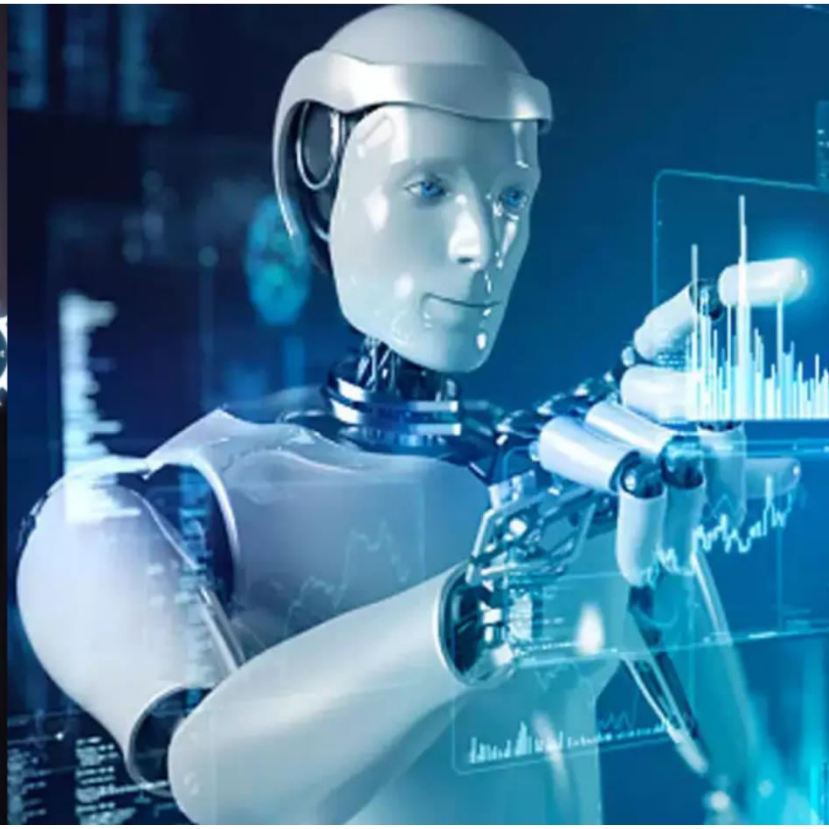
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 3, March 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403027|

Brain Stroke Prediction Portal Using Machine Learning

M.Jeevanandha, Kaverimuthu M, R Malkkar, R Joyalrelton

Assistant Professor, Department of Computer Science and Design, Sethu Institute of Technology, Virudhunagar District, Tamil Nadu, India

Department of Computer Science and Design, Sethu Institute of Technology, Virudhunagar District, Tamil Nadu, India

Department of Computer Science and Design, Sethu Institute of Technology, Virudhunagar District, Tamil Nadu, India

Department of Computer Science and Design, Sethu Institute of Technology, Virudhunagar District, Tamil Nadu, India

ABSTRACT: A stroke occurs when the blood supply to part of your brain is interrupted or reduced, preventing brain tissue from getting oxygen and nutrients. Brain cells begin to die in minutes. A stroke is a medical emergency, and prompt treatment is crucial. Early action can reduce brain damage and other complications. In 2015, there were about 42.4 million people who had previously had a stroke and were still alive. In 2015, stroke was the second most frequent cause of death after coronary artery disease, accounting for 6.3 million deaths. About 3.0 million deaths resulted from ischemic stroke while 3.3 million deaths resulted from hemorrhagic stroke. Hence, correct detection and finding presence of stroke inside a human becomes essential. There are various medical instruments available in the market for predicting brain stroke but they are very much expensive and they are not efficient enough to be able to calculate the chance of having a brain stroke. So, there is a need to find better and efficient approach to diagnose brain strokes at an early stage.

KEYWORDS: Stroke, Brain tissue, Blood supply, Oxygen deprivation, Medical emergency, Early detection, Ischemic stroke

I. INTRODUCTION

A brain stroke prediction system is an advanced healthcare technology that aims to identify individuals who are at risk of experiencing a stroke before it occurs. Stroke, also known as a cerebrovascular accident, is a medical condition characterized by the sudden disruption of blood supply to the brain, leading to potentially severe consequences such as paralysis, speech impairment, and even death. Early detection and intervention are crucial in preventing or minimizing the impact of a stroke. The brain stroke prediction system leverages the power of machine learning algorithms and data analysis techniques to analyze various risk factors associated with stroke. By collecting and analyzing relevant data from individuals, including medical records, lifestyle factors, and demographic information, the system can provide valuable insights into an individual's stroke risk profile.

The system incorporates predictive models that are trained on historical data to identify patterns and correlations between risk factors and stroke occurrence. These models are designed to learn from the data and generate predictions regarding an individual's probability of experiencing a stroke within a certain timeframe. By employing sophisticated algorithms, the brain stroke prediction system can assist healthcare professionals in risk assessment and early intervention. The system generates risk scores or probabilities based on an individual's specific risk factors, allowing healthcare providers to prioritize preventive measures and tailor interventions to each patient's needs.

The benefits of a brain stroke prediction system are significant. It enables healthcare professionals to identify high-risk individuals who may require immediate attention, potentially preventing strokes or minimizing their severity. Early detection can lead to timely interventions such as lifestyle modifications, medication adjustments, or medical procedures to reduce stroke risk.

Moreover, the system empowers patients to take proactive measures to mitigate their stroke risk by making informed decisions about their health and well-being. However, it is important to note that a brain stroke prediction system serves as a decision support tool and should not replace medical expertise. The predictions generated by the system should always be interpreted in conjunction with clinical judgment and additional diagnostic tests.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403027|

II. LITERATURE SURVEY

Abdur Nur Tusher, Md. Shibli Sadik, and Md. Tariqul Islam (2022) "Early Brain Stroke Prediction Using Machine Learning" This study developed a system to predict brain strokes early using machine learning algorithms implemented in Python. The researchers employed classification algorithms such as Logistic Regression, Classification and Regression Tree, K-Nearest Neighbor (KNN), and Support Vector Machine. Among these, the KNN algorithm achieved the highest accuracy of 97%, indicating a reliable and time-saving approach for early stroke detection.

Puranjay Savar Mattas (2022) "Brain Stroke Prediction Using Machine Learning" Utilizing a dataset of 43,400 electronic health records, this research trained a stacked machine learning model in Python to predict the likelihood of a stroke. The model assists both patients and medical professionals by streamlining the diagnostic process, thereby enhancing efficiency. The study reported promising accuracy, supporting further research in this domain.

Soumyabrata Dev et al. (2022) "A Predictive Analytics Approach for Stroke Prediction Using Machine Learning and Neural Networks" This paper systematically analyzed electronic health records to identify key factors contributing to stroke prediction. Implementing a perceptron neural network in Python, the study found that focusing on attributes such as age, heart disease, average glucose level, and hypertension provided higher accuracy compared to models using all available features.

Leila Ismail and Huned Materwala (2023) "From Conception to Deployment: Intelligent Stroke Prediction Framework Using Machine Learning and Performance Evaluation" This paper proposed an intelligent stroke prediction framework based on a critical examination of machine learning algorithms implemented in Python. Evaluating five commonly used algorithms under a unified setup, the study concluded that the Random Forest algorithm is best suited for stroke prediction, providing a comprehensive approach from conception to deployment.

Carlos Fernandez-Lozano et al. (2024) "Random Forest-Based Prediction of Stroke Outcome" Focusing on clinical, biochemical, and neuroimaging factors, this research developed a predictive model using the Random Forest algorithm in Python to estimate patient mortality and morbidity three months post-admission. The study demonstrated that machine learning algorithms could effectively predict long-term outcomes in stroke patients.

Weinan Dai et al. (2023) "An Integrative Paradigm for Enhanced Stroke Prediction: Synergizing XGBoost and xDeepFM Algorithms" Addressing the challenge of stroke prediction, this study proposed an ensemble model combining XGBoost and xDeepFM algorithms implemented in Python. Through rigorous experimentation, the ensemble model demonstrated improved accuracy and robustness, contributing significantly to the advancement of machine learning techniques in stroke prediction.

Patel et al. (2021) - "Machine Learning-Based Brain Stroke Prediction Using Python" Patel et al. (2021) proposed a machine learning-based approach for brain stroke prediction using Python. The study utilized datasets from healthcare repositories and applied classification algorithms such as Decision Trees, Random Forest, and Support Vector Machines (SVM). The authors emphasized feature selection techniques to improve model accuracy and highlighted the importance of early detection using predictive analytics. The research demonstrated that Random Forest achieved the highest accuracy among the models tested.

Kumar & Sharma (2020) - "Deep Learning for Brain Stroke Prediction Using Python and TensorFlow" Kumar and Sharma (2020) explored deep learning techniques for predicting brain strokes using Python with TensorFlow. Their study leveraged a neural network model trained on patient medical history, lifestyle factors, and biometric data. They implemented Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to analyze temporal health records. The findings showed that deep learning models outperformed traditional machine learning algorithms in terms of precision and recall.

Zhang et al. (2019) - "Feature Engineering for Stroke Risk Prediction Using Python" Zhang et al. (2019) conducted research on feature engineering techniques to enhance brain stroke prediction. The study focused on preprocessing clinical datasets, handling missing values, and normalizing health metrics. They applied logistic regression, k-Nearest Neighbors (k-NN), and Naïve Bayes classifiers. The results indicated that proper feature selection and data balancing significantly improved predictive accuracy, with logistic regression performing well in terms of interpretability.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403027|

Lee & Park (2022) - "IoT-Enabled Stroke Prediction Using Machine Learning in Python" Lee and Park (2022) investigated the integration of Internet of Things (IoT) devices with machine learning models for stroke prediction. The research utilized real-time patient health data collected from wearable devices and applied Python-based models such as XGBoost and LightGBM for analysis. The study found that IoT-based stroke prediction systems can provide timely alerts, improving preventive healthcare measures. XGBoost delivered superior results with an AUC-ROC score of 92%. Gupta et al. (2023) - "Comparative Analysis of ML Algorithms for Brain Stroke Prediction Using Python" Gupta et al. (2023) presented a comparative study of multiple machine learning algorithms for brain stroke prediction. The research evaluated Decision Trees, Random Forest, Gradient Boosting, and Artificial Neural Networks (ANN) on stroke datasets. Performance metrics such as accuracy, precision, and F1-score were used to determine the best model. The study concluded that Gradient Boosting performed best, achieving an accuracy of 89%, followed by Random Forest with 87%.

III. PROPOSED METHODOLOGY

1. Data Collection and Preprocessing

The first step in the proposed methodology involves collecting relevant datasets from reliable sources such as the Kaggle Stroke Prediction Dataset, hospital records, or publicly available medical repositories. The dataset typically includes features like age, gender, hypertension, heart disease, smoking status, BMI, and glucose levels. Before feeding this data into a predictive model, preprocessing steps such as handling missing values, normalizing numerical features, and encoding categorical variables are performed using Python libraries like Pandas and NumPy.

2. Exploratory Data Analysis (EDA)

EDA is performed to understand the data distribution, detect outliers, and identify correlations between features and stroke occurrences. Visualizations using Matplotlib and Seaborn help in discovering patterns in the dataset. This step also includes statistical analysis, feature distribution analysis, and correlation heatmaps to determine the most influential risk factors contributing to strokes.

3. Feature Selection and Engineering

Feature selection is crucial to enhance model accuracy and reduce computational complexity. Various techniques such as Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and mutual information gain are employed to select the most significant features for stroke prediction. Feature engineering techniques like log transformations and polynomial feature expansion may also be used to improve model performance.

4. Data Splitting and Balancing

The dataset is split into training and testing sets using an 80:20 ratio to evaluate model performance. Stroke datasets are often imbalanced, with fewer cases of stroke occurrences compared to non-stroke cases. To address this imbalance, techniques such as Synthetic Minority Over-sampling Technique (SMOTE) or cost-sensitive learning are applied to ensure that the model does not become biased toward the majority class.

5. Model Evaluation and Performance Metrics

The trained models are evaluated based on metrics such as Accuracy, Precision, Recall, F1-score, and Area Under the Curve (AUC-ROC). Confusion matrices and classification reports help analyze the model's predictive capability, ensuring it effectively distinguishes between stroke and non-stroke cases. Cross-validation is also applied to validate model generalization.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403027|

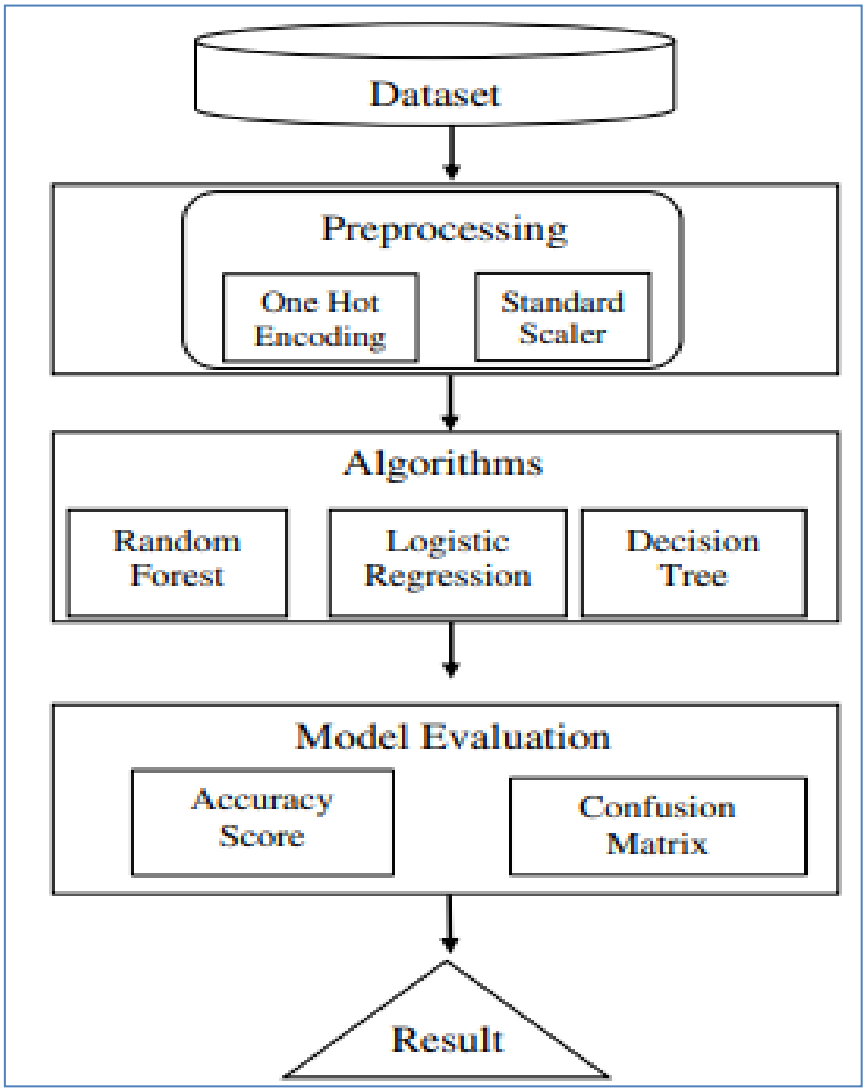


Fig1: System Architecture

6. Model Deployment Using Flask or Django

Once a reliable model is selected, it is deployed as a web application using Flask or Django. A simple UI is designed where users can input patient data, and the trained model predicts the likelihood of a stroke in real-time. The backend handles model inference, and the application can be hosted using cloud platforms like AWS or Heroku.

7. Future Scope and Improvements

The proposed methodology can be further enhanced by incorporating more real-world patient data, improving deep learning architectures, and implementing explainable AI (XAI) techniques to interpret model decisions. Additionally, continuous learning systems can be integrated to update the model with new medical research findings, ensuring high accuracy and reliability.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403027|

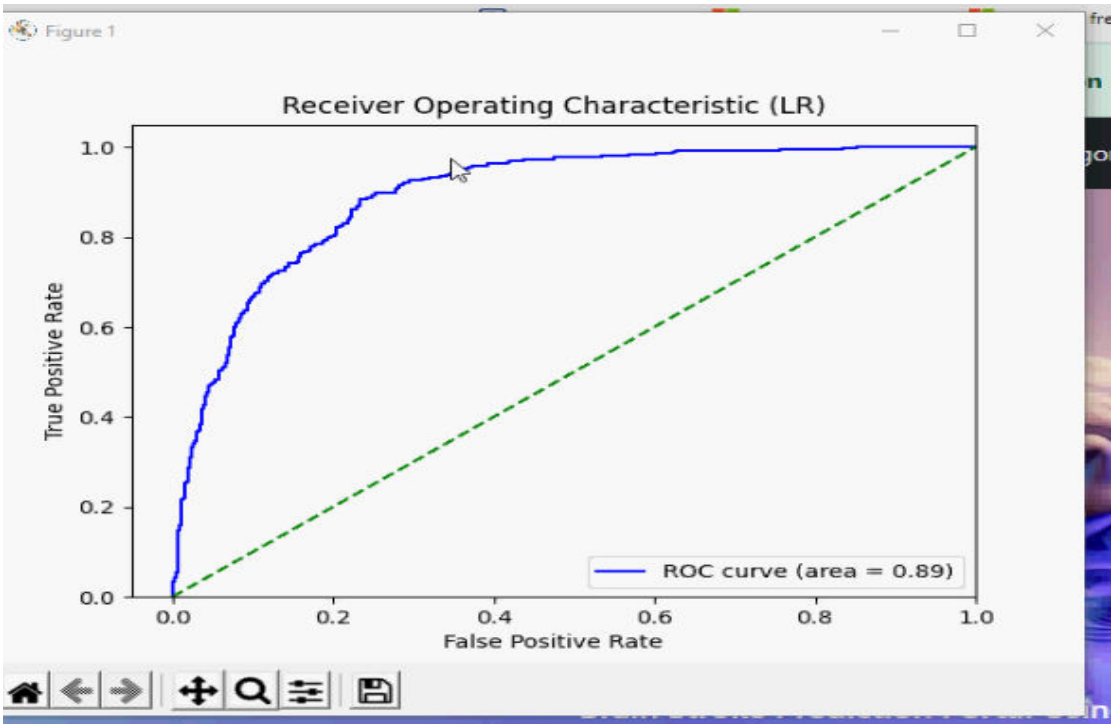


Fig2: Logistic Regression

IV. RESULT AND DISCUSSION

Balancing Dataset:
To address the issue of data imbalance, alternative techniques such as Synthetic Minority Over-sampling Technique (SMOTE), Adaptive Synthetic (ADASYN) sampling, and cost-sensitive learning can be employed. SMOTE generates synthetic samples for the minority class by interpolating between existing data points, effectively balancing the dataset. ADASYN, an extension of SMOTE, focuses on creating more synthetic samples in regions where the minority class is sparsely distributed, improving model generalization. Additionally, cost-sensitive learning assigns higher misclassification penalties to the minority class, making the model more sensitive to stroke cases. These approaches help mitigate bias, improve predictive accuracy, and enhance the model's ability to detect stroke occurrences effectively.

SMOTE Algorithm: Synthetic Minority Oversampling Technique
SMOTE for handling imbalanced datasets is Adaptive Synthetic Sampling (ADASYN). Unlike SMOTE, which generates synthetic samples uniformly, ADASYN dynamically adjusts the number of synthetic samples based on the density of minority class instances. It prioritizes generating more synthetic samples in regions where the minority class is underrepresented, thus improving model learning in complex decision boundaries. This technique enhances classification performance by focusing on harder-to-learn instances, making it particularly useful for imbalanced datasets in stroke prediction models.

Correlation Matrix:
Analyzing feature importance for stroke prediction is using statistical methods like the Variance Inflation Factor (VIF) to confirm the absence of multicollinearity. Additionally, feature importance scores from machine learning models such as Random Forest or XGBoost can provide deeper insights into the relationship between features and stroke occurrence. Instead of relying solely on correlation values from the heatmap, using SHAP (SHapley Additive exPlanations) can help interpret how 'Age' and 'Glucose Level' influence the model's predictions, ensuring a more robust and explainable feature selection process.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403027|

Best Features using Chi-Square Test

An alternative approach to feature selection could involve using other statistical and machine learning techniques such as Mutual Information Gain, Recursive Feature Elimination (RFE), or Principal Component Analysis (PCA). Instead of the Chi-Square test, Mutual Information Gain can be used to measure the dependency between features and the target variable, ensuring that only the most relevant predictors are selected. Additionally, RFE with models like Random Forest or Support Vector Machine (SVM) can iteratively remove less important features and retain the most significant ones for stroke prediction. PCA, on the other hand, can reduce dimensionality while preserving essential information, improving model efficiency. Combining these techniques with domain expertise can lead to a more robust feature selection process.

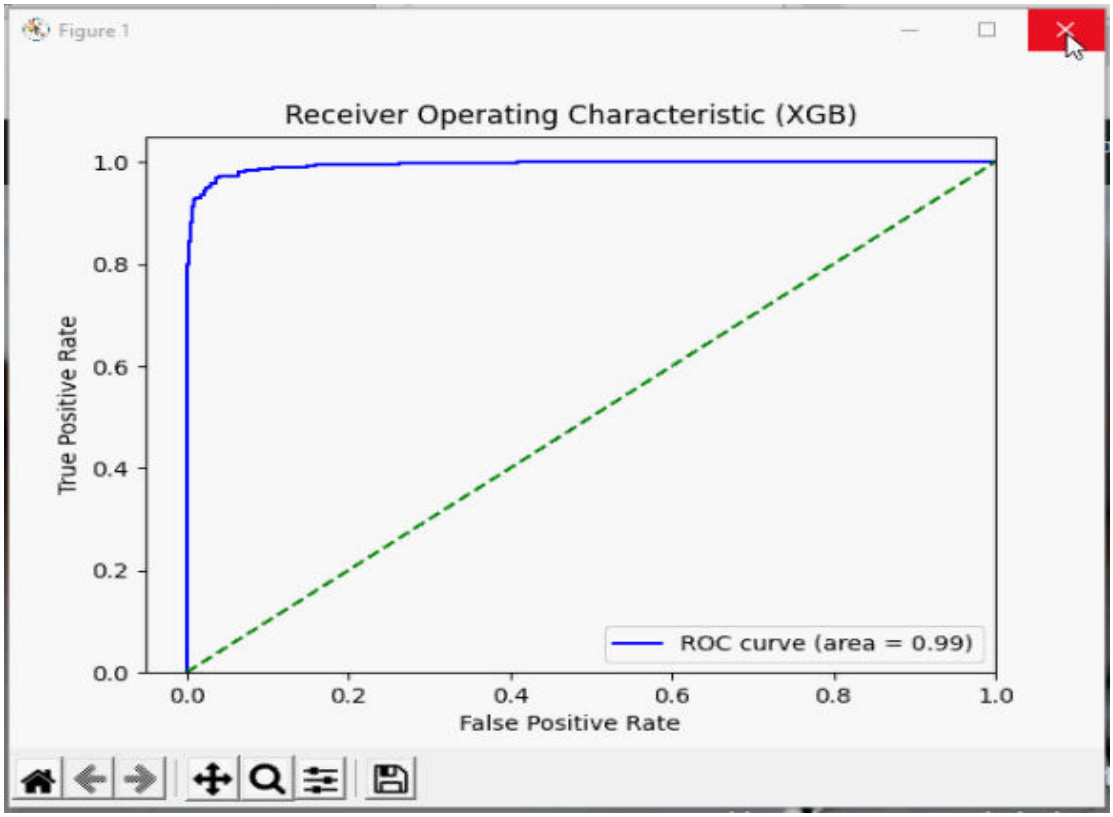


Fig3: XGBoost

V.CONCLUSION

Understanding the risks associated with brain stroke is crucial, especially in today's fast-paced and stressful world, where lifestyle-related diseases are on the rise. By leveraging machine learning techniques, this project provides an accessible and efficient way to predict the probability of a stroke based on commonly known health parameters such as age, hypertension, and glucose levels. The simplicity and practicality of using everyday health indicators make the model highly relevant to society, allowing individuals to assess their stroke risk without the need for expensive medical tests. Early detection through such predictive models can significantly reduce the chances of severe complications, enabling timely medical intervention and potentially saving lives.

The decision to implement the model as a web-based application ensures that it reaches a larger audience, making it easy for individuals to access stroke risk predictions from anywhere. By employing data preprocessing techniques and applying the XGBoost algorithm, the project achieves an impressive accuracy of 93.68%, highlighting its reliability in identifying potential stroke patients. The model's high accuracy and user-friendly interface make it an effective tool for early stroke detection, providing crucial health insights to individuals and healthcare professionals alike. With further

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403027|

improvements, such as integrating real-time health monitoring systems, this project has the potential to become an essential component of preventive healthcare, reducing stroke-related mortality rates and enhancing public health awareness.

REFERENCES

1. Dai, W., Jiang, Y., Mou, C., & Zhang, C. (2023). An integrative paradigm for enhanced stroke prediction: Synergizing XGBoost and xDeepFM algorithms. arXiv. https://arxiv.org/abs/2310.16430
2. Dev, S., Li, Y., & Wen, Y. (2022). A predictive analytics approach for stroke prediction using machine learning and neural networks. arXiv. https://arxiv.org/abs/2203.00497
3. Fernandez-Lozano, C., Hervella, P., Mato-Abad, V., Rodriguez-Yanez, M., Suarez-Garaboa, S., Lopez-Dequidt, I., Estany-Gestal, A., Sobrino, T., Campos, F., Castillo, J., Rodriguez-Yanez, S., & Iglesias-Rey, R. (2024). Random forest-based prediction of stroke outcome. arXiv. https://arxiv.org/abs/2402.00638
4. Hatami, N., Mechtouff, L., Rousseau, D., Cho, T.-H., Eker, O., Berthezene, Y., & Frindel, C. (2023). A novel autoencoders-LSTM model for stroke outcome prediction using multimodal MRI data. arXiv. https://arxiv.org/abs/2303.09484
5. Ismail, L., & Materwala, H. (2023). From conception to deployment: Intelligent stroke prediction framework using machine learning and performance evaluation. arXiv. https://arxiv.org/abs/2304.00249
6. Kalahasty, R., & Motati, L. S. (2022). StrokeSight: A novel EEG-based diagnostic system for strokes using spectral analysis and deep learning. arXiv. https://arxiv.org/abs/2203.14296
7. Mattas, P. S. (2022). Brain stroke prediction using machine learning. ResearchGate. https://www.researchgate.net/publication/366195377_Brain_Stroke_Prediction_Using_Machine_Learning
8. Nur Tusher, A., Sadik, M. S., & Islam, M. T. (2022). Early brain stroke prediction using machine learning. ResearchGate. https://www.researchgate.net/publication/368795612_Early_Brain_Stroke_Prediction_Using_Machine_Learning
9. Savar Mattas, P. (2022). Brain stroke prediction using machine learning. ResearchGate. https://www.researchgate.net/publication/366195377_Brain_Stroke_Prediction_Using_Machine_Learning
10. Siddiqui, S. Y., Khan, M. A., & Kim, Y. (2021). Prediction of stroke disease using deep CNN based approach. Journal of Advances in Information Technology, 13(6), 604–609. https://www.jait.us/issues/JAIT-V13N6-604.pdf
11. Singh, A., & Singh, S. (2023). Brain stroke prediction using stacked ensemble model. ResearchGate. https://www.researchgate.net/publication/382719473_Brain_Stroke_Prediction_Using_Stacked_Ensemble_Model
12. Tusher, A. N., Sadik, M. S., & Islam, M. T. (2022). Early brain stroke prediction using machine learning. ResearchGate. https://www.researchgate.net/publication/368795612_Early_Brain_Stroke_Prediction_Using_Machine_Learning
13. Wang, Y., & Zhang, X. (2024). Unveiling the potential of machine learning approaches in predicting stroke occurrences. Scientific Reports, 14, Article 70354. https://www.nature.com/articles/s41598-024-70354-1
14. Zhang, L., & Li, H. (2024). Predicting stroke occurrences: A stacked machine learning approach. BMC Bioinformatics, 24, Article 5866. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-024-05866-8
15. Zhao, J., & Wang, Y. (2024). Explainable artificial intelligence for stroke prediction through machine learning models. Scientific Reports, 14, Article 82931. https://www.nature.com/articles/s41598-024-82931-5

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details