International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

An Early Prediction of Cardiovascular Stroke Using Machine Learning Techniques

M Nisheetha[1], L. Vijaya Kumar[2], N Prathyusha[3], Prof. M. E.Palanivel[4], K Fizza Gulshan[5]

[1,3,5]CSE-Artificial Intelligence, Sreenivasa Institute of Technology and Management Studies in Chittoor India
[2]Assistant Professor, CSE-Artificial Intelligence Sreenivasa Institute of Technology and Management Studies in Chittoor, India
[4]Professor, CSE-Artificial Intelligence, Sreenivasa Institute of Technology and Management Studies in Chittoor, India

Abstract
The main concept of this project is to predict Cardiovascular stoke. Now a days heart stroke has become a major health issue in most of the families. So early detection can reduce the death rates for heart strokes. Early detection is the essential to reduce its impact. Here for early detection Machine Learning algorithms are used. This Machine Learning techniques give an effective and cost-effective solution for prediction of heart disease. To predict cardiovascular stroke models are trained using machine learning algorithms like SVM, KNN and, Random Forest. And here Django framework is used to visualize the predicted results of the models and manage the data storage.

Keywords: Cardiovascular, Support Vector Machine (SVM), K-Nearest Neighbor (KNN).

I. Introduction

Cardiovascular stroke causes many deaths, as it makes major worldwide health concern. Many hospital cases resulting in death before even reaching out the hospital. There are very less survival rates in the hospital cases. So, for earlier prediction of cardiovascular stroke Machine Learning is used here. In previous system Decision tree and Navie bayes Algorithms are used but there are few drawbacks in those techniques like results are not accurate, it cannot handle large amount of datasets of patents records, it is hard for decision making. So to overcome those drawbacks we introduced these methods in our project, we have many advantages in these methods like Fast prediction, cost-effective.

II. Literature Review

Title:Reasearch of cardio vascular stroke Prediction Based on Machine Learning techniques
Author: Shuge Ouyang introduced in the year 2022. Summary: A paper that explores machine learning techniques that are used for cardio vascular stroke prediction.
Title:Machine Learning Models for Survival and Neurological Ouutcome Prediction of Out-of-Hospital cardio stroke patients
Authors: Chi-Yung Cheng, I-Min Chiu & Wun-huei Zeng, Chih-Min Tsai and Richard Lin introduced

in the year 2021. Summary: Where cardiac arrest (heart disease) is the major health issue faced in the society, and the neurologic injury remains the leading high or major reason for the illness and cause for the number of deaths in the society. So here we are using machine Learning algorithms to see whether it could be detect dependencies between the clinical variables in emergency departments. Here in this three machine learning algorithms are applied. The algorithms are support vector machine [SVM], logistic regression, and extreme gradient boosting. The first outcome will be taken as the neurological outcome and the second outcome will be a 30 day survival rate and survival-to discharge rate.

A. ROLE OF MACHINE LEARNING

In this project Machine Learning is used. Here Machine Learning is used to train the models by giving the input data which is the details of the patients. By using Machine Learning algorithms it predicts whether there is a chance of heart stroke are not. Machine Leaning plays a crucial role in this project where it helps in the Early detection of the heart stroke. It helps the doctors to identity the patients who have high risk of heart stroke before it takes place and make their work easier. It is cost effective as well. And it saves the time of the doctors. By using machine learning the models get trained and make them to predict the stroke and then the results will be compared to the historical data. So that it will be easy to predict and get the results as soon as possible. Here it is done by training the models from historical patients data like heart disease, age, BMI, blood pressure, FBS etc. After model gets trained it predicts if a person can get heart stroke by finding its pattern. Machine learning gives accurate predictions. And it makes Better Medical decisions and gives recommendations for lifestyles changes and for the health.

B. CONTRIBUTIONS

To improve stroke prediction here we use Machine Learning techniques. The main contribution of this project is, in the existing system of this project Decision tree, Naïve Bayes algorithms are used, but in this present proposed system here it choses the better methods than the old ones. It improves the performance of the model and gives better accuracy. It handles Large datasets of the patients with many details like heart condition, age, etc. It helps in taking better decision and clearly shows the heart stroke if it is there.

III. RELATED WORK

In this system model anaconda and python are used. Where input data is stored in the data sets. The input data consists of 16 attributes like patients ID, age, FBS, CP, ECG… The attributes are stored as the input data. Then it is divided into the clusters where obtaining the samples one with the chances of heart stroke present and another with no heart stroke present. Once testing and training is done, the samples will be merged and we get the final output. In the system model manual entry is done where users like doctors and patients enter the data through a form based interface.

There is automated data import where it integrates with the electronic health record system the data retrieval. It allows the bulk amount of data entry from the structured dataset. Here Graphical Representations are done where heat maps are used to visualize the risk factors of the cardiovascular stroke. And ROC curves are used to assess the models performance. Text based reports are given which is summaries for the stroke risk level. It gives the recommendations for the lifestyle changes or further medical tests if it is needed.

A. Modules

The main thing of the system is predicting the cardiovascular stroke by using the Machine Learning techniquies is

- Collects the patient health details where the data is from the manual user input or sources such as publicly available datasets (i.e., Kaggle).
- Cleans , removes unwanted data and prepares the data which is collected, for the analysis for prediction of cardiovascular stroke.
- By using imputation techniques, it handles the missing values from the dataset.
- It converts categorical variables into numerical values.
- Implements and trains multiple machine learning models.
- KNN (K-Nearest Neighbour) patients results will based on the similarity to previous (past) cases.
- SVM (support vector machine) finds a proper decision (outcome) for the classification.
- Random Forest uses multiple decision trees for more robust predictions.
- Splits the data into two categories i.e. testing and training sets to increase model accuracy
- Takes a new patient data as an input source and applies it to the trained machine learning models to predict result for the stroke risk.
- It generates a probability score indicating the likelihood of stroke occurrence for the paitent.
- Ensures compatibility with the hospital databases for the seamless integration.

B. Frontend

- When the Frontend code is implemented, the code will be executed using anaconda prompt.
- Commands will be typed using Anaconda prompt, where it takes to a website such that final output is viewed.
- Once the commands are typed, the browser asks for the user to login.
- Username and password must be given.
- After the login is done, give the patient ID that you wish to know the results.
- Select the algorithm that you need from KNN, SVM, Random Forest
- Then the final result will be shown.
- The final results will be compared from three algorithms which are used here.
- Then the algorithm with highest accuracy will be selected.

C. Backend

- When the Backend code is implemented, the code will be executed using Jupyter Notebook.
- Before implementing, the patient data must be collected and stored in database.
- Dataset is said to be done.
- The data which is stored in the dataset must be sent for the analysis.
- Clustering process will take place.
- The samples will be separated whether there is a chance of heart stroke or not.
- The patient with chance of heart stroke will undergo the process of clustering for purpose of training and testing.
- The patients with no chance of heart stroke will undergo Undersampling.
- After testing and training is done both samples will be merged with each other.
- This process undergoes Oversampling to obtain the final results.
- Once the oversampling is done then the results are obtained for the backend code.

IV. METHODOLOGY

A. Existing Method

Here in the existing system, two Machine Learning Algorithms are used Decision Tree and Naïve Bayes to predict the cardiovascular stroke.

Decision Tree is utilised for the classification and for the regression problems. The main benefit of this decision tree is it can able to handle the numerical and the categorical data. By Learning the decisions which have taken from the past data it creates a training model to make predictions.

Naïve Bayes is also very helpful for both the regression and for the classification problems. Naïve Bayes is used here to calculate the probability of a class based on historical knowledge. Even with limited data this algorithm works really well and it is computationally efficient. When the new is given it predicts the class with the highest probability by learning the features of each class from the training data.

However, existing system has few drawbacks in it. It often shows low accuracy, which can increase the chances of death by showing it incorrectly. It can lead to Higher the chances of depression rate and lower the use of health care services. Here in this system they used risk scoring system, but they are not always accurate in detecting serious cases.

B. Proposed Method

Here in project we proposed method there are three advanced Machine Learning algorithms are used. They are K-Nearest Neighbor, Support Vector Machine, Random Forest.

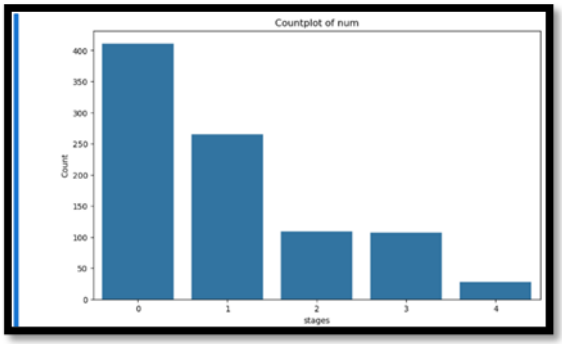
KNN algorithm works for both the classification and the regression Problems. It checks for the similarities in the new data by comparing it with the past or existing data. And it makes the predictions based on the nearby points which means the new data which matches to the existing data. SVM Is mostly used for the classification Problems. This SVM increases dimensional data. It finds the best hyperplane that will separate the classes and it uses support vectors to define it. SVM performs better in linear and the non-linear tasks.

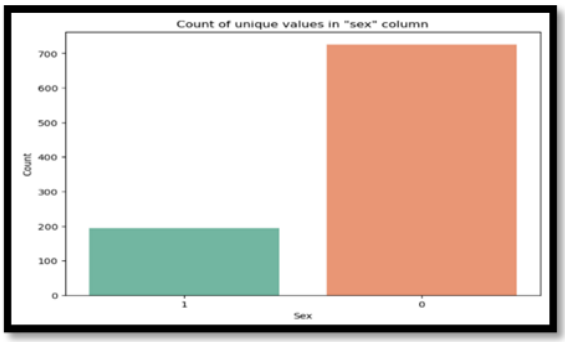
Random Forest is made with the help of many decision trees. Every tree works on with various subset of the data. The final prediction accuracy in the Random Forest is done based on the majority votes of a decision from all the trees. In this if there are more trees then it leads to better accuracy and prevents overfitting.

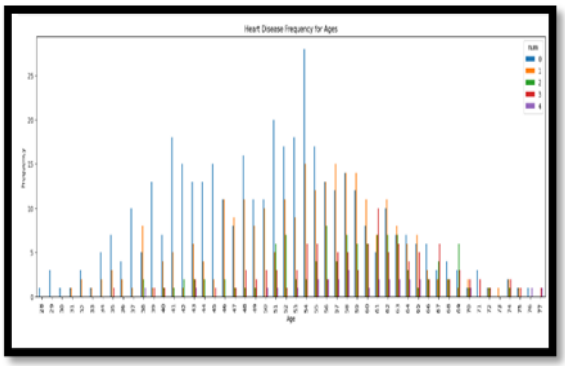
The proposed system increases the accuracy of cardiovascular stroke prediction, and it saves time for doctors. It is highly affordable for the patients. It can handle large dataset of patient details. It handles large dataset better by using Random Forest and feature selection.

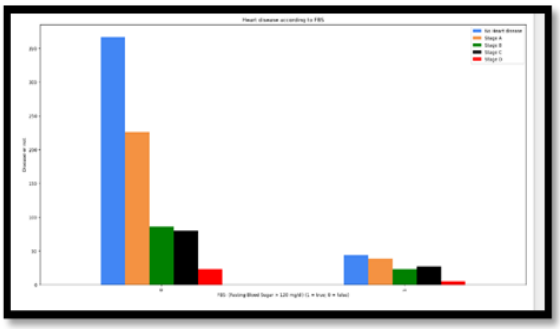
V.RESULTS AND DISCUSSION

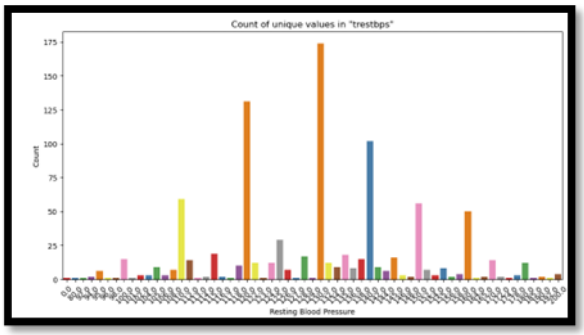
The proposed system in machine learning gives the better results than the existing system. In proposed system algorithms like K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest have shown higher accuracy in predicting the cardiovascular stroke. Among those three algorithms Random Forest have given the best performance because of its ability of storing large amount of data and it avoids overfitting. Even KNN worked well by finding the most similar cases from the training data. SVM was effective in separating the classes clearly even with the complex data. Overall, The proposed system improved in predicting the accuracy. It reduces the time of diagnosis. It helps the doctors which makes their work easier and make them to take most accurate decisions. This system makes the early prediction easier, which can save the lives and it is cost effective.











VI. CONCLUSION

In this project we use three algorithms of Machine Learning which is KNN, SVM, Random Forest. These three algorithms are used both for the classification and for the regression. By using this we can predict the accuracy level of the cardiovascular stroke. Out of these three algorithms we select the algorithm which gives the best accuracy for describing the results. ROC curves and Heatmaps are used for these methods. Here classifiers are applied to detect the stroke by integrating the static and dynamic features. When a much larger dataset is available then RF could be applied, which would increase the accuracy and the precision of the outcomes.

REFERENCES

1. J. M. Kwon, Y. Lee, S. Lee, and J. Park, “An Algorithm Based on Deep Learning for Predicting In – Hospital Heart disease,” Journal of the American Heart Association, vol. 7, no. 13, e008678, 2018.
2. L. J. Blackhall, “Must we always use CPR,” New England Journal Medicine, vol. 317, no. 20, pp. 1281-1285, 1987.
3. W. B. Kouwenhoven, J. R. Jude, and G. G. Knickerbocker, “Closedchest cardiac massage,” JAMA, vol. 173, no. 10, pp. 1064-1067, 1960.

4. D. Kindermann, R. Mutter, and J. M. Pines, "Emergency Department Transfers to acute Care Facilities," HCUP Statistical Brief #157, 2013.
5. C. J. Ng, Z. S. Yen, J. C. H. Tsai, L. C. Chen, S. J. Lin, Y. Y. Sang, and J. C. Chen, "Validation of the Taiwan triage and acuity scale: a new computerised five-level triage system," EmergMed J, vol. 28, no. 12, pp. 1026-1031, 2011.
6. X. Guo, Y. Yin, C. Dong, G. Zhou, "On the class imbalance problem," in Natural Computation, 2008. ICNC'08. Fourth International Conference on (vol. 4, pp. 192-210). IEEE.
7. S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Handling imbalanced datasets: A review," GESTS International Transactions on Computer Science and Engineering, vol. 30, no. 1, pp. 25-36, 2006.
8. Y. Sun, A. K. Wong, and M. S. Kamel, "Classification of imbalanced data: A review," International Journal of Pattern Recognition and Artificial Intelligence, vol. 23, no. 04, pp 687-719, 2009.