Hence the emperor penguins huddling performance includes four stages,

- Generate and describe the emperor penguins huddles the margin.
- Measure the temperatures of the profiles which are around the cluster.
- Describe the space between the emperor penguins.
- Reorganize actual mover.

Hence, the behaviour of huddling is an important feature which reveals that each and every penguin that demonstrates the same chance to cluster warmth.

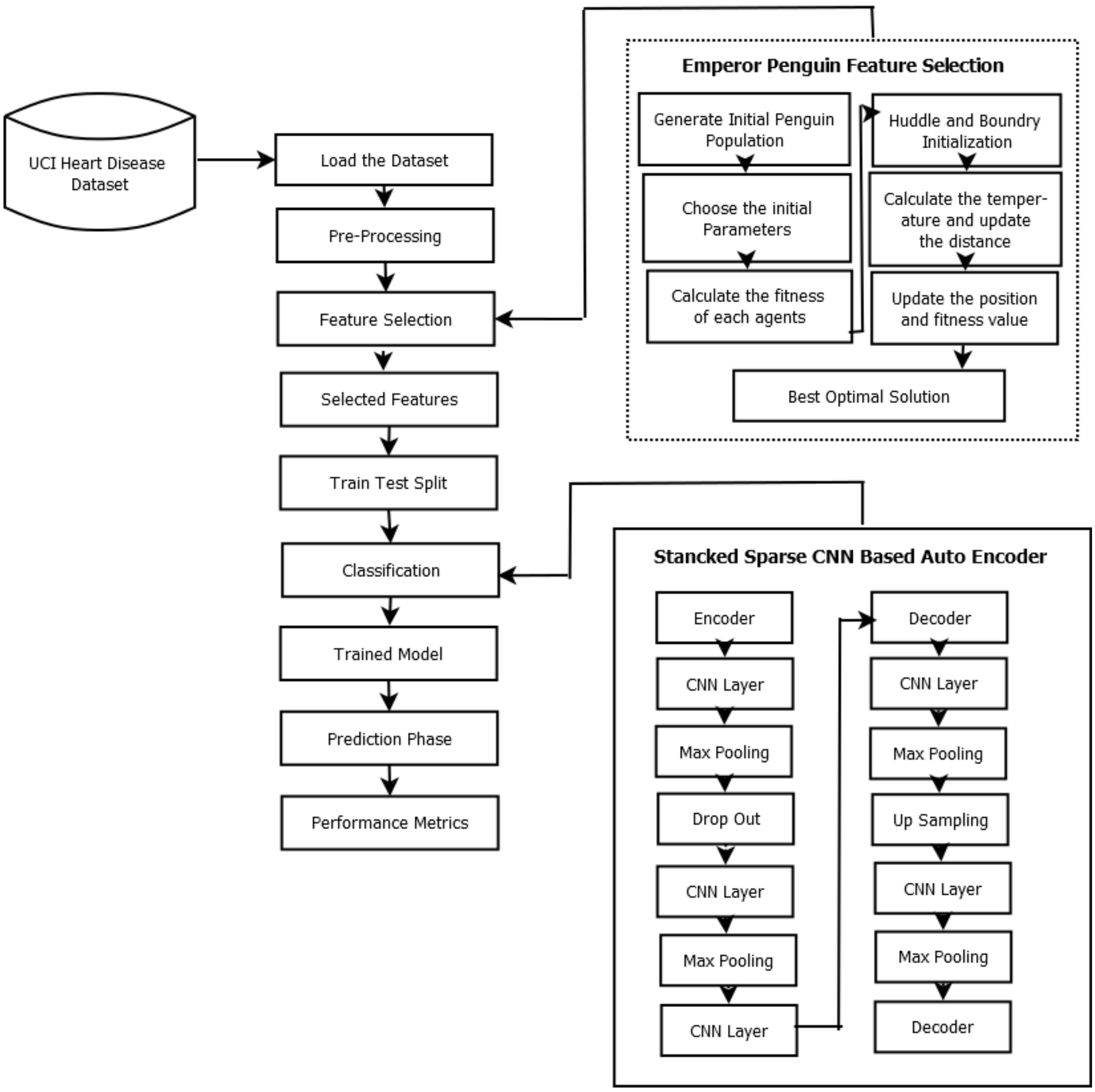

Figure 5.1: Work flow- SSC-AE by Penguin Feature Selection for HD prediction

5.3 Algorithm for Penguin Optimization - Feature Selection

As in consequence, the emperor penguins huddling performance is exhibited statistically. As a result, The primary goal of this study is to identify mover effectively. Consider, the huddle is located on 2-dimensional polygon plane accompanied with L-shape. At the initial stage, the EP makes an unsystematic huddle state line. Hence, the profile temperature is calculated about the huddle. Further, the space among the EP is restrained and assisted for further exploration and exploitation. By concluding the efficient mover is measured as the enhanced optimal solution that is obtained and reconsidered the huddle state line over the emperor penguins or else search agents changed locations. Hence each phases are demonstrated in subsequent divisions.

5.3.1 Creation and definition of the EP huddles boundary

EP typically positioned itself on polygon-shaped grid boundaries during the process of huddling. Emperor penguins are presented with minimum of two neighbours that are selected unsystematically from the group. Wind flow surrounding the huddle with the polygon is used to identify the huddle border. In contrast, the emperor penguin moves more slowly than the wind. Complex factors were used to explain the border of the randomly formed huddle of emperor penguins. Consider the velocity of wind as Φ and gradient is Ψ of Φ

$$\Psi = \nabla\Phi \tag{5.1}$$

By joining vector Ω and Φ complex potential is created

$$G = \Phi + i\Omega \tag{5.2}$$

According to the equation (5.2), G is the analytical function and i represents the imaginary constant on polygon plane. The figure 5.3 displays the equation (5.2) effect in 2-dimensional ambient. At the time of iteration procedure, EP position can be updated unsystematically based upon its position located at L-shaped polygon centre with superior effectual rate of fitness.

5.3.2 Compute the temperature of the profile surrounding the huddle

Huddle created using EP to reduce energy usage also to increase surrounding temperature is depicted in figure 5.2. When the radius of polygon is $R > 1$ and the temperature be $T = 0$, also when radius is $R < 1$ while the temperature $T = 1$, this condition is mathematically described. The temperatures of the profile have been accountable for the emperor penguins in diverse areas for the exploitation and exploring operation. $T^{'}$ - the calculated temperature profile in the vicinity of the huddle is

$$T^{'} = (T - \frac{Max_{iteration}}{-Max_{iteration}})T = 0,\ if\ S > 1 \tag{5.3}$$

In order to identify the best optimum solution in the search space, equation (5.3) defines the present iteration as x, the maximum amount of iterations as the max iteration, time as T and radius as S.

5.3.3 Describe the space between EP

Following the creation of the huddle border, the space between the emperor penguin and the more advantageously attained optimum value is calculated. As a result, fitness value is closer to the optimal value is the current better ideal option. In order to give a better ideal value, the EP or position of the search agents will be adjusted, mathematically defined as:

$$\overrightarrow{C_{ep}} = Abs(R(\overrightarrow{B})\overrightarrow{P(x)} - \overrightarrow{D}.\overrightarrow{P_{ep}}(x)) \tag{5.4}$$

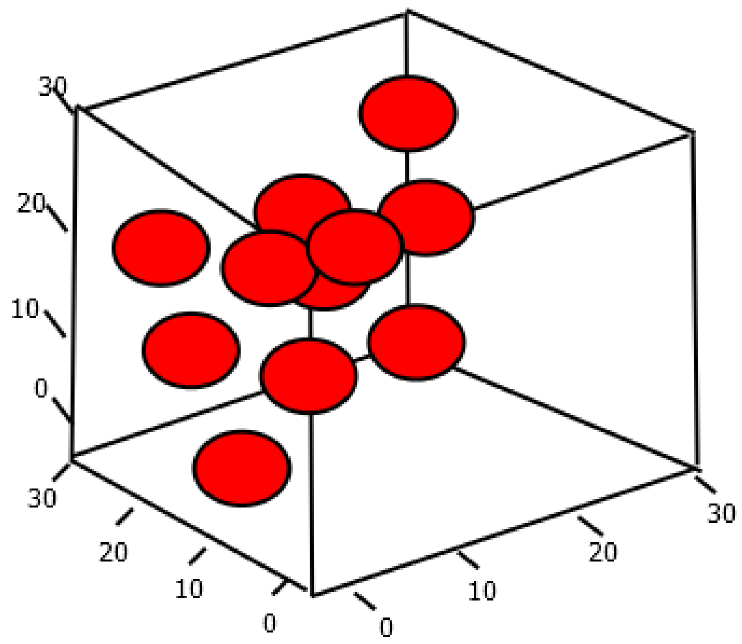

Figure 5.2: Huddles created for maintaining energy and transfer search agents to improved neighbours

$$\overrightarrow{B} = (N(P_{grid}(Accuracy))Rand()) - T^{'} \tag{5.5}$$

$$P_{grid(Accuracy)} = Abs(\overrightarrow{p} - \overrightarrow{P_{ep}}) \tag{5.6}$$

$$\overrightarrow{D} = Rand() \tag{5.7}$$

According to equation (5.5), N is regarded as a movement parameter for collision avoidance, which helps to keep the gap between search agents. N has the value of 2. $T^{'}$ is the temperature profile surrounding huddle, also the accuracy of polygon grid is represented using P grid (Accuracy), and random function is R and () that is displayed in the range [0,1].

Measurement for the function $R()$ is

$$R(\overrightarrow{B}) = (\sqrt{f.e^{-x/l} - e^{-x}})^2 \tag{5.8}$$

According to equation (5.8), e is the expression function. f and l , with values in [2, 3] and [1.5, 2], correspondingly, are the control parameters for exploitation and exploration. The presented algorithm, however, shows superior performance in selecting feature among these series.

5.3.4 Reposition actual mover

Based on the mover, EP positions are updated they are also termed as the attained optimal solution depicted in figure 5.4. The mover is in charge of altering the positions of search agent's within the search space that is provided and for removing the current position. The subsequent equation (5.9) is presented for updating the next position of the emperor penguin.

$$\overrightarrow{P_{ep}}(x+1) = \overrightarrow{P(x)} - \overrightarrow{B}.\overrightarrow{C_{ep}} \tag{5.9}$$

The presented algorithm saves the best solutions through the iteration process. The presented polygon grid method around the solutions, expanded to larger dimensions, defines an L-shaped polygon grid. The candidate solutions that perform erratically in avoiding collisions and search space and within the search space are supported by $\overrightarrow{D}$ and $\overrightarrow{B}$ values. The potential solutions permitted using the suggested distance method that finds the finest fit conceivable EP solution. Convergence behaviour of general optimization algorithms proposes that the pace of convergence increases owing to exploitation while it reduces due to exploration. The prospect of enhanced exploitation and exploration has been achieved by modifying $\overrightarrow{D}$ and $\overrightarrow{B}$ values.

As a result, the behaviour of the EP swarm in the search space influences the EPO.

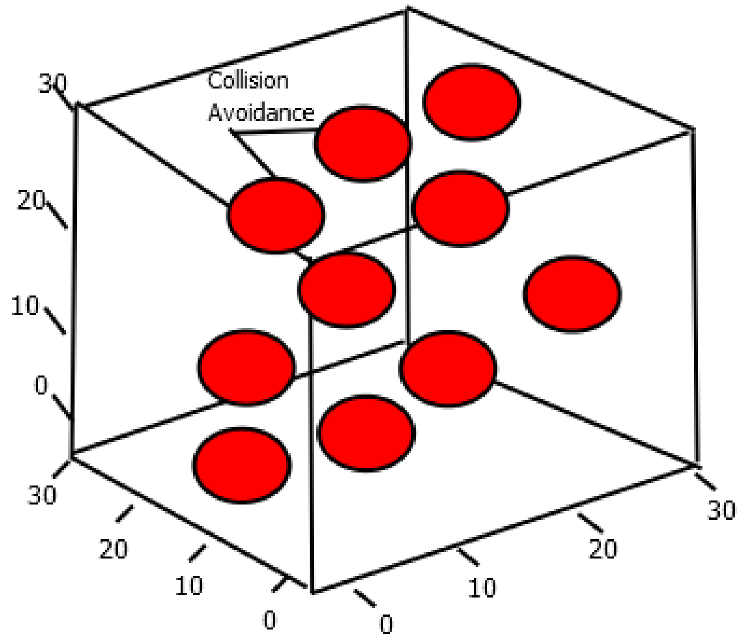

Figure 5.3: Preventing search agent collisions

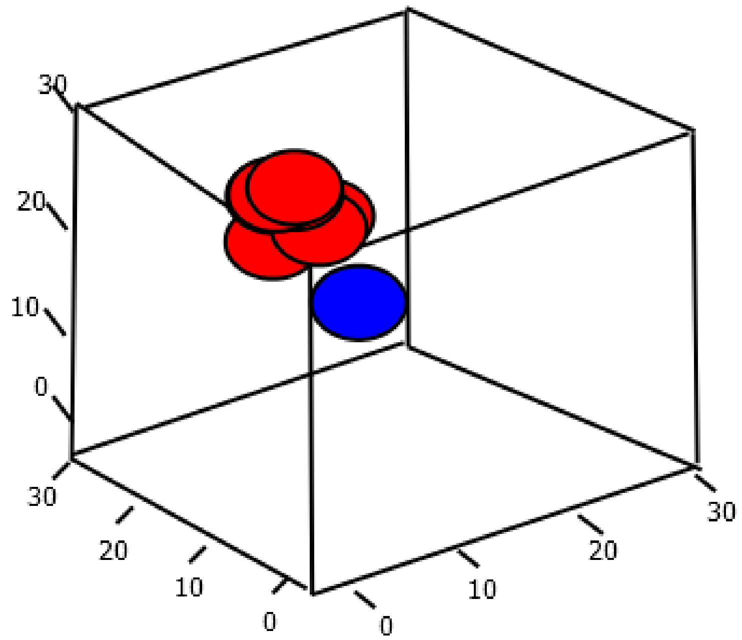

Figure 5.4: Updated positions of EP towards the better search agent

5.4 SSC-AE For Classification

A feedforward neural network with output, input and hidden layers is an auto-encoder. A stacked auto-encoder is used to stack many autoencoders. The SSC-AE approach incorporates noise into input data for the process of learning and training. The SSC-AE method's procedure

Table 5.1: Emperor Penguin Algorithm

step 1. Set the emperor penguin population as $\overrightarrow{P_{ep}}(x)$ where $x = 1, 2, ..., n$.
step 2. Select primary parameters and $T^{'}$, $\overrightarrow{B}, \overrightarrow{D}, R(), S$ and $Max_{iteration}$
step 3. Determine fitness value of each search agent.
step 4. Describe the EP huddle border by employing equations (5.1) and (5.2).
step 5. Using equation (5.3), estimate the $T^{'}$ temperature profile surrounding the huddle.
step 6. By equations (5.4 to 5.8), estimate the distance among the EP.
step 7. By equation (5.9), change the position of the search agents.
step 8. Check to see if any search agents are moving outside of the allocated search space and it is then adjusted.
step 9. The adjusted value of search agent fitness is computed, and the optimal result is updated.
step 10. The algorithm is run till the approach is finished. Otherwise, it will return to the fifth step.
step 11. After the obtained halting process, return the better optimal solution.

is as follows:

- Using non-noise data as the input, unsupervised training reduces output as well as input error, also output value of the hidden layer have been attained.
- The output of the initial autoencoder's hidden layer is fed into the consecutive autoencoder, the output value of the hidden layer is attained, and then the unsupervised training reduces errors between output and input, step 2 should be repeated until every autoencoders are trained.
- The autoencoder weights are extracted to prepare the SSC-AE model weight.
- For unsupervised training, noisy data with low output error also the no-noise data are used. Finally, the SSC-AE model is attained.

figure 5.5 depicts the *SSC_AE* creation process, which includes the two layers of the auto encoder. For the first autoencoder,

$$h_i = f(W_{11}x_i + a_{11}) \tag{5.10}$$

According to equation (5.10), the input variable x_i, h_i while h_i is a hidden variable used in place of the subsequent auto-encoder input. And it is written as,

$$t_i = f(W_{21}h_i + a_{21}) \tag{5.11}$$

The output variable from equation (5.11) is t_i. W_{11} and i^{th} represents the weight matrix then the i^{th} auto encoder one-sided encoding procedure, whereas W_{12} and a_{12} represents weight matrix then the i^{th} auto encoder one-sided decoding procedure. Presented SSC-AE is obtained by incorporating the consecutive autoencoder output into the initial autoencoder decoding section.

$$t_i = f(W_{21}f(W_{11}x_i + a_{11}) + a_{21}) \tag{5.12}$$

$$x_i = s(W_{12}s(W_{22}t_i + a_{22}) + a_{12}) \tag{5.13}$$

Convolutional and pooling layers constitute CNN. Finally, the classifier is incorporated in general. Figure 5.6 depicts a conventional CNN. The convolution process is characterised as the three 2-dimensional matrices with 2 feature maps and 1 convolutional kernel (Y as Output, W as convolutional kernel and X as Input). The convolution kernel proliferates the appropriate weights of neuron on the alike size scale in place of the input graph. The outcome is allocated to the output matrix position. And convolution core is shifted from the left to the right, and the individual stages on the input graph are moved from top to bottom to complete the convolution

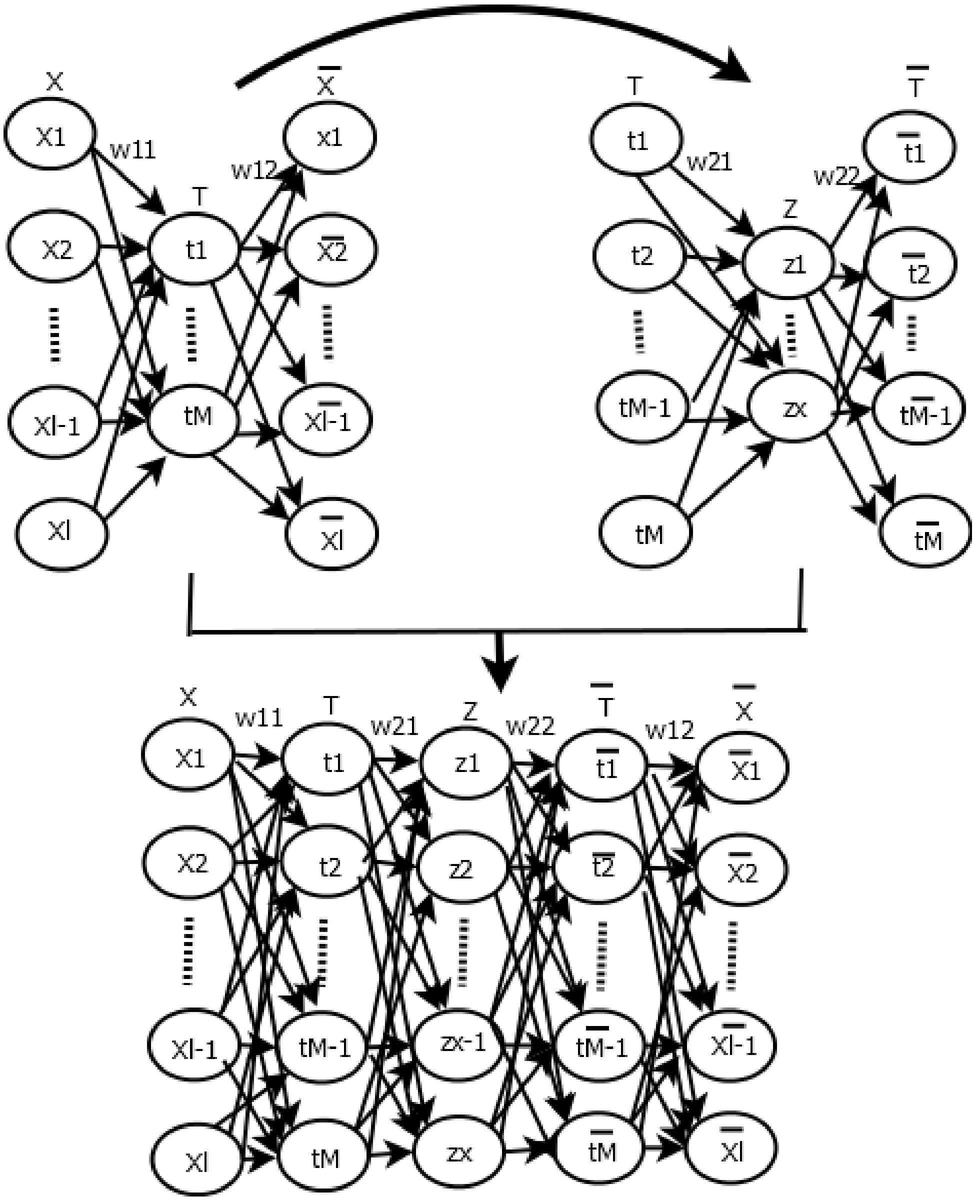

Figure 5.5: SSC-AE Architecture

procedure. Figure 5.7 depicts the convolution process. The softmax model is a logistic regression method extension that is employed for multi-classification issues. Softmax splits the x to j possibility categories as follows:

$$q(y^{(i)} = jx^{(i)}; \theta) = (\sum_{m=1}^{k} e^{\theta_m^T x^i})^{-1} e^{\theta_j^T x^i} \tag{5.14}$$

According to equation (5.14), values of $j = 1, 2, \ldots, k$. Parameter of softmax regression

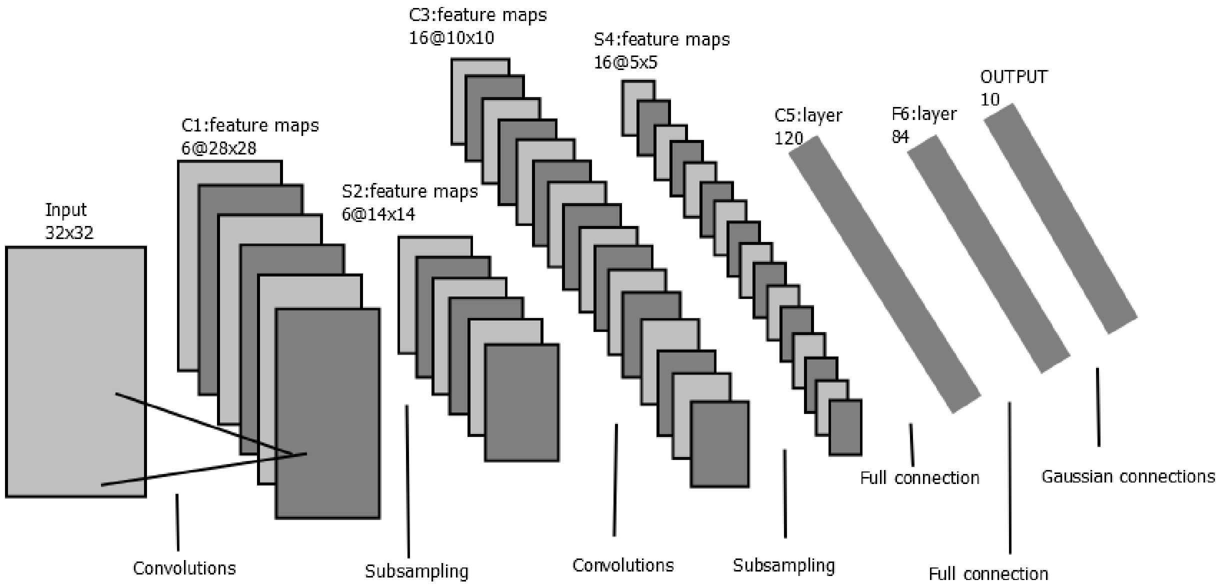

Figure 5.6: CNN Structure

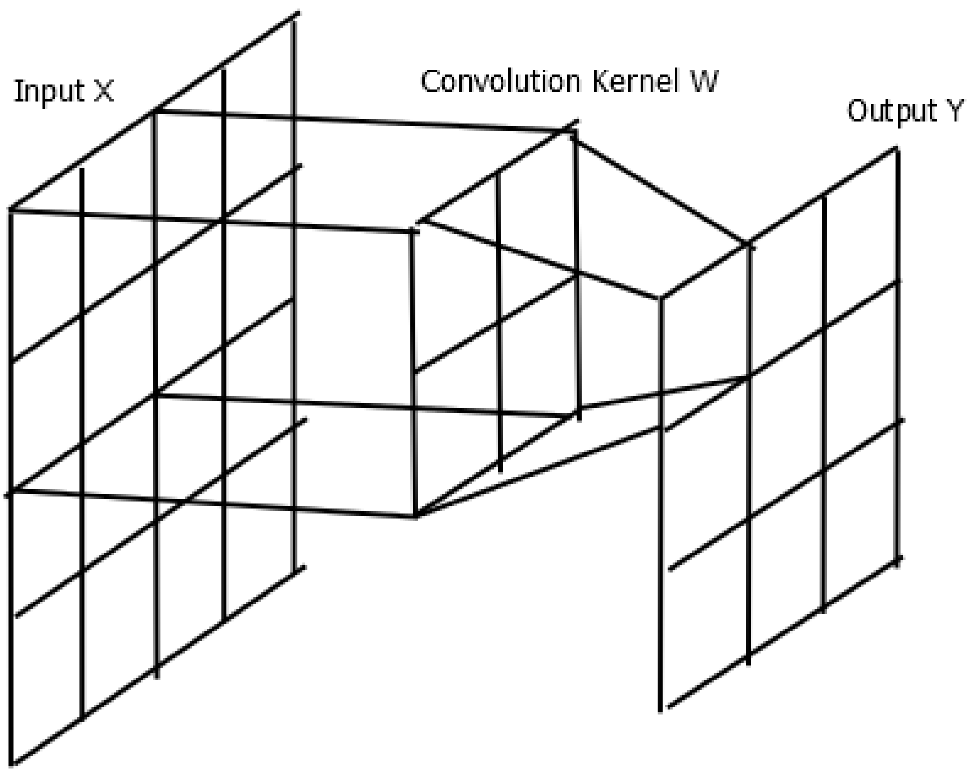

Figure 5.7: Process of Convolution

model is θ^L= $[\theta_1^L, \theta_2^L, \ldots\ldots., \theta_k^L]$. The loss function of softmax regression is,

$$J(\theta) = -\frac{1}{n}[\sum_{i=1}^{n}\sum_{j=1}^{k} I\{y^i = j\} \log \frac{e^{\theta_j^T x(i)}}{\sum_{m=1}^{k} e^{\theta_m^T x(i)}}] \tag{5.15}$$

According to equation (5.15), the j is represented as the indicator function. At times the judgment is right, it returns to 1; otherwise, it returns to 0.

5.5 Results and Discussion

The findings and suitable description of HD prediction utilizing SSC-AE based EPO technique is mentioned in the following section. Details of the dataset are provided for testing and training datasets. In addition, performance measures for deep learning classifiers are provided

5.5.1 Dataset Description

Data-sets of two-kinds were employed in the application stage such as HD analysis of heart failure clinical dataset and UCI dataset

5.5.1.1 Dataset 1: HD diagnosis UCI-Dataset

This dataset comprises of 76 attributes, also holds a whole of 14 subsets. Cleveland dataset is only database employed by the ML researchers till now. The attribute that is targeted stands for identifying HD from the patient data. This value is between 0 and 4, inclusive. The online data set for HD estimation in the presented framework was the cleveland data-set for the ailment. The dataset is accessible online through the UCI-source of ML methods. The following is a list of the data-set attributes.

The field stands for social-security amounts and patients name have been rejected from employed dataset. Rejected values is replaced using fake-values. And, above presented list of attributes list stands for the whole amount of feature fields employed in data-set.

Table 5.2: List of the Data-set 1 Attributes

1. Age
2. Gender
3. The slope of the peak
4. Old peak
5. Fasting blood sugar greater than 120 mg/dl
6. Resting of resting electro-cardiac
7. Major vessels count
8. Type of Chest Pain (4 values)
9. Maximum Heart-rate attained
10. Exercised induced angina
11. Serum cholesterol in mg/dl
12. 7-reversable-defect ,fixed-defect-6, 3 $\rightarrow$ ordinary one

5.5.1.2 Dataset 2: Heart failure clinical dataset

This dataset type is conquered for selecting feature method, enactment evaluation and process of classification is examined in the 2 data-sets that are resulted from expected results. The author of this dataset is giuseppe and davide Chicci, uses ML algorithms that predicts the survival rate of patients having heart-failure problems. Prediction of HD is made with the data from serum-creatinine and ejection-fraction only. The second data-set is the heart failure clinical records dataset. The following attributes are included in this dataset. The data-set's attributes are as follows.

5.5.2 Performance Measures

Prediction model results is evaluated using the performance features where the enactment metrics comprises recall factor, accuracy-factor, f-measure value and precision factor. The metrics is detailed as follows.

Table 5.3: List of the Data-set 2 Attributes

1. Age
2. Gender
3. Anaemia
4. Ejection portion
5. Diabetes
6. Creatinine phosphate
7. Serum-creatinine
8. Platelets
9. Serum-sodium
10. High Blood pressure

5.5.2.1 Accuracy

The accuracy levels are given by the effective categorized data from the testing data-sets revealed values in percentage. Accuracy factor and the formula for achieving accuracy is as follows:

$$Accuracy\ for\ prediction = \frac{TP+TN}{TP+TN+FP+FN} \tag{5.16}$$

5.5.2.2 Precision

Precision values indicate the categorized data of the healthy people that are appropriately assessed. The precision values are evaluated using the formula given below.

$$Precision\ for\ prediction = \frac{TN}{FP+TN} \tag{5.17}$$

5.5.2.3 Recall

The Recall value measure defines the recovered linked proportional instances. As a result, both the accuracy factors and recall have been based on the measurement of the results and relevance of the results. The values are assessed using the formula given in the notation below.

$$Recall\ for\ prediction = \frac{TP}{TP+FN} \tag{5.18}$$

5.5.2.4 F-measure

The recall mean and accuracy values of the results are what make up the f-measure. The values are computed as follows:

$$F1\ Score\ for\ prediction = \frac{2 * Precision * Reall}{Precision + Recall} \tag{5.19}$$

5.6 Performance Analysis

The below session displays the outcome of presented SSC-AE based EPO scheme to predict HD according to 2 datasets i.e. heart failure clinical dataset and HD diagnosis UCI-dataset. Outcome of the presented scheme is related with alike prevailing studies in order to discover the HD prediction effectiveness.

5.6.1 Comparative study- HD diagnosis UCI-Dataset

According to the table 5.4 and figures 5.8 - 5.11, as it is obvious that presented SSC-AE based EPO displayed superior results with regard to accuracy, sensitivity, f-measure and precision as 98.02%, 91%, 98% and 98% correspondingly compared with prevailing studies by UCI

Table 5.4: Comparative study of SSC-AE (proposed) and Prevailing studies for UCI celveland dataset

Algorithm	Accuracy	Precision	Sensitivity	F-measure
KNN	60	61	59	58
LR	78	79	78	78
LDA	78	80	79	79
SVM	79	80	79	79
CART	68	69	68	68
GB	81	79	84	81
RF	83	81	87	84
Proposed Method	98.02	98	91	98

cleveland dataset.

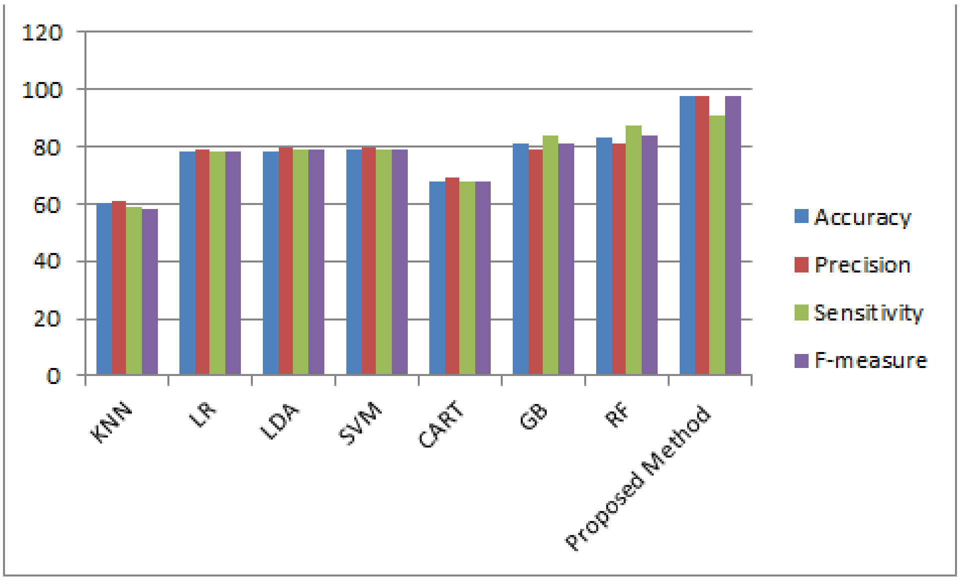

Figure 5.8: Accuracy analysis of the conventional researches and proposed SSC-AE

According to the table 5.5, it is obvious that presented SSC-AE based EPO displayed superior results with regard to AUC score and accuracy as 96.04% and 98.02% correspondingly compared with prevailing studies by UCI cleveland dataset.

According to table 5.6, it is obvious that presented SSC-AE based EPO displayed superior results with regard to accuracy as 98.02% when related with the prevailing studies namely PSO fuzzy expert, RF, CFS-PSO, SMO expert-based, two-tier ensemble learning, boosted C5.0,

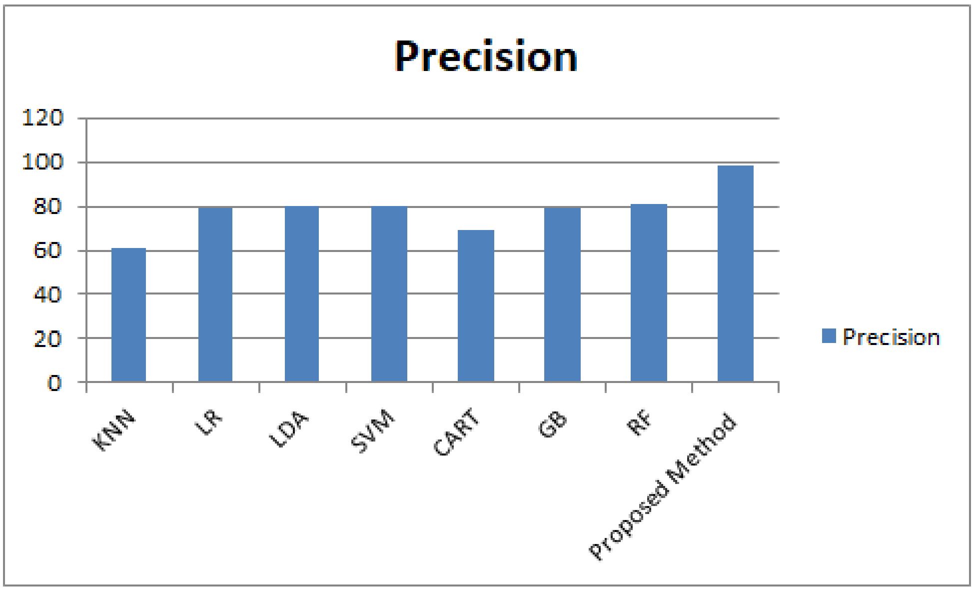

Figure 5.9: Precision study of the SSC-AE (proposed) conventional researches

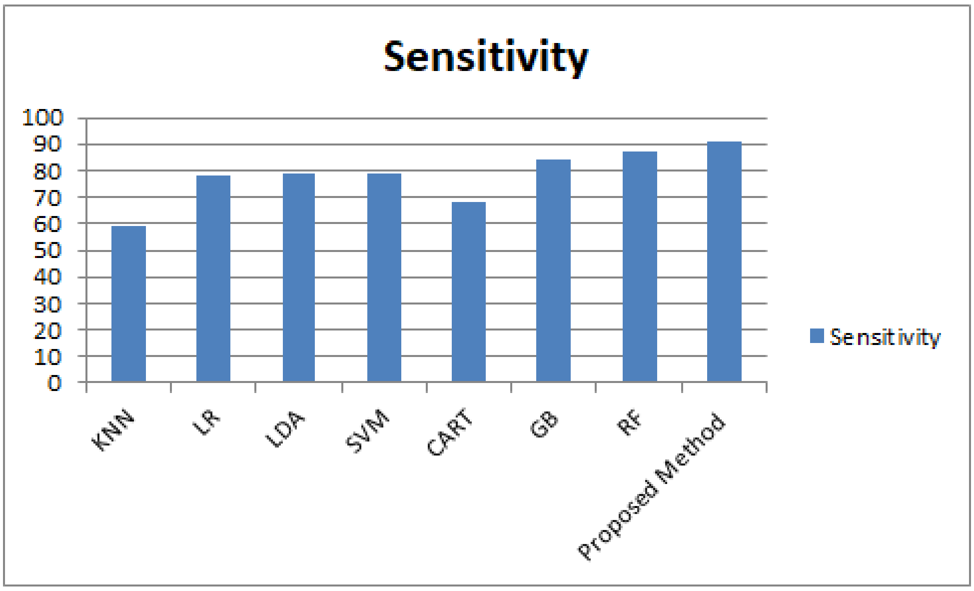

Figure 5.10: Sensitivity analysis of the conventional researches and proposed SSC-AE

Table 5.5: Study of SSC-AE and conventional researches

Algorithms	AUC	Accuracy
Logistic Regression	0.9161585	0.8651685
Stochastic gradient boosting	0.9070122	0.8426966
Support vector machine	0.882622	0.7977528
Proposed Method	0.9604	0.9802

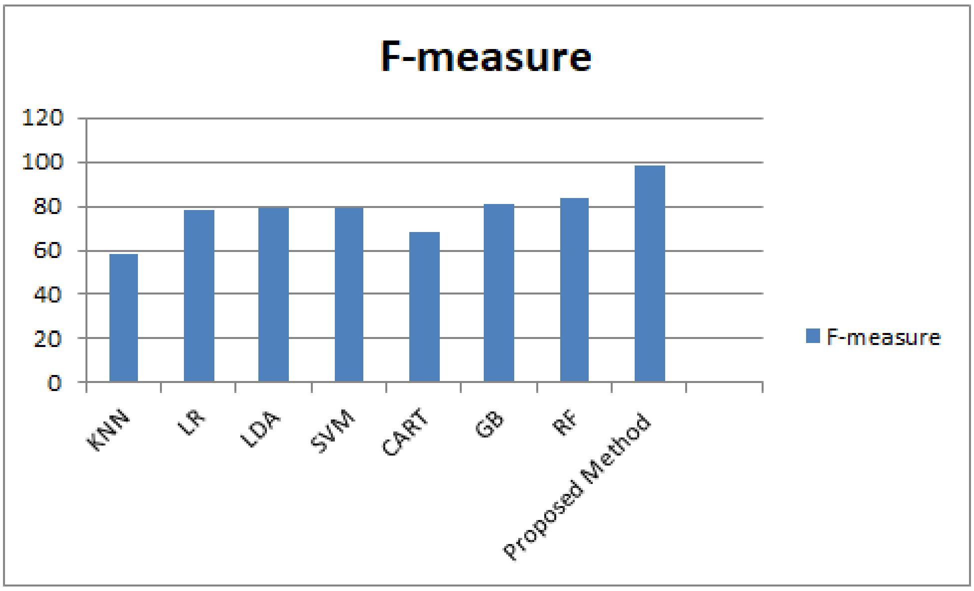

Figure 5.11: F-measure study of the SSC-AE (proposed) conventional researches

Table 5.6: Comparison of proposed SSC-AE with Existing studies

Technique	Features	Acc	Fmeasure	AUC
Rotation forest-J48-CFS	7	84.48	NA	89.5
PSO fuzzy expert systems	76	93.27	NA	NA
SMO-expert-based feature selection	8	84.49	86.2	NA
CFS-PSO-clustering-MLP	5	90.28	NA	NA
Logistic regression-LASSO	6	89	NA	NA
Boosted-C5.0	12	77.8	NA	NA
Neural network	12	81.9	NA	NA
Voting-naive Bayeslogistic regression	9	87.41	NA	NA
Two-tier ensemble PSO based feature selection	7	85.71	86.49	85.86
SSC-AE	12	98.02	98	96.04

neural network algorithms, CNN centred algorithms by UCI cleveland dataset.

Figure 5.12 and figure 5.13 displays that for UCI dataset, also the enactment evaluation outcomes are superior for the SSC-AE presented process when compared to the prevailing algorithms. Performance metrics taken into account are accuracy, f1-score, recall, ROC curve and precision. For the obtained metrics, the presented methods surpasses the prevailing performance values. When it comes to accuracy values, decision tree attained 73.62%, naïve bayes attained

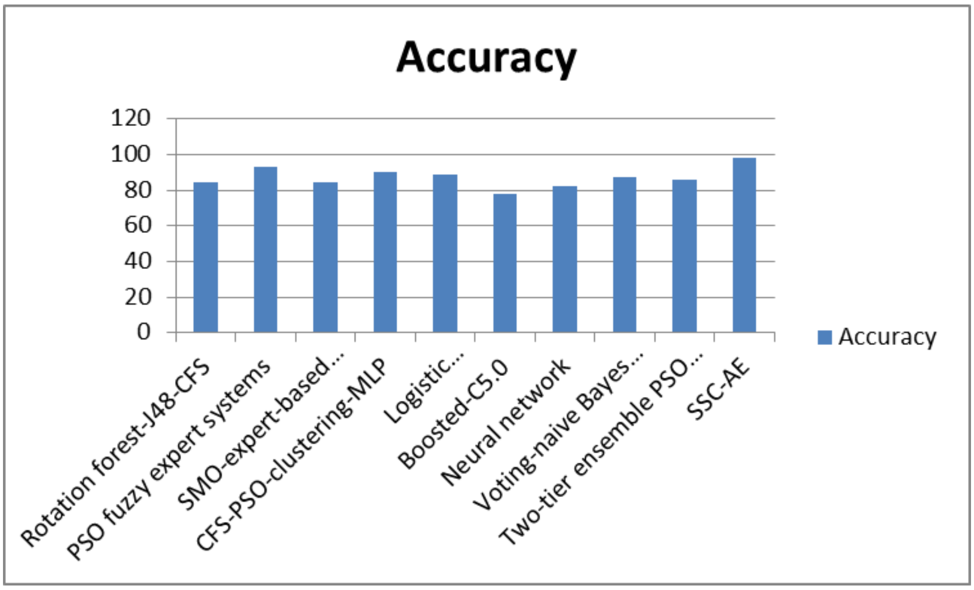

Figure 5.12: comparative study of SSC-AE (proposed) and prevailing studies

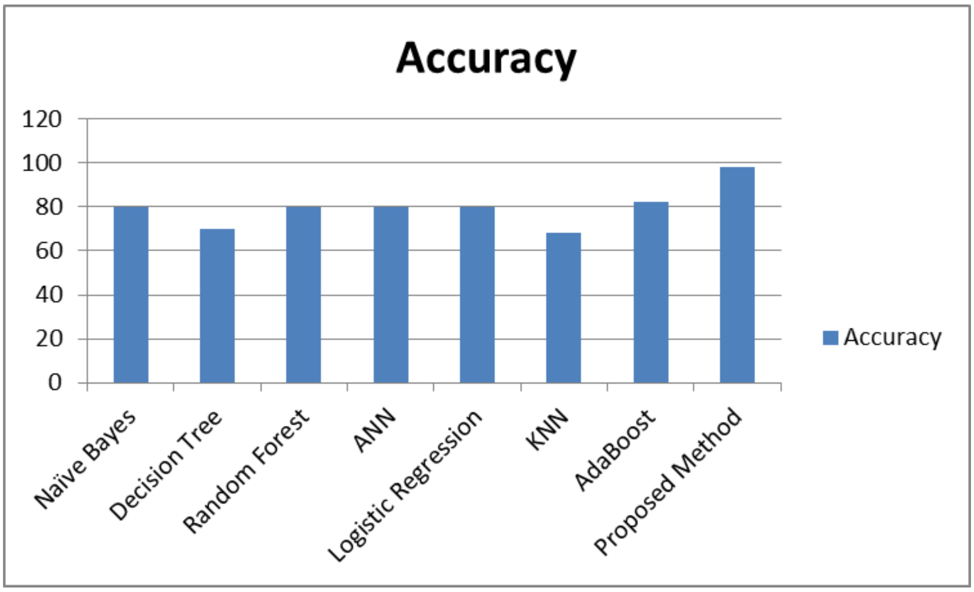

Figure 5.13: Comparison of Accuracy

81.67%, K-NN attained 76%, adaboost attained 83%, logistic regression attained 81.3%, and presented SSC-AE technique attained 98%. Therefore, SSC-AE is regarded as best suitable procedure in predicting HD by UCI cleveland dataset.

In the figure 5.14, accuracy values presented are related with prevailing LSTM, GRU, SVM and KNN techniques. The presented method attains an accuracy value of 98.02 which suppressed the accuracy prediction of all present model. Therefore the presented SSC-AE process is extremely advantageous in supporting clinical practitioners to study the condition of the

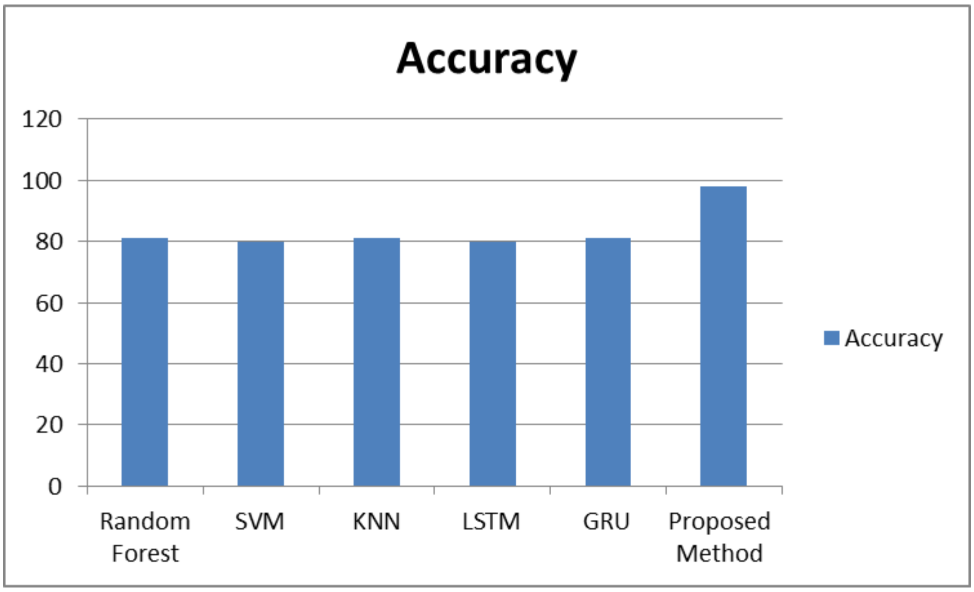

Figure 5.14: Comparison of Accuracy

patients.

5.6.2 Comparative analysis- Heart failure clinical dataset

Table 5.7: Comparative study of SSC-AE(proposed) and Prevailing studies

Method	AUC	Accuracy
LR	59.72	78.03
RF	84	87.12
Naive bayes classifier	82	82
Ensemble model with Boosted C5.0 tree and SVM	NA	84.2
Proposed method	91.12	92

Based on the empirical clinical concept, the conventional predictive heart failure assessment is implemented as in table 5.7.

The brutality and therapeutic impacts of patients are assessed from expert clinicians by concentrating on objective attestation from diverse test results. Also, for numerous patients group, the researchers have established precise models to predict the risk of heart failure fatality. The current model compared with presented SSC-AE method for evaluating the prediction of patients with heart ailment is illustrated in figure 5.15 and table 5.8 via comparing with prevailing RF, SVM Radial. The presented SSC-AE resulted in improving outcome based on AUC and accuracy as 91.2% and 92% respectively.

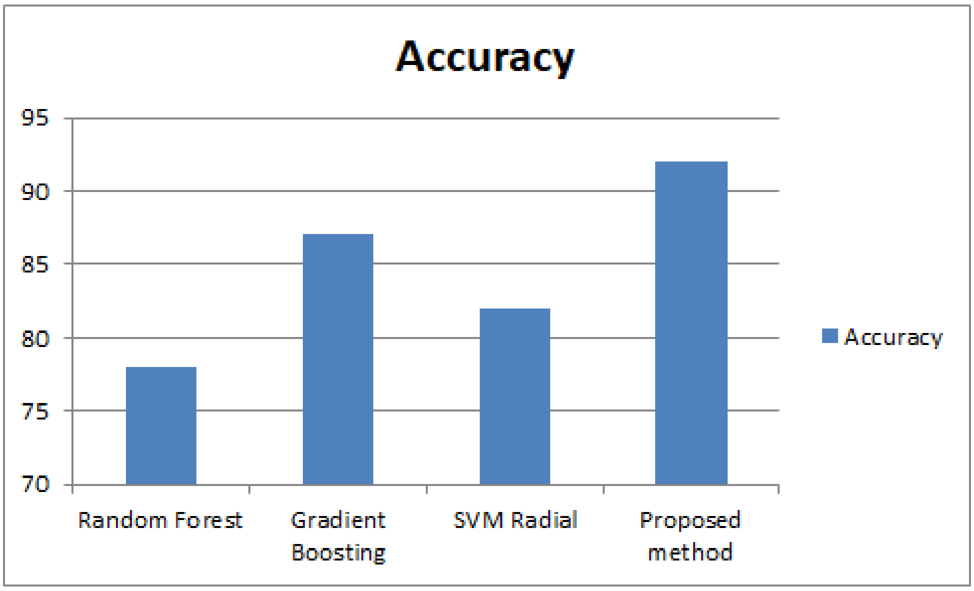

Figure 5.15: Comparison of Accuracy

Table 5.8: Comparative study of SSC-AE (proposed) and Prevailing studies

Method	F-measure	Accuracy
Random forests	0.754	0.585
Gradient boosting	0.75	0.585
SVM radial	0.72	0.543
Proposed method	0.91	0.92

The figure 5.16, displays that the 2 datasets are related in term of f1 score, AUC score and accuracy. The presented SSC-AE based EPO optimizer provides superior enactment from UCI cleveland dataset compared to HD clinical dataset. Especially for cleveland dataset, several optimization methods are employed and feature selection technique features in definite time is displayed in the table 5.9 below.

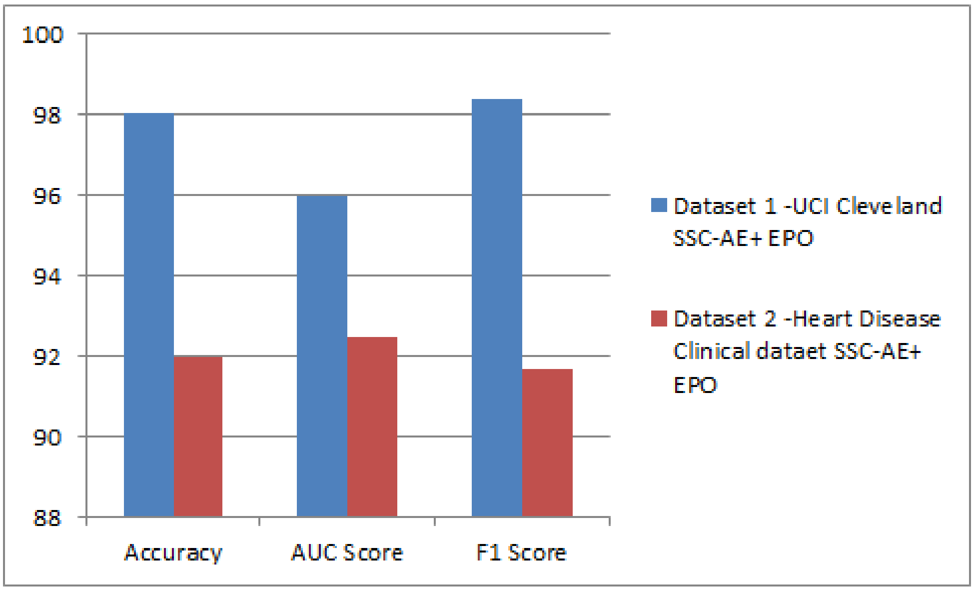

Figure 5.16: Comparison of Accuracy

Table 5.9: Comparative study of optimization methods

Method	Selected feature	Feature count	Accuracy	Time
GA-model	1111111111010	11	0.9016390	47.312710
PSO-model	1110101011001	8	0.9180330	92.512820
ABO-model	1001100101011	7	0.9836065	3.1326892

5.7 Conclusion

The aim of this work is to do effective HD prediction categorization and compare it to diverse ML algorithms with diverse performance indicators. When compared to naive bayes, random forest, and other current algorithms, the presented SSC-AE model outperforms them. The enactment evaluation outcome displays that EPO strategy improves the predicting accuracy of HD datasets. As a result of the performance analysis, it was obvious that the SSC-AE with EPO outperformed UCI cleveland HD dataset and the HD clinical dataset, with accuracy values of 98.02% and 92%, respectively. Furthermore, when the two datasets are compared, the outcome demonstrates that the UCI cleveland dataset produced superior outcomes.

5.8 Summary

In this chapter, efficient HD prediction by using DL based models are discussed. SSC-AE classification algorithm and penguin optimizer algorithm is utilized for effectual classification of feature with greater robustness. Results display that the presented model SSC-AE outperforms when compared with other classification methods which extract discriminant features that has better enactment when compared to state-of-art algorithms.

CHAPTER 6

IOT BASED HEART DISEASE PROGNOSIS MULTI CLASS CLASSIFICATION USING DEEP LEARNING

6.1 Introduction

IoT has undoubtedly changed health platforms by introducing wearable forms of fitness bands, wireless forms of connected devices, BP and cholesterol monitoring cuffs, glucometer etc. This adds up to personalized attention to the patient health care and diagnostic monitoring. These IoT devices enable a constant form of tracking on health conditions. Concurrently, these IoT devices are connected to sensors which are used in tracking the current location of the clinical equipments such as wheelchairs, nebulizers and many other patient care devices. Some of the significant benefits of IoT in healthcare sector includes:

Reduction of Cost- Precise time patient monitoring is done which significantly prohibits several visit to doctors and cases of re-admissions.

Enhanced forms of treatment- enables doctors to make decisions based on evidence and to have absolute forms of transparency over treatments.

Reliable disease diagnosis- ongoing observation and procured real-time data aids in early stage disease detection and prevents complex forms of endurances.

Error Reduction- Smoother and fewer errors in healthcare operations and accurate forms of diagnostic results.

IoT enables physicians to monitor real-time conditions of patients in preventing emergency situations. These are done by collecting and analysing the data from medical devices such as wearables where the healthcare provider can gain increased insights in patient health and can tailor the treatment plans accordingly. IoT in healthcare takes up a broad range of use cases which includes automated forms of diagnosis and management operations. IoT offers a smooth platform to enable various forms of interactions. In this regard, There are three types of IoT in healthcare:

- tracing people and objects
- people authentication and identifications
- Automatic forms of sensing data and collections.

Thus, IoT in healthcare domain completely redefines the healthcare monitoring by means of divides applications and emergency management. IoMT and HIoT which are defined to be Internet of Medical Things and Healthcare. Internet of Things are applied to be the alternatives in integrating the medical applications and devices are able to connect information technology in health care systems and to the IoT-based environment. To this end, IoT provides personalized modes of healthcare enabling remote patient monitoring, diagnosis, diseases prevention and early detection.

IoT is defined to be the computing operation, in which every IoT physical object is equipped with microcontrollers (MC), transreceivers and sensors, which empowers the actions of communication. These devices are built with suitable protocol stacks which makes these devices interact with one other making a viable form of communication between the users. Diverse distributed devices used in IoT-based healthcare diagnostics can analyse, aggregate, and interact real time clinical data with the cloud, which enables the data to be collected, stored and to analyse even large quantity of data.

Moreover, IoT ensures enhanced forms of personalised healthcare services individually for each of the patients. IoT based healthcare monitoring systems are non-invasive, pervasive and is useful in keeping track of the patient data effectively. In these systems, several IoT based devices are gathered, which can analyse, pass real time medical information and are able to activate context dependent alarms.

In view of these advantages in adapting IoT in medicare systems and in healthcare diagnostics, the proposed method used IoT system only for data collection and has made use of the AD8232 sensor framework and arduino uno in aspects of gathering the real-time data by receiving heat rate and ECG rates. This makes the model to be less robust and results in bringing less accurate prediction outcomes. In view of resolving these issues, the proposed study also makes use of several DL algorithms as a primary data- min ing approach, especially for the phase for feature extraction, using algorithms namely GA, particle swarm optimization (PSO), genetic sine algorithm (GSA), african buffalo optimization (ABO) and penguin optimization algorithm (POA). These algorithms are adapted in the proposed study which aims at retrieving vital features of the complete data. Followed by this process, the projected model aims in performing the classification on both binary and in multi-class forms using certain classifiers such as RNN, LSTM, SSAE and using deep progressive attention-RNN+LSTM (DPA-RNN+LSTM). The complete efficiency of the mentioned system is assessed using the probabilistic quality measures constituting accurateness, f1-score, precision and several other parameters such as specificity, sensitivity and recall rates. These are used in validating the model performance in aspects of classifying the binary and multi-class data using IoT based real time data which is collected from the patients constituting heart rate and the ECG signal rates.

6.2 Proposed Design

This section deliberates the overall flow and a short term description for the overall method imposed in the projected study. The data used as input for the projected model is collected using the IoT sensors which are placed with the knowledge of human intervention. These are done in aspects of analysing patient monitoring in real time data and collecting them for diagnosing the disease and their cases of severity.As this data is collected from IoT and is real time data from patients, it contains more of unnecessary attributes and several outliners which are endured with high cases of making the model to be less robust and bringing inaccurate ranges of predictions. During the phase of pre-processing the data, these data are made free from several unnecessary attributes making the model learn only the needed data and bringing faster rates of prediction. This enhances cases of accuracy and reliability of the data. This process is also advantageous in removing the missing inconsistent data values which are resultant of human or computation error. This enhances the computational and accuracy rate within in set of data. Finally, the pre-processed data is loaded in to the model. This makes two different data classification which is possible for making a binary form of classification and is capable of multi-classification. This is depicted in figure 6.1 as a pictorial flow as an overall proposed flow.

This data is further divided into train and test data which are in the ratios of 20:80 respectively, the train data is used for training and test data is used in validating the model. The train data is then moved to the phase of feature scaling. This process enhances the process of optimization which makes the gradient descent flow to be smoother. This process also makes the algorithm to reach the minimum cost function more quick. These are used in normalizing the features of the dataset, which contains varying magnitude and units. This makes the model to converge quickly, and enhance the accuracy. For the process of feature selection, several algorithms are implanted such as GA, PSO, ABO, GSA and POA. These algorithms are adapted for the feature selections endeavors in selecting only the needful features for the model to perform better even in cases of large dataset. Further, the process of classification is handled using the classification models comprising the RNN, LSTM, deep progressive attention RNN+LSTM and stacked sparse encoder. Finally, the overall outcomes of the proposed framework is evaluated using the performance metrics in case of classification of the data collected using the IoT devices that track heartbeat and other parameters.

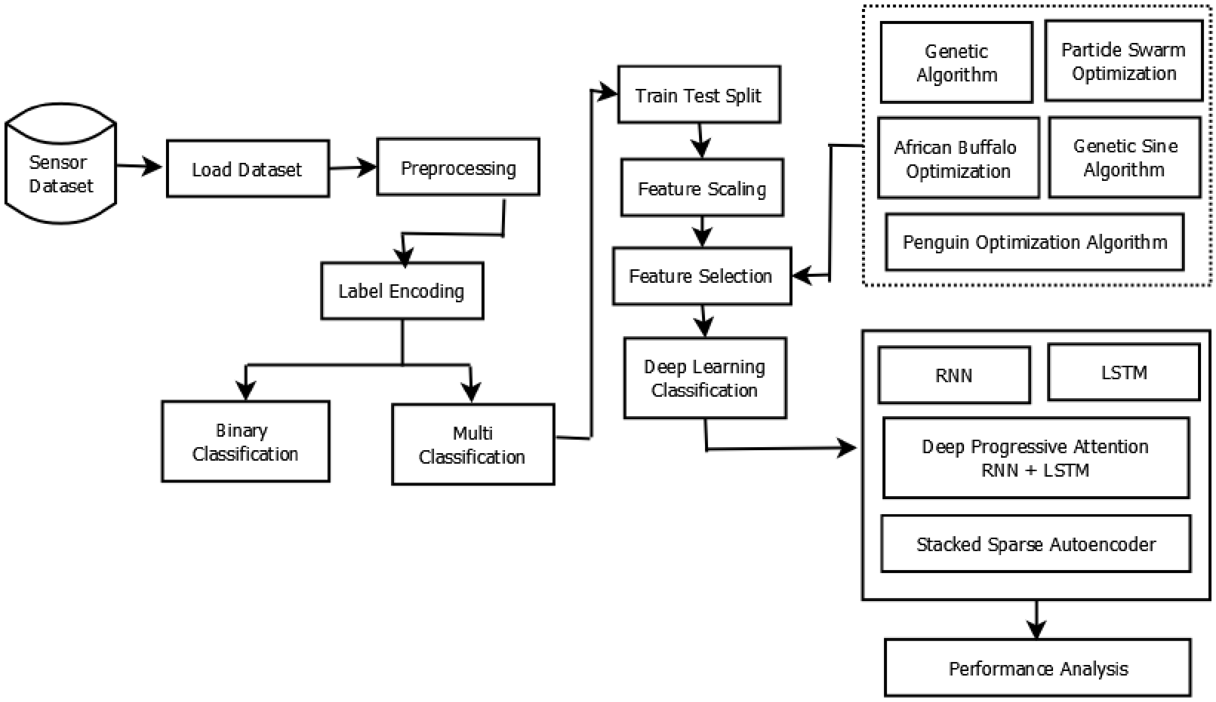

Figure 6.1: Overall Flow of Projected Approach

The data regarding the patients is obtained using the general purpose I/O (GPIO) pins such as arduino uno board which is connected to AD8232 board pins. Concurrently, these sensors are connected to the probes placed on the surface of body. These pins are allotted in two differing manners, in order to obtain ideal ranges of outcomes. This corresponding arduino board is connected to the system via USB port. This board contains a decoded programme before complete installation of the model being loaded to them. A separate software is used in analysing the results obtained using these graphical prototypes which is identical to the ECG devices. By this connectivity, the device points here are able to observe the QT and PR intervals from the individuals. BP sensor recorded bp values, FGA sensor recorded random sugar and haemoglobin values. Finally this arduino is linked to the raspberry-Pi via USB. A web service is also used in attaining the record collected from the patients by connected sensors. These are depicted from figure 6.2-6.5.

Figure 6.2: Prototype of model capturing ECG rates

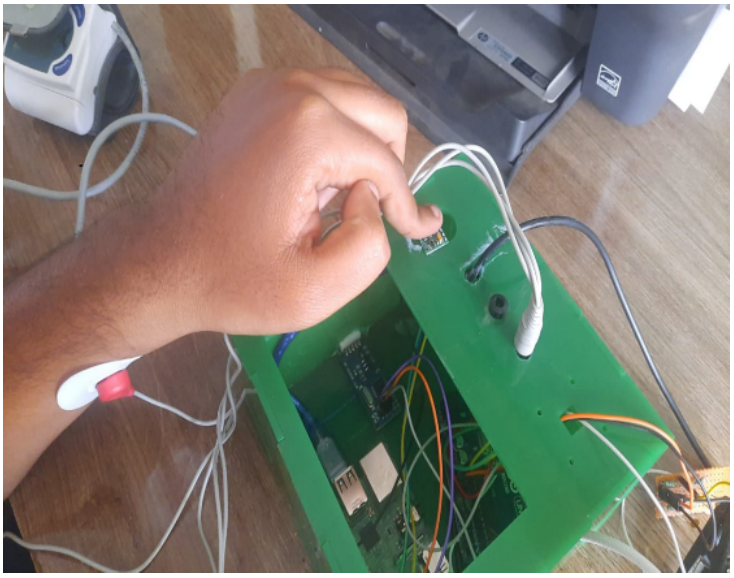

Figure 6.3: Prototype of model capturing SpO2 rates

6.3 Meta-Heuristic Algorithms in Feature Selection

In this section, the GA, PSO, ABO, GSA, and penguin optimization algorithm—meta-heuristic algorithms developed for the study are utilized.

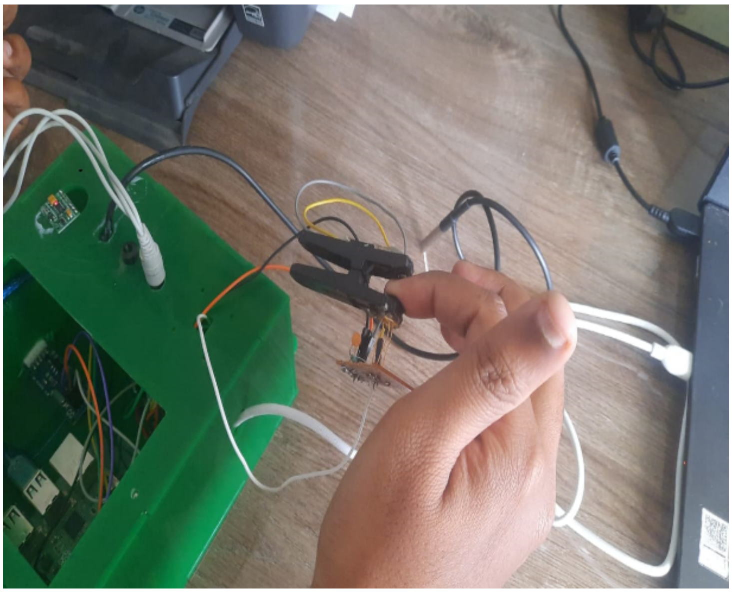

Figure 6.4: Prototype showing the capturing of random sugar and cholesterol rates

Figure 6.5: Prototype depicting of BP range capturing

6.3.1 GA- Genetic Algorithm

The GA algorithm, used in the proposed study by using the IoT heart disease samples of data as the input. This data is selected with higher ranges of adaptability and flexibility. The important aspect of using the GA algorithm is for proper construction and functioning along with the evaluation of fitness function. The process of fitness function produces a value which exposes the levels of optimal of the attained solution. This is done by participation of an individual in reproduction, thereby, producing new population. Thus, the biological functions are carried out on the developed population having probable solutions. This includes general functions such as

- Selection,
- Crossover and
- Mutation

Some of the other advantages of imposing GA are,

- Parallelism
- Has a large set for solution space
- Requires less information and are
- Probabilistic in nature

6.3.2 PSO- Particle Swarm Optimization algorithm

PSO algorithm is bio-motivated algorithm, used in aspects of finding an ideal solution within search space provided. This algorithm requires only the objective function and is also independent of gradient or differential kind of objective values. It makes use of only limited number of parameters. This approach makes use of social interaction concept to overcome an issue. Also, these algorithms are advantageous of being unaffected to scaling process of various design variables. Moreover, it encompasses less algorithmic parameters. Furthermore, it is one of an ideal global search approaches, which is easily parallelized for the aspects of simultaneous processing and are also derivative free. Some of the advantages associated with the PSO algorithm are:

- Insensitive to scaling of the design variables
- Free from derivatives

- Uses few algorithm to derive solution in search space and
- Effective phase of global search algorithm

6.3.3 African Buffalo Optimization (ABO) algorithm

This algorithm is used in the proposed approach is advantageous to resolve the issues related to stagnation or regarding the issues of pre-mature convergence. This is done by ensuring that, the location of individual buffalo is consistently updated with accordance to specific and optimal prior location of buffaloes and current location of ideal buffaloes present in herd. ABO is more capable of accomplishing enough exploitation. Similarly, ABO also assures a set of quick convergence using only limited parameters, which are principally the learning parameters $(l_{p1} and l_{p2})$. These learning parameters authorises movement towards greater ranges of both exploration or exploitation which relies on the concentration of algorithm within a specific iteration. Some of the added advantages of using ABO algorithm are,

- Insensitive to scaling variables
- Effective global search algorithm
- Uses few algorithmic parameters
- Better starter algorithm

6.3.4 GSA (Genetic Sine Algorithm)

GSA is considered to be one of an optimal features in accomplishing effective classification. This algorithm employs genetic concepts and natural selection for procuring solutions for resolving the issues. This algorithm initializes with solutions. Further, it is relocated with the optimal solution. This algorithm terminates the searching process the moment the termination condition seems to be satisfied. This algorithm is one of an optimizer which transposes the notions of natural evolution to digitized forms of computer era that imitates natural evolution.

It employs genetic concepts and natural selection for procuring solutions for issues. Some of the additional advantages in adapting the GSA algorithm are,

- Uses larger set of solution space
- Requires less set of input information
- Provides multiple phases of optimal solution and are with
- Higher adaptability

6.3.5 POA (Penguin Optimization Algorithm)

This algorithm is a bio-inspired which is from penguins residing in colder regions. The process of active aggregation are done to maintain warmth around them to protect themselves from low temperatures. These penguins do not possess any of territory and are devoid of nest to breed in case of emperor penguin. This algorithm is chiefly adopted for

- Resolving the optimization issues.
- The vital process in resolving these optimisation issues are resolved by generating huddle boundary, computing temperature throughout the huddle, computing distance and determining efficient mover.

Some of the other advantages of POA are,

- High-performing hardware is less needed
- Less computational complexity
- Used more in resolving mathematical complex tasks

6.4 Deep Learning Approaches for Classification

This approach makes use of certain DL methods in performing classification. The classification approach lies in making both the binary and theshow the presence or the absence of heart disease in an individual (binary) and the class of severity of heart disease (mild, moderate, normal and severe) classification.

6.4.1 RNN- Recurrent Neural Network

RNN is one of an ANN approaches, used in both time-series and in sequential forms of data. Computation is done by the inclusion of simple memory. Bias and similar weights are added to each of the layers that minimizes the complexity which are entangled with RNN parameters. Moreover, the standard path is provided for memorizing the prior results. This is done by affording prior result as input to subsequent layer. Some of the advantages of implicating the RNN in cases of performing classification are

- Able to handle varying length sequences
- Has an ability to memorise the past data provided as inputs using memory
- Non-linear mapping is possible
- Has higher rates of flexibility and improved accuracy rates
- Model size remains the same even for lengthier input

6.4.2 Long Short Term Memory (LSTM)

LSTM is one of a modified forms of RN for recollecting the existing information stored in memory. Long-term reliance is maintained by the LSTM by inclusion of memory-cell that acts as a container comprising information for enhanced time period. This memory-cell is regulated by three different gates namely input, forget and the resultant gate. Such gates decide on the information to be added, or eliminated from the memory cell. This LSTM is capable of processing both single as well as entire data sequences present. Usual LSTM unit encompasses of cell, input, forget and resultant gate. In this case, input gate manages the information to be included to memory-cell. Following this, forget gate manages the information to be eliminated from memory-cell. In conclusion, output gate controls the information to be attained from memory-cell. LSTM is appropriate for classifying, processing and deter- mining time series provided for unknown time. Some of the benefits in adapting the LSTM in performing classification are

- Adaptive handling in case of long-term dependencies
- Remembers information for extended period of time ranges
- Less susceptible to vanishing gradient issues
- Increased memory power to remember outputs from each model
- Higher scales of flexibility and simple with faster computational power

6.4.3 Deep Progressive Attention RNN+LSTM (DPA-RNN+LSTM)

Attention mechanism is used in unveiling the crucial information present in the data, by filtering out the invalid forms of information for a corresponding task. Using this mechanism, the classification mechanism can be enhanced. Thus, the current approach makes use of the attention method namely, DPA-with RNN+LSTM. In this aspect, both the LSTM and the RNN are considered along with the DPA due to their viable and innate forms of abilities and merits. Hereby, the RNN has the ability to possess the inputs which are of random size also encompasses the ability of recollecting the information in a specific time frame which are assisted in prognosticating the time series, on the other hand the LSTM comprises several parameters as input and the output biases. This makes the model free from fine-tuning mechanism. Taking all these credits into account, the proposed approach makes use of the DPA+RNN+LSTM as a primary approach in resolving the classification issues.

Thus, the score for l longitudinal variable at t^{th} time is evaluated using equation (6.1),

$$a(hid_t^l) = v_l \tanh(we_l^a \odot hid_t^l) \odot hid_t \tag{6.1}$$

Whereby, we_l^a and v_l indicate the weights for transforming results of l^{th} LSTM units as single value. Attention value for l^{th} longitudinal variable within t^{th} time slot computation is performed for finding the contribution of t^{th} time of l^{th} longitudinal value to overall prediction outcomes as expressed in equation (6.2),

$$p_t^l = \frac{exp(p(hid_t^l))}{(\sigma_{(t=1)}^T exp(p(hid_t^l)))} \tag{6.2}$$

Moreover, representation vector corresponding to l^{th} longitudinal variable computation is performed as per equation (6.3),

$$Zout_l = \sum_{t=1}^{T} (p_t^l \odot hid_t^l) \tag{6.3}$$

Some of the additional advantages of using the attention mechanism with the classification mechanism are of,

- Allows primitive guiding for complex systems such as modelling, prediction and for identification.
- Improved ranges of performance and accuracy of the model even using long and complex form of sequences.
- Permits the model to extract the increasing complex characteristics of the data when passes over the network.
- Improved quality in case of quality and in diversity of the output.

6.4.4 Stacked Sparse Auto Encoder (SSAE)

Auto encoder is one of an unsupervised learning method which encompasses of three main layers namely input, hidden and output-layer. The output layer is constructed similar to the input layer that assuages the process of reconstruction error used in extracting the ideal exploration of hidden-layer. This auto-encoder relies on two-vital phases for data processing. Initially, the input data is encoded in the hidden layer, from the input layer. This is given by equation (6.4)

$$b = f(a) = Sig(Weight(a) + bias) \tag{6.4}$$

Whereby, $b = [b^1.b^2b^m]^T$ indicates feature representation of hidden-layer.

$Sig(a)$ represents the sigmoid function which is denoted as $Sig(a) = 1/(1+e-a)$. Subsequently, feature representation (b) is decoded to output-layer from hidden-layer. This is provided by equation (6.5)

$$output = h(y) = Sig(Weight^T b + bias^{'}) \tag{6.5}$$

Feature representation (y) is decoded for attaining reconstructed vector result as $out = [out^1.out^2out^m]^T]$.

SSE, is the enhanced form of auto encoder (AE) which includes sparse restrictions for each of the hidden nodes. This is used in managing the activated neurons. The complexity of the system and the parameters could be minimized as a result of less activated neurons by which spare auto encoder could learn features in a better manner. Equation 6.6 provides the cost function for the spare auto encoder for the entire dataset.

$$P = /frac1m \sum_{i=1}^{m=1} [\frac{1}{2}(a^i - out^i)^2 + \frac{wac}{2} \sum_{i=1}^{m=1} \sum_{j=1}^{m=1} (wgt_{kl})^2 + beta \sum_{j=1}^{n} (\frac{\sigma \log \sigma}{sigma_j}) + \frac{(1-\sigma)\log(1-\sigma)}{((1-sigma_j))} \tag{6.6}$$

Initial term includes square root error (SRE) representing the variance amongst output and input. Subsequent term indicates weight decay utilized for solving over fitting issue, wherein wac denotes the weight attenuation coefficient and wgt_{kl} denotes weight that corresponds to input node (i) and hidden node (j). Final term involves sparse penalty, wherein σ denotes the sparse target, beta (β) represents weights corresponding to the sparse penalty, σ_j indicates average of activation quantity corresponding to hidden unit (j). Back propagation is selected for attaining parameters (W and bias) through minimization of cost function (P), where stochastic gradient descent is employed for training. These parameters in each of the iteration process could be updated as per equation (6.7) and equation (6.8),

$$wgt_{kl} = wgt_{kl} - \varepsilon \frac{d}{(d(wgt_{kl}))} P \tag{6.7}$$

$$bias = bias - \varepsilon \frac{d}{(d(bias))} P \tag{6.8}$$

Wherein ε denotes the learning rate. Further, a forward-propagation is also employed in cases of computing the average activation (σ_j) for obtaining error. This is followed by the application of backward-propagation method used for updating the parameters $(Wandbias)$. Considering the viability, the proposed approach makes use of the SSE, which is one of a NN encompassing of multi-layers of fundamental sparse auto encoder. The results of each individual layers are integrated to inputs of individual succeeding layer.

Thus, some of the advantages lying in using the SSE in classification mechanism are,

- Prevented abundance of data in model during processing.

- Conductive learning of the model is achieved based on unlabelled data.
- Hidden layers are also provided with excess information which are encoded about the training data.
- Addresses complex problems such as adhesion state recognition.
- Required memory is substantially reduced which reduces the run time, and space.

6.5 Results

This section deliberates the outcomes provided by the proposed model with their dataset description, model evaluation using appropriate key measures, the obtained experimental results and their description. The result section is also explained with the dataset which is used in the respective study.

6.5.1 Dataset Description

This research model works on the data being acquired from the 312 individuals using an IoT sensors. These sensors are used in collecting the real time working conditions of the human body. These sensors are linked devices which are able to capture the real time data, translates this data and format it. The data collected from these population is insufficient in making a complete analysis, thus SMOTE method is used in the respective data to augment the number of data. From this approach about 1591 data records are obtained. These data augmentation techniques are used in case of making insufficient quantity of data to be duplicated without missing original sensitivity of the data. The data is created using the original data as the mother data. In this case, about 11 vital attributes of human functioning is considered. This includes, BP rates, Cholesterol levels, LDL, HDL and several others. These are listed in table 6.1.

Table 6.1: Different sensors used in measuring health attributes

Sensors	Attributes
Blood-Pressure (Sunrom BP)	Systolic BP Diastolic BP
Pulse sensor	Heart rate
TSL2561 Module TIL32 LED FGA10 Photodiode	Cholesterol LDL Cholesterol HDL Total cholesterol level Random sugar and HB
ECG (AD 8232)	PR interval QT interval
MAX30100	Oxygen Saturation

6.5.2 Performance Metrics

Performance metrics are used in assessing the model based on their performance which is relied on their behavior, during both training and testing. This corresponding study makes use of certain performance metrics in evaluating their model in case of performing binary and multi-class classification using the current information gathered by IoT sensors.

6.5.2.1 Accuracy

Accuracy is a measure of describing the performance of the model across all of the classes. These are useful in measuring if all the classes are equally considered. This is generally the ratio among the number of accurate prediction made by the model to the total number of predictions made. This is given by equation (6.9)

$$Acc = \frac{(TRN + TRP)}{(TRN + FLN + TRP + FLP)} \tag{6.9}$$

Where, TRN is true negative, TRP is true positive, FLN is false negative, and FLP is false positive.

6.5.2.2 Recall

This metrics is defined to be the ratio among the total number of positive samples which is classified by the model to the total number of positive samples present. This measures the ability of the model to classify the positive samples. If the levels of recall rate are higher, more positive samples are detected. This is given by equation (6.10)

$$Recall = \frac{TRP}{(TRP + FLN)} \tag{6.10}$$

6.5.2.3 Precision

This metrics is defined to be total number of true positives, which are divided by the total number of positive and negative observations in the data. This is provided by equation (6.11)

$$precision = \frac{TRP}{(FLP + TRP)} \tag{6.11}$$

Where, TRP is true positive, FLP is false positive.

6.5.2.4 F1-score

This is one of a metric used in measuring the model accuracy. This shows the measure of models lower false positive and lower false negative measuring ability. F1-score is defined to be the weighted harmonic-mean value of the precision and the recall rates, this is estimated using equation (6.12)

$$F1 - score = 2(RcPc)(Rc + Pc) \tag{6.12}$$

6.5.2.5 Sensitivity

This is one more measure used in measuring the wellness of a model in predicting the positive instances. This is also known to be the true positive rates. This is used in evaluating the performance of the model being able in detecting the positive instances correctly. This is evaluated using equation (6.13)

$$sensitivity = \frac{TRP}{(TRP + FLP)} \tag{6.13}$$

6.5.2.6 Specificity

This is measured to be the proportion of true negative that is appropriately detected by the model. This metrics implies that there are other proportions of an actual negative which is predicted as positive and are termed to be false positives. This is given using equation (6.14)

$$specificity = \frac{TRN}{(TRN + FLN)} \tag{6.14}$$

6.5.3 Experimental Results

This section deliberates the overall outcomes predicted by the proposed model, according to both binary and the multi-class classification. The table 6.2 explains the results obtained for classification using the ABO algorithm under various performance metrics constituting accuracy, precision, f1-score, recall and the rest of other metrics.

6.5.3.1 Results for Binary Classification

Table 6.2: Classification results by ABO for suitable classifiers and with appropriate features

Classifier	Acc	Precision	Recall	F1	Specificity	Sensitivity	features
RNN	99.16	99	99	99	99	99	00111100111
LSTM	99.79	98	98	98	98	98	01111111110
SSAE	98.43	98	98	99	98	98	01111110111
DPA-RNN+ LSTM	100	99.24	99.62	99.43	99.62	99.62	01111100111

From table 6.2, accuracy ranges of ABO on RNN is found to be 99.16%, while ABO-LSTM has shown 99.79%, ABO-SSAE has revealed 98.43%, while DPA-RNN+LSTM model with ABO has exposed 100%. Following this, the precision rates, sensitivity, specificity, f1-score, and recall rate of all the considered algorithms have shown better results in ranges of 99% for ABO-RNN, 98% for LSTM-ABO and for SSA-RNN. But the DPA-RNN+LSTM model have achieved precision rate of 99.2% which outperformed other classifier models. Final, the other such considerable metrics such as recall, F1-scores, specificity and the sensitivity ranges of the DPA-RNN+LSTM are in scales of 99.6%, 99.4%, 99.6% and 99.6%. This showed that the DPA-RNN+LSTM has performed better cases of binary classification using ABO algorithm. The performance relied to the GA with binary classification performed is tabulated in table 6.3.

Table 6.3: Classification results for GA with the considered classifiers and selected features

Classifier	Acc	Precision	Recall	F1	Specificity	Sensitivity	features
RNN	100	98	98.57	98	97	98	01111111111
SSAE	87	99.50	76.72	86.64	76.72	99.13	01111111111
LSTM	100	97.40	98.68	98.68	96.74	96.74	00111111110
DPA-RNN+ LSTM	100	100	100	100	100	100	01111111111

From table 6.3, accuracy value of GA-RNN is found to be 100%, while GA-LSTM has shown 100%, GA-SSAE has revealed 87%, whereas, DPA-RNN+LSTM model performing classification (binary using GA algorithm) has achieved 100%. This is achieved for all metrics such as, the accuracy, sensitivity, f1 score, specificity and recall rates. Nevertheless, DPA-RNN+LSTM system has less-performed when compared to other classification models. Similarly, the outcomes that obtained which corresponds to binary classification using GSA with miscellaneous classifiers are listed in table 6.4.

Table 6.4: GSA Classification results with the measured classifiers and selected features

Classifier	Acc	Precision	Recall	F1	Speci	Sensi	features
RNN	98.43	98.88	98.87	99.43	98.55	98.87	2 3 4 5 6 7 8 9 10 11
LSTM	99.37	98	97.25	98	97.61	97.55	1 2 3 4 5 6 7 8 9 10 11
SSAE	100	95.80	77.40	85.63	77.40	95.74	2 3 4 5 6 7 8 9 10 11
DPA-RNN+ LSTM	100	100	100	100	100	100	2 3 4 5 6 7 8 9 10 11

From table 6.4, accuracy measure of GSA-RNN is to be 98.43%, while GSA-LSTM has shown 99.37%, GSA-SSAE has revealed 100%, but the DPA-RNN+LSTM model performing classification using GSA algorithm has attained an accuracy range of 100%. Moreover, the other performance indicators like the f1-score, sensitivity, precision, specificity and recall rate of all the deliberated algorithms have shown explicit outcomes achieving them in 100% ranges. Nevertheless, DPA-RNN+LSTM model has comparatively performance lag when compared to other classification models. Similarly, the conclusion that agrees to binary classification for POA with varied classifiers are exposed in table 6.5.

Table 6.5: Classification results by POA with the considerable classifiers and features

Classifier	Acc	Precision	Recall	F1	Specificity	Sensitivity	features
RNN hline SSAE	100 100	98 100	97 100	97 100	97 100	97.2 100	10 10
LSTM	100	98	97	97	97	97	10
DPA-RNN+ LSTM	98.74	100	100	100	100	100	10

From table 6.5, accuracy value of POA-RNN is found to be 100%, while POA-LSTM has shown 100%, POA-SSA has revealed 100%, and DPA-RNN+LSTM model using POA in aspects of performing the multi-class classification by the proposed model has achieved markable outcomes under all of the performance metrics such as precision, f1-score, recall, specificity and the sensitivity rates which shows better performance in case of the DPA-RNN+LSTM model. Consistently, the conclusions that approves to binary classification for PSO with different classifiers are deliberated in table 6.6.

Table 6.6: Classification results for PSO with the considered classifiers and selected features

Classifier	Acc	Precision	Recall	F1	Specificity	Sensitivity	features
RNN	99.79	99.24	99.62	99.62	99.07	99.07	01111110111
SSAE	99.70	98.55	77.86	86.99	77.86	98.60	01111111111
LSTM	99.58	99.58	91.22	95.22	99.53	91.22	00111111111
DPA-RNN+ LSTM	99.37	100	99.62	99.80	100	99.62	01111111110

From table 6.6, accuracy value of PSO-RNN is found to be 99.79%, while PSO-LSTM has shown 99.58%, PSO-SSAE has revealed 99.70%, while DPA-RNN+LSTM model with PSO has unveiled 99.37% accuracy. Followed by, the precision rates have achieved by 100%, and sensitivity at ranges of 99.61%. Similarly specificity have also been attained at rates of 100%, recall rate of the DPA-RNN+LSTM model have also produced out ranged results showing 100%. Nevertheless, the DPA-RNN+LSTM system has shown superior outcomes in comparison to other models. Thus, from the empirical outcomes obtained on the aspects of binary classification, the DPA-RNN+LSTM model has shown better outcomes in cases of using GA, GSA and POA by showing 100% accuracy rate.

6.5.3.2 Results for Multi-Classification

Initially, the multi-class classification of the proposed model with suitable classifiers under the probabilistic performance metrics are listed in table 6.7.

From table 6.7, accuracy ranges of ABO-RNN is obtained with scales of 96.46%, while ABO-LSTM has shown 87.54%, in case of ABO-SSAE has attained in ranges of 99.3%, but DPA-RNN+LSTM model with ABO has exposed the accuracy ranges in 98.99%. Following this, the other probabilistic performance metrics constituting precision, f1-score, specificity, sensitivity, and the recall rates of the other considered algorithms have shown better results

Table 6.7: ABO Classification outcomes on considered classifiers and designated features

Classifier	Acc	Precision	F1	Specificity	Sensitivity	Recall	features
RNN	96.46	96.48	96.41	100	97.99	96.53	11110011111
LSTM	87.54	87.38	87.04	96.24	79.07	87.57	11111111101
SSAE	99.3	96.57	96.55	99.25	93.75	96.63	11110011111
DPA-RNN+ LSTM	98.99	99.03	98.99	98.51	100	98.98	11111111011

which are under ranges of 96%, 99%, followed by the specificity and the sensitivity rates in ranges of 99.2% and 93.7% for the SSAE-ABO model. But the DPA-RNN+LSTM model have attained an accuracy rate of 99%, precision values of 99.09%, f1-score in levels of 98.9%, specificity and sensitivity scores of the DPA-RNN+LSTM model are 100% and 98.8% respectively. Followed by, the outcomes corresponding to multi-classification by GA with DPA-RNN+LSTM model using diverse classifiers are represented in table 6.8.

Table 6.8: Classification outcomes for GA with the considered classifiers and selected features

Classifier	Acc	Precision	Recall	F1	Specificity	Sensitivity	features
RNN	97.98	97.93	97.99	97.94	100	96.55	11111111011
LSTM	87.88	84.62	84.80	84.48	94.49	82.17	11111111101
SSAE	99.3	92.03	91.85	91.61	97.01	94.44	11110011111
DPA-RNN+ LSTM	98.48	99	99	99	100	100	11111111011

From the table 6.8, the accuracy ranges of the RNN-GA are in ranges of 97.98%, for LSTM in ranges of 87.88%, similarly for SSAE the accuracy rates are in ranges of 99.3%. Whereas the DPA-RNN+LSTM model has achieved an accuracy rate of 98.48%. Similarly, the DPA-RNN+LSTM model performing the multi-class classification using the GA algorithm under various measurable performance metrics in aspects of, sensitivity and specificity has obtained markable results in ranges of 0.99, for all precision, f1-score and for the recall rates. Finally The DPA-RNN+LSTM model has succeeded 100% in ranges of both sensitivity and specificity for the multi-class classifications. Similarly, the results that corresponds to multi-classification for GSA with diverse classifiers are presented in table 6.9.

Table 6.9: Classification outcomes for GSA with the considered classifiers and selected features

Classifier	Acc	Precision	Recall	F1	Specificity	Sensitivity	features
RNN	96.72	96.66	96.73	96.66	97.85	95.75	0 1 2 4 5 6 7 8 9 10
LSTM	97.22	97.13	97.21	97.15	98.92	95.70	0 1 2 3 4 5 6 7 8 9
SSAE	97.73	97.66	97.69	97.64	100	93.62	0 1 2 3 4 5 6 7 8 9
DPA-RNN+ LSTM	97.47	97.92	97.97	97.92	98.99	98.91	0 1 2 4 5 6 7 8 9 10

Correspondingly, the inferences that agrees to multi-classification for POA with different classifiers are exposed in table 6.10.

Table 6.10: Classification outcomes for POA with the considered classifiers and selected features

Classifier	Acc	Precision	Recall	F1	Specificity	Sensitivity	features
RNN	94.45	92.09	91.49	91.45	98.91	92.41	0 1 2 3 5 6 7 10
LSTM	72.73	72.59	72.59	72.45	89.61	75.31	0 1 3 4 6 9 10
SSAE	93.94	93.88	93.79	93.79	97.75	91.21	0 2 4 5 6 7 8 9 10
DPA-RNN+ LSTM	98.48	98.48	98.48	98.48	98.92	100	0 2 4 5 6 7 9 10

Compatibly, the deductions that supports to multi-classification for PSO with diverse classifiers are explored in table 6.11.

Table 6.11: Classification outcomes for PSO with the considered classifiers and selected features

Classifier	Acc	Precision	Recall	F1	Specificity	Sensitivity	features
RNN	90.24	90.44	90.43	90.21	98.48	91.09	11110011111
LSTM	94.11	94.07	94.07	94.01	95.49	94.48	11110011111
SSAE	94.78	95.02	94.74	94.77	93.98	98.65	11110011111
DPA-RNN+ LSTM	98.82	98.88	98.81	98.82	98.51	100	11110011101

Finally, from table 6.11, the accuracy rates, precision and the recall rates of the DPA-RNN+LSTM model performing the multi-class classification are in scales of 98.82%. Whereas, the recall and the f1-score rates of the proposed model are 4% to 8% greater than the other classifiers performing the classifications. Further, the mentioned model has attained the specificity and the sensitivity rates in ranges of 100% and 98%. The empirical outcomes of the multi-classification performed by the given model has shown better results with GA and ABO algorithm with 99% accuracy.

6.5.4 Performance Analysis

The performance of the proposed model analyzed using the performance analysis comprising for both the binary and the multi-class classification. In this aspect, the analysis are carried out using the confusion matrix which is used in identifying the correct and the incorrect predictions made by the model. This is used in measuring the efficacy of the proposed model. This is also used in evaluating the performance of the trained model, in aspects of finding the model can perform based on the dataset.

6.5.4.1 Performance Analysis for Binary Classification

The performance of the proposed model in performing the binary classification using the GA algorithm is depicted in figure 6.6

From the figure 6.6, it is observed that the right ranges of prediction are made in both the cases of classifying them in cases of normal and the abnormal in ranges of 215 and 262. Moreover, the model had no mispredictions on both the normal and the abnormal cases.

From figure 6.7, the classification performed by the model, clearly depicts that the model has predicted 141 normal cases and 177 abnormal cases using the GSA algorithm in the proposed model. This model also affirms that no (0) mispredictions are made by the proposed model.

From figure 6.8, it is completely obvious that the proposed method performs 215 normal case prediction and 262 abnormal case prediction. Meanwhile, the proposed model has made

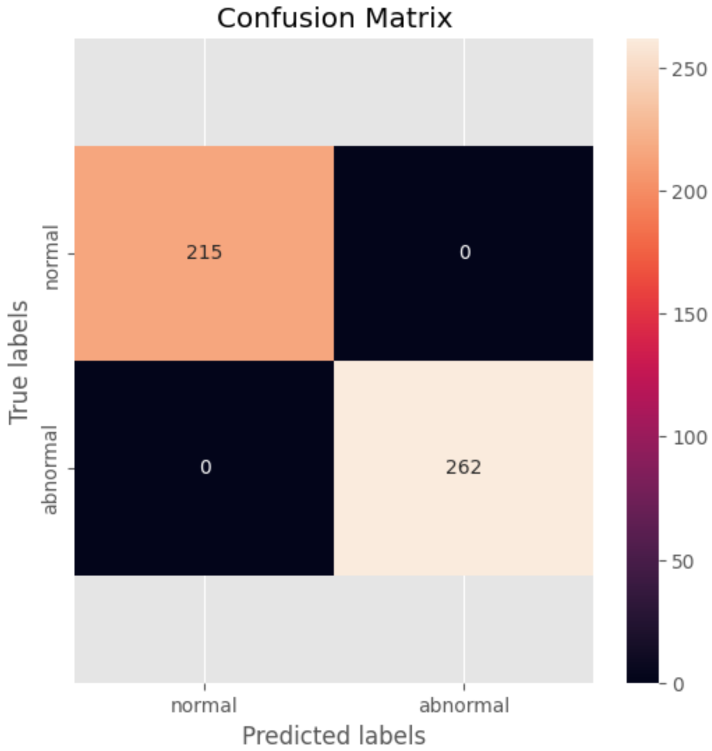

Figure 6.6: Confusion Matrix for GA performing binary classification

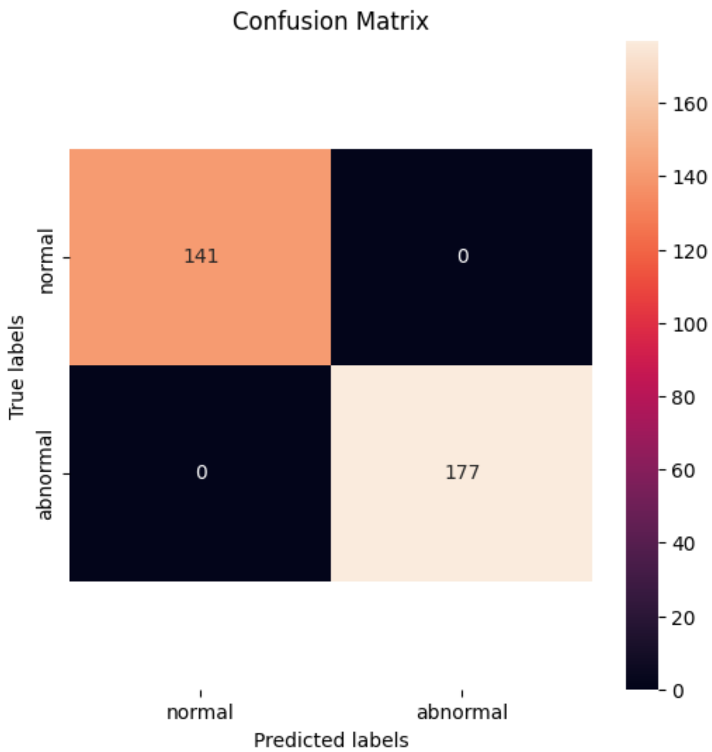

Figure 6.7: Confusion matrix for proposed system using GSA

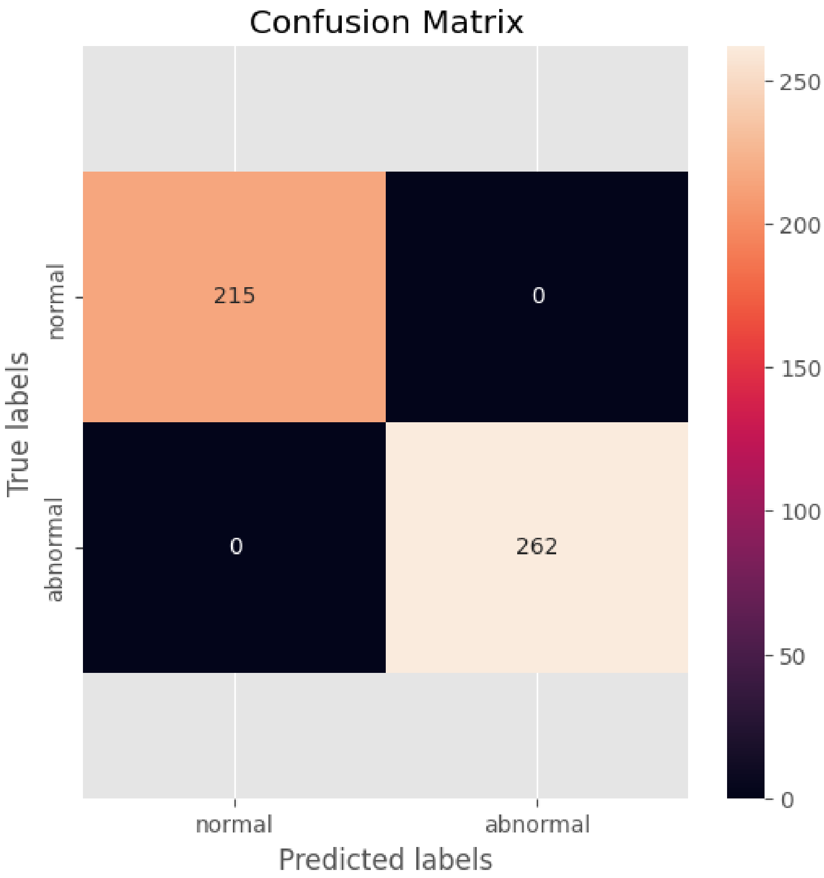

Figure 6.8: Confusion matrix for proposed system using POA

0 mispredictions rates. Thus, on an overall performance analysis of the proposed system performing the binary class classification. This exposes the proposed model ideal performance in handling the prognosis of heart disease.

6.5.4.2 Multi-class Classification performance analysis

The performance analysis has been carried out in this section for evaluating the performance of the proposed model in terms of making the multi-class classification. This is analyzed using the creation of Confusion matrix. It is given in figure 6.9

From figure 6.9 it is clear that the proposed model when performing classification using the ABO algorithm in aspects of multi-class classification have made prediction in scales of 132 normal cases, 149 mild cases, 151 moderate cases and 156 severe cases with heart diseases. This confusion matrix also deliberates that the proposed model also made 2 incorrect prediction in cases of normal heart, no mispredictions were made in cases of mild heart disease ranges.

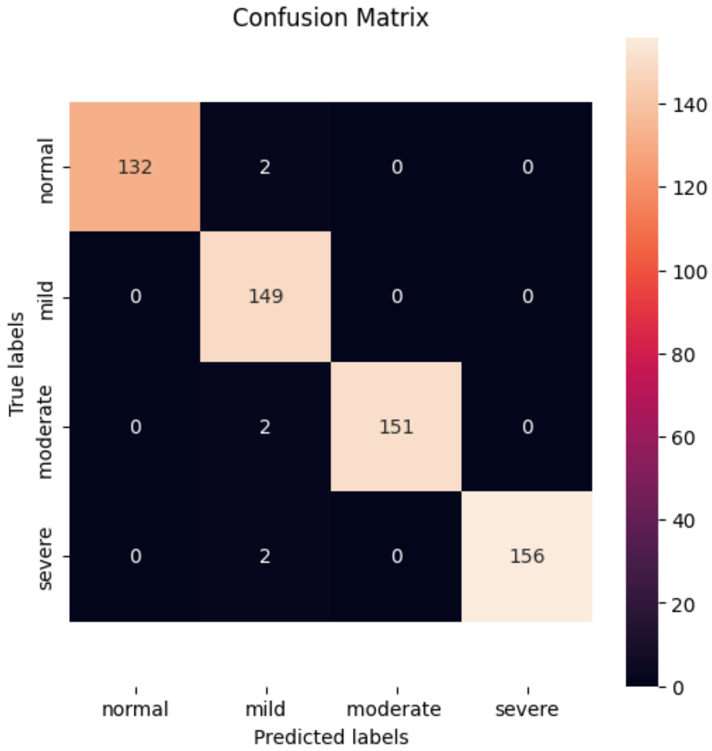

Figure 6.9: ABO Confusion Matrix for Proposed Model

Further, 2 incorrect predictions have been made in cases of moderate cases of heart rate and 2 incorrect predictions have been made in aspects of severe heart disease. This showed the efficacy of the model performing the multi-class classification using the ABO algorithm. Further the corresponding outcomes shows the efficacy of the proposed model.

Figure 6.10, clearly states that the proposed model using the GA model in performing the multi-class classification in cases of diagnosing the heart disease based on their severity. The proposed model resulted in 134 correct predictions in normal case identification, followed by 146 correct predictions in aspects of mild cases. 150 moderate cases and 158 severe case predictions, this model also made 3 incorrect predictions in cases of moderate and as mild heart disease rates. Further, 0 cases of incorrect prediction on normal and in severe rates, expressing the model better case of performance. Thus on an overall view, the model preforming better in aspects of making higher ranges of correct predictions and very few incorrect scales of predictions.

Concurrently, the RNN used in performing the classification is offered by the capability of processing the random size, and has an ability in performing them in a specific time ranges. The LSTM used here, offers the model with numerous parameters which encompasses the learn-

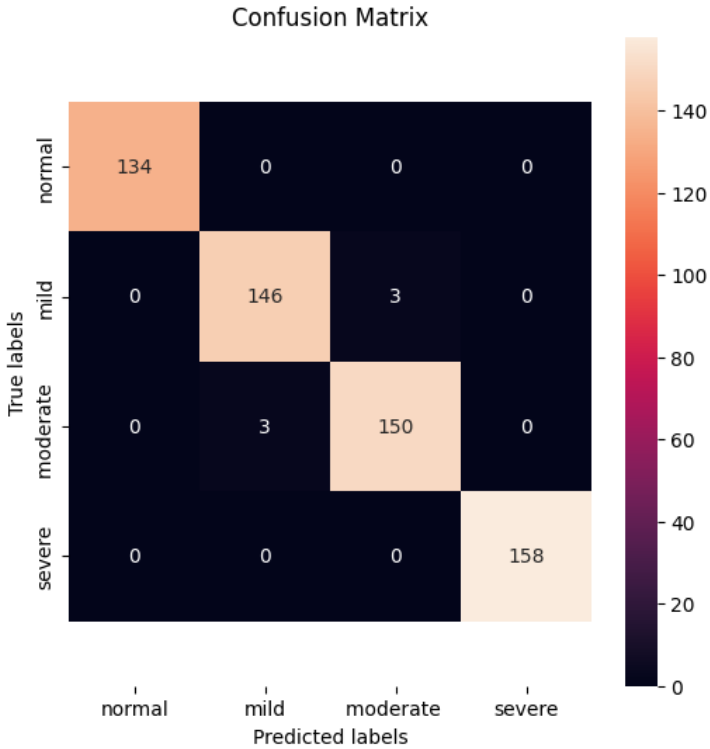

Figure 6.10: GA used in proposed system for multi-class classification

ing rates, with input and output biases. This makes the model to be devoid from fine-tuning. Considering all these advantages into account, the proposed approach has resulted in bringing optimal outcomes when analyzed using performance analysis.

As the proposed approach makes use of the real-time data collected using the IoT sensors, comparative analysis with the other model and state-of-art approaches cannot be presented.

6.5.5 Summary

This chapter deliberated the performance of the proposed models. This approach makes use of the IoT based sensor in diagnosing health attributes for the prognostics of severity of heart issues among populations. This approach makes use of several meta-heuristic algorithms for performing the feature selection. These algorithms include, GA, PSO, ABO, GSA and the POA. The results are classified using the approaches implicating RNN, LSTM, SSAE and DPA-RNN+LSTM. Both the binary and the multi-class classification are performed in this approach. The binary form of classification is used in classifying whether the heart disease exists or not.

Whereas, the multiclass classification is used in classifying them into mild, moderate, normal and severe classes of heart disease occurred in an individual. From the overall results obtained it is evident that the binary classification performed by GA, GSA and POA on proposed classifiers i.e. SSAE and DPA-RNN+LSTM showed ideal performance with 100% accuracy rates. Similarly, for the case of multi-class classification ABO and GA in the proposed system i.e. SSAE deliberated optimal scales of performance with 99.3% accuracy range. Confusion matrix performed for this respective approach shown very less cases of incorrect interpretation and higher ranges of precise scales of classification. All these effectiveness are considered for making the proposed approach efficient in cases of performing the features selection and the classification of heart diseases. These are done based on either presence or absence of the heart disease followed by the severity classification done using the multi-class classification.

CHAPTER 7

IOT BASED HEART DISEASE PREDICTION USING REGRESSION METHODS

7.1 Introduction

HD is considered as a potentially fatal disease as a result of multiple contributory risk factors such as abnormal pulse rate, high bp, diabetes, excessive cholesterol as well. Timely and accurate diagnosis is critical for HD therapy and prevention. Various approaches are used to predict HD, but they are time-consuming and inaccurate.

In phase (I), the presented system has implemented feature selection methodologies like PSO, GA, GSA and ABO to remove unnecessary and duplicate elements from the provided UCI dataset and also to prevent features from becoming trapped in local minima. With the created GSA, features are chosen. The accuracy of HD classification would be improved by this process, which would also produce reliable prediction outcomes. For the classification of specific features, the system has also implemented DL-based RNN combined with LSTM, and DPA-RNN+LSTM has been used to boost the model's classification rate. In addition to determining whether HD is present or not, the developed method aids in this process.

Phase (II) of the presented model employs stacked sparse convolutional based auto-encoder for the classification of HD and POA for feature optimisation. The proposed study begins with pre-processing the data, then moves on to the emperor penguin optimization system, which is used to choose the best features for classification. Finally, a stacked sparse CNN-based auto encoder approach is used to increase classification accuracy and efficiency while minimizing reconstruction error. Using benchmark heart disease datasets like the UCI cleveland dataset and the HD clinical dataset aids in predicting the results more accurately. The web-service is launched from the raspberry-pi in phase (III) using the arduino uno and AD8232 sensor framework. This made it possible to receive the patient's BP, hb and heart rate parameters in real time. Since IoT-based data is compiled from a variety of sources, it may be chaotic and noisy. IoT-data mining is employed to help with tasks like defining typical links between data components and using them to address prognostication concerns. In order to accomplish this, the research makes use of 5 data mining-based algorithms for achieving the ideal results.

The current study uses a flask based web model to identify the heart information on the basis of cardiac risk score, cardiac index and vascular age to implement a system for monitoring an individual's health on real-time IoT based sensor dataset. By entering the input variables such as id of patient, age, etc., prediction of HD is also determined. The results are displayed based on the variables entered, and by clicking, resultant values are calculated by the model. To get better results, the five optimization algorithms are combined with DL techniques. The currently used techniques have implemented multi and binary- class classification utilizing a wide variety of DL and optimization approaches. In order to determine whether an individual is normal or anomalous, the proposed method applies five different optimisation techniques for feature selection, DL-based algorithms for HD classification, and regression of heart-based data. The flask-based web model is based on real-time sensor-based dataset, in conflict with current techniques. The overall workflow is shown in figure 7.1.

DL Approaches

GA

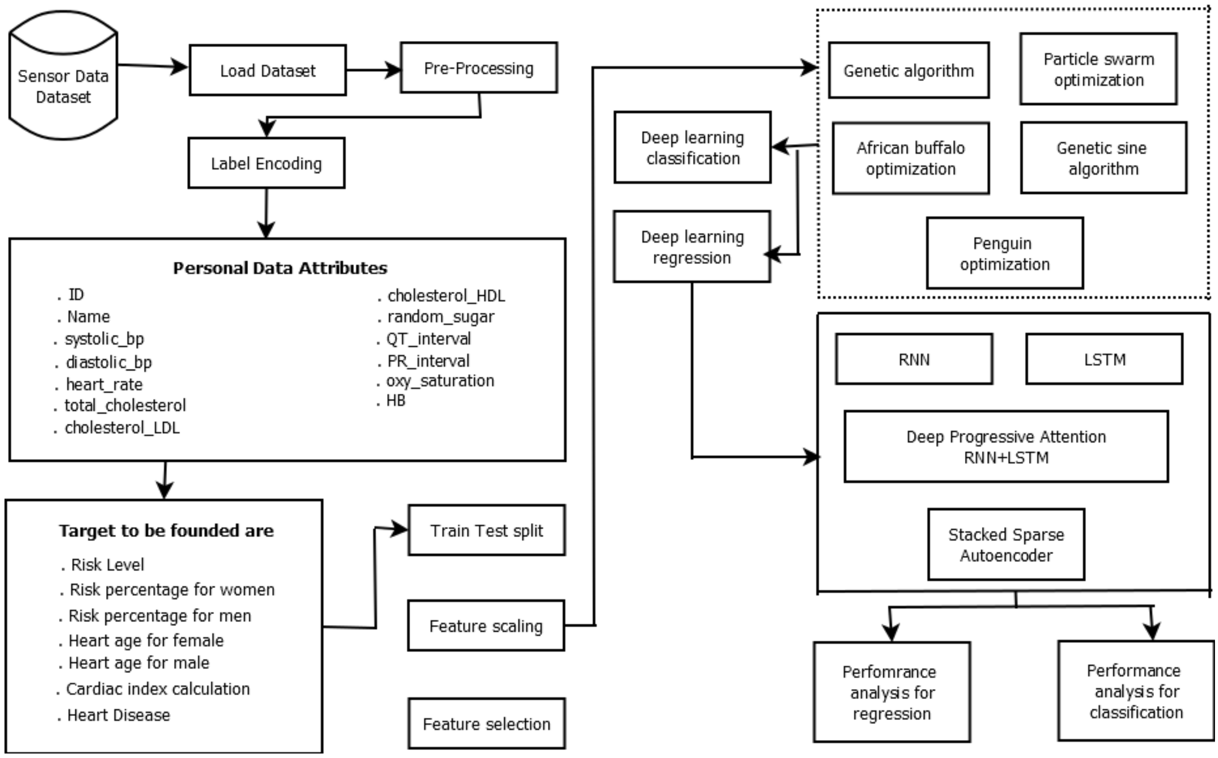

Figure 7.1: Whole process of classification of HD and regression of IoT based Sensor Dataset

The GA is an optimisation algorithm that uses recombination operators to maintain vital data while encoding a possible solution in order to exactly issue on a basic chromosome-like module. In GA, the solutions are referred to as chromosomes. Where the chromosomes are made up of genes, and particular alleles represent the issues. The collecting of issues is referred to as a population. In GA, there are three operators: mutation, crossover and selection. The selection operator is used for choosing individuals for representation. Here, Crossover refers to the process of selecting two parent chromosomes and producing a child. Mutation, on the other hand, is utilised to modify novel approaches in the hunt for the best solutions, preventing GA from becoming locked in a local optimal solution.

PSO

It is a computation process that is imitated from the birds flying activities and their mode of information exchange. In PSO, each specific solution is represented by a bird in search space, which is then copied as a particle. These particles have fitness values that are calculated using a fitness function that has to be enhanced and velocities that direct flying particle. It is used to address various complex optimisation problems. By increasing the generation levels, the model is initially built up with a population of unsystematic solutions also searchers for optima. This method is simple to apply because it does not employ evolution mechanisms like crossover and mutation.

$$data_j^c = (data_1^c, \ldots\ldots, data_d^c, \ldots\ldots.., data_D^c) \tag{7.1}$$

$$data_1^k \varepsilon [lo_n, uo_n], 1 \leqslant d \leqslant D, lo_d, uo_d \tag{7.2}$$

The velocity present, $VV_j^c = vv_1^c, , \ldots..vv_d^c, \ldots\ldots.., vv_D^c$ is constrained by the minimum velocity $VV_min^c = (v_(mini,1)^c, \ldots, v_(mini,d)^c, \ldots\ldots, v_(mini,D)^c)$ and the maximum velocity $VV_max^c(vv_(maxi,1)^c, \ldots., vv_(maxi,d)^c, \ldots\ldots$
$, vv_(maxi,D)^c)$.

Initially, particles are upgraded with two superior values in each search iteration. One among those values are pbest and another is gbest. After computing the two superior values the particles upgrade the position and velocity by employing the equations (7.3) and (7.4),

$$VV_j(c+1) = wi(VV_j(c) + jk_1e_1(PP_j(c) - data_j(c) + jk_2e_2(IP_g(c) - data_j(c))) \tag{7.3}$$

$$data_j(c+1) = data_i(c) + VV_i(c+1) \tag{7.4}$$

Where PP_j represents the pbest of jth particle and IP_g is the gbest of particles. The IP_j and IP_g is shown by equation (7.5),

$$IP_j = \begin{pmatrix} IP_j of(data_i) \geqslant of(IP_j) \\ data_j of(data_i) < of(IP_j)) \end{pmatrix} \tag{7.5}$$

$$IP_g \varepsilon IP_0, IP_1, \ldots .IP_m \mid (of(IP_g) = min(of(IP_0), of(IP_1), \ldots \ldots of(IP_m)) \tag{7.6}$$

where $h \leq H$, and H represents the whole particles from two random sequences in range of $(0,1) : e1 \sim V(0,1); e2 \sim V(0,1)$.

ABO

ABO method's main goal is to elevate an improved, robust, handy and convenient to use meta-heuristic method. This algorithm represents the replication of african buffalo's motion initiating from one site to another over african forest and savannahs, to conquer immense hungers. African buffalos create some sounds to gather animals for manipulating a particular area and uniting altogether to discover diverse region. Based on this technique of african buffalos, ABO algorithm exhibits the unique ability within exploitation and exploration of search space through solving the premature merging problems. Usually, the ABO method follows star topology that entwines the entire buffalos with everyone, as the whole herd behaves as an entity in searching apt solutions, concerning with data propagation. This technique involves three characteristic feature of african buffalos, via facilitating the search process. Also they have a broad data capacity, obliging communication ability, depending on independent nature of great intelligence. Therefore, the vital features of population related method represents the obliging in individuals, in order to achieve numerous exploration and exploitation methods by sharing information between individuals.

GSA

GSA develops a central merging feature. The bad particles of fitness may learn to reduce the impact of good particles existing in population. In GSA, the particles move with regard to the search space centre that makes it convenient to fall into the local minimum. Where the sub-optimum solution exist in the centre area and the global optimum solution is present across boundary.

$$A_j^{u+1} = A_j^u + s_1 \times sin(s_2) \odot ||s_3 Q_j^u - A_j^u| \tag{7.7}$$

$$A_j^{u+1} = A_j^u + s_1 \times cos(s_2) \odot ||s_3 Q_j^u - A_j^u| \tag{7.8}$$

Where A_j^{u+1} is the position of present solution in the jth dimension at the uth iteration. And Qj is the location of place point present in the uth dimension. The three random values are $s1, s_2$ and s3. The equations (7.7) and (7.8) are joined together in order to form the equation (7.9),

$$A_j^{u+1} = \begin{pmatrix} (A_j^u + r_1 \odot \sin(s_2) \odot |s_3 Q_j^u - A_j^u|, s_4 < 0.5 \\ A_j^u + r_1 \odot \cos(s_2) \odot |s_3 Q_j^u - A_j^u|, s_4 \gneqq 0.5) \end{pmatrix} \tag{7.9}$$

According to the equation (7.9), it is represented that in GSA there are 4 central parameters, where s4 represents the random value among 0 and 1. And r1 denotes the motion direction of subsequent position place, r2 displays at which range the motion should be, both outwards or towards from targeted place.

Penguin Optimisation Algorithm (POA)

This algorithm is contemplated as one among the kind of meta-heuristic technique that is developed from the characteristic of emperor penguin huddling. Superior results are produced when related to other algorithms in terms of intensification and diversification. Here, the huddling technique is accountable for better diversification that aids POA to achieve better global search competency. The superior intensification is implemented by moving attributes of POA that enhances the better local search competency. Also it offers superior switch from diversification to the intensification. The POA algorithm is represented as presented below,

$$\theta t = \frac{(\theta t - (iteration_{maxi})}{(d - iteration_{maxi})} \tag{7.10}$$

For emperor penguins, θt denotes temperature profile, d is the present iteration and is represented as iteration_maxi. As the emperor penguins huddle together to maintain temperature, certain protective measures has to be applied to protect from neighbourhood crashes. Thus, two vectors are implied namely as W and Y and is assessed as displayed in the equation (7.11) and (7.12),

$$W = [N * (\theta t) + input_{feat}] \tag{7.11}$$

$$Y = Rand(feat) \tag{7.12}$$

$$input_{feat} = |input - input_{emp}| \tag{7.13}$$

Here N represents the parameter for movements and 'input' is the best optimal with $input_{emp}$ representing position of other the emperor penguins.

$$E_{emp} = |b(w) \odot input(i) - Y \odot input_{emp}(i)| \tag{7.14}$$

$$B(W) = \sqrt{((b \odot exp^{-d/y} - exp^{-d})^2)} \tag{7.15}$$

Equations (7.14) and (7.15) is inferred to assess the distance among the optimum fittest search agent E_{emp} and emperor penguin.

7.2 Classification and Regression based on Hybrid DL methodology

HD classification and regression is done by applying DL approaches namely LSTM, RNN, SSAE and deep progressive attention LSTM and RNN. By employing framingham equation, the regression is assessed in terms of vascular age, cardiac risk and cardiac index.

LSTM

LSTM is a progressive NN that is an extension of RNN and is used as a building component for RNN layers. The LSTM distributes weights to data, which aids RNN in either allowing fresh input in, forgetting it, or giving it relevance to impact the result. It has the capability to remember the inputs for extensive time period since the LSTM stores the data in memory, comparable to a computer's memory. From its memory, it writes, reads and erases data. It is modelled as a gated cell, with the cell deciding whether to delete or save data based on the relevance of the information provided as input. This relevance is assigned using weights that are learned using an algorithm.

RNN

RNN is sort of NN that is comprised of multiple layer of network in loop form that follows progressive data. RNN is best known for its memory, taking data from preceding inputs, to impact the output and input. This loop allows the data to extend, and every network in the loop take data and input from the previous network. It also executes a particular operation and therefore provides outcome through data transferring to successive network. Also it shares the variables with similar weight at every layers of the network. To facilitate reinforcement learning, the weights are accustomed using back-propagation and gradient descent.

Stacked Sparse Auto-encoder

In SSAE network, the hidden layer input is taken from the preceding SAE to the successive SAE. After training numerous SAE the decoded layer is erased, while the learned variables are stocked in hidden layers. And then it is intertwined to softmax classifier that tends to execute data classification. Features fine-tuning is done employing back-propagation method of the complete network, in terms of training and label instances.

Deep Progressive Attention (DPA) RNN-LSTM

In this algorithm, RNN approach acquires the vector as the input and multilayer LSTM creates vector as the output. RNN retains short term memory but when integrated with LSTM it retains long-term memory. Usually, LSTM model is well-known as repeating module has 4 NN layers correlating with every one. This module comprises of 3 gate activation functions namely sig1, sig2 and sig3. Concatenation function is represented as $\bigotimes$. The network selects the capacity of previous data to the flow.

$$cfor_{gate} = sig_1(W_{cfor} \otimes [O_{t-1}, data_{gate}] + bia_{cfor} \tag{7.16}$$

$$Ip_{gate} = sig_2(W_{Ip} \otimes [O_{t-1}, data_{gate}] + bia_{Ip} \tag{7.17}$$

$$out_{gate} = tanh(W_{out} \otimes [O_{t-1}, data_{gate}] + bia_{out} \tag{7.18}$$

Table 7.1: Steps for prediction model

step 1. set lstm units, op units, ip units, and optimizer to describe LSTM network (L)
step 2.standardise the dataset (sf_i) to values between 0 to 1 by $sf_{norm} = (sf - mini(sf))/(maxi(sf) - min(sf))(7.20)$
step 3. choose training size of window (sf_size) and then organize (sf_i) consequently
step 4. for n epochs and batch size do
step 5. Train the $(Propo_{model})$
step 6. end for
step 7. employ $Propo_{model}$ to run Predictions
step 8. Compute the loss function by $funobj = mini(\frac{1}{k}\sum_{i=1}^{k}(ytrue_i - ypred_i)^2)(7.21)$

$$out_{gate} = cfor_{gate} \odot out_{t1} + Ip_t \times out^1_{gate-1} \tag{7.19}$$

By employing 2 network layers, the cell is stowed using new data, where sig2 selects the Ip_{gate}, value that need to be advanced. A new candidate figure out^1_{gate} its vector is developed and the integration of (7.18) and (7.19) is then added in this state and hence the state of the cell is upgraded by the equation (7.19). Predictive unit termed resource manager is then fed in the model. Workload prediction is done by employing the steps below. The predicted and actual values of the time instances are k and i is the data samples number. For classification and regression of HD, the model having minimum mean square error has been employed. Regression for HD is predicted on the basis of vascular age, cardiac index assessment and cardiac score risk. By employing framingham risk score matrix the cardiac risk score is computed and the results showing equal to greater than 20% represents high risk of HD, 6-20% is medium and less than 6% represent low risk of HD. Also for calculating the heart vascular age framingham risk score matrix is employed. Additionally, assessment of cardiac index is computed and if result is 2.5 to $4l/min/m^2$, then it is normal. It is calculated in terms of heart rate, pulse pressure, weight and height of the individual. The prediction model steps are shown in table 7.1.

7.3 Results and Discussion

The outcomes of implementing the proposed model are deliberated in this section with the performance metrics, internal comparison, and experimental results.

7.3.1 Performance Metrics

To estimate the performance of the proposed system, some major parameters are used: precision, accuracy, f1 score, recall, sensitivity and specificity (discussed in chapter II and V).

7.3.2 Data Collection

Sensor dataset data collection is done that behaves like a terminal node to collect signals and IoT communication. Numerous real-time data, concerning patients are gathered via IoT wearable sensors. ECG sensor, SPO2 sensor, BP sensor and glucose sensor are the sensors. Systolic BP is collected using BP sensor, pulse rate and diastolic BP. RR intervals and QT is found using ECG sensors. SPO2 computes the value of oxygen saturation. Additionally, using FGA 10, sensor aids in measuring haemoglobin, high cholesterol, low cholesterol, total cholesterol and random sugar.

7.3.3 Pre-Processing of data

For data pre-processing, the collected data is loaded and fed. Pre-processing of data is the important portion of ML sequence that executes analysis of data and enhances the precision and speed of the scheme. To manage the over-fitting problem, noise is detached also it aids in minimizing the complexity of data and gives clear insights of ailment symptoms. Some unwanted and unrelated features produce noise and vagueness over description that tends to minimize the enactment and efficiency of the presented system. Thus the pre-processing method

Table 7.2: Dataset employed in classification of HD

Attributes	Description	Categorical or Numerical
ID	Integer value	Numerical
Gender	Male/Female	Categorical
Age	Years	Numerical
Height	Centimetres	Numerical
Weight	Kilograms	Numerical
BP treatment (Medicine)	Yes/No	Categorical
Smoker	Yes/No	Categorical
Diabetic	Yes/No	Categorical
Pre-existing HD	Yes/No	Categorical

is inculpated to manage drawbacks and to produce efficient results.

7.3.4 Label Encoding

Label encoding is performed to the data that is pre-processed. The category characteristics are label encoded since numerical data can be handled only by supervised learning algorithms. Converting the gender from male to female, which is encoded as 0 and 1, is an example of label encoding.

For all categorical variables namely name, age, BP treatment, diabetic, smoker and preceding HD, yes (presence) is programmed as 1 and no (absence) is programmed as 0. While, the variables for computing regression are heart vascular age, cardiac index and cardiac score risk.

7.3.5 Data Split

Once labelling is done for the attributes targeted, the data is assembled for the process of data splitting. And the data present here is converted to training and then to testing phase in that for training 80% data is used and for testing 20% data is used. Data testing is employed for assessing the model execution and superior model is chosen on the basis of training as well as testing data.

7.3.6 Feature Scaling

The trained data feature scaling is executed on the trained data that is encoded using the label. This method is employed for monitoring the self-determining attributes of dataset within particular range. Range of self-determining variables is reduced via feature scaling, allowing for comparison of these variables on common platform. The dataset contains data that might not all have the same scale. This could result in inaccurate results in such circumstances. To avoid these problems, feature scaling should be used. This is essential since the suggested system, which calls for data scaling, employs gradient descent as an optimization technique. Standardization and normalization are the main steps in feature scaling. In a scaling strategy known as standardization, values are clustered around the mean and have a unit standard deviation. Features here are rescaled to make sure that the mean and standard deviation are both 1. While normalization is a method that scales and shifts data projecting a range among 1 and 0.

7.3.7 Feature Selection

In data modelling, feature selection is a vital and beneficial stage that assist to remove unrelated and useless data. Also it minimizes the amount of input attributes in the model to reduce cost and enhance the enactment of the presented model. This is done to display the statistical relationship among the target and input variable, it is done by selecting input variables having robust relationship. Even though the employed statistical assessment are impacted by the data type in output and input variables. Here, meta- heuristic processes are employed for learning and also for the method of forecasting accurate solution. GA, PSO, GSA, PO and ABO are the optimisation techniques.

7.3.8 Confusion Matrix of HD Classification

The performance of the recommended model is represented by the confusion matrix which depends upon the accurate and detailed prediction. Hence, it assists in conceptualizing the results, by identifying the defeats which occurred during the execution of classification. Subsequently the figure 7.2 expresses the HD confusion matrix of classification which depends upon GSA with SSEA. It represents the least possible errors. This results in the involvement of better classification of HD.

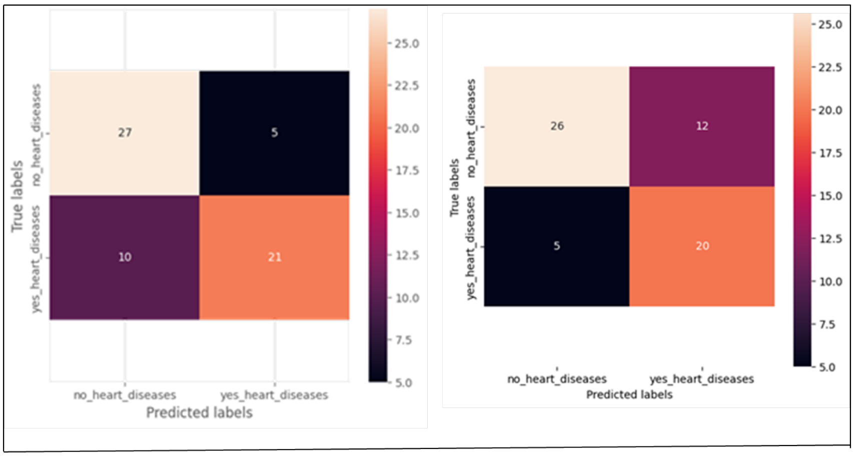

Figure 7.2: POA and GS Confusion Matrix with SSAE

Since, the figure 7.2 inferred a finer accuractness of HD classification which can be predicted by POA, GSA with SSAE method. Hence, the POA-SSAE reads the presence of 5 misclassification of HD combined with 10 absence of HD. While GSA-SSAE misclassifies with presence of 12 misclassified HD and 5 no HD.

7.3.9 Experimental Results

As in consequence, the regression of heart oriented IoT removed the sensor dataset which is considered by using five optimization method and the DL methods where the results are denoted in the form of pictorial representation. Hence, the figure 7.3 represents the prediction of cardiac index using ABO.

Since the figure 7.3 has considered the prediction using regression method which depends

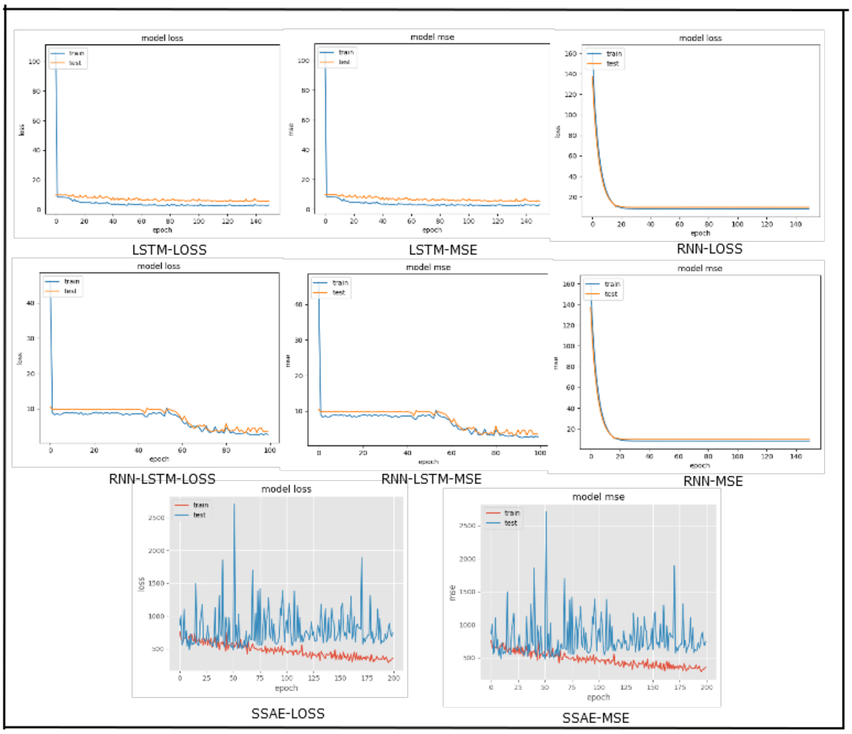

Figure 7.3: Prediction of cardiac index using ABO

upon the IoT repossessed a sensor dataset for $LSTM\_LOSS$, $LSTM\_MSE$, $RNN-LSTM\_LOSS$ and $RNNLSTM\_MSE$ which made additional relationship among the trained and test dataset.

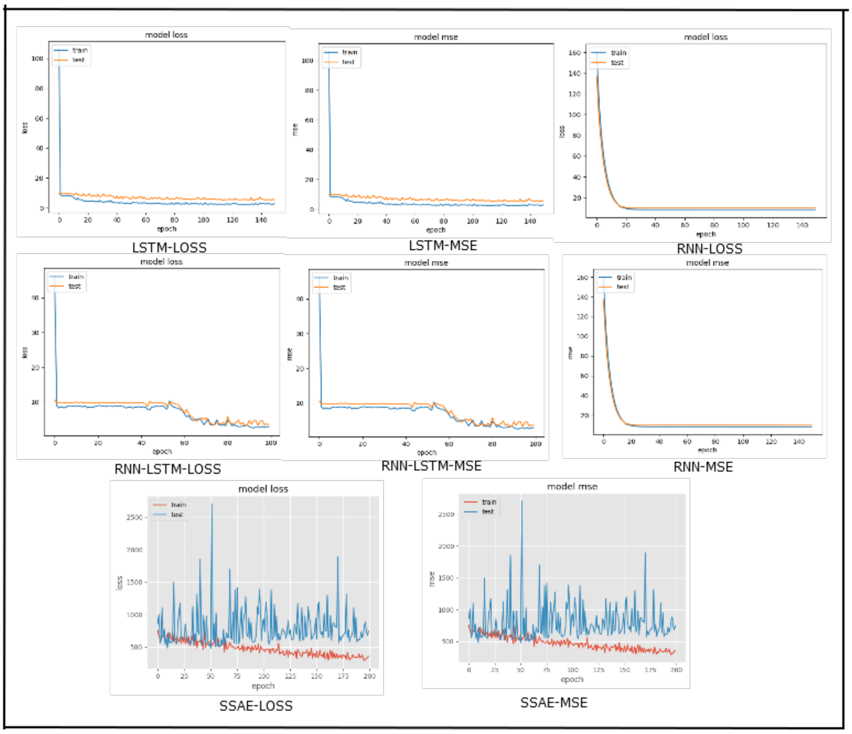

Figure 7.4: Prediction of cardiac index using GA

Subsequently the figure 7.4, shows prediction depends on IoT using regression method reclaimed sensor dataset of $SSAE\_LOSS$, $LSTM\_MSE$ as well as $SSAE\_MSE$ which made further connection among the trained and test dataset.

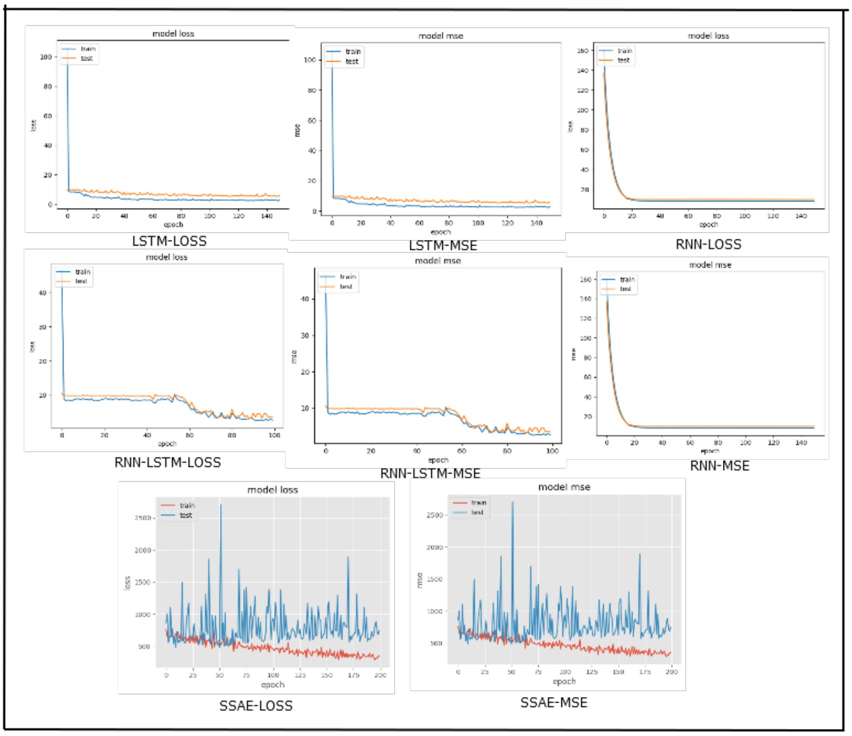

Figure 7.5: Prediction of cardiac index using GSA

As of the figure 7.5 IoT prediction is retrieved using the regression method which uses the sensor dataset of $SSAE\_LOSS$, $SSAE\_MSE$ which creates more link among the trained and test dataset

Following the figure 7.6, the prediction based on IoT retrieved sensor dataset the $RNN-LSTM\_LOSS$ and $RNN-LSTM\_MSE$ produced more correlation between the trained data and test dataset using regression technique.

By the figure 7.7 represents that the prediction based on IoT extracted sensor dataset of RNN_LOSS, RNN_MSE, $RNN-LSTM\_LOSS$ along with $RNN-LSTM\_MSE$ which composes additional connection among the test dataset and trained data.

Subsequent to the figure 7.8 it is evident that the IoT prediction found sensor dataset of RNN_LOSS, RNN_MSE processed additional correlation among the test dataset and trained data using regression technique.

From the figure 7.9 it is conceivable that the prediction depends on IoT reacquires sensible

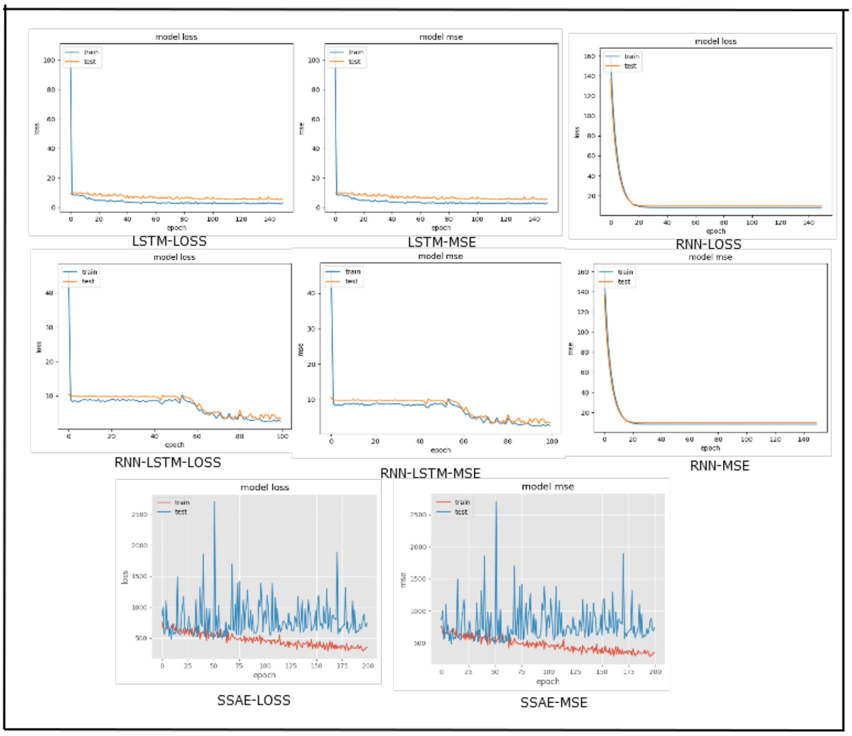

Figure 7.6: Prediction of cardiac index using POA

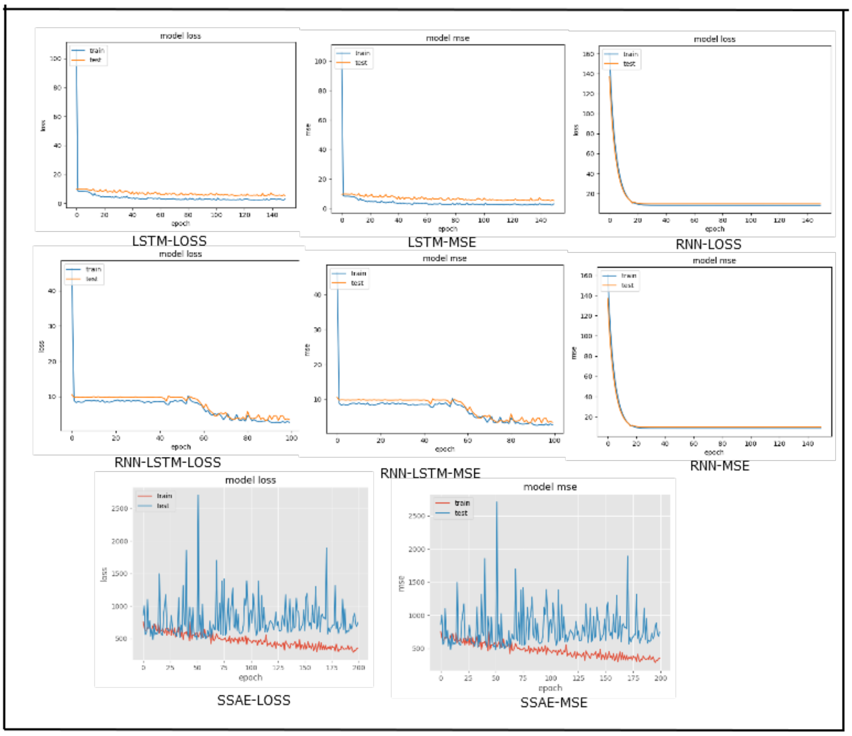

Figure 7.7: Prediction of cardiac index using PSO

dataset of $RNN - LSTM\_LOSS$, $RNN - LSTM_{M}SE$ that constructed with extra concurrence among the data which are trained and test dataset in regression method.

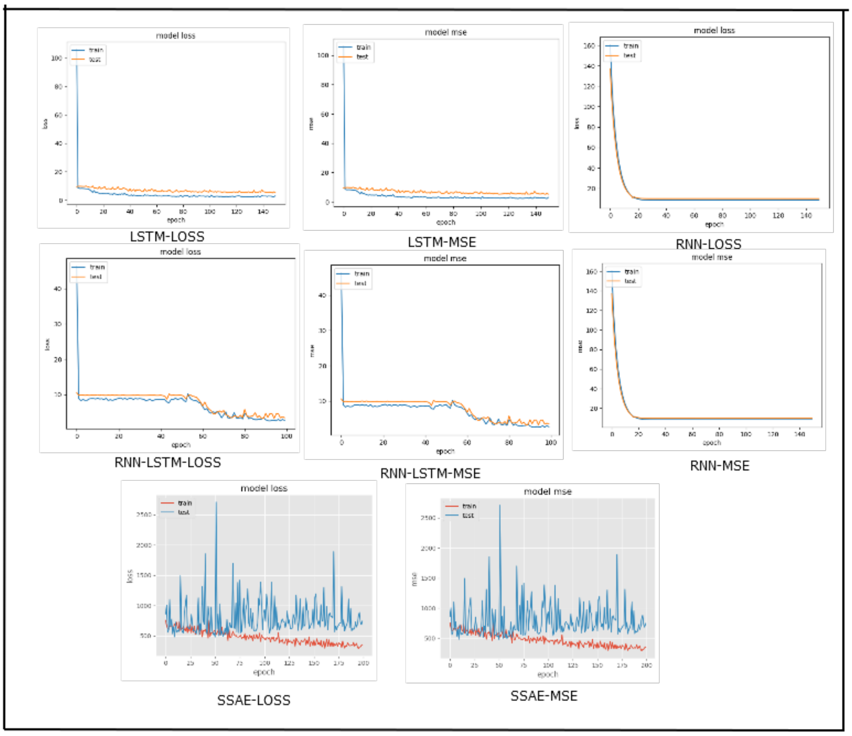

Figure 7.8: Calculation of Heart Age using ABO

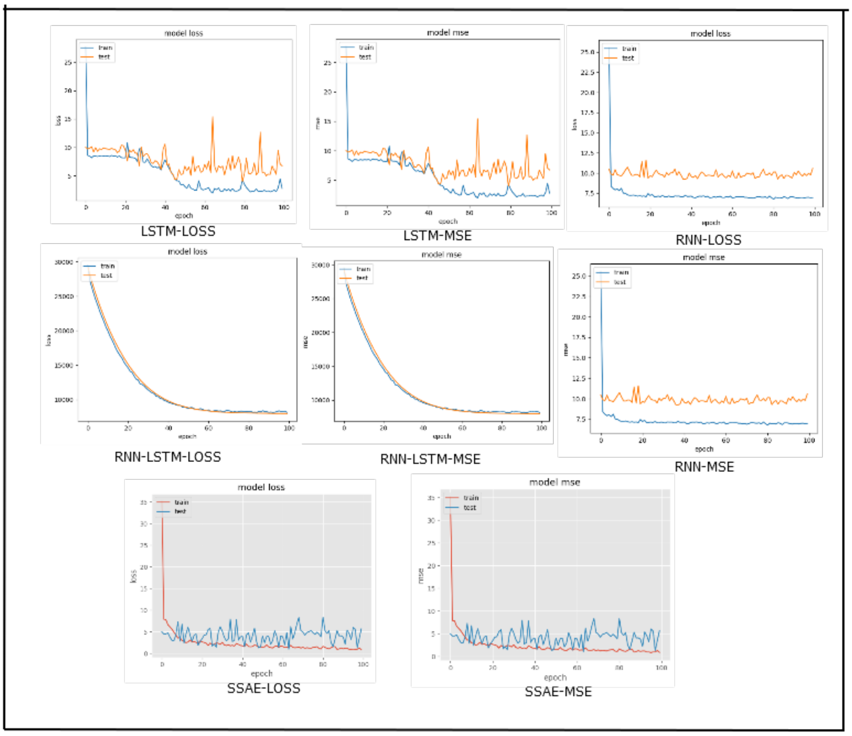

Figure 7.9: Prediction of Heart Age using GA

Through the figure 7.10 it is established, where the prediction of IoT resolved the dataset of $LSTM\_LOSS$, LSTM-MSE, $RNN-LSTM\_LOSS$, $RNN-LSTM\_MSE$ which generated huge

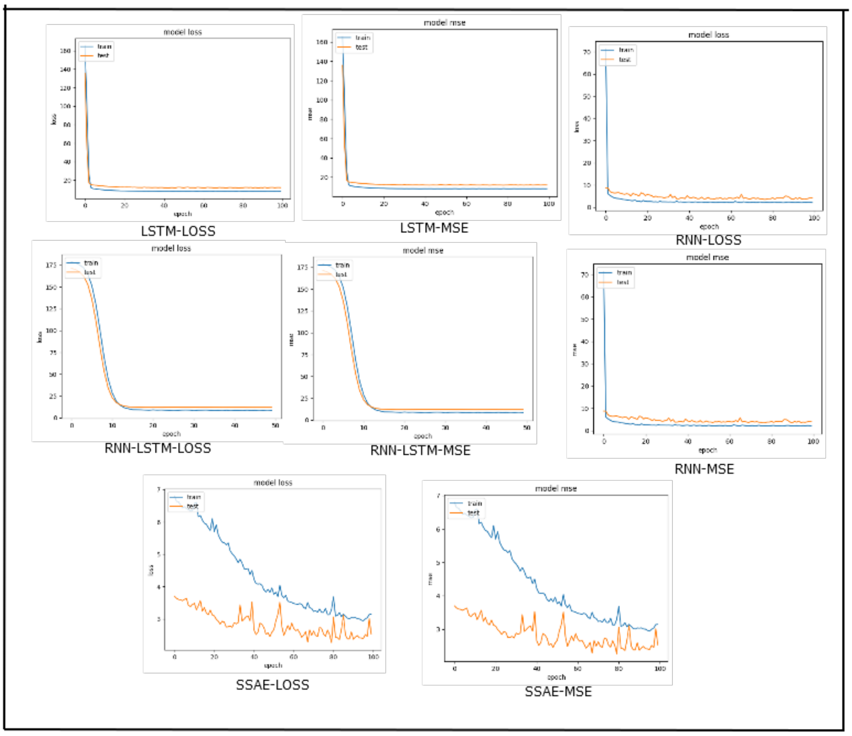

Figure 7.10: Prediction of Heart Age using GSA

number of connection between the test dataset and trained data with regression.

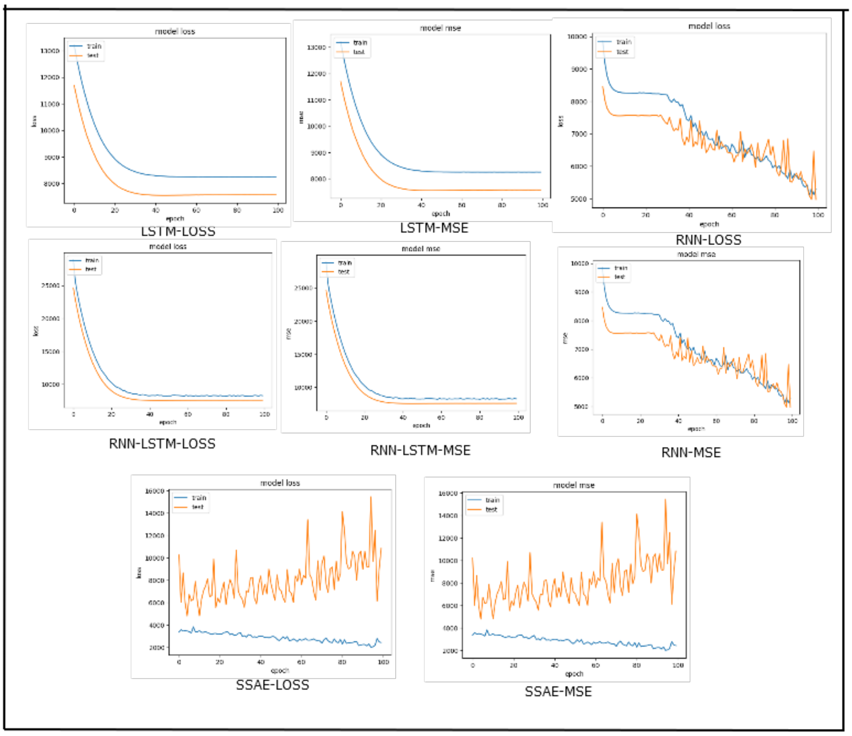

Figure 7.11: Prediction of Heart Age using POA

Subsequent to the figure 7.11 it is evident that the IoT prediction using regression which has taken sensible dataset for $RNN - LSTM\_LOSS$, $RNN - LSTM\_MSE$ can induced more interrelationship among the data which are trained along with test dataset.

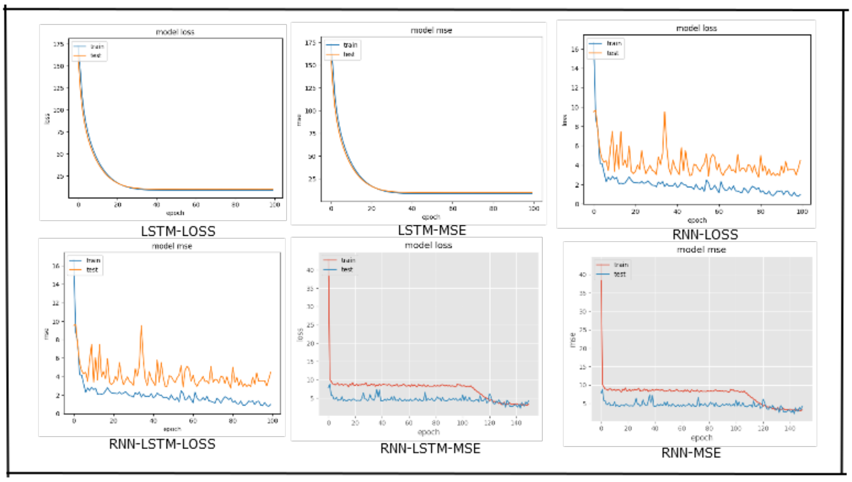

Figure 7.12: Prediction of Heart Age using PSO

By the figure 7.12 it is going to predict the trained data and test data using IoT which extracted the sensitive dataset of $LSTM\_MSE$, $LSTM\_LOSS$ which shows more connection among trained and test dataset in regression.

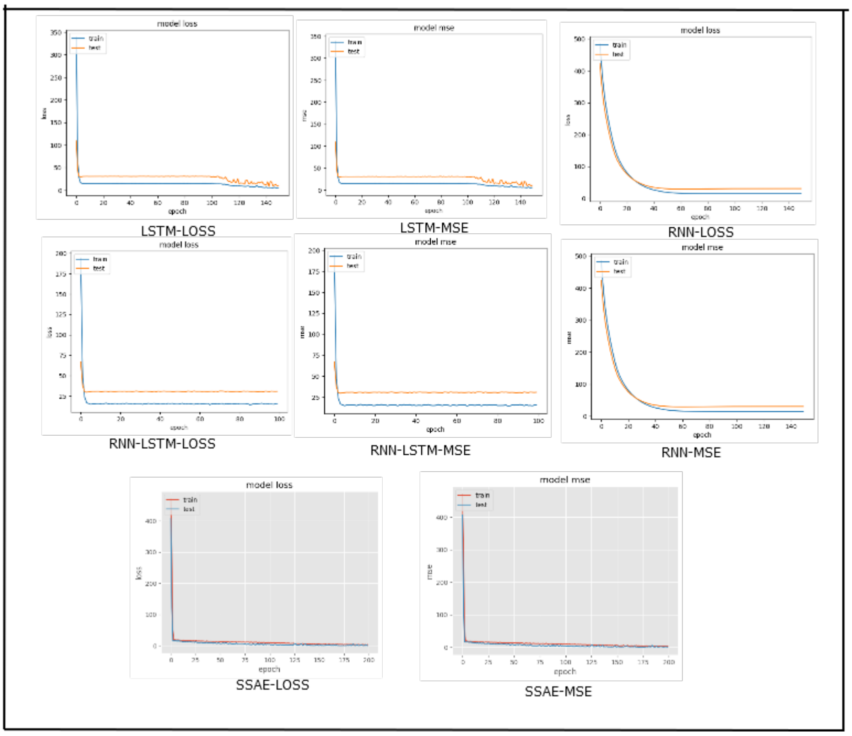

Figure 7.13: Prediction of risk score using ABO

Since the figure 7.13 it is convincible that the IoT prediction regained sensitize dataset the

$SSAE\_LOSS$, $SSAE\_MSE$ which creates more relationship among the trained data, test dataset in regression.

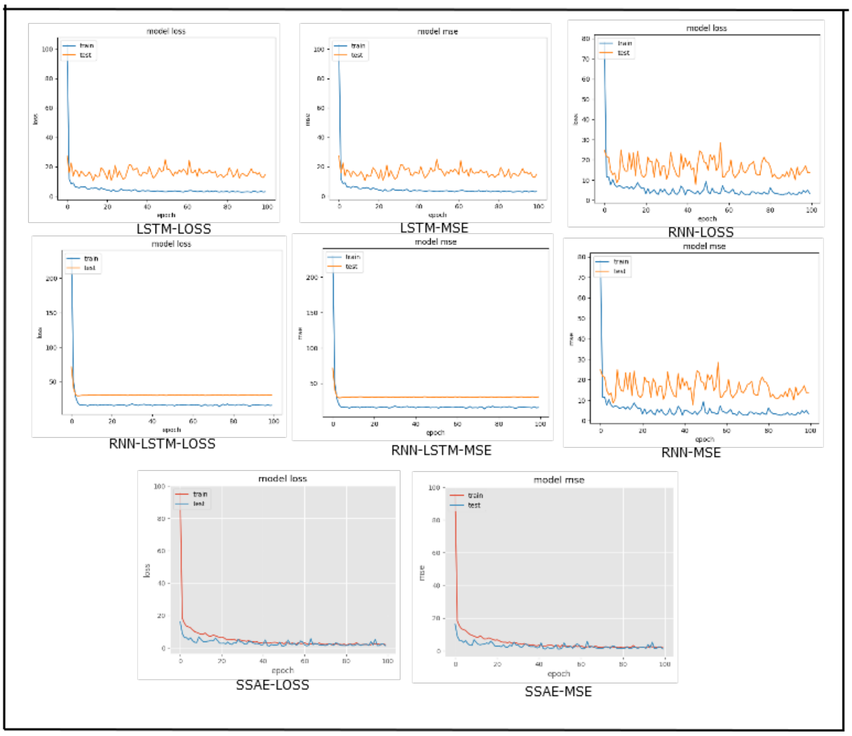

Figure 7.14: Risk score prediction using GA

Following the figure 7.14, in the regression, IoT prediction resolved sensitive dataset which are $SSAE\_LOSS$, $SSAE\_MSE$ have produced extra relationship between data and the dataset.

From the figure 7.15 it is found that the regression prediction based on IoT retrieved sensor dataset the $RNN-LSTM\_LOSS$, $RNN-LSTM\_MSE$ produced more correlation between the trained data and test dataset.

Since the figure 7.16 it is evident that the IoT prediction reacquired sensible dataset of $SSAE\_LOSS$, $SSAE\_MSE$ and the connection was made among the test and trained dataset.

Following the figure 7.17 it is found that the regression prediction based on IoT retrieved sensor dataset the $LSTM\_LOSS$, $LSTM\_MSE$, $RNN-LSTM\_LOSS$, $RNN-LSTM\_MSE$ produced more correlation between the trained data and test dataset. From the figures, it is interpreted that the SSAE model outperformed other models by producing more correlation between train and test dataset.

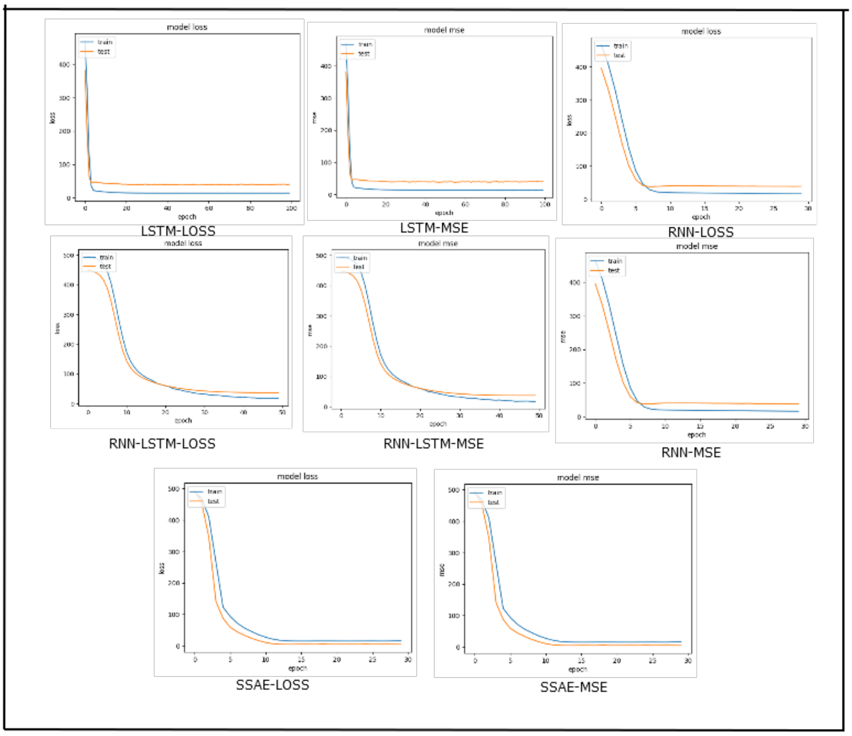

Figure 7.15: Risk score prediction using GSA

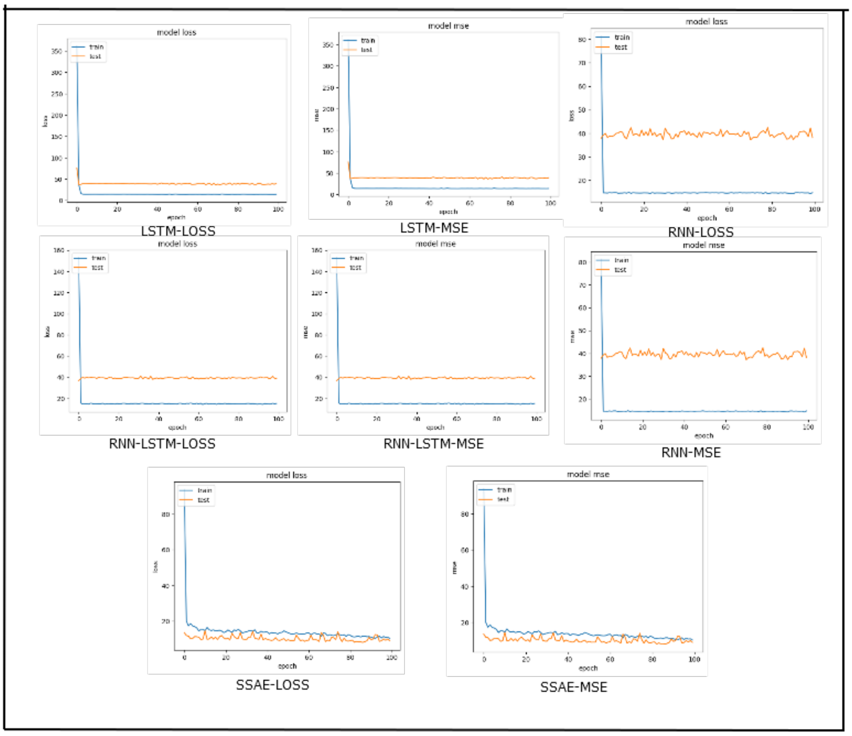

Figure 7.16: Risk score prediction using POA

7.4 Comparison among the models

As, HD classification and regression is achieved by evaluating the optimisation methods combined hybrid DL approaches. Hence, the consequences created by ABO in HD classification

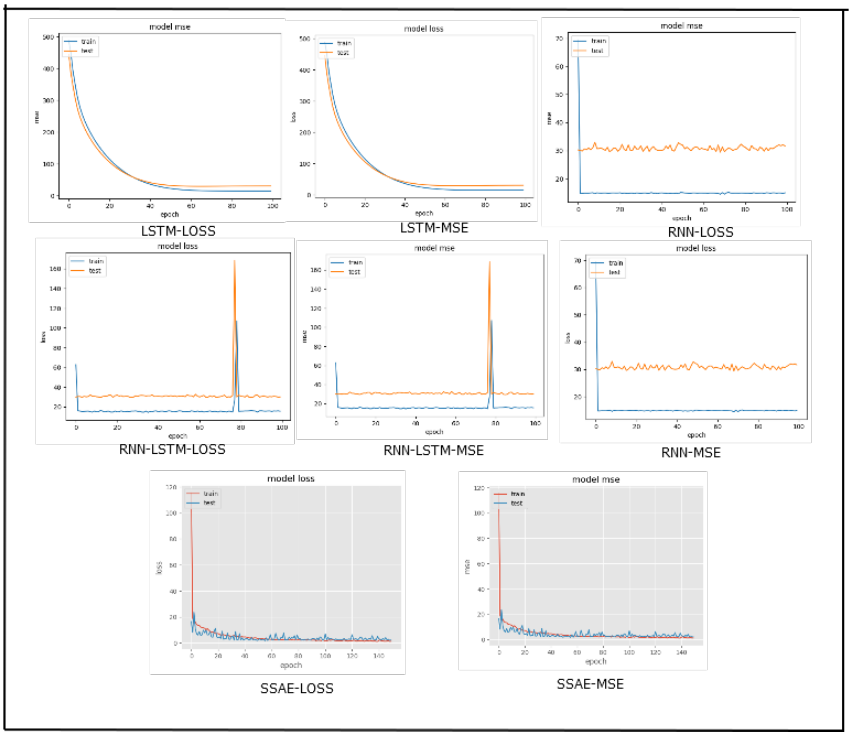

Figure 7.17: Risk score estimation using PSO

developed on RNN, LSTM, SSAE and DPA-RNN+LSTM is stated in table 7.3.

Table 7.3: Heart Disease Classification Attributes - ABO

Classifier	Features	Acc	Preci	Sensi	Speci	F1	Recall
RNN	0111110011101100101000000	71.28	72	80.49	64.15	71	72
LSTM	001001001100111010101101	66	67	41.46	84.90	63	63
SSAE	110011101110000001011010	66	66	68.29	64.15	66	66
DPA-RNN+ LSTM	110010000001001111101101	72	74	85.37	62.26	72	74

Following the table 7.3, ABO involves an accuracy of 71.28%, with RNN, LSTM with 66%, SSAE with 66% and DPA-RNN+LSTM is 72%. Additionally, the performance metrics such as sensitivity, precision, recall, specificity and F1 score of all models are computed and the DPA model shows better results compared to other models.

In the table 7.4, the GA produces an accuracy with RNN is 63.16%, with LSTM as 65.96%, SSAE as 68.09% and DPA as 71.28%. Hence, the performance measures are evaluated with the better results and so the recommended model displays the bettered results.

Table 7.4: Heart Disease Classification Attributes - GA

Classifier	Features	Acc	Preci	Sensi	Speci	F1	Recall
RNN	010001011001110111111110	63.16	63.16	58.54	73.58	60.76	58.54
LSTM	011101010111111111110111	65.96	60.47	63.41	67.92	61.90	63.41
SSAE	111111100010111011111111	68.09	65.71	56.1	77.36	60.53	56.1
DPA-RNN+ LSTM	011111111110101001111111	71.28	71.88	56.1	83.02	63.01	56.1

Table 7.5: Heart Disease Classification Attributes - GSA

Classifier	Features	Acc	Preci	Sensi	Speci	F1	Recall
SSAE	1 3 5 9 11 12 13 14 15 17 22	73.02	62.5	80	68.42	70.18	80
RNN	1 2 3 4 5 10 13 15 16 21 22	60.32	50	32	78.95	39.02	32
LSTM	1 6 8 12 13 14 15 16 18 19 21 22	60.317	50	92	39.47	64.79	92
DPA-RNN+ LSTM	1 3 4 5 9 13 14 17	63.49	53.57	60	65.79	56.60	60

Through the table 7.5, the accuracy results in GSA together with RNN is 60.32%, LSTM with 60.32%, SSAE with 73.02% and DPA with 63.49%. Hence, the performance metrics of overall models are estimated. So, GSA-SSAE illustrates better performance than other models.

Table 7.6: Heart Disease Classification Attributes - POA

Classifier	Features	Acc	Preci	Sensitivity	Specificity	F1	Recall
SSAE	16	76.19	80.77	67.74	84.38	73.68	67.74
RNN	16	75.53	67.31	85.37	67.92	75.27	85.37
LSTM	16	73.40	69.05	70.73	75.47	69.88	70.73
DPA RNN+LSTM	16	68.25	56.76	84	57.89	67.74	84

Succeeding the table 7.6, POA-RNN involves the accuracy of 75.53%, LSTM is 73.40%, SSAE is 76.19% and recommended DPA is 68.25%. Additionally, performance measures of entire models are computed and the outcomes illustrates better results, thus POA combined with SSAE transcends the further models.

Table 7.7: Heart Disease Classification Attributes - PSO

Classifier	Features	Acc	Preci	Speci	F1	Sensi	Recall
SSAE	100111100010100001110010	68.09	61.22	64.15	66.67	73.17	73.17
RNN	000011000000101011001000	48.94	45.68	16.98	60.66	90.24	90.24
LSTM	000001011100000011001000	70.21	70.97	83.02	61.11	53.66	53.66
DPA-RNN+LSTM	111011010111001000011010	70.21	80.95	92.45	54.84	41.46	41.46

In the table 7.7, the accurateness contributed by PSO along with RNN results in 48.94%, and LSTM with 70.21%, SSAE with 68.09% and suggested DPA with 70.21%. Additionally, the recall, precision, f1 score, specificity, and sensitivity of entire models are considered and the outcomes display better outcomes. Hence, POA combined with LSTM and POA-SSAE is superior to the further models.

Table 7.8: Cardiac Index Prediction - ABO

Classifier	Features	MAE	MSE	RMSE
SSAE	000000000000100000100000	21.73610939	891.6980796	29.8613141
RNN	011110111000010101111110	159.6595745	33430.55319	182.8402395
LSTM	100001011000100000000000	159.6595745	33430.55319	182.8402395
DPA RNN+LSTM	000000000000000000000001	159.6595745	33430.55319	182.8402395

From the table 7.8 concluded that the SSAE combined with ABO involves declined mean error rate however the further approaches includes the values of MAE as 21.73, MSE as 891.69 and RMSE as 182.84.

Following the table 7.9, it is calculated that the SSAE including GA results reduced mean error rate while contrasted with further approaches with 22.38 as MAE, 715.69 as MSE and 26.75 as RMSE values.

From the table 7.10, it is concluded that the SSAE among GSA causes lesser mean error rate while measured with further techniques with MAE (65.13), MSE (5806.37) and RMSE

Table 7.9: Cardiac Index Prediction - GA

Classifier	Features	MAE	MSE	RMSE
SSAE	000000000000001000100001	22.38098092	715.6995518	26.75256159
RNN	101110101001000100111001	159.6595745	33430.55319	182.8402395
LSTM	001111110110000001101110	159.6595745	33430.55319	182.8402395
DPA-RNN+LSTM	101111001011100000101101	159.6595745	33430.55319	182.8402395

Table 7.10: Cardiac Index Prediction - GSA

Classifier	Features	MAE	MSE	RMSE
RNN	20	151.8730159	30620.19048	174.986258
LSTM	23	151.8730159	30620.19048	174.986258
SSAE	8	65.13182576	5806.375301	76.19957546
DPA-RNN+LSTM	21	151.8730159	30620.19048	174.986258

(76.199).

Table 7.11: Cardiac Index Prediction - POA

Classifier	RMSE	MAE	MSE	Features
SSAE	29.8613141	21.73610939	891.6980796	0 2 5 9 13 20 21 22
RNN	174.986258	151.8730159	30620.19048	1 3 8 12 23
LSTM	174.986258	151.8730159	30620.19048	0 2 5 6 8 10 15
DPA-RNN+LSTM	174.986258	151.8730159	30620.19048	4 5 10 11 17 20 21

From the table 7.11, it is inferred that the SSAE with POA produces less mean error rate when compared with other methods with MAE of 21.74, MSE of 891.69 and RMSE of 29.81.

Following the table 7.12, it is induced that the SSAE added with PSO involves reserve mean error rate conversely measured up to further techniques in computing the MAE value as 22.53, MSE value as 1629.50 and RMSE value as 40.36.

In the table 7.13, explained that the RNN, LSTM and recommended model with ABO leads to reduction in mean error rate even if it is achieved by SSAE process against 12.85 as

Table 7.12: Cardiac Index Prediction - PSO

Classifier	Features	MAE	MSE	RMSE
SSAE	0111100111111010101011111	22.53844955	1629.503245	40.36710598
RNN	0010010100000111101010000	159.6595745	33430.55319	182.8402395
LSTM	0001001000000000000100100	159.6595745	33430.55319	182.8402395
DPA-RNN+LSTM	1011101101001001110010100	159.6595745	33430.55319	182.8402395

Table 7.13: Heart Age Prediction - ABO

Classifier	Features	MAE	MSE	RMSE
LSTM	1000000110110010000011100	12.85106383	174.9148936	13.22553945
SSAE	1111111111111110111111111	21.73610939	891.6980796	29.8613141
RNN	1111101100001010000110100	12.85106383	174.9148936	13.22553945
DPA-RNN+LSTM	1100010111101111101111100	12.85106383	174.9148936	13.22553945

MAE value, 174.91 as MSE value and 13.22 as RMSE value.

Table 7.14: Heart Age Prediction - GA

Classifier	Features	MSE	RMSE	MAE
LSTM	1010010110000011001100000	174.9148936	13.22553945	12.85106383
SSAE	0101001000000011000000100	3.605284087	1.898758565	0.967520468
RNN	0001000100000000100000000	33430.55319	182.8402395	159.6595745
DPA-RNN+LSTM	1011110010111000001011010	33430.55319	182.8402395	159.6595745

Subsequent to the table 7.14, determined that the SSAE coupled with GA contributes to lesser mean error rate whereas SSAE method provides the MAE value of 0.967, MSE value of 3.605 and RMSE value of 1.898.

From the table 7.15, implicative that the SSAE in conjunction with GSA brings out smaller mean error rate, by contrast observed that the SSAE process including MAE is 1.346, MSE is

Table 7.15: Heart Age Prediction - GSA

Classifier	RMSE	MAE	MSE	Features
RNN	13.12092718	12.66666667	172.1587302	1 2 9 13
LSTM	13.12092718	12.66666667	172.1587302	9
SSAE	2.241315139	1.346952552	5.023493551	0 4 11 15
DPA-RNN+LSTM	13.12092718	12.66666667	172.1587302	2

5.0234 and RMSE is 2.241. Further, the table 7.16, represents the outcomes made by POA in vascular age prediction depends on RNN, LSTM, SSAE and recommended DPA-RNN and LSTM.

Table 7.16: Heart Age Prediction - Penguin Optimization Algorithm

Classifier	RMSE	MAE	MSE	Features
SSAE	81.73751178	60.21714468	6681.020832	0 2 5 9 13 20 21 22
RNN	174.986258	151.8730159	30620.19048	1 3 8 12 23
LSTM	174.986258	151.8730159	30620.19048	0 2 5 6 8 10 15
DPA-RNN+LSTM	174.986258	151.8730159	30620.19048	4 5 10 11 17 20 21

In the table 7.16, it is expected that the SSAE beside POA involves insignificant mean error rate regardtheless competed with SSAE approach further to MAE value of 60.21, MSE value of 6681.02 and RMSE value of 81.73. Hence, the table 7.16, indicates the outcomes stimulates with PSO in vascular age prediction determined from RNN, LSTM, SSAE and suggested DPA-RNN and LSTM.

Table 7.17: Heart Age Prediction - PSO

Classifier	Features	MAE	MSE	RMSE
LSTM	010010100001000000000010	12.85106383	174.9148936	13.22553945
SSAE	010000000001000000000000	0.83869494	3.57895521	1.891812678
RNN	100000100101010011000000	12.85106383	33430.55319	13.22553945
DPA-RNN+LSTM	100000000000010010100100	159.6595745	33430.55319	182.8402395

From the table 7.17, it is inferred that the SSAE with PSO produces less mean error rate when compared with SSAE method with MAE of 0.838, MSE of 3.5789 and RMSE of 1.8918. The table 7.18, denotes the results produced by ABO in risk score prediction based on RNN, LSTM, SSAE and proposed deep progressive attention RNN and LSTM.

Table 7.18: Risk Score Prediction using African Buffalo Optimization

Classifier	Features	MAE	MSE	RMSE
LSTM	11110101011001110111001	20.82978723	463.2553191	21.52336682
SSAE	01110000000100110101010 0	1.74714334	13.65392241	3.695121434
RNN	01001010100001011001110 0	20.82978723	463.2553191	21.52336682
DPA-RNN+LSTM	11101110111111011111111 1	20.82978723	463.2553191	21.52336682

Within the table 7.18, it is induced that the SSAE together with ABO processes fewer mean error rate which is directly measured with SSAE approach among the value of MAE is 1.747, the value of MSE is 13.653 and the value of RMSE is 3.695. Further, the table 7.19, signifies the consequences constructed during GA in risk score prediction developed on RNN, LSTM, SSAE and recommended DPA- RNN combined with LSTM.

Table 7.19: Risk Score Prediction - GA

Classifier	Features	MAE	MSE	RMSE
SSAE	100000000011000000100000	2.118983715	23.86134546	4.884807618
RNN	101010100000001010111000	20.82978723	463.2553191	21.52336682
LSTM	110100000000000001010000	20.82978723	463.25531916	21.52336682
DPA-RNN+LSTM	111111101110011100110 11	20.82978723	463.2553191	21.52336682

In table 7.19, SSAE added to GA ends in inferior mean error rate however correlated with SSAE technique including MAE(2.1189), MSE (23.861) and RMSE (4.884). Moreover the table 7.20, represents the resultants triggered over GSA in risk score prediction derived from RNN, LSTM, SSAE and suggested DPA- RNN in addition with LSTM.

Since the table 7.20, has implied that the SSAE and GSA culminates in declined mean

Table 7.20: Risk Score Prediction - GSA

Classifier	Features	MAE	MSE	RMSE
LSTM	10	20.31746032	449.49206352	21.20122788
SSAE	4	3.15146755	40.20119416	6.340441165
RNN	3	20.31746032	449.4920635	21.20122788
DPA-RNN+LSTM	2	20.31746032	449.4920635	21.20122788

error rate as analysed with SSAE technique including MAE value of 3.1514, MSE value of 40.201 and RMSE value of 6.340. Hence, the table 7.21, represents the outcomes composed through POA in risk score prediction which depends upon RNN, LSTM, SSAE and Suggested DPA-RNN and LSTM.

Table 7.21: Risk Score Prediction- POA

Classifier	Features	MAE	MSE	RMSE
SSAE	5 6 14 15 19 23	3.309573885	40.41550715	6.357319179
LSTM	3 5 7 17 18 19 23	20.31746032	449.4920635	21.20122788
RNN	0 2 4 13 15 17 20	20.31746032	449.4920635	21.20122788
DPA-RNN+LSTM	5 9 11 13 18 20 23	20.31746032	449.4920635	21.20122788

Subsequent to the table 7.21, it is implicit that the SSAE accompanied by POA compose reduced mean error rate while contrasted with SSAE method including 3.3095 of MAE value, 40.415 of MSE value and 6.3573 of RMSE value. Hence the Table 7.22, represents the outcomes developed through PSO on risk score prediction depends on RNN, LSTM, SSAE and proposed Deep Progressive Attention (DPA) RNN and LSTM.

Since the Table 7.22 has concluded that the SSAE with PSO results minimum mean error rate while compared with SSAE technique, where MAE is 2.046, MSE is 19.298 and RMSE is 4.39.Hence the resultant shows the SSAE model exceeds other methods which produce minimum error rate and supports in improvement of system performance.

Table 7.22: Risk Score Prediction - PSO

Classifier	Features	RMSE	MAE	MSE
LSTM	1000000000011000000010100	21.52336682	20.82978723	463.2553191
SSAE	0100000111010000010001000	4.392986359	2.046273226	19.29832915
RNN	1101100101000010010111101	21.52336682	20.82978723	463.2553191
DPA-RNN+LSTM	0011000000100001110001000	21.52336682	172.1587302	463.2553191

7.5 Deployment of Web Model

As a result of training, testing process, validation of accuracy and minimum error rate of models are utilized in training and so the model with greater accuracy can produce prediction. Hence, the model which is trained is reserved and fetched into the net. The flask web application structure is engaged to execute and construct web surfers to fetch HD attributes of the persons. The prediction displays whether the patient is affected with HD or not. Correspondingly, the regression is predicted depends upon the cardiac risk score, vascular age along with cardiac index measurements. Hence, the figure 7.18 denotes the web pages which displays the presence of HD or not and the figure 7.19 represents the whether the cardiac risk is high or low. Correspondingly, the figure 7.20 illustrates the vascular age and figure 7.21 shows the cardiac index value is in normal or abnormal condition based upon the chosen attributes.

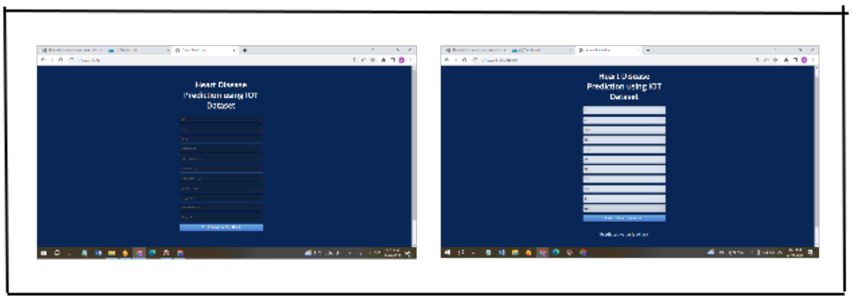

Figure 7.18: Heart Disease Prediction

Hence, the user can distinguish the information of heart based upon vascular age, cardiac risk score and cardiac index. Further, HD prediction is recognised by giving the input attributes like patient Id, age, etc. Hence on clicking, the model predicts the consistent values, and the

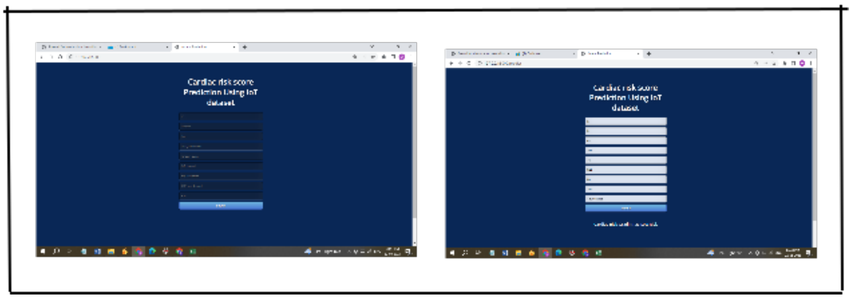

Figure 7.19: Cardiac Risk score Calculation

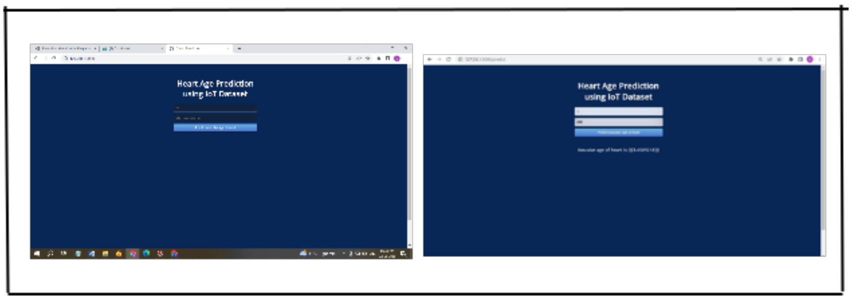

Figure 7.20: Vascular Age Assessment

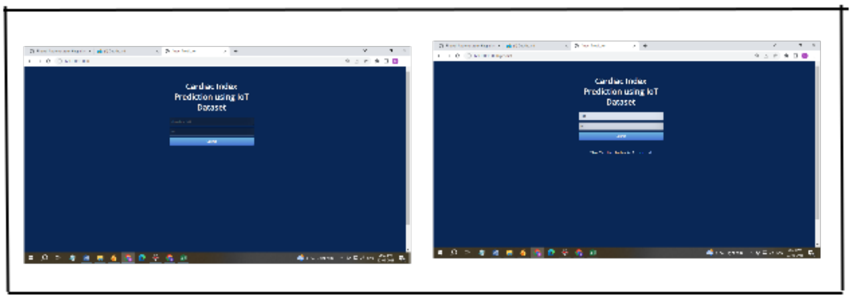

Figure 7.21: Cardiac Index Estimation

outcome will be displayed based upon the entered features.

7.6 Summary

In this chapter, the recommended model utilized five dissimilar optimisation methods such as GA, GSA, ABO, PSO and POA and DL approaches like RNN, LSTM, DPA-RNN and LSTM, SSAE for HD classification of and regression. Hence, the regression is predicted by calculating the parameters like vascular age, cardiac risk score and cardiac index values by utilizing framingham equation. Further, based upon these issues, the person is recognised to be normal or abnormal. However, the model with reduced mean error rate is predicted and utilized for HD classification and regression. Hence, the efficacy of the recommended system is computed based upon performance metrics. Further, the precision measured through the SSAE model combined with POA is 76.19% for HD classification. While the regression is performed for heart healthiness position and is established to measure lesser error rate with cardiac index value of SSAE along with POA is 21.73, value of vascular age of SSAE-PSO is 0.83 and SSAE combined with ABO cardiac risk score value is 1.74. Hence the prediction inferred that, the SSAE model shows better results when compared to the other models in connection with classification and regression. Following the outcomes acquired, it is established that the recommended model ends in finer results than conventional algorithms.

CHAPTER 8

SUMMARY AND CONCLUSIONS

8.1 Conclusion

Heart disease or an abnormality in heart is referred to be one of the deadly diseases of all age groups worldwide. Globally, heart diseases named as cardio vascular abnormality (CVD) results in 70% of death rates. Lack of medication and delayed intervention of these diseases results in increased ranges of complexity, leading to greater mortality rates. Age and Genetics are the causes and these factors cannot be controlled by humans. Heart diseases are caused because of several important factors such as negligence and mode of lifestyle. Both power and middle income nations are crucially affected by the rise of this chronic illness. It is essential for an accurate and a timely diagnosis to treat heart failure and in preventing complexity.

In most of the aspects, traditional method using the past medical history could lead to several drawbacks such as misidentification, mispredictions, overfitting of the model, computational complexity, local optima and premature convergence of the model. Furthermore, devices such as ECG, CT scans are essential for detecting CVD, which is expensive and infeasible in cases of middle income countries. The widespread use of health records leads to big data. In most of the cases, the clinical and the pathological data being a complex combination can lead to higher medical expenses that affect the level of medical care. Thus, data mining is predominantly used in medical fields in determining heart disease.

There has been a recent significant advancement of data mining and in machine learning domains in healthcare industry especially in fields of medical cardiology. The accumulation of medical data has given a possibility for researchers to train and validate net cases of algorithms in this particular field. As heart disease is a leading factor in death in developing nations, the identification and detection of the risk factors and early sign detection of these diseases are important aspects of this proposed approach.

This proposed study is designated to four different phases which constitutes the usage of ML and the DL algorithms and approaches in detection and in identification of the severity of heart diseases. This is done by implicating both the feature selection and the classification mechanism which are used in eliminating the irrelevant data and incorporating additional forms of needed features in case of making superior model performance.

In case of these data mining and the DL approaches in medical and in healthcare fields, there have been several advantages such as effective assessment, best treatment plans at lower costs. Advancements in clinical decision making was enhanced. Moreover, data mining enhanced the risks of drug interactions such as adapting big-data analytics. Increased ranges of diagnostic precision was achieved.

The proposed phase constitutes four phases. The first phase of the proposed study makes use of two various datasets. Initially, the UCI-dataset is used in performing the heart disease prognostics. This is one of the cleveland dataset, which has about 14 subsets. Followed by the second dataset is the HF- clinical dataset. This dataset has various set of health attributes which were primarily used in determining heart disease. This corresponding phase makes use of metaheuristic algorithms in aspects of making feature selection, such as GA, PSO, ABO and GSA. These algorithms are used for making an appropriate feature selection mechanism. DL approaches are carried forward in making classification within ranges of binary and the multi-class classification, by adapting RNN, LSTM with the Attention Mechanism.

This is followed by the second phase constituting use of HD clinical dataset and the UCI cleveland dataset by applying effectual forms of optimization and classification procedure, in order to depict the utility in aspects of early detection of cardiovascular disease of initiation. This aims in avoiding further cases of complex situations and emphasizes heart abnormality. Generally, optimization algorithms result in multi-face non-linear issues with their ability to adaptability and tractability. To overcome these issues, the respective phase makes use of the EPO algorithm in performing the feature selection mechanism. This algorithm is effective in picking up the necessary attributes from the data using the SSC-AE approach. As a case of consequence, SSC-AE is used under the aspects of performing the classification of heart diseases. The recommended technique has attained accuracy rate in ranges of 98.02%, with a precision value of 98%. Followed by the f1-score range in scales of 58% and finally the sensitivity value in range of 91%. Concurrently, this phase of approach using the SSC-AE achieve an AUC in rates of 96.04%.

Predicting the ranges of abnormality is difficult resulting in false positive and false negative rates. To overcome these laybacks, this analogous phase makes use of the real time data which is collected from the patients based on several health determining attributes. These include the adaptation of IoT based sensors, which are used in measuring the levels of cholesterol, BP and other vital health attributes. This data is then forwarded to feature selection approaches which are deployed using GSA, PSO, POA, GA and ABO. Finally, the data is classified using DL based approaches namely RNN, LSTM, SSA and DPA-RNN+LSTM. In this phase, both the binary and the multi-class classifications are done using adaptive approaches. The binary classification is done in aspects of determining whether or not heart disease will develop in advance in an individual. Whereas, the multiclass classification is done in aspects of determining cardiovascular disease severity into normal, mild, moderate and the severe cases. From the results, it is found that, for binary classification, GA, GSA and POA with proposed system shows ideal performance with 100% accuracy. Whereas, for multi-classification, ABO and GA with proposed system explores optimal performance with 99% accuracy. As the study considers real-time data, comparison with conventional works is not probable in this phase.

Phase IV makes use of real time dataset, by the data collected from the wearable sensors such as, bp sensor, FGA 10 and many others. This phase also procures the web-deployment of the data which is collected from the patient after the classification and the process of regression. The method of feature selection is carried out by five varying optimisation techniques such as GA, GSA, ABO, PSO and POA and DL methods such as RNN, LSTM, Deep progressive attention of RNN and LSTM, stacked Sparse Auto- encoder for categorisation of cardiovascular disease and regression. The regression is predicted by evaluating the parameters such as cardiac risk score, vascular age and cardiac index values using framingham equation. Based on these factors, the person is identified to be normal or abnormal. The model with less mean error rate is predicted and used for classification and regression of heart disease. The performance of the proposed system is calculated based on specificity, precision, recall, sensitivity, f1-score and accuracy. The precision produced by the SSAE model with POA is found to be 76.19% for heart disease classification. Followed by the process of regression by SSAE with various algorithms resulted in, SSAE with POA in ranges of 21.73, vascular age value of SSAE with PSO is 0.83 and cardiac risk score value of SSAE with ABO is 1.74.

8.2 Scope for Future Research

As a part of future, the proposed model can be performed with classification and regression using various other optimization algorithms and DL approaches. These phases of the projected model can also be validated using other such different datasets in assessing the model effectiveness. The proposed model in future can be increased with their ranges of accuracy and other possible performance metrics. This same model can be used in deterring other such disease abnormality ranges with their complexity ranges. Further aspects for future recommendation are:

- Unlabelled data can also be used in case of making the model learn and perform classification and feature selection.
- Treatment and recovery of diseases can also be worked using the proposed approach.
- Image related dataset can be used as the input in determining the disease complication by feature selection and classification approaches.
- Performances of the model against active learning algorithms can be handled partially using labelled instances.
- Coupling automated forms of framework to detect CFD, by investigating the influence

of the amount of data used in initial training and exploring techniques can be further automated.

LIST OF PUBLICATIONS

INTERNATIONAL JOURNAL

1. **Vaishali Baviskar**, Madhushi Verma, Pradeep Chatterjee, Gaurav Singal, 2023 "Efficient Heart Disease Prediction Using Hybrid Deep Learning Classification Models ", *Innovation and Research in Biomedical Engineering (IRBM), Elsevier*, https://doi.org/10.1016/j.irbm.2023.100786 (SCI, IF: 5.5).

2. **Vaishali Baviskar**, Madhushi Verma, Pradeep Chatterjee, Gaurav Singal, Thippa Reddy Gadekallu, 2023 "Optimization using IoA based SSC-AE: Stacked Sparse CNN Auto Encoder Model for Heart Disease Prediction ", *Expert Systems, Wiley*, 10.1111/exsy.13359 (SCI, IF:3.09).

3. **Vaishali Baviskar**, Madhushi Verma, Pradeep Chatterjee, 2023 "IoT based Heart Disease Prognosis Multi Class Classification using Deep Learning ", *ACM Transactions on Internet Technology*, doi:10.1016/j.compbiomed.2021.104432 (Communicated, SCI, IF: 5.15).

4. **Vaishali Baviskar**, Madhushi Verma, Pradeep Chatterjee, 2023 "IoT Based Heart Disease Prediction using Regression Methods ", *Neural Computing and applications*, (Communicated, SCI, IF:5.6).

INTERNATIONAL CONFERENCE

1. **Baviskar V.**, Verma M., Chatterjee P., 2021 "A Model for Heart Disease Prediction Using Feature Selection with Deep Learning", *Proceedings of the 10th Springer International Advance Computing Conference (IACC 2020)*, December 5-6, Goa, Panaji, India, pp. 151-16, https://doi.org/10.1007/978-981-16-0401-0

2. **V. Baviskar**, M. Verma and P. Chatterjee, 2021 "Improving classification performance of deep learning models using bio-inspired computing", *13th International Conference on Contemporary Computing (IC3- 2021)*, 05-07 August, Noida, India, https://doi.org/10.1145/3474124.3474174.

3. **Baviskar Vaishali**, Y.Dwivedi, Mishra Manika, Verma Madhushi, Chatterjee Pradeep, 2022 "Design of an Augmented Ensemble Heart Failure Prediction Model using Multi Parametric Analysis ", *in I2CT Proc of the int conf of latest discoveries in Convergence in Technology, IEEE Bombay Section*, India, Pune April 07-09, https://doi.org/10.1109/I2CT54291.2022.9823979

4. **Vaishali Baviskar**, Aditya Ojha, Ishaan Verma, Sahajpal Mahla, Madhushi Verma, Pradeep Chatterjee, Sunil Kumar, 2023 "H2EMDCO: Design of a Hyper Ensemble Model for Evaluation of Heart Diseases from Clinical Observations,", *Procedia Computer Science, Volume 218, 2023, Pages 1288-1294, ISSN 1877-0509*, International Conference on Machine Learning and Data Engineering, September 07-08, 2022, https://doi.org/10.1016/j.procs.2023.01.107

BOOK CHAPTER

1. **Vaishali Baviskar**, Dr. Madhushi Verma, Dr. Pradeep Chatterjee, Divya Srivastava, Sunil Jangir, Manish Kumar, 2022"IOT Enabled Heart Disease Prediction using ML, ", *Industrial Internet of Things: Technologies and Research Directions*, CRC Press (Taylor & Francis Group), DOI: 10.1201/9781003145004-7

2. **Vaishali Baviskar**, Dr. Madhushi Verma, Dr. Pradeep Chatterjee, 2021 "Accurate Heart

Disease Prediction using AI based hybrid classification models", *A step Towards Society 5.0 : Research, Innovations, and Development in Cloud-Based Computing Technologies*, CRC Press (Taylor & Francis Group), https://doi.org/10.1201/9781003138037

PATENT

1. **V. Baviskar**, M. Verma., P. Chatterjee and G. Singal, "Mechanism for Non-invasively Predicting Heart Disease ", *Submitted to Patent Office, India on April 30, 2021*, Published on October 22, 2021, Application number: 202111019854.

REFERENCES

[1] M. Abdar, W. Książek, U. R. Acharya, R. S. Tan, V. Makarenkov, and P. Pławiak. A new machine learning technique for an accurate diagnosis of coronary artery disease. *Computer Methods and Programs in Biomedicine*, 179:104992, 07 2019.

[2] H. Ahmed, E. M. G. Younis, A. M. Hendawi, and A. A. Ali. Heart disease identification from patients' social posts, machine learning solution on spark. *Future Gener. Comput. Syst.*, 111:714–722, 2020.

[3] L. Ali, A. Niamat, J. Khan, N. Amiri Golilarz, X. Xingzhong, A. Noor, R. Nour, and S. A. C. Bukhari. An optimized stacked support vector machines based expert system for the effective prediction of heart failure. *IEEE Access*, 7:54007–54014, 04 2019.

[4] M. Alkhodari and L. Fraiwan. Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. *Computer Methods and Programs in Biomedicine*, 200:105940, 01 2021.

[5] Y. An, N. Huang, X. Chen, F. Wu, and J. Wang. High-risk prediction of cardiovascular diseases via attention-based deep neural networks. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 18(3):1093–1105, aug 2019.

[6] K. Arumugam, M. Naved, P. P. Shinde, O. Leiva-Chauca, A. Huaman-Osorio, and T. Gonzales-Yanac. Multiple disease prediction using machine learning algorithms. *Materials Today: Proceedings*, 80:3682–3685, 2023. SI:5 NANO 2021.

[7] A. Arya, S. Kumar, P. Tyagi, D. Gola, and A. Mishra. *Forest-Based Medicinal Plants for Cardiovascular Diseases*, pages 213–227. 07 2021.

[8] G. Bazoukis, S. Stavrakis, J. Zhou, S. Chandra, G. Tse, Q. Zhang, J. Singh, and A. Armoundas. Machine learning versus conventional clinical methods in guiding management of heart failure patients—a systematic review. *Heart Failure Reviews*, 26, 01 2021.

[9] J. W. Benjamins, T. Hendriks, J. Knuuti, L. Juárez-Orozco, and P. van der Harst. A primer in artificial intelligence in cardiovascular medicine. *Netherlands Heart Journal*, 27, 05 2019.

[10] R. Bharti, A. Khamparia, D. M. Shabaz, G. Dhiman, S. Pande, and P. Singh. Prediction of heart disease using a combination of machine learning and deep learning. *Computational Intelligence and Neuroscience*, 2021, 07 2021.

[11] G. T. Brown JC and K. E. Risk factors for coronary artery disease. *Treasure Island*, 2023.

[12] J. I.-Z. Chen and P. Hengjinda. Early prediction of coronary artery disease (cad) by machine learning method - a comparative study. 2021.

[13] H. Dai, H.-G. Hwang, and V. S. Tseng. Convolutional neural network based automatic screening tool for cardiovascular diseases using different intervals of ecg signals. *Computer Methods and Programs in Biomedicine*, 203:106035, 2021.

[14] A. J. Delima and J. C. Arroyo. An optimized neural network using genetic algorithm for cardiovascular disease prediction. *Journal of Advances in Information Technology*, 13:95–99, 01 2022.

[15] O. Deperlioglu, U. Kose, D. Gupta, A. Khanna, and A. K. Sangaiah. Diagnosis of heart diseases by a secure internet of health things system based on autoencoder deep neural network. *Computer Communications*, 162:31–50, 2020.

[16] R. J. Desai, S. V. Wang, M. Vaduganathan, T. Evers, and S. Schneeweiss. Comparison of Machine Learning Methods With Traditional Models for Use of Administrative Claims With Electronic Medical Records to Predict Heart Failure Outcomes. *JAMA Network Open*, 3(1):e1918962–e1918962, 01 2020.

[17] G.-P. Diller, A. Kempny, S. Babu-Narayan, M. Henrichs, M. Brida, A. Uebing, A. Lammers, H. Baumgartner, W. Li, S. Wort, K. Dimopoulos, and M. Gatzoulis. Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: Data from a single tertiary centre including 10 019 patients. *European heart journal*, 40, 01 2019.

[18] M. Diwakar, A. Tripathi, K. Joshi, M. Memoria, P. Singh, and N. Kumar. Latest trends on heart disease prediction using machine learning and image fusion. *Materials today: proceedings*, 37, 10 2021.

[19] A. K. Dubey, A. K. Sinhal, and R. Sharma. An improved auto categorical pso with ml for heart disease prediction. *Engineering, Technology amp; Applied Science Research*, 12(3):8567–8573, Jun. 2022.

[20] A. Dutta, T. Batabyal, M. Basu, and S. Acton. An efficient convolutional neural network for coronary heart disease prediction. *Expert Systems with Applications*, 159:113408, 05 2020.

[21] A. Ed-daoudy and K. Maalmi. A new internet of things architecture for real-time prediction of various diseases using machine learning on big data environment. *Journal of Big Data*, 6, 11 2019.

[22] M. G. Vishnupriya, T. Pradeep and M. VJ. Prediction of pre-cardiac disease using ml dl techniques in t-kinter. *Journal of Survey in Fisheries Sciences*, 10:1874–1883, 2023.

[23] T. Gadekallu, P. Reddy, K. Lakshman, D. Rajput, R. Kaluri, and G. Srivastava. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evolutionary Intelligence*, 13, 06 2020.

[24] X.-Y. Gao, A. Ali, H. Shaban, and E. Anwar. Improving the accuracy for analyzing heart diseases prediction based on the ensemble method. *Complexity*, 2021:1–10, 02 2021.

[25] P. Gemke, N. Spicher, and T. Kacprowski. An lstm-based listener for early detection of heart disease. In *2022 Computing in Cardiology (CinC)*, volume 498, pages 1–4, 2022.

[26] P. Ghosh, S. Azam, M. Jonkman, A. Karim, F. M. Shamrat, E. Ignatious, S. Shultana, A. Beeravolu, and F. De Boer. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques. *IEEE Access*, PP:1–1, 01 2021.

[27] C. Harrison and C. Sidey-Gibbons. Machine learning in medicine: a practical introduction to natural language processing. *BMC Medical Research Methodology*, 21, 07 2021.

[28] S. Hiriyannaiah, S. G M, K. M, and K. Srinivasa. A comparative study and analysis of lstm deep neural networks for heartbeats classification. *Health and Technology*, 11, 04 2021.

[29] S. H. e. a. Joloudari JH, Joloudari EH. Coronary artery disease diagnosis; ranking the significant features using a random trees model. *International journal of environmental research and public health*, 17(3), 01 2020.

[30] H. k. Kalluri and S. Tulasi Krishna. A deep learning method for prediction of cardiovascular disease using convolutional neural network. *Revue d intelligence artificielle*, 34:601–606, 12 2020.

[31] R. Kannan and V. Vasanthi. *Machine Learning Algorithms with ROC Curve for Predicting and Diagnosing the Heart Disease*, pages 63–72. Springer Singapore, Singapore, 2019.

[32] E. Karregat, J. Himmelreich, W. Lucassen, W. Busschers, H. van Weert, and R. Harskamp. Evaluation of general practitioners' single-lead electrocardiogram interpretation skills: a case-vignette study. *Family practice*, 38, 08 2020.

[33] R. Katarya and S. K. Meena. Machine learning techniques for heart disease prediction: A comparative study and analysis. *Health and Technology*, 11:87–97, 2020.

[34] Y. Khourdifi and M. Bahaj. Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. *International Journal of Intelligent Engineering and Systems*, 12, 02 2019.

[35] S. Krishnan, P. Magalingam, and R. Ibrahim. Hybrid deep learning model using recurrent neural network and gated recurrent unit for heart disease prediction. *International Journal of Electrical and Computer Engineering*, 11:5467–5476, 12 2021.

[36] C. Krittanawong, K. Johnson, R. Rosenson, Z. Wang, M. Aydar, U. Baber, J. Min, W. Tang, J. Halperin, and S. Narayan. Deep learning for cardiovascularmedicine: A practical primer. *European heart journal*, 40, 02 2019.

[37] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of Ambient Intelligence and Humanized Computing*, 14:1–28, 01 2022.

[38] M. Lee, T.-G. Song, and J.-H. Lee. Heartbeat classification using local transform pattern feature and hybrid neural fuzzy-logic system based on self-organizing map. *Biomedical Signal Processing and Control*, 57:1–9, 11 2019.

[39] N. Masih, H. Naz, and S. Ahuja. Multilayer perceptron based deep neural network for early detection of coronary heart disease. *Health and Technology*, 11, 11 2020.

[40] A. Mehmood, M. Iqbal, Z. Mehmood, A. Irtaza, M. Nawaz, T. Nazir, and M. Masood. Prediction of heart disease using deep convolutional neural networks. *ARABIAN JOURNAL FOR SCIENCE AND ENGINEERING*, 46, 10 2020.

[41] I. D. Mienye, Y. Sun, and Z. Wang. Improved sparse autoencoder based artificial neural network approach for prediction of heart disease. *Informatics in Medicine Unlocked*, 18:100307, 2020.

[42] K. Modepalli, G. Gnaneswar, R. Dinesh, Y. Sai, and R. Suraj. Heart disease prediction using hybrid machine learning model. pages 1329–1333, 01 2021.

[43] D. Mohamed, B. Youcef, M. Ladjal, and A. Bilal. Heart disease prediction using mlp and lstm models. pages 1–5, 09 2020.

[44] S. Mohan, C. S. Thirumalai, and G. Srivastava. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, PP:1–1, 06 2019.

[45] U. Nagavelli, D. Samanta, and P. Chakraborty. Machine learning technology-based heart disease detection models. *Journal of Healthcare Engineering*, 2022:1–9, 02 2022.

[46] S. Nurmaini, R. Umi Partan, W. Caesarendra, T. Dewi, M. Naufal Rahmatullah, A. Darmawahyuni, V. Bhayyu, and F. Firdaus. An automated ecg beat classification system using deep neural networks with an unsupervised feature extraction technique. *Applied Sciences*, 9(14), 2019.

[47] C. R. Olsen, R. J. Mentz, K. J. Anstrom, D. Page, and P. A. Patel. Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. *American heart journal*, 229:1—17, November 2020.

[48] D. Overman. Ai predicts heart disease risk using chest x-ray. *AXIS Imaging News*, 2022.

[49] K. K. S. P. Singh, N. Singh and A. Singh. Diagnosing of disease using machine learning. *Machine learning and the internet of medical things in healthcare*, pages 89–111, 2021.

[50] Y. Pan, M. Fu, B. Cheng, X. Tao, and J. Guo. Enhanced deep learning assisted convolutional neural network for heart disease prediction on the internet of medical things platform. *IEEE Access*, 8:189503–189512, 01 2020.

[51] K. B. e. a. Raju. Smart heart disease prediction system with iot and fog computing sectors enabled by cascaded deep learning model. *Computational Intelligence and Neuroscience*, Jan 2022.

[52] A. Rath, D. Mishra, G. Panda, and S. C. Satapathy. Heart disease detection using deep learning methods from imbalanced ecg samples. *Biomedical Signal Processing and Control*, 68:102820, 2021.

[53] A. Raza, A. Mehmood, D. S. Ullah, M. Ahmad, G. S. Choi, and B.-W. On. Heartbeat sound signal classification using deep learning. *Sensors*, 19, 11 2019.

[54] S. Roy, T. Meena, and S.-J. Lim. Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine. *Diagnostics*, 12:2549, 10 2022.

[55] F. Rustam, A. Ishaq, K. Munir, M. Almutairi, N. Aslam, and I. Ashraf. Incorporating cnn features for optimizing performance of ensemble classifier for cardiovascular disease prediction. *Diagnostics (Basel, Switzerland)*, 12, 06 2022.

[56] H. O. Seidu and S. Seidu. Lifestyle or pharmacotherapy in cardio-metabolic disease prevention. *Therapeutic Advances in Cardiovascular Disease*, 17:17539447231177175, 2023.

[57] S. B. Shuvo, S. N. Ali, S. Swapnil, M. Alrakhami, and A. Gumaei. Cardioxnet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings. *IEEE Access*, PP:1–1, 03 2021.

[58] L. Stevens, B. Mortazavi, R. Deo, L. Curtis, and D. Kao. Recommendations for reporting machine learning analyses in clinical research. *Circulation: Cardiovascular Quality and Outcomes*, 13, 10 2020.

[59] O. Taylan, A. S. Alkabaa, H. S. Alqabbaa, E. Pamukçu, and V. Leiva. Early prediction in classification of cardiovascular diseases with machine learning, neuro-fuzzy and statistical methods. *Biology*, 12(1), 2023.

[60] S. Telemedicine. The drawbacks of current heart disease detection technology. 2023.

[61] R. Tusher, A. Beeravolu, M.-H. Islam, A. Karim, S. Azam, and S. Mukti. A performance based study on deep learning algorithms in the efficient prediction of heart disease. 12 2021.

[62] W. A. e. a. Ullah M, Hamayun S. Smart technologies used as smart tools in the management of cardiovascular disease and their future perspective. *Current problems in cardiology*, page 101922, 10 2023.

[63] U. D. V. K. V. Shankar, V. Kumar and K. Rohitaksha. Heart disease prediction using cnn algorithm. *SN Computer Science*, 1:170, 2020.

[64] Z. Wang, S. Stavrakis, and B. Yao. Hierarchical deep learning with generative adversarial network for automatic cardiac diagnosis from ecg signals. *Computers in Biology and Medicine*, 155:106641, 2023.

[65] H. C. e. a. Willemink MJ, Koszek WA. Preparing medical imaging data for machine learning. *Radiology*, 295(1):4–15, 2020.

[66] B. Xiao, Y. Xu, X. Bi, J. Zhang, and X. Ma. Heart sounds classification using a novel 1-d convolutional neural network with extremely low parameter consumption. *Neurocomputing*, 392:153–159, 2020.

[67] M. Y and T. M. et al. Early and accurate detection and diagnosis of heart disease using intelligent computational model. *Scientific reports*, 10(1):17, 2020.

[68] L. Yang, H. Wu, X. Jin, P. Zheng, S. Hu, X. Xu, W. Yu, and J. Yan. Study of cardiovascular disease prediction model based on random forest in eastern china. *Scientific Reports*, 10, 03 2020.

[69] H. Yoo, S. Han, and K. Chung. A frequency pattern mining model based on deep neural network for real-time classification of heart conditions. *Healthcare*, 8:234, 07 2020.

[70] Y. Zhang, J. Wang, J. Zhao, G. Huang, K. Liu, W. Pan, L. Sun, J. Li, W. Xu, C. He, Y. Zhang, S. Li, H. Zhang, J. Zhu, and Y. He. Current status and challenges in prenatal and neonatal screening, diagnosis, and management of congenital heart disease in china. *The Lancet Child Adolescent Health*, 7(7):479–489, 2023.