Measuring the Heart Attack Possibility using Different Types of Machine Learning Algorithms

Maria Sultana Keya
Daffodil International University
Dhaka, Bangladesh
maria15-1215@diu.edu.bd

Muhammad Shamsojjaman
Daffodil International University
Dhaka, Bangladesh
shamsojjaman15-1484@diu.edu.bd

Faruq Hossain
Daffodil International University
Dhaka, Bangladesh
faruqhossain15-2700@diu.edu.bd

Farzana Akter
Daffodil International University
Dhaka, Bangladesh
farzana15-2579@diu.edu.bd

Fakrul Islam
Daffodil International University
Dhaka, Bangladesh
fakrul15-1316@diu.edu.bd

Minhaz Uddin Emon
Daffodil International University
Dhaka, Bangladesh
minhaz15-1294@diu.edu.bd

***Abstract*—The heart seems to be a very complicated organ. And if some of it may have been seriously damaged, the remainder of the heart is still functioning. But as a result of the injury, the heart can be weakened and unable to pump as much blood as normal. With timely detection of multiple possible hamstring issues, proper care, and dietary changes after a heart attack, and additional injury can be reduced or avoided. In this paper different types of machine learning algorithms are used for measuring the heart attack possibility. They are logistic regression, random forest, bagging, MLP, and decision tree. Finding the best algorithm with this paper also shows the correlation matrices, visualizes the feature, and AUC. In this paper logistic regression is the best model with an Accuracy is 80% and also gives the best AUC which is 87%.**

***Keywords*—Hamstring issues, Machine learning algorithms, Correlation matrices, Accuracy, AUC.**

I. INTRODUCTION

Cardiovascular disorder identifies a number of heart disorders. Disease under the umbrella of cardiovascular disease encompasses disorders of the vessel's blood, including coronary artery disease; heart rhythm disorders, and heart defects in which you are born, like (congenital heart defects). The heart is the most important organ in our body. The heart is our body's lifeline. Heart disease is a common disease. One of the common illnesses is a heart attack. The supply of blood does not act perfectly when a heart attack takes place. Fat, cholesterol and other compounds are most commonly accumulated after a heart attack. As part of the cardiac muscle does not get enough oxygen, an attack, or myocardial infarction also occurs. Heart disease symptoms: Pressure or discomfort in the abdomen. Much heart attacks cause emotional or left chest pain, lasting longer than a few minutes or leaving and returning. Uncomfortable strains, rubbing, fullness or pain may feel discomfort. Weakness, faintness, or weakness. You might break into a cold sweat as well. Mask, neck, or back pressure or irritation. In either or both arms or shoulders, pain or irritation. Breath shortening. Sometimes this leads to heart pain, but lack of oxygen can also arise before heart attacks. Statistics for heart disease: 735,000 Americans had heart attack last year The main cause of death in the US Cardiovascular Disease . Only over 0.3% of men and women aged 20 to 39 suffer from heart disease The first heart attack averages 65 years for men and 72 years for women.

Factor of risk: The key risk factors for control of a heart attack include: Alter. Alter. Men 45 or older and females 55 or older are more likely than younger men and women to have a heart attack. Smoking. Smoking. Smoking and exposure to second-hand smoke was included. The hypertension in the blood. High blood pressure can damage your heart's arteries over time. High blood pressure, such as obesity, high cholesterol, or diabetes, raises the chances even more. Higher amounts of serum or triglyceride cholesterol. Low density lipoprotein is potentially associated with high cholesterol ("bad" cholesterol) (LDL). A high amount of triglycerides, a type of blood fat associated with your diet, also raises your risk of a heart attack. The risk would be minimized by a large volume of high-density cholesterol ("good" cholesterol") lipoprotein (HDL). Abesthood. Obesity is related to elevated serum cholesterol levels, high triglyceride concentrations, high blood pressure and diabetes. By losing just 10 percent of the body weight, this risk is reduced. Diabetes. Diabetes. Not enough or insulin secreted hormones allow the body to raise blood sugar levels, thereby raising your risk of a heart attack. The effect of insulin does not affect the amount of blood sugar you get. Substantial metabolism. This is because you have smoking, asthma and high blood sugar. For metabolic syndrome you will have heart disease twice as often as you do not.

Heart assaults family past. You may be prone to greater risk because your siblings, parents or ancestors have suffered early heart attacks (55 years old for men, 65 years old for women). Failure to workout. Blood levels are increased with inactive serum cholesterol and obesity. People who practice daily have improved cardiovascular wellbeing, including lower blood pressure. Stress: In order to increase the chance for a heart attack, you should respond to stress. Illegal use of drugs. A spasm of the coronary artery will cause a heart

attack if relaxing narcotics like cocaine or amphetamines are used. A pre-eclampsia history. This syndrome raises the risk of lifetime heart failure, causing elevated blood pressure during pregnancy. A state of autoimmune. A disease like rheumatoid arthritis or lupus will increase the risk.
So, heart attack possibility defends this disease to analysis them.
The rest of the paper organized as follows:

- Related work describe in section II.
- Algorithms description describe in section III.
- In section IV proposed model are shown.
- In section V result discussion and analysis are performed
- And finally section VI conclusion are described.

II. RELATED WORK

In 1940, cardiovascular disorder had become the main source of mortality among Americans [1]. When President Franklin D. Roosevelt died of hypertensive heart disease in 1945, Americans became interested in research on heart disease [5]. In 1948, The Framingham Heart Study was first established in the United States to study heart disease and its various risk factors [2]. Since then, there has been constant research on heart failure and its various risk factors on how to better prevent it using modern technology. Fizar Ahmed [3] explained the architecture of heart attack rates and used the IoT concept to predict future heart attack patients in his paper and also he used one of the most popular machine learning algorithms like kNN (k Nearest Neighbour) to complete his work, achieving better accuracy. Prince Kansal et. al. [4] mainly they are focused on how early to predict heart disease using various data mining techniques and they used almost 4 algorithms of machine learning. The attributes they used in their dataset like age, sex, blood pressure, and blood sugar. They achieved better accuracy from the decision tree. Asha Rajkumar [5] manually diagnosed heart disease is very time consuming and required experts. They focused on her paper fast diagnosis of heart disease using data mining techniques. They used three algorithms and tanagra tools. It took 609ms to diagnose the heart disease using the Naive Bayes algorithm. Salman, Issam [6] mainly focused on mortality prediction because of a heart attack using different machine learning approaches. To complete his work he used real-time data. For achieving better accuracy he used different types of classifier NB-tree, Bayesian Network etc. Jesse Davis et al. [7] eMedical has revolutionized our traditional treatment and made it much easier to get patient data online. As a result of eMedical, we nowadays treat many diseases online. We often do not understand whether this medicine will have any other side as a result of us?. This paper focuses on the risk of heart attack in eMedical. Himanshu Sharma et al. [8] focused on summarised state of art technologies and available approaches for prediction of heart attack disease. They have used both machine learning and deep learning concepts and basically focused on comparative analysis of heart disease detection. Beant Kaur et al. [9] provide reviews on many heart disease prediction research and achieving better results from neural networks is nearly 100% in one paper [10]. Gayathri et al. [11] focused on diagnosis of different types of heart diseases (cardiovascular disease, coronary artery disease, heart failure, ischemic heart disease etc) using soft computer technology and data mining approaches.The main limitation for them is that when the dataset gets bigger, good results do not come. Previously, much research has been done on heart attack possibility prediction but we focused on how to measure heart attack possibility using new machine learning approaches.

III. ALGORITHMS DESCRIPTION

The algorithms used for this research are listed below. Descriptions are discussed. In this research total number of five algorithms are used to evaluate performance

- **Logistic regression :** Logistic regression is a well-known machine learning algorithm that falls under the technique of supervision learning. Since logistic regression and linear regression are supervised by nature, these algorithms use labeled datasets to make predictions. Variables are regulated in the regression analysis to separate the relationship. When using linear regression and logistic regression in the case of nonlinear regression ships, only logistic regression can be used. In order to solve regression problems, linear regression is used, while logistic regression is used to solve classification problems [12].
- **Bagging :** The bootstrap aggregation, also known as bagging, is a meta-algorithm for machine learning that improves the consistency and accuracy of algorithms used in statistical classification and regression. Here we have a bag of hypotheses. It's a preferred symbol of hypotheses that might be produced by different algorithms. We were going to see later on how we can obtain them but let's say for now that we have different hypotheses now if we want to make a classification based on different hypotheses then a simple way could just be to get each hypothesis to make a prediction and then simply return the class that has that is the most popular among those predictions, in other words, the majority class [13].
- **Random forest :** Random forest or forest random decision is a strategy that works during the training process by creating multiple decision trees. The random forest selects for the final decision the decision of the majority of trees [14]
- **Decision tree :** A shaped diagram used to evaluate a course of action is a Decision Tree. Each branch of the tree is a potential option, occurrence or reaction. We consider the whole training set as the root at the beginning. For data benefit, attributes are assumed to be categorical and attributes are assumed to be continuous for the gini index. Records are recursively distributed on the basis of attribute values. As a root or internal node, we use statistical methods to order attributes [15].
- **MLP :** A multicolored (MLP) sensor is a kind of artificial neural feed forward network. A MLP has at least three node layers: an input, a hidden layer and a display layer. Of node is an anti linear activation function with the

exception of the input nodes. Every node. MLP uses a supervised learning method known as teaching back propagation. The MLP varies from a linear perceptron by several layers and non-linear activation. Data that cannot be linearly segregated can be distinguished [16] [17] [18].

IV. PROPOSED MODEL

Figure 1 indicates that each segment of our proposed model is seen in a separate succinct description.

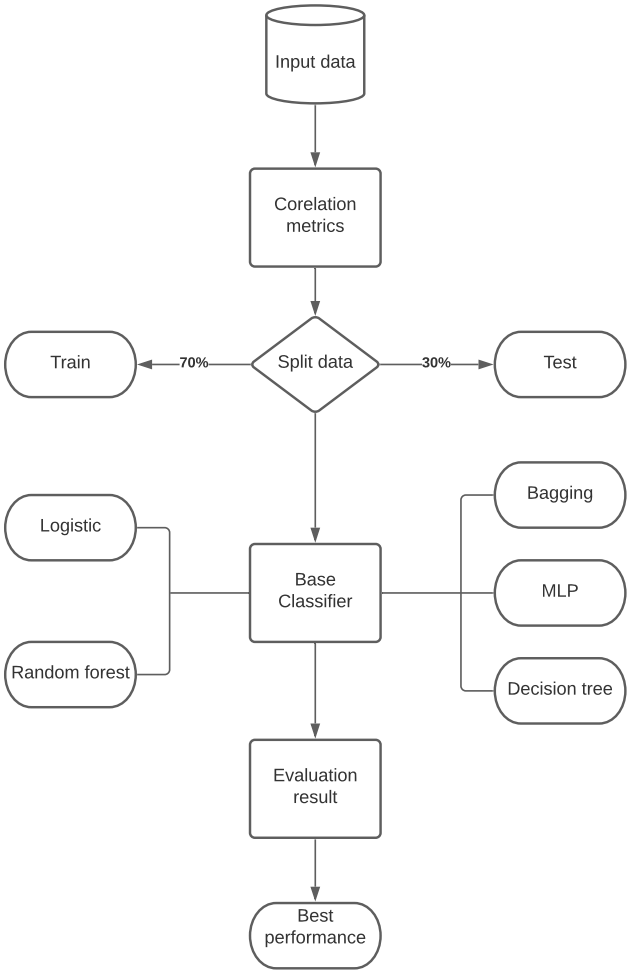

Fig. 1. Proposed Model Step By Step Procedure

1) **Input Data:** In this research work with 303 data. Where 12 attributes are observable and then all of them are floating data. And there's a decision class/class variable. This data was collected from UCI machine learning repository.
2) **Corelation Metrics** In fig 2, corelation metrics are shown
3) **Split Data:** In this research 70% data use for train model and 30% data use for testing purpose.
4) **Base Classifier** logistic regression, random forest, bagging, MLP, and decision tree are used.

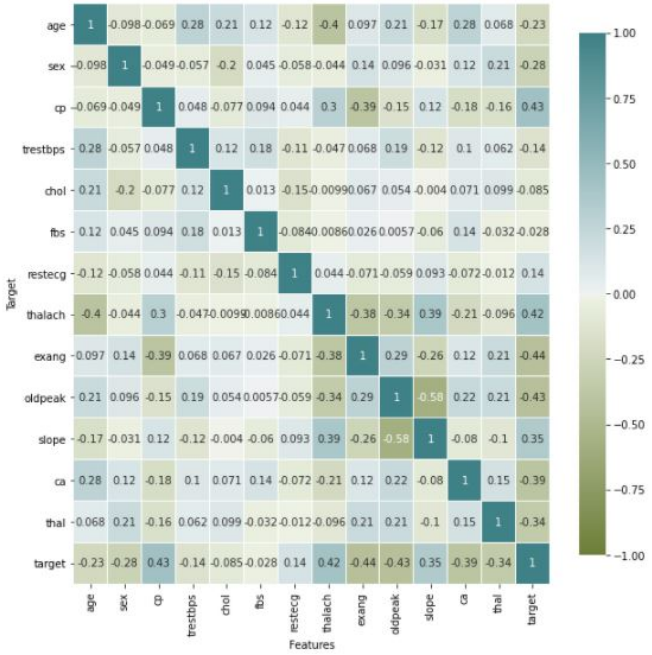

Fig. 2. Corelation metrics

5) **Evaluation Result** In the classification report we were able to find out the confusion matrix, precision, recall and many result are shown for all algorithms
6) **Best Algorithms:** In this analysis the result depends on some part of this research. However, which algorithm gives the best true positive, false positive, true negative, and false negative are the best algorithms in this analysis.

V. RESULT DISCUSSION AND ANALYSIS

Results: In this research table I represent the confusion matrix of applying classifier. The best result shown in logistic regression. In table II classification result are performed. The Precision, recall, f measure, and AUC(area under curve) are analysis here.

- **Precision**: Precision is the fraction of records obtained that are important to the query in the field of information retrieval. For eg, the number of correct results divided by the number of results returned in a text search on a collection of documents is precise.
- **Recall**: Recall is a fraction of the related records which are effectively collected while obtaining information. For ex: the number of correct results divided by the number of results to be retrieved is recalled for a text search of a collection of documents.
- **F measure**: The F-score or F-measure is a measure of the accuracy of a test in the statistical study of the binary classification.
- **AUC**: In all practicable classification threshes, the AUC offers a detailed measure of success. One of the ways AUC can be viewed is that the model ranks higher than a random negative random positive example

TABLE I
APPLYING CLASSIFIER CONFUSION METRICS ANALYSIS.

Model Name	Accuracy(%)	Label	Predictive Negative	Predictive Positive
logistic	80	**Actual Negative**	30	11
		Actual Positive	7	43
random forest	75	**Actual Negative**	29	12
		Actual Positive	11	39
bagging	74	**Actual Negative**	28	13
		Actual Positive	11	39
mlp	78	**Actual Negative**	26	15
		Actual Positive	15	45
decision tree	70	**Actual Negative**	30	11
		Actual Positive	16	34

TABLE II
APPLYING CLASSIFIER RESULT ANALYSIS.

Model Name	Class	Precision(%)	Recall(%)	F measure(%)	AUC(%)
logistic	**0**	81	73	77	87
	1	80	86	83	
random forest	**0**	72	71	72	84
	1	76	78	77	
bagging	**0**	72	68	70	81
	1	75	78	76	
mlp	**0**	84	63	72	82
	1	75	90	82	
decision tree	**0**	66	73	79	71
	1	76	68	72	

VI. CONCLUSION

In this research analysis the heart attack possibility performance are shown. The performance analysis have figure out various categories such as confusion metrics, precision, recall, f measure, and auc. In overall performance logistic regression gives the best accuracy from other, and its accuracy is 80%. Some machine learning technique are perform to figure out best performance from our dataset. In future this study will be added deep learning and artificial intelligent method to analysis and predict heart attack possibility.Also added more data as needed.

REFERENCES

[1] Kannel, William B. "Contribution of the Framingham Study to preventive cardiology." Journal of the American College of Cardiology 15.1 (1990): 206-211.

[2] Mahmooda, S. S., Levy, D., Vasan, R. S., Wang, T. J. (2014). The Framingham Heart Study and the epidemiology of cardiovascular diseases: a historical perspective. Lancet, 383(9921), 999-1008.

[3] Ahmed, Fizar. "An Internet of Things (IoT) application for predicting the quantity of future heart attack patients." International Journal of Computer Applications 164.6 (2017).

[4] Methaila, A., Kansal, P., Arya, H., Kumar, P. (2014). Early heart disease prediction using data mining techniques. Computer Science Information Technology Journal, 53-59.

[5] Rajkumar, Asha, and G. Sophia Reena. "Diagnosis of heart disease using data mining algorithm." Global journal of computer science and technology 10.10 (2010): 38-43.

[6] Salman, Issam. "Heart attack mortality prediction: an application of machine learning methods." Turkish Journal of Electrical Engineering Computer Sciences 27.6 (2019): 4378-4389.

[7] Davis, Jesse, et al. "Machine learning for personalized medicine: Will this drug give me a heart attack." The Proceedings of International Conference on Machine Learning (ICML). 2008.

[8] Sharma, Himanshu, and M. A. Rizvi. "Prediction of heart disease using machine learning algorithms: A survey." International Journal on Recent and Innovation Trends in Computing and Communication 5.8 (2017): 99-104.

[9] Kaur, Beant, and Williamjeet Singh. "Review on heart disease prediction system using data mining techniques." International journal on recent and innovation trends in computing and communication 2.10 (2014): 3003-3008.

[10] Dangare, Chaitrali, and Sulabha Apte. "A data mining approach for prediction of heart disease using neural networks." International Journal of Computer Engineering and Technology (IJCET) 3.3 (2012).

[11] Gayathri, P., and N. Jaisankar. "Comprehensive study of heart disease diagnosis using data mining and soft computing techniques." (2013).

[12] Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., Klein, M. (2002). Logistic regression. New York: Springer-Verlag.

[13] Bühlmann, P., Yu, B. (2002). Analyzing bagging. The Annals of Statistics, 30(4), 927-961.

[14] Pal, M. (2005). Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), 217-222.

[15] Rokach, L., Maimon, O. (2005). Decision trees. In Data mining and knowledge discovery handbook (pp. 165-192). Springer, Boston, MA.

[16] Taud, H., Mas, J. F. (2018). Multilayer perceptron (MLP). In Geomatic Approaches for Modeling Land Change Scenarios (pp. 451-455). Springer, Cham.

[17] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, M. S. A. Mamun and M. S. Kaiser, "Performance Analysis of Machine Learning Approaches in Stroke Prediction," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2020, pp. 1464-1469, doi: 10.1109/ICECA49313.2020.9297525.

[18] M. A. Islam, M. S. Islam, M. S. Hossen, M. U. Emon, M. S. Keya and A. Habib, "Machine Learning based Image Classification of Papaya Disease Recognition," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2020, pp. 1353-1360, doi: 10.1109/ICECA49313.2020.9297570.