MUTHAYAMMAL ENGINEERING COLLEGE
(An Autonomous Institution)
Approved by AICTE, New Delhi | Affiliated to Anna University, Chennai
Rasipuram - 637 408. Namakkal Dt., Tamil Nadu
Estd. 2000

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Predicting Stroke Risk with Machine Learning and a Web based Early Intervention Tool

First Review

Batch Members
Surya A (20CS167)
Venkatesan J(20CS172)
Vikram C (20CS177)

Guided by
Dr.T.Aravind M.E, Ph.D.
Assistant Professor.

Abstract

- Stroke is a dangerous medical disorder that occurs when blood flow to the brain is disrupted.
- Stroke is a medical emergency that occurs when a section of the brain's blood supply is cut off.
- It is a big worldwide threat with serious health and economic implications.
- The number of people at risk for stroke is growing as the population ages, making precise and critical.
- The brain cells die when they are deprived of the oxygen and glucose needed for their survival.

Introduction

- The rising global incidence of stroke emphasizes the urgency for early intervention to prevent long-term disability and mortality.
- Traditional methods are time-consuming and error-prone, underscoring the need for advanced, efficient solutions.
- Machine learning algorithms demonstrate remarkable accuracy in predicting stroke risk by analyzing various clinical factor.
- Interpretable machine learning models offer clinicians valuable insights into factors contributing to stroke risk.

3

Introduction

- Recently, Machine learning algorithms have shown great promise in accurately predicting stroke risk based on various risk factors.
- Additionally there is a growing need for transparency and explainability in machine learning models in healthcare.
- The World Stroke Organisation estimates that 13 million people worldwide experience a stroke each year, leading to 5.5 million fatalities.
- Stroke can impact both men and women, lowering their quality of life and putting a load on public health resources.

4

Literature Survey

Title of the paper	Description	Publication details
A systematic review of machine learning models for predicting outcomes of stroke with structured data	The complexity of a condition such as stroke potentially lends itself well to the use of ML methods which are able to incorporate a large variety of variables and observations into one predictive framework without the need for preprogrammed rules	PLoS ONE, vol. 15, no. 6 Jun. 2020
Explainable artificial intelligence model for stroke prediction using EEG signal	EEG data from the stroke subjects were collected in the active state no later than three months after the diagnosis of ischemic stroke.	Sensors, vol. 22, no. 24, p. 9859, Dec. 2022

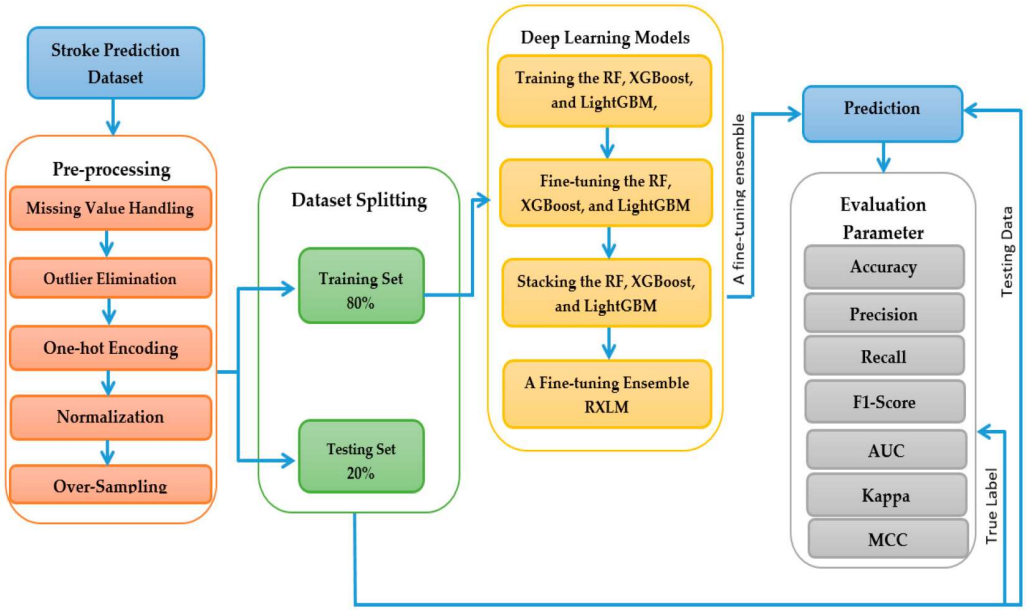
5

Literature Survey

Title of the paper	Description	Publication details
Different medical data mining approaches based prediction of ischemic stroke	10-fold cross validation resampling method was utilized, and model performance evaluation metrics were accuracy, area under ROC curve (AUC), sensitivity, specificity, positive predictive value and negative predictive value	Comput. Methods Programs Biomedvol. 130, pp. 87–92, Jul. 2016
Machine learning for brain stroke: A review	The aim of this work is to classify state-of-arts on ML techniques for brain stroke into 4 categories based on their functionalities or similarity, and then review studies of each category systematically	J. Stroke Cerebrovascular Diseases, vol. 29, no. 10, Oct. 2020

6

Block Diagram



7

Existing System

- Machine learning is increasingly used in medical diagnostics, particularly in stroke prediction
- The main objectives are to increase diagnostic precision and classification efficiency
- Traditional methods are time-consuming and prone to errors, leading to delayed intervention and worsened patient outcomes
- Naive Bayes has the lowest accuracy scores among all the algorithms, with 76.03% accuracy.

8

Existing System

- Several approaches have been explored for automatic stroke prediction.
- Common techniques include logistic regression, decision trees, support vector machines, and neural networks.
- Feature selection plays a crucial role,with variables like age,blood pressure,cholesterol levels and lifestyle factors being significant.
- Dataset quality is essential, and researches often use medical records or specific stroke databases.

9

Existing System

- For the implementation of the stacking model, four base classifiers were combined.
- More specifically, naive Bayes, random forest, J48 and RepTree were selected.
- As for the majority voting,we considered the same models with the stacking method, except for naive Bayes.
- The minimum number of instances per leaf node was set to the default value.

10

Proposed System

- The study aims to develop a reliable machine learning model for predicting stroke disease, addressing class imbalance issues
- It uses Mutual Information Score, Chi-Square Score, and ANOVA(Analysis of Variance) tests to identify key features contributing to stroke risk
- The technique is compared with six classifiers and demonstrated its effectiveness in generalization and prediction accuracy
- The study uses SHAP and LIME explainable techniques to understand the decision-making process

11

Proposed System

- **Naive Bayes algorithms** are mostly used in sentiment analysis, spam filtering, recommendation systems etc.
- In most of the real-life cases, the predictors are dependent; this hinders the performance of the classifier.
- **Artificial Neural Network**- Neural networks are a set of algorithms, modelled loosely after the human brain, that are designed to recognize patterns.
- **Decision Tree-** A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, resource costs, and utility

12

Proposed System

- **Data pre-processing -**is necessary prior to model construction in order to eliminate a dataset's undesirable noise and outliers.
- This phase deals with all the issues that keep the model from operating more effectively.
- Data must be cleansed and processed for model development after the pertinent dataset has been collected.
- The column id is firstly ignored because its inclusion has no impact on model creation.

13

Software Components

- Operating System : Windows 11
- Platform : Python
- Tool : Spyder, Python 3.6
- Front End : Anaconda
- Back End : Python Anaconda Script

14

Hardware Components

- System : I5 1.3ghz
- Hard Disk : 100 GB Minimum
- Monitor : 15 Inch VGA Color
- Mouse : Logitech Mouse
- Ram : 512 Mb
- Keyboard : Standard Keyboard

15

References

[1] T. Elloker and A. J. Rhoda, "The relationship between social support and participation in stroke: A systematic review," *Afr. J. Disability*, vol. 7, pp. 1–9, Oct. 2018.

[2] M. Katan and A. Luft, "Global burden of stroke," Seminar Neurol., vol. 38, no. 2, pp. 208–211, Apr. 2018.

[3] I. Mosley, M. Nicol, G. Donnan, I. Patrick, and H. Dewey, "Stroke symp- toms and the decision to call for an ambulance," Stroke, vol. 38, no. 2, pp. 361–366, Feb. 2007.

[4] A. K. Boehme, C. Esenwa, and M. S. V. Elkind, "Stroke risk factors, genet- ics, and prevention," Circ. Res., vol. 120, no. 3, pp. 472–495, Feb. 2018.

[5] L. Gibson and W. Whiteley, "The differential diagnosis of suspected stroke: A systematic review," J. Roy. College Physicians Edinburgh, vol. 43, no. 2, pp. 114–118, Jun. 2013.

16

THANK YOU