International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 09 | Sept - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Stroke Prognosis Using ML Classification Algorithms

Dr. Gnanapriya S[1] , Dinesh Kumar V[2]

[1]Assistant Professor (SG), Department of Computer Applications, Nehru College of Management, Coimbatore, Tamilnadu, India.

[2]II MCA student, Department of Computer Applications, Nehru college of Management, Coimbatore, Tamilnadu, India.

1. ABSTRACT:

Stroke is one of the worst diseases on the planet, responsible for many fatalities both directly and indirectly. Numerous data mining techniques are used by the healthcare industry to aid in early disease identification and diagnosis. Many stroke risk factors are examined in the body of current research. We start by examining the characteristics of people who have strokes more frequently than others. The dataset is sourced from a publically available source, and various classification techniques are used to predict the likelihood of a stroke. Extreme Gradient Boosting was used to achieve accurate percentage. Reducing bias and enhancing model performance by applying a range of tactics, such as Random oversampling, a method that duplicates examples of the minority class to balance the distribution of classes.

2. INTRODUCTION:

Because of the internet's tremendous expansion and use in recent years, there is an enormous amount of data available on it in many different forms. Alongside these developments in technology, there has been an increase in the use of this data for the investigation of other applications. Alongside these developments in technology, there has been an increase in the use of this data for the investigation of other applications. Big data is a processed collection of several types of data that is rich in information. Machine learning (ML), a branch of artificial intelligence, is the process of teaching a computer using a labelled dataset. The machine then makes predictions about the future using this learned dataset. The project is gathered and comprises two datasets: the fruit dataset and the vegetable dataset. There are six [6] different categories in the fruit data collection, where two datasets for fruits are categorized as healthy, one for diseases impacted, and one for vegetables. There are six [6] classifications for the fruit data collection, which includes apple, corn, and peach datasets that are divided into healthy and disease-affected categories.

3. LITRATURE REVIEW:

Gangavarapu Sailasya and Gorli L. Aruna Kumari [1] the proposed system is called Analyzing the Performance of Stroke Prediction Using ML Classification Algorithms. Using a range of physiological parameters and machine learning methods, this study trained five different models for successful prediction: Naïve Bayes classification, K-Nearest Neighbors, Support Vector Machine, Decision Tree classification, and Logistic Regression. In this

task, the Naïve Bayes algorithm did better than the others, with an accuracy of about 82%. Sharmin Akter Tanna, Taslima Mostafa Alif Ema, Iamil Shoily, Sumaiya Jannat, Tajul Islam, and Tasfia Romana Rahman [2] Identification of Stroke Illness The suggested system was called Using Machine Learning Algorithms. Four machine learning algorithms—Naive Bayes, J48, k-NN, and Random Forest—were used in this study to determine the type of stroke that might have happened based on information from the patient's medical record and physical attributes. To solve their problem, they make use of a substantial number of entries that they have collected from the hospitals.The outcome is suitable for use in a real-time medical report, as shown by the classification result. They proposed that machine learning algorithms could help improve understanding of diseases and be a helpful tool in the healthcare industry. The Performance Analysis of Machine Learning Approaches in Stroke System was presented by Shamim Al Mamun, M. Shamim Kaiser, Md. Mahfujur Rahman, Tamara Islam Meghla, Minhaz Uddin Emon, and Maria Sultana Keya [3].

4. METHODOLOGY:

- **Data Collection**:

Describe the dataset used (e.g., hospital records, publicly available databases).

- **Preprocessing**:

Detail the steps taken to clean and prepare the data (handling missing values, feature selection, normalization).

- **Modeling Techniques**:

Discuss the machine learning models used (e.g., logistic regression, SVM, decision trees, RNN) and justify your choices.

- **Evaluation Metrics**:

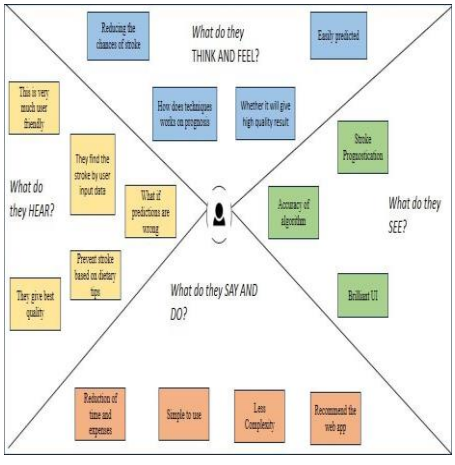
Define the metrics used to evaluate model performance (e.g., accuracy, AUC-ROC, F1 score).

5. EXISTING SYSTEM:

Three research papers provide a thorough analysis of machine learning techniques for stroke detection and prediction in the past. A set of techniques, including Decision Tree Classification, K-Nearest Neighbors, Support Vector Machine, Naive Bayes Classification, and Logistic Regression, were carefully used in the first study publication. Notably, the Naive Bayes Classification method produced an excellent 82% accuracy rate. However, the possible effect of unbalanced datasets on the models' performance was a major shortcoming noted in this work. This is an important factor to take into account when using these algorithms in real-world healthcare environments where the distribution of stroke cases may not be uniform.

6. PROPOSED SYSTEM:

Worldwide, stroke ranks as the third most common cause of death. When cerebral blood flow is disturbed, a serious disease known as cerebral haemorrhage occurs, which often leads to severe illness, incapacity, or even death. Predicting strokes accurately is crucial to lowering the death rate from strokes and to delivering timely treatment. Our research offers a machine learning-based technique to enhance stroke diagnosis, particularly in situations with uneven data distributions. To address this imbalance, we apply the RandomoverSampling(ROS)technique.



The performance of five distinct classifiers is assessed in this study: Random Forest classification, K-nearest Neighbor, Decision Tree classification, Support Vector Machine, and XGBoost. To maximize the prediction accuracy score for each model, a rigorous cross-validation and hypermeter tweaking procedure is applied. Especially, our findings illustrate the exceptional precision reached

7. DATASET DESCRIPTION:

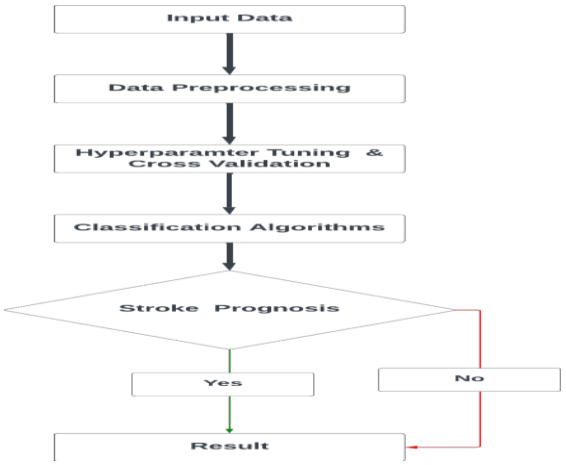
Stroke Dataset:

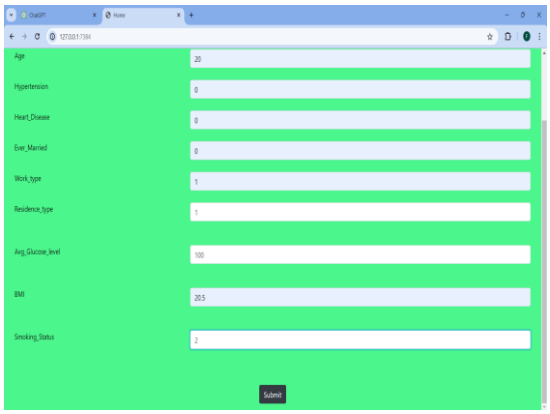
This is the primary data set that we make use of in our project. It contains 11 classes of Data. This shall be various aspects of individual's health and demographics that we are going to train and test for building our model called **xg.sav.** The 11 classes of individual Data that we are going to use for training and testing the model are:

1. Gender
2. Age
3. Hypertension
4. Heart_disease12
5. Ever married
6. Work type
7. Residence type
8. Average_glucose_level
9. BMI
10. Smoking status
11. Stroke

This datasets is employed to perform prognostication, specifically to determine whether an
Individual is at risk of experiencing a stroke or not.

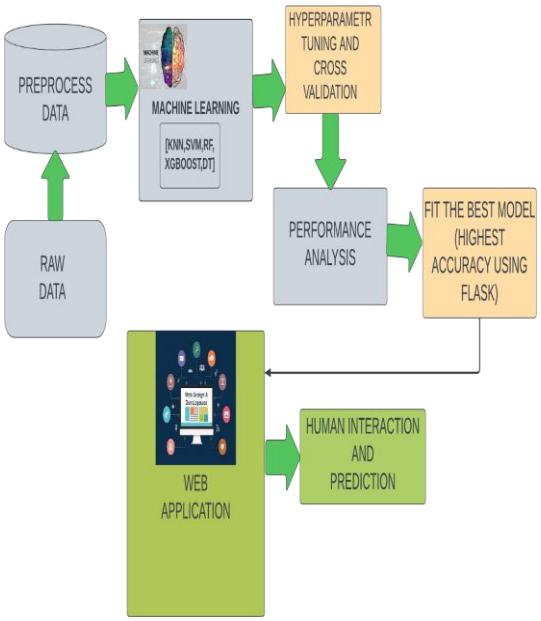
8. DATA FLOW DIAGRAMS:





Data flow diagrams are used to graphically represent the flow of data in a business information system. DFD describes the processes that are involved in a system to transfer data from the input to the file storage and reports generation. Data flow diagrams can be divided into logical and physical. The logical data flow diagram describes flow of data through a system to perform certain functionality of a business. The physical data flow diagram describes the implementation of the logical data flow.

9. ARCHITECTURE DESIGN:



10. RESULTS:

- Present the results of your models, including performance metrics.
- Use tables and figures to summarize key findings.
- Discuss the significance of your results in relation to your research questions.

11. DISCUSSION:

- Interpret your results in the context of the existing literature.
- Discuss the implications for clinical practice and future research.
- Acknowledge limitations of your study (e.g., sample size, data quality).

12. CONCLUSION:

There are several key reasons why machine learning-based stroke prediction is significant in the medical field. First off, it is critical to bettering patient outcomes by assisting medical practitioners in identifying individuals who are at higher risk of experiencing a stroke. Through the assessment of patients' vital signs, medical history, and relevant data, machine learning models are able to identify risk factors and early warning signals in patients. This makes quick intervention possible. Individualized therapy and early discovery can significantly increase a patient's chances of survival and recovery, ultimately leading to an improvement in their overall quality of life. Moreover, stroke prognostication based on machine learning reduces healthcare costs.

Long-term damage is largely caused by strokes, and paying for long-term care, treatment, and rehabilitation can be expensive for both patients and healthcare systems.

13. REFERENCE:

· **Kwan, J., & Hand, P. J. (2010).**

- · Title: "Predicting the outcome of patients with acute stroke using a machine learning algorithm."
- Journal: BMC Medical Informatics and Decision Making.
- DOI: 10.1186/1472-6947-10-43

· **Gong, Y., & Zhang, Y. (2021).**

- · Title: "A novel machine learning approach for stroke risk prediction: A nationwide cohort study."
- Journal: Journal of Stroke and Cerebrovascular Diseases.
- DOI: 10.1016/j.jstrokecerebrovasdis.2020.104745

.

Alonso-Fernández, F., et al. (2021).

- · Title: "Machine learning methods for predicting outcomes in stroke: A systematic review."

- Journal: Neurology.
- DOI: 10.1212/WNL.0000000000000547

· **Hussain, M., et al. (2022).**

- · Title: "Predicting the severity of stroke outcomes using ensemble machine learning techniques."
- Journal: Frontiers in Neurology.
- DOI: 10.3389/fneur.2022.752844

· **Hwang, H. S., & Yoon, S. H. (2019).**

- · Title: "Deep learning for stroke prognosis prediction: A systematic review."
- Journal: Journal of Healthcare Engineering.
- DOI: 10.1155/2019/9564791

Cheng, Y., et al. (2020).

- · Title: "Stroke outcome prediction using machine learning: A systematic review."
- Journal: International Journal of Medical Informatics.
- DOI: 10.1016/j.ijmedinf.2020.104234

· **Khatri, P., & Kothari, S. (2018).**

- · Title: "Machine learning for prediction of stroke outcomes: A systematic review."
- Journal: Stroke.
- DOI: 10.1161/STROKEAHA.117.018989