Heart Failure Prediction

Satwik Goyal and Gagan Biradar
The University of Massachusetts, Amherst

***Abstract*—In the United States, about 6.2 million adults have heart failures and in 2018 about 379,800 death certificates (13.4%) mentioned heart failures as the primary cause of death. The correct prediction of heart failures can help prevent them and hence save lives, whereas incorrect prediction of heart failures could be life-threatening. Heart Failure Prediction is a field where the use of deep learning models has increased in recent years. In this research paper, the focus is from a data analytical point of view on heart diseases. This study aims at exploring different non-neural and neural network deep learning algorithms for heart failure prediction using the Heart Failure Prediction Dataset by *fedesoriano* on Kaggle. The experiment starts with a pre-processing of the dataset and then applying different deep learning algorithms to find the model that predicts heart failures with the highest accuracy. We found that our MLPNN model best predicts heart failures with an accuracy of 90.5%.**

I. INTRODUCTION

The Health Care field is the world's largest and most important industry. Being such an important field there is an intense amount of research being conducted to develop deep-learning models that can diagnose diseases and predict their effects on a person.

Currently, there are experts in the field that look at patient reports and suggest procedures accordingly. Most of the time such examinations are very expensive and are not affordable by many people. Depending on the test-time complexity, most deep-learning models have the capability to go through data faster than humans. Previous studies have shown that Artificial Intelligence can do as well or even better than experts in the field of medical diagnosis[6]. Developing accurate deep learning models for heart failure prediction would make these clinical examinations more scalable, affordable, and efficient.

In the United States, a staggering 6.2 million adults are reported to have had heart failure. Moreover, in 2018 about 379,800 death certificates (13.4%) mentioned heart failures as the primary cause of death[2],[3]. The correct prediction of heart failures can help prevent them and hence save lives, whereas incorrect prediction of heart failures could be fatal. We can use deep learning models to predict heart failures based on factors such as Age, Sex, Chest Pain Type, Cholesterol, Resting Blood Pressure, and Resting ECG. If deep-learning models are built that can predict heart failures with high accuracy, then precautionary measures can be taken to prevent them and hence save lives.

A. MODELS

This research offers to approach the problem by providing a comparative analysis of using K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Naive Bayes (NB), Decision Tree (DT), and a custom engineered Multilayer Perceptron Neural Network (MLPNN) to predict heart failures and find which model(s) provides the best results.

B. AIM

The aim of this research is to build efficient scalable models that can predict cardiovascular diseases that may lead to heart failure with higher accuracy than existing systems.

II. BACKGROUND/RELATED WORK

A. INTELLIGENT HEART DISEASE PREDICTION USING NEURAL NETWORK BY SOUMONOS MUKHERJEE AND ANSHUL SHARMA[5]

In this literature the authors propose to develop a Neural Network Model that outperforms existing systems for predicting cardiovascular disease.

TABLE I
PERFORMANCE ANALYSIS OF PREVIOUS SYSTEMS DESIGNED

	Machine Learning Model	Accuracy
1	Naive Bayes	73%
2	Decision Tree	68%
3	K-Nearest Neighbours	70%
4	Support Vector Machine	80%
5	Clustering (Foggy k-means)	75%

The dataset that was used in this research were obtained from University of California Irvine (UCI) and Physionet data repositories. Their dataset contained 76 attributes for each of 303 patients from 3 countries. After using reduction techniques they end up using 14 principal attributes from the dataset for each data point. In their research Mukherjee and Sharma build a 1D convolutional neural network in crystal with a kernel size 16, and 64 filters and a default stride of 1 using the Tensorflow framework. Using this model they were able to achieve a ground-breaking testing accuracy of 97%. For this study the heart data of patients was sequential (data points are correlated/dependent on each other). A 1D convolutional neural network works best with sequential data. The authors of this paper had the ECG of patients over a time period, however, we do not have any sequential data in our dataset. Therefore, we plan to build a custom engineered neural network and run it on our dataset to see if we can achieve high accuracies.

B. *IMPLEMENTATION OF MACHINE LEARNING MODEL TO PREDICT HEART FAILURE DISEASE BY FAHD SALEH ALOTAIBI*[7]

In this study, the author discussed previous approaches to Heart Failure Prediction by the University of California Irving using machine learning models in Rapid Miner, Matlab, and Weka.

TABLE II

PERFORMANCE ANALYSIS OF PREVIOUS SYSTEMS DESIGNED BY UCI

	Machine Learning Model	Rapid Miner	Matlab	Weka
1	Decision Tree	82.22%	60.9%	67.7%
2	Logistic Regression	82.56%	65.3%	67.3%
3	Random Forest	84.17%	X%	X%
4	Naive Bayes	84.24%	X%	X%
5	Support Vector Machine	84.45%	67%	63.9%

The dataset used that was used in this research were obtained from University of California Irvine data repository. The dataset had about 14 attributes for each data point representing 303 patients.
To improve the accuracies from previous works this study chose to use 10 Fold Cross Validation over previous 5 Fold Cross Validation technique. The research further tuned hyperparameters using random number generation and normalization techniques. The highest accuracy achieved by this research was 93.19% using a Decision Tree Model. In our study we plan to use random number generation and various normalization techniques to achieve a high accuracy on out our dataset.

TABLE III

PERFORMANCE ANALYSIS OF SYSTEMS DESIGNED BY ALOTAIBI

	Machine Learning Model	Accuracy
1	Decision Tree	93.19%
2	Logistic Regression	87.36%
3	Random Forest	89.14%
4	Naive Bayes	87.27%
5	Support Vector Machine	92.30%

C. *A DATA MINING APPROACH FOR PREDICTION OF HEART DISEASE USING NEURAL NETWORKS BY CHAITRALI S. DANGARE AND SULABHA S. APTE FROM INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND TECHNOLOGY*[1]

In this study, the authors build a neural network-based Decision Support System for Heart Disease Prediction. The dataset used contains 15 attributes for each patient. The data contains patients' smoking habits, obesity, sex, blood pressure, and cholesterol-like attributes. The authors use a Multilayer Perceptron Neural Network (MLP or MLPNN) to model the data in their experiment. The authors are able to achieve an accuracy of nearly 100%. However, this high accuracy could have been a result of some bias such as a certain random state that returns high accuracy, overfitting dimension(s), and hazy line of demarcation. So no comments can be made about the reproducibility of the results. We plan to build our own Multilayer Perceptron Neural Network and model our dataset.

III. APPROACH AND METHODOLOGY

A. *PREPROCESSING*

We preprocess the ordinal data by feature scaling by subtracting each attribute column by its means and then diving it by its standard deviation. We preprocess the non-ordinal data by using One-Hot-Encoder provided by sklearn and transform the data for Sex, Chest Pain Type, Resting Electrocardiogram, Exercise Angina, and ST Slope into multiple binary columns.

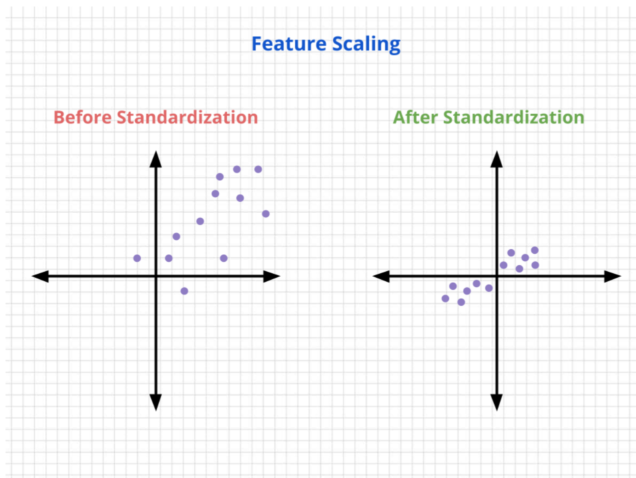

Fig. 1. Feature Scaling (Made using Google Draw)

B. *K-FOLD CROSS VALIDATION*

Since our dataset has only 918 data points, K-fold cross validation will enable us to perform a resampling procedure to evaluate the models. We first divide the training dataset into k sets. Then we train the data on k-1 sets, test it on 1 remaining set, and repeat the procedure for all possible combinations. At the end of the procedure the accuracy through all iterations/combinations is averaged and the model with the highest average accuracy is selected. We are using k = 5. To perform K-fold cross validation we will use the sklearn framework.

C. *K-NEAREST NEIGHBOURS (KNN)*

K-Nearest Neighbours finds the distance of a test data point from all training data points and classifies the test data points based on the label with highest frequency among the k-nearest training data points. For example, in the figure below we can see that for the new test data point the model first finds the 5 closest training data points to it. Since, 4 of those 5 closest training data points are labeled 'No Heart Disease', the new data point will be labelled 'No Heart Disease' as well. To implement the K-Nearest Neighbours algorithm we will use the sklearn KNeighborsClassifier.

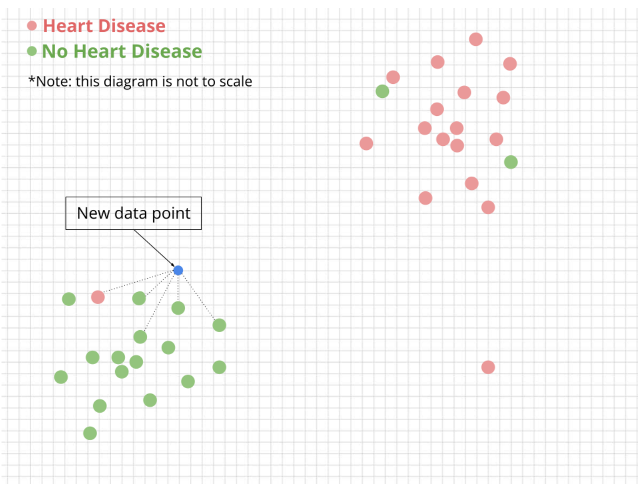

Fig. 2. KNN Classification Example (Made using Google Draw)

D. SUPPORT VECTOR MACHINE (SVM)

A Support Vector Machine plots all training data points on a D-dimensianl plane where D is the dimension of one data point (i.e. number of attributes). Then it separates the data points for the 2 classes by creates a line or hyperplane between them such the margin between the 2 classes is the largest. Depending on which side of the line/hyperplane the testing data point falls, SVM classifies the data point. For example, in the figure below SVM finds a line such that it maximizes the margin between the points from the 2 classes. Since the new point falls on the left side of line, it is classified as 'No Heart Disease'. To implement Support Vector Machine algorithm we will use the sklearn LinearSVC.

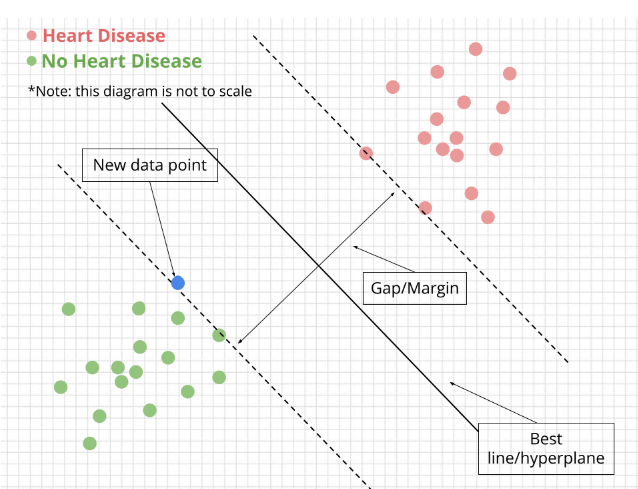

Fig. 3. SVM Classification Example (Made using Google Draw)

E. DECISION TREE (DT)

As the name suggests a decision tree is a tree like structure where each non-leaf node corresponds to an attribute and their branches corresponds to a decision to be made between the values (or range of values) that attribute can take. The leaf nodes corresponds to the classes. The tree model is incrementally made by running over the dataset and updating the probabilities under each branch.

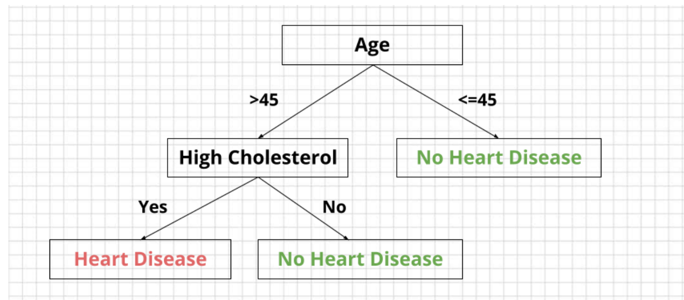

Fig. 4. DT Classification Example (Made using Google Draw)

F. LOGISTIC REGRESSION (LR)

Logistic Regression is a linear method that uses a logistic/sigmoid function representation to predict the test data point's class. It is usually used when the classification problem is binary. Logistic regressions models the probability of each class using the dataset. The data points from both classes are then divided by a certain threshold, any test point that lies below the threshold belongs to class 0, and any test point on or above the threshold belongs to class 1. To implement Logistic Regression model we will use the sklearn LogisticRegression.

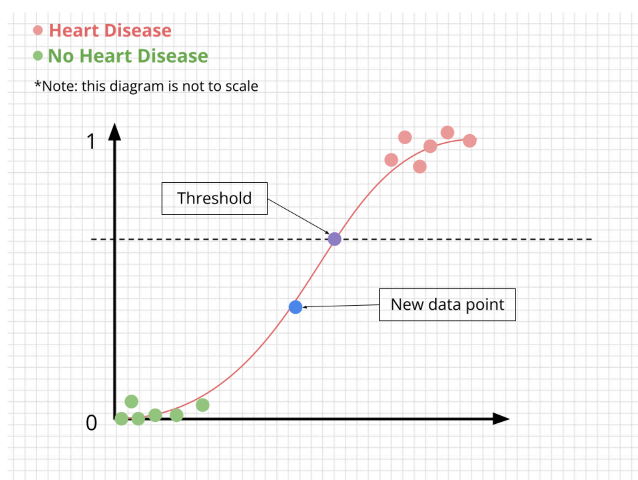

Fig. 5. LR Classification Example (Made using Google Draw)

G. NAIVE BAYES (NB)

Naive Bayes classifiers use the Bayes Theorem to find the probability of a class given some attributes. Then the class with the highest probability is returned. These probability distribution models are constructed using the training data points. To perform Naive Bayes classification we will use the sklearn GaussianNB.

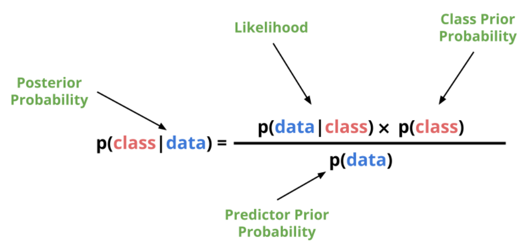

Fig. 6. The Naive Bayesian Model (Made using Google Draw)

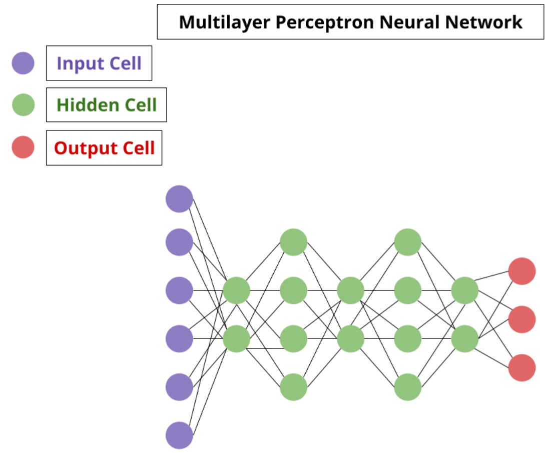

Fig. 7. MLPNN Example (Made using Google Draw)

H. MULTILAYER PERCEPTRON NEURAL NETWORK (MLPNN)

A Multilayer Perceptron Neural Network consists of an input layer, multiple hidden layers, and and output layer. All these layers are connected through weighted connection links. During the training process data is passed through the network from the input layer and as they go through the layers the data is multiplied by the weights and added up and sent to the next layer. Finally, depending on the loss function, scores are assigned to each class. Gradients are then calculated with respect to these scores so that the weights and biases can be tuned according to a optimization function, so as to minimize errors; this process is known as Backpropagation.

1) MLPNN 1
   – To implement Multilayer Perceptron Neural Network 1 we will use the sklearn MLPClassifier. For this network we have two hidden layers each of size 10, with relu activations between the layers.
2) MLPNN 2
   – For this network we used the TensorFlow package to construct a network with 5 layers with sizes 80, 120, 80, 120, 80 respectively. Since, this has a lot of layers, as the data goes through each layer it becomes really sensitive to the initial weight scale. So, to prevent the output to have scores really close to zero or really high values, we standardized the inputs for each mini-batch after each layer and then rescale them, this process is called batchnormalization. Lastly, we also perform Dropout, a regularization technique used to prevent overfitting by omitting neurons while training.

I. DATASET

We will be using the Heart Failure Prediction Dataset by fedesoriano on Kaggle[4]. The dataset has the following attributes/features: Age, Sex, Chest Pain Type, Resting BP, Cholesterol, Fasting Blood Sugar, Resting Electrocardiogram, Maximum Heart Rate, Exercise Angina, Old Peak, ST Slope, and Heart Disease. Below in the table is a description for each attribute and the distinct values it may have.
There are a total of 918 patient samples out of which 508 samples correspond to patients with a heart disease that may lead to heart failure and 410 samples of patients with no heart disease. 70% (i.e. 642 data points) of the data will be used to generate the training dataset.

IV. EXPERIMENT

A. EVALUATION

We plan to test the models based on 3 scoring metrics: accuracy, recall, and precision. The models will be tested on both Training and Testing datasets. 30% of the data (i.e. 276 data points) of the data will be used to generate the testing dataset.
Accuracy is one of the most important metrics to judge a model's performance. The accuracy of a model gives us an insight into how correctly the model can predict heart failures. Precision and recall are other two important metrics that help judge a model's performance. Both these metrics are based on relevance. Precision gives us an insight into relevant instances among the retrieved instances, while recall gives us an insight into relevant instances that were retrieved.

B. RESULTS

We achieved really high accuracies across all models, over 86 % to be exact. The KNN model achieved slightly above 89% with K = 11 neighbours. Using SVMs we achieved a slightly higher accuracy of 89.49% when we picked $\lambda = 1$ as the regularization parameter. Logistic regression with a regularization strength of 5 on the other hand shares a similar accuracy as that of the KNN model. With the Naive Bayes model the accuracy

TABLE IV
ATTRIBUTE DESCRIPTIONS AND RANGE OF VALUES

Attribute Description	Distinct Possible Values
Age: Quantifies how many years old the patient is	[28, 77]
Sex: Gender of the patient in a binary system	{1: Male, 0: Female}
Chest Pain Type: Type of chest pain a patient is suffering, i.e. ASY, ATA, NAP, or TA as one hot encoded vectors	{[1, 0, 0]: ATA, [0, 1, 0]: NAP, [0, 0, 0]: ASY, [0, 0, 1]: TA}
Resting Blood Pressure: Blood Pressure of the patient upon admission	[80, 200]
Cholesterol: Cholesterol Level of the patient upon admission	[85, 603]
Fasting Blood Sugar: Desribes if the patients fasting blood sugar level is above 120mg/dl or not	{1: Yes, 0: No}
Resting Electrocardiogram: Type of electrical activity of the heart, i.e. Normal, ST, LVH as one hot encoded vectors	{[1, 0]: Normal, [0, 1]: ST, [0, 0]: LVH}
Maximum Heart Rate: The rate at which the heart is beating	[60, 202]
Exercise Angina: Describes if exercise produces angina or not	{1: Yes, 0: No}
Old Peak: Quantifies the patient's depression status	[-2.6, 6.2]
ST Slope: Patient's condition during peak exercise, i.e. Up sloping, Flat, Down sloping	{[0, 1]: Up, [1, 0]: Flat, [0, 0]: Down}
Heart Disease: Describes whether the patient has the heart disease or not	{1: Yes, 0: No}

plunges below the 89% mark and attains an accuracy of 88.7%. Compared to other models we noticed that the decision tree model with a depth of 4 performs the worst across all metrics compared to all other models. We didn't notice any significant increase in performance as the depth of the tree was increased. Our MLPNN models achieved the highest accuracies, especially our 5-layer network which crossed the 90% barrier. Moreover, across all our models it can be seen that our testing accuracies were really close to our training accuracies which implies that our models generalized really well.

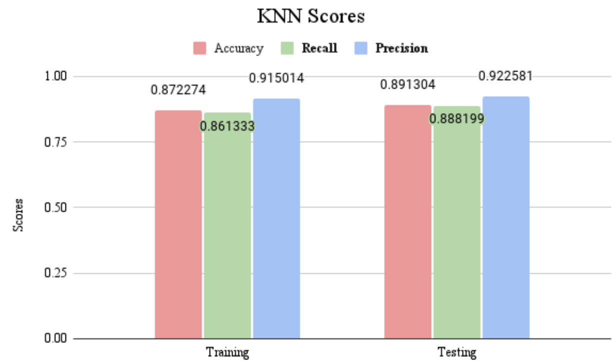

Fig. 8. KNN Scores

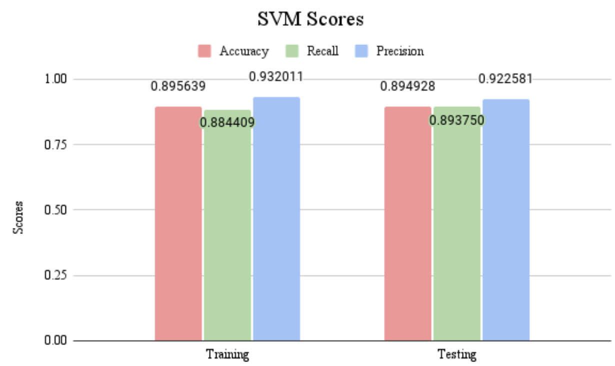

Fig. 9. SVM Scores

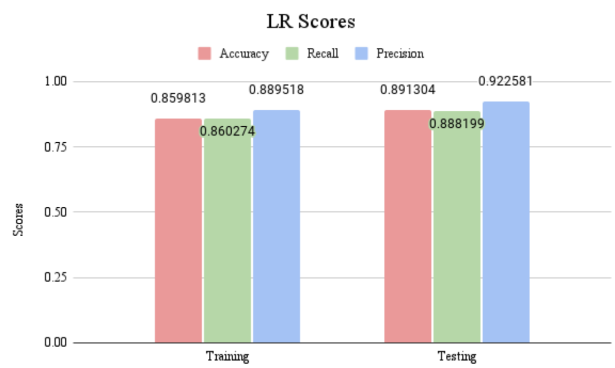

Fig. 10. Logistic Regression Scores

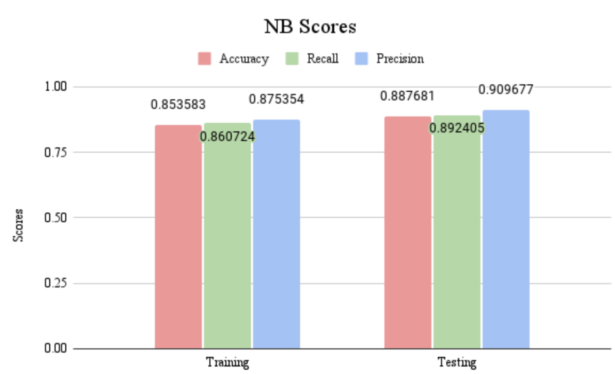

Fig. 11. Naive Bayes Scores

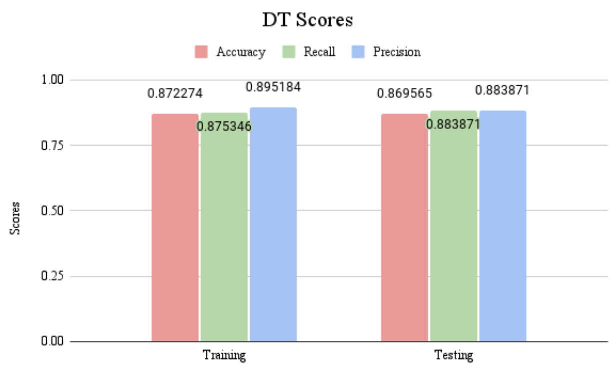

Fig. 12. Decision Tree Scores

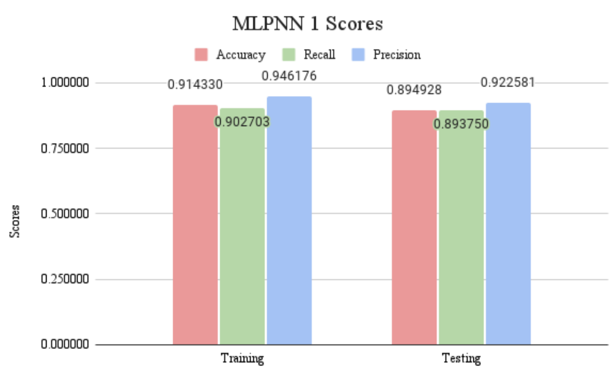

Fig. 13. MLPNN 1 Scores

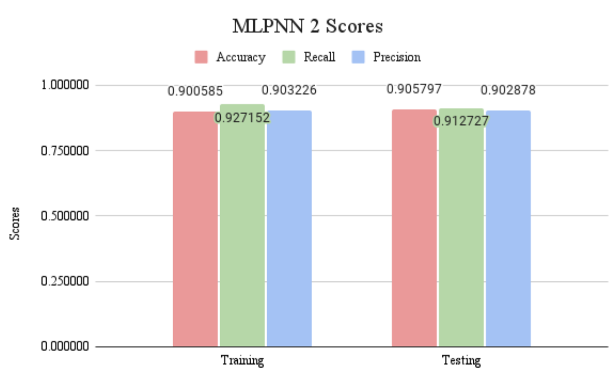

Fig. 14. MLPNN 2 Scores

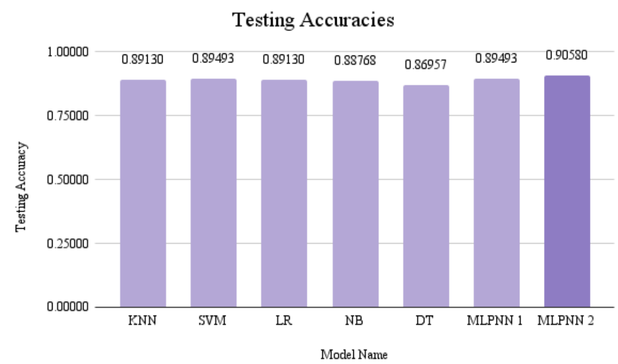

Fig. 15. Testing Accuracies

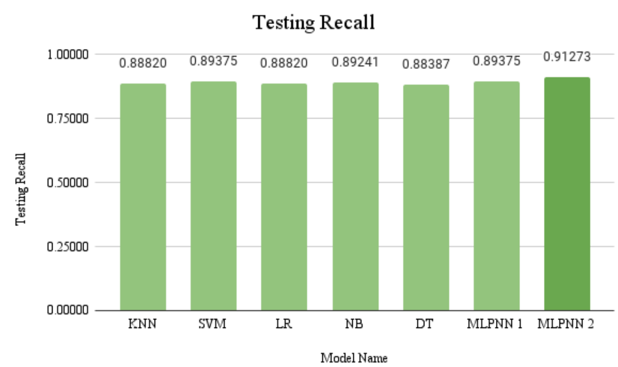

Fig. 16. Testing Recall

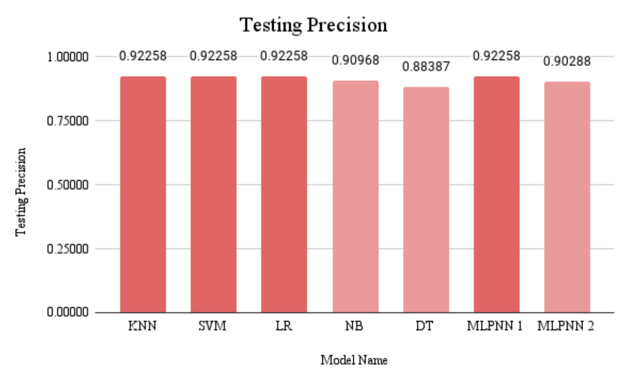

Fig. 17. Testing Precision

V. CONCLUSION

To make health care more accessible to the masses it is imperative to have fast and cheap systems. Machine learning approaches paves us one such path, which was shown extensively in this study. We discussed different machine learning models and demonstrated their performance on diagnosing heart diseases among patients. Similar to other studies, we managed to achieve an accuracy of 90.5% using our Neural Network architecture on a more extensive dataset. Though this study showed high performance on classifying whether a patient has a heart disease or not, efforts need to be put into amplifying the size of the dataset. Moreover, using other forms of data in tandem could also result in higher accuracies, such as using ECG records. It would also be interesting to see AI algorithms achieving high performance in detecting potential disorder amongst other parts of the bodies as well. Thus, this could allow future doctors and surgeons to go over timely fully body checkup quickly and efficiently.

REFERENCES

[1] Dangare, Chaitrali, and Sulabha Apte. "(PDF) A Data Mining Approach For Prediction Of Heart Disease Using Neural Networks". Researchgate, 2021, https://www.researchgate.net/publication/254938414_A_Data_Mining_Approach_for_Prediction_of_Heart_Disease_Using_Neural_Networks.

[2] "Heart Failure | Cdc.Gov". Centers For Disease Control And Prevention, 2021, https://www.cdc.gov/heartdisease/heart_failure.htm.

[3] "Heart Disease Facts | Cdc.Gov". Centers For Disease Control And Prevention, 2021, https://www.cdc.gov/heartdisease/facts.htm#:~:text=About%20659%2C000%20people%20in%20the,1%20in%20every%204%20deaths.&text=Heart%20disease%20costs%20the%20United,year%20from%202016%20to%202017.

[4] "Heart Failure Prediction Dataset". Kaggle.Com, 2021, https://www.kaggle.com/fedesoriano/heart-failure-prediction.

[5] Mukherjee, Soumonos, and Anshul Sharma. Researchgate, 2021, https://www.researchgate.net/publication/331112206_Intelligent_heart_disease_prediction_using_neural_network.

[6] Shen, Jiayi et al. "Artificial Intelligence Versus Clinicians in Disease Diagnosis: Systematic Review." JMIR medical informatics vol. 7,3 e10010. 16 Aug. 2019, doi:10.2196/10010

[7] Alotaibi, F., 2019. Implementation of Machine Learning Model to Predict Heart Failure Disease. 6th ed. International Journal of Advanced Computer Science and Applications, pp.261-268. Available at: https://pdfs.semanticscholar.org/a74f/d8c51251e8c6126a1527e545bd78860a10f9.pdf.