ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

Prediction of Heart Disease Utilising Machine Learning Techniques

Dr. D.V. Chandra Sekhar,

Professor, P.G. Department of Computer Science, TJPS College, Guntur

dvchand.info@gmail.com

U. Harita,

Assistant Professor, Koneru Lakshmaiah Education Foundation, Vaddeswaram Guntur

uharita@gmail.com

Abstract

The utilisation of machine learning and artificial intelligence has proven to be valuable across multiple academic fields throughout their progression, particularly in response to the substantial growth of data in recent times. Utilising disease forecasts can enhance the reliability and expediency of decision-making processes. Machine learning algorithms are progressively being utilised to forecast diverse phenomena.The topic of discussion pertains to various illnesses and medical conditions. The construction of a model can also facilitate the visualisation and analysis of diseases, hence enhancing the consistency and accuracy of reporting.This study has examined the utilisation of several machine learning algorithms for the purpose of detecting cardiac disease. The research conducted in this paper has demonstrated a two-step procedure. The dataset pertaining to heart illness is initially formatted in a manner that is suitable for utilisation in machine learning methods. The UCI repository is utilised to collect medical records and other pertinent patient information. The dataset pertaining to heart illness is afterwards utilised to ascertain the presence or absence of cardiovascular ailments among the patients. Furthermore, this paper presents a multitude of significant findings. The efficacy of machine learning methods, including Logistic Regression, Support Vector Machine, K-Nearest Neighbours, Random Forest, and Gradient Boosting Classifier, is assessed by means of the confusion matrix. Based on current research, it has been shown that the Logistic Regression method exhibits a notable accuracy rate of 95% in comparison to alternative algorithms. Additionally, it demonstrates superior accuracy in terms of f1-score, recall, and precision compared to the other four distinct algorithms. Nevertheless, a crucial aspect of this research involves the pursuit of enhancing the precision rates of machine learning algorithms

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

to a range of around 97% to 100%. This endeavour represents a significant area of focus and presents a formidable challenge for future investigation.

Keywords: Machine Learning, Artificial Intelligence, Heart Disease, Linear Regression, Support Vector Machine, K-Nearest-Neighbors, Random Forest, Decision Tree, Gradient Boosting

1. Introduction

Machine learning (ML) is a component of artificial intelligence (AI) that enables a software application to enhance its predictive accuracy through iterative processes, without the need for explicit programming. Machine learning algorithms utilise previous data as input in order to predict new output values [1]. The field of machine learning is of great importance and exhibits a wide range of diversity, with its boundaries and practical implementation continuously developing on a daily basis. Due to this rationale, machine learning has emerged as a pivotal factor for gaining a competitive edge in numerous organisations. Machine learning encompasses many types of classifiers, namely supervised, unsupervised, and ensemble learning, which are employed to make predictions and evaluate the precision of a given dataset. Machine learning algorithms have the capability to construct a model using a set of sample data, commonly referred to as training data, in order to make informed decisions or predictions [1, 2].

The current work examines the application of machine learning techniques in the medical field. The primary objective is to replicate certain human behaviours or cognitive processes and accurately identify disorders using diverse input sources [3].The phrase "cardiovascular disease" encompasses a range of medical diseases that impact the functioning of the heart. Based on studies from the World Health Organisation, it has been observed that cardiovascular illnesses have emerged as the primary cause of mortality on a global scale, with an estimated 17.9 million deaths [4, 5]. Numerous studies have been conducted and various machine learning algorithms have been employed in the investigation and diagnosis of cardiac disorders. Ghumbre et al. (year) utilised machine learning and deep learning techniques for the purpose of predicting heart illnesses in the UCI dataset [3]. The researchers reached the conclusion that machine learning algorithms had superior performance in doing this investigation. In a

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

publication by Rohit Bharti et al., the authors discuss the application of machine learning techniques for the prediction of heart illness. The study concludes that the utilisation of various data mining and neural systems is necessary to accurately assess the severity of heart disease in patients [4]. Certain analysis has prompted consideration of the potential application of a predictive data mining approach on the identical dataset [5]. The study conducted by Jee S H et al. focuses on the application of machine learning techniques for the prediction of cardiac disease.

The training and testing datasets are utilised through the application of a specific methodology.The neural network algorithm is discussed in reference [6]. The K-Nearest Neighbour algorithm is a machine learning technique that is commonly used for classification and regression tasks.

The article authored by Mai Shouman et al. [7] was examined in order to assess its effectiveness in diagnosing cardiac disease. Several efficient algorithms have been employed for the detection of Huntington's disease (HD), which The findings demonstrate that each algorithm possesses distinct strengths in achieving the specified objectives [8]. Raihan M et al. [9] conducted a study in which they utilised a supervised network for the purpose of diagnosing HD. This study concept has garnered global attention and stimulated scholarly discourse through the publication of numerous works [10-15].

This article aims to develop a machine learning predictive model for the analysis of heart disease based on medical history.

The data is obtained from the UCI repository, which contains patients' medical records and associated information. The dataset would be employed for the purpose of predicting the presence or absence of cardiac disease in patients. This article examines 14 patient variables in order to diagnose the HD dataset. The classification process determines the presence or absence of a condition and can aid in the diagnosis of diseases with reduced reliance on medical interventions [1, 5]. This study examines multiple patient features, including age, sex, serum cholesterol levels, blood pressure, and the presence of exercise-induced angina (exang). Five distinct machine learning (ML) methods, including Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest-Neighbors (KNN), Random Forest (RF), and Gradient Boosting Classifier (GBC), are utilised to classify and forecast instances of cardiac disease. This article presents a number of advantageous outcomes. The dataset's attributes have been learned using

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

the following algorithms. A comparison analysis of algorithms has been conducted to assess the accuracy rate based on the characteristics of the HD dataset. All of the machine learning algorithms that were chosen demonstrate efficiency by their accuracy, which exceeds 80%. The approach that demonstrates the highest level of efficiency is Logistic Regression (LR), yielding an accuracy rate of roughly 95%. In this study, the Logistic Regression (LR) method will be utilised to make predictions and diagnose the presence of heart disease in patients.

The present article has been reorganised in a sequential manner. Section 2 of the document provides a comprehensive discussion of the methods employed in the study. Section 3 provides a concise overview of several machine learning algorithms. The findings and subsequent examination are presented in the next section. In this section, we will discuss the fourth component. In the results section, a comparison is made between algorithms based on the evaluation of the confusion matrix. Section 5 encompasses the final analysis and the identification of future prospects.

2. Methodology

This section explains the procedure and analysis that are part of this study. The first steps in this investigation are data gathering and attribute selection. The necessary data is then preprocessed into the desired format. The provided information is subsequently partitioned into a training dataset and a test dataset.Algorithms are applied, and the provided data is utilised to "train" the model. This model's precision is determined by its performance in testing. Multiple modules, including data collecting, attribute selection, data pre-processing, data balancing, and disease prediction, are used to power the processes in this investigation.

2.1 Data Collection

The dataset utilised in this article was obtained from the UCI repository, a widely recognised source for research analysis as acknowledged by multiple authors [4, 7]. The initial phase involves organising the dataset obtained from the UCI repository in order to make predictions regarding heart disease. Subsequently, the dataset is partitioned into two distinct regions, namely the training set and the testing set. This article use 80% of the available data as a training dataset, while the other 20% is allocated for testing reasons.

2.2 Dataset and Attributes

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

The attributes of a dataset refer to the inherent qualities of the dataset that are crucial for analysis and prediction within the context of our inquiry. Several patient variables, including as gender, chest discomfort, serum cholesterol, fasting blood pressure, and exang, are taken into account when predicting illnesses. The correlation matrix can be utilised for attribute selection in order to develop a model.

Table 1. Attributes used are listed.

Sl. No.	Attributes	Description	Values
1.	Age	Patients age in years	Continuous
2.	Sex	Sex of subject (male-0, female-1)	Male/Female
3.	CP	Chest pain type	Four types
4.	Trestbps	Resting blood pressure	Continuous
5.	Chol	Serum cholesterol in mg/dl	Continuous
6.	FBS	Fasting blood pressure	< or >120 mg/dl
7.	Restecg	Resting Electrocardiograph	Five values
8.	Thalach	Maximum heart rate achieved	Continuous
9.	exang	Exercise Induced Angina	Yes/No
10.	oldpeak	ST Depression introduced by exer.	Continuous
11.	slope	Slope of Peak Exercise ST segment	up/flat/down
12.	Ca	Number of major vessels	0-3
13.	thal	Defect type	Reversible/Fixed/Normal
14.	Targets	Heart disease	1 (disease), 0 (no disease)

2.3 Pre-procesing of Data

In order to get precise and optimal outcomes, it is important to engage in the process of data cleaning, which involves the elimination of missing or noisy values from the dataset. By employing established methodologies in Python version 3.8, it is possible to address the issue of missing and noisy values, as demonstrated in reference [16]. Next, it is necessary to modify our dataset by taking into account the normalisation, smoothing, generalisation, and aggregation of the dataset.

Integration is a pivotal stage in the process of data preprocessing, wherein several factors are taken into account for the purpose of integration. Occasionally, the dataset may exhibit a higher degree of complexity or present challenges in terms of comprehension. In this scenario, it is necessary to transform the dataset into a prescribed format in order to achieve optimal outcomes.

2.4 Balancing of Data

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

The act of balancing the dataset is crucial in order to enhance the efficacy of machine learning algorithms. In the context of dataset composition, a balanced dataset is characterised by an equal distribution of input samples across different output categories .The class, also referred to as the target class. The imbalance within a dataset can be rectified through the implementation of two strategies, namely under sampling and over sampling.

2.5 Prediction of Disease

This article presents the implementation of five distinct machine learning algorithms for the purpose of classification. A comprehensive examination of the algorithms has been conducted. This paper examines a machine learning system that demonstrates the highest accuracy rate in predicting cardiac illness, as depicted in Figure 1.

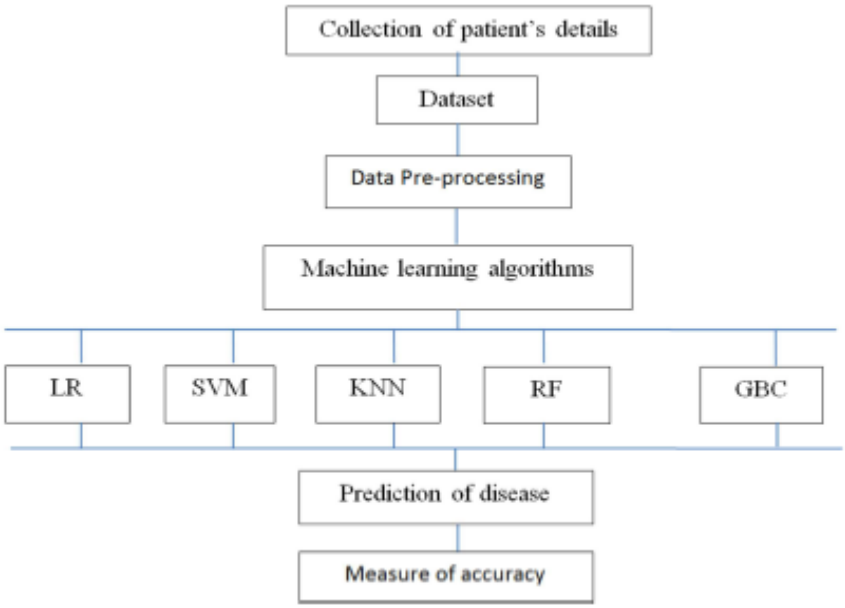

Figure 1. Architecture of prediction models.

3. Machine Learning Algorithms

Machine learning is a data analysis technology that facilitates the automated generation of analytical models. This study examines the accuracy of five distinct algorithms in order to determine the most effective one.

3.1 Logistic Regression Model

The machine learning model frequently employed for classification and predictive analysis is commonly referred to as logit regression [16]. It is also employed to estimate discrete values, such as binary outcomes, based on a set of independent factors. A binary outcome refers to a

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

situation in which there are only two possible results: the occurrence of an event (denoted as 1) or the non-occurrence of the event (denoted as 0).

The next section outlines the operational processes of the Logistic Regression model.

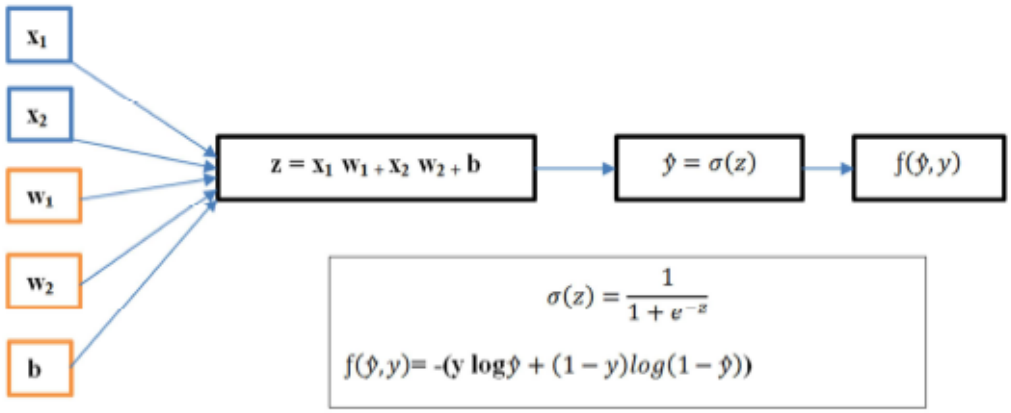

Figure 2. Logistic Regression model.

The function z is dependent on the variables x1, x2, w1, w2, and b. The variable z represents a linear equation that is utilised as an input for a sigmoid function in order to make predictions about the outcome. The loss function is utilised to assess the performance of the model through calculation. The cross-entropy loss function is utilised in this particular scenario [17].

3.2 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is well recognised as a prominent supervised machine learning technique that is commonly employed for both classification and regression tasks [23]. While this technique is largely focused on classification problems in machine learning, it can also be used to other domains. The primary objective of the Support Vector Machine (SVM) technique is to generate an optimal decision border or hyperplane that effectively separates classes within an n-dimensional space. This enables efficient classification of fresh data points into their respective categories. The hyperplane [23] is commonly referred to as the optimal decision boundary. Support Vector Machines (SVM) are a class of supervised learning algorithms that aim to identify the optimal hyperplane by selecting the most influential vectors, known as support vectors.

The support vectors, which are the extreme vectors, are associated with the support vector machine method. Presented below is a diagram illustrating the Support Vector Machine (SVM)

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

algorithm, wherein the decision boundaries, also known as hyperplanes, effectively classify distinct categories.

The dataset used for training consists of a pair of vectors, denoted as (x2, x1), where x1 represents the x-axis vector and x2 represents the target vector. Please refer to Figure 3 for a visual representation.

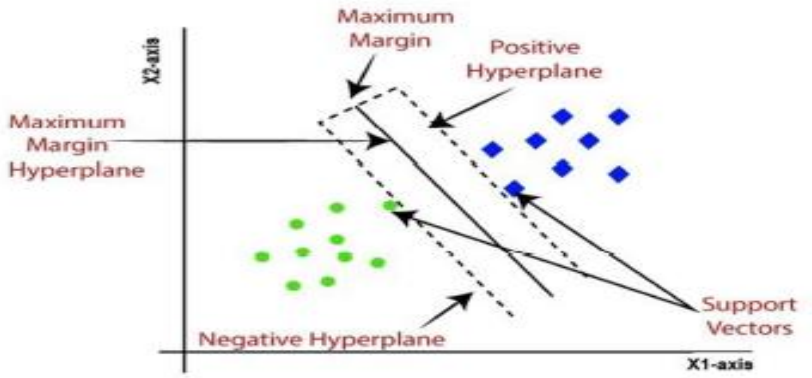

Figure 3. Support Vector Machine.

3.3. K- Nearest Neighbours (K-NN)

The K-Nearest Neighbours (K-NN) method is a classification technique that is widely used in supervised learning. It is considered to be the most straightforward approach in this field. However, the K-Nearest Neighbours (K-NN) algorithm can also be employed for regression tasks, although its primary application is in classification tasks [18]. The classification of a new data point is determined by the K-Nearest Neighbours (K-NN) algorithm, which relies on the degree of similarity between the new data and the previously stored data. The observation suggests that the K-NN algorithm has the ability to efficiently identify newly encountered data that falls inside an appropriate category, as depicted in Figure 4.

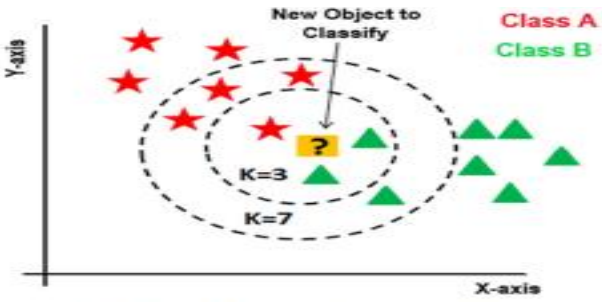

Figure 4. K-Nearest Neighbors.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

In this case, the x-axis represents an independent variable, while the y-axis represents a dependent variable in a function. The K-NN method is illustrated in its simplest form in Figure 4. This algorithm expects the test sample (Yellow Square with what symbol) to be labelled as either a green triangle or a red star. Since green triangles outnumber red stars by a large margin, the yellow square becomes a green triangle when k=3 is examined in a tiny dash circle. The number of red stars is four, and the number of green triangles is three, so the yellow square would be red stars if we assume k=7, which is in a wide dash circle. According to Figure 5, this situation highlights the significance of the regional majority vote.

3.4 Random Forest

The Random Forest (RF) technique is widely utilised in supervised machine learning for the purposes of classification and regression. However, its primary application is in the realm of categorization problems. The RF algorithm is founded upon the principle of ensemble learning. Ensemble learning is a widely applicable machine learning technique that can enhance predicted performance by combining multiple learning algorithms [2, 19]. The Random Forest (RF) technique involves generating many decision trees based on the data samples. Each tree provides a forecast, and the final solution is determined by evaluating the majority voting outcome. It has been observed that the ensemble technique surpasses a single decision tree in terms of performance, since it effectively addresses the issue of over-fitting by aggregating findings through averaging. The utilisation of a substantial quantity of decision trees in Random Forest (RF) facilitates the attainment of higher accuracy levels and serves as a preventive measure against overfitting issues. The subsequent steps are executed by the RF algorithm, as depicted in Figure 6.

In the initial step, a random sample of n numbers is taken from a provided dataset.

In the second step, a decision tree will be generated for each individual.

In the third step, every decision tree will generate a prediction for the output.

Step 4: The final outcome is determined through a process of majority vote or averaging.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

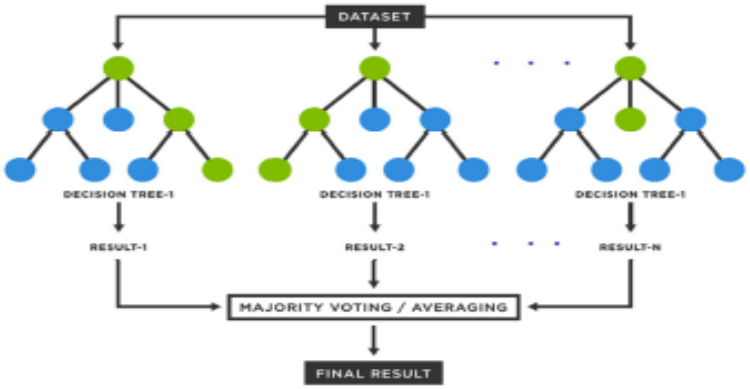

Figure 5. General procedure of Random Forest.

Gradient Boosting (GB) is a machine learning methodology commonly employed in the context of classification and regression tasks, similar to other approaches. The approach holds significant computational capabilities within the domain of machine learning [21]. It is widely recognised that faults in machine learning algorithms can be categorised into two distinct types: bias error and variance error. The Gradient Boosting Classifier (GBC) aids in the reduction of bias error in the model, as depicted in Figure 6. A diagram is defined as follows below.

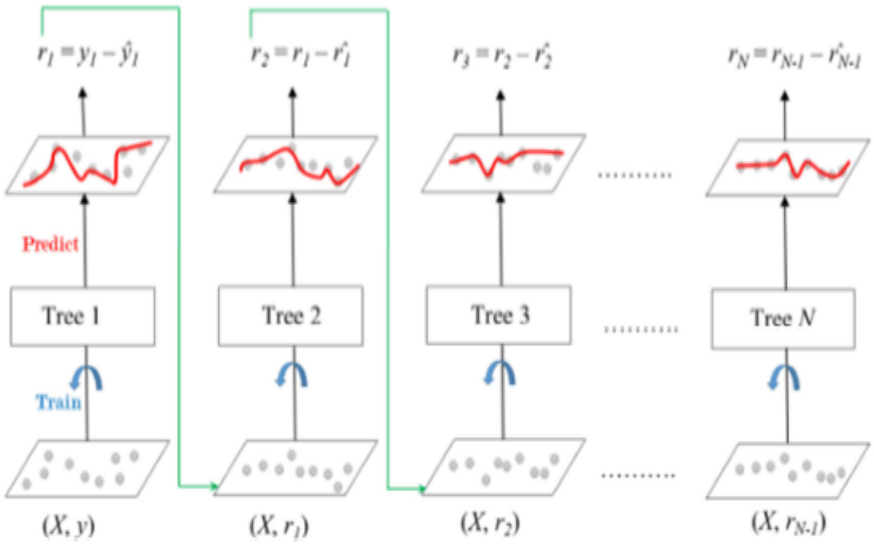

Figure 6. Diagram Gradient Boosting (Source [22]).

It is evident that the ensemble is comprised of N trees, as depicted in Figure 6. The feature matrix X and the labels y are utilised in the training process of Tree 1. The training set residual error r1 is calculated using the labelled predictions. Next, Tree 2 is trained by utilising the feature matrix X and employing the residual errors r1 from Tree 1 as the labels. The calculation of the residual error r2 involves the utilisation of predictive error, as illustrated in Figure 6.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

4. Result Analysis

4.1 Analysis of Heart Disease Dataset

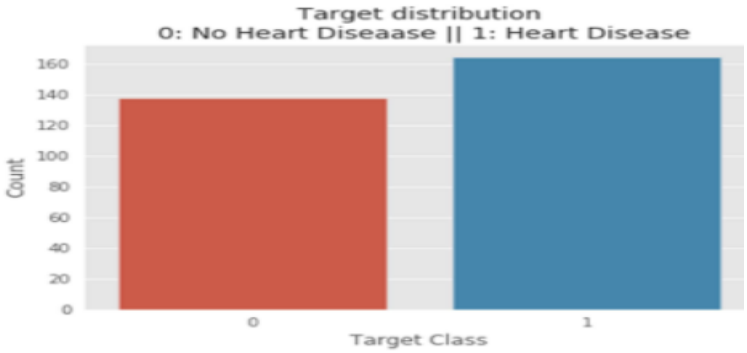

Figure 7. Target class.

Prior to delving into the examination of machine learning methods in this study, the analysis of the features of the heart disease dataset will be the primary focus. The total number of observations in the target attributes is 1025, with 499 instances indicating the absence of heart disease. The individuals in the study were classified into two groups: those without heart disease (0) and those with heart disease (1), as shown in Figure 7. The prevalence of individuals without heart disease is 45.7%, while the prevalence of individuals with heart disease is 54.3%, as depicted in Figure 8(a). The data demonstrates that the prevalence of heart disease is higher than the prevalence of those without heart disease.

The sex feature of the HD dataset is examined in Figure 8(b) by considering its relationship with the target feature. In terms of the sex attribute, there are 312 individuals identified as female and 713 individuals identified as male.

The number of males is more than twice the number of females. The data presented in Figure 8(b) illustrates a higher prevalence of heart disorders among males compared to females.

Likewise, the prevalence of cardiac disease in males does not exceed that in females. According to the findings in Figure 8(b), it can be concluded that males experience a higher degree of suffering compared to females. For further details, please refer to Figure 8(b).

ISSN PRINT 2319 1775 Online 2320 7876

Research paper

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

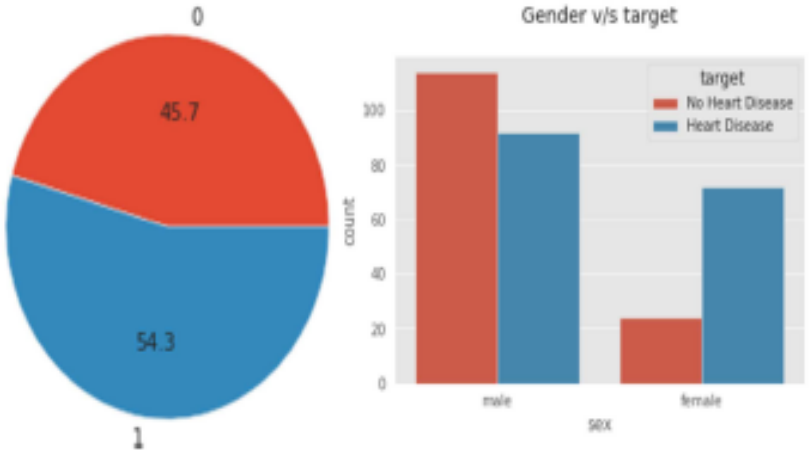

Figure 8. (a) Percentage of no heart disease and heart disease, (b) Comparison between sex and target feature.

The correlation between age and cholesterol is depicted in Figure 9(a). These characteristics from the dataset are considered at random for the experiment. Cholesterol levels between 200 and 300 mg/dl are associated with a reduced risk of cardiovascular disease in people aged 55 to 68. KDE figure 9(b) is analysed, and it's found to have very comparable statistical properties.

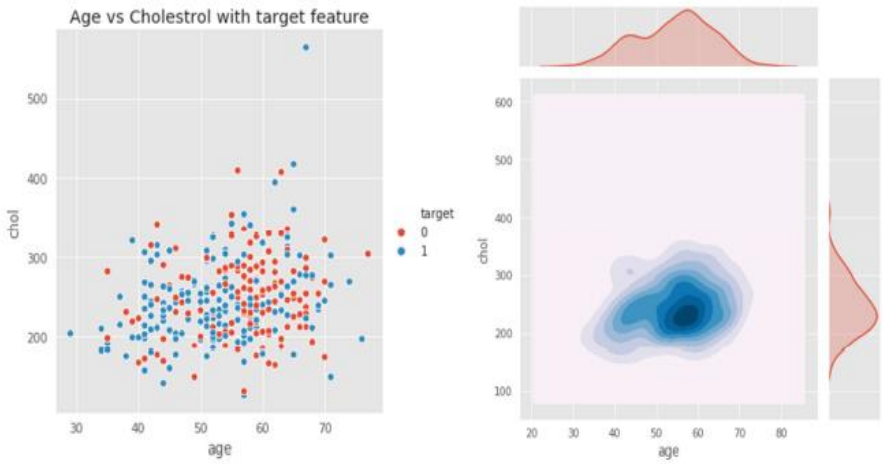

Figure 9. (a) Age v/s Cholesterol with the target feature, (b) Kernel density estimate (kde) plot of age v/s cholesterol.

Figure 10 depicts the association between the features. The correlation plot's primary use is to establish the presence or absence of a positive or negative relationship between the features. However, it presupposes that figuring out the strong and weak association using Figure 10 is difficult. This article includes Figure 11 to facilitate quickly obtaining these associations.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

Figure 11 shows that cp, thalamus, and slope all have favourable correlations with target attributes.

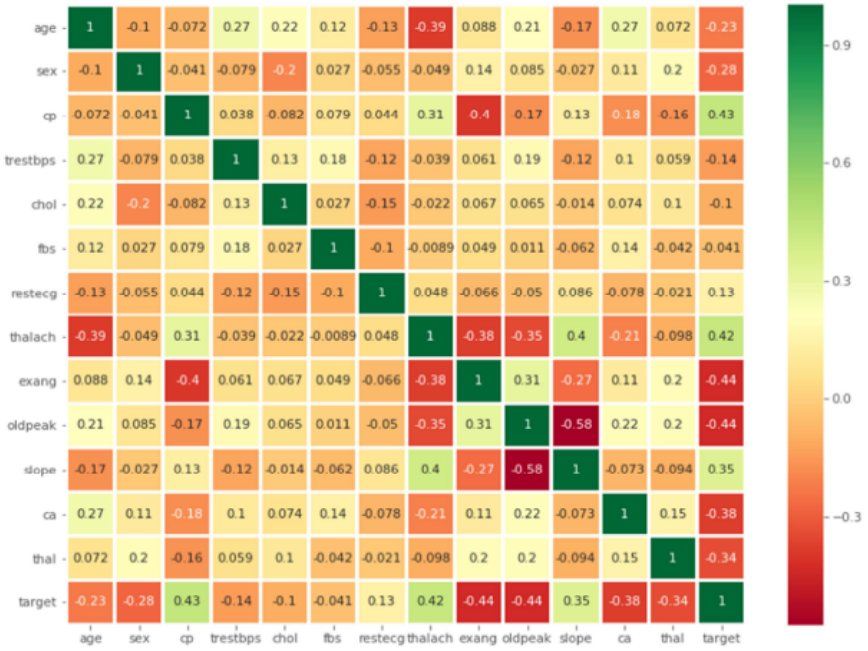

Figure 10. Correlation matrix of the attributes.

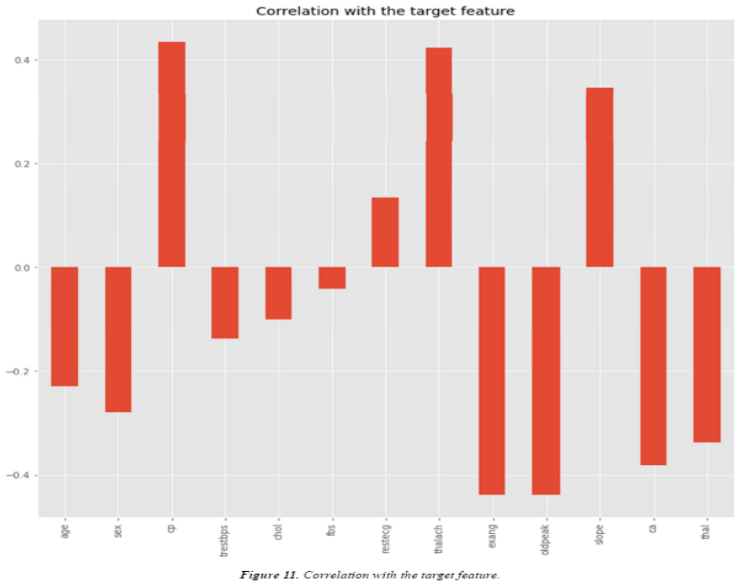

Figure 11. Correlation with the target feature.

The statistical significance of the two major correlations, cp and slope, with the desired trait is investigated. Figure 12(a) shows that hypertension does not occur when the cp is above 350, but that heart disease is more likely to persist between 200 and 250. Figure 12(a) also

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

demonstrates the absence of sickness when the slope is within the range 300 slope-1 350. On the other hand, 300 slope-2 350 is the heart disease threshold for slope-2.

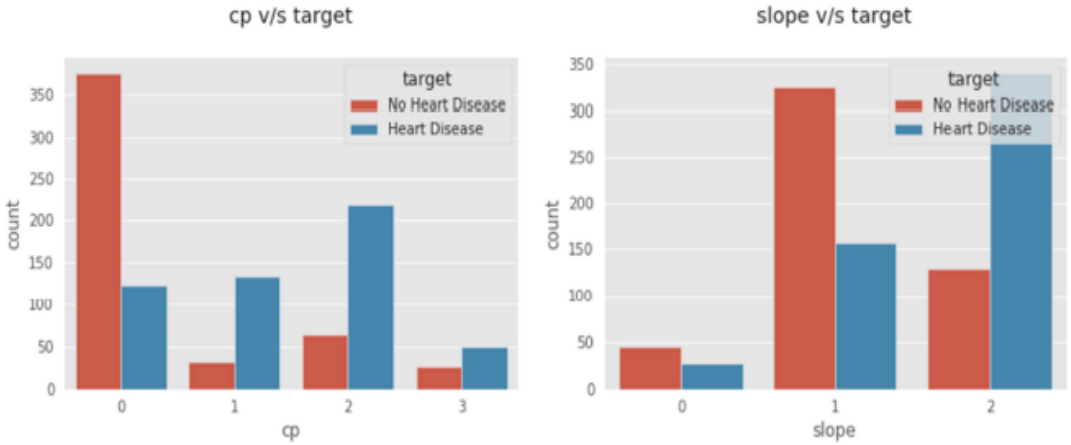

Figure 12. (a) cp v/s target, (b) slope v/s target.

4.2 Performance Analysis

This article discusses a range of machine learning algorithms, like as Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbours (KNN), Random Forest Classifier (RF), and Gradient Boosting Classifier (GBC) have been extensively investigated in the literature.

To forecast the occurrence of cardiovascular disease. The accuracy rate of each algorithm has been quantified, and the algorithm with the best accuracy has been chosen. The accuracy rate refers to the proportion of valid predictions in relation to the total number of datasets provided. The text can be reformulated in an academic manner.

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

Where, TP = 100% Authentic

What Does "True Negative" Mean?

F. P. = Fake Positive

Synonym: True Negative

Once the machine learning algorithms have been applied to the dataset for training and testing, the accuracy rate may be used to determine which algorithm performed better. A confusion matrix is used to determine the precision rate.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

Table 2 shows that when compared to other ML methods, Logistic Regression has the highest accuracy.

Table 2. Accuracy comparison of algorithms.

Algorithms	Accuracy
Logistic Regression (LR)	0.95
Support vector machine (SVM)	0.90
K-Nearest-Neighbors (KNN)	0.87
Random Forest Classifier (RF)	0.79
Gradient Boosting Classifier (GBC)	0.80

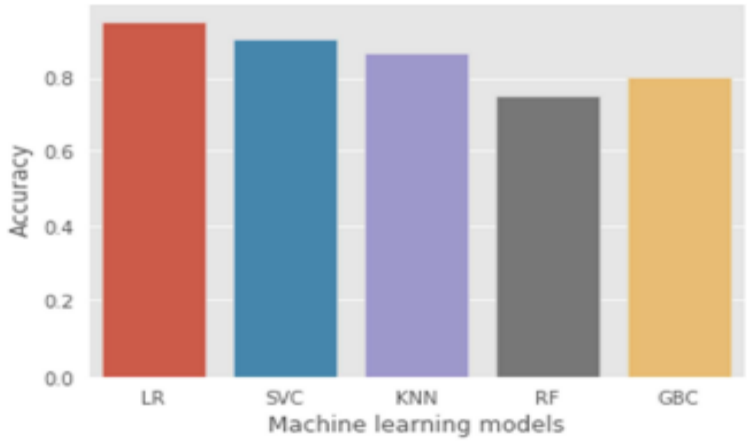

***Figure 13.** Accuracy comparison of machine learning algorithms by bar diagram.*

The LR machine learning framework has seen more research on this topic using a confusion matrix and a f1-score algorithm. According to the confusion matrix, the most likely outcome is Figure 14 depicts the 95% confidence interval for the f1-score calculation.

$$f_1 = 2 * \frac{1}{\frac{1}{P}+\frac{1}{R}} = \frac{2PR}{P+R}$$

Where precision

$$P = \frac{TP}{TP+FP}$$

And Recall

$$R = \frac{TP}{TP+FN}$$

ISSN PRINT 2319 1775 Online 2320 7876

Research paper

 UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

Figure 14. Confusion matrix of LR algorithm.

	precision	recall	f1-score	support
0	0.94	0.97	0.96	34
1	0.96	0.93	0.94	27
avg / total	0.95	0.95	0.95	61

Figure 15. f_1-score, precision, and recall of LR algorithm.

5. Conclusion and Future Scope

The heart is a crucial organ, but heart disease is becoming increasingly common, making it a global health crisis. If we have a model that can anticipate the onset of heart disease, we can treat it. Therefore, it is necessary to develop a machine learning model that can aid in the diagnosis of cardiac disease with more certainty and at a lower cost. It has the potential to be the first line of defence in diagnosing heart problems. This article examines the accuracy rate of the confusion matrix as a means of predicting cardiovascular disease.

In this vein, we use the provided algorithm statistics to evaluate the statistics across machine learning algorithms and estimate the accuracy rate of the confusion matrix. After evaluating five different algorithms, it was determined that the Logistic Regression algorithm provided the highest rate of accuracy. The Logistic Regression model has a 95% accuracy rate, suggesting that machine learning algorithms will soon be considered a standard tool for detecting cardiovascular illness. Logistic Regression also has a f1-score, recall, and precision

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

rate of 95%, 95%, and 95%, respectively. These predicted numbers are highly indicative of this algorithm's precision.

These results provide strong evidence that machine learning algorithms can be taught to make accurate disease forecasts. Our research might be expanded to include the detection of additional disorders. It's possible that we'll dig through data archives and integrate it with study methods enhanced by machine learning. This research has the potential for a wide range of future applications, including but not limited to: the prediction of cardiovascular disease, diabetes, breast cancer, tumours, and many diseases.

References

[1] [1] Wikipedia contributors. (2018, June 22). Machine learning. In Wikipedia, The Free Encyclopedia. Retrieved 06:31, June 26, 2002, from https://en.wikipedia.org/w/index.php?title=Machine_learning &oldid=1094363111.

[2] [2] Victor Chang, Vallabhanent Rupa Bhavani, Ariel Qianwen Xu, MA Hossain. An artificial intellegence model for heart disease detection using machine learning. Healthcare Analytics, volume 2, November 2002, 100016. https://doi.org/10.1016/j.health.2022.100016.

[3] [3] Ghumbre, S. U., & Ghatol, A. A. (2012). 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 (pp. 217-225). Springer, Berlin, Heidelberg.

[4] [4] Rohit Bharti, Aditya Khamparia, Mohammed Shabaz, Gaurav Dhiman, Sagar pande, and Parneet Singh. Prediction of Heart Disease Using a combination of Machine Learning and Deep learning. Hindawi Computational Intelligence and Neuroscience, Volume 2001, Article ID 8387680, 11 pages. https://doi.org/10.1155/2021/8387680.

[5] [5] Khaled Mohamed Almustafa. Prediction of heart disease and classifiers sensitivity analysis. Almustafa BMC Bioinfirmatics (2000) 21: 278. https://doi.org/10.1186/s12859-020-03626-y.

[6] [6] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014), A coronary heart disease prediction model. The Korean Heart Study. BMJ open, 4 (5), e005025.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 3, 2019

[7] [7] Mai Shouman, Tim Turner, and Rob Stocker. Applying kNearest Neighbour in diagnosis heart disease patients.. International Journal of Information and Education Technology, vol. 2, No. 3, June 2012.

[8] [8] Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Arnlov J & Ingelsson E (2013). Multilocus genetic risk scores for coronary heart disease prediction. Arteriosclerosis, thrombosis, and vascular biology, 33 (9), 2267-72.

[9] [9] Raihan M, Mondal S, More A, Sagor M O F, Sikder G, Majumder M A & Ghosh K (2016, December). Smartphone based ischeme heart disease (heart attact) risk prediction using clinical data and data mining approaches, a prototype design. 19th International conference on Computer and Information Technology (ICCIT) (pp. 299-303). IEEE.

[10] [10] Acharya U R, Fujita H, Oh S L, Hagiwara Y, Tan J H & Adam M (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Information Sciences, 415, 190-8.

[11] [11] Takci H (2018). Improvement of heart attack prediction by the feature selection methods. Turkish Journal of Electrical Engineering & Computer Sciences, 26 (1), 1-10.

[12] [12] Brown N, Young T, Gray D, Skene A M & Hampton J R (1997). Inpatient deaths from acute myocardial infarction, 1982-92: analysis of data in the Nottingham heart attack register, BMJ, 315 (7101), 159-64.

[13] [13] Soni J, Ansari U, Sharman D & Soni S (2011). Predictive data mining for medical diagnosis: an overview of heart disease prediction. International Journal of Computer Applications, 17 (8), 43-8.

[14] [14] Bashir S, Qamar U & Javed M Y (2014, November). An ensemble-based decision support framework for intelligent heart disease diagnosis. International Conference on Information Sociaty (i-Sociaty 2014) (pp. 259-64). IEEE.

[15] [15] Ordonez C (2006). Associate rule discovery with the train and test approach for heart disease prediction. IEEE Transaction on Information Technology in Biomedicine, 10 (2), 334-43.