Extended Firefly Prediction Model for Prognosis of Heart Disease

Jyoti Thakur
Student
Dept. CSE
Sri Sai University Palampur (H.P), India

Munish Katoch
Asst.Professor
Dept. CSE
Sri Sai University Palampur (H.P), India

ABSTRACT
In all the diseases, cardiovascular disease or CVD is the main reason for the death all over the world. 1 to 5 in every one thousand persons suffers from the heart disease. Even though advances have been performed to get better surveillance treatment but yet Heart failure diagnosis has been occurred 1.7 years in men and 3.2 years in women. Several techniques have been proposed till now to find the effect of disease at earlier stage but still it is under consideration. Data mining is used for the extraction of significant, meaningful and desired information from the datasets of patients. The different classification algorithms were used in the existing systems for the prediction of heart disease, in which the attributes of data mining are fed. However, it has been analyzed that there is no single classifier that produces best result for dataset and not a single data mining technique that gives consistent results for all types of health related data. Therefore, in this paper, a novel classifier i.e. fa-ANN, is proposed that can provide the optimal results for the healthcare data than other classifiers in terms of accuracy, precision and recall.

Keywords
ANN, CVD, KNN, SVM, RF.

1. INTRODUCTION
Heart is a significant human body organ whose output assesses the fitness or inadequacy of an individual. To see that an individual's life completely relies on how effectively the heart works. In case of an issue in the heart's working, brain, kidney or other components of the body are immediately affected. The heart's primary function is to simply pump blood into the body. Heart conditions therefore refer to heart disease containing the blood vessel system.

Also, the rate of diseases is enhanced with the advancement in technology. In various nations, diseases of heart attack are the leading cause of mortality; however, untimely detection can stop this attack and allow patients to survive longer. Fifty per cent of all fatalities owing to heart disease in Europe have been studied and grow until today as the population of the elderly increases.

Data are collected from medical resources in the forecast of diseases, where there is also concealed information. First, it converts raw data into a dataset on which a number of data mining artificial intelligence is crucial because it can predict accurately and assist healthcare professionals. Different classification methods have been used for modelling and then it decides on the information it obtains.

In the field of healthcare techniques can be used with the help of which the accurate prediction can be made

1.1Classification Techniques
ANN This algorithm is designed to emulate the human brain's neurons. It comprises of some linked nodes or neurons, and one node's output is another. Each node gets several inputs, but only one output value [1].The MLP consists of an input layer, hidden layers, and an output layer and it is a widely used type of ANR. During distinct circumstances, separate numbers of neurons are allocated to every layer.

K-nearest neighbors (KNN) This algorithm forecasts a fresh example on the basis of the majority of votes from its nearest neighbors. It calculates the distance of an attribute from neighbors with Euclidean distance. The K-nearest neighbors (KNN) is a straightforward, parametric, lazy classification system [2]. If all characteristics are constant, KNN is preferred.

Support vector machine (SVM) The classification precision of the algorithm is helpful. It is described as a vector space of the finite dimensions, consisting of a dimension for each object feature. A vector support machine is a model form used for data analysis and discovery of patters in the regression analysis. When your information has precisely two classes, support vector machine (SVM) is used [3].

Random forest (RF) One of the most efficient classification approaches is the Random Forestry Algorithm [4]. The forecast and the estimate of probability have been using the RF algorithm. The RF comprises of a large number of decision-making trees. In 1995, the Tin Kam HO of bell laboratories suggested random forest item for the first time. RF combines bagging and random features selection.

1.2 Data Set
The data set is taken from the UCI. There are 76 characteristics in this database, but all published studies refer to the use of a subgroup of 14. The only data base used by the ML researchers to date is in specific the Cleveland database. The field of "objective" relates to the patient's presence of heart disease. It is rated integer between 0 (no presence) and 4. Cleveland experiments focused on just trying to differentiate existence (values 1, 2, 3, 4) from lack (value 0).

Only 14 attributes used:

Table 1: Attribute Information

1	Age	Age in years
2	Sex	Male or female
3	cp	Chest pain type
4	thestbps	Resting blood pressure
5	chol	Serum cholesterol
6	Restecg	Resting electrographic results
7	fbs	Fasting blood sugar
8	thalach	Maximum heart rate achieved
9	exang	Exercise induced agina
10	Oldpeak	ST depression induced by exercise relative to rest
11	solpe	Slope of the peak exercise ST segment
12	ca	Number of major vessels colored by floursopy
13	thal	Defect type
14	Num	diagnosis of heart disease (angiographic disease status)

2. LITERATURE REVIEW

The number one problem in the world is heart disease. Heart condition occurs in the first heart attack more than fatalities of individuals. However, some problems for breast cancer, lung cancer, and ventricle have not been attacked only for a heart attack. The framework research must be done to acknowledge the incidence of cardiovascular disorders instantaneously in thousands of samples. In **[5],** the author accessed the forecast of heart disease for the potential for nine (9) classification methods. The native Bayesian neural network, namely decision tree SVM, ANN, and KNN. In the forecast of cardiovascular disease, the author had proposed an algorithm of the SVM (support vector machine). Health care profiles like age, gender, blood pressure, form of chest pain, blood sugar fasting was used. It can predict how patients become cardiovascular.

The paper [6] used the classification methods for the forecast of cardiopathy. In this paper, various authors had proposed the methods for prediction of cardiovascular diseases.

In paper [7], the author had created an anticipatory system for cardiac diseases that can help medical experts to predict the status of cardiac disease based on patient clinical information. There are three steps to the presented approach. First, the author selected a total of 13 clinical characteristics such as age, trastorne, gender, type of chest pain, cholesterol, maximum heart rate, fasting blood sugar, ecg-relieving, angina-induced exercise, ancient peak, colored vessel, slope and thal. After that, the author had created an algorithm based on these clinical characteristics for classifying neural cardiac diseases. The forecast precision is close to 80%. Lastly, a user friendly scheme for predicting heart disease (HDPS) was developed by author.

In research [8], the author had analysed several algorithms that play a crucial role when identifying or predicting heart disease (such as the PSO, genetic algorithm, Artificial Neural Network, etc.). In the presented paper, the author explained first of all the fundamental ideas of these three algorithms and evaluated how they help to predict cardiac diseases.

Recently, Mamatha Alex P and Shaicy P Shaji in [9] followed 'Data mining' technique in which attributes are fed in to KNN, SVM, Random forest, and ANN classification Algorithms for the heart diseases' prediction. In this paper the authors assessed various classification techniques in heart disease diagnosis. It can perform as reliably as in the diagnosis of heart disease. Early detection of heart disease is the primary benefits of this study. It can be diagnosis correctly on time, providing treatment with reasonable cost.

3. PRESENT WORK

In the existing system the researchers worked on different classifier techniques such as SVM, ANN, KNN, Random forest etc, among which ANN is the most efficient one with the help of which the accurate prediction can be made. However, they did not provide the efficient results for every type of data set that will affect the diagnosis process.

Many researchers have been training and testing their classificatory on the renowned UCI heart disease dataset and others have used information from affordable hospitals.

Finally, machine-learning can assist health experts as well as patients in diagnosing cardiovascular diseases.

4. PROPOSED WORK

As per the statement defined in the aforementioned section, it is observed that there is a provision to update the existing system. In the existing system the classifiers are used for accurate prediction. However, it has been analyzed that there is no single classifier which gives best result for every dataset and not a single data mining techniques which provide accurate results for cardiovascular disease related data.

Thus, an algorithm is required which can provide optimal result. For this, the proposed methodology will have a classifier which will hybridize with the optimization technique named, firefly.

This is inspired from the literature survey in which it is explained that optimization algorithm can improve the factors of the classifiers to improve their classification or prediction property. For instance, in neural network if initial weights of the network are updated, the classification accuracy is also going to vary. So those weight values can be optimized using different optimization algorithms.

These findings motivate the proposed work to provide a hybrid model as best of classifier achieved (ANN) with traditional approach along with the firefly optimization (fa-ANN).

5. METHODOLOGY

The entire methodology can be summarized with the following flow chart pictographically.

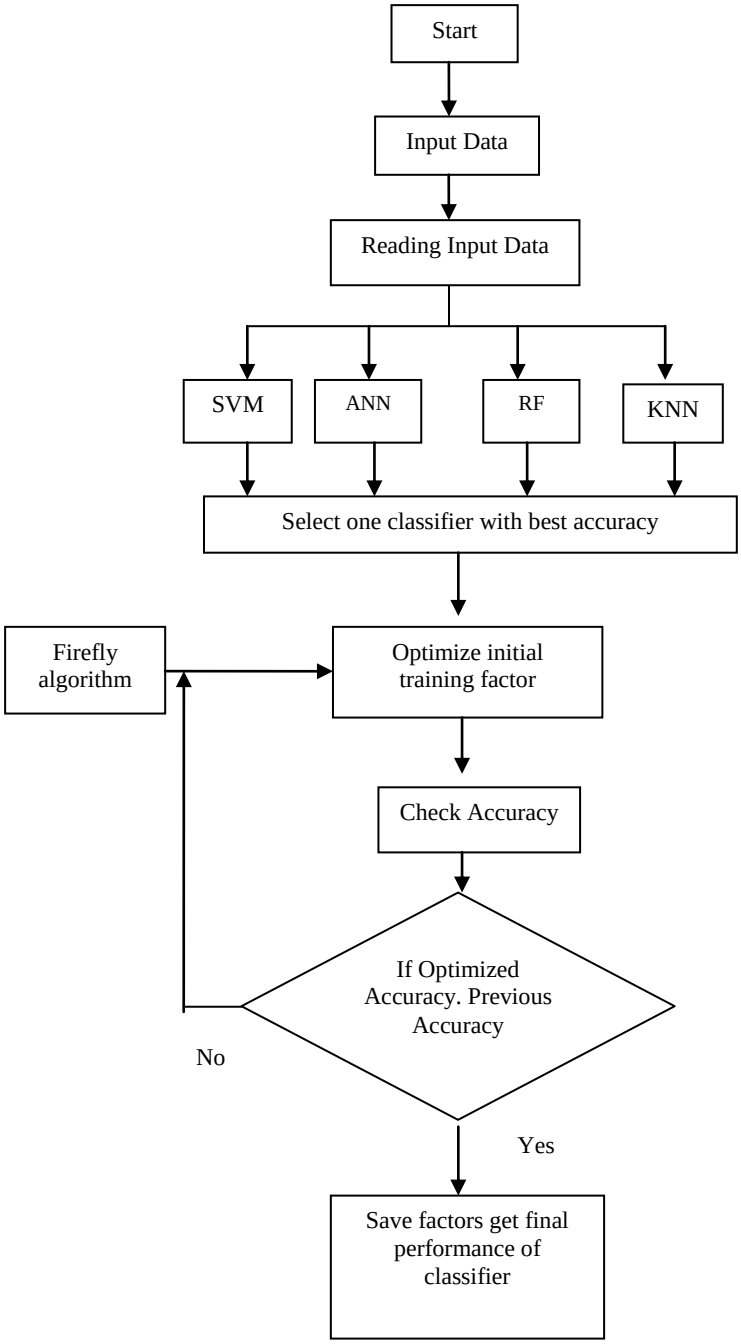

Figure 1: Workflow of proposed approach

Description of flow chart:
Step 1: Collection of input data from UCI to train the software.

Step 2: Read the input data.

Step 3: Applying four different classifier algorithms namely, SVM, ANN, KNN, and Random Forest.

Step 4: Outputs given from the results of four algorithms.

Step 5: Select the best classifier in terms of accuracy.

Step 6: Apply firefly optimization for selected classifier factor updation.

Step 7: Retrain the network to improve accuracy.

Step 8: Get final results from proposed model.

Step 9: Comparison with traditional classifier.

6. RESULT AND DISCUSSION

In this section, the comparison results of various existing classifiers with the proposed one i.e. fa-ANN are represented. The various classifiers i.e. SVM, ANN, RF, KNN and fa-ANN are compared in terms of three parameters i.e. Accuracy, Precision and Recall and the obtained results are discussed below:

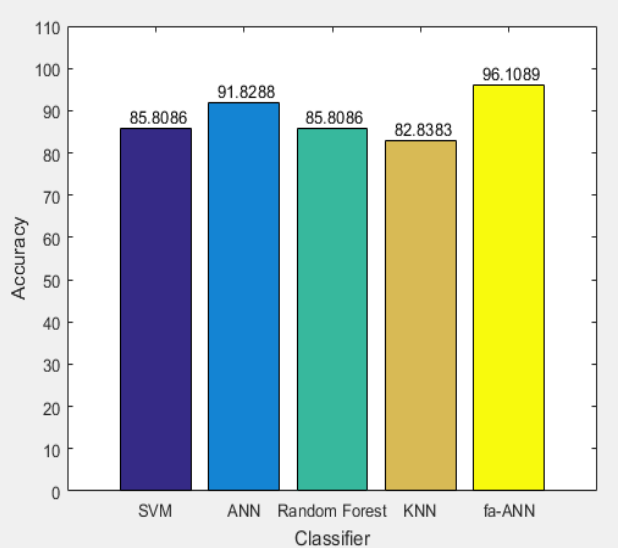

Figure 2: Accuracy value of different classifiers

The graph in figure 2 depicts the comparison analysis of various classifiers with respect to accuracy parameter. In the graph, the y-axis calibrates the value of accuracy and x-axis represents the different classifier i.e SVM, ANN, RF, KNN, fa-ANN. The obtained results demonstrates that proposed classifier i.e. fa-ANN classifier has the highest accuracy rate in contrast to all other classifiers 96.1089.

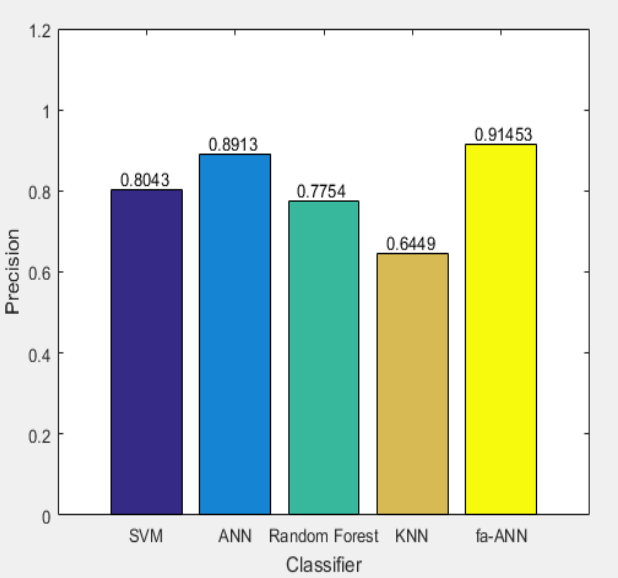

Figure 3: Precision value of different classifiers.

The results obtained on performing the comparison analysis of various classifier with respect to precision, is exemplified in the graph of figure 3. The precision value and various classifiers are represented along y-axis and x-axis respectively. It is clearly observable from the graph that KNN has the lowest precision i.e. 0.6449 and fa-ANN has the highest precision value i.e. 0.91453 which implies that fa-ANN is the most efficient classifier than other classifiers.

The recall value of different classifiers is illustrated in figure 3. In graph, the y-axis calibrates the recall value which varies from 0 to 1.2 and x-axis shows different classifiers. From the graph, the sequence of various classifiers in terms of recall value can be depicted i.e. fa-ANN>KNN>ANN>RF>SVM which implies that fa-ANN classifier has the highest recall value and SVM classifier has the least recall.

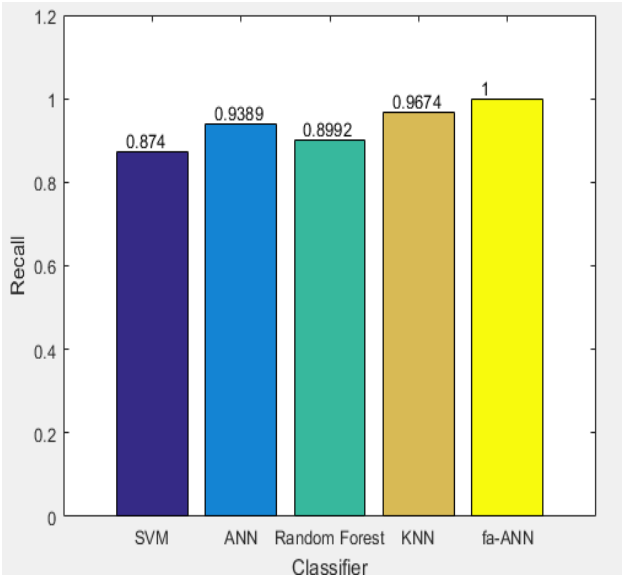

Figure 4: Recall values of different classifiers

The values of the various conventional classifiers and proposed classifier in terms of three parameters i.e. accuracy, precision and recall is represented in the table 1.

Table 2: Comparison values of different classifiers with respect to different parameters:

Classifiers	Accuracy	Precision	Recall
SVM	85.8086	0.8043	0.874
ANN	91.8288	0.8913	0.9389
RF	85.8086	0.7754	0.8992
KNN	82.8383	0.6449	0.9674
Fa-ANN	96.1089	0.91453	1

The above table illustrates the values of proposed and existing classifiers with respect to accuracy, precision and recall parameters. It is clearly comprehensible from the table 2 that the proposed classifier i.e. fa-ANN has the highest accuracy, precision and recall in contrast to all the other classifiers. Therefore, it demonstrates that the proposed classifier is the most efficient one as it outperforms the other classifiers in terms of various parameters.

7. CONCLUSION AND FUTURE SCOPE

Heart diseases are the most common cause of death across worldwide. It is required to predict the heart disease at the early stage so that it can be prevented. It has been analysed from the previous studies that among various classification techniques such as, SVM, ANN, RF, KNN, an efficient performance is achieved by using ANN. However, it is not efficient for every type of dataset. Therefore, in this paper, a novel classifier is proposed in which the ANN classifier is hybridized with the optimization technique named, firefly, in order that the optimal results can be achieved for the heart disease data. The proposed classifier (fa-ANN) is contrasted

with the existing classifiers and three parameters are considered for performing the comparison analysis i.e. accuracy, recall and recall and it is demonstrated from the obtained results that the proposed classifier is efficient one and provides the optimal results in terms of accuracy, precision and recall. These classification algorithms can be further improved by increasing the number of attributes in it with the help of which more accurate prediction can be done. Many possible improvements to enhance the scalability and precision of this forecast scheme could be studied.

8. REFERENCES

[1] Al-Janabi Mahmoud ,H.Qutqut, Mohammad Hijjawi, "Machine learning classification techniques for heart disease prediction: a review", vol. 7, 2018

[2] Jabbar MA, "Prediction of heart disease using k-nearest neighbor and particle swarm optimization", Volume 28, Issue 9,

[3] Mythili T., et al, "A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL)", vol. 68, 2013

[4] .A.Jabbar, "Intelligent heart disease prediction system using random forest and evolutionary approach", Volume 4 (2016) pp. 175-184

[5] C. Sowmiya and P. Sumitra, "Analytical study of heart disease diagnosis using classification techniques," 2017, pp. 1-5

[6] Shafquat Perween , Suraiya Parveen , "Analysis of Heart Disease Prediction Techniques", Volume 7, Issue 4, April 2018

[7] H. Chen, et al , "HDPS: Heart disease prediction system," 2011 Computing in Cardiology, Hangzhou, 2011, pp. 557-560

[8] S. Bharti and S. N. Singh, "Analytical study of heart disease prediction comparing with different algorithms," 2015, pp. 78-82.

[9] Mamatha Alex P and Shaicy P Shaji, "Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique", 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0848-0852.

9. AUTHORS PROFILE

Jyoti Thakur. She received B.Tech degree in Computer Science Engineering from the Sri Sai University Palampur Himachal Pradesh in2017. Currently she is pursuing in M.Tech in Computer Science Engineering from the Sri Sai University Palmpur.

Munish Katoch. He received his B.Tech degree in Computer Science Engineering from the Lovely Professional University. He also received his M.Tech degree in Computer Science Engineering from Lovely Professional University. Now he is working as Assistant Professor in Sri Sai University Palampur