2021 Asian Conference on Innovation in Technology (ASIANCON)
Pune, India. Aug 28-29, 2021

A Novel Approach for Prediction of Heart Disease using Machine Learning Algorithms

Akanksha Kumari
Dept. of Computer Applications
National Institute of Technology Jamshedpur
Jamshedpur 831014, India
akankshakumariee@gmail.com

Ashok Kumar Mehta
Dept. of Computer Applications
National Institute of Technology, Jamshedpur
Jamshedpur 831014, India
akmehta.ca@nitjsr.ac.in

***Abstract*— Heart Disease is the leading causes of death and hospitalization in the world. With advancement of technology and the contribution of computer engineering, it is easy to detect heart disease and thus treatment is fast and effectively done. Machine learning nowadays is very popular in predicting disease in the medical field. In this paper, authors have tried to predict heart disease using seven machine learning algorithms and attempted to improve the accuracy of weak performing algorithms using ensemble methods like AdaBoost and voting ensemble method. The performance of Linear Discriminate Analysis is good among other algorithms, it's mean value is approximately 0.847 and mean absolute error is 0.185, the false acceptance rate is lowest among all i.e.; 0.33 and the false recognition rate is 0.076, accuracy is somehow coming 80% which is less if compared with Logistic Regression.**

Keywords*— *Confusion Matrix; Ensemble Method; Heart Disease Prediction; Machine Learning;

I. INTRODUCTION

In heart disease, usually, the heart is unable to pump the required amount of blood to other major organs and restrict the functionalities of the body. Due to this, ultimately the heart stops working or failed [1].

ML is one of the most stimulating technologies that have ever invented. As it is obvious from the name, it makes the computer more similar to humans. **Arthur Samuel**, a vanguard in the field of artificial intelligence and computer gaming, composed the term **"Machine Learning"**. He defined machine learning as – **"Field of study that gives computers the capability to learn without being explicitly programmed".**

Machine learning is the brain where all the learning takes place. Just like human learn from their past experiences and mistakes, machine learns. So there is some set of training data. Machines are trained using these dataset and then they create model which takes inputs and make predictions. To make an accurate prediction, the use of statistical methods, algorithms are trained to make classifications or predictions.

II. PROBLEM STATEMENT

The prediction of heart disease helps specialists make more accurate decisions regarding patients' health. Therefore, the use of machine learning (ML) is a key to figure out the symptoms related to heart disease.

Heart disease can be handled persuasively with a consolidation of lifestyle changes, medicine and, in some cases, medical surgery. The predicted results can be used to prevent and thus diminish the cost of surgical treatment and other expenses [2].

Many of scholars have used machine learning algorithms and ensemble methods. Some of them have made a mobile app and related system to predict heart disease. The overall objective of this paper will be to compare the evaluation and performance metrics of some machine learning algorithms as well as improving the performance of weak algorithms using the ensemble method.

III. HEART DISEASE PREDICTION SYSTEM

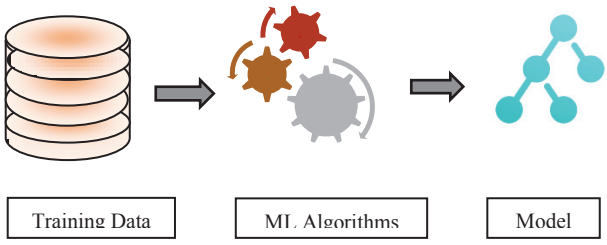

Fig. 1. Heart Disease Prediction Process

Heart disease prediction should be based on the doctor's institution and experience but this practice is to help in clinical decision making and cost reduction [2].

Fig1 tells what the process behind the prediction of disease is and how machine learning works. Our first task is to collect a dataset of heart disease [11] and in this paper, the used dataset is taken from *Kaggle*. After collecting the dataset and visualizing of data it needs to prepare it for machine learning. This is called *data preprocessing* [3] which involves rescaling data, standardizing data, normalizing data. It is done with the help of scikit-learn libraries.

Then Feature selection is done. Irrelevant features can negatively impact performance [8].

The following features have been extracted in a large set of features of heart disease (see table 1) [5].

A. Evaluation of Algorithms

The authors have evaluated and compared seven types of machine learning algorithms in this paper, these are as follow [12]:

- LR-Logistic Regression
- LDA-Linear Discriminate Analysis

2021 Asian Conference on Innovation in Technology (ASIANCON) | 978-1-7281-8402-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ASIANCON51346.2021.9544544

- KNN-K Neighbors Classifier
- CART-Decision Tree Classifier
- GNB-Gaussian Naïve Bayes
- SVM-Support Vector Machine
- RF-Random Forest classifier

TABLE I. CLINICAL FEATURES & THEIR DESCRIPTIONS[7]

Clinical features	Description
Age	In years
Sex	Gender-male or female
Cp	Constrictive pericarditis (a type of chest pain)
Trestbps (mmHg)	Resting blood pressure(on admission to hospital)
Chol (mg/dl)	Serum cholesterol measurement
Fbs	Blood sugar on fasting
Restecg	Resting electrocardiographic results
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina
Oldpeak	ST depression induced by exercise relative to rest
Slope	The slope of the peak exercise ST segment
Ca	Number of major vessels (0–3)
Thal	Blood disorder called thalassemia

Some algorithms like Decision tree classifier and Gaussian Bayes have weak mean value so to improve their performance two types of ensemble methods have been used 1)AdaBoost and 2)Voting method. Voting method is also applied in 3 different ways.

B. *Performance Metrics*

Performance of algorithms can be observed by calculating mean, mean absolute error, FAR, FRR, accuracy, precision etc. Fig 2 shows the graph of the mean for all algorithms. Fig 3 shows the mean absolute error.

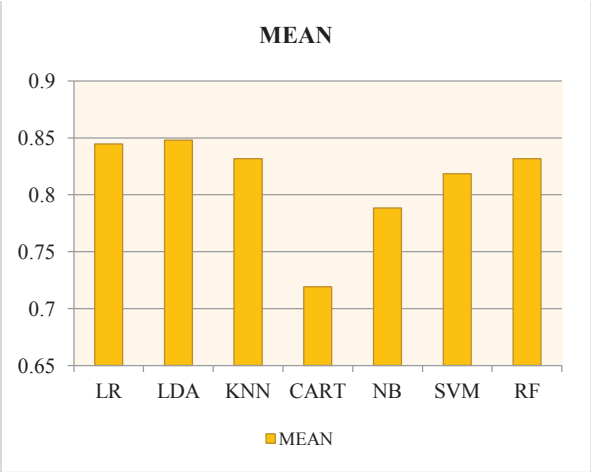

Fig. 2. Graph for the mean of all algorithms

- Mean Absolute Error: It is the difference between the true value and the predicted value. The mean absolute error of a model with respect to a test set is the mean of the absolute values of the individual prediction errors on all instances in the test set.

*Mean Absolute Error (*MEA) is used to evaluate the performance with regression models [10].

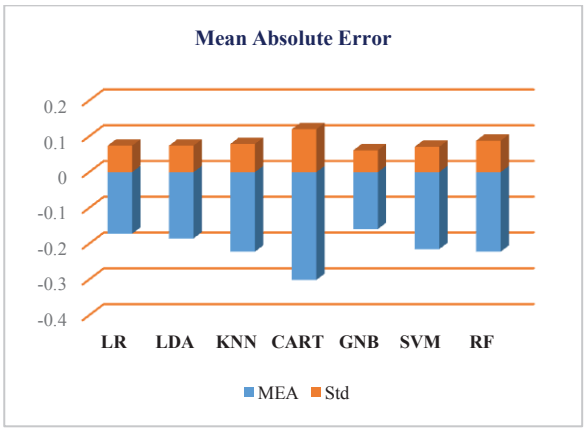

Fig. 3. mean absolute error for all algorithms

From fig 3 Gaussian Naïve Bayes has the lowest mean error (ignoring negative sign due to absolute value) and also a standard deviation from mean error is small among all whereas Decision tree classifier has large (poor) mean error.

- *Confusion Matrix:* A confusion matrix is used to evaluate the performance of the classification model. It is n*n matrix where n is the number of target classes. It is a table with 4 different combinations of predicted and actual values.

  The confusion matrix is for discrete data which means for classifiers and not for continuous data (regression). But here authors have applied this method to all seven algorithms including logistic regression. And here the outcome is.

```
from sklearn.metrics import confusion_matrix
results = []
variable = []
for variable, model in score:
    model.fit(X_train,y_train)
    y_predicted=model.predict(X_test)
    cm=confusion_matrix(y_test,y_predicted)
    results.append(cm)
    print(model,'\n',cm)
```

```
LogisticRegression()
 [[48  0]
 [ 0 52]]
LinearDiscriminantAnalysis()
 [[32 16]
 [ 4 48]]
KNeighborsClassifier()
 [[22 26]
 [13 39]]
DecisionTreeClassifier()
 [[48  0]
 [ 0 52]]
GaussianNB()
 [[48  0]
 [ 0 52]]
SVC()
 [[17 31]
 [ 4 48]]
RandomForestClassifier()
 [[48  0]
 [ 0 52]]
```

After calculating the confusion matrix it is shown as in fig 4[6].

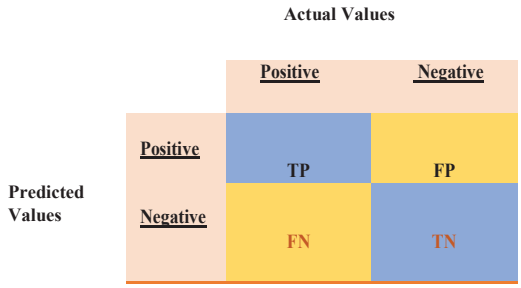

Fig. 4. confusion matrix

True Positive: This means Sick people correctly predicted as sick by the model. It is predicted that a person is suffering from heart disease and she/he is.

True Negative: This means Sick people incorrectly predicted as not sick by the model. It is predicted that a man is not suffering from heart disease t and he is not.

False Positive: (Type 1 Error) Healthy people incorrectly predicted as sick by the model. It is predicted that a man is suffering from heart disease but he is not.

False Negative: (Type 2 Error) Healthy people correctly predicted as not sick by the model. It is predicted that a man is not suffering from heart disease but he is.

Here TRUE/FALSE implies was prediction right

Positive /Negative implies what was predicted

TP, TN, FP & FN has been calculated manually as following:

$$TN = CM\,[0]\,[0]$$

$$FN = CM\,[1]\,[0]$$

$$TP = CM\,[1]\,[1]$$

$$FP = CM\,[0]\,[1]$$

Where CM stands for confusion matrix

TABLE II. FAR, FRR VALUE FOR ALL ALGORITHMS[15]

Classifier	TN	FN	TP	FP	FAR	FRR
LDA	32	4	48	16	0.333	0.07692
KNN	22	13	39	26	0.54167	0.25
CART	48	0	52	0	NAN	NAN
SVC	17	4	48	31	0.6458	0.05063

Since LR, CART, GNB & RF have the same confusion matrix (This is because of the low dataset) so in the table, it's calculated performance metrics for CART only (see table III).

- *False Acceptance Rate*:
  FAR stands for False Acceptance Rate which means the probability that a system accepts and incorrectly predict that person has not heart disease.
  FAR = imposter scores or fraction exceeding threshold/all imposter scores.
  Imposter scores exceeding threshold = FP
  All imposter scores = FP+TN

$$\mathbf{FAR = FPR = \frac{FP}{FP+TN}}$$

- *False Recognition Rate:*
  FRR stands for False Rejection Rate which is the probability of a system to rejects and incorrectly predicts that a person is suffering from heart disease.
  FRR = genuine scores exceeding threshold/all genuine scores
  Genuine scores exceeding threshold = FN
  All genuine scores = TP+FN

$$\mathbf{FRR = FNR = \frac{FN}{TP+FN}}$$

- *Accuracy*
  Due to low dataset accuracy for LR, CART, SVC and RF are generated to be 100% (fig 5), but there are other parameters like mean, mean absolute error, precision, recall, f1_score which shows the performance of algorithms.

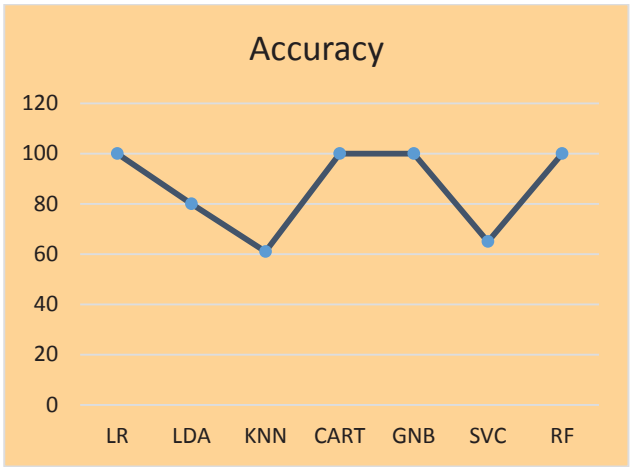

Fig. 5. Accuracy in % for all algorithms

- *Precision:* It tells how many of the correctly predicted cases turned out to be positive that means the used model is reliable or not [9].

$$\text{Precision} = \frac{TP}{TP+FP}$$

- Recall: It tells how many of the actual positive cases can be predicted correctly [9].

$$\text{Recall} = \frac{TP}{TP+FN}$$

- *F1 score:* It is a harmonic mean of Precision and Recall. It measures accuracy but since accuracy calculation is not advantageous for unevenly distributed data so there f_score is used [9].also f_score is useful if there is higher cost associated with particular type of error.

TABLE III. PRECISION, RECALL, F_ SCORE FOR POSITIVE (1) & NEGATIVE (0) OUTCOMES FOR ALL ALGORITHMS

		PRECISION	RECALL	F1_SCORE
LR	0	1	1	1
	1	1	1	1
LDA	0	0.89	0.67	0.76
	1	0.75	0.92	0.83
KNN	0	0.63	0.46	0.53
	1	0.60	0.75	0.67
CART	0	1	1	1
	1	1	1	1
GNB	0	1	1	1
	1	1	1	1
SVC	0	0.81	0.35	0.49
	1	0.61	0.92	0.73
RF	0	1	1	1
	1	1	1	1

$$\text{F1\_score} = \frac{2}{\left(\frac{1}{\text{Recall}}\right)+\left(\frac{1}{\text{Precision}}\right)}$$

- *Macro average:* It is the average of precision/recall/F1_score [9].

Table III shows the performance metric of algorithms for positive and negative outcomes.

C. *Ensemble Algorithms:*

This method combines several algorithms together to bring out better predictive results, as compared to using a single algorithm. In the Ensemble method variance and bias decreases and predictions improve in a developed model. In other words, it helps in avoiding overfitting.

- *Overfitting:* when a model train data too much, it learns irrelevant data as well as noise which affects learning negatively. That means noise and random fluctuations in the training data is elected and learned as concepts by the model which is not practiced to new data. So if training is continued for long time, the performance on the training dataset continues to decrease [8].
- *Underfitting*: opposite to overfitting when model learns too small that no more model the training data and not at all generalized new data.

- *Variance:* Variance express how much the estimated target function vary if different training data were used. In other words, variance express how much a random variable differs from its predictable value. Variance can lead to overfitting [13]

- *Bias:* bias is an error in which model's prediction conflict with the target value, compared to the training data. When there is a high bias error, it results in a very simplistic model that does not consider the variations very well. It leads to underfitting.

There are various ways to combine algorithms, here authors have used boosting method and the voting ensemble method.

AdaBoost Ensemble Algorithm:

Adaptive Boost is used for classification rather than regression [4]. Used to boost the accuracy of the Decision tree classifier. Initially, the mean was 0.752 and then after using the AdaBoost method overall mean is found 0.791

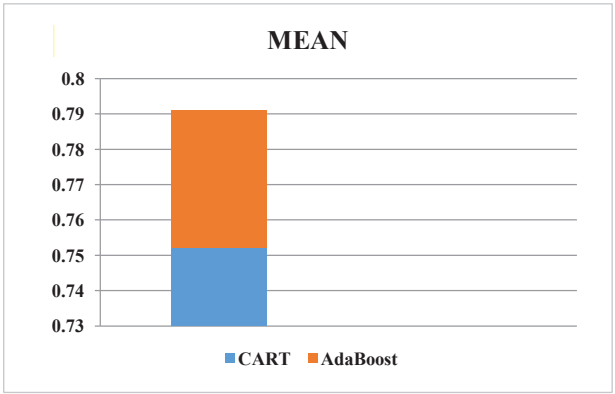

Fig. 6. mean value after applying AdaBoost algorithm

Voting Ensemble Algorithm:

It is one of the simplest ways to combining the prediction from multiple machine learning algorithms. Voting Ensemble does not require the base models to be consistent.

In other words, It can train diverse base learners, for example, a Decision Tree, Logistic Regression, & Support Vector Machines, and then use the Voting Ensemble to combine the results[14].

- Firstly authors have combined all algorithm(LR, LDA, KNN, CART, GNB, SVC, RF) and tried to enhance overall performance using voting ensemble algorithm, so the overall mean value was 0.81
- Later only three algorithms having better result among all i.e.; LR, LDA, KN have been chosen and then the voting ensemble algorithm is applied and found overall mean is 0.82
- Then authors have chosen the weakest performing classifiers NB & Decision tree classifier and the overall mean value became 0.84

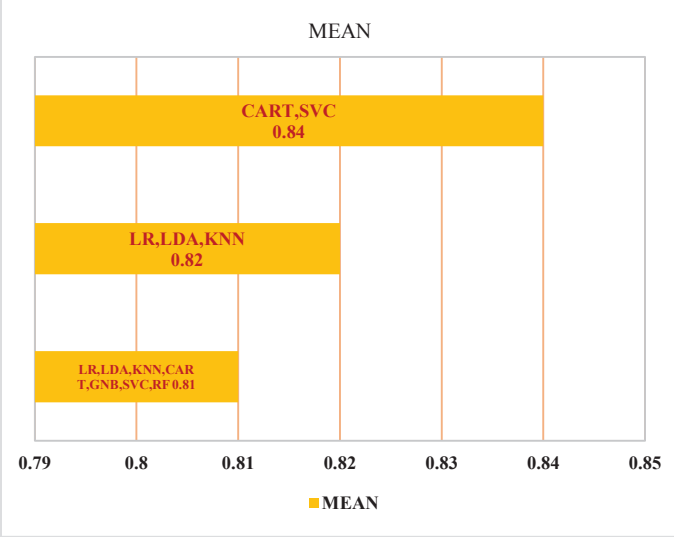

Fig. 7. mean values with & without voting ensemble algorithms

IV. CONCLUSION

The paper is all about predicting heart disease using machine learning algorithms. There are an immense diversity of algorithms and among them, if the individual analysis is to be considered then Logistic Regression, Linear discriminant analysis and KNN have better performance. When compared to Seven Algorithms i.e.; LR, LDA, CART, KNN, NB, SVM, RF the three algorithms LR, LDA and KNN gives a better mean value.

The Decision Tree classifier has a minimum value that can be improved using AdaBoost Algorithm.

The Voting Ensemble algorithm improves the mean value of the combination of Naïve Bayes and decision tree classifier. The Voting Ensemble technique has been used in three ways and from fig 7 it can be concluded that it is best suitable for enhancing the performance of weak algorithms.

It is found that LR, CART, GNB, RF algorithms gives 100% accuracy. Soon machine learning algorithms can be more efficient and useful in medical and healthcare for taking decision.

REFERENCES

[1] A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms Amin Ul Haq , 1 Jian Ping Li , 1 Muhammad Hammad Memon , 1 Shah Nazir , 2 and Ruinan Sun 1 1 School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China 2 Department of Computer Science, University of Swabi, Khyber Pakhtunkhwa, Pakistan

[2] KENNEDY NGURE NGARE BSC/LMR/5551/17 project proposal submitted for the study leading to a project report in partial fulfilment of the requirements for the award of a Bachelor of Science in Computer Science at St. Paul's University.

[3] Heart Disease Prediction Using Machine Learning Algorithms 1.Archana Singh Computer Science and Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur 2.Rakesh Kumar Computer Science and Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur

[4] P. Ghosh et al., "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques," in IEEE Access, vol. 9, pp. 19304-19326, 2021, doi: 10.1109/ACCESS.2021.3053759.

[5] Prediction of Heart Disease Using Machine Learning Algorithms Rajesh N1, T Maneesha2, Shaik Hafeez3, Hari Krishna4 1,2,3,4Computer Science Engineering, K L E F, Guntur, India

[6] Reddy, Viswanath & Sajjan, Guttappa. (2019). Design and Development of Real-Time Heart Disease Prediction System for Elderly People Using Machine Learning. 10.13140/RG.2.2.12199.50081.

[7] A. H. Chen, S. Y. Huang, P. S. Hong, C. H. Cheng and E. J. Lin, "HDPS: Heart disease prediction system," 2011 Computing in Cardiology, 2011, pp. 557-560.

[8] Jason Brownlee Master Machine Learning Algorithms Discover How They Work and Implement Them From Scratch.

[9] A. Bhandari article "Everything you Should Know about Confusion Matrix for Machine Learning".

[10] L. Mendo, "Estimation of a probability with guaranteed normalized mean absolute error," in IEEE Communications Letters, vol. 13, no. 11, pp. 817-819, November 2009, doi: 10.1109/LCOMM.2009.091128.

[11] D. Paikaray and A. K. Mehta "An extensive approach towards heart stroke prediction using machine learning with ensemble classifier"International Conference on PCCDS 2021.

[12] N. Fatima, L. Liu, S. Hong and H. Ahmed, "Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis," in IEEE Access, vol. 8, pp. 150360-150376, 2020, doi: 10.1109/ACCESS.2020.3016715.

[13] M. Zhu, Y. Shi, A. Li and P. Zhang, "A Bias-Variance Analysis of Multiple Criteria Linear Programming Classification Ensembles," 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008, pp. 447-450, doi: 10.1109/WIIAT.2008.283.

[14] R. Atallah and A. Al-Mousa, "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method," 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS), 2019, pp. 1-6, doi: 10.1109/ICTCS.2019.8923053.

[15] A. Kumari and A. K. Mehta "Effective prediction of COVID-19 using supervised machine learning with Ensemble Modeling" International Conference on PCCDS 2021.