UGBOMEH, O., YIYE, V., IBEKE, E., EZENKWU, C.P., SHARMA, V. and ALKHAYYAT, A. 2024. Machine learning algorithms for stroke risk prediction leveraging on explainable artificial intelligence techniques (XAI). In *Proceedings of the 2024 International conference on electrical, electronics and computing technologies (ICEECT 2024), 29-31 August 2024, Greater Noida, India*. Piscataway: IEEE [online], article 10739320. Available from: https://doi.org/10.1109/ICEECT61758.2024.10739320

Machine learning algorithms for stroke risk prediction leveraging on explainable artificial intelligence techniques (XAI).

UGBOMEH, O., YIYE, V., IBEKE, E., EZENKWU, C.P., SHARMA, V. and ALKHAYYAT, A.

2024

Machine Learning Algorithms for Stroke Risk Prediction Leveraging on Explainable Artificial Intelligence Techniques (XAI)

Ogochukwu Ugbomeh
School of Creative and Cultural Business
Robert Gordon University
Aberdeen, United Kingdom
o.ugbomeh@rgu.ac.uk

Versse Yiye
School of Creative and Cultural Business
Robert Gordon University
Aberdeen, United Kingdom
v.yiye@rgu.ac.uk

Ebuka Ibeke
School of Creative and Cultural Bus.
Robert Gordon University
Aberdeen, United Kingdom
e.ibeke@rgu.ac.uk

Chinedu Pascal Ezenkwu
School of Creative and Cultural Business
Robert Gordon University
Aberdeen, United Kingdom
p.ezenkwu@rgu.ac.uk

Vandana Sharma
Computer Science Department
Christ University
Delhi-NCR Campus, India
vandana.juyal@gmail.com

Ahmed Alkhayyat
College of Technical Engineering
The Islamic University
Najaf, Iraq
ahmedalkhayyat85@gmail.com

***Abstract*—Stroke poses a significant global health challenge, contributing to widespread mortality and disability. Identifying predictors of stroke risk is crucial for enabling timely interventions, thereby reducing the increasing impact of strokes. This research addresses this imperative by employing Explainable Artificial Intelligence (XAI) techniques to pinpoint stroke risk predictors. To bridge existing gaps, six machine learning models were assessed using key performance metrics. Utilising the Synthetic Minority Over-sampling Technique (SMOTE) to minimize the impact of the imbalanced nature of the dataset used in this research, the Random Forest algorithm emerged as the most effective among the algorithms with an accuracy of 94.5%, AUC-ROC of 0.95, recall of 0.96, precision of 0.93, and an F1 score of 0.95. This study explored the interpretation of these algorithms and results using Local Interpretable Model-agnostic Explanations (LIME) and Explain Like I'm Five (ELI5). With the interpretations, healthcare providers can gain insight into patients' stroke risk predictions. Key stroke risk factors highlighted by the study include *Age, Marital Status, Glucose Level, Body Mass Index, Work Type, Heart Disease,* and *Gender*. This research significantly contributes to healthcare and healthcare informatics by providing insights that can enhance strategies for stroke prevention and management, ultimately leading to improved patient care. The identified predictors offer valuable information for healthcare professionals to develop targeted interventions, fostering a proactive approach to mitigating the impact of strokes on individuals and the healthcare system.**

***Index Terms*—Explainable Artificial Intelligence (XAI), Explain Like I'm Five (ELI5), Local Interpretable Model Agnostic Explanation (LIME), Machine Learning**

I. INTRODUCTION

As described by Walter et al. [1] cerebrovascular disease, commonly known as stroke, occurs when arteries carrying oxygenated blood to the brain become damaged, resulting in reduced blood supply to various brain regions and eventual impairment of blood vessels within the brain. The World Health Organisation recognises stroke as a substantial global health concern, implicated in approximately 17.5 million fatalities worldwide ranking among the principal contributors to mortality associated with non-communicable diseases. Additionally, stroke has been identified as a leading cause of disability and the second leading cause of death globally [2]. A study by Feign et al. [3] highlights that the global population of stroke survivors exceeds 30 million, with an estimated 8.8 million stroke-related fatalities occurring annually. A substantial portion of these figures involves males and females aged 35 years and older. The burden of stroke lies both in its high mortality and morbidity where about half of its survivors become chronically disabled. Although stroke is a disease of public health importance, its escalating prevalence has received inadequate attention, particularly in the United Kingdom, where over 115,000 individuals annually experience strokes, leading to profound socioeconomic implications.

According to Derek et al. [4] the 2022 global stroke factsheet reveals a significant increase in the lifetime risk of stroke, reflecting a 50% rise over the past 17 years, projecting that approximately 1 in 4 individuals will likely experience a stroke during their lifetime. From 1990 to 2019, there has been a substantial 70% surge in stroke incidence, resulting in a 43% increase in stroke-related fatalities, a 102% upturn in stroke prevalence, and a remarkable 143% rise in Disability Adjusted Life Years (DALY). Notably, the global burden of stroke is concentrated, with 86% of stroke-related deaths and 89% of DALYs situated in low and lower-middle-income countries.

Based on the findings of Patel et al. [5] anticipating the potential epidemiological and socioeconomic consequences of strokes lays the groundwork for the early identification of risk factors, prevention, treatment, and provision of support services. Also, Johnston et al. [6] study emphasises the significance of implementing proactive interventions that strategically target modifiable risk factors, including hypertension, elevated lipid levels, and diabetes. Simultaneously, addressing risks associated with lifestyle choices such as smoking, sedentary behaviour, poor dietary habits, and obesity has demonstrated considerable success in reducing mortality rates attributed to stroke.

Gianfrancesco et al. [7] propose the integration of XAI to clarify the outcomes of machine learning algorithms in predicting strokes. This approach demonstrates substantial potential for mitigating mortality, morbidity, and the incidence of strokes. However, Gawsalyan et al. [8] suggest that mining large volumes of electronic medical record (EMR) data can help identify trends and patterns in patient information. Analysing these trends and patterns from EMRs assists medical practitioners in developing models for the early detection, prevention, and treatment of various illnesses. Scho¨nberger's [9] study argues that despite recognizing the importance of AI and machine learning in healthcare, challenges persist in their widespread adoption, attributed to their inherent opacity and limited ability to explain results to human experts. These challenges are addressed using XAI methodology which creates a traceable link between model predictions and clinical outcomes. XAI promotes interpretability and model transparency while supporting diagnostics and analytics processes.

Overall, the need to understand machine learning models' predictive results gave birth to XAI. In this research, XAI is deployed to enhance the understanding and interpretation of the variables that lead to stroke outcomes prediction, fostering trust, confidence, and informed decision-making in healthcare.

II. Literature Review

Machine learning has experienced substantial growth and attained widespread acceptance, particularly in the medical field, where it plays a critical role in advanced stroke prediction. By leveraging sophisticated algorithms, machine learning is substantially improving the accuracy and efficiency of predicting stroke risk. Most research on stroke risk prediction has focused predominantly on using conventional quantitative and qualitative methods to analyse factors influencing susceptibility to strokes. In a study by Kim et al. [10] the authors investigated the use of natural language processing (NLP) and machine learning algorithms to categorise brain MRI radiology reports. Their systematic approach involved random partitioning, tokenization, and bias correction through cross-validation. The meticulous manual annotation process used to identify clinical notes related to arterial ischemic stroke (AIS) significantly enhanced the methodological robustness of the study. While a key strength of the research lies in the comprehensive comparison of multiple binary classifiers, with a focus on the F1 measure as the primary evaluation metric, the exclusive emphasis on a limited set of classifiers may potentially limit the generalizability of the findings.

Monteiro et al. [11] explored Decision Trees, Random Forests, and Multi-layer Perceptron algorithms to train a stroke prediction model. The accuracy values for these approaches were 74.31%, 74.53%, and 75.02%, respectively, with the Multi-layer Perceptron demonstrating a better accuracy. The study's reliance on accuracy as a performance metric is questionable. Concerns emphasise the overlooking of model interpretability and the clinical significance of accuracy values. Also, the study lacks a comprehensive examination of dataset imbalances and does not address the impact of false positives and false negatives in stroke prediction, which are crucial considerations for medical applications.

Qiu et al. [12] demonstrated the efficacy of promptly discerning stroke risk through the utilization of five classifiers and 12 clinical characteristics. Random Forest and XG-Boost demonstrated significant predictive capabilities, identifying a sedentary lifestyle as a primary predictor. However, it is crucial to acknowledge certain limitations in the study, including the inability to differentiate between ischemic and haemorrhagic strokes. Also, findings may have limited generalizability given they were derived from a region in China.

Lumley et al. [13] developed a 5-year stroke prediction model using the Cox proportional hazard model, yielding commendable results. Concerns arose about its reliance on preselected features and sensitivity to variable selection. Similarly, Khosla et al. [14] predicted stroke outcomes using the Cox model with a novel feature selection algorithm combined with a Support Vector Machine (SVM), showing promise but faced challenges with model performance. Dritsas and Trigka [15] study conducted a comprehensive evaluation of seven machine learning models, including stacking, random forest, majority voting, and 3-NN, for stroke prediction. Stacking and random forest exhibited superior performance, with both models achieving AUC and Accuracy values of 98% and 97%, respectively, indicating strong discrimination capability. Nevertheless, the study is subject to limitations concerning optimisation, cross-validation, missing data handling, and model interpretability assessments, as well as the need for further validation of real clinical data before these models could be clinically useful.

The literature review sheds light on several studies utilising machine learning to analyse risk factors contributing to stroke occurrence. However, a challenge persists due to the multitude of risk factors associated with stroke. To enhance the prediction performance, it is crucial to carefully select relevant risk factors, eliminate misleading and overfitting attributes and reduce data noise. Zihni et al. [16] research is a pronounced recognition of the indispensability of XAI in the development of Clinical Decision Support Systems (CDSS). This integration aims to enhance the effectiveness and efficiency of medical diagnosis and treatment. The synergistic fusion of XAI and CDSS not only furnishes clinicians with a heightened understanding but also simplifies intricate models, thereby benefiting both healthcare practitioners and patients.

III. Methodology

This section elucidates the methodology and approach employed to attain the research objective. The study unfolded across three pivotal phases: Data Profiling, Machine Learning Algorithms, and Explainable Artificial Intelligence, each furnishing unique perspectives on stroke risk prediction. The research flow chart as shown in Figure 1 illustrates methodical steps and systematic progression used in the research.

Fig. 1. Research Process Flow Chart

A. Data Description

The stroke dataset utilised in this study is open-sourced from Kaggle [17]. The data contains numeric and categorical variables with 5110 rows and 11 columns. The target variable of the dataset (i.e., 'stroke') shows the presence ('1'- 249) or absence ('0'- 4861) of a stroke. Thus, indicating that 95% of patients had 'no-stroke' and 5% were 'stroke', thereby creating a data imbalance leading to algorithm bias. To rectify this imbalance and augment the credibility of results for predictive models, data pre-processing techniques, such as the Synthetic Minority Over-sampling Technique (SMOTE), are implemented in the training set to increase and equalize under-represented classes. The incorporation of synthetic data through the SMOTE technique substantially contributes to attaining a balanced distribution of classes, thereby augmenting the model's capacity to effectively learn from the minority class and mitigating prediction bias thereby enhancing the overall predictive capabilities of machine learning models [18]. Table I describes the various attributes used for stroke risk prediction in the dataset.

B. Results and Discussion of Findings

A comprehensive evaluation has been undertaken to appraise the efficacy of six distinct models, utilizing a diverse set of performance metrics including accuracy, precision, recall, F1 score, and Area Under the ROC Curve (AUC-ROC). Recognizing that reliance solely on accuracy as an evaluation metric can be misleading, particularly in the context of imbalanced datasets. The imbalanced data, with "No-Stroke" instances outweighing "Stroke", poses a challenge, as high accuracy may not fully represent the model's true capabilities. Precision signifies the proportion of accurately classified stroke cases among all predicted positives, while recall indicates the proportion of correctly predicted strokes among all actual positives. As the harmonic mean, the F1-score furnishes a holistic measure of predictive performance. AUC-ROC assesses classification accuracy by comparing predicted probabilities to actual outcomes, measuring the balance between false positives and true positives, with values ranging from 0 to 1. This multifaceted approach ensures a balanced assessment of the model's effectiveness while acknowledging and mitigating the challenges of imbalanced data.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{F-Measure} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{2}$$

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{AUC\_ROC} = \int_0^1 \text{TPR}\, d(\text{FPR}) \tag{5}$$

C. Evaluation of the Predictive Models

Considering the skewed class distribution in the dataset, utilising accuracy alone as an evaluation metric introduces a risk for potential bias, which may not precisely reflect an accurate model performance. Consequently, Recall, F1-score, Precision, and AUC-ROC were incorporated to buffer accuracy as key evaluation metrics for the assessment of the model behaviour. However, before employing SMOTE the machine learning models exhibited significant class imbalance, this is reflected by the low recall, F1-scores, and precision values shown in Table II. These stems directly from the bias towards the majority class in the heavily skewed dataset. The integration of synthetic minority class data via SMOTE significantly contributes to achieving a more balanced class distribution, thereby enhancing the models' capacity to effectively learn from minority 'stroke' cases and mitigating prediction bias. Table II illustrates the performance of various machine learning models before the application of SMOTE.

Furthermore, addressing the class imbalance with SMOTE improves the overall predictive capabilities of the machine learning models, as shown by the improved metrics illustrated in Table III. Post-SMOTE retraining, these models exhibited substantial improvements across all metrics compared to pre-SMOTE results. The synthesis of additional stroke cases through SMOTE is pivotal in overcoming the dominance of the majority "non-stroke" class, enabling the models to proficiently identify critical minority positive cases. Thus highlighting the significant positive impact of employing balancing techniques such as SMOTE on the performance of machine learning models. Through the synthesis of additional minority class instances, SMOTE augments the models' capacity to discern predictive patterns related to stroke risk, thereby bolstering their proficiency in distinguishing between 'Stroke' and 'Non-stroke' outcomes. This advancement enhances the models' reliability and applicability within the healthcare domain. The Random Forest Classifier and XG-Boost exhibited superior performance across multiple metrics such as Accuracy, AUC-ROC, F1-score, and Precision as illustrated in Table III. However, Naïve Bayes demonstrates a higher recall for stroke prediction, with a value of 0.99 compared to 0.96 for Random Forest Classifier and XG-Boost. Despite Naive Bayes' leading recall for stroke prediction, Random Forest Classifier and XG-Boost outperform other models in precision and F1 scores, suggesting their efficacy in making highly

TABLE I
DESCRIPTION OF ATTRIBUTES

Attribute	Description
GENDER	Specifies the gender of the patient ({'Male':0,'Female':1})
AGE	Age of the Patient
HYPERTENSION	Indicates the presence or absence of hypertension ({'No':0,'Yes':1})
HEART DISEASE	Indicates the presence or absence of heart disease ({'No':0,'Yes':1})
EVER MARRIED	Indicates whether the patients are married or not ({'No':0,'Yes':1})
WORK TYPE	Indicates the type of employment of the patient ({'Private':0,'Self-employed':1,'Govt_job':2,'children':-1,'Never_worked':-2})
RESIDENCE TYPE	Indicates the type of residence where the patient lives ({'Rural':0,'Urban':1})
AVERAGE GLUCOSE LEVEL	Indicates the patient's average blood glucose levels.
BODY MASS INDEX	Indicates the patient's body mass index
SMOKING STATUS	Indicates the patient's smoking status ({'never smoked':0,'smokes':1,'formerly smoked':-1,'Unknown':-2})
STROKE	Presence (1) or Absence (0)

TABLE II
SUMMARY OF MACHINE MODELS PERFORMANCE PRE-SMOTE APPLICATION

Results	Random Forest	Support Vector Machine	Logistic Regression	Decision Tree	Naïve Bayes	XG Boost
Accuracy	**94.0%**	93.9%	93.9%	91.5%	84.1%	93.7%
AUC-ROC	0.51	0.50	0.50	0.58	**0.70**	0.54
Recall	0.016	0.0	0.0	0.19	**0.55**	0.10
F1 Score	0.032	0.0	0.0	0.22	**0.29**	0.16
Precision	**1.0**	0.0	0.0	0.25	0.20	0.43

TABLE III
SUMMARY OF MACHINE MODELS PERFORMANCE POST-SMOTE APPLICATION

Results	Random Forest	Support Vector Machine	Logistic Regression	Decision Tree	Naïve Bayes	XG Boost
Accuracy	**94.5%**	82.7%	80.6%	90.5%	70.9%	94.4%
AUC-ROC	**0.95**	0.83	0.81	0.91	0.71	0.94
Recall	0.96	0.89	0.84	0.91	**0.99**	0.96
F1 Score	**0.95**	0.84	0.81	0.91	0.77	**0.95**
Precision	**0.93**	0.79	0.79	0.90	0.63	**0.93**

accurate positive predictions. Additionally, the Random Forest Classifier achieves the highest accuracy rate, highlighting its proficiency in identifying many positive instances correctly. Despite a slightly lower F1-score and AUC-PR, Decision Tree still exhibits commendable performance in accurate stroke predictions.

However, compared to previous studies, a study conducted by Dritsas and Trigka [15] comprehensively assessed seven machine learning models, including ensemble models such as stacking, random forest, majority voting, and k-nearest neighbours for their effectiveness in predicting strokes. The Stacking ensemble model, featured carefully selected base models that exhibited superior performance across all evaluation metrics, including AUC, F-measure, precision, recall, and accuracy, when compared to alternative models such as random forest, 3-NN, and Decision Tree. Specifically, the stacking ensemble achieved top-tier results of 99% AUC, 97% F-measure, precision, recall, and 98% accuracy, indicating exceptional discrimination between 'Stroke' and 'Non-stroke' classes. Although Random Forest also demonstrated strong performance, it marginally trailed behind the Stacking ensemble across the metrics. The high AUC values, approximately 99%, attained by both Stacking and Random Forest showcase their robust predictive capabilities in distinguishing stroke cases. Despite the comparatively lower performance, the 3-NN and Decision Tree models still showcased utility for stroke prediction, as evidenced by their reasonably high AUC and F-measure scores. Nevertheless, the study is subject to limitations concerning optimization, cross-validation, missing data handling, and model interpretability assessments.

D. Application of XAI

This section examines the practical implementation of XAI techniques, specifically LIME and ELI5, to elucidate the critical factors driving stroke risk prediction. The application of XAI serves to validate the basis for the model's projected appraisal of a patient's stroke probability. The subsequent tables delineate the LIME and ELI5 outcomes for all models, utilising sample patient data for evaluation purposes. Table IV provides a concise summary of the consistent results obtained from LIME and ELI5 analyses across all models' predictors for stroke risk.

1) Local Interpretable Model-agonistic Explanation (LIME): A LIME plot elucidates the impact of individual features on predicting stroke risk, thereby augmenting comprehension of the model's decision-making process. This visual representation utilises a diverging bar chart

TABLE IV
SUMMARY OF KEY CONTRIBUTORS TO STROKE RISK USING XAI

Machine Learning Models	LIME Features	ELI5 Features
Random Forrest	Age Marital Status Nature of Employment	Age = 39% Smoking Status = 19.2% Marital Status = 14.75% Average Glucose Level = 7.7% Type of Residence = 5%
Logistic Regression	Age Marital Status Average Glucose Level Nature of Employment BMI	Age = 22% Smoking Status = 3.5% Average Glucose Level = 2.2%
Decision Tree	Age Marital Status Nature of Employment	Age = 44.4% Smoking Status = 21.3% Marital Status = 13.1% Average Glucose Level= 4.8% Type of Residence= 4.1%
Naïve Bayes	Age Average Glucose Level	
XG-Boost	Age Marital Status Nature of Employment	Age = 21% Nature of Employment = 12.6% Heart Disease = 12.3% Gender = 10.3% Average Glucose Level = 9.9%

format, with the most influential features for predicting stroke positioned on the left and those for 'Non-stroke' on the right. The length of the bars, denoting higher values, corresponds to the level of importance ascribed to the features in the predictive model as shown in Figure 2. Figure 2 illustrates a visual representation of key factors contributing to stroke risk using a LIME plot.

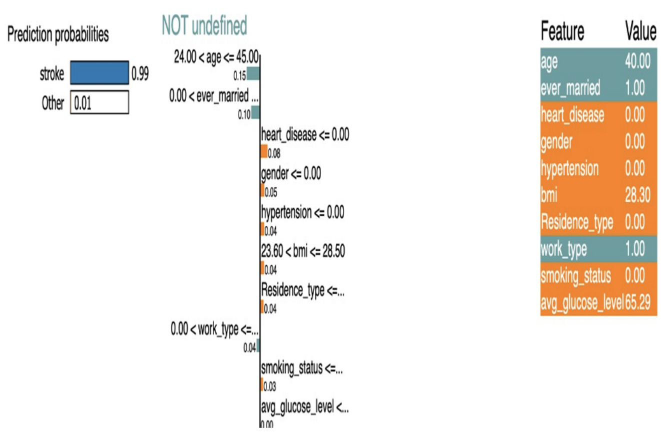
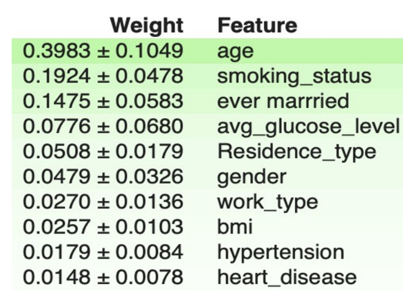


Fig. 2. Visual Representation of a LIME Plot

The analysis of the LIME plots across machine learning models demonstrates that "Age" is a predominant factor in determining stroke risk, suggesting a positive correlation between advancing age and stroke probability. While Age consistently exhibited significant influence, other attributes like Glucose level, BMI, Marital status, and Employment type also contributed to stroke prediction across models, though with differing extents of impact. Specifically, for Logistic Regression, key predictors were Age, Marital status, Work type, Glucose level, and BMI, while Random Forest and XGBoost highlighted age, marital status and work type. The Naive Bayes model showed age and glucose level as most dominant, though with stronger predictive power for non-stroke cases. Decision Tree again emphasized age, marriage, and work type. In addition to illuminating the most influential features, LIME plots also revealed a nuanced interplay between diverse patient risk factors highlighting the complexity inherent in stroke risk prediction. Some attributes like 'Gender', 'Hypertension', 'Heart Disease', and 'Smoking Status' seemingly showed reduced predictability for stroke likelihood across models.

2) Explain Like I'm Five (ELI5): The ELI5 technique is strategically utilised to augment the accessibility and comprehension of weighted factors produced by machine learning algorithms for stroke risk prediction. The ELI5 analysis aims to reflect the critical importance of certain features in the hierarchical structure using the permutation importance of relevant weighted factors. However, among the examined algorithms; Logistic Regression, Decision Tree, Random Forest, and XGBoost - there is a distinct emphasis on the significance of specific integrated factors as shown in Figure 3. This reflects the pivotal role of these elements in ensuring accurate stroke risk prediction. Figure 3 illustrates a visual representation of key features and their weighted permutation importance to stroke risk using an ELI5 plot.



Weight	Feature
0.3983 ± 0.1049	age
0.1924 ± 0.0478	smoking_status
0.1475 ± 0.0583	ever marrried
0.0776 ± 0.0680	avg_glucose_level
0.0508 ± 0.0179	Residence_type
0.0479 ± 0.0326	gender
0.0270 ± 0.0136	work_type
0.0257 ± 0.0103	bmi
0.0179 ± 0.0084	hypertension
0.0148 ± 0.0078	heart_disease

Fig. 3. Visual Representation of an ELI5 Plot

The ELI5 and LIME analysis of feature importance across the Random Forest, Decision Tree, XGBoost, and Logistic Regression models consistently identify Age as the most influential factor for stroke prediction, with importance scores ranging from 22% to 44.4% across models. Other top variables include 'Smoking status', 'Ever married status', 'Average Glucose level', 'Work type', 'Heart Disease', 'Gender', and 'Hypertension' though their significance varies between models. Notably, 'Heart disease and Hypertension' have low importance in Random Forest and Decision Tree models. The consistent emergence of Age as the dominant predictor aligns with the insight that stroke risk rises with increasing age. While specific variables differ in importance between models, ELI5 and LIME collectively highlight Age, Smoking, Marital Status, Glucose, Work Type, Heart Disease, Gender, and Hypertension as key factors influencing stroke likelihood predictions. The prominence of these features emphasizes their significance and roles within the machine learning models for assessing a patient's cerebrovascular event risk. These techniques not only identified but quantified the proportional contributions of specific features, providing valuable insights into the key factors that impact the likelihood of both stroke and non-stroke occurrences. Understanding these insights enhances the model's decision-making process, empowering informed decision-making and precise interventions by medical professionals in the realm of stroke prevention and management.

IV. Conclusion and Recommendation

Predicting stroke risk remains a critical health concern for healthcare professionals given the significant implications for mortality and morbidity. This work enhances the timely identification of at-risk individuals, enabling prompt interventions before exacerbation. The study acknowledges the potential of machine learning algorithms and XAI techniques like LIME and ELI5 for predicting stroke risk by uncovering pivotal influential features for each model. A valuable recommendation based on the findings would be integrating patient CT scan image data. Deep learning has shown promise in capturing intricate patterns for stroke risk prediction. Strategic integration of brain CT scan images into EMR could:

- Harness the delicate structural and spatial information inherent in these images by incorporating this rich dataset into deep learning models.
- Enable comprehensive and multi-dimensional understanding of stroke risk, as these models may unveil subtle, yet substantial correlations between image patterns and the probability of stroke occurrence.

A notable constraint within this study is rooted in limitations associated with the dataset. Although the dataset utilised for analysis is diverse, it lacks the depth and intricacy typically inherent in comprehensive hospital records. The augmentation of more expansive records, particularly incorporating patient imaging results, holds the potential to significantly enhance the efficacy of analytical models aimed at evaluating stroke risk. Another constraint of the study is that the data is specific to a geographic region and therefore the observed outcome may not be accurate in another geographic region. Therefore, data from other regions would be useful in expanding the scope of the study; however, this would be time and labour-intensive. Further challenges of the study are about privacy and ethical issues in accessing medical records.

References

[1] W. Johnson, O. Onuma, M. Owolabi, and S. Sachdev, "Stroke: a global response is needed," *Bulletin of the World Health Organization*, vol. 94, no. 9, p. 634, 2016.

[2] W. H. Organization *et al.*, "Global status report on noncommunicable diseases 2014," no. WHO/NMH/NVI/15.1, p. 265, 2014.

[3] V. L. Feigin, M. H. Forouzanfar, R. Krishnamurthi, G. A. Mensah, M. Connor, D. A. Bennett, A. E. Moran, R. L. Sacco, L. Anderson, T. Truelsen *et al.*, "Global and regional burden of stroke during 1990–2010: findings from the global burden of disease study 2010," *The lancet*, vol. 383, no. 9913, pp. 245–255, 2014.

[4] D. King, R. Wittenberg, A. Patel, Z. Quayyum, V. Berdunov, and M. Knapp, "The future incidence, prevalence and costs of stroke in the uk," *Age and ageing*, vol. 49, no. 2, pp. 277–282, 2020.

[5] A. Patel, V. Berdunov, Z. Quayyum, D. King, M. Knapp, and R. Wittenberg, "Estimated societal costs of stroke in the uk based on a discrete event simulation," *Age and ageing*, vol. 49, no. 2, pp. 270–276, 2020.

[6] S. C. Johnston, S. Mendis, and C. D. Mathers, "Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling," *The Lancet Neurology*, vol. 8, no. 4, pp. 345–354, 2009.

[7] M. A. Gianfrancesco, S. Tamang, J. Yazdany, and G. Schmajuk, "Potential biases in machine learning algorithms using electronic health record data," *JAMA internal medicine*, vol. 178, no. 11, pp. 1544–1547, 2018.

[8] G. Sivapalan, K. K. Nundy, S. Dev, B. Cardiff, and D. John, "Annet: A lightweight neural network for ecg anomaly detection in iot edge sensors," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 16, no. 1, pp. 24–35, 2022.

[9] D. Schönberger, "Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications," *International Journal of Law and Information Technology*, vol. 27, no. 2, pp. 171–203, 2019.

[10] R. Islam, S. Debnath, and T. I. Palash, "Predictive analysis for risk of stroke using machine learning techniques," in *2021 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*. IEEE, 2021, pp. 1–4.

[11] S. Kawakura, M. Hirafuji, S. Ninomiya, and R. Shibasaki, "Adaptations of explainable artificial intelligence (xai) to agricultural data models with eli5, pdpbox, and skater using diverse agricultural worker data," *European Journal of Artificial Intelligence and Machine Learning*, vol. 1, no. 3, pp. 27–34, 2022.

[12] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan *et al.*, "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–33, 2023.

[13] C. Kim, V. Zhu, J. Obeid, and L. Lenert, "Natural language processing and machine learning algorithm to identify brain mri reports with acute ischemic stroke," *PloS one*, vol. 14, no. 2, p. e0212778, 2019.

[14] M. Monteiro, A. C. Fonseca, A. T. Freitas, T. P. e Melo, A. P. Francisco, J. M. Ferro, and A. L. Oliveira, "Using machine learning to improve the prediction of functional outcome in ischemic stroke patients," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 15, no. 6, pp. 1953–1959, 2018.

[15] Y. Qiu, S. Cheng, Y. Wu, W. Yan, S. Hu, Y. Chen, Y. Xu, X. Chen, J. Yang, X. Chen *et al.*, "Development of rapid and effective risk prediction models for stroke in the chinese population: a cross-sectional study," *BMJ open*, vol. 13, no. 3, p. e068045, 2023.

[16] T. Lumley, R. A. Kronmal, M. Cushman, T. A. Manolio, and S. Goldstein, "A stroke prediction score in the elderly: validation and web-based application," *Journal of clinical epidemiology*, vol. 55, no. 2, pp. 129–136, 2002.

[17] A. Khosla, Y. Cao, C. C.-Y. Lin, H.-K. Chiu, J. Hu, and H. Lee, "An integrated machine learning approach to stroke prediction," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2010, pp. 183–192.

[18] E. Dritsas and M. Trigka, "Stroke risk prediction with machine learning techniques," *Sensors*, vol. 22, no. 13, p. 4670, 2022.