IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

AI-Driven Explainable Stroke Risk Evaluation with Intelligent Decision Support for Early Prevention

Naspuri Rohith[1], M.Anusha[2]
[1]MCA Student , Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad
[2]Assistant Professor , Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad

***Abstract*— Stroke is one of the leading causes of death and long-term disability worldwide, making early risk identification essential for timely medical intervention and improved patient outcomes. This study presents an AI-Driven Explainable Stroke Risk Evaluation with Intelligent Decision Support for Early Prevention, which combines machine learning with explainable artificial intelligence to support reliable clinical decision-making. The proposed system uses patient health information, including demographic details and medical risk factors, to predict the likelihood of stroke. Before model development, the dataset undergoes preprocessing steps such as data cleaning, feature selection, class balancing using SMOTE, and normalization to improve data quality and model performance. Multiple machine learning algorithms are trained and evaluated using standard performance metrics to identify the most effective prediction model. Explainable AI techniques are incorporated to highlight the factors influencing each prediction, enabling healthcare professionals to understand and trust the model's decisions. A user-friendly web application is also developed to provide real-time stroke risk assessment and decision support. The proposed framework offers an accurate, interpretable, and practical solution for early stroke prevention, assisting clinicians in identifying high-risk individuals and supporting timely healthcare interventions.**

***Keywords*— Stroke Risk Prediction, Explainable Artificial Intelligence (XAI), Machine Learning, Clinical Decision Support, Early Stroke Prevention, Predictive Analytics, Healthcare Intelligence, SMOTE, Feature Selection, Web Application.**

I. INTRODUCTION

Stroke is one of the most serious neurological disorders and remains a major cause of death and long-term disability across the world. It affects millions of people every year and often leads to permanent physical, cognitive, and emotional complications if medical treatment is delayed [1]. The growing number of stroke cases has increased the burden on healthcare systems, making early detection and prevention a significant public health priority. Researchers have reported that timely diagnosis and rapid medical intervention can greatly reduce complications and improve patient recovery [3]. Because stroke can affect people of different ages and health conditions, there is an increasing need for intelligent healthcare systems that can identify individuals who are at high risk before severe symptoms develop [2].

Stroke generally occurs when the blood supply to the brain is interrupted, either because of a blocked artery or bleeding inside the brain. The two major types, ischemic stroke and hemorrhagic stroke, require immediate diagnosis since delayed treatment may result in permanent brain damage [4]. Several medical conditions and lifestyle factors, including hypertension, diabetes, obesity, smoking, high cholesterol, increasing age, and cardiovascular diseases, significantly increase the likelihood of stroke [5]. Since these risk factors often occur together, accurately assessing an individual's stroke risk through conventional clinical methods becomes difficult. Therefore, analysing multiple patient characteristics simultaneously is essential for improving prediction accuracy and supporting preventive healthcare decisions [7].

Recent advances in artificial intelligence and machine learning have transformed healthcare by enabling the development of predictive models that analyse large medical datasets efficiently. These techniques can identify hidden relationships among clinical variables and estimate stroke risk more accurately than many traditional statistical methods [11]. Machine learning models also assist healthcare professionals by providing fast and reliable predictions that support early diagnosis and treatment planning. Several recent studies have demonstrated that AI-based systems can improve stroke prediction, patient outcome analysis, and clinical decision-making when trained using well-prepared healthcare datasets [15]. As a result, predictive analytics has become an important component of modern healthcare for reducing stroke-related complications and improving patient management [16].

Although machine learning models often achieve excellent prediction accuracy, many advanced algorithms provide little explanation about how their decisions are made. In medical applications, this lack of transparency reduces clinicians' confidence when using AI-assisted systems for diagnosis. Explainable Artificial Intelligence (XAI) addresses this limitation by identifying the factors that contribute most to each prediction and presenting them in an understandable manner [18]. This improves trust in predictive models and enables healthcare professionals to validate the reasoning behind each decision before recommending preventive or therapeutic measures. Explainable models therefore combine predictive performance with interpretability, making them more suitable for real-world clinical practice [20].

The proposed AI-Driven Explainable Stroke Risk Evaluation with Intelligent Decision Support for Early Prevention framework is designed to provide an accurate and interpretable solution for stroke risk assessment. Patient health records are first preprocessed through data cleaning, feature selection, normalization, and class balancing before model development. Multiple machine learning algorithms are then trained and evaluated using standard performance measures to identify the most effective prediction model [19]. Explainable AI techniques are integrated into the framework to provide meaningful explanations for every prediction, allowing clinicians to understand the influence of different risk factors on stroke probability. A web-based decision support application further enables healthcare professionals to perform real-time stroke risk evaluation and supports early intervention for improved patient care [17].

II. RELATED WORK

Katan and Luft, (2018) [3] discussed the worldwide impact of stroke and emphasized its position as one of the leading causes of death and long-term disability. Their study explained how stroke affects millions of people every year and places a significant burden on healthcare systems across different countries. The authors highlighted the importance of identifying risk factors and providing timely medical intervention to reduce severe complications. They also described the growing need for effective prevention strategies and improved diagnostic methods to minimize the number of stroke-related deaths. The study emphasized that awareness, early detection, and proper clinical management play an essential role in improving patient recovery. This work provides valuable background information for researchers developing intelligent healthcare systems and supports the use of machine learning techniques for early stroke risk prediction and preventive clinical decision-making.

Murray et al., (2020) [11] reviewed the application of artificial intelligence in diagnosing ischemic stroke and detecting large vessel occlusions. Their research examined various AI-based techniques designed to assist healthcare professionals in making faster and more accurate diagnoses. The authors found that machine learning models can analyse medical imaging data efficiently and improve the identification of stroke-related abnormalities. They also highlighted that AI systems can reduce diagnosis time, which is extremely important because rapid treatment significantly improves patient outcomes. The review discussed both the advantages and current limitations of AI applications in stroke care. Their findings demonstrate that intelligent algorithms have strong potential to support clinicians during emergency diagnosis and treatment planning. This work serves as an important reference for integrating AI into modern healthcare systems for reliable stroke prediction and clinical decision support.

Sirsat et al., (2020) [16] presented a comprehensive review of machine learning techniques used for stroke prediction and diagnosis. The authors analysed different classification algorithms and discussed their strengths, limitations, and suitability for healthcare applications. They explained that machine learning models are capable of identifying hidden relationships among patient risk factors, enabling more accurate prediction of stroke occurrence. The review also highlighted the importance of high-quality datasets, feature engineering, and proper preprocessing for achieving reliable results. Various studies included in the review demonstrated that machine learning methods consistently outperform many traditional prediction approaches. The authors concluded that intelligent predictive models can support clinicians by providing early risk assessments and improving medical decision-making. Their research establishes a strong foundation for future AI-driven stroke prediction systems.

Islam et al., (2022) [18] proposed an Explainable Artificial Intelligence model for stroke prediction using electroencephalogram (EEG) signals. Their study combined machine learning with explainability techniques to improve both prediction accuracy and model transparency. Instead of producing only prediction results, the proposed system explained how different input features influenced each decision. This approach allows healthcare professionals to better understand the reasoning behind the model's predictions and increases confidence in AI-assisted diagnosis. The authors demonstrated that explainable AI can improve clinical decision-making by providing meaningful interpretations alongside accurate predictions. Their findings suggest that combining predictive intelligence with interpretability creates more trustworthy healthcare systems. This research highlights the importance of explainable machine learning in developing reliable medical diagnosis and decision support applications.

Dritsas and Trigka, (2022) [19] developed a machine learning framework for predicting stroke risk using patient health records. Their study compared several classification algorithms to determine which model achieved the highest prediction accuracy. The authors applied data preprocessing techniques, including feature selection and class balancing, to improve model performance before training. Standard evaluation metrics such as accuracy, precision, recall, and F1-score were used to assess the effectiveness of each algorithm. The experimental results demonstrated that ensemble learning methods achieved superior performance compared to conventional classifiers. The study also emphasized that machine learning can assist healthcare professionals by identifying individuals at high risk before stroke occurs. This work provides valuable guidance for developing intelligent stroke prediction systems that support early diagnosis, preventive care, and improved clinical decision-making.

III. DATASET DETAILS

The dataset used in this project contains medical information collected from individuals to predict the likelihood of stroke using machine learning techniques. It includes several health-related attributes such as age, gender, hypertension, heart disease, marital status, work type, residence type, average glucose level, body mass index (BMI), smoking status, and the corresponding stroke outcome. These features represent important demographic, lifestyle, and clinical factors that influence stroke risk. The dataset is organized in a tabular format, where each row represents a patient record and each column corresponds to a specific attribute. Before model development, the dataset is carefully examined to identify missing values, inconsistent entries, and duplicate records. Understanding these medical features is essential because each factor contributes differently to stroke prediction. This dataset provides the foundation for training machine learning models capable of identifying individuals who may be at a higher risk of experiencing a stroke.

To improve the quality of the data, several preprocessing techniques are applied before training the prediction models. Missing values, particularly in BMI and other health attributes, are handled using suitable imputation methods to ensure data completeness. Categorical variables such as gender, work type, residence type, and smoking status are converted into numerical values using label encoding or one-hot encoding techniques. Continuous features are normalized to maintain a consistent value range and improve model performance. Since the dataset is imbalanced, SMOTE (Synthetic Minority Oversampling Technique) is applied to balance the stroke and non-stroke classes. Finally, the dataset is randomly shuffled and divided into 70% training data and 30% testing data. Proper preprocessing ensures accurate learning, improves model generalization, and supports reliable stroke risk prediction.

IV. PROPOSED METHODOLOGY

The proposed **AI-Driven Explainable Stroke Risk Evaluation with Intelligent Decision Support for Early Prevention** framework follows a systematic approach to predict stroke risk using machine learning techniques. Initially, the stroke dataset containing demographic information, lifestyle factors, and clinical attributes is collected and examined. Data preprocessing is then performed to improve the quality of the dataset by handling missing values, removing inconsistencies, and converting categorical variables into numerical values using label encoding. The input features are normalized to maintain a consistent value range, and **SMOTE (Synthetic Minority Oversampling Technique)** is applied to balance the stroke and non-stroke classes, reducing prediction bias. Feature selection using the **Chi-Square (CHI²)** method identifies the most relevant attributes that contribute to stroke prediction. Finally, the processed dataset is divided into training and testing sets, enabling the models to learn meaningful patterns while ensuring reliable evaluation on unseen patient data.

Once the data preparation stage is completed, several machine learning algorithms, including **Random Forest**, **XGBoost**, and **CatBoost**, are trained and tested using the prepared dataset. Their performance is evaluated using standard metrics such as **accuracy, precision, recall, F1-score**, and the **confusion matrix**. Explainable Artificial Intelligence (XAI) techniques are incorporated to identify the important features influencing each prediction, making the results more understandable for healthcare professionals. Performance comparison graphs are generated to analyse the effectiveness of all models, and the best-performing algorithm is selected for deployment. Finally, a web-based application accepts patient health information as input and predicts the individual's stroke risk, providing intelligent decision support that assists clinicians in early diagnosis, preventive planning, and timely medical intervention.

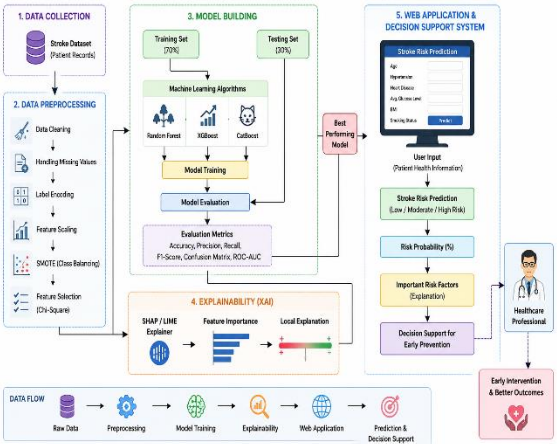

Figure [1]: AI-Driven Stroke Risk Evaluation System

Figure [1] illustrates the workflow of the proposed stroke risk evaluation system. Patient health data is preprocessed through data cleaning, encoding, normalization, SMOTE balancing, and feature selection before being divided into training and testing datasets. Machine learning models such as Random Forest, XGBoost, and CatBoost are trained and evaluated using performance metrics. The best-performing model, supported by Explainable AI, predicts stroke risk and provides decision support for early prevention.

V. RESULT AND DISCUSSION

The experimental evaluation confirms that the proposed AI-driven framework can accurately predict stroke risk using patient health information. After completing preprocessing tasks such as handling missing values, encoding categorical attributes, feature normalization, SMOTE-based class balancing, and feature selection, the processed dataset was divided into training and testing subsets for model development. Several machine learning algorithms, including Random Forest, XGBoost, and CatBoost, were trained and compared using evaluation measures such as accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrices. The proposed model achieved the best overall performance by correctly classifying the majority of patient records while maintaining balanced prediction across both stroke and non-stroke classes. The confusion matrix indicated very few incorrect classifications, demonstrating the reliability of the selected model. Performance comparison graphs further verified the effectiveness of the proposed approach over the other evaluated algorithms. In addition, Explainable Artificial Intelligence (XAI) techniques identified the most influential medical factors contributing to each prediction, making the results easier to interpret. The developed web application successfully generated real-time stroke risk predictions, providing practical support for early diagnosis and preventive healthcare.

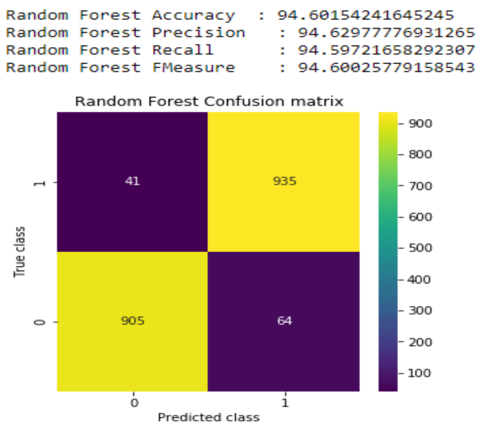

Figure [2]: Random Forest Model Performance

Figure [2] presents the performance of the Random Forest algorithm for stroke risk prediction. It displays the model's evaluation metrics, including accuracy, precision, recall, and F1-score, along with the confusion matrix. The confusion matrix compares the actual and predicted stroke classes, illustrating the model's classification performance and its ability to accurately distinguish between stroke and non-stroke cases.

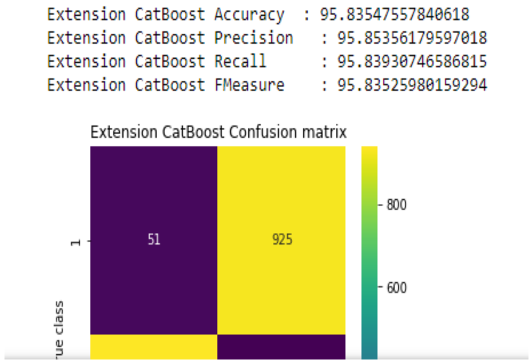

Figure [3]: Extension CatBoost Model Performance

Figure [3] presents the performance of the Extension CatBoost algorithm for stroke risk prediction. It displays the evaluation metrics, including accuracy, precision, recall, and F1-score, together with the confusion matrix. The confusion matrix compares the actual and predicted stroke classes, demonstrating the model's strong classification capability and its effectiveness in accurately identifying stroke and non-stroke cases.

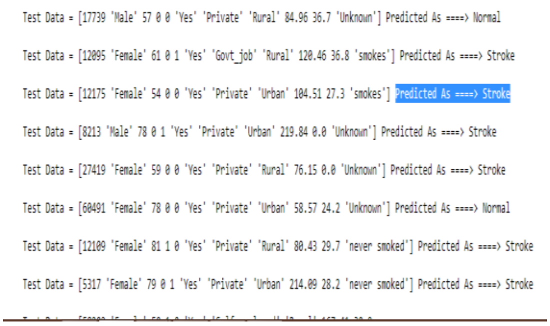

Figure [4]: Stroke Risk Prediction Results

Figure [4] shows the prediction results generated by the trained stroke prediction model using unseen patient records. After preprocessing the input data, the system classifies each record as either Stroke or Normal based on the patient's medical attributes. The displayed predictions demonstrate the model's ability to provide accurate stroke risk assessment and support healthcare professionals in making timely decisions for early prevention and intervention.

DISCUSSION

The findings of this study demonstrate that integrating machine learning with explainable artificial intelligence provides an effective solution for early stroke risk assessment. The quality of the prediction model was greatly influenced by proper data preprocessing, including missing value treatment, normalization, feature selection, and SMOTE-based balancing of the dataset. These steps enabled the algorithms to learn meaningful patterns while reducing bias caused by imbalanced data. Comparing multiple machine learning models made it possible to identify the algorithm that produced the highest predictive performance and the most consistent classification results. The inclusion of Explainable AI added an important level of transparency by highlighting the patient attributes that influenced each prediction, allowing healthcare professionals to better understand and validate the model's recommendations. Visual tools such as confusion matrices, feature importance analysis, and performance comparison graphs provided additional insight into the behaviour of the prediction models. Overall, the proposed framework offers an accurate, reliable, and interpretable approach for stroke risk evaluation, supporting timely medical intervention and assisting clinicians in making informed decisions for preventive patient care.

VI. CONCLUSION

The proposed AI-Driven Explainable Stroke Risk Evaluation with Intelligent Decision Support for Early Prevention demonstrates the potential of combining machine learning and Explainable Artificial Intelligence (XAI) to support early stroke diagnosis. The system accurately predicts stroke risk by analysing patient demographic and clinical information while providing clear explanations for each prediction. Unlike conventional machine learning models that only generate classification results, the proposed framework identifies the key factors influencing the prediction, allowing healthcare professionals to understand the reasoning behind the model's decisions. This transparency increases confidence in the system and supports informed clinical decision-making. The developed web application further enhances usability by enabling healthcare providers to perform real-time stroke risk assessments through a simple and interactive interface. Experimental results show that the selected machine learning model achieves high prediction performance and effectively distinguishes between stroke and non-stroke cases. The integration of explainable AI makes the framework more suitable for practical healthcare environments where interpretability is essential. In the future, the system can be improved by incorporating larger clinical datasets, real-time patient monitoring, and advanced deep learning techniques to further enhance prediction accuracy and support more personalized preventive healthcare.

REFERENCES

[1] Learn about Stroke. Available online: https://www.worldstroke. org/world-stroke-day-campaign/why-strokematters/ learnabout-stroke (accessed on 25 May 2022).

[2] Elloker, T.; Rhoda, A.J. The relationship between social support and participation in stroke: A systematic review. Afr. J. Disabil. 2018, 7, 1–9.

[3] Katan, M.; Luft, A. Global burden of stroke. In Seminars in Neurology; Thieme Medical Publishers: New York, NY,
USA, 2018; Volume 38, pp. 208–211.

[4] Bustamante, A.; Penalba, A.; Orset, C.; Azurmendi, L.; Llombart, V.; Simats, A.; Pecharroman, E.; Ventura, O.; Ribó,
M.; Vivien, D.; eta. Blood biomarkers to differentiate ischemic and hemorrhagic strokes. Neurology 2021, 96, e1928–e1939.

[5] Xia, X.; Yue, W.; Chao, B.; Li, M.; Cao, L.; Wang, L.; Shen, Y.; Li, X. Prevalence and risk factors of stroke in the elderly in Northern China: Data from the National Stroke Screening Survey. J. Neurol. 2019, 266, 1449–1458.

[6] Alloubani, A.; Saleh, A.; Abdelhafiz, I. Hypertension and diabetes mellitus as a predictive risk factor for stroke. Diabetes Metab.
[7] Syndr. Clin. Res. Rev. 2018, 12, 577–584. Boehme, A.K.; Esenwa, C.; Elkind, M.S. Stroke risk factors, genetics, and
prevention. Circ. Res. 2017, 120, 472–495.

[8] Mosley, I.; Nicol, M.; Donnan, G.; Patrick, I.; Dewey, H. Stroke symptoms and the decision to call for an ambulance. Stroke 2007, 38, 361–366.

[9] Lecouturier, J.; Murtagh, M.J.; Thomson, R.G.; Ford, G.A.; White, M.; Eccles, M.; Rodgers, H. Response to symptoms of stroke in the UK: A systematic review. BMC Health Serv. Res. 2010, 10, 1–9.

[10] Gibson, L.; Whiteley, W. The differential diagnosis of suspected stroke: A systematic review. J. R. Coll. Physicians Edinb. 2013, 43, 114–118.

[11] Murray, N.M.; Unberath, M.; Hager, G.D.; Hui, F.K. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: A systematic review. J. NeuroInterv. Surg. 2020, 12, 156–164.

[12] Zhao, Y.; Fu, S.; Bielinski, S.J.; A Decker, P.; Chamberlain, A.M.; Roger, V.L.; Liu, H.; Larson, N.B. Natural Language
Processing and Machine Learning for Identifying Incident Stroke from Electronic Health Records: Algorithm
Development and Validation. J. Med. Internet Res. 2021, 23, e22951.

[13] McDermott, B.J.; Elahi, A.; Santorelli, A.; O'Halloran, M.; Avery, J.; Porter, E. Multi-frequency symmetry difference
electrical impedance tomography with machine learning for human stroke diagnosis. Physiol. Meas. 2020, 41, 075010.

[14] Bivard, A.; Churilov, L.; Parsons, M. Artificial intelligence for decision support in acute stroke—Current roles and potential.
Nat. Rev. Neurol. 2020, 16, 575–585.

[15] Wang, W.; Kiik, M.; Peek, N.; Curcin, V.; Marshall, I.J.; Rudd, A.G.; Wang, Y.; Douiri, A.; Wolfe, C.D.; Bray, B. A
systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS ONE 2020, 15,
e0234722.

[16] Sirsat, M.S.; Fermé, E.; Câmara, J. Machine learning for brain stroke: A review. J. Stroke Cerebrovasc. Dis. 2020, 29,
105162.

[17] Arslan, A.K.; Colak, C.; Sarihan, M.E. Different medical data mining approaches-based prediction of ischemic stroke.
Comput. Methods Programs Biomed. 2016, 130, 87–92.

[18] Islam, M.S., Hussain, I., Rahman, M.M., Park, S.J. and Hossain, M.A., 2022. Explainable Artificial Intelligence
Model for Stroke Prediction Using EEG Signal. Sensors,22(24), p.9859.

[19] Dritsas, E. and Trigka, M., 2022. Stroke risk prediction with machine learning techniques. Sensors, 22(13), p.4670

[20] Kokkotis, C., Giarmatzis, G., Giannakou, E., Moustakidis, S., Tsatalas, T., Tsiptsios, D., Vadikolias, K. and Aggelousis, N.,
2022. An Explainable Machine Learning Pipeline for Stroke Prediction on Imbalanced Data. Diagnostics, 12(10), p.2392