Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

INVESTIGATION OF HYBRID SVM –EBO METHOD BASED HEART DISEASE PREDICTION

KAVITHA C*
B.E., MTech, Associate Professor in the Department of Computer Science, Vijaya Vittala Institute of Technology, Bangalore, Karnataka, India. Email: kc98cs@gmail.com

T.N ANITHA
Professor, Department of Information Science and Engineering, Atria Institute of Technology, Bangalore. Email: Anithareddy72@gmail.com

ABSTRACT:

The deadliest disease among humans is the heart disease, and it's prevalent among all the ages of the people. Therefore, there is an urgent need for early detection of the heart disease, for both prevention and therapy purpose. However, the clinical methods used to diagnose HD today are expensive and frequently need for a high degree of intervention. Recently, varieties of intelligent systems for the automated diagnosis of HD have been designed by researchers as a solution to this problem. But, still there is a huge research gap in terms of accurate forecasting. This fact serves as the impetus for the current study's introduction of a new optimized Support Vector Machine (SVM) model for HD prediction. Pre-processing, feature extraction, feature selection, and heart disease prediction are the four main stages of the projected model. The initial pre-processing of the obtained data uses the Antonyan Vardan Transform (AVT) method. The features are then retrieved from the pre-processed data using the Interclass sub-Space Clustering method. The necessary characteristics are then retrieved using the Partially Differential Equation (PDE) from the extracted features. Finally, the heart disease is accurately predicted using the new SVM model that has been improved. The SVM is trained with new Enhanced Bat Optimization (EBO). The optimized SVM yields the final predicted result.

KEYWORDS: Heart Disease Prediction; Antonyan Vardan Transform; Interclass sub-Space Clustering; Partially Differential Equation (PDE); SVM; EBO

1 Introduction

The heart is a muscular organ in the human organism [1] that sits in the mid of the chest and leans slightly to the left. It draws deoxygenated blood from every region of the body and carries it to the lungs, where it is converted into oxygenated blood and carbon dioxide is expelled. As a result, oxygenated blood is then circulated throughout the body from the lungs. The heart circulates over 72, 00 litres of blood throughout the body every day, beating 3 billion times throughout its lifetime. In addition to moving blood, the heart also moves other substances around the body, including hormones created by tissues and nutrients from digestion that are dispersed to entire body cells [2]. Heart disease claims the lives of 17.5 million people per annum [3]. The leading cause of death and shortening of life expectancy in modern society is heart disease [4], which is a common condition. Because the heart is a portion of our body that is absolutely necessary for life to exist, life is dependent on how well it functions. Heart illness may

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

be calamitous or make a patient miserable before passing away and affects how well the heart works. One of the main problems with heart malady is estimating a person's likelihood of not getting enough blood to the heart [5]. An important problem in clinical data analysis is the prediction of cardiovascular disease (CVD) [6]. Based on indications such as pulse rate, sex, age, and many others, heart disease is predicted [7].

World Health Organization estimates that 17.9 million people died as a result of heart disease in 2016. Heart conditions include coronary heart disease, hypertension, and cerebral infarction are the leading causes of mortality in the USA. In the US, coronary heart disease alone causes around 366,800 deaths each year, accounting for one in every seven fatalities. The estimated number of heart attacks in the US is 7.9 million, or around 3% of all adult heart attack cases in the country. In the same nation, heart attacks claimed the lives of 114,023 people in 2015 [8]. Although there are children who have comparable health difficulties, cardiovascular illnesses are more common in males than in women, especially in middle or late age. Throughout 17.9 million people, globally die from CVDs each year [9]. According to the European Cardiology Society (ESC), 3.6 million individuals worldwide receive a heart disease diagnosis each year, totalling 26 million persons with the condition [10]. A heart attack can be brought on by arterial plaque build-up, which is a condition known as heart disease. By adopting a healthful lifestyle, such as cutting back on salt intake, increasing fruit and vegetable consumption, engaging in regular physical exercise, and giving up alcohol and cigarette use, these risk factors can be reduced, thus lowering the chance of developing heart disease [11]. Unhealthy eating habits, unhealthy diets, and abusing liquor and cigarettes to excess are a few risk factors that might result in heart disease [12].

The use of advanced technology helps to enhance the present health-care system by assisting in the discovery of patterns in vast quantities of data that individuals are unable to discern. In order to anticipate the outcome of new data that is input into the system, a machine-learning model first learns from the documented data it gets and then develops prediction algorithms. The effectiveness and quantity of the input data determine how accurate these models are. A large amount of data will help in the creation of a more accurate model that accurately forecasts the outcome. [13]. In order to take action to save death, an early, accurate, and effective medical diagnosis of heart disease is essential [14]. Massive amounts of data from many different sectors, including the medical industry, may be analysed using machine learning techniques. It serves as an replacement to conventional prediction modelling methods employing computers to comprehend complicated and non-linear interactions between many components by minimizing errors in expected and actual results [15].

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

2 Related Work

Ali et al. [16] had presented life-threatening nature of cardiac disease; it might result in heart attacks, which can have deadly consequences. Data mining and machine learning approaches might be utilised to anticipate the illness's incidence because of its potential for accurate disease prediction rates. Usefulness of ML techniques to heart disease prediction via a heart disease dataset and discovered that three classification algorithms, KNN, RF, and DT, performed incredibly well with 100% accuracy was tested. For every method used, with the exception of MLP and KNN, feature significance ratings for each feature were also computed. Based on the score for feature relevance, these features were prioritised.

Ali et al. [17] had introduced a system for intelligent healthcare surveillance that combines an ensemble deep learning model with feature fusion techniques to increase the precision of heart disease prediction and support doctors in production prompt and accurate diagnoses of heart patients. Many reasonable issues are covered, along with physiological data collection through wearable sensors and medical tests, FRF extraction from EMRs, feature fusion-based transformation of extracted data into a useful dataset, important feature selection by means of information gain, identification of the significance of features by feature weighting method, heart disease prediction through an ensemble deep learning model, and ontology-based dietary plans and actigraphy. The suggested approach offers a prediction system that identifies the most significant risk indicators in high-dimensional healthcare data and critically evaluates them to precisely forecast heart disease before a heart attack occurs.

Javeed et al. [18] had presented the overfitting issue with newly developed approaches for heart failure prediction, and suggested a unique learning system to help with heart failure prediction. Two algorithms are combined in the learning system. A random search algorithm was employed in the first method to find a subset of attributes that include additional information concerning heart failure. Based on the chosen subset of characteristics, the second algorithm, random forest, predicts heart failure.

Guo et al. [19] had organised the raw data and to provide a new and improved knowledge of cardiac disease, machine-learning techniques have been used. It is crucial and extremely difficult in the realm of medicine to forecast cardiac disease. But the death rate can be significantly reduced if the sickness was identified early on and the preventative evaluation was made as soon as feasible. By combining the characteristics of the linear model and random forest, the suggested RFRF-ILM approach was used. When predicting cardiac disease, RFRF-ILM has great accuracy.

Rani et el. [20] had presented clinical characteristics of the patient and proposed a hybrid decision support system that can help in the early detection of cardiac disease.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

The method for handling the missing values is multivariate imputation through chained equations. In order to select the most appropriate features from the supplied dataset, a hybridised feature selection approach that combines the Genetic Algorithm (GA) and recursive feature elimination has been implemented.

Saqlain et al. [21] had put forth the feature selection algorithms, MFSFSA, FFSA, RFSA, and FSSA, a system for diagnosing heart illness. The suggested algorithms choose the features in the feature subsets based on their respective Fisher scores and MCC scores. Each of the proposed feature selection algorithms chooses one feature subset, and among all of them, it is wise to choose the feature subset with the highest MCC score and fewer dimensions. For heart disease prediction, a binary class RBF kernel-based SVM was employed.

Ismail et al. [22] had presented sizable percentage of coronary heart pollution accumulation frameworks. According to the examination, a fake neural network model was adequate for gathering data on a sizable amount of medical information. The results of our suggested ANN-based complete framework implementation show at least a 5% improvement in spotlight extraction and forecast. When equipped with sufficient data, the classifier used was more effective than neural networks or SVM.

Ashish et al. [23] had analysed an automatic IHD categorization and localization system. The appropriate for post-surgical and early diagnostics were discussed. Both the training and testing of the data use the Random Forest machine learning technique. The procedure of cross-validation achieves the maximum accuracy, and by utilising this, the same database beats other machine learning algorithms. Locations of stenosis are correlated with magnetic field patterns, particularly pole features. After pre-processing the data, the machine learning methods of SVM and XGB classifier were initially used to the data set of ZAlizadeh Sani, heart disease.

Ali et al. [24] had detected the diagnosis of heart failure based on stacked SVMs. While the second SVM model served as a prediction model, the first SVM model was utilised to remove unnecessary characteristics. With the use of a hybrid grid search technique, both models were improved. It was demonstrated that the suggested strategy outperformed eleven well-known methods that were used and other cutting-edge machine learning models.

Alfaidi et al. [25] had proposed the effectiveness of machine learning methods for foretell the likelihood of cardiovascular illness. Databank of cardiovascular illness with 70000 individuals was used. The performance of the models was assessed based on their accuracy and outlined a few procedures for pre-processing the dataset. Additionally detailed features was selected that affect how well the models perform.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

3. Materials and Methods of Proposed System

In this research work, a novel heart disease prediction model is introduced. The projected model includes four major phases: (a) pre-processing, (b) feature extraction, (c) feature selection and (d) heart disease prediction. Let the collected input data be denoted as $\boldsymbol{D_i^{inp}; i = 1, 2, \ldots n}$. Here, $\boldsymbol{n}$ denotes the count of data.

Step 1 : Initially, the collected data $\boldsymbol{D_i^{inp}}$ is pre-processed via Antonyan Vardan Transform (AVT) approach. The data acquired after pre-processing is pointed as $\boldsymbol{D_i^{pre}}$.

Step 2 : Then, from the pre-processed data $\boldsymbol{D_i^{pre}}$, the features are extracted using the Interclass sub–Space Clustering approach. The features extracted from $\boldsymbol{D_i^{pre}}$is denoted as $\boldsymbol{F_i}$.

Step 3 : Subsequently, from the extracted features $\boldsymbol{F_i}$, the relevant ones are extracted using the Partially Differential Equation (PDE). The selected optimal features are denoted as $\boldsymbol{G_i}$

Step 4 : Finally, the heart disease is precisely forecasted using the new optimized SVM model, which is trained via Enhanced Bat Optimization (EBO). The final predicted outcome is acquired from optimized SVM.

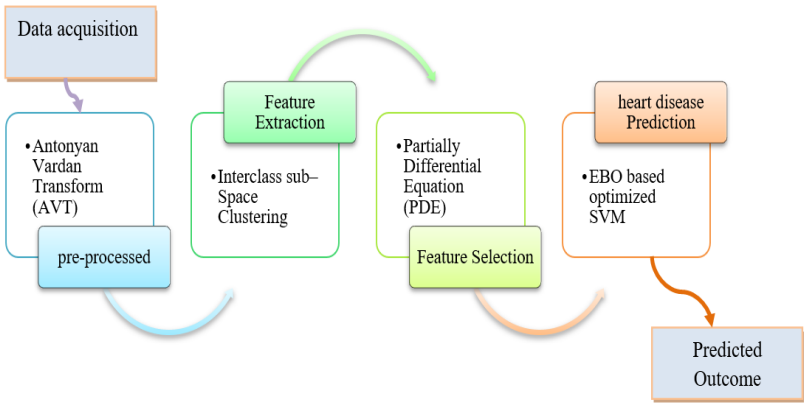

Figure 1: Architecture of the projected disease prediction model

3.1 Dataset Collection

The gathering of data is the first stage of this study project. The information was carefully gathered from 74 ICU patients' medical records. The information in these records includes things such "surgery information, ICU information, pharmacological information, and the process parameters have been chosen by skilled physicians and nurses. D_i^{inp} is used to represent the obtained data.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

3.2 Data Pre-processing- Antonyan Vardan Transform (AVT)

Pre-processing is done on the obtained raw data D_i^{inp} using Antonyan Vardan Transform (AVT). The benefit of this technique is that it does not totally reliant on Gaussian noise distribution. It can quickly recognise and remove occasional random noise. A data point with a change in standard deviation higher than one is removed by this technique as an outlier. This filters out useful data for the subsequent stage or result. Because it doesn't rely on the raw data's Gaussian distribution, it enhances the quality of the data that is provided for final usage. The following stages may be used to implement the algorithm, which is straightforward:

1. Gather N data samples.

2. Determine the dataset's average and standard deviation.

3. Eliminate any data with a standard deviation that is higher or lower than average.

4. Compute the average for the remaining group of data and display the results. $D_i^{inp(AVT)}$ stands for the obtained result.

The AVT algorithm's phases, which include the heart disease row j_row and the disease column k_col, outline the pre-processing of data to eliminate noise and undesirable data. In Algorithm 1, the Antonyan Vardan Transform (AVT) pseudo-code is displayed.

Algorithm 1: Pseudo-code of Antonyan Vardan Transform (AVT)

Begin

$For(j\_row = 0;\ j\_row < length(\ [data]\ \_row\ );j++)$

$For(k\_col = 0; k\_col < length(\ [data]\ \_col\ );k++)$

$if(lenth(data)[j_{row}][k_{col}] = 0)$

$data_{processing}\,[j_{row}][k_{col}]$

$$AVT = \frac{1}{2}\sum_{j}^{k} \log(j_{row}.k_{col})$$

Endif

End for

End for

Terminate

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

3.3 Missing Value removal

Then, in $D_i^{inp(AVT)}$, the missing values are removed. The handling of missing data is very important during the preprocessing of the dataset as many machine learning algorithms do not support missing values. The data acquired after missing value removal is denoted as $D_i^{inp(miss)}$.

3.4 Normal distribution transformation

These $D_i^{inp(miss)}$ are normalized within the range 0 to 0.5. the data acquired after normalization is denoted using the symbol D_i^{pre}. This D_i^{pre} is the final pre-processed data. From D_i^{pre}, the features are extracted.

3. 3 Feature Extraction- Interclass sub-Space Clustering

The internal characteristics related to certain class names are aggregated by Interclass sub-Space Clustering, and features like Age, Gender, Cholesterol, and treetops are retrieved from D_i^{pre}. Due to the volume and dimension of the data, it is essential to categorise it. The feature depth similarity clustering method is described to accomplish this. The approach assesses the degree of similarity between characteristics at several levels. The approach calculates the total number of samples in each class, the number of samples that get near to the feature, and the feature depth similarity (FDS) based on similarity. The approach computes the Interclass sub-Space Clustering measures to accomplish clustering after allocating a set of samples to every disease class clusters.

The feature like Age, Gender, Cholesterol, treetops are extracted from D_i^{pre}, and the internal features relevant specific class names are grouped via Interclass sub-Space Clustering. As the dimension and volume of data is huge in size, it is necessary to group them under specific class. To perform this, the feature depth similarity clustering approach is presented. The method measures the similarity among the features in multiple level. According to the similarity, the method computes the feature depth similarity (FDS), the number of samples gets close on the feature and the total number of samples in any class. Initially, a set of samples are assigned towards each disease class clusters and further the method computes the Interclass sub-Space Clustering measures to perform clustering. Algorithm 2 demonstrates the Interclass sub-Space Clustering pseudo-code.

Algorithm 2: Interclass sub-Space Clustering

Input: Pre-processed data

Output: Cluster Set Cs

Step1: Start

Step 2: Read phy-Icmr.

Step3: Initialize number of clusters Nc = $\sum Disease\ Classes$

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

For each disease class n

Assign random samples.

CS(n) = $\sum Random\ Sample\ (phy, Icmr)$

End

For each sample s

For each cluster c

For each cluster sample cls

Compute Feature Depth Similarity FDS = $\frac{\sum_{i=1}^{size\,(S)} S(i).value == Cls(i).value}{size(S)}$

End

Compute cumulative feature depth similarity CFDS.

CFDS = $\sum_{i=1}^{size(c)} C(i).FDS \Big/ size(C)$

End

Step 3: Compare the internal centroids to form cluster mediod weight

Cluster c = Choose the cluster with maximum CFDS.

Index the data point to the selected cluster.

End

Stop

According to feature similarity, the huge data has been grouped using the aforementioned technique. The degree of feature similarity between various data points in each cluster that is available has been used to quantify similarity. Data points have been indexed to a cluster based on the value of feature depth similarity, which has been chosen. The features extracted from D_i^{pre} is denoted as F_i.

3.4 Feature selection – Partially Differential Equation (PDE)

There are often many feature learning techniques, however these techniques have significant limitations, especially when working with tiny training sets. In this study, PDE is used for feature learning to get around this problem. The Extracted features F_i are the PDE's (first) input. The PDE can be written as per Eq. (1)

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

$$\begin{cases} \frac{\partial u}{\partial t} = f(u,x,y,t), \quad (x,y,t)\epsilon Q, \\ \quad u(x,y,t) = 0, \quad (x,y,t)\epsilon T \quad \cdots \\ u|t = 0(x,y,t) = F \ (x,y)\epsilon\Omega \end{cases} \tag{1}$$

Where I stands for the Extracted Features, Ω represents the number of features that I have extracted, and T represents the time at which the PDE has completed feature extraction. The learnt feature map is the evolution's outcome, u | t = T. The chosen ideal attributes are indicated by the symbol as $\boldsymbol{G_i}$.

Formulated the PDE: The $\boldsymbol{f}$ (u, x, y, t) in Eq. (2) is unknown

$$\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{t}} = \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{\nabla u}, \boldsymbol{Hu} \ldots \tag{2}$$

Different PDEs result from different f values. This intuition is used by the researcher to create a unique F_i for diverse data processing issues. To properly illustrate the feature extraction procedure for classification issues, is challenging. They often derive the property of F_i to narrow their search space rather than finding the correct form of the PDE straight away. To achieve approximation invariance in lighting, each basic differential invariant is made virtually invariant by the addition of the nonlinear mapping $g(x) = \frac{x}{1+|x|}$ Therefore, {g(inv$i$(u)}^{5}i =0 continue to be fundamental differential invariants. f is therefore a linear mixture of these modified basic differential invariants.

$$f(u,x,y,t) = \sum_{i=0}^{5} a_i(x,y,t)g(inv_i(u(t))) \ldots \tag{3}$$

Here a_i (x,y,t)5 i=0 are parameters to be determined

When $f(u,x,y,t)$ in Eq. (2) is chosen using Eq. (3), As follows, PDE is a simplified form of the PDE system. The entire PDE model is formulated as follows by integrating Feature learning and classification.

$$\min_{F,W} E = \frac{1}{M}\left||H - W.U|t = T\right|2_F + \lambda||W|2_F \ldots \tag{4}$$

$$\frac{\partial um}{\partial t} = \sum_{i=0}^{5} a_i(x,y,t)g(inv_i)\big(u(t)\big), \quad (x,y,t)\epsilon\ Q \ \ldots \tag{5}$$

$$u_m(x,y,t) = 0, (x,y,t)\epsilon T$$

$$u_m|t = 0(x,y,t) = \ I_m\ , (x,y)\epsilon\Omega$$

As a {a (t)$_j$}^{5} j=0 is set to minimize the loss function of the training data, the PDE extracts discriminative feature. The selected optimal features are denoted as $\boldsymbol{G_i}$.

3.5 Optimized SVM

The heart disease is precisely forecasted using the new optimized SVM model, which is trained via Enhanced Bat Optimization (EBO). The final predicted outcome is acquired

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

from optimized SVM.

SUPPORT VECTOR MACHINE

To divide numerous classes, a common wrapper-based sorter called Support Vector Machine (SVM) is used. Since SVM can categorise with trustworthy accuracy while utilising minimal computer resources, it is commonly employed in the data analytics industry. To do this, the basic data from the original input space is mapped using a function into a higher-dimensional space that allows for linear data separation. One discovers a hyper plane in this higher-dimensional space 0 with the greatest margin for establishing the borders between the input classes. The two most significant issues, however, are selecting a basic function that is acceptable and modifying the approach's parameters. The best space where categories are typically split linearly by a single disorderly transformation is found by a kernel function with the assistance of an optimization problem that is mostly used in logic to solve the challenge of selecting the optimal decision plane. Considering that the kernel functions are described as, $r(b_i, b_j) = \emptyset\ (b_i)^L \emptyset\ (b_j)$ and several common kernel functions in SVM models, including:

- Linear kernel, where $\emptyset = b_j * b_i$
- kernel of a polynomial in degree deg, $\emptyset = (b_j * b_i + 1)^{deg}$
- RBF kernel $\emptyset = \exp(-\left|\left|b_j - b_i\right|\right|^2 / 2\sigma^2)$,
- Sigmoid kernel $\emptyset = \tanh(b_j * b_i + 1)$

The polynomial and RBF (Radial Basis Function) kernels are frequently more suitable and acceptable for non-linear problems in a range of domains, such as intrusion detection, classification, and image processing, since they have been demonstrated to be beneficial in providing superior performance. Polynomial kernel provides good performance by requires less processing time, particularly for high-dimensional dataset.

Enhanced Bat Optimizer

The standard bat algorithm has been enhanced to achieve the objective of heart disease prediction accuracy. The way that bats measure distances using echolocation has an impact on EBO. When foraging at night, bats frequently employ quick, powerful sound impulses to locate obstacles or prey. The steps followed in EBO model is manifested below:

Step 1: Compute the Bat parameters.

Step2: Compute the fitness function (Fit) of every search agent using Eq. (1).

$$Fit = \max(Accuracy) \tag{6}$$

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

Step 3: Upgrade the global best position X^*, frequency of the pulses f_i, speed, and position x_i^t of the ith bat as per Eq. (7) to Eq. (9). The velocity of the search agents are updated using the newly projected expression given in Eq. (8). The projected model considers the global best position and global worst position of the search agents. Therefore, the search agents do not get trapped into the local optima. In addition, the position of the search agent x_i^{t+1} is also updated using the new expression given in Eq. (9). As per the proposed model, the position best of the solutions are considered, and therefore the convergence can be increased.

$$f_i = f_{min} + (f_{max} - f_{min})\,\beta;\ \beta \in [0,1] \tag{7}$$

$$V_i^{t+1} = \frac{V_i^t + (x_i^{t-1} - X_{best}^g) f_i}{X_{worst}^g} \tag{8}$$

$$x_i^{t+1} = \frac{(x_i^t + V_i^t)}{X_{best}^P} \tag{9}$$

Where X_{best}^g represents the global best position, X_{worst}^g denotes the global worst position and X_{best}^p points to the position best.

Step 3: The following equation generates a new solution for the bat if the random number is greater than r_i.

$$A_i^{t+1} = (\alpha A_i^t) * X_{best}^p \tag{10}$$

where A^t is the average volume of all bats at time t, and α is a random number between [1, 1].

Step 4: If the random number falls below A_i and $f(N_i) < f(N^*)$ next the modified approach is approved. Then, update A_i and r_i as per Eq. (10) to Eq. (12), respectively.

$$x_{new} = x_{old} + \varepsilon A^t \tag{11}$$

$$r_i^{t+1} = r_i^0 [1 - \exp(-\gamma t)] \tag{12}$$

Here, $A_i^t \rightarrow 0$, $r_i^t \rightarrow r_i^0$ as $t \rightarrow \infty$,

A_i^{t+1} and A_i^t denote the loudness at time t and $t+1$ respectively. r_i^t and r_i^0 are the initial pulse rate and pulse rate at time t , respectively.

Step 5: Find the current ideal solution X^* by classifying the bats according to their fitness.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

Step 6: Once the maximum number of iterations has been reached, go back to Step 2 and output the globally optimal answer.

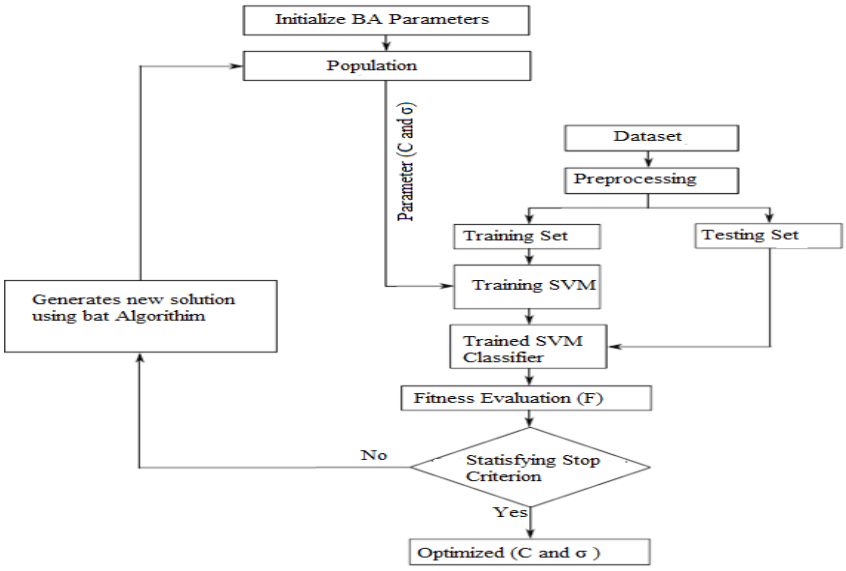

Fig. 2: Flowchart of Proposed System

3. Simulation Results and Discussions

Heart disease and cardiovascular illness remain the primary causes of death many developed countries- and middle-income countries. Many predispositions, such as the genetic susceptibility of such personal and professional life habits are earmarked as the reasons. Early diagnosis of heart disease, medical efficiency, and accuracy play an important role in preventing death. These refer to the feature extraction datasets such as Age, Gender, Cholesterol, treetops, etc. As required by the medical field, large Heart disease prediction data sets from various data set such as Age, Gender, Cholesterol, treetops etc. are useful in heart disease prediction. Artificial learning is one of the most rapidly developing method and it allows the analysis of large amounts of data from various fields and is an important tool in the medical field. These can be done by reducing the error between the actual results and predicted which replaces the conventional prediction modeling methods using a computer to understand the complex non-linear interaction between various factors. The simulation parameters used for this work is listed in table 1.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

Table 1: Simulation parameters

Parameters	Values used
Tool	Matlab
Language	C
Name of the dataset	UCI Machine Learning Repository dataset
Input data	Heart disease data set
No of Datas	800
Age group	(33-87)
Training dataset	500
Testing dataset	300

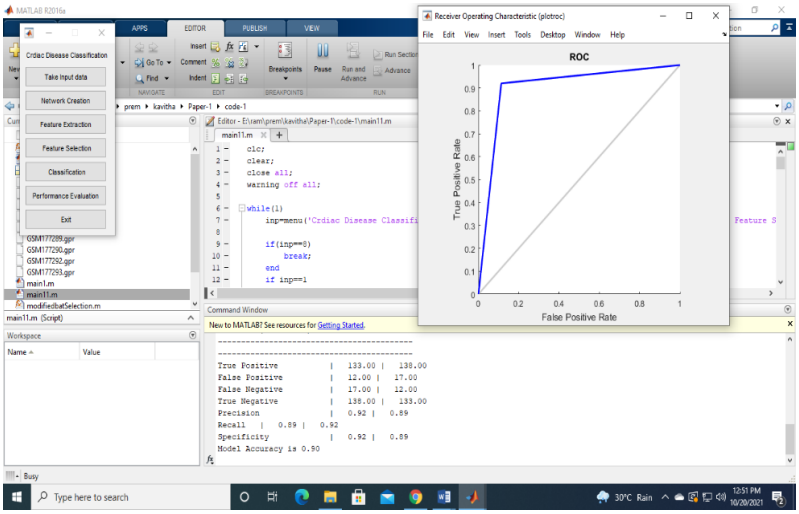

Fig. 3: Simulation Screen shot

The matlab simulation screen shot of proposed SVM-EBO based heart disease predcition is shown in Figure 3. In this work use UCI Machine Learning Repository dataset to validate the performance of heart disease prediction.

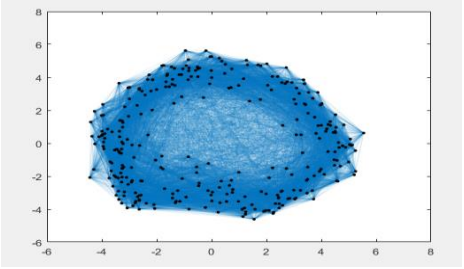

Fig. 4: Link prediction network

The figure 3 shows the simulation result of link prediction network. The purpose of this

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

link prediction network identify where nodes are the diseases and edges represent the contemporarily presence of two illnesses in a patient, is built

	Actual Classes	
Predicted Classes	0	1
0	135.0	12.0
1	15.0	138.0

Fig. 5: Result of Confusion Matrix

The figure 5 shows the simulation result of confusion matrix of proposed system. In this work use 300 data's for testing, based on this data's the confusion matrix result is obtained

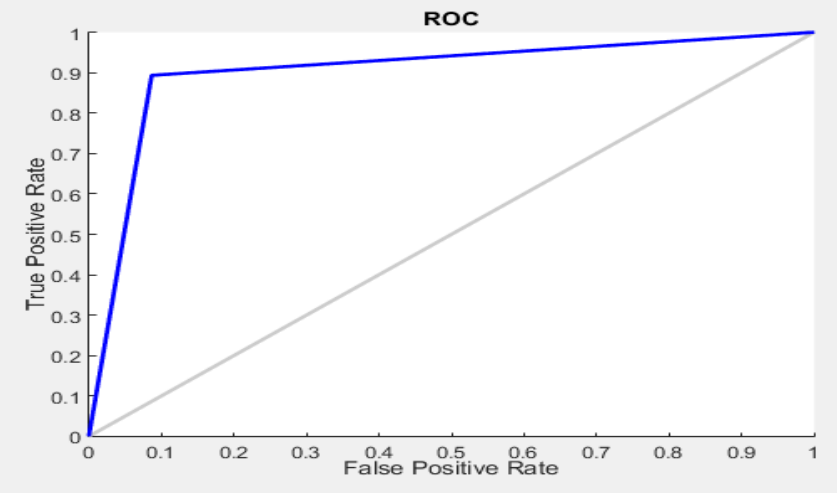

Fig. 6: ROC characteristics

The simulation results of Receiver Operating Characteristic (ROC) response of proposed SVM-EBO method based heart disease prediction is shown in above figure 6.

4.1 Performance Evaluation

The most important factor and the index for heart disease patients are the patient information. The accurate information helps in improving the robustness and the medical community's diagnostic performance. The following parameters are used to validate the performance of proposed system.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

Precision

Precision refers to the percentage of relevant results. This is usually expressed in percentage terms.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (8)$$

Recall

This is another statistical measurement that can be used to indicate the correct rate of identification. This is referred to as the true positive rate. This is usually expressed in percentage terms.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (9)$$

F-Measure

F-measure is also called F1-score, it measures a model's accuracy on a dataset. It evaluates binary classification systems, which categorize examples into 'positive' or 'negative. The best value of the F1 score is unity.

$$F1 = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \quad (10)$$

Accuracy

The degree of accuracy is defined by measuring the value to be consistent with a true or accepted value. This is usually expressed in percentage terms.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (11)$$

Where

TP = True Positive, TN = True negative, FP = False Positive and FN = false Negative

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

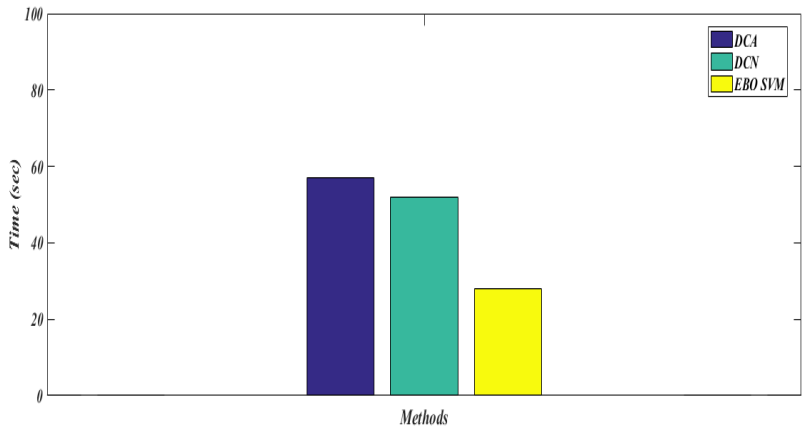

Fig. 7: Performance analysis of Time Complexity

The time complexity analysis of proposed SVM-EBO method with existing Difference of Convex functions Algorithm (DCA) and difference of Convex Neural-network DCN methods are shown in Figure 7. In this evaluation clearly shows the proposed SVM-EBO obtained low time complexity. The time complexity of DCA,DCN and SVM-EBO are 62Sec, 58sec and 26Sec.

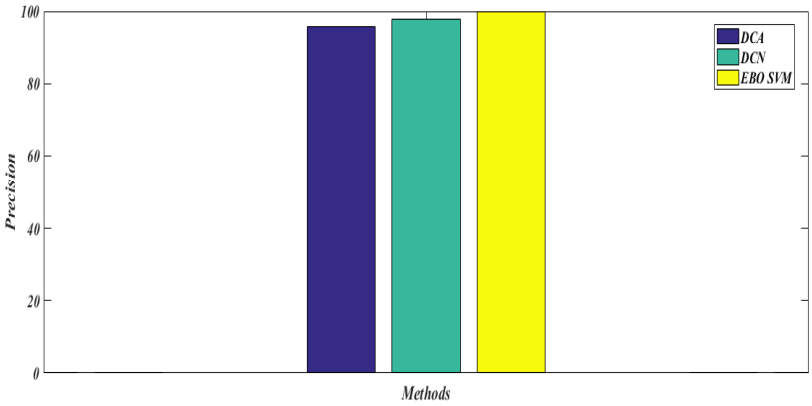

Fig. 8: Performance analysis of Precision

The precision analysis of proposed SVM-EBO method with existing Difference of Convex functions Algorithm (DCA) and difference of Convex Neural-network DCN methods are shown in Figure 8. In this evaluation clearly shows the proposed SVM-EBO obtained good precision response compared with other methods. The precision of DCA, DCN and SVM-EBO are 83%, 86%c and 91%.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

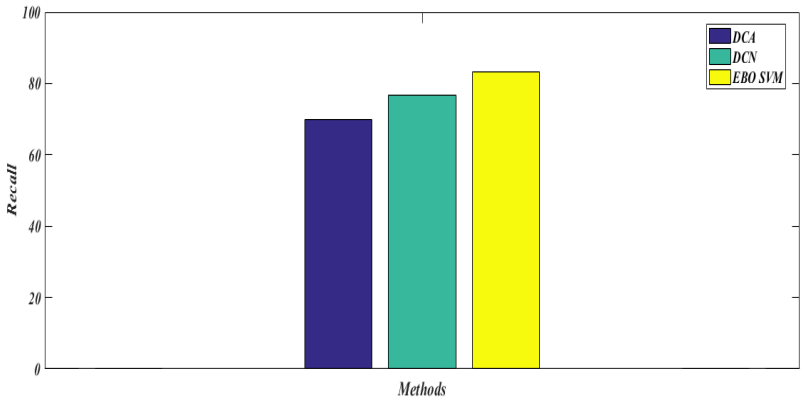

Fig. 9: Performance analysis of Recall

The recall analysis of proposed SVM-EBO method with existing Difference of Convex functions Algorithm (DCA) and difference of Convex Neural-network DCN methods are shown in Figure 9. In this evaluation clearly shows the proposed SVM-EBO obtained good recall compared with other methods. The recall of DCA,DCN and SVM-EBO are 75%, 79%c and 86%.

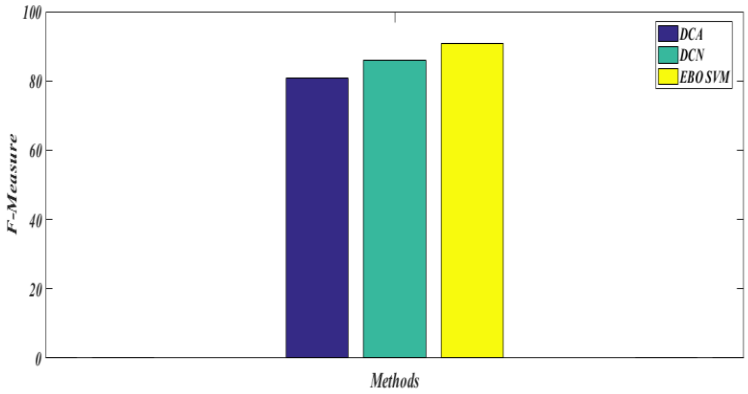

Fig.10: Performance analysis of F-Measure

The F-measure analysis of proposed SVM-EBO method with existing Difference of Convex functions Algorithm (DCA) and difference of Convex Neural-network DCN methods are shown in Figure 10. In this evaluation clearly shows the proposed SVM-EBO obtained good F-measure compared with other methods. The F-measure of DCA, DCN and SVM-EBO are 79%, 82%c and 91%.

The Accuracy analysis of proposed SVM-EBO method with existing Difference of Convex functions Algorithm (DCA) and difference of Convex Neural-network DCN methods are shown in Figure 11. In this evaluation clearly shows the proposed SVM-

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

EBO obtained high accuracy compared with other methods. The accuracy of DCA, DCN and SVM-EBO are 81%, 83%c and 92%.

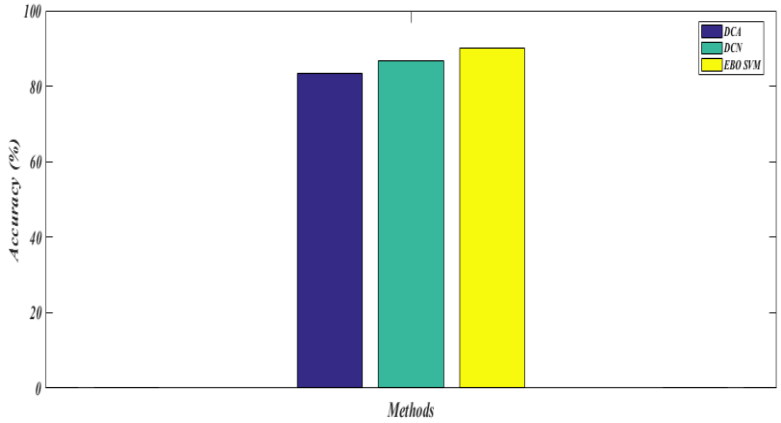

Fig.11: Performance analysis of Accuracy

5. Conclusion

The processing of the raw medical data is used to detect long-term abnormalities in the early stages of people's lives to save them from heart disease. Artificial Intelligence (AI) is used to process the raw data and provide a new approach to heart disease detection. To provide prediction results to the user, employing artificial intelligence, a prediction system algorithm is very essential in heart disease. In the latest development technologies, SVM-EBO in heart disease prediction using the proposed algorithm, it gives reliable output based on the input provided by the user. Management is required to prevent and treat universal and comprehensive system cardiovascular diseases for recording data. Medical records are the most important data classified as one of the technologies to be processed conveniently and quickly. The goal of the current methods to improve the health policy is to provide a classification system for treating cardiovascular disease as against the prevention of cardiovascular disease. These features run the test of heart diseases that are used as inputs to these input samples. Test results of the proposed algorithm SVM-EBO shows that it has a superior performance compared to conventional algorithms. The time complexity, precision, recall, F-measure and Accuracy of SVM-EBO method are 26Sec, 91%, 86%, 91%, 92%. The SVM-EBO method gives good accuracy but is consume more memory, so in future introduce sequential learning neural network method to reduce the memory consumption and enhance the overall performance of heart disease prediction.

References

[1] Gupta, V., Mittal, M. and Mittal, V., 2020. Chaos theory: an emerging tool for arrhythmia detection. Sensing and Imaging, 21(1), pp.1-22.

[2] Shah, D., Patel, S. and Bharti, S.K., 2020. Heart disease prediction using machine learning

ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

techniques. SN Computer Science, 1(6), pp.1-6.

[3] . Stapleton, L., Zhu, Y., Woo, Y.P.J. and Appel, E., 2020. Engineered biomaterials for heart disease. Current opinion in biotechnology, 66, pp.246-254

[4] Afzal, M., 2021. Recent updates on novel therapeutic targets of cardiovascular diseases. Molecular and Cellular Biochemistry, 476(1), pp.145-155.

[5] Diwakar, M., Tripathi, A., Joshi, K., Memoria, M. and Singh, P., 2021. Latest trends on heart disease prediction using machine learning and image fusion. Materials Today: Proceedings, 37, pp.3213-3218.

[6] Hasan, N. and Bao, Y., 2021. Comparing different feature selection algorithms for cardiovascular disease prediction. Health and Technology, 11(1), pp.49-62.

[7] Mohan, S., Thirumalai, C. and Srivastava, G., 2019. Effective heart disease prediction using hybrid machine learning techniques. IEEE access, 7, pp.81542-81554.

[8] Ahmed, H., Younis, E.M., Hendawi, A. and Ali, A.A., 2020. Heart disease identification from patients' social posts, machine learning solution on Spark. Future Generation Computer Systems, 111, pp.714-722.

[9] Babatunde, O.A., Olarewaju, S.O., Adeomi, A.A., Akande, J.O., Bashorun, A., Umeokonkwo, C.D. and Bamidele, J.O., 2020. 10-year risk for cardiovascular diseases using WHO prediction chart: findings from the civil servants in South-western Nigeria. BMC cardiovascular disorders, 20(1), pp.1-10.

[10] Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F.J.M., Ignatious, E., Shultana, S., Beeravolu, A.R. and De Boer, F., 2021. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access, 9, pp.19304-19326.

[11] Fitriyani, N.L., Syafrudin, M., Alfian, G. and Rhee, J., 2020. HDPM: an effective heart disease prediction model for a clinical decision support system. IEEE Access, 8, pp.133034-133050.

[12] Balwan, W.K. and Kour, S., 2021. Lifestyle Diseases: The Link between Modern Lifestyle and threat to public health. Saudi J Med Pharm Sci, 7(4), pp.179-84.

[13] Ahmad, G.N., Fatima, H. and Saidi, A.S., 2022. Efficient Medical Diagnosis of Human Heart Diseases using Machine Learning Techniques with and without GridSearchCV. IEEE Access.

[14] Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V. and Nappi, M., 2021. Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE access, 9, pp.39707-39716.

[15] Shah, D., Patel, S. and Bharti, S.K., 2020. Heart disease prediction using machine learning techniques. SN Computer Science, 1(6), pp.1-6.

[16] Ali, M.M., Paul, B.K., Ahmed, K., Bui, F.M., Quinn, J.M. and Moni, M.A., 2021. Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. Computers in Biology and Medicine, 136, p.104672.

[17] Ali, F., El-Sappagh, S., Islam, S.R., Kwak, D., Ali, A., Imran, M. and Kwak, K.S., 2020. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Information Fusion, 63, pp.208-222.

[18] Javeed, A., Zhou, S., Yongjian, L., Qasim, I., Noor, A. and Nour, R., 2019. An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection. IEEE Access, 7, pp.180235-180243.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 09-2022
DOI 10.17605/OSF.IO/39QPN

[19] Guo, C., Zhang, J., Liu, Y., Xie, Y., Han, Z. and Yu, J., 2020. Recursion enhanced random forest with an improved linear model (RERF-ILM) for heart disease detection on the internet of medical things platform. IEEE Access, 8, pp.59247-59256.

[20] Rani, P., Kumar, R., Ahmed, N.M. and Jain, A., 2021. A decision support system for heart disease prediction based upon machine learning. Journal of Reliable Intelligent Environments, 7(3), pp.263-275.

[21] Saqlain, S.M., Sher, M., Shah, F.A., Khan, I., Ashraf, M.U., Awais, M. and Ghani, A., 2019. Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines. Knowledge and Information Systems, 58(1), pp.139-167.

[22] Ismail, M., Vardhan, V.H., Mounika, V.A. and Padmini, K.S., 2019. An effective heart disease prediction method using artificial neural network. International Journal of Innovative Technology and Exploring Engineering, 8(8), pp.1529-1532.

[23] Ashish, L., Kumar, S. and Yeligeti, S., 2021. Ischemic heart disease detection using support vector machine and extreme gradient boosting method. Materials Today: Proceedings.

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

[25] Alfaidi, A., Aljuhani, R., Alshehri, B., Alwadei, H. and Sabbeh, S., 2022. Machine Learning: Assisted Cardiovascular Diseases Diagnosis. International Journal of Advanced Computer Science and Applications, 13(2).