JOURNAL OF BASIC SCIENCE AND ENGINEERING

ACCURACY ASSESSMENT OF MACHINE LEARNINGALGORITHMFOR PREDICTION OF HEART STROKE

Dr.R.Venkatesh, Mithun.V ,HariVenkatesh.M , Hari Prathiunun.R and Kishore.S
Asso.Prof, Dept. of Computer Science and Engineering, Karpagam College of Engineering, Myleripalayam, Coimbatore-641 032, Tamil Nadu, India
Venkatesh@kce.ac.in

Abstract: This paper conducts a relative analysis of machine learning algorithms applied inthe prediction of heart strokes. It is necessary to automate the heart stroke prediction procedure because it is a hard task to reduce risks and warn the patient well in advance. The cardiac stroke dataset is used in this work. The suggested work uses various approaches including Decision Tree , Random Decision Forest , Extreme Gradient Boosting, and Hybrid model to forecast the likelihood of Heart Stroke and categories patient risklevel. These algorithms are crucial in automating the prediction process, leading to enhanced patient outcomes and greater efficiency in healthcare delivery. The core of our investigation revolves around assessing the execution of these algorithms in respect of predictive precision, sensitivity, specificity, and other relevant grade. Our methodology includes meticulous data preprocessing, model training, and rigorous evaluation using standard healthcare performance metrics. The results of our testing reveal that, among the various machine learning algorithms considered, the RF approach exhibits the greatest level of accuracy. This suggests that Random Forest has the potential to be a valuable tool in early heart stroke prediction and risk stratification. The testing outcomes indicate that among the ML algorithms utilized, the RF approach demonstrates the highest accuracy..

Keywords:Decision Tree, Hybrid Model, Machine Learning Algorithms, Classification, Accuracy, Comparitive Analysis, XGBoost, Supervised Machine Learning Algorithms

1. Introduction

Heart stroke prediction is a crucial area of research in healthcare, as it offers the potential to enhance patient care and reduce the risks associated with cardiacevents.[1] Given the intricate nature of heart stroke prediction, automating this process is imperative to provide early warnings to patients and healthcare providers. In this research, we undertake an extensive comparative analysis of different ML algorithms to predict the probability of heart strokes and classify patient risk levels. To accomplish this, we leverage the cardiac stroke dataset, a valuable resource for this critical research.

Our study incorporates multiple machine learning techniques, such as Classification Tree, Random Decision Forest, and Extreme Gradient Boosting, and a novel Hybrid figure. The primary goal on our probing is to determine which on these algorithms exhibits the highest accuracy in predicting heart strokes and identifying patient risk levels.

The results of our testing reveal that, among the machine learning algorithms considered, the Random Decision Forest approach consistently outperforms the others, showcasing its superior predictive capabilities in this context. This finding underscores the potential of Random Decision Forest as a valuable tool in improving heart stroke prediction and patient risk assessment.[5]

Heart stroke prediction is a multifaceted process that encompasses several essential steps. Initially, pertinent data, including medical history, lifestyle factors, and demographics, is collected through surveys, medical records, and physical examinations. This information forms the basis for assessing problems, such as hypertension, sugar problem, and family origins, each assigned a weight according to significance.

To ready the data for study, thorough preprocessing is necessary, involving purification, handling missing values, and normalizing variables. Feature selection aids in narrowing down the most relevant variables for the predictive model, improving its efficiency. Choosing the appropriate machine learning algorithm, like Classification Trees, Random Decision Forest, or Extreme Gradient Boosting, depends on the dataset and problem complexity.[3]

Once chosen, these models undergo training using actual information to establish correlations between proposed variables and the likelihood of heart strokes. Thorough model validation ensures their ability to make true prognosis on imaginary data, with metrics like efficiency, precision, recall, and ROC curves employed for assessment.

The core of the prediction process lies in risk assessment, where each patient is assigned a risk score or category. This risk level informs personalized recommendations, which may include lifestyle changes, medication, or further diagnostic tests. Continuous monitoring and feedback are essential for refining preventive measures over time.

The efficacy of heart stroke prediction models hinges on data quality, algorithm selection, and rigorous validation. Ongoing research endeavors to enhance these models, ultimately aiming to reduce the burden of heart disease on public health and save lives.

2. Benefits of Heart Stroke prediction:

1. Early Intervention: Heart stroke prediction allows for early identification of individuals at risk. This enables healthcare providers to intervene and initiate preventive measures well in advance, potentially preventing the occurrence of a stroke or reducing its severity.

2. Improved Patient Outcomes: Early detection and timely intervention can significantly improve patient outcomes. Patients at risk can receive appropriate treatment and care, leading to a higher chance of recovery and a better quality of life after a stroke.

3. Reduced Mortality: Predicting heart strokes helps reduce mortality rates. Quick identification of high-risk patients can lead to faster medical attention and life-saving interventions.

4. Cost Savings: Early prediction and intervention can lead to cost savings in healthcare. Preventing strokes or minimizing their impact can reduce the financial burden on both patients and healthcare systems.

5. Resource Allocation: Healthcare resources, including medical personnel and facilities, can be allocated more efficiently when at-risk individuals are identified in advance. This prevents overburdening healthcare systems during stroke emergencies.

6. Personalized Care: Prediction models can help tailor treatment and care plans to the individual patient's risk level. This leads to more personalized and effective care.

7. Risk Factor Management: Heart stroke prediction models often take into account various risk factors, such as hypertension, diabetes, and smoking. Identifying these risk factors can motivate patients to make lifestyle changes and manage their health more effectively.

8. Research and Development: Data generated from heart stroke prediction studies can be used for further research and development of medical interventions and treatments. This contributes to advancements in stroke prevention and management.

9. Public Health Initiatives: Heart stroke prediction can inform public health

initiatives aimed at educating the public about stroke risk factors and prevention strategies.

10. Healthcare Planning: Predictive models assist in long-term healthcare planning. Understanding the prevalence and distribution of stroke risk in a population can guide public health policies and healthcare resource allocation.

3.Proposed Work:

After examining every tactic now in use, multiple researchers described the numerous advantages of each suggested methodology and remarked on a number of constraints that are still tied to real-world strategies and have a big influence on how well the tactics work. Rigidity, which makes creating a simulation time consuming, alternate parameters, inaccurate assessments, and other associated issues are some of the major challenges.

Data Gathering: Gather and assemble pertinent data on people with heart disease, such as demographic data, lifestyle choices, medical histories, and test findings. Data Pre-processing involves cleaning up the data, managing outliers, and transforming information that is categorical into numerical data in order to make it ready for analysis.

Variable Selection: Using variable selection approaches like Recursive Feature Elimination (RFE), like Feature Importance, choose the data set's most pertinent characteristics for prediction.

By contrasting the reliability, precision, recollection, and F1-score of various methods, such as logit model, classification trees, random decision forests, and support vector networks, it is possible to choose the most suitable categorization models for predicting coronary artery disease.

Model Ensemble: Create a hybrid model by combining the predictions of many models using methods like bagging, boosting, or stacking. provides a more accurate and robust prediction. Model Validation: Using cross-validation methods like crossvalidation using k-folds or stratified k-fold cross-validation, validate the efficacy of the hybrid model.

Model Deployment: To allow healthcare practitioners to utilise the completed model for real-time cardiac disease prediction, deploy it in a mobile or website app. Based on the outcomes of the model selection, approval, and deployment processes, this plan may be improved and optimised. To forecast cardiac disease and enhance patient outcomes, a useful and effective tool has to be made available.

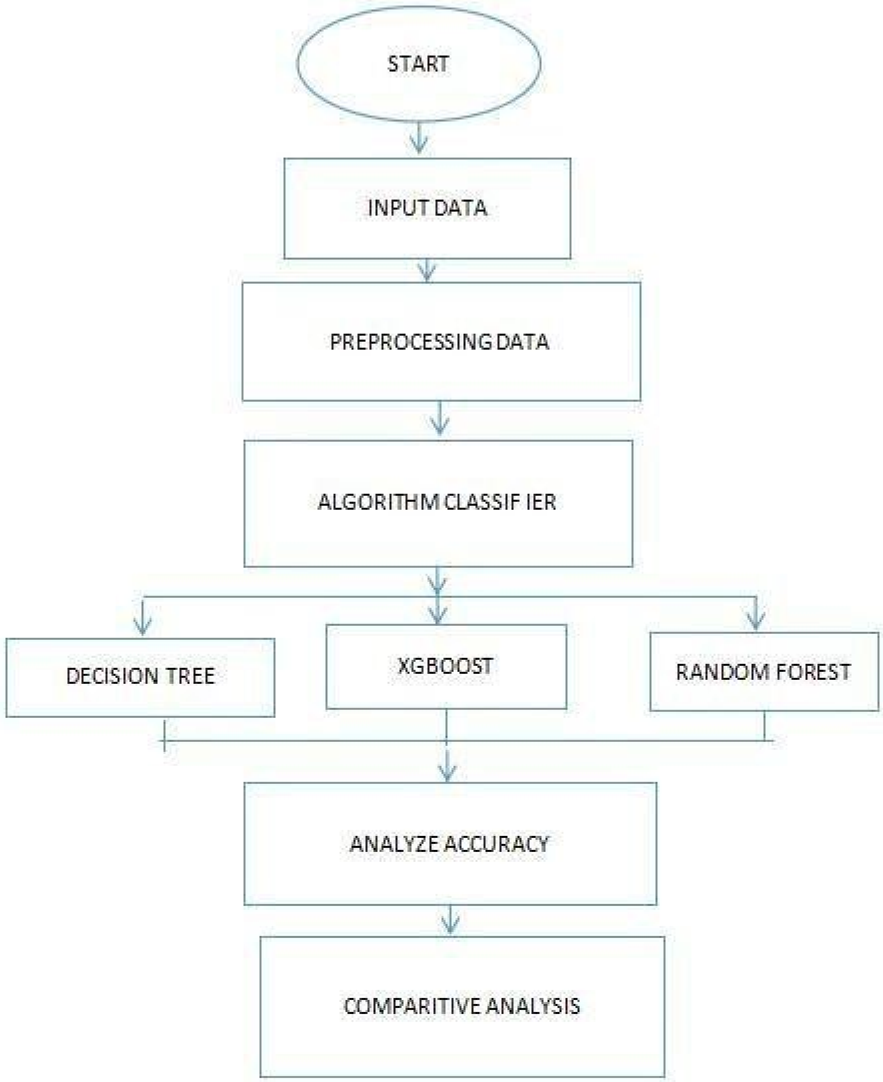

Figure 1.SYSTEM FLOW DIAGRAM

4.Results and Discussion:

<u>STEPS TO EXECUTE THE PROJECT</u>

1. Install the necessary packages

2. Define the problem statement

3. Develop a User Interface based on Django.

4. Connect a Comma-separated values(CSV) file

5. Display figures

6. Choose a exemplary

7. Visualize Graph

8. Generate predictions.

5. ALGORITHMS

5.1 DECISION TREE:

A Classification tree is a flowchart representation that provides a fair pathway to a decision. In the realm of data analytics, it's an algorithm incorporating conditional control statements for data classification. Starting from a single point or node, it diverges into two or more directions, providing different outcomes depending on decisions and chance events until reaching a conclusive result. Decision trees are useful for dissecting intricate data into manageable components, frequently employed for predictive analysis, data categorization, and regression.

Advantages of Classification trees:

- Effective for visually interpreting data.
- Capable of handling both Quantitative and qualitative data.
- Simple to define protocols such as 'yes, no, if, then, else...'
- Need token data for training or purification.
- Useful for assessing perfect, bad, and likely case plot.
- Easily combinable with other decision-making techniques.

Detriment of classification trees:

- Prone to overfitting if overly complex.
- Not suitable for continuous variables.
- Computational complexity increases with many chance variables.
- Vulnerable to bias with imbalanced datasets.

5.2 RANDOM DECISION FOREST:

Random Decision forest is a Flexible, Easy to use machine learning algorithm known for producing excellent outputs without extensive hyper-parameter tuning. It's widely used for both classification and regression tasks. Operating as a Guided learning algorithm, it builds an ensemble of decision trees typically skilled using the bagging method, where a combination of learning models enhances overall performance.

5.3 XGBOOST:

XGBoost stands out as one of the foremost popular and efficient ML algorithms. It enhances the gradient boosting framework by introducing extreme gradient boosting, which incorporates both linear model solver and tree learning algorithms. XGBoost's exceptional speed stems from its ability to perform simultaneous computation on a single machine, Enabling speeds of at least ten times faster than current implementations, it accommodates a range of utility functions, including regression, Categorization, and grading.

Benifits of XGBoost:

- Built-in regularization with L1 and L2 to prevent overfitting.
- Utilizes multiprocessing for faster computation.
- Handles missing values effectively.
- Allows cross-validation at each boosting iteration.
- Implements effective tree pruning to enhance performance.

6.Conclusion:

The objectives of the work completed for this investigation are to improve QoS, appropriateness, and effectiveness. In order to create a more successful strategy, the literature review examined the drawbacks of current strategies. Four different algorithms, which include Random Forest, the algorithm the XGBoost including a decision treevariation (J48), are investigated in the proposed study. These algorithms are instrumental in automating the prediction process, leading to enhanced client outcomes and increased efficiency in medical care delivery.The core of our investigation revolves around assessing the efficiency of these algorithms in terms of predictive precision, sensitivity, specificity, and other relevant criteria. Our methodology includes meticulous data preprocessing, model training, and rigorous evaluation using standard healthcare performance metrics. The results of our testing reveal that, among the various ML algorithms considered, the Random Decision Forest approach exhibits the highest accuracy. This suggests that Random Forest has the potential to be a valuable tool in early heart stroke prediction and risk stratification. The testing results show that, when compared to other ML algorithms used, the Random Forest approach has the best accuracy

References:

[1] A. K. Garate-Escamilla, A. H. E. Hassani and Andres,"Classification models for heart disease prediction usingfeatureselection and pca," Informatics Medicine Unlocked, vol. 19, Article ID 100330, 2020.

[2] A predictive analytics approach for stroke prediction using machine learning and neural network soumyddbrata Dev a,b , Hewei Wang c,d , Chidozie Shamrock Nwosu , Nishtha Jain , Bharadwaj Veeravalli , Deepu John Healthcare Analytics 2 (2022) 100032.

[3] Analyzing the Performance of Stroke Prediction using ML Classification Algorithms Gangavarapu Sailasya1 , Gorli L Aruna Kumari2 (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 6, 2021.

[4] C. Ricciardi, A. S. Valente, K. Edmund et al., “Linear discriminant analysis and principal component analysis topredict coronary artery disease, ” Health Informatics Journal,vol. 26, no. 3, pp. 2181–2192, 2020.

[5] D Balakrishnand, TD Rajkumart, S Dhanasekaran, “ An intelligent and secured heart rate monitoring system using IOT ” Materials Today: Proceedings.

[6] D. C. Yadav and S. Pal,“Prediction of heart disease using feature selection and random forest ensemble method, ” International Journal for Pharmaceutical Research Scholars,vol. 12, pp. 56–66, 2020.

[7] Fahd Saleh Alotaibi: Implementation of Machine Learning Model to Predict Heart Failure Disease. In: International Journal of Advanced Computer Science and Applications (IJACSA) (2019).

[8] Jeena R.S and Dr.Sukesh Kumar “Stroke prediction using SVM”, International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, 2016.

[9] Menaria H.K., Nagar P., Patel M. (2020) Tweet Sentiment Classification by Semantic and Frequency Base Features Using Hybrid Classifier. In: Luhach A., Kosa J., Poonia R., Gao XZ., Singh D. (eds) First International Conference on Sus tainable Technologies for Computational Intelligence. Advances in Intelligent Systems and Computing, vol 1045. Springer, Singapore. https://doi.org/10.1007/978- 981-15-0029-9_9.

[10] Nwosu, C.S., Dev, S., Bhardwaj, P., Veeravalli, B., John, D.: Predicting stroke from electronic health records. In: 41st Annual International Conference of the IEEE Engineerin