An Efficient System for the Prediction of Coronary Artery Disease using Dense Neural Network with Hyper Parameter Tuning

Debabrata Swain, Santosh Kumar Pani, Debabala Swain

Abstract: ***Diagnosis of heart disease is considered as one of the challenging problems in medical science in the current decade. Coronary artery disease is a type of heart disease in which the arteries of the heart gets affected. Hence many researchers propose a number of intelligent solutions to improve the predictability towards the identification of Coronary artery disease. If the disease can be identified at an early stage, then precautions can be taken for its recovery. In the proposed system, an efficient deep learning technique is used for improving accuracy towards the identification of the disease. The proposed system is built using a Dense Neural Network which is a type of deep learning network. Here the experimentation is done using Cleveland Heart disease data set present in the UCI repository. The system has three stages. In the first stage data cleaning and feature selection is performed. In the second stage model training is done using hyper parameter tuning. In the last stage, the trained model is used for prediction of coronary artery disease using test data set. The proposed model results in the classification accuracy of 96.03% during training and an accuracy of 94.91% during testing, which is best among all the discussed methods.***

Keywords: ***KNN method, Random forest, Fuzzy logic, SVM, PCA, ANFIS***

I. INTRODUCTION

Present-days heart disease is treated as one of the major cause of human death in the world. 10% of the total death occurs in the world is due to heart disease only. Hence the disease has become one of the biggest concerns in various countries of the world. As per Japan death rate statistics, heart disease occupies the second position [1]. Due to the improvement of technology and the availability of automation, people perform very less physical work and use the mental ability which makes them prone to get heart disease. Due to this people are getting addicted to smoking, alcohol which leads to have big bellies. As per University of Rochester's Medical centre view the major source for Heart disease are overweight, lack of physical activity, fatness, consumption of malnutrition and tobacco [2]. As heart disease is widely accepted as the major source of death hence medical analysis of heart disease becomes a regular need for every human being.

Due to the number of ambiguity and risk factor the prediction of the disease became a very tough task for every physician. If the heart attack can be identified earlier then the life of the patient can be saved through proper medication and also harm to the heart can be saved up to a large extent [3]. Heart diseases are of various types like Coronary artery disease, Valvular heart disease, Cardiomyopathy. These diseases mainly affect the arteries of the heart, blood in and out valves of heart, heart muscle squeezing [4]. Proper heart functioning is really a highly essential thing for a healthy life. In Coronary artery disease cholesterol, calcium and some other substance getting deposited in the veins through which blood circulation is done. As a result of which some blockage is being created against the smooth passage of blood. Due to this the heart muscles will not get an adequate amount of oxygen which creates discomfort in the patient's chest and results as chest pain with the patient [5]. As per WHO report up to 2030 around 23.6 million people in the world will die due heart disease. So there is a need to take some preventive steps to minimize the threats of heart disease. Practitioners mainly view the symptoms, expressions and medical test to identify the occurrence of the disease with the patient [6,13]. For coronary artery disease identification doctors mainly uses SPECT and ECG methods. In SPECT method radioactive tracers are being injected in to blood for generating images of heart which are used by the doctors to know about the identification of coronary artery disease and also the prediction of the heart attack. ECG reports are used to know about abnormality in heart beating [7]. The diagnosis made by the doctors about any disease is not always 100% correct. Hence various computerised tools are used in the healthcare domain. These tools are used to identify critical parameters for the diagnosis of the disease [8].Improvement in the health condition of any patient can be known by analysing various critical parameters related to their disease. The main goal of the intelligent systems assisting medical diagnosis is to predict the presence of any disease accurately [9,12]. A number of input symptoms are used to indicate the occurrence of Coronary artery disease. Out of all age, sex and family medical history cannot be changed. But others symptoms like smoking, blood pressure, cholesterol level, physical exercise can be changed to reduce the possibility of Coronary artery disease. As so many parameters are involved while diagnosing the disease so the practitioner uses the present medical report of the patient.

Revised Manuscript Received on December 22, 2018

Debabrata Swain, PhD Scholar, School of Computer Engineering, KIIT University Bhubaneshwar

Santosh Kumar Pani, Associate Professor, School of Computer Engineering, KIIT University Bhubaneshwar

Debabala Swain, Associate Professor, Computer Science Department, Ramadevi women's University, Bhubaneshwar

After going through the report doctors adopt the same method which was used for any previous patient having a similar type of test report [10,14]. Some modern genetic algorithms now days are trying to find out some important data which are mainly contributing towards the occurrence of the heart disease instead of analysing a number of data. Through this the algorithms trying to identify the disease with an optimal number of parameters and consuming very less time [11].

Recent development in the field of Neural Network contributes a lot to medical science towards the development of intelligent systems. In last few decades various computational systems have developed which were helpful for the physicians in improving their diagnosis decisions. In the motivation towards the requirement of an expert system here an efficient heart disease prediction system is proposed. In the proposed system Dense Neural Network is used which is a type of Artificial Neural Network. Dense Neural Network is a type of feed forward artificial neural network where some hidden layers are added as intermediate layers between the input and the output layer. Here every node is connected to every other node in the next layer.

All neurons have a linear activation function which linearly maps input weights to the output of each neuron. The same process is continued up to the output layer to obtain the final output. The remaining portion of the paper discusses various heart disease prediction systems used by various researchers, proposed system and performance analysis of various discussed techniques with the proposed system in terms of accuracy.

The goal of the proposed system is –

- Understanding the data findings like finding out outliers, anomalies and data imputation in data set.
- Building an accurate predictive model by tuning hyper parameters in the cost function to avoid over fitting and under fitting.
- Finding the optimal value of features by minimizing the cost function through which the accuracy of the model can be improved.

II. LITERATURE REVIEW

Alberto Palacios Pawlovsky et al [1] has used an ensemble based KNN method which employs distance based heart disease prediction. Here a two stage method has been used. In the first stage different K values are chosen. For each K point we put all the 5 different distance formula like Euclid, Manhattan, Chebyshev, Canberra and Mahalanobis and obtained distance value from the test instance and noted the class of each k neighbour. We also found the classification accuracy by taking different cross validation size 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% of the total records. In the second stage of the algorithm we put majority based voting algorithm to assign the class to an unknown instance as per the majority class is chosen.

Yeshvendra K. Singh et al [2] used Random Forest method for detection of heart disease. The system uses all the 13 important input features for the prediction of heart disease provided UCI repository. The prediction system is implemented by removing the features between whom no correlation can be established. The enhancement in the accuracy is achieved by tuning various linearly dependent variables of a random forest like randomness, the number of trees, the minimum number of splits and the minimum number of leaf nodes. Initially the number of trees and the minimum number of splits are considered to find a better correlation with accuracy. Highest accuracy obtained when numbers of splits are 20 and the number of trees are 75.

TanmayKasbe et al [3] have employed an expert heart disease prediction system using fuzzy based logic. Here a fuzzy indicator functions like triangular and trapezoid are used for the implementation of the fuzzy expert system and fuzzy rule base. The fuzzy expert system first does categorization of independent features and dependent feature. In this stage, the features are experimented to observe its value range and its equivalent class. In the next stage Fuzzy rule base over data is employed by the different combination of single or several features with AND,OR operator. Here the system development is done using total 86 fuzzy based rules with all possible arrangements. In the rule base, all the possible input variables with different combinations and its corresponding output value are used for the output level calculation. A relationship is established between all independent features value and their corresponding dependent feature values are set. This relation will be helpful to find out the class for an unknown instance.

Purushottam et al [6] used an efficient heart disease prediction system which uses a large dataset from Cleveland which contains data about coronary diseases. First, the database pre-processing is done using all Possible-MV algorithms. It is used mainly used to fill the missing values in the data set. After that classification decision rules are generated to do the proper classification through original rules, pruned rules, and rules without duplicates, classified rules and polish.

B.Subanya et al [7] used Feature selection which is used to filter the dataset. It is used to select the required attributes from the dataset. In medical diagnosis, it is necessary to remove the redundant features. Heuristic techniques help to select the best features. Artificial Bee Colony(ABC) is a meta heuristic algorithm. It selects feasible solutions by sharing the information between the bees in the population. ABC follows a unique solution update mechanism. Due to this, the results converge to the optimal solution swiftly. Its fitness is judged by the SVM classification. The ABC-SVM uses the Wrapper Technique for classification.

Saba Bashire et al [8] used an ensemble based method for the prediction of heart disease. First data cleaning and data division is done in to training and testing set. Then model training is done using various classifiers (Decision Tree, Naive Bayes, Support vector machine). Here various types of classifiers are used in the model to increase the competence of the system. The classification rules of different classifiers are defined during the time of training of the model which will be accumulated as knowledge of the system. Then testing data feed to the classifiers for classification and class of any test instance is identified using a majority voting algorithm. Finally the unknown tuple assignment is done to the highest voted class.

Bhuvaneswari [9] used Principal Component Analysis and Adaptive Neuro Fuzzy Inference System to develop an efficient system for the identification of heart disease. First the data from the dataset are pre-processed. After that feature reduction is performed to find only useful and independent parameters for the enhancement of system performance using principal component analysis. The number of input parameters selected by PCA are 7. And in the next phase the reduced resultant data used for classification with the help of Adaptive Neuro Fuzzy Inference System which includes 5 layers.

Resul Das et al [10] used SAS based software 9.13 which consists of 2 different programs First the SAS enterprise guide 4.3 used for processing of data. Second the SAS enterprise miner 5.2 used for analysing and recognizing heart disease by combining several neural networks with ensemble nodes. The system consists of 5 components like heart database component, Data partitioning component, Variable selection component, neural network component and ensemble component. The database component holds the parameters list to classify between healthy and sick person. The partitioning component is responsible for dividing the input data into a training and validation set. The Variable selection component is used for filtering out the unrelated input parameters and rejects them in successive nodes in the process. The neural network component is responsible for doing the prediction using the related features. Three types of neural network models are used to form the component. The multi-layer feed forward network is used for the forecasting. The Ensemble based component is used to form a new classifier model after combing the output of all previous classifiers.

Sumit Bhatia et al[11] developed a heart disease prediction system using Simple Support vector machine and Integer Coded Genetic Algorithm. Integer Coded Genetic algorithm is used to select vital and related parameters and rejecting the unrelated and dependent parameters. Out of total M parameters N parameters are used to form 50 chromosomes. Each chromosome consists of input parameters array. After that tournament selection method is used to find the winner which will be chosen for crossover operation. In tournament selection method winner is selected by comparing classification accuracy of two consecutive chromosomes. Then one point crossover is performed on two consecutive chromosomes after interchanging randomly generated crossover sites. After crossover operation is done mutation operation is performed to generate the next generation of population of a chromosome by arbitrarily altering the gene sequence with a certain probability. The repeated genes are removed from a chromosome using Remove repetition operation. Finally the population prior crossover and post mutation are arranged according to their fitness value. Now top half individuals are considered for the creation of the next generation. This process will repeat for 25 iterations to generate the optimum set of parameters. After this Simple Support vector machine prediction started by changing the penalty term to achieve the best classification accuracy. Here an RBF kernel and polynomial kernel performance is tested with different parameter values. At last the model outperforms with RBF kernel at parameter value=0.025 and penalty parameter=150.

III. DESCRIPTION OF DATASET

The data is collected from the Cleveland heart disease data set available in the UCI repository [15]. It is having various inputs feature which are used for the identification of heart disease. The dataset consists of 303 records and 76 features. Out of 76 features 14 are identified as important features. Out of 14 features in the dataset 13 are independent features and 1 is dependent feature which specifies the status of the disease (Present or absent). Following table provides the details of the attributes. The parameters which are of the discrete type have some categorical or classical values. Like in case of sex values can be male or female. But for continuous or numeric or real type there is no limit of values for the features. The parameters are mentioned in Table:1.

Table. 1 Dataset Details

R No	Attribute Name	Characteristic
	Age in year	Numeric
	Sex	Discrete
	Chest pain	Discrete
	Resting blood pressure	Numeric
	Serum cholesterol	Numeric
	Fasting blood sugar	Discrete
	Maximum heart rate achieved	Numeric
	Resting ECG results	Discrete
	Exercise-induced angina	Discrete
0	ST depression	Real
1	Slope of the peak exercise ST segment	Discrete
2	Fluoroscopy result	Discrete
3	Thalassemia	Discrete
4	Heart disease	Discrete

IV. DENSE NEURAL NETWORK

Dense Neural Network is one of the most used data mining model. It is used to mine some blueprint in an intuitive and reliable manner and has been significantly implemented to find models which explain data co-relations.

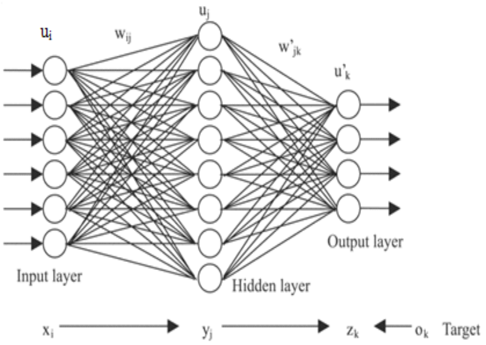

Fig. 1 Architecture of Dense Neural Network

Where u_i,$u_j$$u_k$ are the nodes in the input layer, hidden layer and output layer.
W'$_{ij}$ is the weight of edges between the input layer and hidden layer
W'$_{jk}$ is the weight of edges between the hidden layer and output layer.

A dense neural network consists of one or more layers between the input and the output layer shown in Figure1. Here feed forward neural network is used which is a special distinguished, non-linear regression model. Now a day various predictive systems are designed for real world problems using a neural network as it is able to identify the complex non-linear co-relation between data. The most commonly used feed forward neural network is called back propagation network. Using the back propagation method the error gradient in case of a feed forward network can be calculated. Back propagation method is commonly used by various training methods for computation of their gradient. Using any one of the gradient descent technique generally back propagation network is trained. Gradient descent technique finds the derivative of the cost function with respect to network parameters and then adjusting the parameters in gradient direction.

A Neural Network (NN) can be formed by taking many Processing Elements (PEs), which are also known as neurons and weighted edges among the PEs. Each PE does the weighted sum of its input edges and finds an output value that will send to PEs in the next layer. During training of a Neural Network the weights of the connecting edges are updated to obtain the required output. The Dense Neural Network (DNN) is an approach used to find a solution for data mining based problems. A Neural Network can be created by taking a number of neurons in a strongly connected network. It is available with variety types of structures which show similar behaviour like a biological neural network. A vital feature of the neural network is a continuous learning process in which instances are fed as input one by one to the network and weights assigned with the input are updated. When all the instances are given, the same method repeats again. During this repeated process the network learns by adjusting the weights so that it can forecast the appropriate class for the input instance. These changes are stored as knowledge. When an unknown input is given to a neural network then it produces a new result which is generated using its experience.

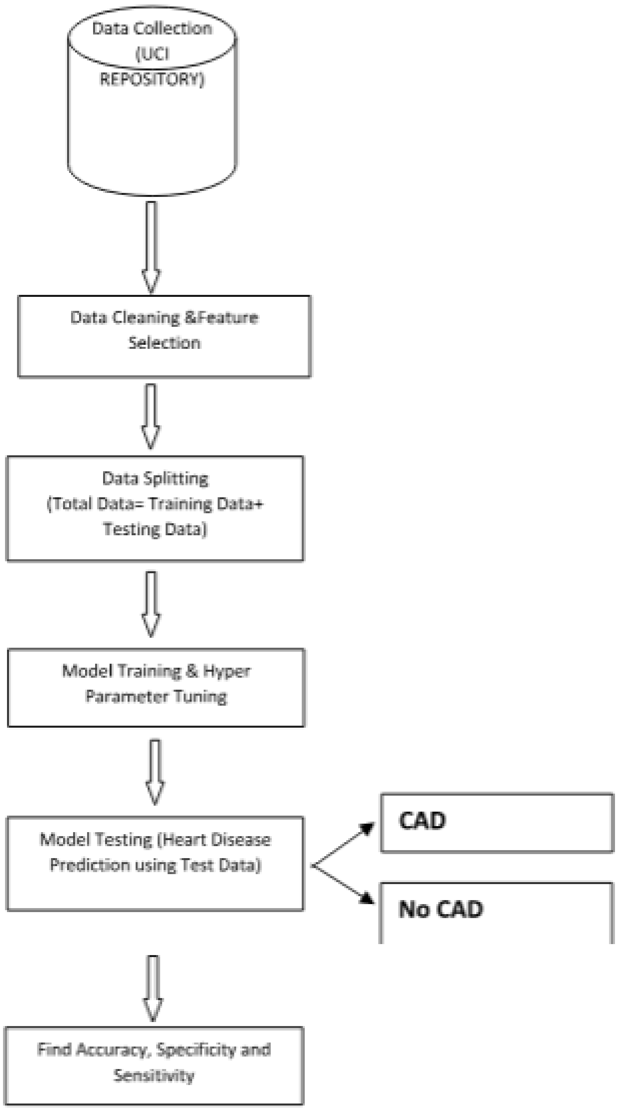

Fig. 2 Proposed System Architecture

V. PROPOSED ALGORITHM

Following steps are used for disease diagnosis using the proposed system shown in Figure 2.

Step 1: Data Collection- In this step we are importing the data from UCI Switzerland, Long Beach and Cleveland data set.

Step 2: Data Cleaning and Feature Selection

This stage we are performing 2 processes.

- Data Cleaning- In this process the whole data set is scanned to find missing values, outliers. Missing values are identified using hit map present in the seaborn library. Outliers are identified using box plot present in matplotlib library. No outliers are found in the data set. The missing values are normally replaced with mean or median or mode. Sometimes the missing value contained rows are also deleted from the dataset to make it clean. The dataset contains total 303 entries. Out of which 6 records are containing missing values which are identified using hit map. After that these rows are removed.

Hence the total number of rows became 297. The status column contains values like 1,2,3,4,5 representing different levels of heart disease. But as in the proposed system binary classification is planned for the disease prediction so all the output levels are converted from 1,2,3,4,5 to only 1 which indicates the presence of the disease.

- Feature Selection – It is a process of selecting useful and important features out of all features present in a dataset. Here the backward feature selection method is used to find the important features. Logistic regression model is used for this purpose. Initially all the 76 features were considered. Gradually each feature were removed which are not contributing towards the prediction of the model. For the purpose of feature selection model accuracy, F1 score, precision and false negative rate are considered.

Step 3: Data Splitting

After getting the clean data from the previous stage data splitting is done. Here for model training 80% of data (238 records) are used and for model testing 20% of data (59 records) are used.

Step 4: Training

Here Dense Neural Network classifier's training is performed. The training data set is provided to the model for training with hyper parameter tuning. The hyper parameter tuning is used to get the best values of the parameter which outperform on the training set and gives good accuracy on the testing data. 10-Fold cross validation is carried out to evaluate the model during its training.

Step 5: Prediction

Here the testing data is given to find the predictability of the trained model. In the proposed model DNN classifier is used to categorize the test data. The categorization is done into 2 classes like patients' having or not having coronary artery disease (1=Presence, 0=Absence).

Step 6: Find the accuracy, recall and precision of the model.

VI. MODEL DETAILS AND ACTIVATION FUNCTION

The size of the input layer is taken as 13 because the total numbers of input features are 13. The size of the hidden layer is taken as 20. It is mainly done for better computation, reducing the complexity and improving accuracy. As Binary classification is used in the model hence the size of the output layer is taken as 2.

Every layer of the dense neural network contains one or a few neurons which directly connected with the neurons form the last and the next layer. Every neuron of the dense neural network is of a similar type. All neurons are having some input edges and some output edges. The value of any neuron of a layer is calculated after applying activation function on the output of previous layer neurons connected to it. The output is obtained by the sum of bias term and accumulated sum of the product of the value of all connected neurons and weight assigned to their edges. The activation function used here is g.

In other words, given the outputs Pj of the layer k, the outputs Qi of the layer k+1can be calculated as shown in Eqn 1 & 2.

Here Hyperbolic Tangent activation function is used. The main job of the activation function is to transform input in the range of (- to +) to (0 to +1). As the goal of the system is to classify the target variable in to 2 categories hence the target value is converted to either 0 or +1 which is demonstrated in the below graph Figure 3 & Eqn 3.

Hyperbolic Tangent Where A & B are referred as free parameters.

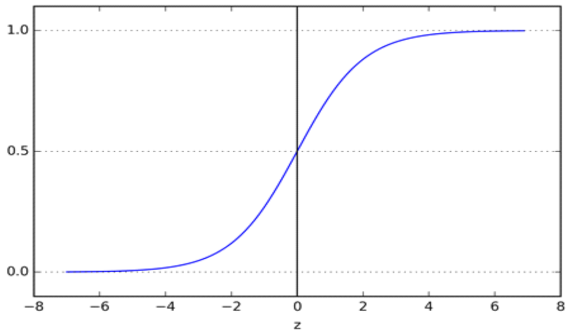

Fig. 3 tanh(x) Graph

In classification problem Error means how the predicted output varies from the actual output of the data set. So Error can be calculated by using the following Eqn 4 & 5.

Where, = Hyper parameter which needs to be tuned, = Cost function

Hyper parameter Tuning

It is the problem of selecting a group of optimal hyper parameters for a learning algorithm. Hyper parameter tuning functions by running several trails in a single training job. Each trail is a complete execution of training application. These parameters are set with some specified values. During training it keeps track of the result of each trail and does modifications for successive trails. After all the trails are over the most effective configuration of values according to specific criteria are known from the summary of the trails. Hyper parameter tuning has to deal with 2 types of issues.

- When =0 it means the Cost function (is not getting protected against over fitting. So there is a possibility of over fitting.
- When is too large (e.g. when =100), there is a possibility of under fitting.

Hence biased term will dominate. Therefore - parameters need to tune to avoid under fitting and over fitting to get the best values of unknown parameters (that perform well for the training set. Here the value of is taken as 38.

Cost Function- It is a function which is used to measure the average squared difference between actual and predicted values of a classification model. The cost function used is given in Eqn 6.

)ln(1-(())]+ [

Cost Function Eqn(6)

Here the model is having 3 layers which are an input layer, a hidden layer and an output layer. In order to calculate the kth activation function value (() forward propagation has been used to represents the weight. It computes the gradient (also called partial derivative) of the cost function with respect to all weights in the dense neural network for minimizing the value of the cost function.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Optimization of cost function results in minimizing the average squared difference between actual and predicted values. For this purpose the BFGS algorithm is used. The BFGS algorithm is used by invoking the **fmin_bfgs** function from **Scipy library of Python**.

VII. RESULTS AND DISCUSSION

Table. 2 Cross Validation Score of 10 Folds

Fold No	Accuracy
1	92
2	93
3	94.8
4	95.4
5	96.8
6	96.3
7	97.4
8	98
9	98.6
10	98.4
Average Accuracy	**96.03**

Cross validation

It is a technique of validating a model with rotation. It is a method of evaluating a model for finding how the output of predictive analytics can be used to provide a generic solution for an unknown data set. So during training process the training data are divided in to k parts which are otherwise known as k fold. So out of k folds 1-fold used for validating the model and remaining k-1 fold are used for training the model. Here the value of k is taken as 10. During each time the fold which is used for evaluation changes. The accuracy of all the k-folds is noted and the average is calculated. The accuracy of all the 10- Folds are shown in the following table-2 & Fig-3. After 10 fold cross validation are done the average training accuracy reported by the system as 96.3%.

Testing Accuracy

In this section the performance of the system is evaluated by using the test data set. Out of the total 59 records of test data 56 records are correctly predicted by the system which is given in the following confusion matrix table-3. The efficiency of the system is measured using its classification accuracy, precision and recall.

Table. 3 Confusion Matrix

Total Test Records	59
Correctly Predicted	56
TP (True Positive)	30
TN (True Negative)	26
FP (False Positive)	1
FN (False Negative)	2

Accuracy value is the ratio between correct predictions to the total prediction done by the system. Precision indicates the ratio of total positive predicted records to total correctly predicted records. Recall indicates the ratio of total positive predicted records to the total actual positive records. The accuracy, precision and recall of the system are found as 94.91%, 96.77%, and 93.75% respectively.

Performance Evaluation

The accuracy of any classification model state how it is able to find any generalized solution to classify any unseen data in to any correct class. This knowledge of classification is obtained by the model during its training. Here the recital of the system is assessed by considering accuracy as a metric. In the following table no-4 the accuracy value of different literatures are compared with the accuracy of the proposed model. It is witnessed that the proposed model is having the best accuracy as compared to all the discussed literature by taking the same Cleveland data set which is shown in Fig-4. In addition to accuracy other performance metric like precision and recall percentage of the proposed method is also calculated for finding efficiency. The good percentage of recall indicates how the proposed model has identified the positive cases with less error. In the proposed model 13 features used. In comparison to this in some literatures less number of features are used. In fuzzy logic method 10 features are used, in artificial bee colony method 7 features are used, in principal component analysis method 7 features are used and in SVM Based integer-Coded Genetic Algorithm method 6 features are used. The main aim of using feature reduction is to remove the redundant and dependent feature in order to increase the accuracy. But in the proposed model no features are reduced. The proposed model gives the better accuracy without reducing any feature which is 94.91%. Hence it shows the efficiency of the model as compared to other literatures using feature reduction. In the literature, for diagnosing the coronary artery disease the researchers have used various traditional methods. These methods basically rely on Hypothesis to increase the accuracy level with huge data training to identify the pattern. The high accuracy is achieved using machine learning classification algorithms. But it has the biggest drawback like overfitting. The biggest advantage of the proposed system is that it is not only enhancing the classification accuracy of the system but also properly deal with overfitting. This makes the system more useful for the real-time environment.

Table 4

Model Name	**Accuracy obtained**
Ensemble based KNN(EKNN)	84.8
Random Forest(RF)	85.81
Heart Disease Diagnosis System using Fuzzy Logic(FL)	93
Efficient Heart Disease prediction system(DPS)	86.7
Artificial Bee Colony(ABC)	86.76

Ensemble-based Decision Support Framework(EDSF)	82
Principal Component Analysis and Adaptive Neuro-Fuzzy Inference System(ANFIS)	93.2
Effective diagnosis of heart disease through neural networks ensembles(NNS)	89.01
SVM Based Decision Support System with Integer-Coded Genetic Algorithm(GA)	72.55
Proposed System	**94.91**

VIII. CONCLUSION AND FUTURE WORK

IN previous sections, several literatures are analysed on the basis of their accuracy for predicting coronary artery disease. In the discussed approaches to solve the above stated problem, different solutions were proposed by the researchers to find better accuracy values using the UCI machine learning repository datasets. In the proposed methodology, an efficient solution is proposed using a dense neural network with hyper parameter tuning for the diagnosis of Coronary artery disease. Investigations were carried out on coronary artery disease data set in a self-regulated manner. The result obtained by the proposed system is 94.91% accuracy, 93.75% recall and 96.77% precision during the timing of prediction of coronary artery disease. The proposed methodology provides an efficient and useful solution for the identified problem. It gives a prospect to optimize the error and boost the recital of the system through the implementation of hyper parameter tuning. The actual enhancement achieved in accuracy due to the auto-generation of optimal values for the hyper parameters and then outperforming the selected hyper parameters. In addition to this use of regularization term offers the real power to the system by not allowing the system to over fit and under fit in any circumstance. In addition to this use of the dominant network structure helpful in improving the flow of information and gradient through the network which is really helpful in easily training the network.

REFERENCES

1. Alberto Palacios Pawlovsky ‘An Ensemble Based on Distances for a kNN Method for Heart Disease Diagnosis’ 2018 International Conference on Electronics, Information, and Communication (ICEIC)
2. Yeshvendra K. Singh, Nikhil Sinha., and Sanjay K. Singh., ‘Heart Disease Prediction System Using Random Forest’; First International Conference, ICACDS 2016 , November 11–12, 2016, pp 613-623
3. TanmayKasbe , Ravi Singh Pippal ‘Design of Heart Disease Diagnosis System using Fuzzy Logic’ 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017)
4. https://www.medicinenet.com/heart_disease_coronary_artery_disease/article.htm
5. M. Nikhil Kumar, K. V. S. Koushik, K. Deepak, ‘Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools’ International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307
6. Purushottam,Prof. (Dr.) Kanak Saxena,Richa Sharma, ‘Efficient Heart Disease Prediction System’., Procedia, Computer Science 85 (2016) 962 – 969
7. B.Subanya, Dr.R.R.Rajalaxmi ‘Feature Selection using Artificial Bee Colony for Cardiovascular Disease Classification’ 2014 International Conference on Electronics and Communication System (lCECS -2014)
8. Saba Bashir, Usman Qamar, M.YounusJaved., ‘An Ensemble based Decision Support Framework for Intelligent Heart Disease Diagnosis’ International Conference on Information Society (i-Society 2014), pp 259-264
9. BhuvaneswariAmma N G, ‘An Intelligent Approach Based on Principal Component Analysis and Adaptive Neuro Fuzzy Inference System for Predicting the Risk of Cardiovascular Diseases’., 2013Fifth International Conference on Advanced Computing (ICoAC), pp 241-245
10. Resul Das, Ibrahim Turkoglu. and AbdulkadirSengur. (2009), ‘Effective diagnosis of heart disease through neural networks ensembles’, Expert Systems with Applications 36 (2009) 7675–7680
11. Sumit Bhatia, Praveen Prakash, and G.N. Pillai., ‘SVM Based Decision Support System for Heart Disease Classification with Integer-Coded Genetic Algorithm to Select Critical Features’, Proceedings of the World Congress on Engineering and Computer Science 2008 WCECS 2008, October 22 - 24, 2008, San Francisco,USA
12. Meherwar Fatima, Maruf Pasha ‘Survey of Machine Learning Algorithms for Disease Diagnostic’ Journal of Intelligent Learning Systems and Applications, 2017, 9, 1-16 ISSN Online: 2150-8410 ISSN Print: 2150-8402
13. Hakizimana Leopord, Dr. Wilson Kipruto Cheruiyot, Dr. Stephen Kimani ‘A Survey and Analysis on Classification and Regression Data Mining Techniques for Diseases Outbreak Prediction in Datasets’ The International Journal Of Engineering And Science (IJES)|| Volume || 5 || Issue || 9 || Pages || PP -01-11 || 2016 || ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805
14. M. Chandralekha and N. Shenbagavadivu, ‘Performance Analysis Of Various Machine Learning Techniques To Predict Cardiovascular Disease: An Empirical Study’(2018) Appl. Math. Inf. Sci. 12, No. 1, 217-226 (2018)
15. http://archive.ics.uci.edu/ml/datasets/heart+Disease