Heart Disease Prediction using Deep Neural Network Models

Sadman Jahin,Md. Moniruzzaman, Miyad Bhuiyan,

CSE Department

Ahsanullah University of Science & Technology
141 & 142, Love Rd, Dhaka 1208

170204115@aust.edu

170204067@aust.edu

170204110@aust.edu

I. Introduction

Cardiac disorders also known narrowly as “heart diseases” are the cause of most deaths worldwide. Heart disease has become a cause of increasing concern for this country with patients enduring several sorts of related illnesses. Death is inevitable if some of the related diseases are diagnosed too late.

In our project, we will try to generate a predictive model of heart diseases which will be used for early detections. Our focus is to find the pre-processing techniques best for specific models, improving the existing models, creating combined predictions from two or more datasets.Our Focus is to implement the model and increase the accuracy of the model done previously.

II.Motivation

A. Why is this project important?

Heart disease is a fatal human disease, rapidly increases globally in both developed and undeveloped countries and consequently, causes death. Normally, in this disease, the heart fails to supply a sufficient amount of blood to other parts of the body in order to accomplish their normal functionalities. Early and on-time diagnosing of this problem is very essential for preventing patients from more damage and saving their lives.

B. Which aspects of our societies does this project solve?

The problem we find out is that major of the patients who need surgery remain outside of the purview of service. If we detect them at early stages, half of the surgeries can be avoided through medication and proper lifestyle. People in accessible areas bear a lot of heart-related diseases due to the distance of adjacent clinics.

III. Methodology

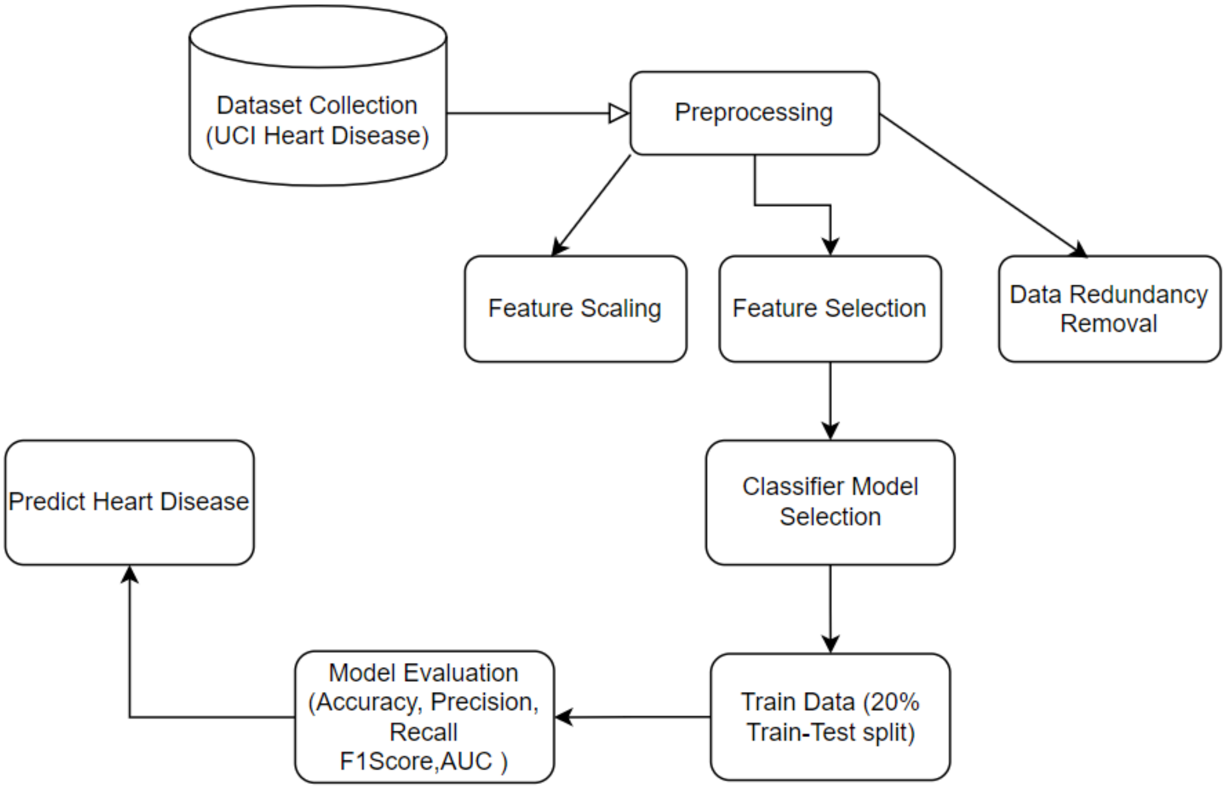


Fig: Proposed Methodology for this work

A. Model Selection

We have selected Artificial Neural Network(ANN), Convolutional Neural Network(CNN) and Long Short Term Memory (LSTM) for training our data based on paper study. Our Focus is to implement the model and increase the accuracy of the model done previously.

Model 4 (ANN)

Input Data Shape : 13. Input Data is scaled. No. hidden layer : 1. Neurons: 300

Activation:

Hidden Layer: Relu Output Layer: Sigmoid

Loss: Binary Cross Entropy with LogitLoss

Model 5 (ANN)

Input Data Shape : 13. Input Data is scaled. No. hidden layer : 1. Neurons: 300

Activation:

Hidden Layer: Relu Output Layer: Sigmoid

Loss: Binary Cross Entropy with LogitLoss

Output Shape: 1

B. Design of the Network Structure

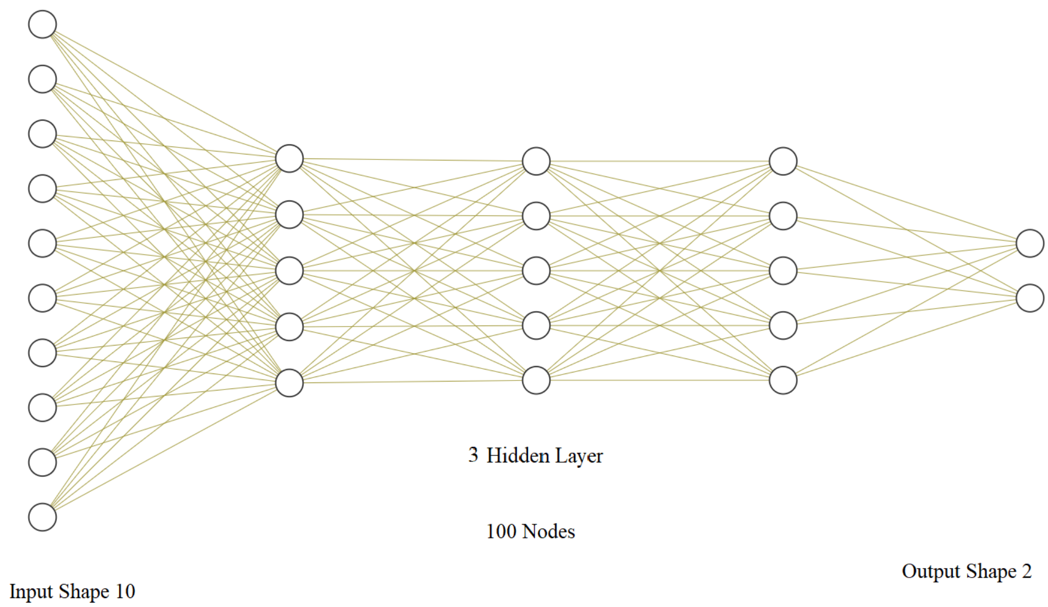


Fig: Proposed Model 4

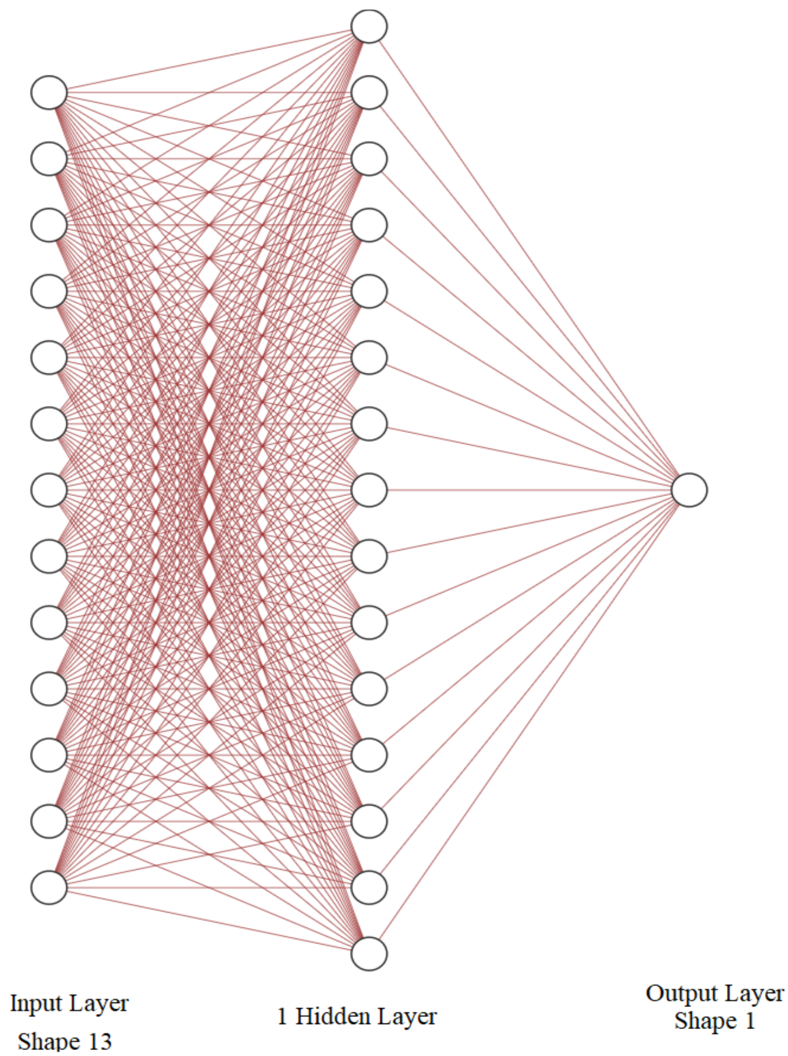


Fig: Proposed Model 5

IV. Experiments

This dataset was taken from the UCI machine learning repository. The heart disease dataset is made up of 75 raw features from which 13 features were published. These features are very vital in the diagnosis of heart diseases. The 13 features considered in this research work are stated below :

C. Dataset Collection

Sl	Attributes	Description
1.	Age	age in years
2.	**Sex**	1 = male; 0 = female
3.	cp	chest **pain** type (4 values)
4.	trestbps	resting blood pressure (in mm Hg on admission to the hospital)
5.	chol	serum cholesterol in mg/dl
6.	fbs	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7.	restecg	resting electrocardiographic results
8.	Thalach	maximum heart rate achieved
9.	Exang	exercise induced angina (1 = yes; 0 = no)
10.	Oldpeak	ST depression induced by exercise relative to rest
11.	Slope	Heart rate slope
12.	Ca	Count of major vessels (value 0-3) coloured by fluoroscopy.
13.	Thal	Thal: 3= normal; 6 = fixed defect; 7 = reversible defect.

The data set had 13 features and 303 rows. No NULL values or duplicate values found in the dataset. Dataset contains 164 (54.3%) heart disease (target = 1) patients and 138 (45.7%) non heart disease (target = 0) patients. Fig.1 represents a balanced dataset. Among these 31.79 % are Female Patients and 68.21 % are Male Patients. The average age of patients is 53. Fig.2 shows the patients affected in cardiac disease at different ranges of age. From Fig.3 visualized that the affected rate of male patients is higher than the rate of female patients.

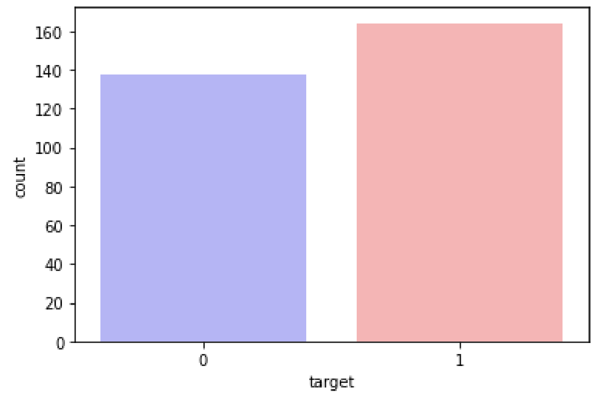


Fig.1: Heart disease(1) and Non heart disease(0)

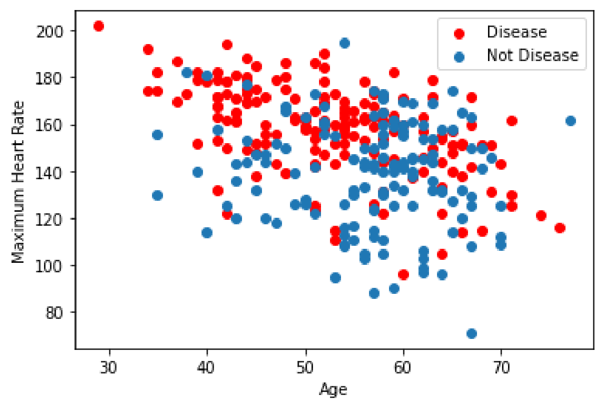


Fig.2: Age vs Cardio Disease.

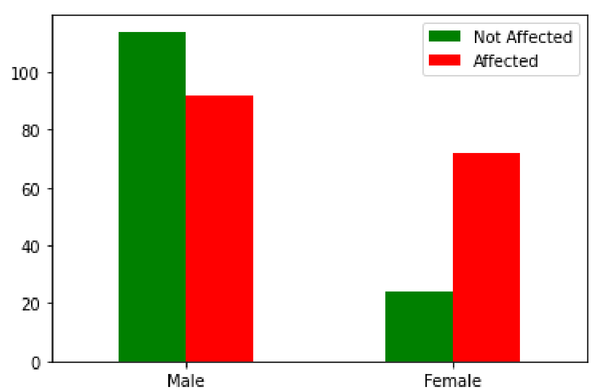


Fig.3: Affected patient based on sex.

D. Data Pre-Processing

To increase the performance and stability need to pre processing the data. The SelectKBest method selects the features according to the k highest score. Applying KBest chol (2.002%),fbs (0.2160%), trestbps (6.55%) have low scores and drop these features. Now it has 10 features and 210 rows.

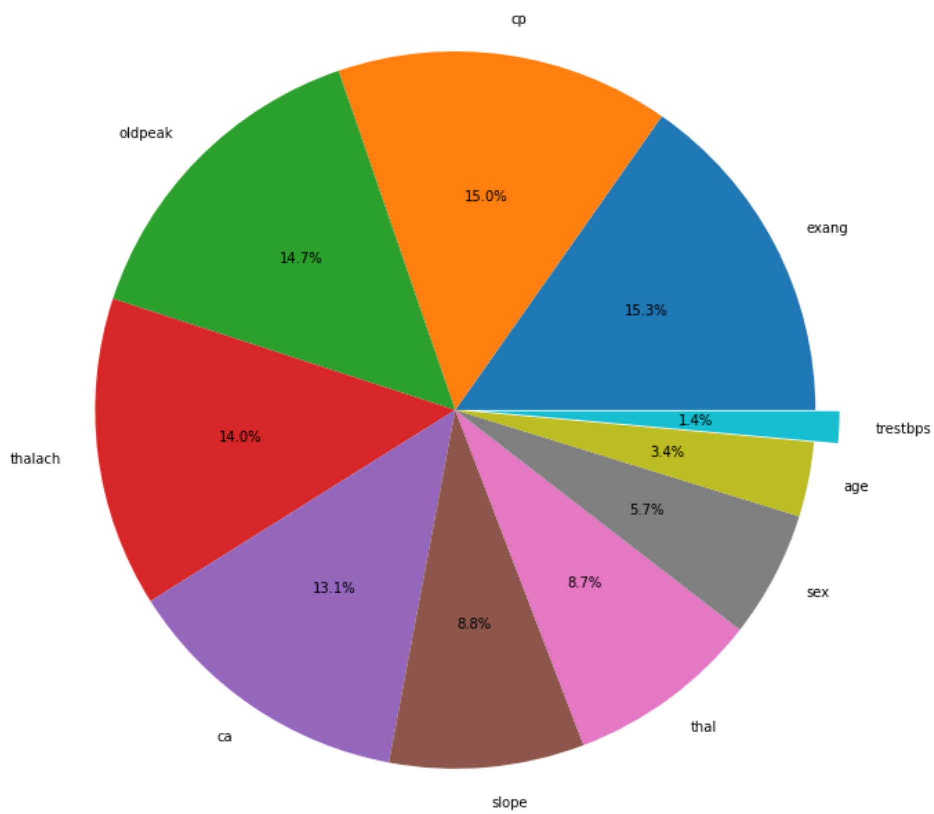


Fig.4: Feature Score (Least Important Selected)

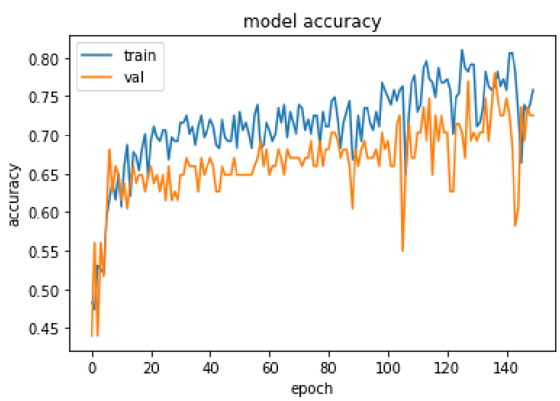
E. Performance Metrics

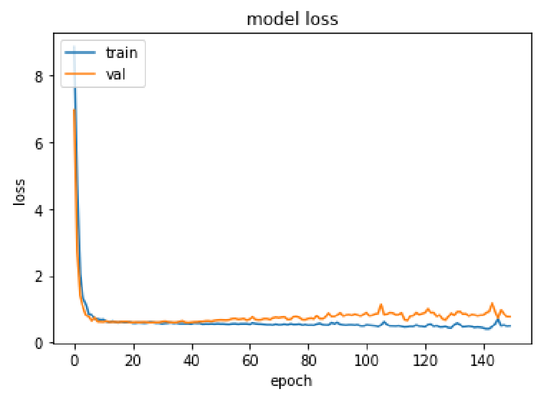
For *Performance Metrics we have taken Accuracy , Precision , Recall, F1 Score for evaluation of models.*

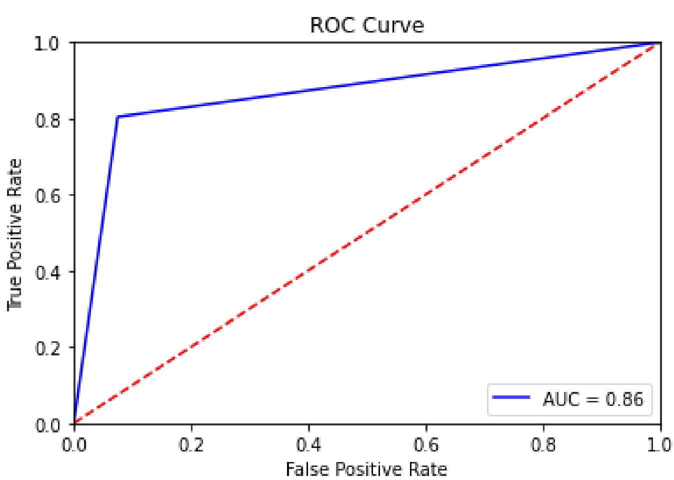
The paper we have chosen to improve firstly was Neural network diagnosis of heart disease (2015). Our expected result is 85% accuracy after implementing the mentioned structure.
As we successfully improved the model given and increased the accuracy of the selected paper we tried to select two more papers with better performance metrics. As we know, only accuracy can not be a good performance metric for heart disease prediction.

Our Result:

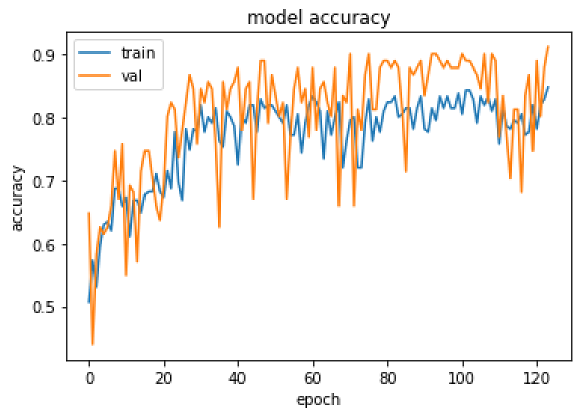
ANN(Paper Structure) : Accuracy 85.71%

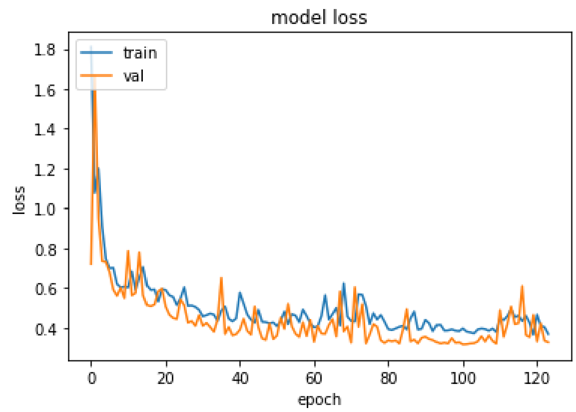


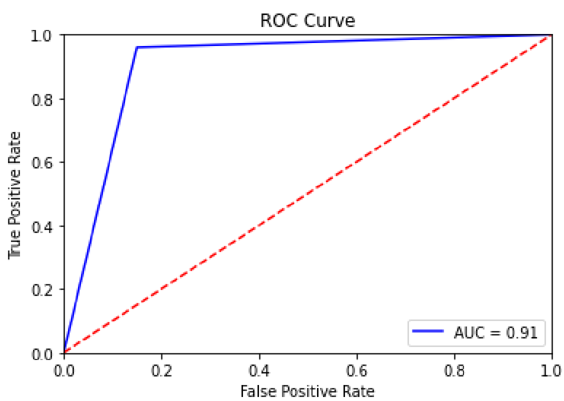




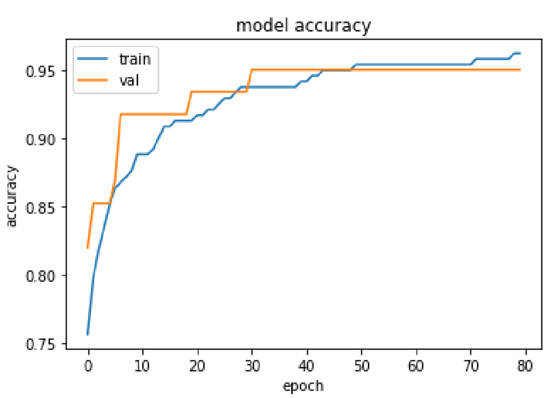
ANN(Model 4) : Accuracy 91.08%

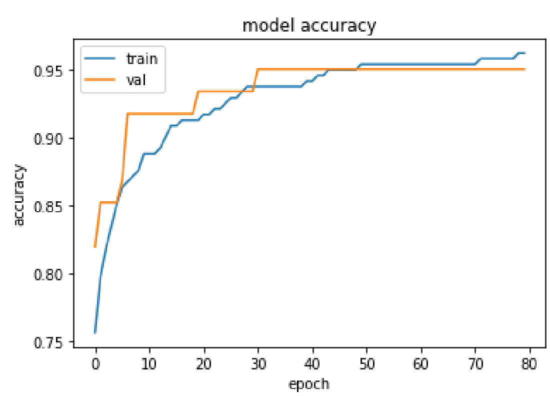


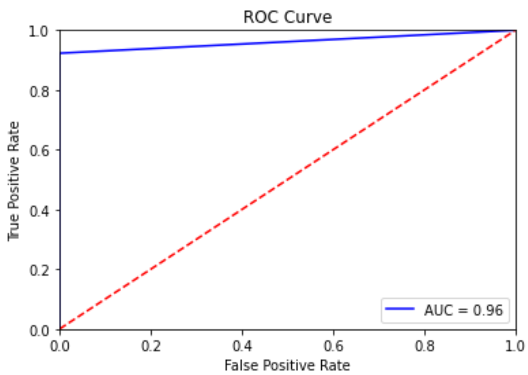




ANN(Best Result) : Accuracy 95.08%







Confusion Matrix

	Predicted Positive	*Predicted Negative*
Actual Positive	*35*	*0*
Actual Negative	*3*	*23*

Result Comparison

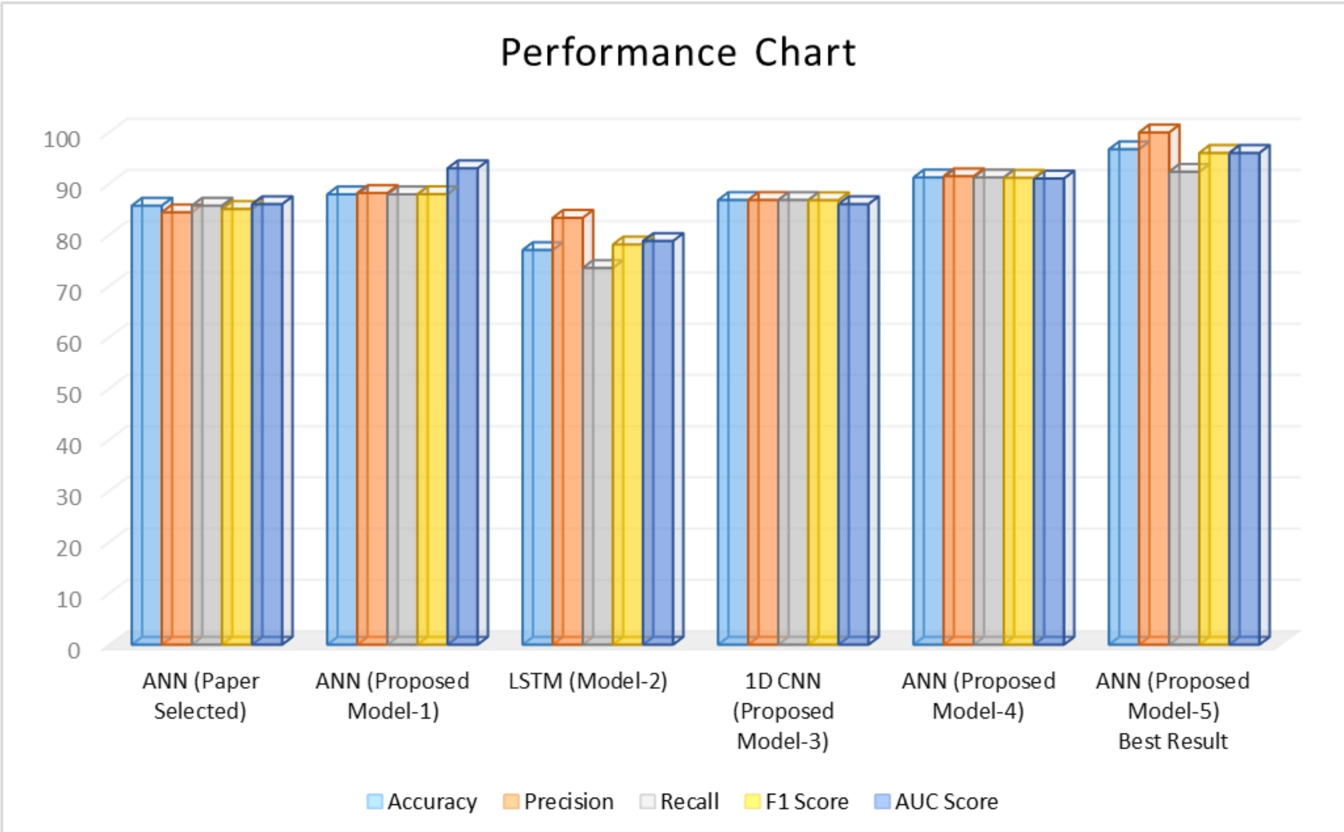
Paper	Model	Accuracy	Precision	Recall	F1-Score	No Of Hidden Lair
Olaniyi, E. O., Oyedotun, O. K., Helwan, A., & Adnan, K. (2015). Neural network diagnosis of heart disease.	ANN	85%				6
Lin, C.-H., Yang, P.-K., Lin, Y.-C., & Fu, P.-K. (2020). On Machine Learning Models for Heart Disease Diagnosis. 2020 IEEE 2nd Eurasia	ANN	91.26%				1
	CNN	83.50%				3
Terrada, O., Cherradi, B., Hamida, S., Raihani, A., Moujahid, H., & Bouattane, O. (2020). Prediction of Patients with Heart Disease using Artificial Neural Network and Adaptive Boosting techniques. 2020	ANN	91.41%	79.67%	70.36%	75.98%	3

Table: *Previous Result*

F. Evaluation

Model Name	Preprocessing	No of Input Layer	No of Hidden Layer	Neurons /Filters	Activation Function & Loss Function	Optimizer & Learning Rate	Epoch	Result				
								Accuracy	Precision	Recall	F1 Score	AUC Score
ANN (Paper Selected)	Scaling	13	6	5	*Sigmoid* *MSE*	*Adam* *0.0032*	2000	85.71%	84.41%	85.71%	85.05%	86%
ANN (Proposed Model-1)	Scaling	13	3	12	*Relu* *BCE*	*Adam* *0.01*	200	87.91%	88.17%	87.91%	87.95%	93%
LSTM (Model-2)	Scaling	13	4	100	*Relu* *BCE*	*Adam* *0.001*	90	77.04%	83.33%	73.53%	78.12%	78%
1D CNN (Proposed Model-3)	Scaling	13	2	128	*Relu* *BCE*	*Adam* *0.01*	15	86.81%	86.81%	86.81%	86.77%	86%
ANN (Proposed Model-4)	Feature Selection with Scaling	10	3	100	*Relu* *BCE*	*Adam* *0.01*	125	91.21%	91.44%	91.21%	91.14%	91%
ANN (Proposed Model-5) Best Result	Scaling	13	1	300	*Relu* *BCE with LogitLoss*	*SGD* *0.01*	80	96.72%	100%	92.31%	96%	96%

Table: *Our Result*



V. Discussion

Our Experiment yielded good results.We have successfully obtained better results than above mentioned papers. In future performing hyperparameter tuning may increase our result.

References

1. Neural Network Diagnosis of Heart Disease https://ieeexplore.ieee.org/document/7323241
2. A Comparative Study on Heart Disease Prediction Using Data Mining Techniques and Feature Selection https://ieeexplore.ieee.org/document/9331158
3. Prediction of Patients with Heart Disease using Artificial Neural Network and Adaptive Boosting techniques |https://ieeexplore.ieee.org/document/9199620
4. https://www.ibm.com/topics/logistic-regression
5. Decision tree - Wikipedia
6. https://www.analyticsvidhya.com/blog/2021/04/simple-understanding-and-implementation-of-knn-algorithm/
7. https://www.ibm.com/cloud/learn/random-forest