

Diagnosis of Coronary Artery Disease using 1-D Convolutional Neural Network



Debabrata Swain, Santosh Kumar Pani, Debabala Swain

Abstract: Heart disease is treated as one of the noxious diseases at the present time. Coronary artery disease is a kind of heart syndrome which is statistically growing day by day in the society. It is very tough for medical practitioners to predict Coronary artery disease as it is a complicated task that needs experience and acquaintance. For the detection of the disease doctors normally prescribe various invasive and non-invasive methods like angiography, ECG and echocardiogram. These methods are very expensive and sometimes not able to discover a number of undiagnosed symptoms. Due to these, it is not possible to detect the disease accurately at an early stage. The medical sector today contains a number of useful data that is helpful to detect a disease accurately. Using this data many researchers have proposed a number of intelligent systems for the detection of the disease. In this work, a competent system is implemented using deep learning for the better detection of the disease. The system is constructed using Convolutional Neural Network. The system has three phases. In the first phase data cleaning, data imputation and important feature selection are performed. In the second phase model training and hyperparameter tuning is performed. Finally, in the last phase, the model prediction is performed using the test data. The data set used for experimentation is Cleveland, Hungary, Switzerland and Long beach heart disease data present in the UCI repository. The proposed system gives a classification accuracy of 96.49% during testing, which is highest among all the discussed methods.

Index Terms: 1-D Convolution Neural Network (1-D CNN), Artificial neural network (ANN), Coronary artery disease (CAD), Decision tree (DT), K Nearest Neighbor, Naïve Bayes (NB), Principal Component Analysis (PCA), SSVM algorithm, Support vector machine (SVM),

I. INTRODUCTION

The mortality rate of heart disease is approximately 10% of the total death rate in the world. According to global death rate statistics, heart disease is identified as the major source of demise. In Japan heart disease recognized as the principal basis of mortality rate. [1] The availability of modern gadgets, automation of tools, lack of exercise and consumption of junk food upturns slothfulness with a person. This laziness creates a big belly in the human body which further becomes the major cause of heart disease. [2]

Revised Manuscript Received on 30 July 2019.

* Correspondence Author

Mr. Debabrata, Research Scholar, Department of Computer Engineering, K.I.I.T. University, Bhubaneswar, Odisha India.

Dr Santosh Kumar, Associate Professor Department of Computer Engineering, K.I.I.T. University, Bhubaneswar, Odisha, India.

Dr Debabala, Associate Professor Department of Computer Science, Ramadevi Women's University, Bhubaneswar, Odisha, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

According to the survey of WHO every year millions of populations of the world are dying and suffering from heart disease. Therefore, prediction of the disease became an ultimate need for society. [3] Generally, heart disease indicates any exceptional conditions which alter the smooth and healthy functioning of the heart. Heart disease is having different types like coronary artery disease (CAD), valvular heart disease (VHD) and cardiomyopathy. In CAD the arteries through which blood supply happens mainly gets affected. In this disease, cholesterol gets deposited in the arteries which create an obstacle for the free flow of blood. In VHD the valves which control the in and out process of the blood of the heart gets affected. In cardiomyopathy, the heart muscle compression and expansion process get affected. [4] The coronary artery disease is again further divided into several types like unstable angina, complete heart block, and myocardial infarction. In the case of unstable angina, the heart does not get suitable quantity of oxygen and less amount of blood flow happens through it. As a result of this heart muscle becomes feeble which is also known as Myocardium. The primary causes of unstable angina are more blood pressure, heavy LDL cholesterol, old age, smoking addiction. In the case of complete heart blockage, the rhythms generated by the heart are incapable to reach from the heart's upper chamber to the lower chamber. Few symptoms of this disease are unsteadiness, exhaustion, chest pain and breathing problem. In the case of myocardial infarctions, some exceptional behavior observed in the proper functioning of the heart. As a result of this disease discomfort, ness in chest and tiredness observed with the patient. [5] Normally a person dies because of heart disease without any prior sign. Around 50% of heart stroke happens without any prior indication. When some person gets severe chest pain then only the physicians normally suggest for various invasive or non-invasive tests like ECG and ECHO. Through these tests, coronary disease can be identified. But sometimes these test reports contain a number of ambiguities. [6] As coronary artery disease is associated with a number of complex factors and vagueness hence its detection becomes a complicated task for the physicians. The early detection of the syndrome can save the life of the patient from death. [7] The ladies of greater age have more possibility of getting heart disease.



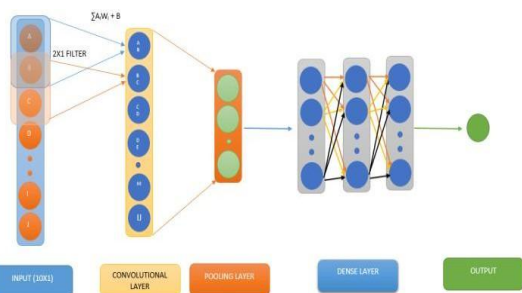
Hence a lot of lives can be saved if the disease detection time can be reduced and detection accuracy can be enhanced. [8] During the analysis of coronary artery disease physician generally observe the various symptoms, expressions and bodily examination report of any patient. After that, the physician uses his experience for the diagnosis of the disease which is not always 100% correct. [9] Normally the physicians refer to the previous diagnosis decisions that they took in the past for any patient with the same physical parameters. These parameters evaluation for the identification of the disease generally makes the diagnosis process very much complicated. Hence for the accurate identification of Coronary artery disease, a physician must need high experience and a lot of skills with them. [10] The medical data of the patient plays a vital role in doing the accurate identification of the syndrome. Hence it is required to identify the useful data or bioindicators (attribute) from the remedial data of any patient. [11] Once the useful data are identified it can be used for training a classification model. Once the model training is done it can be used for taking an efficient decision which is otherwise known as machine learning. When a model gets trained it acquires the ability to recognize the pattern which is really helpful in classification-based problems. [12] Some generic algorithms are used in identifying the critical features of the disease as well as model parameter optimization to develop an efficient solution for the disease prediction. [13] Currently, a number of intelligent healthcare applications are developed due to the advancement in the field of computational intelligence. Since the last few years, various computational systems have been developed which advanced the skills and capabilities of the doctors for taking the accurate diagnosis of patients. Inspired by the requirement of an intelligent system here a competent prediction system is proposed using the 1-D convolutional neural network (CNN). 1-D CNN is a type of CNN. In 1-D CNN so many 1-D convolutional filters are used to mine local and global features from data. Deep CNN is used subject to the nature of the data. A 1-D CNN is mainly used for stemming features from a stable-length division of the overall dataset where feature presence or absence in the segment is not so significant. Here the goal of the proposed system can be set as:

1. Removing outliers, anomalies and use of proper imputation to handle the missing values to clean and standardize the data set
2. Selection of proper CNN model among the existing CNN models according to the experimentation data set.
3. Determining the right dimension of convolutional neural network filters using different accuracy measures.
4. Selecting the full architecture of the convolutional neural network for better prediction.

II. LITERATURE REVIEW

Alberto Palacios Pawlovsky et al [1] proposed a cardiovascular disease diagnosis method using an ensemble-based K-nearest neighbor algorithm. In the beginning, the value of K was selected which was the number of neighbors. Then 5 distance finding algorithms like Euclid, Manhattan, Chebyshev, Canberra, and Mahalanobis were used to find K nearest neighbors and their classes. Finally, the voting algorithm was implemented to select the class having the highest majority for the test instance. Hamidreza Ashrafi Esfahani et al [2] proposed a cardiovascular disease prediction system using an ensemble-based classifier method. Here the system was divided into 5 phases. In the first phase, data were collected from the UCI repository. In the second phase data pre-processing was carried out to handle the null values in the data. In the third phase, feature selection was performed to identify the important features using the Karl Pearson coefficient of correlation. In the fourth phase model prediction was carried out using the ensemble-based classifier. The model consisted of decision tree, support vector machine, neural network, logistic regression, naïve bayes, and rough set classifiers. Finally, the voting algorithm was used to find the class having the highest majority. Tulay Karayilan et al [3] implemented a neural network-based system for the identification of CAD. The neural network contained 3 layers. The Input layer, hidden layer, and the output layer were containing 13, 12 and 2 neurons respectively. The system designed for binary prediction which is presence or absence of disease. During model training, the backpropagation algorithm was used for weight updation. Here sigmoid activation function was used. Out of the total data, 70% used for model training, 15% used for validation and rest 15% used for model testing. Yeshvendra K. Singh et al [6] implemented a random forest based cardiac disease detection system. The algorithm primarily tried to find the non-linear relation between the features in the data set. Hence the non-correlated data were simply separated out of the data set. Boosting of accuracy was achieved by doing proper adjustment of the random forest arguments like randomness, number of trees and optimum number of splits. With 20 splits and 75 trees, the highest accuracy was obtained. Tanmay Kasbe et al [7] implemented a fuzzy logic based cardiac disease prediction system. Here the system was formed by 2 components namely – fuzzy expert system (FES) and fuzzy rule base (FRB). In FES various membership functions like triangular and trapezoid were used to find the membership value for each feature. Then according to the values of the features, the equivalent classes were assigned.

In FRB various rules were formed using AND, OR operators with different combinations. Finally, a total of 86 different fuzzy rules were formed and their equivalent output classes were assigned. Radhimeenakshi et al [8] developed a CAD detection system by taking two classifiers such as SVM and ANN. SVM did prediction by forming a hyperplane which served as a separator between the healthy and sick test instances. In the second method, neural network was used for prediction. During neural network training, cross-validation was used for evaluating the model during its training. Purushottam et al [9] have formed a decision tree for the classification of heart disease. In the beginning data pre-processing was performed to get the clean data. In the next stage, a decision tree was constructed by considering various factors like confidence, MinItemset, and threshold. The different rules used for classification were original rules, pruned rules, rules without duplicates, classified rules and polish. Resul Das et al [10] have implemented an ensemble based neural network method for the better identification of heart disease. The system was divided into five different components like heart database, Data partitioning, Variable selection, neural network and ensemble component. Heart database consisted of input features. Partitioning component did the partition of data for training and testing purposes. The variable selection component found the unrelated features and removed them successively from the data set. The neural network component was used for predicting the correct class of each test record after their training was over. Ensemble component composed of three different types of feedforward neural networks. After each neural network did their prediction for any test instance this component collected the prediction class of all the neural networks and used voting algorithm to select the class having majority. Saba Bashir et al [11] implemented an ensemble-based method for the detection of the cardiac syndrome. The ensemble model consisted of DT, NB and SVM classifiers. All three classifiers predicted each unseen test instances. Finally, the voting algorithm is used for selecting a class having majority



voting. T. Santhanam et al [12] developed an efficient cardiovascular sickness prediction system using PCA and feedforward neural network (FFNN). PCA discovered the most relevant feature out of all the features present in the data set. PCA used to invent the co-relation between the features by creating the correlation matrix. After the reduced number

Fig-1 Convolutional Neural Network Architecture
of features was identified they were used for the model training and testing. The FFNN was used to detect the class of any test record. S.Bhatia et al [13] implemented a cardiovascular detection system using SVM and integer coded genetic algorithm. For quick detection of support vectors, the SSVM algorithm was used. For the identification of crucial parameters integer coded genetic algorithm was used. Finally, the classification of the records is done using the SVM algorithm.

III. DATASET DESCRIPTION

The data is gathered from the Cleveland, Hungarian, Switzerland and Long beach medical data present in the UCI repository of datasets. The dataset is consisting of a number of health related biomarkers used for the detection of Coronary artery disease. Cleveland contains 303 records, Hungarian contains 294 records, Switzerland contains 123 records and Long beach contains 200 records. All the datasets are having a total of 76 attributes. Out of 76 attributes 14 are recognized as crucial features using feature selection method. Among the 14 attributes, 13 are independent features and 1 is the target feature. The target feature is the dependent one which works as an indicator of the presence or absence of the disease. All the features are having 2 types of values such as discrete and continuous values. The discrete values are divided into many classes. Continuous values are real or integer values having no limit. Blood sugar is a discrete value having 2 classes like true or false. All the features are shown in Table-1.

Sr No	Feature	Value type
1	Age	Numeric
2	Sex	Discrete
3	Chest pain	Discrete
4	Resting BP	Numeric
5	Cholesterol	Numeric
6	Fasting sugar	Discrete
7	Resting ECG	Discrete
8	Maximum rate	Numeric
9	Induced angina	Discrete
10	ST Depression	Real
11	Slope	Discrete
12	Vessels colored	Discrete
13	Thalassemia	Discrete
14	Heart disease	Discrete

I Data set details

1-D CONVOLUTIONAL NEURAL NETWORK

A 1-D CNN is a kind of dense neural network which was designed from the biological driven models. It is considered as a better model for pattern recognition type problems. There are some input data given which is nothing but the features. These features are present in the form of a one-dimensional array of the order (n×1). Where n is the number of features.

These input data are processed through a number of layers as shown in figure-1.



Diagnosis of Coronary Artery Disease using 1-D Convolutional Neural Network

1. **Input Layer-** This layer contains the set of feature value present in the data set. Generally, Input Layer is the layer from where the processing starts.
2. **Convolution Layer-** It
3. will take a patch of data from the input layer to the convolution layer and apply a set of 1-D filters. This layer consists of a number of filters which are nothing but an array of weights. These filters are applied to a given patch of input data features. Dot product performed upon the input feature patch and filter. The same input data patch is getting processed by a number of different filters. Here a RELU activation function is used. RELU function is applied upon the result which is generated after filter applied on input. The main job of RELU function is to add non-linearity in the decision boundary.
4. **Pulling Layer-** After the data is processed by the Relu function it is given to this layer. It is a kind of non-linear down sampling layer which reduces the input data size. This layer focuses on capturing important features. Again data will go through a bunch of activation filters in the further convolution layers and get some more information. Again downsampling is performed in the further pulling layers.
5. **Dense Layer-** This layer is also referred as the fully-connected layer. Here each and every node in one layer is connected with the nodes in the next layer. At the end of the fully connected layer, it will follow a layer of softmax or logistic regression for probability distribution and come up with the top 3 probabilities for prediction.

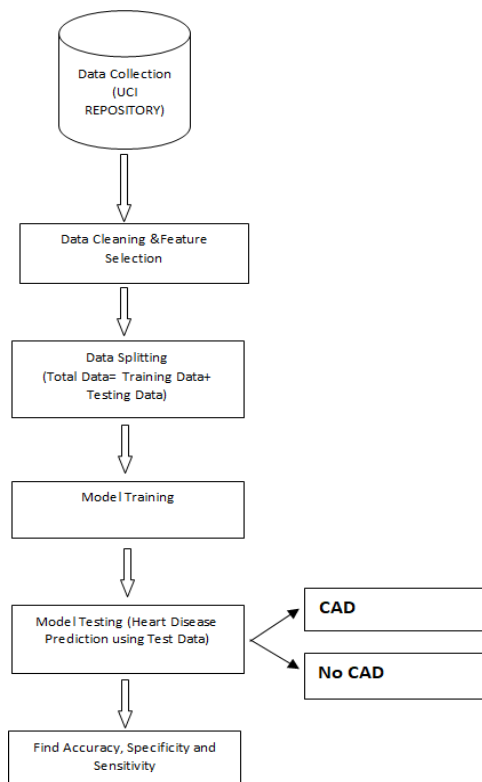


Fig-2 Proposed System Architecture

IV. PROPOSED SYSTEM ALGORITHM

The proposed system architecture is shown in figure -2.

1. **Data Collection-** Data is collected from Cleveland, Hungarian, Switzerland and Long beach data set.
2. **Data Cleaning-** The total number of records read from all the data sources was 920. After reading the data from the data sets a number of missing values were found. Especially the slope of the peak exercise, major vessels colored by fluoroscopy and thalassemia are containing a number of missing values. The slope of the peak exercise contained 308, major vessels colored by fluoroscopy contained 606 and thalassemia 477 number of missing values. The missing values in every feature are found using the visualization tool like heat map. The total count of missing values is found using isnull().sum() function. The missing values in Cholesterol are replaced by the mean value. Whereas missing values in fasting sugar, vessels colored and thalassemia are replaced by their mode values respectively. The remaining missing values are found in Resting BP, maximum rate, induced angina, ST Depression and Slope. The count of the missing values for these features is very less. These records are deleted because their deletions are not having any major effect on the improvement of the accuracy. The total number of records after data cleaning is 854 as shown in figure 6. The target feature is containing values like 1, 2, 3, 4 representing different stages of the diseases. As in the proposed system binary prediction is targeted hence all these values 1, 2, 3, 4 are replaced by 1 indicating the presence of the disease. Out of all 76 features, 14 important features are identified using the backward feature selection method. For the addition or removal of a feature logistic regression algorithm is used. The null values in the features before, after imputation through heat map are shown in figure 3, 4 & 5.

```
In [35]: #df.drop(['ca', 'thal'],axis=1,inplace=True)
sns.heatmap(df.isnull(),yticklabels=False,cbar=False,cmap="viridis")
Out[35]: <matplotlib.axes._subplots.AxesSubplot at 0x103b8240>
```

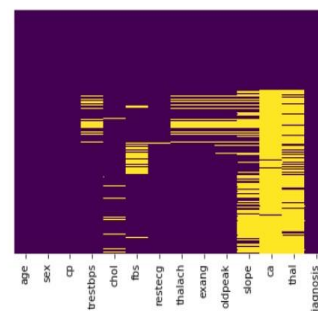


Fig-3 (Data set before Imputation)

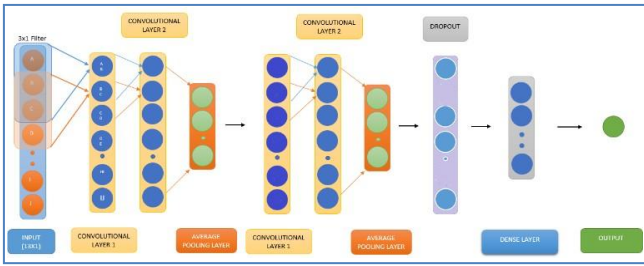


Fig-4 (Data set after Imputation)

```
In [37]: df.isnull().sum()
Out[37]: age      0
sex        0
cp         0
trestbps   0
chol       0
fbs        0
restecg    0
thalach    0
exang      0
oldpeak    0
slope      0
ca         0
thal       0
diagnosis  0
dtype: int64
```

Fig-5 (Features after imputation)

```
In [39]: dftemp=df.abs()
temp=dftemp['oldpeak']
dftemp=dftemp[['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs',
'restecg', 'thalach', 'exang', 'slope', 'ca', 'thal',
'diagnosis']].astype('int')
dftemp['oldpeak']=temp
len(dftemp)
Out[39]: 854
```

Fig-6 (Total Records after cleaning)

- Data Splitting-** In this step the total data set are divided into training and testing set. For training 80% (683) data records are used whereas for testing 20% (171) data records are used.
- Training-** In this step, the 1-D Convolutional neural network is trained using the training records. During training, the weights of the filters are updated for better prediction during testing.

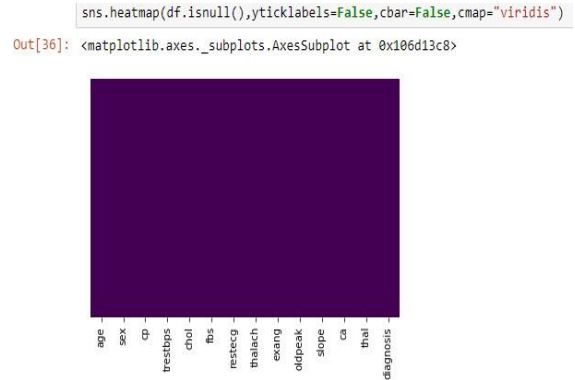


Fig-7 (Modified CNN Structure)

Hyperparameters are tuned in this step to find the best set hyperparameters which will result in better prediction during testing.

- Testing-** In this step, the trained model is evaluated using test records. Here the 1-D CNN is classifying the records into 2 categories of classes where 1 indicates the presence of disease and 0 indicates the absence of the disease.
- In this step, the accuracy of the model is calculated.

Model Details

Here the proposed model contains four convolution layers, two pooling layer, and one dense layer. Dropout is used to avoid overfitting. The modified CNN is shown in figure-7. In all the convolution layers 32 filters are used. Each filter is having size 3. The activation function which is used in all the convolution layers is Relu. Relu is otherwise known as Rectified linear unit. The main job of the Relu activation function is to provide non-linearity to the output generated by the filters in the convolution layer. The 13 features are given as input to this layer. The size of this layer is 13x32. Mathematically it can be expressed as

$$A = \text{Max}(0, B) \tag{1}$$

(Relu function)

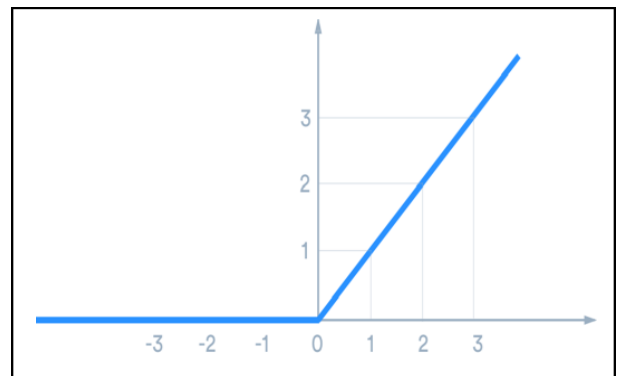


Fig-8 (Graph for Relu Function)

Diagnosis of Coronary Artery Disease using 1-D Convolutional Neural Network

Relu function gives the linear output B if the value of B>0 otherwise for all values of B≤0 it gives the output 0 as shown in figure 4 and equation 1. As different 32 filters are applied hence the convolution layer generates 32 different output patches of size 3.

Max Pooling layer- This layer is responsible for doing down-sampling by applying maximum criteria upon each patch of size 3 generated from the previous layer. So out of all the 3 values present in the patch whichever is maximum that will be selected by the max poll layer. Hence this layer reduces the data size which comes from its previous layers as shown in figure 9 and equation 2.

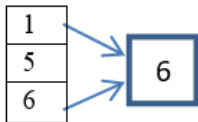


Fig- 9 (Max Pooling)

$$\text{MaxPooling} = \text{MAX} (x_1, x_2, x_3, \dots, x_n) \quad (2)$$

(Max Pooling)

Average Pooling- This layer gets the input from the previous convolution layer. In the average polling layer again downsampling is performed by finding the average of the data present in the patch of size 3. The 3 data values present in every patch is replaced by their average as shown in figure 10 and equation 3.

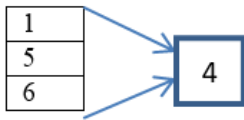


Fig- 10 (Average Pooling)

$$\text{Average Pooling} = \frac{1}{n} (\sum_{i=1}^n x_i) \quad (3)$$

(Average Pooling)

Dropout- It is a method of regularization that prevents the model to get overfit. Dropout is the process of removing the connections between an output neuron of a layer and its connected input neurons of the previous layer. For a particular class of an output neuron, it finds the probability of every input neuron. Here the probability value specified for drop out is 0.2. It means the input neurons whose

probability value is less than 0.2 are disconnected from the output neuron. Dropout rule is given in following equation 4.

$$P(x_i | y_i=1) < 0.2 \quad (4)$$

(Drop out Rule)

Dense Layer- There is only one dense layer used in the model. This is the last layer of the model. It is generating a single output which is the status of the disease. For output

activation sigmoid function is used. The sigmoid function always generates output between 0 and 1 as shown in figure 11 and equation 5.

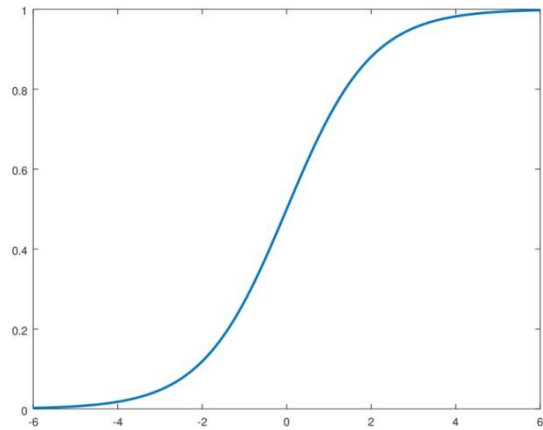


Fig-11 (Sigmoid Function)

$$y(x) = \frac{1}{1+e^{-x}} \quad (5)$$

(Sigmoid Function)

Loss Function- It calculates the average error of any classifier. Error measures the deviation of predicted and actual outputs. Loss function mainly finds the average error for every training point. An optimizer is used for minimizing the error by finding the optimum values for the weights of the filters. Here Adam optimizer is used for finding the best value of the parameters. As in the proposed system binary prediction is targeted hence binary cross entropy loss function is used. So Loss is calculated by taking the logarithmic probability of both the positive and negative classes of the classification problem which is demonstrated in equation 6.

$$L(q) = \frac{1}{M} r_j \log(P(r_j)) + (1 - r_j) \log(1 - P(r_j)) \quad (6)$$

(Binary cross entropy)

Where,

L(q) = Loss function

r_j = output label of positive class

P(r_j) = Probability of positive class

(1- r_j) = output label of negative class

1-P(r_j)= Probability of negative class

Layer	Input Size	Number of filters	Output Size
Convolution Layer1	13	32	13×32
Convolution Layer	13×32	32	13×32

Max Pooling	13×32	-	4×32
Convolution Layer	4×32	32	4×32
Convolution Layer	4×32	32	13×32
Average Pooling	4×32	-	32
Dense Layer	32	-	1

II (Model layer details)

VI. RESULT

To evaluate the trained model the test data is used. The efficiency of the model is found using the confusion matrix as shown in table-3. The accuracy of any classification model is the ratio between correctly predicted cases and total test cases. The accuracy of the model is found as 96.49%. The precision of any model is the ratio between true positive and the addition of true positive and false positive. Precision is found as 93%. The recall is the ratio between true positive and the addition of true positive and false negative. The recall is found as 100%.

TP	82
FP	6
TN	83
FN	0

III (Confusion Matrix)

VII. PERFORMANCE ANALYSIS

The performance of the system is computed using classification accuracy as a metric. The classification accuracy of any model shows how it predicts against any unseen data. The proposed system shows a classification accuracy of 96.49% which is the highest among all the discussed literature. In the proposed system backward feature selection is conducted which identified 13 important features. The actual aim of feature reduction is to remove the irrelevant and redundant features from the dataset for improving the accuracy of the classifier. The comparison between the proposed system and other discussed literature are shown in the table-4 and fig-7. In fuzzy logic based prediction 10 features are used

and prediction using PCA and FFNN 8 features are used. But both the case accuracy found as 93.33% and 95.2%. In some classification systems while improving the accuracy sometimes the model leads towards overfitting. But the use of the Relu activation function prevents the proposed system from overfitting. The use of Adam optimizer also boosts the result of the system by finding the optimum value of the weights in the filters. Finally, the use of a robust classification model like 1-D convolutional neural network facilitates learning all the features properly and resulted in improved accuracy during testing. The best part of the model is the value of false negative which is found as 0. It indicates that the wrongly negative prediction case is 0. If a person is wrongly identified as not having heart disease then he may not go for medication. Hence it may create a problem for that person. It shows the proposed model is very efficient.

Model Name	Accuracy
Ensemble based kNN method (EKNN)	85
Cardio Vascular Disease detection using a new ensemble classifier (NEC)	89
Prediction of Heart Disease Using Neural Network(NN)	95
Random Forest (RF)	85.81
Design of Heart Disease Diagnosis System using fuzzy Logic (FL)	93.33
Classification and prediction of heart disease risk using data mining techniques of SVM and ANN (ANN)	85.6
Efficient Heart Disease Prediction System (EHDPS)	86.3
Neural networks ensembles (NNE)	89.01
An Ensemble-based decision Support Framework (EBDSS)	82
Classification using PCA and Feed Forward Neural Networks (FFNN)	95.2
SVM Based Decision Support System (SVM)	90.57
Proposed System	96.49

IV (Accuracy Comparison Table)

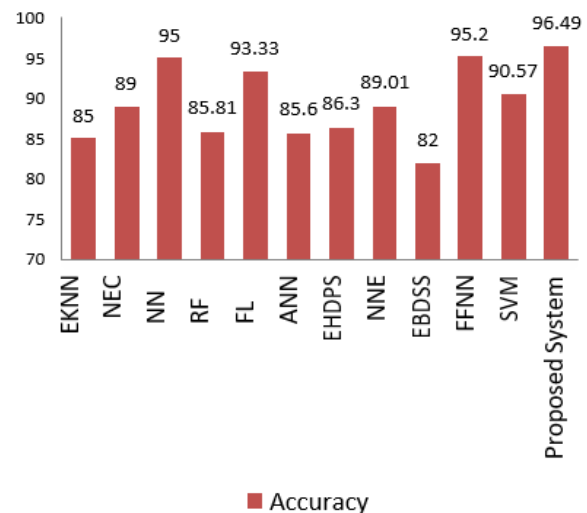


Fig-12 (Prediction Accuracy Comparison Of Different Methods With Proposed System)

ROC Curve- It gives a graphical representation of the predictability of any binary classifier. It can be formed by drawing a graph between true positive rate and false positive rate at various threshold values. The area under the curve (AUC) shows the performance measure of the classifier at different threshold values. Its value varies from 0.5 to 1.0. If the value is 1.0 or closer to it then it indicates a good classifier. In the proposed model the AUC value is 0.98 which indicates the efficiency of the system in correctly predicting the test record. The ROC graph is shown in figure-13.



Diagnosis of Coronary Artery Disease using 1-D Convolutional Neural Network

```
fpr, tpr, thresholds = roc_curve(y_testID, probs)
plot_roc_curve(fpr, tpr)
```

AUC: 0.98

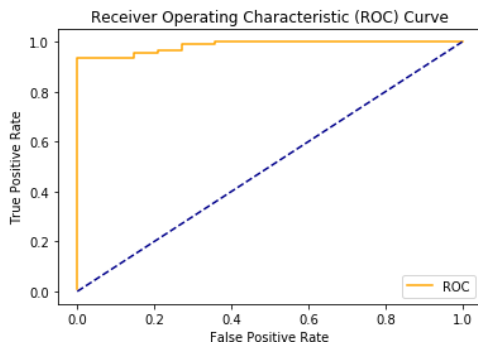


Fig-13 (ROC Curve)

VIII. CONCLUSION

In the previous section, the performance of different systems is compared with the proposed system for the detection of coronary artery disease. In all the literature different researchers have proposed their solution for the better identification of the disease using the UCI heart disease data set. In the proposed methodology a competent detection system like 1-D CNN is used. The analysis was carried out upon the CAD dataset in a self-regulated manner. The accuracy, precision and recall of the proposed system are found as 96.49%, 93% and 100% respectively. The proposed system provided a useful and robust solution for the detection of CAD. Error optimization and performance boosting has been done by the use of Adam optimizer. The optimizer enhanced accuracy by finding the best values of the hyper parameters. The actual improvement in the system is achieved by the use of the Relu function. Relu function handled the overfitting and underfitting problem successfully. The use of drop out in the proposed system helps in making the model simple by removing weak connections. Also the overfitting issue of the model is properly handled by the use of drop out. The use of large and properly pre-processed dataset boost the efficiency of the system by covering a number of different case studies during the training of the system. This helps the system in providing a generalized solution for the prediction of coronary artery disease. The good percentage of AUC shows the efficiency of the classifier in doing correct prediction against unseen test data. Finally, the powerful and adaptive structure of the 1-D CNN boosted the model learning of different feature patterns of the disease which can be highly applicable to the real-time environment.

REFERENCES

1. Alberto 'An Ensemble Based on Distances for a kNN Method for Heart Disease Diagnosis' ICEIC 2018
2. Hamidreza 'Cardiovascular disease detection using a new ensemble classifier' KBEI 2017)
3. Tulay 'rediction of heart disease using neural network' International conference on computer science and engineering 2017)
4. https://www.medicinenet.com/heart_disease_coronary_artery_diseases/article.htm
5. K Raj Mohan 'rediction and diagnosis of cardio vascular disease- A critical survey' (WCCCT 2014)
6. Y. Singh 'Heart Disease rediction System Using Random Forest';

ICACDS 2016.

7. Kasbe 'Design of Heart Disease Diagnosis System using Fuzzy Logic' ICECDS 2017.
8. Radhimeenakshi 'Classification and prediction of heart disease risk using data mining techniques of SVM and artificial neural network' Indiacom 2016
9. Purushottam, 'Efficient Heart Disease Prediction System', Procedia, Computer Science 85 (2016)
10. Resul 'Effective diagnosis of heart disease through neural networks ensembles', Expert Systems with Applications 36 (2009)
11. Bashir, 'An Ensemble based Decision Support Framework for Intelligent Heart Disease Diagnosis' (i-Society 2014)
12. Santhanam 'Heart Disease Classification Using PCA and Feed Forward Neural Networks' MIKE 2013, LNAI 8284, pp. 90–99, 2013.
13. Sumit, 'SVM Based Decision Support System for Heart Disease Classification with IntegerCoded Genetic Algorithm to Select Critical Features', WCECS 2008
14. [http://archive.ics.uci.edu/ml/datasets/heart+Dis ease- UCI Data Set](http://archive.ics.uci.edu/ml/datasets/heart+Dis+ease-+UCI+Data+Set)

AUTHORS PROFILE



Mr. Debabrata Swain is currently working as a PhD research scholar in the School of Computer Engineering, K.I.I.T. University, Bhubaneswar, Odisha. His areas of interest are Machine Learning, Data analytics, Deep Learning.



Dr Santosh Kumar Pani is currently working as an Associate professor at School of Computer Engineering, K.I.I.T. University, Bhubaneswar, Odisha. His area of interest are program analysis, internet of things, cloud computing, block chain technology and machine learning.



Dr Debalala Swain is currently working as an Associate Professor in Computer Science Department, Ramadevi Women's University, Bhubaneswar, Odisha. Her area of interest is Computer Architecture, Computer Security, Artificial Intelligence and Machine Learning