Comparative Analysis of Two Artificial Intelligence Based Decision Level Fusion Models for Heart Disease Prediction

Hafsa Binte Kibria[a], Abdul Matin[a] and Sanzida Islam[a]

[a]*Department of Electrical & Computer Engineering, Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh*

Abstract

Artificial Intelligence is currently a popular theme in health care technology predictions. Machine learning is an artificial intelligence (AI) implementation that automatically learns and builds processes from experience. The leading cause of death worldwide at present is cardiovascular disease. The rate of death could be minimized by detecting the risk early. A lot of machine learning models have been developed to predict heart disease. We also introduced a fusion model to produce a better performance than the existing models. In this study, the proposed method analyzes two decision-making fusion models using five and ten-fold cross-validation to estimate the existence and absence of heart disease. The Jupyter Notebook, Scikit-learn, Tensorflow, and Keras were used as implementation tools. Three machine learning algorithms have been used here: the deep neural network (DNN), logistic regression(LR), and decision tree(DT). The decision tree and the logistic regression were merged separately with the deep neural network to form two fusion models (DNN+DT) and (DNN+LR), model-1 and mode-2. Five performance measurements have been used to compare model performance: accuracy, recall accuracy, f1-score, and AUC score. A significant improvement was found in performance parameters after fusing the algorithms in this work. The fusion occurred at the decision level by adding the decision scores of two algorithms. The main target is to enhance the fused model's performance by combining the individual model's decision for better classification. For both model-1 and 2, the accuracy has increased. Model-2 has obtained 87.12% classification accuracy for 10-fold cross-validation, which is the highest accuracy.

Keywords

Deep neural network, decision tree, logistic regression, heart disease, decision level fusion

1. Introduction

The number one worldwide cause of death is cardiovascular diseases (CVDs). CVDs are a category of problems of the cardiac and blood vessels, including stroke, heart attack, rheumatic heart disease, coronary artery disease, and other illnesses. Due to strokes and cardiac attacks, there are four out of five deaths, and people below the age of 70 die prematurely because of it, which is one-third of these deaths[1]. There are several reasons for CVD: unbalanced diet, smoking, stress, alcohol, fast foods, and inactive lifestyles. A study surveyed in 2016 that over 17 million individuals' deaths are because of cardiovascular disease by the world health organization, which accounts for over 30 percent of deaths worldwide. The same survey found that the number of mortality in underprivileged and medium-income countries is more than 70%.

The good news is that heart diseases can be prevented by avoiding some critical factors, such as poor diet habits and insufficient physical exercise. To control their general state of health and avoid sudden cardiac failure, prompt detection and predictive mechanisms are needed for people who are at risk of high cardiovascular disease. Speaking about predicting heart disease, one of the well-known prediction is machine learning. AI showed promising outcomes in healthcare. In the Journal of Clinical Analysis [2] in a 2012 study; it was reported that machine learning plays a significant role in automatically detecting intricate patterns in radiology applications, and it helped radiologists make smart decisions. Moreover, in 2015[3], the researchers showed that machine learning is essential to improve our understanding of cancer progression. It also has a significant improvement in their accuracy and efficiency in decision making.

Clinical diagnosis is a diagnostic task whereby evaluating the qualities of a variety of features; a doctor tries to classify disease. Traditional approaches to treat heart disease, including ECG, blood pressure, level of cholesterol, etc., are costly and require a lot of time. So, to drop the death rate. It is necessary to design a heart disease diagnosis system that is computerized. The number of physicians is low compared to patients. That is why it is essential to develop a medical diagnosis system to identify heart disease that is built on machine learning. It gives more precise results than traditional ways and reduces cost [4]. Researchers are continually doing their hardest to identify more reliable smart models for successful treatment of this disorder, and a variety of smart models based on professional ML methods have been established with the passing of time. To make the

ISIC'21: International Semantic Intelligence Conference, February, 25–27, 2021, New Delhi, India
hafsabintekibria@gmail.com (H. B. Kibria); ammuaj.cseruet@gmail.com (A. Matin); mousanzida7@gmail.com (S. Islam)
https:https://www.ece.ruet.ac.bd/AbdulMatin (A. Matin)

medical diagnosis easy and cheap, we choose to build a fusion model to classify disease effectively. Here, a decision level fusion approach has been presented for heart disease disorder. Decision level fusion has implemented using the summation of scores of separate models[5]. The decision score of the separate models has been taken from the trained models. As the fusion occurs at the decision, so it is referred to as decision level fusion. The supervised algorithms- decision tree and logistic regression are fused with a deep neural network separately to develop a better fusion model that can tell the existence and non-existence of heart disease by analyzing the data. This work contributes by making a fusion model besides different machine learning algorithms to achieve better accuracy than them. And in our work, there was a notable improvement in the accuracy after fusion.

This paper is presented as follows. The concern for the issue of heart disease is discussed in the first segment. The second segment contains the previous work. Materials and methodology are presented in the third segment. In part four, the paper explains the experimental outcomes of our suggested architecture for identifying heart disease and also a comparative analysis with previous work. After that, we have ended with a conclusion and prospects.

2. Related Study

A lot of work has been performed on the classification of heart disease. The goal of all researchers was to improve accuracy. Various algorithms have been applied to the dataset to observe which model performs best. Now the researchers are trying to build a fusion model for effectively classifying disease efficiently. Different algorithms are now making using fuzzy logic or merging algorithms to create a new one. Some of the relevant works have been discussed here, along with their accuracy.

In [6] for different types of disease prediction, four different machine learning approaches were analyzed: LSTM, which is a kind of the recurrent neural network, XGBoost (XGB), random forest (RF), and Logistic regression (LR). Here, XGBoost is performing better than LSTM. In another study by [7], a fuzzy k-NN classifier was used, and the performance of the fuzzy k-NN classifier was way better than k-NN classifier. To remove the data's uncertainty, Fuzzy k-NN was used, and it provides higher accuracy than the k-NN classifier.

In a comparative study [8], six machine learning algorithms were applied to Cleveland dataset using six data mining tools, and then the result of those algorithms was compared to each other. Based on the data examination and the effects of the output data taken, the best output was given by ANN which was implemented in Matlab, and the best performing tool was Matlab.

Following that, a study by [9] several machine learning algorithms were used for medical diagnosis. These methods have been validated by tenfold cross-validation. Logistic regression gives an accuracy of 85% that was highest with an F1 score of 79%. And ANN with an accuracy of 84% with an F1 score of 84%. Among other things, those two algorithms worked better. In [10], Indian Pima diabetics dataset was used for classification. It was predicted if a person has diabetes or not. The prediction was made on test data with three-fold cross-validation. In this paper, Gradient Boosting offers a maximum accuracy of 86%, which is higher than Naive Bayes and Logistic Regression. Both Naive Bayes and Logistic Regression showed an accuracy of 77% and 79%, respectively.

In this study, [4] heart disease was predicted, and a multilayer perceptron was implemented in the suggested architecture. The architecture of the neural network contains 13 features derived from the Cleveland dataset. The proposed system gives 95% accuracy, and different accuracy has been found with the variation of the number of hidden layers. The improvement in our work compared those that we have used the fusion model to identify the disorder, which results in improved outcome.

In [11] researchers focused on creating a hybrid model to diagnose lung disease from X-ray images. CNN, Vgg16, and other techniques have been applied to predict lung disease. As CNN gave a poor performance, that's why a hybrid model was implemented for better results. The hybrid model VDSNet outperforms current methods in terms of the evaluation metrics.

3. Materials and Methodology

3.1. Data Collection

The dataset for heart disease was taken from the UCI ML repository [12]. Cleveland dataset was used, and the target of this study is to classify the patients with and without heart disease efficiently. This depository was developed in 1987 and currently holds 507 datasets. It has 13 features and 303 instances with some missing values. The dataset we used has been described in detail in table 1[12].

3.2. Data Preprocessing

The possibility is pretty good that data collected from any archive may have incomplete values or could even contain outliers. The probability of missing data values increases for the medical dataset [10]. In the Cleveland dataset, some values are also missing. For example, the values of major vessels and thalassemia are missing. There are numerous ways to handle missing data. Here for imputation, the strategy of most frequent was used.

Table 1
Descriptions of Features

Attribute	Description	Values
Age	Age in years	Continuous
Sex	Male/ Female	1=male,0=female
ChestPainType	Chest pain class	1 = typical type 1 2 = typical type angina 3 = non-angina pain 4= asymptomatic
RestBloodPressure	Resting blood pressure	Continuous value in mm hg
SerumCholestoral	Serum cholesterol	Continuous value in mm/dl
FastingBloodSugar	Resting electrographic results	0 = normal 1 = having_ST_T wave abnormal 2= left ventricular hypertrophy
ResElectrocardio-graphic	Fasting blood sugar	1 $\geqslant$ 120 mg/dl 0 $\leqslant$ 120 mg/dl
MaxHeartRate	Highest rate of heart	Continuous value
ExerciseInduced	Exercise induced angina	0= no, 1 = yes
Oldpeak	ST depression induced by exercise associated with rest	Continuous value
Slp	Peak exercise's slope	1 = un sloping 2 = flat 3= down sloping
MajorVessels	Number of major vessels colored by floursopy	0-3 value
Thal	Defect type	3 = normal 6 = fixed 7= reversible defect

The dataset contains some categorical and numerical values. Eight of the 14 attributes are categorical, and six are numerical. These eight categorical values were converted into numeric data, and for that label encoding was used, where based on alphabetical ordering, a unique integer is assigned for each label. It ensures that machine learning does not assume that higher numbers are more significant than lower numbers.

3.3. Data Partitioning

After the preprocessing and cleaning of the data, data were cross-validated. Here the five & ten-fold cross-validation was used, where the dataset was divided into five & ten equal parts. Four portions were used as training, and the left portion was used as validation at five-fold.

3.4. Feature Extraction

Feature extraction is the process of reducing the number of attributes in a dataset. Irrelevant features that tend to increase the model's accuracy are reduced. Here the top ten attributes have been displayed in the figure 1, contributing the most to the target output.

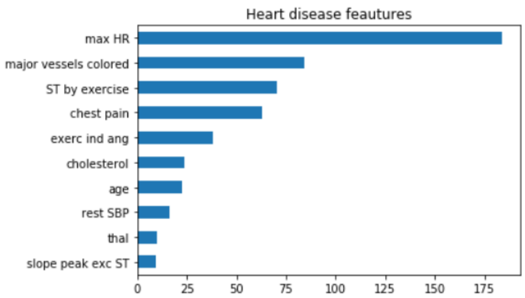


Figure 1: Top 10 Features of Heart Disease Dataset According to Chi-Square Score

3.5. Proposed Approach

This study's main objective is to form a fusion model with improving accuracy to classify cardiovascular disease's presence and absence. By merging the decision of different algorithms using summation, a new fusion model has been developed with high accuracy. The Jupyter Notebook, Scikit-learn, Tensorflow, and Keras have been used as implementation tools.

In figure 2, the proposed approach contains three main

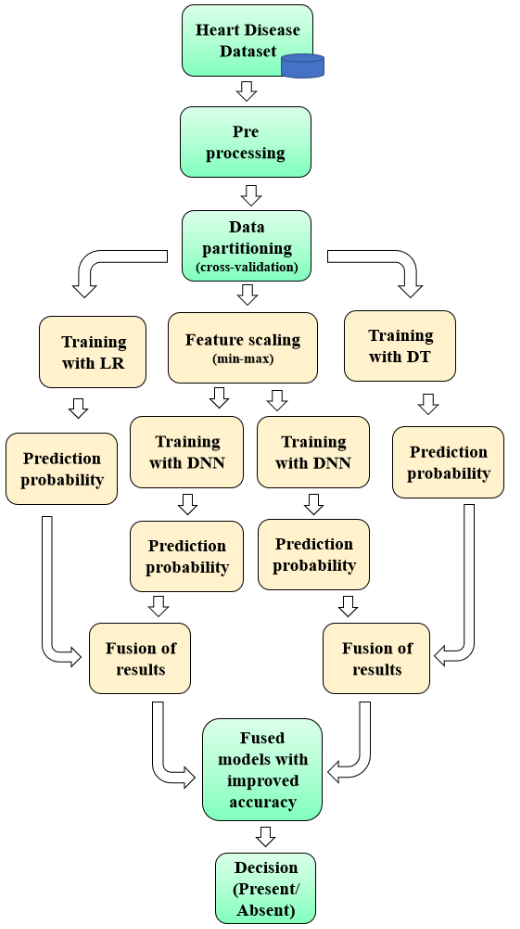


Figure 2: Proposed Architecture

segments; first, data were pre-processed and partitioned, then the accuracy was obtained using an individual classifying algorithm, and the following was model fusing. Three data mining algorithms DNN, DT, and LR, were applied to the heart dataset to evaluate efficiency and accuracy. After pre-processing, data was normalized only for the deep neural network to make all the features on the same scale. Min-max normalization was used for feature scaling. The equation for min-max is given in 1.

$$t' = \frac{t - min(t)}{max(t) - min(t)} \tag{1}$$

Where t is an original value, t' is the normalized value.

Data were not normalized for the other two algorithms, decision tree and logistic regression, as there is no impact on it. The same ordered dataset was used for deep neural network and decision tree algorithms. After training, those two algorithms' output probability was merged to get a decision level fusion model's output. After normalization, the data were trained with the three algorithms. After training, we took two classifiers to make a new fusion model. The fusion was done by a simple summation of the probability score from two algorithms. If p is the number of algorithms that we used in decision level fusion, and S_k is the decision score from the individual algorithms after training. In that case, the final decision score for the fusion model can be represented by S_m for each fold using the following equation.

$$S_m = \sum_{k=1}^{p} S_k / p \tag{2}$$

Algorithm 1: Algorithm for decision level fusion

Input : Float value $S[i][k]$, two int n and p. $S[i][k]$ is the decision score of a single algorithm of each fold, n is the fold number in cross-validation and p is the number of algorithms used in fusion

Output : S_f, the final average decision score for fusion model

St=0;
for $i \leftarrow 0$ **to** n **do**
  $S_m = 0$
  for $k \leftarrow 0$ **to** p **do**
    $S_m = S[i][k] + S_m$
  end
  $S_m = S_m/p$; `// fusion model's decision score of each fold`
  $S_t = S_m + S_t$; `// summation of fusion model's decision score of all folds`
end
$S_f = S_t/n$

A generalized algorithm-1 is displayed here. The sum of the decision score (S_m) needs to be divided by the number of algorithms (p) in fusion. As we took two algorithms for the fusion of model-1 and model-2, so we needed to divide the sum by 2. Using the final score, the fusion model predicted the output for test data. As both algorithms' impact is present in our new fusion score, it decreases the rate for miss-classification. This process continues five times in five-fold and ten times in ten-fold cross-validation. After that, we took the average value from all folds and noticed an improvement in the average score comparing to the individual algorithm's scores. To illustrate the procedure clearly, all fold values have been displayed in tables 2 and 4.

The fusion model's output is based on two algorithms (DNN+DT) & (DNN+LR). We took the DNN and DT decision scores, and by summing them together, we got the fusion model's decision score. This summation was done at every stage of the folds in cross-validation. At last, the performance of two individual models and fusion models was compared. This process is the same for both models for the same dataset with a different order. This new, improved fused model was used to diagnose heart disease.

3.6. Decision Tree (DT)

A decision tree is a tree-like structure consisting of branches, nodes, and leaf nodes. It is a branching graph that functions like a splitting rule for each particular attribute. Each feature is treated as a branching node. These nodes build a rule, and, based on the rule, values are grouped into different classes. In the decision tree, the leaf decides some decisions at ending, and the topmost is the root, which partitions the tree based on attribute value. Building A DT is easy and simple, and the results are predicted more accurately.

DT is sensitive to overfitting. It happens when the model is very good at identifying trained data but gives poor performance for test data. It becomes over skilled for training data by having minimal impurity in the leaf node. That's why pre-pruning is needed to minimize the number of leaf nodes that are not that important for model building. It gives better accuracy for prediction. Information gain is another important criterion, and attributes with the highest information gain split first. For that method, features with the lowest entropy are selected for splitting [13]. In the proposed design gini was used as criterion, maximum dept of the tree was 8, and minimal leaf size was 10. Pre-pruning was done to get the best parameters. The Gini index and entropy are split criteria for a decision tree, and they are arranged by applying (3) and (4).

$$Gini\ :\ Gini(E) = 1 - f(x) = 1 - \int_{j=1}^{n} Pj^2 \tag{3}$$

$$Entropy\ :\ H(E) = 1 - f(x) = -\int_{j=1}^{n} Pj \log_2 Pj \tag{4}$$

Here Pj is the percentage of a class in a node.

3.7. Deep Neural Network (DNN)

An ANN has more than one layer, then is named a deep neural network (DNN). Like humans, a deep learning (DL) system can teach itself and learn through several hidden layers. DL is a model that is based on a multilayer feed-forward perceptron and uses back-propagation to train with stochastic gradient descent. There are four layers with nodes and neurons for the designed network, which has a uni-direction. It has two hidden layers, and there is a single way connection between every node and the next node. For training, stochastic gradient descent was used with back-propagation. It has been suggested as a very useful tool in different medical sectors for decision making. Figure 3 displays the structure of our proposed DNN.

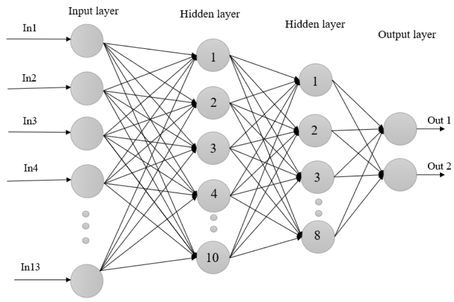


Figure 3: Proposed Architecture of Deep Neural Network

In this proposed approach, an input layer was used for data entrance, an output layer for predicting output(presence/absence), and two hidden layers. One of the most challenging tasks is model optimization for machine learning implementation. Model optimization reduces the test error; however, deep learning tunes the parameters outside of the model but has a significant influence on results and classification. The mother of all hyperparameters is the learning rate. The speed of learning of the model depends on the learning rate. The number of hidden units is vital as it regulates the model's representational capacity. The number of first and second hidden layers in our model is 10 and 8, respectively. L2 regularization was used to block overfitting. Bias, which is a constant, was used to help the model in a way that can fit best for the given data. Batch size ten was used.

3.8. Logistic Regression (LR)

LR is a supervised machine learning technique with continuous/discrete predictors. LR is generally used for binary classification. It is a statistical model that represents the relation between the binary dependent variable's logit transformation and independent variables (one or more than one) by determining the best fitted linear model. This model is a simple prediction approach compared to other non-parametric models of machine learning with baseline accuracy scores provided by the model [14].

3.9. Classification Performance Measurement

Heart disease has been classified using three ML models. After training the three models, two decision level fusion models have been developed using the trained models' decision score. Five and ten-fold cross-validation was used for output performance. Five quality measures have been measured for the classification of the model.

Here,

- True positive ($True_P$): Case and prediction both are positive.
- True negative ($True_n$): Case and prediction both are negative.
- False positive ($False_p$): Case negative but prediction is positive.
- False negative ($False_n$): Case positive but prediction is negative

The performance parameters are explained as below:

Classification accuracy: Accuracy is the ratio of predictions that model got right.

$$Accuracy = \frac{True_P + True_n}{True_p + True_n + False_p + False_n} \tag{5}$$

Precision or positive predictive value: Precision is the fraction of the real positive (absence labeled as absent) of all cases classified as positive (total cases labeled as absence) . Recall calculates the amount of the real positive that are accurately classified.

$$Precision = \frac{True_P}{True_P + False_p} \tag{6}$$

$$Recall = \frac{True_P}{True_P + False_n} \tag{7}$$

F1 score measures the weighted score of precision and recall. The value of one indicates the greatest performance of f1-score, where zero indicates the worst.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{8}$$

ROC is a plot between the rate of true and false positive where the performance and quality of diagnosis heart disease are shown. The values between 0 and 1 are taken for the ROC curve, and the classifier is considered an ideal classifier, which takes a value of 1.

4. Performance Evaluation

Two decision level fusion models were constructed by applying an artificial neural network, logistic regression, and decision tree. Three algorithms were applied individually on the same dataset, and then the decision scores of the DNN and DT algorithm were combined as well as the scores of DNN and LR.

The fusion models (DNN+DT) and (DNN+LR) gave a better result than individual achievement. Tables 2 and 3 have shown the performance parameters of model-1, and table 4,5 has displayed the outcome of model-2, using five and ten-fold cross-validation, respectively. The performance of the fusion model has increased.

In table 2, various performance parameters were measured. To observe the improvement of the parameters, the value of each fold has been shown. We can see that each fold's parameters improved than the separate model,

Table 2
The Performance of Deep Neural Network and Decision Tree (Model-1) with Five-Fold Cross-Validation

		ACC	Prec	RCL	F1-score	AUC score
Fold-1	DNN	81.97	82	82	82	81.70
	DT	81.97	84	81	81	80.90
	DNN+DT	**83.61**	**84**	**83**	**83**	**84.29**
Fold-2	DNN	85.24	88	84	85	84.20
	DT	85.24	86	85	85	84.74
	DNN+DT	**86.89**	**88**	**86**	**87**	**87.77**
Fold-3	DNN	80.33	80	80	80	80.47
	DT	80.33	80	80	80	79.92
	DNN+DT	**83.61**	**83**	**83**	**83**	**83.50**
Fold-4	DNN	88.33	88	88	88	88.38
	DT	86.67	87	86	86	86.20
	DNN+DT	**86.67**	**87**	**87**	**87**	**86.53**
Fold-5	DNN	86.7	90	85	86	85.19
	DT	81.7	83	81	81	80.64
	DNN+DT	**85**	**88**	**84**	**84**	**87.5**
Average	DNN	84.51	85.6	83.8	84.2	83.99
	DT	83.17	84	82.6	82.6	82.48
	DNN+DT	**85.15**	**86**	**84.6**	**84.8**	**85.92**

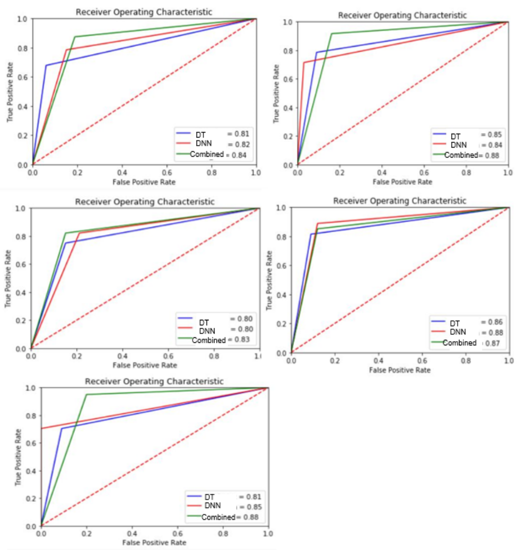


Figure 4: ROC Characteristics of Five Folds Respectively of Model-1

which helped increase the overall performance of the fused model. Though the performance of fold-5 slightly decreased, overall, we found improved accuracy in the merged model.

Figure 4 represents the ROC curve of model-1 of every fold. Only the average value of ten-fold cross-validation has been displayed in table 3. It gave 84.21% accuracy for fused model-1, which improved with respect to DNN and DT.

Table 3
Average Performance of Ten-Fold Cross-Validation of(Model-1)

	ACC	Prec	RCL	F1-score	AUC-score
DNN	83.34	83.7	82.9	82.8	80.24
DT	81.12	82.7	82.5	82.4	82.40
DNN+DT	**84.21**	**84.3**	**83.8**	**83.88**	**84.62**

Table 4 shows that the performance did not improve for all the folds among the five-folds. For model-2, the performance of two folds has neither increased nor decreased. But there is a notable improvement in all other folds in the model, which makes the average score better and allows the model accurate at classifying.

Table 4
The Performance of Deep Neural Network and Logistic Regression (Model-2) with Five-fold Cross-Validation

		ACC	Prec	RCL	F1-score	AUC score
	DNN	77.05	77	76	77	76.35
Fold-1	LR	78.69	79	78	78	78.14
	DNN+LR	**80.33**	**81**	**80**	**80**	**80.86**
	DNN	83.61	84	83	83	82.95
Fold-2	LR	83.61	84	83	83	83.23
	DNN+LR	**85.25**	**86**	**85**	**85**	**85.67**
	DNN	83.61	87	82	83	82.41
Fold-3	LR	83.61	85	83	83	82.68
	DNN+LR	**83.61**	**87**	**82**	**83**	**86.52**
	DNN	91.67	92	91	92	91.41
Fold-4	LR	86.67	87	87	87	86.53
	DNN+LR	**88.33**	**88**	**88**	**88**	**88.17**
	DNN	85	86	84	85	84.38
Fold-5	LR	90	90	90	90	89.73
	DNN+LR	**90**	**90**	**90**	**90**	**90.27**
	DNN	84.19	85.2	83.2	84	83.50
Average	LR	84.51	85	84.2	84.2	84.06
	DNN+LR	**85.50**	**86.4**	**85**	**85.2**	**86.30**

Figure 5 displays the ROC characteristics of every fold of the model-2. The roc curve of DNN, LR, and (DNN+LR) has been shown in every diagram.

Table 5 shows that the average performance of 10-fold cross-validation is shown, and it gave a good accuracy of 87%, whereas DNN and LR individually gave 86.10% and 86.15%, respectively. Also, it gave a better performance than model-1.

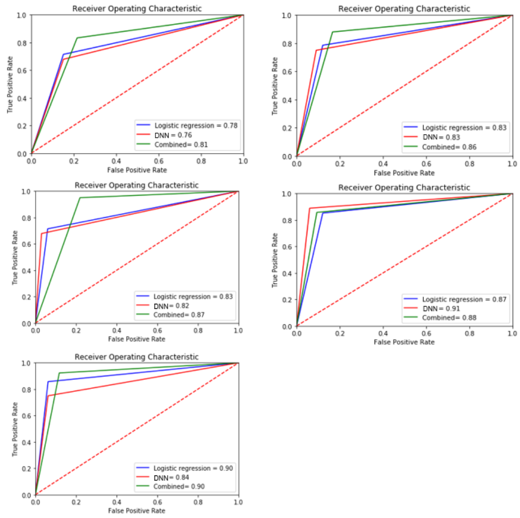
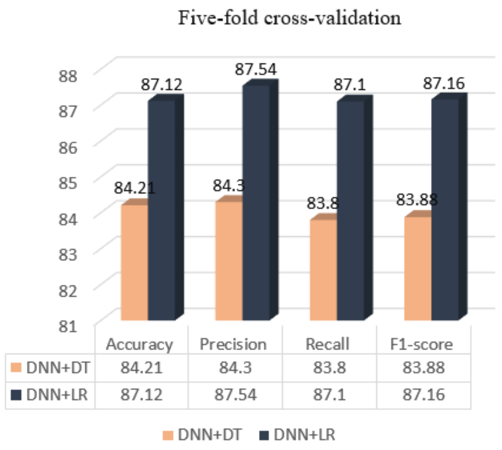


Figure 5: ROC Characteristics of Five Folds Respectively of Model-2

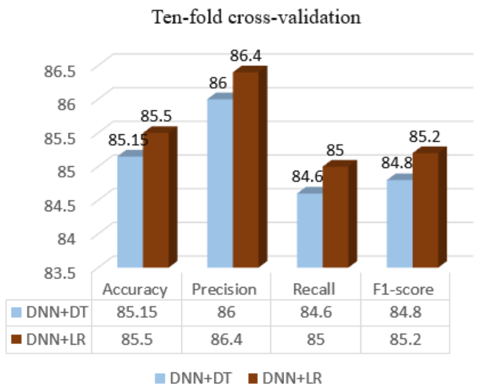
Table 5
Average Performance of Ten-Fold Cross-Validation of (Model-2)

	ACC	Prec	RCL	F1-score	AUC-score
DNN	86.10	86.8	86	86.02	85.59
LR	86.15	86.7	86.14	86.34	85.62
DNN+LR	**87.12**	**87.54**	**87.1**	**87.16**	**87.54**

We can see from all the values of tables 2, 3, 4 , and 5 that all the parameters slightly improved after the algorithms' fusion in some folds. That is why the hybrid model's average performance is higher than the algorithms' average individual performance. Finally, it was found that model-2 (DNN+LR) had the best result for 10-fold cross-validation from all fusion models. And also, for 5-fold, model-2 performed better than model-1. And after fusion, improvement happened to all models. And using these decision level fusion models, decision making will be more efficient. As the decision probability of both algorithms was fused, so the chance of more accurate prediction increases. If one algorithm has a low probability for specific testing data (having heart disease). In that case, there is a chance that other algorithms will have a higher probability, which will help the model to predict correctly. After merging the output probabilities, the model can easily predict the person with heart disease. So, those data that have a chance for miss-classification may get correct classification after fusion, and thus a better model will be constructed.



(a) Five-fold Cross Validation



(b) Ten-Fold Cross Validation

Figure 6: Comparison of Performance Parameters

Figure 6 represents the graphical chart for model-1 and model-2 for both five and ten-fold cross-validations along with the value of the performance parameters.

4.1. Comparison with Previous Work

This section discusses the work that has already been done with the same dataset as ours. In table 6, the output result of those works has been displayed with their approach. In every work, the researchers used various algorithms to classify disease. We have only shown the highest accuracy and the name of the algorithm to obtain that accuracy. Table 6 shows that the accuracy we found for our fusion model is a little more than their models. The outcome of both of our models using the same configuration has also be shown here. Some researchers used both five and ten-fold to calculate accuracy. Both results have been shown in the table to compare correctly with our work.

As we used both five and ten-fold cross-validations, our accuracy was 85% for model-1 and 85.5% for model-2 using five-fold cross-validation. For five-fold cross-validation, the highest accuracy from table 6 is 83.83%. So both of our fusion models performed well. The highest accuracy for ten-fold cross-validation was 84.15%, which was also obtained by a fusion model of MLP and SVM. We got 84.21% and 87.12% for model-1 and 2, respectively. So also, in ten-fold, our fused model gave better accuracy for decision making. In paper[10], they used three-fold cross-validation, so to compare our work with them, we have also calculated the accuracy and got 85.20% and 86.13% from model-1 and model-2, respectively, which are higher than them.

5. Conclusion

At present, machine learning is a must, particularly in the health sector. This research aims to predict patients suffering from heart disease with greater accuracy using machine learning algorithms. Two fusion models have been developed to classify the presence and absence of heart disease. A comparative comparison was made to see which model performs better. Here, three algorithms have been used, and then two new decision level fusion models were created by combining the three algorithms. Both fused models gave a better performance than separate algorithms, and between the two fused models, model-2, which is the combination of DNN and LR, performed better than model-1(DNN+DT). Model-2 has an average accuracy of about 86%, where model-1 provided 85% accuracy with five-fold cross-validation. And for ten-fold, model-1 gave an accuracy of 84%, and model-2 gave 87.12%. In terms of other parameters, both models' parameters improved. Both models' performance is almost the same for five-fold cross-validation, and in ten-fold, model-2 performed much better than model-1. And in both models, we obtained improved accuracy after fusion, which was our main goal. The classifier can be used for different datasets for other medical diagnoses in the future, and the fusion model can be built with more than two algorithms.

References

[1] H. Ayatollahi, L. Gholamhosseini, M. Salehi, Predicting coronary artery disease: a comparison between two data mining algorithms, BMC public health 19 (2019) 1–9.

Table 6
Previous Work with Same Dataset

Reseacrchers	Year	Approach	Fold number	Accuracy	Our accuracy Model-1	Model-2
I Tougui et al. [8]	2020	LR	10	83.50%	84.62%	87.54%
R Birjais et al. [10]	2019	LR	3	79.20%	85.20%	86.13%
S. Pouriyeh et al.[15]	2017	SVM +MLP	10	84.15%	84.62%	87.54%
Y K Singh. et al [16]	2016	LR	5	83.83%	85.92%	86.30%
			10	83.17%	84.62%	87.54%
		DT	5	79.54%	85.92%	86.30%
			10	79.21%	84.62%	87.54%
S U Ghumbre et al. [17]	2012	MLP	5	82.24%	85.92%	86.30%

[2] S. Wang, R. M. Summers, Machine learning and radiology, Medical image analysis 16 (2012) 933–951.

[3] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, D. I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Computational and structural biotechnology journal 13 (2015) 8–17.

[4] T. Karayılan, Ö. Kılıç, Prediction of heart disease using neural network, in: 2017 International Conference on Computer Science and Engineering (UBMK), IEEE, 2017, pp. 719–723.

[5] A. Matin, F. Mahmud, T. Ahmed, M. S. Ejaz, Weighted score level fusion of iris and face to identify an individual, in: 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, 2017, pp. 1–4.

[6] T. Christensen, A. Frandsen, S. Glazier, J. Humpherys, D. Kartchner, Machine learning methods for disease prediction with claims data, in: 2018 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, 2018, pp. 467–4674.

[7] V. Krishnaiah, G. Narsimha, N. S. Chandra, Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review, International Journal of Computer Applications 136 (2016) 43–51.

[8] I. Tougui, A. Jilbab, J. El Mhamdi, Heart disease classification using data mining tools and machine learning techniques, Health and Technology (2020).

[9] A. K. Dwivedi, Performance evaluation of different machine learning techniques for prediction of heart disease, Neural Computing and Applications 29 (2018) 685–693.

[10] R. Birjais, A. K. Mourya, R. Chauhan, H. Kaur, Prediction and diagnosis of future diabetes risk: a machine learning approach, SN Applied Sciences 1 (2019) 1112.

[11] S. Bharati, P. Podder, M. R. H. Mondal, Hybrid deep learning for detecting lung diseases from x-ray images, Informatics in Medicine Unlocked 20 (2020) 100391.

[12] D. Dua, C. Graff, UCI machine learning repository, 2017. URL: http://archive.ics.uci.edu/ml.

[13] H. Naz, S. Ahuja, Deep learning approach for diabetes prediction using pima indian dataset, Journal of Diabetes & Metabolic Disorders 19 (2020) 391–403.

[14] D. R. Cox, The regression analysis of binary sequences, Journal of the Royal Statistical Society: Series B (Methodological) 20 (1958) 215–232.

[15] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia, J. Gutierrez, A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease, in: 2017 IEEE Symposium on Computers and Communications (ISCC), IEEE, 2017, pp. 204–207.

[16] Y. K. Singh, N. Sinha, S. K. Singh, Heart disease prediction system using random forest, in: International Conference on Advances in Computing and Data Sciences, Springer, 2016, pp. 613–623.

[17] S. U. Ghumbre, A. A. Ghatol, Heart disease diagnosis using machine learning algorithm, in: Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012) held in Visakhapatnam, India, January 2012, Springer, 2012, pp. 217–225.