

# ARTIFICIAL INTELLIGENCE IN IMPROVING STROKE DIAGNOSIS: FOCUS ON MACHINE LEARNING MODELS AND EXPLAINABLE AI APPLICATION

Natacha Usanase<sup>1,2,\*</sup>, Consolée Uwamahoro<sup>3</sup> and Dilber Uzun Ozsahin<sup>4,5,1</sup>

<sup>1</sup>Operational Research Center in Healthcare, Near East University, Nicosia/TRNC, Turkey

<sup>2</sup>Department of Biomedical Engineering, Near East University, Nicosia, Turkey

<sup>3</sup>Department of Biomedical Laboratory Sciences, Faculty of Applied Fundamental Sciences, Ines-Ruhengeri, Rwanda

<sup>4</sup>Department of Medical Diagnostic Imaging, College of Health Science, University of Sharjah, Sharjah, UAE

<sup>5</sup>Research Institute for Medical and Health Sciences, University of Sharjah, Sharjah, UAE

Received: 07.10.2025.

Accepted: 12.11.2025.

## Corresponding author:

### Natacha Usanase

Operational Research Center in Healthcare, Near East University, Department of Biomedical Engineering, Near East University, Nicosia, Turkey

E-mail: natacha.usanase@neu.edu.tr

## ABSTRACT

Stroke is a major and deadly health concern on a global scale, requiring fast and precise methods for effective management. The current research explores six machine learning models: logistic regression, k-nearest neighbors (KNN), support vector machine (SVM), decision tree, random forest, and eXtreme Gradient Boosting (XGBoost), to improve stroke diagnosis. By applying Local Interpretable Model-agnostic Explanations (LIME), this work bridges the gap of interpretability in conventional machine learning models, making it easier for healthcare experts to understand generated model predictions. 5,109 clinical cases with features including age, gender, hypertension, heart disease, average glucose level, and details of patient lifestyle as risk variables, were used to train the applied algorithms. As a result, Logistic Regression had an accuracy of 77%, whereas KNN and SVM had accuracies of 92% and 89%, respectively. The decision tree classifier achieved high precision and accuracy of 95%; however, the random forest and XGBoost models achieved the highest accuracy (97%) and AUC (99%), respectively, outperforming all applied classifiers. The importance of various attributes for each prediction was assessed using LIME, supporting a clear and transparent understanding of the model predictions. Case-based analyses revealed that age, gender, BMI, average glucose level, as well as stressful lifestyle conditions were the major risk factors for stroke. This study highlights the importance of explainable artificial intelligence in assisting healthcare professionals and offering transparent, reliable, and effective personalized treatments relative to specific patient needs.

**Keywords:** Stroke, machine learning, data analysis, diagnosis, prediction, LIME.

UDK: 616.831-005.1:004.891

Eabr 2025; 26(4):391-397

DOI: 10.2478/eabr-2025-0023

## INTRODUCTION

Globally, stroke is among the most common causes of death and long-term disability, whereby each year, it affects over thirteen million people and kills over five million people [1]. A stroke is a neurological condition characterized by blood vessel obstruction, also known as a brain clot, or the bursting of a blood vessel, causing bleeding, damage to brain cells, or even death [1]. It is mainly classified into two primary categories, which are ischemic stroke, the most prevalent form of stroke, accounting for 87% of all stroke incidents [2]. The prevalence of ischemic infarctions grew significantly between 1990 and 2016 [3]. In this category, a portion of the brain cannot receive blood or oxygen because of a blood clot [1]. Secondly, a hemorrhagic stroke, this type of stroke results when a blood vessel ruptures, frequently caused by aneurysms or venous abnormalities [4].

While prevention relies on managing risk factors, identifying stroke symptoms is essential for early diagnosis, urgent triage, and immediate medical support. The symptoms include stuttering or the prodromal symptoms of basilar artery blockage, especially in people with intracranial artery blockage [5]. Furthermore, stroke incidence has been reported to rise with age, whereby after age 55, the incidence of stroke doubles. Between 1990 and 2016, the percentage of stroke cases worldwide among adults aged 20 to 54 surged from 12.9% to 18.6%. Conversely, throughout the same period, age-standardized related mortality rates lowered by 36.2% [5]. Although stroke is controllable and treatable, there is a high possibility of significantly reducing its prevalence as well as its long-term effects, because according to predictions, stroke continues to be one of the top causes of mortality and impairment [6]. Therefore, it is crucial to focus on early detection and prevention to control the severity of this condition. Frequently, traditional risk assessments depend on parameters like clinical records and epidemiology. Still, these do not cover how several aspects of life and health contribute to the cause of a stroke. For this reason, robust data-based strategies that consider diverse factors such as demographics, clinical, and lifestyle factors are crucial to identify early signs of stroke, thereby effectively assessing possible risks to improve its diagnosis and treatment as well as minimize its mortality rates.

Over the last couple of years, machine learning (ML), an artificial intelligence (AI) subfield, has emerged as a powerful tool in clinical decision support systems. ML algorithms are capable of recognizing unknown patterns in large datasets and complex health conditions, and they offer predictive approaches that can outperform traditional statistical techniques. Such algorithms help to accurately determine the risk of diseases based on complex multidimensional data, such as demographic, radiological, omics, clinical records, and lifestyle-related metrics [7]. However, regardless of their predictive strength, most ML algorithms lack the explainability concerning how input parameters affect the model output. This gap causes challenges to healthcare facilities, where transparency, trust, and accountability play significant roles.

As a result, to address such issues, explainable ML models, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), have been developed to balance model complexity and model interpretability [8]. These techniques provide explanations of individual predictions by modeling the behavior of the system over every input variable, thus allowing physicians to understand the specific features impacting a prognosis, making doctors more informed before making a clinical decision, and gaining confidence in AI-driven systems [9].

The application of explainable AI models in stroke prediction is still a developing field. Recent studies have applied decision trees, random forests, support vector machines (SVM), and deep learning algorithms in predicting stroke occurrences based on electronic health records and structured data [10], [11], [12]. Bentley et al. [13] assessed the performance of SVM for the prediction of acute ischemic stroke, whereby 116 instances were considered. The training dataset was composed of 106 instances, whereas the testing dataset was composed of 10 instances. From their findings, the area under the curve (AUC) of SVM was found to be 0.744, which showed a better performance than the compared prognostic scores. Furthermore, Han et al. [14] employed ML methods to build a classification model that is more accurate in predicting short-term stroke probability than the conventional scores. The optimum result in the testing set was accomplished by ensemble algorithms, random forest, and a convolutional neural network (CNN), which outperformed CHA2DS2-VASc by up to 0.14 in the AUC. These results demonstrate that the atrial fibrillation burden signature technique can add benefits for risk prediction, particularly for the short-term risk of stroke. AI has played an essential role in areas like radiology, pathology, and genetics, and has also reduced errors and allowed more personalized treatment [15], [16], [17]. With this, the incorporation of AI in the healthcare field is revolutionizing diagnostics, treatment, and overall disease management.

Stroke management is demanding and advanced, thus requiring a new technology that may support a quick decision-making process and personalized treatment plans. Though quite effective, the traditional methods of diagnosing and treating strokes are constrained by the availability of specialists and the speed at which necessary information is processed. ML offers the possibility to overcome such limitations by enhancing the accuracy of radiological images, predictive power, and tailoring treatment plans to the individual patient level in a short period. Even though these models hold some potential concerning their ability to predict with high sensitivity and specificity, they still tend to lack the ability to explain specific results. The current study aims to perform stroke risk prediction using six ML algorithms (logistic regression, k-nearest neighbors (KNN), SVM, decision trees, random forests, and extreme gradient boosting (XGBoost)) and incorporating LIME to generate local predictions of different stroke cases and interpret the effects of each variable (gender, age, hypertension, heart disease, marital status, work type, residence type, average glucose level in blood, BMI

(body mass index), smoking status) on stroke cases. This is especially needed in healthcare, where patients, their families, and practitioners have to understand the outputs of a model, especially in life-threatening scenarios. Moreover, LIME provides a more accurate evaluation of risk factors in diverse populations when integrated with ML models, as heterogeneous features can be easily explained using this approach. The creation of robust models depending on specific pathophysiology is essential for directing diagnosis and treatment, as well as setting reasonable expectations for patients, considering the multitude of parameters involved in the decision system and their divergent connections with the final result [18].

The remaining parts of the manuscript are arranged as follows: the 2<sup>nd</sup> part elaborates more on the methodology applied as well as the dataset used in this study, the 3<sup>rd</sup> part provides the results generated after applying the considered ML models, which are then discussed in the 4<sup>th</sup> part of the text, and the whole study is concluded in part 5.

## MATERIALS AND METHODS

The current study applied a data-driven approach to develop prediction tools for diagnosing stroke using a publicly available dataset (<https://www.kaggle.com/datasets/fedoriano/stroke-prediction-dataset>) from Kaggle. The dataset consists of 5,109 cases of patients with information on their age, gender, medical history (hypertension and heart disease), lifestyle habits (smoking and type of work), and clinical test records (glucose level and BMI), with a binary value for the target variable (stroke) indicating if the patient had a stroke (1) or not (0). Ethical approval was not required because the data is anonymized and available online for research purposes.

The data was preprocessed using basic analytical methods to confirm its quality and the possibility of using it in ML processes. This analysis included checking the dataset for missing data values, replications, and clinically irrelevant instances. Consequently, some data points for BMI were missing and were thus replaced using the median value found within the range. There were no duplicate patient entries; thus, each entry represented a unique patient case. All the categorical variables were encoded to ensure a good fit with the ML models while maintaining the original information so as not to affect the performance of the algorithms. As a result, binary variables, such as “gender”, “ever married”, “Residence type”, were encoded using the one-hot encoder technique into 0 and 1, whereas all multiclass categorical variables in the data, including “work type”, “smoking status”, were processed using the label encoding technique. The converted dataset consists only of numbers, which allows for easier application of ML tools.

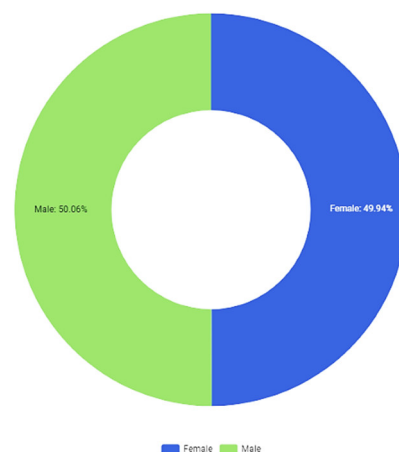
The data was split into training and testing sets in an 80:20 ratio after preprocessing. With the class imbalance in the dataset, 4,860 patients without stroke and only 249 patients with stroke, it is essential to use a strong resampling

process to ensure equal representation of each class during training as such a situation may cause classifiers to give more weight to the majority and ignore the minority class, which can be less effective clinically since patients with stroke are the minority group in the dataset used. As a result, the synthetic minority over-sampling technique (SMOTE), which helps increase the number of instances from the minority category, was applied.

SMOTE fills in the gaps between minority class instances and their nearest neighbors in the feature space by generating new samples to make the boundary between classes more distinct. With SMOTE, the non-stroke and stroke classes were made equal, leading to a 1:1 class balance. Various ML models, including logistic regression, KNN, SVM, decision trees, random forests, and XGBoost, were tested to predict stroke diagnostic cases. Furthermore, the measures used in the evaluation of the models were accuracy, precision, recall, F1-score, and area under the curve (AUC). LIME, along with medical knowledge from health experts, was applied to select the important features and confirm their relevant contribution to the model outcomes.

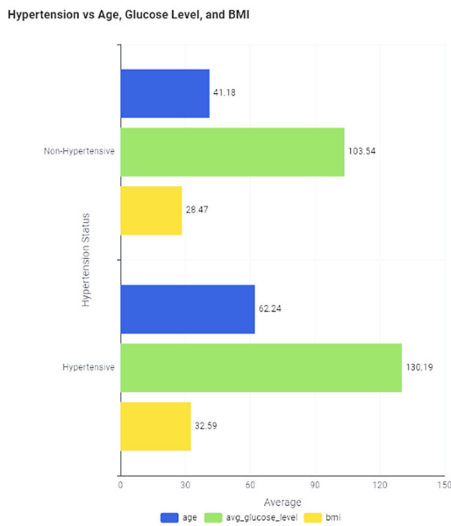
## RESULTS

Exploratory data analysis (EDA) was performed on a dataset of 5,109 people to examine how important stroke risk-related clinical factors were distributed. The average age of the patients was found to be  $43 \pm 23$  years, and the glucose level test showed an average of  $106.1 \pm 45.3$  mg/dL, pointing to a higher than average glucose reading. In addition, the BMI variable had an average of  $28.9 \pm 7.7$  kg/m<sup>2</sup>, which is overweight for the population, and a few outliers showed severe obesity. Figure 1 displays a donut plot illustrating a roughly even gender breakdown in the dataset (Male: 50.06%, Female: 49.94%). Gender equilibrium is important for model training stability and plausibility, reducing the chance of gender bias that is likely to affect the prediction. This distribution ensures that the results and the used models can be applied across genders, adding to the dependability of the current study.



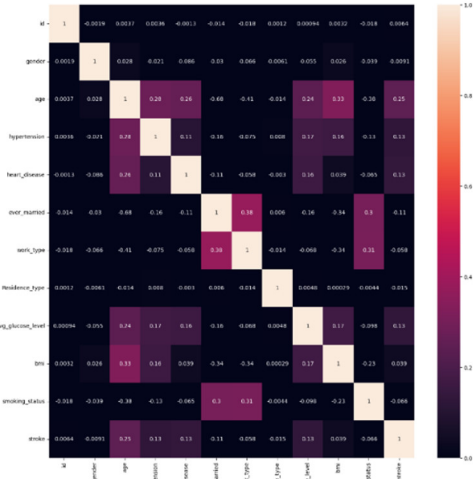
**Figure 1.** The gender distribution in the stroke dataset

Figure 2 compares people with and without hypertension based on three stroke risk factors: age, average glucose level, and BMI. Approximately, the data showed that patients with hypertension are 62 years old on average, while those who do not have hypertension are 41 years old on average, which aligns with the relationship between ageing and hypertension risk reported in previous studies [19]. Furthermore, the majority of hypertensive patients had higher BMI (32.59 kg/m<sup>2</sup>), which is classified as obese, and was associated with higher average glucose level (130.19 mg/dL) compared to non-hypertensive cases (glucose levels 103.54 mg/dL, BMI = 28.47 kg/m<sup>2</sup>). Clinically, patients having a high BMI with hypertension may have poor cardiovascular health [20], which is why different risk factors should be examined when assessing stroke risk [21]. By referring to these descriptive data, we gained basic knowledge about the population within the dataset, which was later explored through ML modeling.



**Figure 2.** Bar chart illustrating the association between hypertension, age, glucose level, and BMI

Figure 3 shows a heatmap matrix that clearly illustrates the relationship between demographic and health factors associated with stroke diagnosis. The matrix displays a Pearson correlation coefficient for each pair of variables, ranging from -1 (a strong negative correlation) to 1 (a strong positive correlation). A strong positive relationship is observed between "age" and "BMI" ( $r=0.33$ ), and "ever married" and "work type" ( $r=0.38$ ). Likewise, a strong negative correlation is seen between "work type" and "age" ( $r=-0.41$ ) as well as "age" and "ever married" ( $r=-0.68$ ). The small correlation coefficients indicate that there is not much multicollinearity among features, which is likely to speed up training and reduce the risk of models learning from noise, which is likely to lead to overfitting. Identifying how different variables are linked is vital during feature selection, especially for stroke prediction, because related variables can influence the overall performance of the models as well as their interpretability.



**Figure 3.** Correlation Matrix of the features within the stroke dataset

Different models were used and tested using an 80% training and 20% test set, as well as SMOTE to address the unequal distribution of classes, for predicting stroke outcomes. Logistic Regression achieved an accuracy of 77%, with precision, recall, AUC, and F1-scores of 76%, 81%, 85%, and 78%, respectively. KNN and SVM were more accurate than Logistic Regression, with KNN reaching 92% accuracy and 87% precision, and SVM achieving 89% accuracy and a precision of 90%. Furthermore, the decision tree model had an accuracy rate of 95%, a precision and recall of 95% and 94%, respectively. However, even though KNN achieved the highest recall score (refer to Table 1), both Random Forest and XGBoost ensemble models outperformed all the algorithms with a 97% accuracy.

**Table 1.** The overall performance of the six applied ML models

Models	Accu racy	Precis ion	Recall	F1	AUC
Logistic Regression	0.775	0.757	0.809	0.782	0.846
KNN	0.920	0.874	0.981	0.925	0.973
SVM	0.894	0.906	0.878	0.892	0.968
Decision Tree	0.947	0.952	0.941	0.947	0.947
Random Forest	0.970	0.992	0.947	0.969	0.994
XGBoost	0.970	0.988	0.952	0.970	0.992

As illustrated in Figure 4, the applied ML methods show strong performance on the ROC curve, with AUC ranging from 0.85 to 0.99. The XGBoost and random forest, being the best classifiers, have an AUC of 0.99, followed by KNN and SVM with an AUC of 0.97. All the curves located in the top-left part signify that each model is highly sensitive in detecting false positives by effectively classifying stroke patients from non-stroke patients. Clinically, tree-based ensemble approaches such as XGBoost and random forest can

detect complex, non-linear relationships [22], [23]. Most of the AUC values are fairly high, suggesting that the used algorithms can be effective in improving clinical decisions for assessing stroke risk.

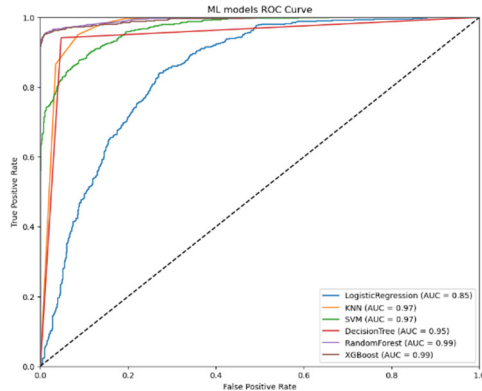
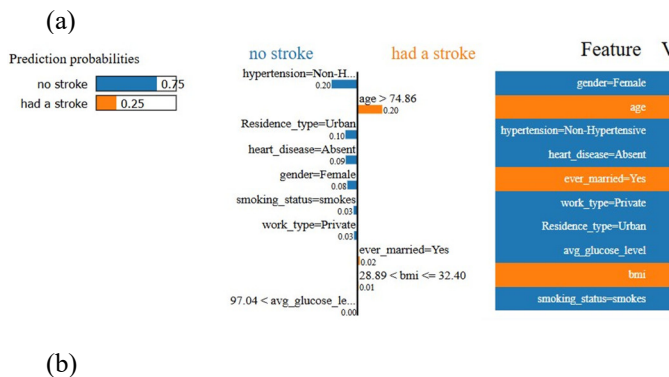


Figure 4. The Area Under the Curve for all applied models for stroke prediction

Using LIME during stroke classification helps discover how specific features affect the model predictions. As illustrated in Fig. 5. (a), an obese female patient of 77 years old with a 25% probability of having a stroke demonstrates that some factors have a high chance of lowering the risk of having a stroke. Even though the patient is old and smokes, and has a high BMI of 31.0 kg/m<sup>2</sup>, LIME highlights that in such a case, it is unlikely to have a stroke when associated with having normal blood pressure, absence of heart diseases, as well as a normal lifestyle. On the other hand, Fig. 5. (b), an 80-year-old male patient has a 99% probability of having a stroke. Though he was non-hypertensive, it was found that advanced age, gender (being a male patient), BMI, and a severe spike in glucose level were the major reasons behind the increased risk, as well as being self-employed, which is likely to trigger stress. According to these results, age, gender, BMI, and glucose level majorly increase stroke risk, especially when associated with a complex lifestyle that prompts stress. Additional health conditions and lifestyle factors that are not too demanding mitigate the risk, suggesting several variables can interact and impact the overall stroke prediction.



(b)

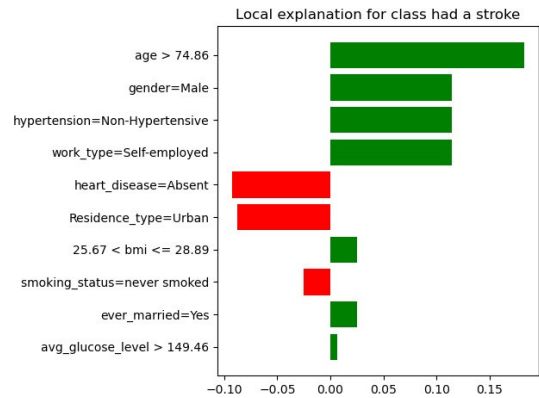


Figure 5: LIME results of two different scenarios (patient at low risk vs patient at high risk) and the factors influencing the outcome. The higher the absolute amount of weight, the more significant the impact on the projected result.

### Discussion

Stroke prediction using ML has been quite promising, especially in situations where early detection and sorting out higher-risk patients can impact the results of treatment. The use of ML methods on structured health records revealed that using ensemble models, like random forest and XGBoost, effectively delivers accurate predictions than single models in stroke detection. Even though such models offer strong performance, they are not easy to explain, thus not favorable for inexperienced users. To solve this, LIME, an explainable AI method, was used to identify the features that affect model outcomes and are valuable in clinical practice. Consistently, feature significance analysis across diverse patient cases revealed that age is the top predictor of stroke. Subsequently, the next health-related factors addressed were average glucose levels, BMI, and hypertension, known to be related to cardiovascular and cerebrovascular diseases [24]. This aligns with recent clinical findings [25], [26] and enhances the model's transparency, encouraging professional and inexperienced users to find it more suitable. Moreover, while most studies target a small population, the current study focuses on a larger dataset and relies on common hospital information, making it more generalizable and easily applicable in health systems, including medical facilities with limited resources.

AI models, ML algorithms in particular, contribute greatly to making clinical decisions and the early detection of disease risk. The current study expands the evidence that ML models and ensemble models, in particular, work better than standard analytical methods in handling complex medical analysis. Additionally, computational algorithms that can analyze how clinical factors work in combination with other factors, like demographic characteristics and patient lifestyle, are highly needed in the medical field. The findings of the current study highlight that advanced computer models can find weak but essential patterns that normal methods may overlook, subsequently making their predictions about stroke risk highly sensitive and specific. Another essential goal of this study was to take medical data and refine it for easy

interpretation. Through the use of LIME, the contribution of each feature was assessed, which is necessary for physicians to understand the rationale behind the outcome of the applied models, as many healthcare professionals are not yet confident in AI algorithms, as they are hard to understand. Therefore, being able to see the results of a model boosts confidence and encourages its use in healthcare settings. In clinical work, it is important to detect patients at risk promptly, as this can strongly affect treatment outcomes. Therefore, choosing the right algorithm in medical applications involves weighing how accurate, generalizable, and reliable it is and how explainable it must be for critical medical choices.

Our research findings indicate that ML-based stroke risk prediction is a promising application; however, it has numerous limitations. The data used doesn't include some critical vascular and metabolic risk factors, such as diabetes mellitus, dyslipidemia, atrial fibrillation, and chronic kidney disease, which have been linked to influencing stroke rates and clinical outcomes. This can lead to unmeasured confounding, therefore, reducing the predictive power and generalizability of the used models. Moreover, the etiological subtyping based on the classification of the TOAST was not used in the study, which can limit the ability to identify specific ischemic stroke processes. As a result, future research can include more types of biomarkers, other comorbidities associated with stroke, and more detailed background variables, such as sociodemographic and behavioral factors, to enhance the understanding of lifestyle and clinically related risks of stroke.

Furthermore, this study relies on structured tabular data and does not use multimodal sources (neuroimaging, voice/gait analysis, and genetic analysis), thus limiting its prognostic performance. This could significantly reduce the quality and clinical value of stroke risk assessments, especially concerning the early detection of patients at high risk. Therefore, longitudinal or real-time datasets of hospitals can be incorporated in future research to test how variations of risk factors affect stroke onset. By addressing these limitations and improving the input features, further studies could develop more effective explainable AI models that align computational intelligence with realistic medical decision-making, thus promoting the initial diagnosis and personalized stroke treatment.

## CONCLUSION

This study highlights the use of explainable AI algorithms such as LIME in stroke risk prediction. Even though ML algorithms, especially ensembled models, offer accurate predictions, their structure limits their application in healthcare, particularly in disease diagnosis and/or treatment. Therefore, LIME allowed personalized stroke predictions by understanding the influence of clinical features such as demographic parameters, BMI, average glucose level, presence or absence of some other health conditions, and patient lifestyle on model decision-making. As a result, healthcare providers can better explain such computationally generated decisions to patients with confidence and understanding, thus

improving patient satisfaction. Furthermore, our findings reveal that ML algorithms associated with explainable AI models can be helpful in medical diagnostics and can contribute to more precise stroke risk categorization. More studies should be done to confirm these results in diverse populations and guide medical decision-making by using real-time data.

## ACKNOWLEDGMENTS

None

## CONFLICT OF INTEREST

None

## REFERENCES

1. Best pain medication for stroke patients: What to know n.d. <https://www.medicalnewstoday.com/articles/best-pain-medication-for-stroke-patients#anti-convulsants> (accessed September 7, 2022).
2. D. Laurent, C.N. Small, M. Goutnik, B. Hoh, "Ischemic stroke. acute care neurosurgery by case management: pearls and pitfalls" pp. 159–172, February 2025. [https://doi.org/10.1007/978-3-030-99512-6\\_13](https://doi.org/10.1007/978-3-030-99512-6_13).
3. D.W. Howells, M.J. Porritt, S.S. Rewell, V. O'collins, E.S. Sena, B. Van Der Worp, et al. "Pathophysiology and treatment of stroke: present status and future perspectives," *International Journal of Molecular Sciences* vol. 21, no. 20, p. 7609, October 2020. <https://doi.org/10.3390/IJMS21207609>.
4. B. Norrving, J. Barrick, A. Davalos, M. Dichgans, C. Cordonnier, A. Guekht, et al. "Action plan for stroke in europe 2018–2030," *Eur Stroke J* vol. 3, no. 4, pp. 309–336, December 2018. <https://doi.org/10.1177/2396987318808719>.
5. R. Hurford, A. Sekhar, T.A.T. Hughes, K.W. Muir, "Diagnosis and management of acute ischaemic stroke," *Pract Neurol* vol. 20, no. 4, p. 304, August 2020. <https://doi.org/10.1136/PRACTNEUROL-2020-002557>.
6. Stroke Association | Home n.d. <https://www.stroke.org.uk/> (accessed September 7, 2022).
7. N. Usanase, A.G. Usman, D.U. Ozsahin, L.R. David, I. Ozsahin, B. Uzun et al., "Hybridized Paradigms for The Clinical Prediction of Lung Cancer," *Proceedings - International Conference on Developments in eSystems Engineering, DeSE*, pp. 417–422, 2024, doi: 10.1109/DESE63988.2024.10911964.
8. S.R.A. Parisineni, M. Pal, "Enhancing trust and interpretability of complex machine learning models using local interpretable model agnostic SHAP explanations," *Int J Data Sci Anal* vol. 18, no. 4, pp. 457–466, October 2023. <https://doi.org/10.1007/S41060-023-00458-W>.
9. F.A. Almisned, N. Usanase, D.U. Ozsahin, I. Ozsahin, "Incorporation of explainable artificial intelligence in ensemble machine learning-driven pancreatic cancer diagnosis," *Scientific Reports* vol. 15, no. 1, pp. 1–15,

- April 2025. <https://doi.org/10.1038/s41598-025-98298-0>.
10. M.E. Marketou, G.P. Tsironis, E. Dritsas, M. Trigka, "Stroke risk prediction with machine learning techniques," *Sensors*, vol 22, p. 4670, June 2022. <https://doi.org/10.3390/S22134670>.
  11. S. Dev, H. Wang, C.S. Nwosu, N. Jain, B. Veeravalli, D. John, "A predictive analytics approach for stroke prediction using machine learning and neural networks," *Healthcare Analytics* vol. 2, p. 100032, November 2022. <https://doi.org/10.1016/J.HEALTH.2022.100032>.
  12. N. Biswas, K.M.M. Uddin, S.T. Rikta, S.K. Dey, "A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach," *Healthcare Analytics* vol. 2, p. 100116, November 2022. <https://doi.org/10.1016/J.HEALTH.2022.100116>.
  13. P. Bentley, J. Ganesalingam, A.L. Carlton Jones, K. Mahady, S. Epton, P. Rinne, et al. "Prediction of stroke thrombolysis outcome using CT brain machine learning," *Neuroimage Clin* vol. 4, pp. 635–640, 2014. <https://doi.org/10.1016/J.NICL.2014.02.003>.
  14. L. Han, M. Askari, R.B. Altman, S.K. Schmitt, J. Fan, J.P. Bentley, et al. "Atrial fibrillation burden signature and near-term prediction of stroke: a machine learning analysis," *Circ Cardiovasc Qual Outcomes*, vol. 12, no. 10, October 2019. <https://doi.org/10.1161/CIRCOOUTCOMES.118.005595>.
  15. D.U. Ozsahin, N. Usanase, I. Ozsahin, "Advancing pancreatic cancer management: the role of artificial intelligence in diagnosis and therapy," *Beni-Suef University Journal of Basic and Applied Sciences*, vol. 14, no. 1, pp. 1–18, April 2025. <https://doi.org/10.1186/S43088-025-00610-4>.
  16. N. Usanase, D. U. Ozsahin, L. R. David, B. Uzun, A. J. Hussain, and I. Ozsahin, "Deep learning-based CT-scan image classification for accurate detection of pancreatic cancer: A Comparative Study of Different Pre-Trained Models," *Proceedings - International Conference on Developments in eSystems Engineering, DeSE*, pp. 358–363, 2024, doi: 10.1109/DESE63988.2024.10912050.
  17. I. Ozsahin, B. Uzun, M.T. Mustapha, N. Usanese, M. Yuvali, D.U. Ozsahin, "BI-RADS-based classification of breast cancer mammogram dataset using six stand-alone machine learning algorithms," *Artificial Intelligence and Image Processing in Medical Imaging*, pp. 195–216, January 2024. <https://doi.org/10.1016/B978-0-323-95462-4.00008-X>.
  18. A. Bivard, L. Churilov, M. Parsons, Artificial intelligence for decision support in acute stroke — current roles and potential. *Nature Reviews Neurology*, vol. 16, no. 10, pp. 575–585, August 2020. <https://doi.org/10.1038/s41582-020-0390-y>.
  19. C.C. Karayiannis, C.C. Christopher Karayiannis, "Hypertension in the older person: is age just a number?" *Intern Med J*, vol. 52, no. 11, pp. 1877–1883, November 2022. <https://doi.org/10.1111/IMJ.15949>.
  20. W. Qiao, X. Zhang, B. Kan, A.M. Vuong, S. Xue, Y. Zhang, B. Li, et al. "Hypertension, BMI, and cardiovascular and cerebrovascular diseases," *Open Medicine (Poland)*, vol. 16, no. 1, pp. 149–155, January 2021. [https://doi.org/10.1515/MED-2021-0014/DOWNLOADASSET/SUPPL/MED-2021-0014\\_SM.PDF](https://doi.org/10.1515/MED-2021-0014/DOWNLOADASSET/SUPPL/MED-2021-0014_SM.PDF).
  21. R. She, Z. Yan, Y. Hao, Z. Zhang, Y. Du, Y. Liang, et al. "Health-related quality of life after first-ever acute ischemic stroke: associations with cardiovascular health metrics," *Quality of Life Research*, vol. 30, no. 10, pp. 2907–2917, October 2021. <https://doi.org/10.1007/S11136-021-02853-X/FIGURES/1>.
  22. C. Odiakaose, F.O. Aghware, M.D. Okpor, A.O. Eboka, A.P. Binitie, A.A. Ojugo, et al. "Hypertension detection via tree-based stack ensemble with SMOTE-Tomek data balance and XGBoost meta-learner," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 3, pp. 269–283, December 2024. <https://doi.org/10.62411/FAITH.3048-3719-43>.
  23. C. Kern, T. Klausch, F. Kreuter, "Tree-based machine learning methods for survey research," *Surv Res Methods*, vol. 13, no. 1, p. 73, 2019. <https://doi.org/10.18148/srm/2019.v13i1.7395>.
  24. Z. Wang, T. Yang, H. Fu, "Prevalence of diabetes and hypertension and their interaction effects on cardiovascular diseases: a cross-sectional study," *BMC Public Health*, vol. 21, no. 1, p. 1224, June 2021. <https://doi.org/10.1186/S12889-021-11122-Y/TABLES/7>.
  25. C.A. Couch, Z. Ament, A. Patki, N. Kijpaisalratana, V. Bhave, A.C. Jones, et al. "Sex-associated metabolites and incident stroke, incident coronary heart disease, hypertension, and chronic kidney disease in the REGARDS Cohort," *J Am Heart Assoc*, vol. 13, no. 9, May 2024. [https://doi.org/10.1161/JAHA.123.032643/SUPPL\\_FILE/JAH39589-SUP-0001-DATA\\_S1.ZIP](https://doi.org/10.1161/JAHA.123.032643/SUPPL_FILE/JAH39589-SUP-0001-DATA_S1.ZIP).
  26. Q. Chen, M. Wu, Q. Tang, P. Yan, L. Zhu, "Age-related alterations in immune function and inflammation: focus on ischemic stroke," *Aging Dis*, vol. 15, no. 3, p. 1046, June 2024. <https://doi.org/10.14336/AD.2023.0721-1>.