

***University of Ibadan Journal of
Science and Logics in ICT
Research (UIJSLICTR)***

ISSN: 2714-3627

A Journal of the Faculty of Computing, University of Ibadan, Ibadan, Nigeria

Volume 17 No. 1, June 2026

journals.ui.edu.ng/uijslictr

<http://uijslictr.org.ng/>

uijslictr@gmail.com



Fuzzy Logic-Based Model for Stroke-Risk Prediction Using Non-Invasive Risk Factors: A Mamdani Inference Approach

¹✉ Hassan J. A., ²Shólànké T. F., ³Ìdòwú I. T., ⁴Ìdòwú P. A.

¹Department of Intelligent Systems Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria

^{2,4}Department of Computer Science and Cyber Security, Obafemi Awolowo University, Ile-Ife, Nigeria

³Department of Nursing Science, Elizade University, Ilara Mokin, Nigeria

¹hassanja@oauife.edu.ng, ²tsholanke@oauife.edu.ng, ³idowuiyinoluwa056@gmail.com, ⁴paidowu@oauife.edu.ng

Corresponding author: hassanja@oauife.edu.ng

Abstract

Stroke-risk assessment can support preventive intervention by identifying individuals whose combined demographic, behavioural and health-related characteristics may indicate elevated risk. This study develops a Mamdani fuzzy inference model for preliminary stroke-risk classification using eight non-invasive inputs: age, gender, family history of stroke, smoking history, alcohol consumption, body mass index (BMI), exercise frequency and hypertension status. Each input was represented by a crisp value mapped to linguistic categories, with triangular membership functions used to model the defined intervals. Age, gender, BMI and exercise frequency were assigned two linguistic labels each, while family history, smoking, alcohol consumption and hypertension were assigned three labels each, giving $3^4 \times 2^4 = 1,296$ possible antecedent combinations. A 1,296-rule IF–THEN base was formulated with input from a domain expert. The system was implemented in MATLAB using the Fuzzy Logic Toolbox and tested with one representative input vector, [0, 1, 1, 0, 0, 1, 1, 0]. The defuzzified output was 0.16, which falls within the defined Low Risk interval. The result demonstrates the feasibility of representing accessible risk factors with an explicit fuzzy rule structure. However, the study is a proof-of-concept: a single simulated case cannot establish predictive accuracy, discrimination, calibration, clinical utility or generalisability. Independent rule-base review and validation against appropriately characterised patient data are therefore required before clinical application.

Keywords: Stroke-risk prediction; fuzzy logic; non-invasive risk factors; Mamdani inference; clinical decision support

1. Introduction

Stroke is a major cause of death and long-term disability, and prevention depends in part on identifying people whose combination of risk factors places them at increased risk. Recent systematic evidence shows that stroke-risk prediction models can achieve useful discrimination, but many models have limited external validation and substantial variation in performance across populations [1], [2]. A meta-analysis of stroke-risk models reported a pooled c-statistic of 0.78 and noted that many models had not been tested in independent populations [3].

Machine-learning approaches have expanded the range of methods available for stroke-risk prediction. Their performance can be strong in

some datasets, but evidence reviews continue to identify problems involving data quality, calibration, risk of bias, explainability and implementation evidence [1], [4], [5]. These concerns are particularly relevant when a model is intended for use outside the population in which it was developed.

Fuzzy logic provides a different modelling perspective. Instead of requiring every clinical state to be represented as a sharply separated category, fuzzy inference permits linguistic descriptions and explicit rules to connect input conditions with an output class. Reviews of interpretable fuzzy inference systems describe this rule-based structure as useful when transparent reasoning is important in medical decision support [6]. Recent evidence also shows that Mamdani systems remain common in clinical fuzzy applications, although methodological quality and validation remain important limitations [7].

The present study investigates whether readily obtainable, non-invasive risk factors can be

Hassan J. A., Shólànké T. F., Ìdòwú I. T., Ìdòwú P. A. (2026) Fuzzy Logic-Based Model for Stroke-Risk Prediction Using Non-Invasive Risk Factors: A Mamdani Inference Approach. *University of Ibadan Journal of Science and Logistics in ICT Research (UIJSLICTR)*, Vol. 17 No. 1, pp. 213 – 220.

organised into an interpretable Mamdani fuzzy inference system for preliminary stroke-risk classification. The model uses eight inputs and three output classes. Its purpose is not to replace established clinical risk scores or provide a clinically validated probability of stroke, but to demonstrate a transparent computational representation that can be tested and subsequently validated with local data. This distinction is important because evidence from African populations remains limited in the stroke-prediction literature [8], [9].

2. Related Works

Recent reviews provide a useful benchmark for interpreting the present prototype. Asowata et al. [3] found a pooled c-statistic of 0.78 across stroke-risk prediction models and reported limited external validation in much of the literature. Asadi et al. [4] similarly found that Random Forest and other ensemble methods often performed well, while emphasising that reported performance depended on the dataset, sample and validation strategy.

Evidence quality is also a recurring issue. Abdulazeem et al. [1] found that most primary-care machine-learning studies they reviewed were retrospective and that many models were developed without external validation. Rakers et al. [5] reported a scarcity of evidence across the AI implementation life cycle in primary care, particularly concerning model quality and effectiveness in clinical workflows. Heseltine-Carp et al. [10] likewise identified concerns about dataset quality, calibration and explainability in stroke-risk prediction from routine hospital data.

Recent stroke-specific work illustrates the range of reported performance. Duan et al. [11] developed and validated models using regional healthcare data; their Random Forest model achieved a validation AUC of 0.980, while the study also examined calibration and factor importance. Eini et al. [12] reported a pooled AUC of 0.94 for machine-learning models in

hypertensive populations, but also found substantial heterogeneity. Zhang et al. [13] reported pooled AUC values around 0.77 for several machine-learning approaches in community primary care and found that machine learning did not clearly outperform conventional regression.

Interpretability has received increasing attention. Tang et al. [14] compared several machine-learning models with SHAP-based explanations, while Yao et al. [15] developed a SHAP-interpretable model using circulating microRNA biomarkers. These approaches improve post-hoc explanation, whereas fuzzy inference provides interpretability directly through the rule structure. Zheng et al. [16] also noted the relevance of fuzzy approaches for uncertain medical data. Taken together, the literature supports investigation of transparent rule-based methods, but it does not justify assuming that interpretability alone establishes clinical validity.

3. Methods

The modelling procedure comprised five computational stages: selection and coding of non-invasive risk factors, fuzzification, construction of the fuzzy rule base, Mamdani inference and defuzzification. The stages were implemented in MATLAB using the Fuzzy Logic Toolbox. Figure 1 summarises the workflow.

Figure 1 shows how each patient's crisp input values are first converted into linguistic membership values. These memberships activate the corresponding IF–THEN rules in the rule base. The Mamdani inference engine aggregates the activated consequents, after which defuzzification converts the resulting fuzzy output into a single numerical value. The numerical value is then mapped to one of three predefined risk levels: Low Risk, Moderate Risk or High Risk. This arrangement makes the computational pathway explicit and allows each stage to be inspected independently.

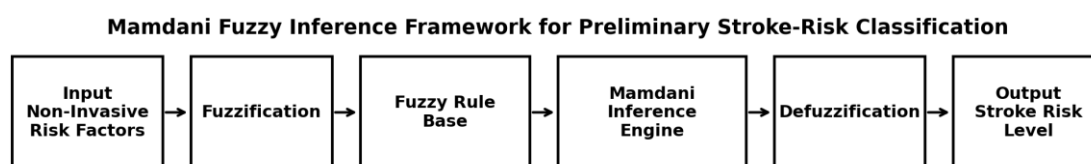


Figure 1: Methodological framework for the Mamdani fuzzy stroke-risk classification model

3.1 Identification of Associated Risk Factors

Eight non-invasive risk factors were selected: age, gender, family history of stroke, smoking history, alcohol consumption, BMI, exercise frequency and hypertension status. The factors were retained from the original model and were coded so that larger crisp values corresponded to categories defined in the study as more strongly associated with stroke risk. The coding scheme is shown in Table 1.

Table 1: Crisp and linguistic values assigned to the eight input risk factors

Risk factor	Linguistic values	Crisp value
Age	Below 50 years; Above 50 years	0; 1
Gender	Female; Male	0; 1
Family history	None; Second-degree relative; First-degree relative	0; 1; 2
Smoking history	No; Previously; Yes	0; 1; 2
Alcohol consumption	No; Previously; Yes	0; 1; 2
BMI	Normal; Abnormal	0; 1
Exercise frequency	Yes; No	0; 1
Hypertension	No; Previously; Yes	0; 1; 2

3.2 Fuzzification and Membership Functions

Each coded input was represented by triangular membership functions. For a triangular membership function with lower base point a, centre b and upper base point c, the membership degree was defined by:

$$\mu(x;a,b,c) = \max \{ \min[(x-a)/(b-a), (c-x)/(c-b)], 0 \}.$$

For two-label variables, the centres were 0 and 1; for three-label variables, the centres were 0, 1 and 2. The corresponding intervals were (−0.5, 0.5), (0.5, 1.5) and (1.5, 2.5). The output variable used the same scale, with 0, 1 and 2 representing

Low Risk, Moderate Risk and High Risk, respectively.

Table 2: Triangular membership-function parameters used for input and output coding

Crisp value	Interval	A	b	c
0	(−0.5, 0.5)	−0.5	0	0.5
1	(0.5, 1.5)	0.5	1	1.5
2	(1.5, 2.5)	1.5	2	2.5

3.3 Fuzzy Inference System Design

The fuzzy rule base combines the eight inputs as antecedents and stroke risk as the consequent. A representative rule is: IF Age is Below 50 AND Gender is Female AND Family history is None AND Smoking is No AND Alcohol consumption is No AND BMI is Normal AND Exercise frequency is Yes AND Hypertension is No, THEN Stroke risk is Low Risk.

Four variables have three linguistic labels and four have two labels. Consequently, the complete antecedent space contains $3^4 \times 2^4 = 1,296$ combinations. All 1,296 combinations were assigned a consequent risk class and reviewed by the domain expert consulted for the study. The full rule base is not reproduced because of its length, but the implementation is illustrated in Figure 6.

3.4 Simulation Environment

The model was implemented in MATLAB using the Fuzzy Logic Toolbox. The graphical tools were used to define the membership functions, construct the rule base and inspect the system through the rule viewer and related inference interfaces. The model was then tested with the representative input vector used in the original study.

4. Results and Discussion

4.1 Model Formulation and Membership Functions

The implemented system retained the eight input variables and their original linguistic partitions. Figures 2–4 present the triangular membership functions used for the binary and three-level inputs. The plots show the intended separation of the coded linguistic categories, with adjacent triangular functions meeting at their boundary points.

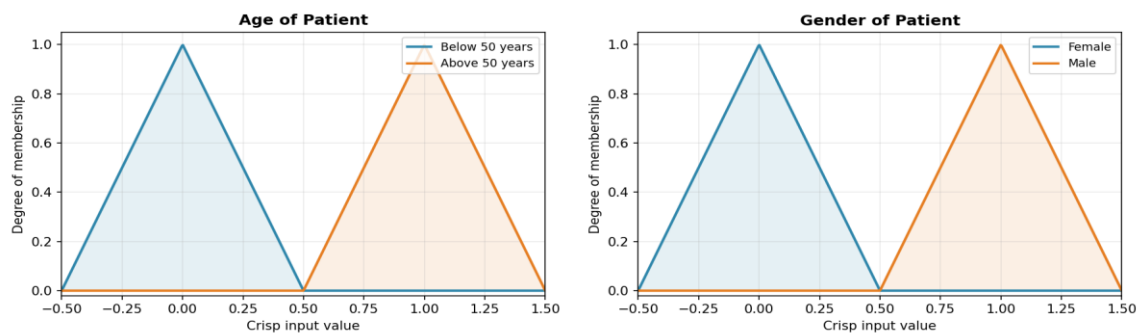


Figure 2: Membership functions for age and gender

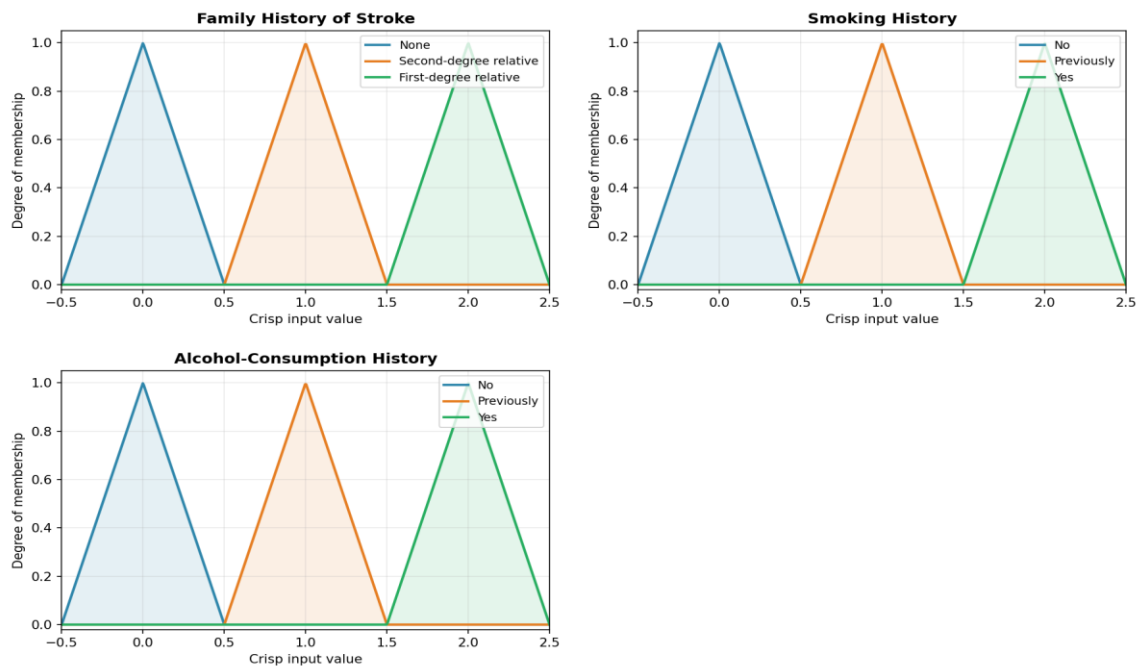


Figure 3: Membership functions for family history, smoking and alcohol-consumption history

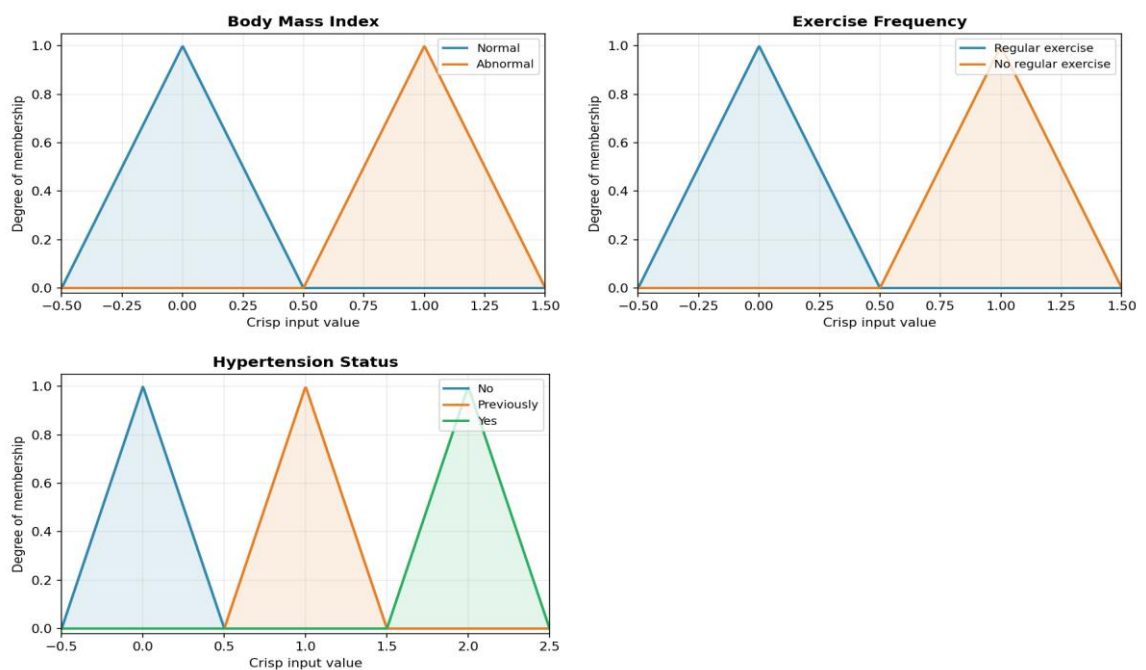


Figure 4: Membership functions for BMI, exercise frequency and hypertension status

The output variable used three triangular functions centred at 0, 1 and 2 for Low Risk, Moderate Risk and High Risk. Figure 5 also shows the complete fuzzy inference configuration. The output classes therefore remain directly connected to the numerical intervals defined in the model.

4.2 Inference Engine and Test Result

With eight antecedent variables, the rule base contains 1,296 possible combinations. Figure 6 shows the rule-base interface used to load these combinations into the MATLAB inference engine. The model was then tested using the

input vector [0, 1, 1, 0, 0, 1, 1, 0]. In the coding scheme of the study, this corresponds to below 50 years, male, a second-degree relative affected, no smoking history, no alcohol-consumption history, abnormal BMI, no regular exercise and no hypertension.

The inference engine produced a defuzzified output of 0.16. Because 0.16 lies within the defined Low Risk interval (−0.5, 0.5), the system returned Low Risk. Figure 7 shows the rule viewer response to the test input. The output is consistent with the corresponding expert-supplied rule.

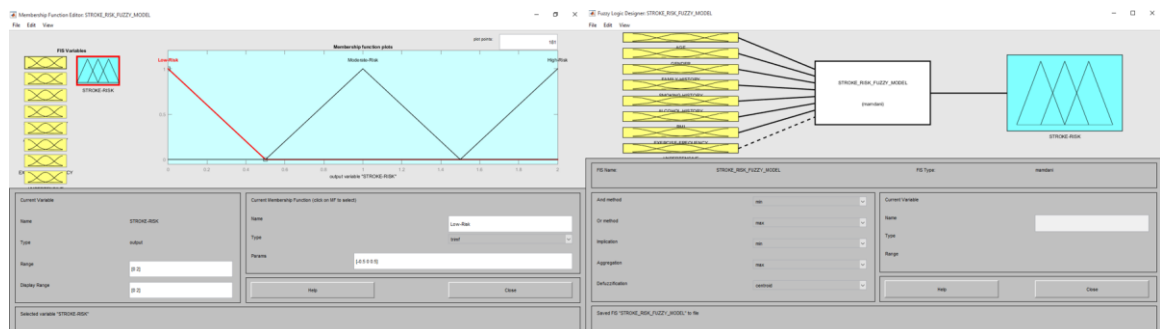


Figure 5: Output membership functions and complete fuzzy inference system

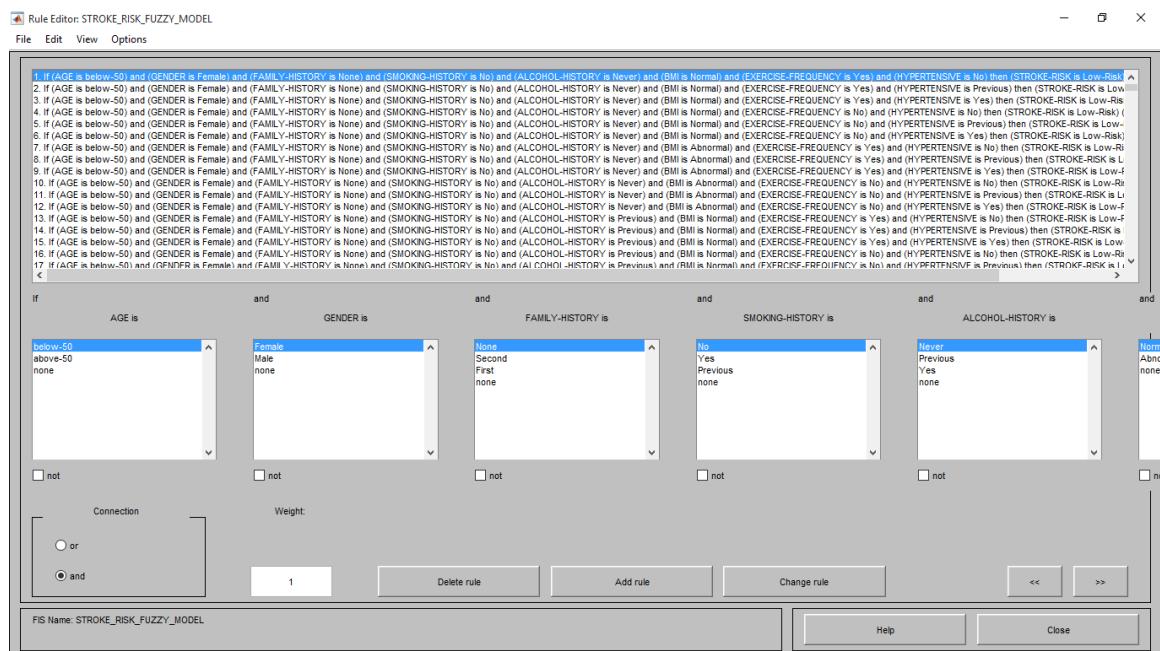


Figure 6: Rule-base interface of the Mamdani fuzzy inference system



Figure 7: Rule viewer response to the representative test input

4.3 Discussion of Results

The principal result is an internally consistent fuzzy classification rather than a measured clinical prediction performance. The value 0.16 is produced by the specified membership functions, rule base, inference procedure and defuzzification scheme. It demonstrates that the eight coded risk factors can be passed through the complete Mamdani pipeline to obtain a stable categorical output for the selected test case. It does not, however, indicate that the model correctly predicts stroke in a patient population.

A direct accuracy comparison with recent stroke-prediction studies is therefore inappropriate at this stage because the present study used one simulated input and did not evaluate sensitivity, specificity, AUC, calibration or predictive value against observed outcomes. For context, Asowata et al. [3] reported a pooled c-statistic of 0.78 across existing stroke-risk models, while Duan et al. [11] reported a validation AUC of 0.980 for their Random Forest model. Eini et al. [12] reported a pooled AUC of 0.94 in hypertensive populations, whereas Zhang et al. [13] found pooled AUCs around 0.77 for several approaches in community primary care. These values come from empirical datasets and cannot be assigned to the present fuzzy model.

The comparison nevertheless clarifies the contribution of the present work. The proposed model makes its decision mechanism explicit

through linguistic variables and 1,296 inspectable rules, rather than deriving the classification solely from fitted statistical parameters. This is consistent with the interpretability motivation reported for fuzzy inference systems [6], [7]. At the same time, recent systematic reviews show that interpretability does not remove the need for rigorous validation; inadequate validation and high risk of bias remain common in both machine-learning and fuzzy clinical studies [5], [7], [10], [13].

From a clinical decision-support perspective, the model may be useful as a prototype for exploring how accessible risk-factor information can be translated into an explicit rule-based representation. Its limitations are substantial. First, the rule base was reviewed by one domain expert, so inter-rater agreement has not been assessed. Second, the membership boundaries are design choices rather than parameters estimated from patient outcomes. Third, the test comprised a single hand-selected input vector. Fourth, the model does not currently incorporate several potentially important clinical variables used in empirical stroke models, such as diabetes, glucose measures or blood-pressure measurements beyond the categorical hypertension variable. Finally, no prospective or external validation has been performed. These limitations mean that the system should be treated as a proof-of-concept decision-support model.

Future evaluation should therefore use appropriately characterised patient datasets with known outcomes, independent clinical review of the rule base, assessment of discrimination and calibration, and comparison with established risk instruments and machine-learning baselines. Such evaluation should also examine performance across relevant population subgroups before any clinical deployment is considered.

5. Conclusion

This study developed a Mamdani fuzzy inference model for preliminary stroke-risk classification using eight non-invasive risk factors: age, gender, family history of stroke, smoking history, alcohol consumption, BMI, exercise frequency and hypertension status. Two or three triangular membership functions were assigned to each input according to its linguistic categories, and the output was represented by Low Risk, Moderate Risk and High Risk. The resulting antecedent space contained 1,296 IF–THEN rules, implemented in MATLAB with input from a domain expert.

For the representative input vector [0, 1, 1, 0, 0, 1, 1, 0], the system generated a defuzzified output of 0.16 and classified the case as Low Risk. The result supports the internal computational consistency of the constructed rule system, but it is not evidence of predictive accuracy. The model should therefore be regarded as a proof-of-concept framework. Independent expert review, empirical validation against real patient outcomes, calibration analysis, and comparison with established and data-driven prediction methods are required before any clinical use.

Areas for Future Research

Future work will test the model with ethically approved patient data, supported by appropriate informed consent or an approved waiver where applicable. The fuzzy rules and reported defuzzified output would also be independently checked to confirm that the MATLAB implementation matches the model specification. Clinical experts from relevant disciplines could then review the 1,296 rules and help refine them. Further studies may include additional factors, such as diabetes, measured blood pressure, glucose and lipid levels. More realistic representations would avoid reducing BMI and physical activity to simple yes/no categories.

For BMI, future versions could use several clinically meaningful fuzzy ranges, such as underweight, normal weight, overweight and obesity, with overlapping membership functions to reflect gradual transitions between categories. For physical activity, the model could consider frequency, duration and intensity, for example distinguishing sedentary, low, moderate and high activity based on weekly activity patterns. The model should subsequently be assessed using sensitivity, specificity, area under the receiver operating characteristic curve (AUC-ROC), and calibration and compared with established stroke-risk tools and machine-learning models. All future patient-based studies should obtain appropriate ethical clearance, protect participant data and comply with applicable research and data-protection requirements.

6. References

- [1] Abdulazeem, H.; Whitelaw, S.; Schauburger, G.; Klug, S.J. A systematic review of clinical health conditions predicted by machine learning diagnostic and prognostic models trained or validated using real-world primary health care data. *PLOS ONE* 2023, 18, e0274276. <https://doi.org/10.1371/journal.pone.0274276>.
- [2] Asadi, F.; Rahimi, M.; Daechini, A.H.; Paghe, A. The most efficient machine learning algorithms in stroke prediction: A systematic review. *Health Sci. Rep.* 2024, 7, e70062. <https://doi.org/10.1002/hsr2.70062>.
- [3] Asowata, O.J.; Okeunle, A.P.; Olaiya, M.T.; Akinyemi, J.; Owolabi, M.; Akpa, O.M. Stroke risk prediction models: A systematic review and meta-analysis. *J. Neurol. Sci.* 2024, 460, 122997. <https://doi.org/10.1016/j.jns.2024.122997>.
- [4] Cao, J.; Zhou, T.; Zhi, S.; Lam, S.; Ren, G.; Zhang, Y.; Wang, Y.; Dong, Y.; Cai, J. Fuzzy inference system with interpretable fuzzy rules: Advancing explainable artificial intelligence for disease diagnosis—A comprehensive review. *Inf. Sci.* 2024, 662, 120212. <https://doi.org/10.1016/j.ins.2024.120212>.
- [5] Rakers, M.M.; van Buchem, M.M.; Kucenko, S.; de Hond, A.; Kant, I.; van Smeden, M.; Moons, K.G.M.; Leeuwenberg, A.M.; Chavannes, N.; Villalobos-Quesada, M.; van Os, H.J.A. Availability of evidence for predictive machine learning algorithms in primary care: A systematic review. *JAMA Netw. Open* 2024, 7, e2432990. <https://doi.org/10.1001/jamanetworkopen.2024.32990>.
- [6] Zheng, Y.; Xu, Z.; Wu, T.; Yi, Z. A systematic survey of fuzzy deep learning for uncertain medical data. *Artif. Intell. Rev.* 2024, 57, 230. <https://doi.org/10.1007/s10462-024-10871-7>.

- [7] Kettle, T.; McKeever, T.M.; Gonem, S.; Figueredo, G.; Bogdanovica, I. Fuzzy logic in respiratory medicine: A systematic review of predictive and diagnostic applications. *Int. J. Med. Inform.* 2026, 212, 106347. <https://doi.org/10.1016/j.ijmedinf.2026.106347>.
- [8] Nweke, M.; Oyirinnaya, P.; Nwoha, P.; Mitha, S.B.; Mshunqane, N.; Govender, N.; Ukwuoma, M.; Ibeneme, S.C. Development of a high-performing, cost-effective and inclusive Afrocentric predictive model for stroke: A meta-analysis approach. *BMC Neurol.* 2025, 25, 282. <https://doi.org/10.1186/s12883-025-04229-x>.
- [9] Soladoye, A.A.; Aderinto, N.; Popoola, M.R.; Adeyanju, I.A.; Osonuga, A.; Olawade, D.B. Machine learning techniques for stroke prediction: A systematic review of algorithms, datasets, and regional gaps. *Int. J. Med. Inform.* 2025, 203, 106041. <https://doi.org/10.1016/j.ijmedinf.2025.106041>.
- [10] Heseltine-Carp, W.; Courtman, M.; Browning, D.; Kasabe, A.; Allen, M.; Streeter, A.; Ifeakor, E.; James, M.; Mullin, S. Machine learning to predict stroke risk from routine hospital data: A systematic review. *Int. J. Med. Inform.* 2025, 196, 105811. <https://doi.org/10.1016/j.ijmedinf.2025.105811>.
- [11] Duan, Y.; Wang, R.; Sun, Y.; Zhu, W.; Li, Y.; Yu, N.; Zhu, Y.; Shen, P.; Sun, H. Development and validation of a stroke risk prediction model using regional healthcare big data and machine learning. *Int. J. Nurs. Sci.* 2025, 12, 558–565. <https://doi.org/10.1016/j.ijnss.2025.10.011>.
- [12] Eini, P.; Rezayee, M.; Kassulke, M.; Tremblay, J. Efficacy and comparative performance of machine learning models for stroke risk prediction in hypertensive patients: A systematic review and meta-analysis. *Int. J. Cardiol. Cardiovasc. Risk Prev.* 2026, 28, 200564. <https://doi.org/10.1016/j.ijcrp.2025.200564>.
- [13] Zhang, Z.; Zhang, Y.; Li, Z.; Tian, C.; Liang, J.; Xie, J.; Lei, J. Machine learning based prediction models for first stroke in community primary care: A systematic review and meta-analysis. *BMC Med. Inform. Decis. Mak.* 2026. <https://doi.org/10.1186/s12911-026-03636-5>.
- [14] Tang, X.; Tang, M.; Liu, W.; Cui, S. Explainable machine learning for stroke risk prediction: A comparative study with SHAP-based interpretation. *Front. Neurol.* 2026, 16, 1716984. <https://doi.org/10.3389/fneur.2025.1716984>.
- [15] Yao, Q.; Sun, B.; Shen, J.; Wang, C.; Yao, W.; Chen, Y. Development and validation of a SHAP-interpretable machine learning model for stroke risk prediction using circulating microRNA biomarkers. *J. Mol. Neurosci.* 2026, 76, 83. <https://doi.org/10.1007/s12031-026-02540-x>.