Enhanced Heart Disease Prediction Using Machine And Deep Learning Techniques

Naveen M
Dept of Computer Science and Engineering,
VIT University, Vellore, India

Abstract- *Nowadays heart disease is the very most critical disease affected by many peoples. Hence the ML classification techniques have chosen authentic production choices and forecasts with vast quantity of data. Machine learning and feature extraction are playing very vital role in internet sector and health care domain. DL is also the one of the branches in ML something else ANN algorithms simulate from the human mind. In this project I proposed the logical correct system to predict the heart diseases to using the various types of machine learning and Deep learning techniques. Additionally, I proposed the one deep learning algorithm to predict the heart disease. And this project is based on some categorical and numerical data to predict the heart disease of patients. In this case I used the accuracy as a Random Forest algorithm (Hard and Soft method), Decision tree and the LSTM algorithm. The proposed method results shows that the high accuracy data contrast to the previous author works. The results shows that the dataset have 93% accuracy through the prediction of the cardio vascular disease.*

Keywords- Machine Learning, Heart Disease, Deep Learning, Cardiovascular Disease, Classification Algorithms

I. INTRODUCTION

Heart disease is the most critical disease several peoples are affected in the world. HD contains occur with major symptoms like the chest pain, high blood pressure, stroke, breathing issues etc., The HD mainly affected by the blood vessels. Hence the problem statement can be applied to the machine learning classification techniques. ML contains the ground of the training bounces processers capacity to study without presence clearly programmed. In ML classification can be separated as supervised leaning and unsupervised learning. The High blood pressure contains the damages the arteries to less the passes of the blood and the oxygen from the heart to conduct the heart disease. In other case, the less blood flow from the heart can causes the angina. The WHO shows the list the heart disease patients can widely 21.8% death cases through every year. The bulk of the people are affected by the heart disease in the America. Moreover, the most of the methods can deficient to detecting the HD Patients.

To address the problem statement, I proposed the establish of HD Prediction created on the ML & DL Classification procedures to gives accuracy and accurate results. The main aim of the project is to detect the heart disease in the primitive stage. It is very easy to prevent the patient's life in the early stage. The Proposed method can have been tested as Stat log (Heart) Dataset to identify the heart disease form the patients.

In Machine learning technique classification models include the RF like hard and soft method, DT to detect the HD. Additionally I proposed the Deep Learning Classification like LSTM Recurrent Neural Network algorithm to predict the HD. Additionally, I used the Confusion matrix technique to forecast the HD correctly. Furthermore, performance measure and the exactness and give right outcomes of the proposed work very well than the former research papers. In the ML & DL algorithms are very useful to predict heart disease using patient's dataset. It is very useful to predict before the heart attack and the heavy chest pain is high. It is very helpful to measure the performance of the algorithms to predict the heart disease from the supervised learning to solve the various types of real time problems.

II. RELATED WORK

In Literature survey the authors are mostly applied the Machine learning and Data mining techniques to predict the Heart Disease Prediction. I study the some of the main machine learning algorithms important to the proposed work. Kapil Wankhade [1] proposed the Decision support system to predict the heart disease using Support Vector Machine (SVM) Classification as 80.41% accuracy. They show the results in the Multilayer Perceptron neural network (MLPNN) to predict the heart disease. HumarKahramanli [2] proposed the Artificial neural network (ANN) and Fuzzy neural network (FNN) algorithm to classify the data. And they tried the k fold cross validation techniques to predict the heart disease. They used the prima Indians Diabetes dataset it gives the 84.24% accuracy. Akhil Jabbar [3] proposed the Artificial Neural Networks and the feature selection algorithm to classify the heart disease. They used the whole Andhra Pradesh state heart disease database to predict the heart diseases. That Heart

disease dataset gives the 82.23% accuracy. Xia Liu [4] proposed the 2 feature subsystems first one is RFRS feature Selection system and second one is Classification system in the ensemble Classifier. They used the Statlog dataset it gives the 92.59% accuracy. Amin UI Haq [5] proposed the various types of Machine learning classifications to predict the heart disease dataset and the backward Feature selection algorithm used to increase the accuracy and the performance level. They used the Cleveland heart disease dataset to predict the heart disease using the algorithm like KNN and gave the accuracy as 90% correctly. Gamal G.N. Geweid [6] proposed the new method using the SVM Machine learning algorithm to Heart Failure Detection HFD. They identified the heart disease using the ECG signals. They show that the results as the given good accuracy 94.97%. Gregg C. Fonarow [7] they proposed the epidemiological dataset on the Heart failure HF and decrease the mounting heart failure HF disease. The dataset contains give the results correctly. SamehGhwanmeh [8] they proposed the ANN and the Decision support system to detecting the three main heart disease like the mitral stenosis, aortic stenosis and the ventricular spectral. Compare the previous research work these algorithm gives the accuracy as 92%.

Manasa . S [9] they proposed the Machine learning algorithms like the LR, KNN, SVM, DT, NB, RF to predict the heart disease. And they also used the feature selection algorithm to increase the accuracy of the Machine learning algorithm classification to gave the accuracy 93%. SellappanPalaniappan [10] to predict the heart disease using the data mining techniques like the decision tree, naïve bayes, and the neural networks. They find the heart disease using the Intelligent Heart disease prediction system (IHDPS) in the application. They used the .NET platform to detect the heart disease and gave the accuracy as 90%. Khashman Adnan [11] proposed the Neural networks attribution to predict the heart disease. They used Intelligent system like Machine learning techniques to diagnose the heart disease to prediction. They used the Statlog heart dataset to finding the heart disease of the human beings to using the multilayer perceptron and the SVM and gave the accuracy as 87.5%.

Gautam Srivastava [12] proposed a novel approach and applying the machine learning algorithm classification techniques improved the accuracy to finding the heart disease prediction. They used the UCI dataset to predict the heart disease. It gave the accuracy as 88.7% to predict the heart disease via the hybrid random forest linear model (HRFLM). Oluwarotimi Williams Samuel [13] proposed a new technique like fuzzy_AHP and the ANN to predict the HDF. In this paper they used the UCI dataset like 13 attributes to apply the algorithm to predict the heart disease prediction. And they newly proposed the performance features are very popularly to predict the heart disease and gave the accuracy as 89.01%. Resul Das [14] they proposed the algorithm like the neural networks to predict the Effective diagnosis of heart disease. They used a new SAS software to predict the heart disease with the help of UCI dataset. This model contains the Posterior probabilities to predict the multiprocessors models and gave the accuracy as 80.95%. Amin UI Haq [15] proposed the new technique for a hybrid Intelligent System and the Machine Learning algorithms and feature selection algorithms, cross validation method to predict the heart disease prediction. They also used the Decision support system to diagnosis the heart disease and gave the accuracy as 91.06%.

III. MATERIALS AND METHOD

About the data source:

In the Clinical databases have been collected the powerful amount of the information of the patients details from the doctor. Most of the peoples have affected the heart disease increasing day by day. The world clinical record can be increased by the heart disease only. The term “heart disease” emphases the conditions of the heart and the blood vein and the way the blood pushed and the circulated through the entire body.

This project I used the Cleveland Heart disease dataset. In these 13 attributes. In this attribute it has the numerical data which is stored in the csv file. In this dataset contains 1025 records based on the age and the gender wise. The following table contains the attributes of the patients which I working.

Age	Age in years
Gender	0 = female 1 = male
Cp (Chest Pain Types)	0 = Typical Angima 1 = Atypical Anigma 2 = Non – Anginmal Pain 3 = Asymptopic
Trestbps (Resting Blood Pressure)	in mm Hg on admission to the hospital
Chol (Cholestrol)	serum cholesterol in mg/dl
Fbs (Fasting Blood Sugar)	(Fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
Restecg	resting electrocardiographic results 0 = Normal 1 = Having ST-T Wave Abnormality 2 = Showing Probable or definite left Ventricular Hypertrophy
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina 1 = yes 0 = no
Oldpeak	ST depression included by exercise relative to rest
Slope	The slope of the peak exercise ST Segment 1 = upsloping 2 = flat 3 = down slopping
Ca	Number of major vessels (0-3) colored by fluoroscopy
Thal (Thalassemia)	1 = normal 2 = fixed defect 3 = reversable defect
target	Diagnosis of heart disease (angiographic heart disease status) Value 0 : < 50% diameter narrowing Value 1 : > 50% diameter narrowing

Table 1 : Cleveland Heart Disease Dataset Definitions

Proposed System

This system is mainly developed for identification of the heart disease patients. The performance of the several ML & DL classification techniques to identified the heart disease identification. The algorithms like the Random Forest (Hard and Soft Method), Decision tree and LSTM Recurrent Neural Network to predict the heart disease. The LOSO technique like the confusion matrix (True positive, False Positive, True Negative, True Negative) is the best way to predict the heart disease correctly. And finally, to measure the accuracy performance I conclude the best algorithm classification technique. The Proposed system implements into the step of the Patient's Dataset, Confusion Matrix, Machine learning algorithm classification, Deep Learning Classification and Accuracy Performance analysis.

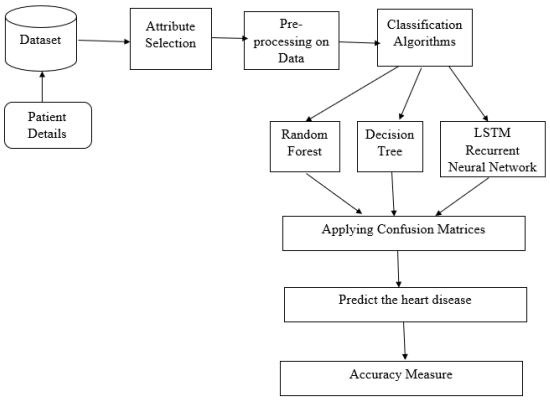


Figure 1 : Proposed Architecture

The Proposed work as follows. In the Supervised Classification experiments have been orderly first collect the dataset from the patient records. The Next step contains to select the attributes selection from the patient dataset to predict the heart disease. After identifying the heart disease to pre-process the data like cleaning, selection, made in the form accordingly. After pre-processing the data to apply the classification algorithms like random forest (hard and soft method), Decision tree and the LSTM Recurrent Neural Network. After applying the classification algorithm to apply the confusion matrix (True positive, False Positive, True Negative, False Negative) to predict the heart disease. After applying the confusion matrix to test performance of the test the accuracy value computed from the different classifiers.

Materials and Methodology

Data Pre-processing

The medical records of patients from four different hospitals are collected for heart disease analysis based on machine learning techniques. Heart disease dataset from the various patient's details is collected after pre-processed. The raw dataset contains some missing values, noise and null standards. Those values take indifferent as of datasets, the residual patient records are applied in the Technique. Finally, pre-processed data are combined, and then the combined dataset is displayed.

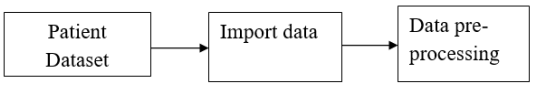


Figure 2: Data pre-processing

Feature Selection Technique

It is the procedure of finding most relevant attribute for a predicting model. The dataset contains 13 attributes of the patients like the age and gender used to identify the personal information. Those attributes are considered that they important for clinical records. It is the process of finding the most relevant attribute for a predicting model. As previously mentioned, several ML & DL techniques are used namely (Decision tree, Random Forest and LSTM Recurrent Neural Network). Hence the classification process repeated with different ML & DL techniques using optimized features. To find the relevant set of attributes, correlation analysis is performed.

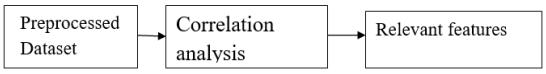


Figure 3 : Feature Selection

Classification Modelling

In direction to make identification, a various classification models' requirement to be accomplished with the patient record dataset and then produce a classification model which is breastfed with a new empty record and the identification of heart disease. The classification technique of datasets completes the foundation of variables, measures of DT features. Each classifier techniques are put in from every dataset command to the evaluation of the concert value. Finest execution models are recognized after the above results.

- Random Forest Algorithm
- Decision Tree Algorithm
- LSTM Recurrent Neural Network Algorithm

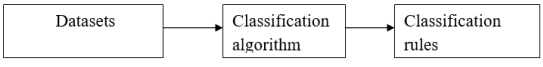


Figure 4 : Classification Algorithm

Random Forest Algorithm

This algorithm is a type of classification techniques in ML classification techniques. Random forest algorithm is ensemble machine learning technique procedure for classification technique, regression constructed by a multiple decision tree at the training dataset time. It is also a Supervised Machine Learning Algorithm.

In Random Forest called because every tree in the forest is built randomly choose the case of the data. In random forest contains the two types one is Soft Method and another one is Hard Method. Based on the classifiers can be used as hard and soft voting method is to predicting or improving and gives the higher accuracy.

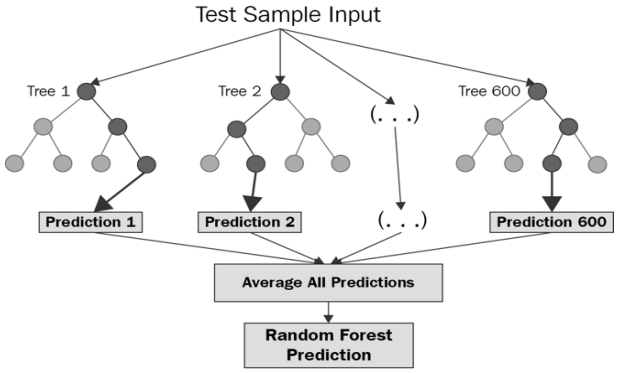


Figure 5: Random Forest Classifier

Decision Tree Algorithm

This algorithm is the one of the flowchart structures. It is also the one of the Machine Learning Classification Method. It contains the leaf node and the branches. A decision tree contains the flowchart shape which contains the internal node and the test on attributes. Decision tree can be used for the classification technique and regression popular the Machine learning and data mining techniques. It separates the small subset compare to the dataset. Figure shows that the decision tree structure. It contains three branches like the young, Middle aged and the senior. It looks like the flowchart structure. In this part the decision tree contains the Yes or no value and fair or excellent.

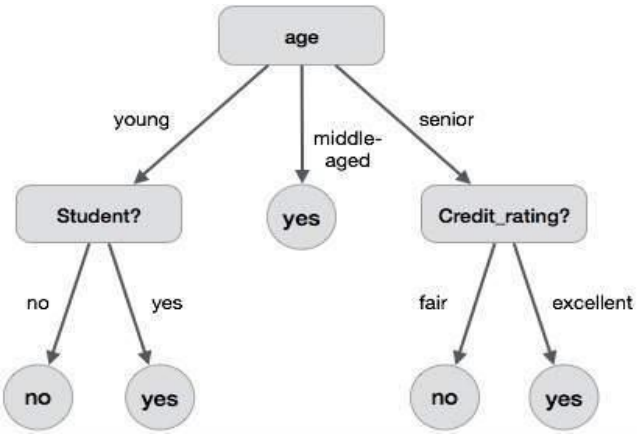


Figure 6 : Decision Tree

LSTM Recurrent Neural Network Algorithm

LSTM can be abbreviated as Long Short Term Memory. It is one of the Deep Learning Algorithm Classifier. It is the artificial Recurrent Neural Network using in the deep learning Methods. It has the feedback connections in the artificial neural network. It is fulfil by the entire sequence of the data.

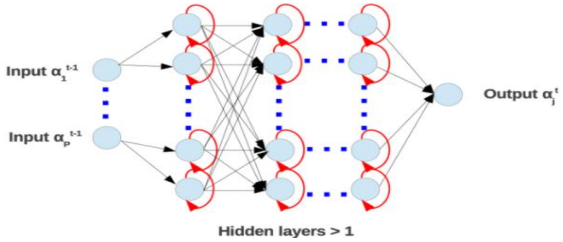
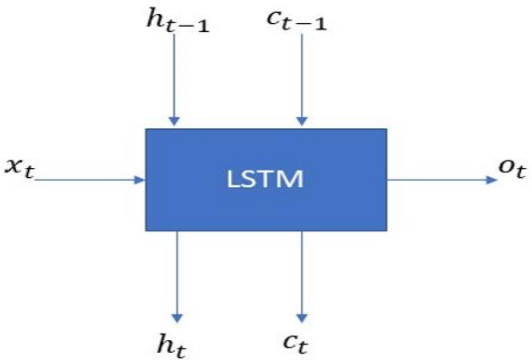


Figure 7 : LSTM Recurrent Neural Network

Mathematically represented as,



Performance Measures

Some normal performance evaluations, exactness, recall, f1-measure support classifying technique can be measured for calculation of performance effectiveness. Thus, outcome of the Predictive model can be provided with the

improved performance level. To recognize the important attributes of HD, performance metrics are applied will support in well thoughtful behaviour several groupings FS technique. I introduce random forest with LSTM Recurrent Network, which generates heavy correctness and low classifying fault of forecasting the HD. To presentation in each separate is assessed separately and all outcomes are appropriately greater more enquiry. These metrics calculation is used with the support of confusion matrix. From these metrics calculation formals are given below :

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100$$

$$Sensitivity = \frac{TP}{(TP + FN)} \times 100$$

$$Specificity = \frac{TN}{(TN + FP)} \times 100$$

$$Precision = \frac{TP}{(TP + FP)} \times 100$$

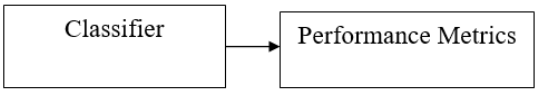


Figure 8 : Performance Metrics

IV. RESULTS AND DISCUSSION

In this section I have discussed about the proposed method results. In this section I shows the screenshot of the results of the proposed method.

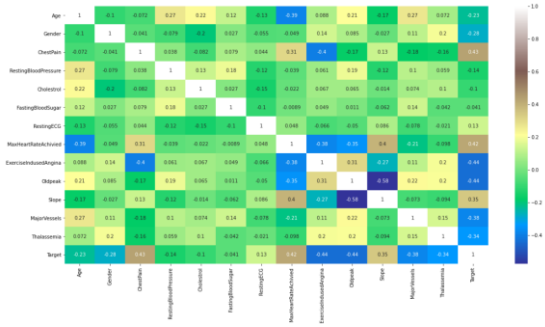


Figure : Heat map for the Correlation Analysis of Heart disease Prediction Cleveland Dataset

Figure shows the heat map for the Correlation Analysis of heart disease prediction Cleveland dataset. The heart what is the relation of age. So, this is negative correlation that means if I have a heart disease it's likely that are of higher age. Gender is of so the negative correlation is with each gender then Resting blood Pressure Cholesterol is very low. Positive correlation between target and the chest pain the more chest pain likely to have the heart attack maximum heart rate achieved. Slope negative correlation between target and gender exchange, major vessel, thalassemia and the Old Peak.

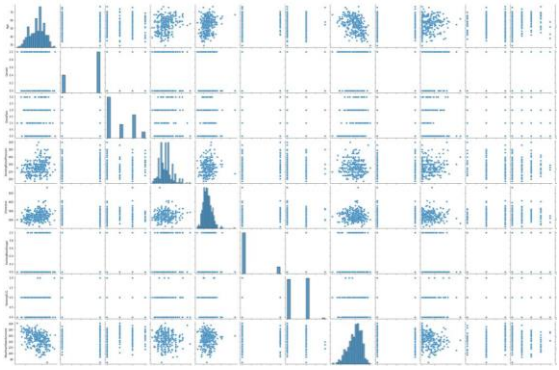


Figure : Data distribution for the correlation between two variables

Figure Shows the Data distribution for the correlation between the two variables with the help of the pair plot. It will be plotted for all the 13 features. I have and it some scattering over here some relationship which its actually is no direction here.

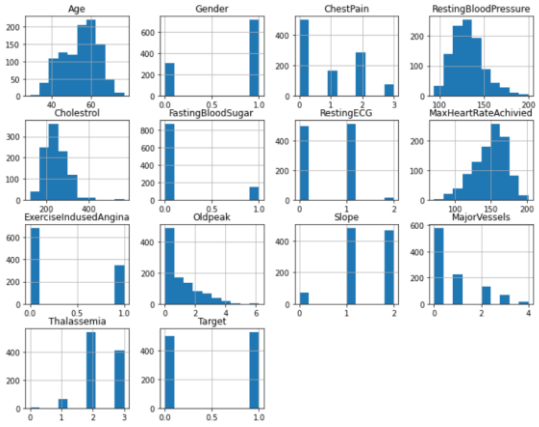


Figure : Cleveland dataset histogram for the heart disease prediction

Figure shows that the shape of my data how does my data looks like in the histogram graph.

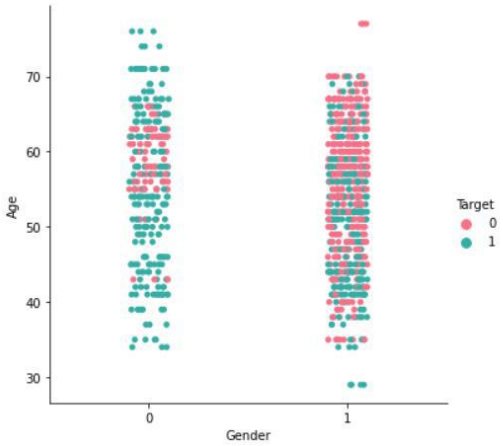


Figure : Gender wise affected heart disease prediction

Figure shows the female having heart disease is plot with the green colour and the range between the age like most of the females are in the age of like more than 42 to 70. Male having sicks doesn't not like extreme low value and also males not having heart attack in the age greater than 70 and see in males that percentage of males not having heart disease is greater as compared to percentage of females.

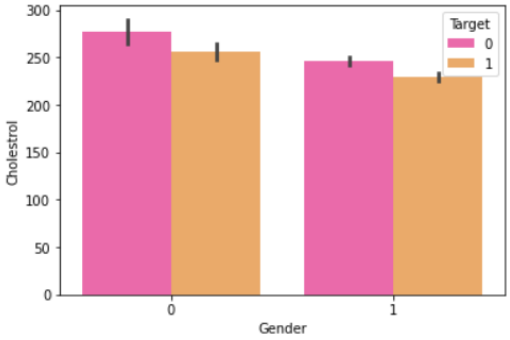


Figure : Gender wise affected by cholesterol

Figure shows that the cholesterol level of various males and females they have. It shows that the females have cholesterol level having heart disease not having heart disease having heart disease greater than 250 not having heart disease less than 270. Male not having heart disease is bust in the cholesterol levels are higher the cholesterol is not that much. Figure 6 It shows that the various type of chest pain affected by based on the age.

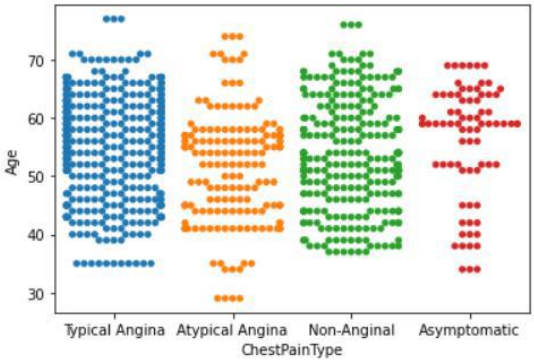


Figure : Different types of chest pain based on age

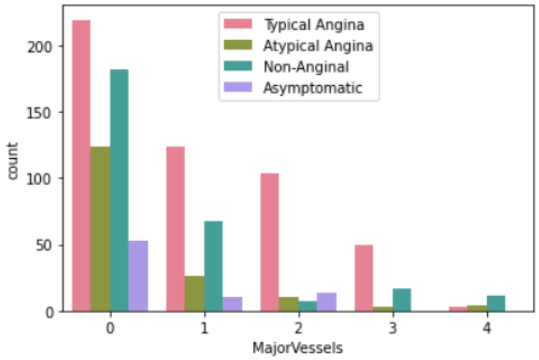


Figure : Types of chest pain based on major vessels

Figure shows that the various type of chest pain affected and check the relation of major vessels is there in the chest pain

	Predicted Active Person 0 (Negative)	Heart disease Predicted Person 1 (Positive)
Actual Active Person 0 (Negative)	True Positive	False Negative
Actual Heart disease Predicted Person 1 (Positive)	False Positive	True Negative

Confusion Matrix

The above table shows it has the confusion matrix. It shows that the True Positive so the True Positive are people who were predicted as having the heart disease in the actually. It shows that the True negative value contains so these people were predicted as the heart disease but they do not have the heart disease. False Positive contains the people who are predicted as not the heart disease but they actually have a heart disease. False Negative contains the Peoples were predicted as not having the heart disease that they actually have a heart disease.

Soft Voting Method

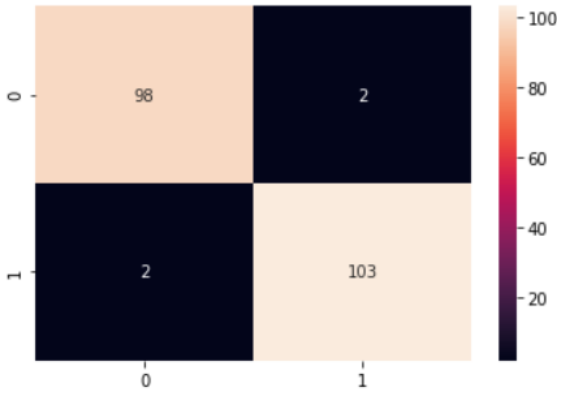


Figure : Heart Disease Prediction using Confusion Matrix for Soft Voting

Figure shows that the Heart disease Prediction using the Random forest Soft Voting Method. 103 refers that the data value request to heart disease presented and my model is also predicted as heart disease predicted correctly. 98 refers that the data value request to heart disease not presented and my model is also predicted as heart disease as heart disease not presented correctly. It shows that the True Positive so the True Positive are people who were predicted as having the heart disease in the actually. It shows that the True negative value contains so these people were predicted as the heart disease but they do not have the heart disease. False Positive contains the people who are predicted as not the heart disease but they actually have a heart disease. False Negative contains the Peoples were predicted as not having the heart disease that they actually have a heart disease.

Hard Voting

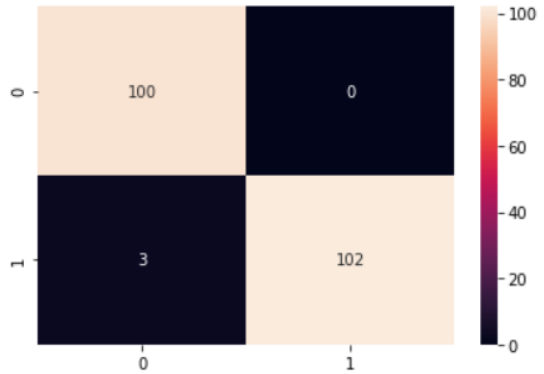


Figure : Heart disease Prediction using Confusion Matrix in Hard Voting

Figure shows that the Heart disease Prediction using the Random forest Hard Voting Method. 102 refers that the data value request to heart disease presented and my model is also predicted as heart disease predicted correctly. 100 refers that the data value request to heart disease not presented and my model is also predicted as heart disease as heart disease not presented correctly. It shows that the True Positive so the True Positive are people who were predicted as having the heart disease in the actually. It shows that the True negative value contains so these people were predicted as the heart disease but they do not have the heart disease. False Positive contains the people who are predicted as not the heart disease but they actually have a heart disease. False Negative contains the Peoples were predicted as not having the heart disease that they actually have a heart disease.

DECISION TREE

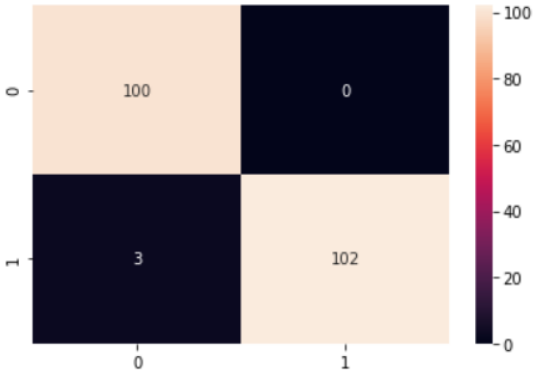


Figure : Heart disease Prediction using Confusion Matrix in Decision Tree

Figure shows that the Heart disease Prediction using the Decision Tree. 102 refers that the data value request to heart disease presented and my model is also predicted as heart disease predicted correctly. 100 refers that the data value request to heart disease not presented and my model is also predicted as heart disease as heart disease not presented correctly. It shows that the True Positive so the True Positive are people who were predicted as having the heart disease in the actually. It shows that the True negative value contains so these people were predicted as the heart disease but they do not have the heart disease. False Positive contains the people who are predicted as not the heart disease but they actually have a heart disease. False Negative contains the Peoples were predicted as not having the heart disease that they actually have a heart disease.

Lstm_reccurent neural network

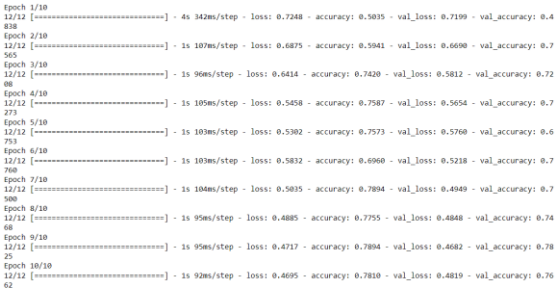


Figure : Training data in LSTM Neural Network

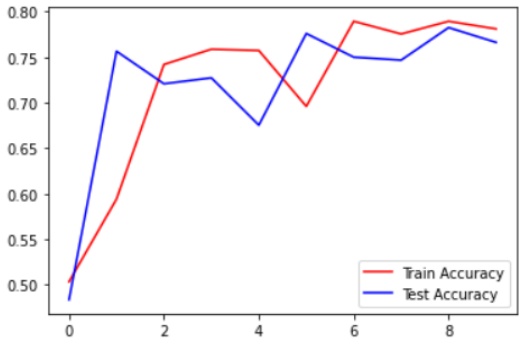


Figure :**Trained Accuracy for the LSTM Recurrent Neural Network**

The figure shows that the test accuracy and trained accuracy. The blue colour contains the test accuracy. The red colour contains the Train accuracy. It shows that the overall testing and training accuracy. It gives as the good trained accuracy value of the system. The best accuracy for the 4 models is Long Short-Term Memory. It gives the overall accuracy results it shows that the Machine learning and Deep Learning classification Method.

Method	Accuracy
Heart Disease using Naïve Bayes	86.12
Artificial Neural network diagnosis	89.02
Decision Tree to finding Heart disease	89.28
Random Forest algorithm to finding the Heart disease	91.89
Using RF (Hard & Soft) and LSTM Neural Network(Proposed Method)	93.31%

Figure : Accuracy Comparison for Existing and Proposed Method

V. CONCLUSION

In this project, an effective machine learning algorithm classification techniques and deep learning technique developed the heart disease prediction. In Machine learning I used Random Forest Hard and soft voting method, Decision tree and in the Deep Learning technique I used the one classifier like LSTM Recurrent Neural Network to designed this project. I conclude that classifying the dispensation of raw health care of HD data will support in the lasting valid of persons breathing and primary stage finding irregularities in kernel circumstances. I used the ML classification and DL classification algorithm methods forecast HD. The Accuracy also increased by using the ML & DL classification Systems. In future I will use the more deep learning system methods to increase Performance to Predict Heart disease. I work on to Predict and recovery the heart disease in future to predict the critical heart disease such as Inflammation of veins, Endocarditis and the Palpitation.

REFERENCES

[1] Gudadhe, M., Wankhade, K., &Dongre, S. (2010, September). Decision support system for heart disease based on support vector machine and artificial neural network. In 2010 International Conference on Computer and Communication Technology (ICCCT) (pp. 741-745). IEEE.

[2] Kahramanli, H., &Allahverdi, N. (2008). Design of a hybrid system for the diabetes and heart diseases. Expert systems with applications, 35(1-2), 82-89.

[3] Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2013). Classification of heart disease using artificial neural network and feature subset selection. Global Journal of Computer Science and Technology Neural & Artificial Intelligence, 13(3), 4-8.

[4] Liu, X., Wang, X., Su, Q., Zhang, M., Zhu, Y., Wang, Q., & Wang, Q. (2017). A hybrid classification system for heart disease diagnosis based on the RFRS method. Computational and mathematical methods in medicine, 2017.

[5] Haq, A. U., Li, J., Memon, M. H., Memon, M. H., Khan, J., &Marium, S. M. (2019, March). Heart disease prediction system using model of machine learning and sequential backward selection algorithm for features selection. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (pp. 1-4). IEEE.

[6] Geweid, G. G., & Abdallah, M. A. (2019). A new automatic identification method of heart failure using improved support vector machine based on duality optimization technique. IEEE Access, 7, 149595-149611.

[7] Bui, A. L., Horwich, T. B., &Fonarow, G. C. (2011). Epidemiology and risk profile of heart failure. Nature Reviews Cardiology, 8(1), 30-41.

[8] Ghwanmeh, S., Mohammad, A., & Al-Ibrahim, A. (2013). Innovative artificial neural networks-based decision support system for heart diseases diagnosis.

[9] porkodi, k., Aishwarya, a., Divya, r., Indhuja, k., &Manasa, s. Efficient classification of heart disease using machine learning algorithm.

[10] Palaniappan, S., & Awang, R. (2008, March). Intelligent heart disease prediction system using data mining techniques. In 2008 IEEE/ACS international conference

on computer systems and applications (pp. 108-115). IEEE.

[11] Olaniyi, E. O., Oyedotun, O. K., & Adnan, K. (2015). Heart diseases diagnosis using neural networks arbitration. International Journal of Intelligent Systems and Applications, 7(12), 72.

[12] Das, R., Turkoglu, I., &Sengur, A. (2009). Effective diagnosis of heart disease through neural networks ensembles. Expert systems with applications, 36(4), 7675-7680.

[13] Samuel, O. W., Asogbon, G. M., Sangaiah, A. K., Fang, P., & Li, G. (2017). An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. Expert Systems with Applications, 68, 163-172.

[14] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE access, 7, 81542-81554.

[15] Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018.