International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 02 Feb 2024, pp: 395-398 www.ijaem.net ISSN: 2395-5252

Heart Stroke Prediction using Machine Learning

S. Ganorkar[1], Hardik Choudhary[2], Nayan Baghel[3], Harshal Pardhi[4], Hitesh Meshram[5], Shampa[6]

[1]Assistant professor,[23456]BTech Students Department OfElectronics and Telecommunication
Smt. Radhikatai Pandav College of Engineering, Nagpur, Maharashtra, India

Date of Submission: 18-02-2024 Date of Acceptance: 29-02-2024

ABSTRACT: Anticipating heart attacks stands as a formidable challenge within the contemporary medical landscape, given the alarming frequency of heart strokes resulting in approximately one fatality per minute. The healthcare industry, grappling with vast datasets, necessitates the application of data science for effective processing. Automating the heart stroke prediction process becomes imperative, considering the complexity of mitigating risks and providing timely warnings to patients. In this study, the cardiac stroke dataset serves as the foundation, and various data mining techniques, Random Forest, Neural Network are employed to predict the likelihood of heart strokes and categorize patient risk levels. A comparative analysis of the efficacy of different machine learning algorithms is undertaken, revealing that the Random Forest approach outshines others with an outstanding accuracy rate of 99.17%. These findings underscore the potential of advanced algorithms in revolutionizing heart attack prediction, paving the way for enhanced patient care and risk management in the realm of cardiovascular health.

Index Terms -Random Forest, Neural Network, Heart Stroke Prediction

I. INTRODUCTION

At the core of the human anatomy lies the heart, and any deviations from its regular rhythm can potentially amplify discomfort in our bodily regions. In the contemporary world, heart attacks stand as a prominent contributor to global mortality rates. Unhealthy practices such as smoking, excessive alcohol consumption, and the intake of high-fat diets are all recognized as precursors to heart attacks and hypertension. According to estimates from the World Health Organization, over 10 million people worldwide succumb to heart attacks annually. The mitigation of this dire situation hinges upon adopting a healthy lifestyle and instituting early detection measures to avert the onset of heart attacks. Present-day healthcare systems grapple with significant challenges in delivering top-tier services and achieving precise, reliable diagnoses [1].

Nowadays it is the major issue to face these medical challenges situation. To solve this problem machine learning algorithm can be very useful. Machine learning, particularly through methods like Decision Trees, and Random Forest, plays a pivotal role in uncovering discrete patterns and conducting thorough analyses of the provided data. By employing these ML approaches, this study scrutinizes their effectiveness in early prediction of cardiac strokes [2]. By using this machine learning approaches cardiac stroke can be predict earlier

II. LITERATURE SURVEY

Numerous studies have explored machine learning approaches for anticipating cardiac attacks. In Govindarajan et al.' research, data from 507 patients were utilized to classify heart stroke disease using a combination of text mining and a machine learning classifier. The SGD method yielded the highest accuracy at 95% [3]. Amini et al. studied 807 individuals, categorizing 50 risk factors for stroke, including diabetes, cardiovascular disease, smoking, hyperlipidaemia, and alcohol consumption. Employing the c4.5 decision tree algorithm and K-nearest neighbour, they achieved accuracies of 95% and 94%, respectively [4]. Cheng et a1. estimated heart stroke prognosis using data from 82 patients, employing two ANN models with precision rates of 79% and 95%. Cheon et al. focused on forecasting stroke patient death, utilizing deep neural networks on a cohort of 15,099 patients and achieving an AUC value of 83% [5]. Singh et al. applied artificial intelligence for cardiac stroke prediction, utilizing a novel technique on the CHS dataset. Their decision tree approach and neural network

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 02 Feb 2024, pp: 395-398 www.ijaem.net ISSN: 2395-5252

classification achieved a model accuracy of 97% [6]. Chin et al. automated early cardiac stroke detection with a CNN-based system, achieving a 90% accuracy after training and evaluating the model with 256 images. Upon reviewing these studies, the core concept for the proposed system is to develop a heart attack prediction system based on inputs. To identify the optimal classification algorithm for heart disease prediction, an analysis was conducted on Neural Network, and Random Forest algorithms, considering Accuracy, Precision, Recall, and f-measure scores .

III. PROPOSED WORK

The proposed framework scrutinizes the effectiveness of the four classification methods detailed earlier, conducting a thorough performance analysis for the prediction of heart attacks. The primary objective of this study is to achieve precise predictions regarding a patient's susceptibility to a heart attack. The healthcare provider is responsible for inputting the patient's health report data, which is then assimilated into a predictive model designed to assess the likelihood of experiencing a heart attack. The entire procedural workflow is illustrated in Fig. 1, providing a visual representation of the comprehensive process.

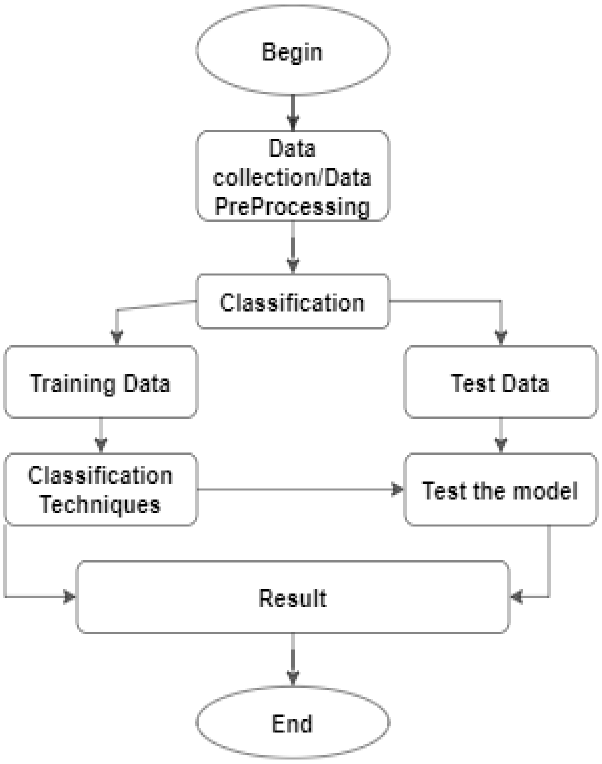

Fig1: Generic Heart Stroke Prediction Model

IV. METHODOLOGY:

This section comprises two main components: Data Description and Machine Learning Classifiers. The detailed procedures are as follows:

A) Data Description:

In this investigation, we utilized the cardiac stroke dataset available on the Kaggle website, encompassing a total of 12 attributes. The characteristics used in this study are summarized below:

1. id: Person's ID represented by numerical values.
2. Age: Person's age represented by numerical values.
3. Gender: Person's gender with categorical data.
4. Hypertension: Indicates whether the person has hypertension, represented by numerical values.
5. Work Type: Represents the person's work environment with categorical data.
6. Residence Type: Shows the individual's living situation with categorical data.
7. heart disease: Indicates whether the person has heart disease, represented by numerical values.

8. Avg Glucose Level: Indicates a person's glucose level with numerical data.
9. BMI: Represents a person's body mass index with numerical data.
10. Ever Married: Denotes a person's marital status with categorical data.
11. Smoking Status: Represents a person's smoking history with categorical data.
12.Stroke: Indicates whether a person has ever had a stroke, represented by numerical values. The 'Stroke' attribute serves as the answer class, and the remaining attributes contribute to the choice class.

B) Machine Learning Classifiers:

Attributes are fed into various ML algorithms, including Random Forest, Decision Tree, and KNN. The input dataset is divided into 80% training data and 20% test data. The training dataset is used to train the model, while the testing dataset evaluates the model's performance. Metrics such as accuracy, precision, recall, and F-measure are employed to analyse the performance of each algorithm. The investigated algorithms are as follows:

i. Random Forest: Utilized for both regression and classification, this method organizes data into a tree, making predictions based on the tree. Random Forest is robust even with a substantial amount of missing data, and it involves two stages: generating a random forest and making predictions using the produced classifier.

ii. 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.

V. RESULTS AND DISCUSSION

This section displays the outcomes of applying Random Forest, Neural Network. The metrics Accuracy score, Precision (P), Recall (R), and F-measure are used to analyses the algorithm's performance. The correct measure of positive analysis is provided by the precision metric (stated in equation (1)). The measure of actual positives that are right is defined by recall, which is mentioned in equation (2). F-measure (shown in equation (3)) evaluates precision.

Precision=(TP)/(TP+FP) (1)
Recall=(TP)/(TP+FN) (2)
F–Measure= (2*Precision*Recall)/(Precision+Recall) (3)

- TPTruepositive:The test is positive, and the patient has the illness.
- FPFalsepositive:Although the test is positive, the patient does not have the disease.
- TNTruenegative:The test is negative, and the patient does not have the condition.
- FNFalsenegative:Despite the test coming back negative, the patient has the illness.

The pre-processed dataset is used in the experiment to conduct the tests, and the techniques are investigated and used. The confusion matrix is used to derive the performance indicators previously described. The model's performance is described by the confusion matrix. The proposed model's confusion matrix for various methods is displayed in Table 1 below. Table 2 below displays the accuracy score for the Random Forest, Neural Network classification algorithms.

TableI CONFUSION MATRIX VALUES OBTAINED USING DIFFERENT ALGORITHM

Algorithm	TruePositive	FalsePositive	FalseNegative	TrueNegative
Neural Network	915	44	0	500
RandomForest	947	12	0	500

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 02 Feb 2024, pp: 395-398 www.ijaem.net ISSN: 2395-5252

TableII ANALYSIS OF VARIOUS MACHINE LEARNING METHODS

Algorithm	Precision	Recall	F-measure	Accuracy
Neural Network	0.97	0.97	0.97	96.25%
RandomForest	0.99	0.99	0.99	99.17%

VI. CONCLUSION

Establishing a system capable of accurately and efficiently predicting heart attacks has become imperative, given the escalating fatalities associated with heart strokes. The paramount objective of this research was to identify the most effective machine learning (ML) algorithm for heart stroke diagnosis. Employing the Kaggle dataset, this study meticulously assessed the accuracy scores of the Random Forest, Neural Network algorithms in predicting heart attacks. The findings unequivocally demonstrate that the Random Forest algorithm outperforms its counterparts, achieving an impressive accuracy score of 99.17% for heart attack prediction.

While these results present a significant advancement, the study's potential for enhancement lies in future endeavours. A promising avenue involves the development of a web application grounded in the Random Forest method, coupled with the utilization of an even more extensive dataset than employed in this analysis. This strategic refinement is poised to yield superior outcomes, providing healthcare professionals with a robust tool for precise and efficient cardiac disease forecasting. As technology evolves, embracing these advancements could contribute substantially to proactive healthcare measures and further elevate the efficacy of heart attack prediction systems.

ACKNOWLEDGMENT

We would like to thank Prof. Shital Ganorkar, for her guidance through each stage of the process.

REFERNCES

[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 learningand logistic regression for multilabel classification. MachineLearning

[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.