Rai University AHMEDABAD
RISING - A Journal of Researchers
Bi-Annuals Peer Review Journal
Vol.9 Issue 1 (Jan 2025)
Impact Factor : 5.999
RISING A journal of researchers
ISSN No : 2456-7752

HEART STROKE PREDICTION USING MACHINE LEARNING: A COMPARATIVE ANALYSIS AND IMPLEMENTATION

1 Sagar Vakhare, 2 Arpit Chopra, 3 Ashutosh pandey, 4 Hemangee Sonara

1, 2, 3 & 4 Assistant Professor, Rai School of Engineering, Rai University, Ahmedabad

ABSTRACT:

One of the most challenging tasks in the medical field these days is predicting heart attacks. A heart stroke claims the life of one person every minute in the modern era. Processing vast amounts of data in the healthcare industry requires the use of data science. Given how difficult it is to predict heart attacks, the prediction process must be automated in order to minimize risks and notify patients well in advance. In this paper, the heart stroke dataset is used. The suggested work uses various data mining techniques, including SVM, Neural Network and Naïve Bayes to predict the likelihood of a heart attack and categorize the patient's risk level. This paper analyzes the performance of various machine learning algorithms in order to present a comparative study. When compared to other ML algorithms that were used, the trial results confirm that the Naïve Bayes algorithm had the highest accuracy, at 96.30 percent.

KEYWORDS: SVM, Neural Network, Naïve Bayes, Heart Stroke Prediction.

INTRODUCTION:

The primary focus of the work presented in this paper is on the different data mining techniques used in the prediction of heart attacks. A person's heart is the main organ in their body. Any heart condition has the potential to upset other body parts. Heart attacks and strokes are among the leading causes of death in the modern world. Over 10 million people worldwide lose their lives to heart attacks each year, according to the World Health Organization. Heart attacks can be caused by unhealthy lifestyle choices, smoking, drinking, and eating a lot of fat, which can raise blood pressure. The only ways to prevent heart stroke are to lead a healthy lifestyle and to detect the condition early. Today's healthcare system faces a major challenge in providing high-quality services and accurate diagnosis. The proposed work aims to identify heart strokes early on to prevent catastrophic outcomes. Techniques for data mining are ways to glean hidden and valuable information from the vast amount of data that is available. Machine Learning (ML), a branch of data mining, effectively manages large, well-formatted datasets. Numerous diseases can be diagnosed, detected, and predicted using machine learning in the medical field. In order to help physicians, treat patients effectively and prevent serious outcomes, the primary objective of this paper is to give them a tool for early detection of heart stroke. In order to uncover the hidden, machine learning is essential. In order to predict heart stroke early on, this paper presents performance analysis of several machine learning techniques, including SVM, Neural Network and Naïve Bayes.

LITERATURE REVIEW:

Machine learning techniques for predicting heart attacks have been the subject of numerous studies. In Govindarajan and associates. In this study, a machine learning classifier and text mining were used to classify heart stroke disease using data from 507 patients. With a 95 percent accuracy rate, the SGD approach produced the best results [3]. Amini et al. studied 807 people and categorized 50 risk factors for stroke, including diabetes, cardiovascular disease, smoking, hyperlipidemia and alcohol consumption. Using the c4.5 decision tree algorithm and the K-nearest neighbor algorithm, they achieved accuracies of 95% and 94%, respectively [4]. Cheng et al.1. estimated prognosis for myocardial infarction from data from 82 patients using two ANN models with precision rates of 79% and 95%. Cheon et al. focused on predicting the death of stroke patients, using deep neural networks on a cohort of 15,099 patients and achieving an AUC value of 83% [5]. Singh et. al. used artificial intelligence to predict heart attacks using a new method on the CHS dataset. Their neural network classification and

decision tree method produced a 97 percent model accuracy [6]. Chin et al. CNN-based system for automated early cardiac stroke detection achieved 90% accuracy after 256 images were used for model evaluation and training. Once these studies have been reviewed, the main idea behind the suggested system is to create an input-based heart attack prediction system. The Neural Network and Random Forest algorithms were analyzed in order to determine the best classification algorithm for heart disease prediction, taking into account the f-measure scores, accuracy, precision, and recall.

PROPOSED WORK:

In order to predict heart attacks, the suggested framework thoroughly analyzes the performance of the three classification techniques previously described. The main goal of this research is to get accurate estimates of a patient's risk of having a heart attack. The data from the patient's health report must be entered by the healthcare professional. This data is then incorporated into a predictive model that determines the patient's risk of having a heart attack. A complete workflow of the procedure is shown in Fig. 1. Offering a graphic depiction of the all-encompassing process.

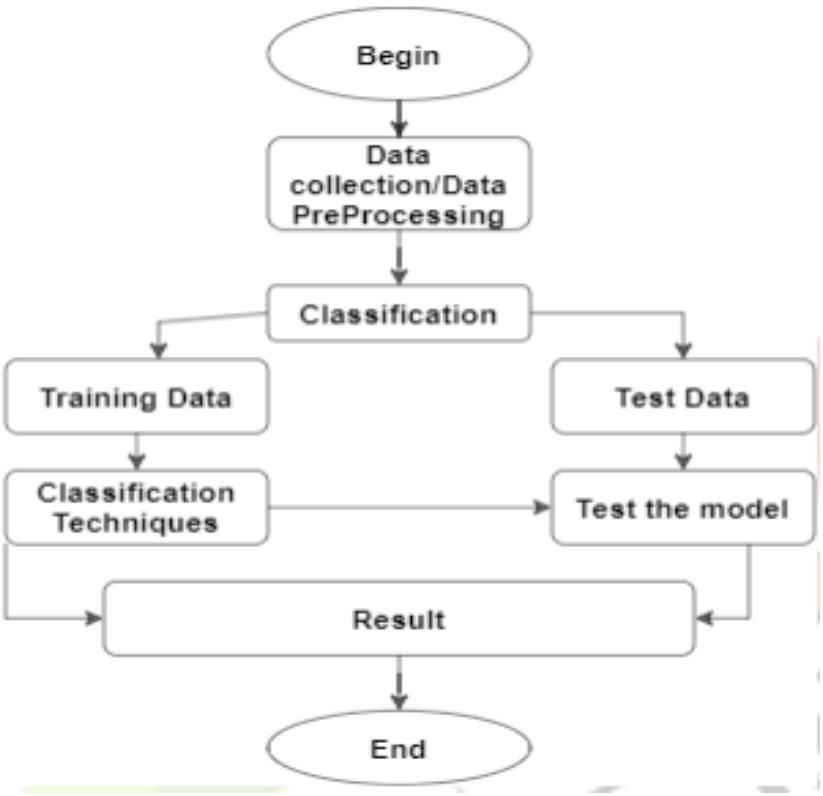

Fig.1: Generic Heart Stroke Prediction Model.

METHODOLOGY:

Machine Learning Classifiers and Data Description are the two primary parts of this section. The following are the detailed procedures.

A. Data Descriptions:

In this research, we used the heart stroke dataset available on the Kaggle website, which includes a total of 08 attributes. The characteristics used in this study are summarized below:

Age: This attribute indicates the age of a person. This is numerical data.

Gender: This attribute indicates the gender of a person. This is categorical data.

Heart Rate: This attribute indicates the frequency of the heartbeat, measured by the number of heart contractions per minute. This is numerical data.

Systolic Blood Pressure: This attribute indicates the maximum blood pressure during contraction of the ventricles. This is numerical data.

Diastolic Blood Pressure: This attribute indicates the minimum pressure measured immediately before the next contraction. This is numerical data.

Blood Sugar: This attribute indicates the main sugar found in your blood. This is numerical data.

Creatine kinase-MB (CK-MB): It is a form of an enzyme found primarily in heart muscle cells. This is numerical data.

Troponin: This attribute indicates the protein that's found in the cells of your heart muscle. This is numerical data.

Stroke: This attribute indicates whether a person has previously suffered a stroke or not. It is numerical data. The entire attribute, stroke is the decision class and the rest of the attributes are the response class.

B. Machine Learning Classifiers:

The mentioned attributes are provided as input to the various ML algorithms such as SVM, Neural Network and Naïve Bayes. The input data set is divided into 75% of the training data set and the remaining 25% into the test data set. A training data set is the data set used to train a model. The test data set is used to check the performance of the trained model. For each of the algorithms, performance is calculated and analyzed based on various metrics used such as

accuracy, precision, recall and F-measure scores as described below. The various algorithms examined in this research are listed below.

SVM: A Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression tasks. While it can be applied to regression problems, SVM is best suited for classification tasks. The primary objective of the SVM algorithm is to identify the optimal hyperplane in an N-dimensional space that can effectively separate data points into different classes in the feature space. The algorithm ensures that the margin between the closest points of different classes, known as support vectors, is maximized.

Neural Network: Represented by a central node denoting dataset property, the Neural Network algorithm provides results through its outer branches. Neural Network are preferred for their speed, reliability, simplicity, and minimal data preparation requirements. The class label prediction in a Neural Network is based on the root attribute's value, compared with the record's attribute, and navigating the corresponding branches.

Naïve Bayes: Naïve Bayes is part of a family of generative learning algorithms, meaning that it seeks to model the distribution of inputs of a given class or category. Unlike discriminative classifiers, like logistic regression, it does not learn which features are most important to differentiate between classes. n statistics, naive Bayes classifiers are a family of linear "probabilistic classifiers" which assumes that the features are conditionally independent, given the target class. The strength of this assumption is what gives the classifier its name. These classifiers are among the simplest Bayesian network models.

RESULTS AND DISCUSSION:

This section presents the results of applying the Random Forest and Neural Network algorithms. The performance of these algorithms is evaluated using the following metrics: Accuracy score, Precision (P), Recall (R), and F-measure. Precision, as defined in equation (1), provides the correct measure of positive analysis. Recall, mentioned in equation (2), represents the measure of actual positives that are correctly identified. The F-measure, shown in equation (3), assesses the precision of the algorithms.

Precision=(TP)/(TP+FP) …….. (1)

Recall=(TP)/(TP+FN) …….. (2)

F–Measure= (2*Precision*Recall)/(Precision+Recall)……(3)

- **TP True positive:** The test result indicates the presence of the condition, and the patient indeed has the illness.
- **FP False positive:** Despite the positive test result, the patient does not actually have the disease.
- **TN True negative:** The test is negative, and the patient does not have the condition.
- **FN False negative:** The test result is negative, but the patient is still afflicted with the illness.

The pre-processed dataset is utilized in the experiment to carry out the tests, and the techniques are examined and employed. The confusion matrix is used to derive the previously described performance indicators. The model's performance is characterized by the confusion matrix. The proposed model's confusion matrix for different methods is presented in Table 1 below. Table 2 below shows the accuracy score for the Random Forest and Neural Network classification algorithms.

Algorithm	Precision	Recall	F Measure	Accuracy
SVM	0.62	0.63	0.62	63.50%
Naïve Bayes	0.97	0.97	0.97	96.30%
Neural Network	0.80	0.78	0.78	78.40%

Table 1: Analysis of Different Machine Learning Algorithms.

Algorithm	True Positive	False Positive	False Negative	True Negative
SVM	52.00%	31.33%	48.00%	68.70
Naïve Bayes	98.3%	3.3%	1.7	96.7
Neural Network	63.33%	11.99%	32.70	88.10%

Table 2: Values obtained for Confusion Matrix of Different Algorithms.

CONCLUSION:

Creating a system that can accurately and efficiently forecast heart attacks has become essential due to the rising number of fatalities linked to heart strokes. The primary goal of this research was to determine the most effective machine learning (ML) algorithm for diagnosing heart strokes. Using the Kaggle dataset, this study thoroughly evaluated the accuracy rates of SVM, Neural Network and Naïve Bayes algorithms in predicting heart attacks. The results clearly indicate that the Naïve Bayes algorithm surpasses its peers, attaining an outstanding accuracy rate of 96.30% in predicting heart strokes.

REFERENCES:

1. Sayad AT, Halkarnikar PP. Diagnosis of heart disease using neural network approach.
2. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques.
3. P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, "Classification of heart stroke disease using machine learning algorithms," Neural Computing and Applications, vol. 32, no. 3, pp. 817– 828, Feb. 2020.
4. L. Amini, R. Azarpazhouh, M. T. Farzadfar, S. A. Mousavi, F. Jazaieri, F. Khorvash, R. Norouzi, and N. Toghianfar, "Prediction and control of heart stroke by data mining," International Journal of Preventive Medicine, vol. 4, no. Supply 2, pp. S245–249, May 2013. (Heart stroke prediction using machine learning).
5. Cheng W, Hüllermeier E. Combining instance-based learning and logistic regression for multilabel classification. Machine Learning.
6. Singh P., Kaur A., Batth R.S., Kaur S., Gianini G. Multi-disease big data analysis using beetle swarm optimization and an adaptive neuro-fuzzy inference system.