Grenze International Journal of Engineering and Technology, June Issue

Performance Analysis of Heart Disease Prediction System using Machine Learning Approach

Praveen Singh Tomar[1], Prof Kalpana Rai[2] and Dr Mohd Zuber[3]
[1-3]Madhyanchal Professional University Computer Science and Engineering Department , Bhopal. India
Email: kalpana.rai123@gmail.com, mzmkhanugc@gmail.com, psinghtomar@gmail.com

***Abstract*— Heart disease is becoming the leading cause of death in worldwide. In today's world due to work load and changes in the human lifestyle lead to many diseases. Early detection of any disease prevents human life and save it from other damage also, artificial intelligence techniques provide the greatest results for every sector, health care or medical science is one of them. In this paper, we present cardiovascular disease prediction using machine learning techniques such as logistic regression, random forest, and k-nearest neighbor classifier, our result shows that better than existing techniques, here results are evaluated in the form of quality of performance parameters measured such as accuracy, precision, recall, and F1-score, all the results are based on respective confusion matrix which is based on some parameters like true positive rate, false positive rate, true negative rate, and false negative rate.**

***Index Terms*— Machine Learning, Kaggle Dataset, Artificial Intelligence, Accuracy, Prediction System**

I. INTRODUCTION

Your goal is to simulate the usual appearance of papers in a Conference Proceedings or Journal Publications. We are requesting that you follow these guidelines as closely as possible.

World health organization conveys that the death rate from diseases is very high in recent years, and heart disease has a very high rate of death. Heart disease happens because of an unfortunate way of lifestyle, high rate of smoking and drinking, high admission of fats, hypertension, and diabetes. Heart disease also called cardiovascular sickness may happen sometime due to diabetics and sometimes due to coronary disease. World health organization suggests that if we predict heart patients at an early stage so the chances of life may increase and we also reduce the death rate of heart disease. Over the last two decades, many researchers work in the healthcare sector, and applied data mining techniques, artificial intelligence, and optimization techniques.

Recently Artificial intelligence, Machine learning, and deep learning models have been successfully developed in various applications, healthcare sector is one of them, and these models are used to process biomedical data, detection of various diseases, and survival outcomes, and prediction of disease at an early stage. Especially machine learning algorithms are widely applied in clinical diagnosis system, medical imaging, and disease diagnosis and prediction systems to develop computer-aided diagnosis systems.

In this work, we divided the dataset into the train and test dataset in a ratio of 7.5:2.5, where train data work as the teacher on test data. After splitting the dataset pre-processing techniques are applied to remove the noise, identify missing value, and null value, and remove any types of duplicity in the dataset (if exist). Here confusion matrix a gives co-relation between the actual value and predicted value of respective datasets which is based on some

Grenze ID: 01.GIJET.9.2.15

parameters like true positive rate, true negative rate, false positive rate, and false negative rate, where it provide the value in the form of 0 and 1, 0 (zero) if no heart disease present and 1 (one) if any heart disease present. The rest of this paper is organized as follows, the next section presents the literature survey work in the form of a comparative table and graph specific for accuracy parameters, in the next section we present the dataset and division of data i.e. train and test data, in the next section we discuss the performance parameter used in our simulation comparative performance evaluation, in the next section we present the result analysis and finally we present the conclusion and future scope of these work.

II. Existing Work

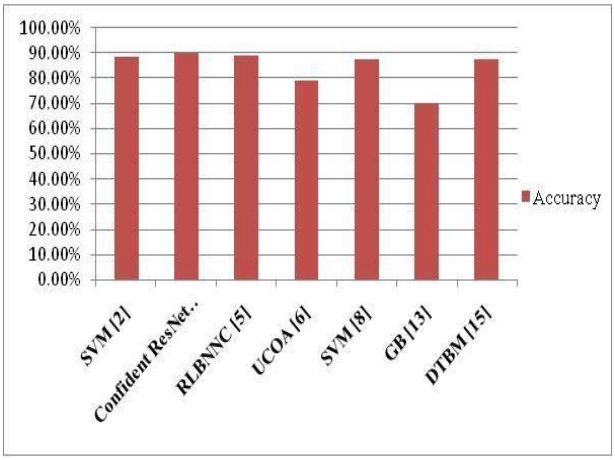
In this section we present the literature survey work which is based on existing work done by researchers, here we include existing researchers' work including used techniques and, published years, and here also include some performance parameters like accuracy to justify their works. Here we discuss the comparative existing work in the form of a table and graph.

TABLE I: STUDY OF EXISTING WORK WITH THEIR ACCURACY AND LIMITATIONS.

Ref No	Techniques	Published Year	Accuracy
[2]	Support Vector Machines	2022	88.4 %
[3]	Confident Res Net	2022	90.2%
[5]	Regressive Learning-Based Neural Network Classifier	2021	89.2 %
[6]	Under sampling Clustering Oversampling Algorithm	2020	79.3 %
[8]	Support Vector Machine	2020	87.7 %
[13]	Gradient Boosting	2021	70.1 %
[15]	Decision Tree Bagging Method	2021	87.5 %

Figure 1: The above figure shows the comparative accuracy for the existing work in health care.

III. Datasets And Attributes



In this section we discuss the heart disease dataset, extracted from the kaggle dataset used for the experimental process, to train the machine learning algorithm, we mention the target column in the data set, then we divide the data set into two small data sets. Training-set to train the algorithm is the Test-set to test it. The below figure explains how we did to apply the Machine Learning algorithm, Where we first decompose our dataset in two parts as mentioned before, then in Training Set we divide the data again, for training and validation.

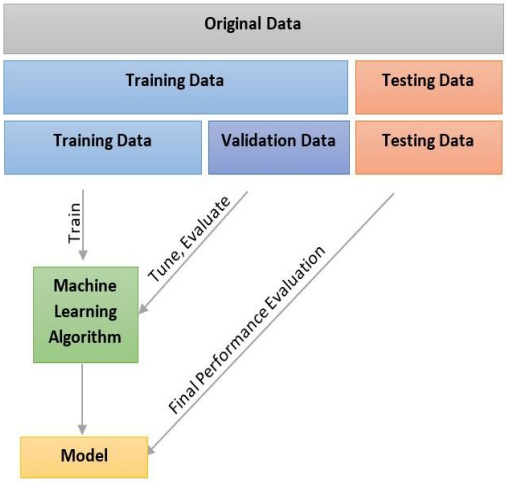


Figure 2: splitting dataset [10].

IV. Performance Parameter

Confusion Matrix(CM) is used to define the delicacy criteria. The CM interprets the items about the work done. The confusion matrix components for the heart disease classification process examined in the study are defined as follows; In the proposed study, various metrics are utilized to compute the performances of MLA techniques. These metrics are calculated on CM. TP (True Positive): Data count, which is essentially true and categorized as true. FN (False Negative): Data count, which is essentially true but categorized as false. FP (False Positive): Data count, which is essentially false but categorized as true. TN (True Negative): Data count, which is essentially false and categorized as false. In the below tables, 0 means "no heart attack risk", and 1 means "heart attack risk".

Table II: Complexity Matrix.

	0	1
0	TP	FN
1	FP	TN

Accuracy: Usually a measure of how often classifiers are estimated correctly. The accuracy rate of the developed algorithm is calculated according to Equation.

$$\text{Accuracy} = TP+TN/TP+TN+FP+FN$$

Sensitivity (Recall): Indicates the rightly classified positive values from the positive samples' total count.

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

Precision: It shows how many of the values we estimated as positive are essentially positive.

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

F1 Score: Precision and sensitivity are the harmonic means, so it's a powerful metric for calculating model performance. The precision in calculating the F-Score is represented as P and R respectively.

$$\text{F1-Score} = 2 * P * R/P+R$$

V. Result Analysis

ML (machine learning) algorithm plays a vital role in every sector nowadays, ML algorithms are a subset of artificial intelligence techniques, in this work we used some machine learning algorithms like logistic regression, random forest, and k-nearest neighbor classifier to predict the heart disease in a patient's and compare performance parameters value like accuracy, precision, recall and, F1-score. that is one supervised, random forest, is used for regression as well as for classification, it is primarily preferred for classification goal. Here we present experimental results and observed that logistic regression gives better results than other machine learning algorithms.

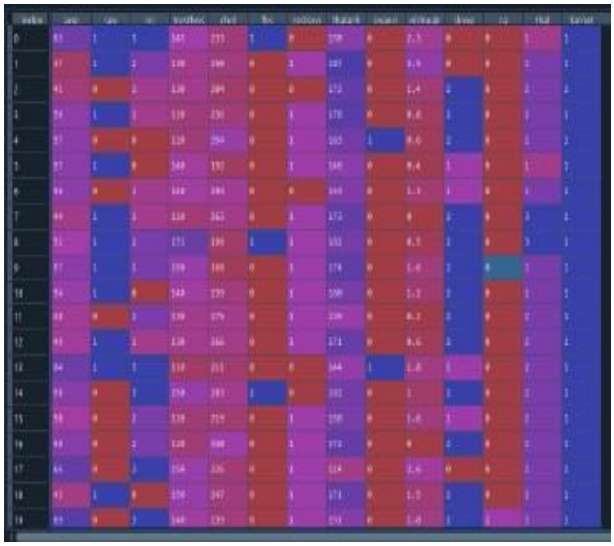


Figure 3: The above picture shows the dataset data frame.

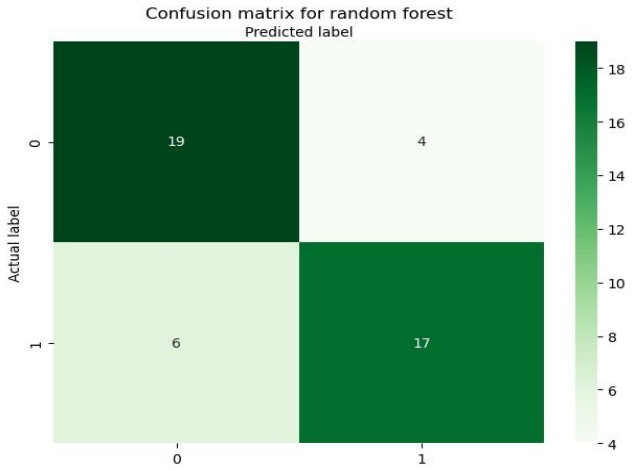


Figure 4: The above picture shows the confusion matrix of random forest models

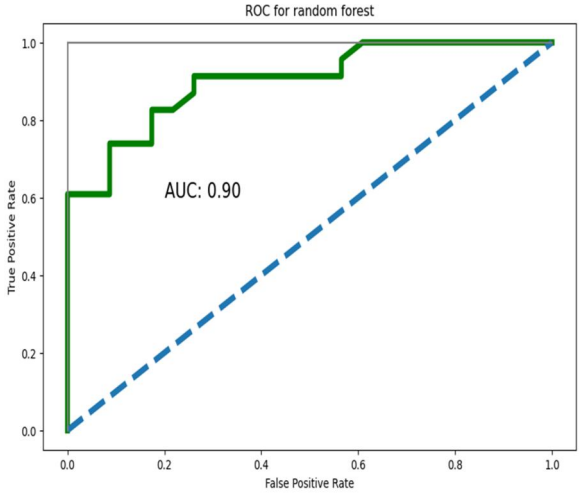


Figure 5: The above picture shows the ROC of random forest models.

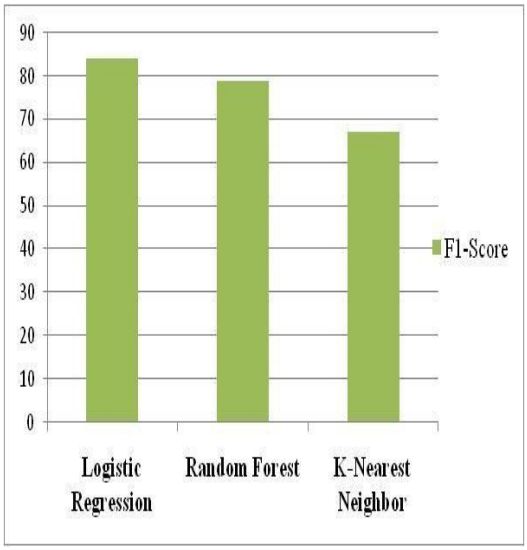


Figure 6: The above picture shows the final F1-score comparison of three different models i.e. logistic regression, random forest, and K-nearest neighbor classifier.

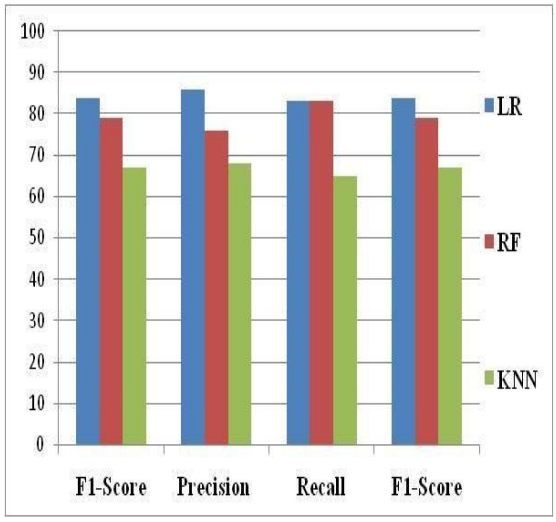


Figure 7: The above picture shows the final Accuracy, precision, recall and, F1-score comparison of three different models i.e. logistic regression, random forest and, K-nearest neighbor classifier.

VI. Conclusion

The result of this work depicts that the logistic regression method provides better results than random forest and, k-nearest neighbor classifier, In this work, we evaluate the performance parameters like accuracy, precision, recall and, f1-score with all these classification techniques with a confusion matrix. In future work, we also suggest some feature reduction techniques and deep learning models with the kernel improve results.

References

[1] Vigneshwaran Namasivayam, Nithyashree Senguttuvan, “Artificial Intelligence and Its Application in Cardiovascular Disease Management”, 2022, pp. 189-207.

[2] Rustem Yilmaz, Fatma Hilal Yagın, “Early Detection of Coronary Heart Disease Based on Machine Learning Methods”, MEDICAL RECORDS-International Medical Journal, 2022, pp. 1-6.

[3] Peng Wang, Zihuai Lin, Xucun Yan, Zijiao Chen, Ming Ding, Yang Song, Lu Meng, “A Wearable ECG Monitor for Deep Learning Based Real-Time Cardiovascular Disease Detection”, IEEE 2022, pp. 1-9.

[4] Deepak Kumar Rathore, Dr. Praveen Kumar Mannepalli, “A Review of Machine Learning Techniques and Applications for Health Care “, International Conference on Advances in Technology, Management & Education, 2021, IEEE proceeding, 978-1-7281-8586-6/21.

[5] A. Sheryl Oliver, Kavithaa Ganesan, S. A. YuvarajT , Jayasankar, Mohamed Yacin Sikkandar, “Accurate prediction of heart disease based on bio system using regressive learning based neural network classifier”, Journal of Ambient Intelligence and Humanized Computing, 2021, pp. 1-10.
[6] Meng Wang, Xinghua Yao, “An Imbalanced-Data Processing Algorithm for the Prediction of Heart Attack in Stroke Patients”, IEEE Access, 2021, pp. 25394-25404.
[7] Aqsa Rahim, Yawar Rasheed, “An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases”, IEEE Access, 2021, pp. 106575-106589.
[8] K. Divya, Akash Sirohi, Sagar Pande, “An IoMT Assisted Heart Disease Diagnostic System Using Machine Learning Techniques”, Springer, 2021, pp. 145-162.
[9] Deepak Kumar Rathore, Dr. Praveen Kumar Mannepalli, “Recent Trends in Machine Learning for Health Care Sector”, International Journal of Innovative Research in Technology and Management, Vol-5, Issue-2, 2021.
[10] Maria-Alexandra Pana,S, tefan-Sebastian Busnatu, “Reducing the Heart Failure Burden in Romania by Predicting Congestive Heart Failure Using Artificial Intelligence: Proof of Concept”, Appl. Sci. 2021, pp. 1-13.
[11] Sharmila Zope, Nandini Chaudhari, Nita Patil, “Heart Disease Prediction using Machine Learning Techniques Comparative Study”, 2021.pp. 1-7.
[12] D.R. Krithika, K.Rohini, “Comparative Intrepretation Of Machine Learning Algorithms In Predicting The Cardiovascular Death Rate For Covid-19 Data”, International Conference on Computational Intelligence and Knowledge Economy, 2021, pp. 394-400.
[13] Hana H. Alalawi, Manal S. Alsuwat, “Detection of Cardiovascular Disease using Machine Learning Classification Models”, IJERT, 2021, pp. 151-156.
[14] Sandeep Chandra Bollepalli, Ashish Kumar Sahani, “An Optimized Machine Learning Model Accurately Predicts In Hospital Outcomes at Admission to a Cardiac Unit”, Diagnostics 2022, pp. 1-14.
[15] Pronab Ghosh Sami Azam, “Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms with Relief and LASSO Feature Selection Techniques”, IEEE Access, 2021, pp. 19304-19326.
[16] R Dubey, D Rathore, D Kushwaha, JP Maurya, “An empirical study of intrusion detection system using feature reduction based on evolutionary algorithms and swarm intelligence methods”, International Journal of Applied Engineering Research 12 (19), 2017. pp. 8884-8889.
[17] Prasannavenkatesan Theerthagiri, Usha Ruby A, “Diagnosis and Classification of the Diabetes Using Machine Learning Algorithms”, 2021, pp. 1-24.
[18] Deepak Kumar Rathore, Praveen Kumar Mannepalli, “Diseases Prediction and Classification Using Machine Learning Techniques”, AIP Conference Proceedings 2424, 070001 (2022); https://doi.org/10.1063/5.0076768.
[19] Jasjit S. Suri, Mrinalini Bhagawati, “Understanding the bias in machine learning systems for cardiovascular disease risk assessment: The first of its kind review”, Computers in Biology and Medicine, 2022, pp. 1-24.