

Prediction of Early Functional Outcome After Acute Ischemic Stroke Using Real-World Clinical Data in Vietnam and Indonesia: Retrospective Cohort Study

Annisa Ristya Rahmanti^{a,b}, Lutfan Lazuardi^a, Cong Minh Tran^c, Hanifah Wulandari^a, Ika Agustin A Putri^a, Farida Niken AN Hati^d, Mai Duy Ton^e, and Huan X Nguyen^{b,1}

^a *Department of Health Policy and Management, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia*

^b *London Digital Twin Research Center, Middlesex University, London, UK*

^c *Nuffield Department of Medicine, University of Oxford, United Kingdom; School of Computing and Mathematical Sciences, University of Leicester, UK*

^d *UGM Academic Hospital, Yogyakarta, Indonesia*

^e *Stroke Center, Bach Mai Hospital; Hanoi Medical University; VNU-University of Medicine and Pharmacy, Hanoi, Viet Nam*

ORCID ID: AR Rahmanti <https://orcid.org/0000-0001-9478-6267>,

MD Ton <https://orcid.org/0000-0002-7210-3522>,

HX Nguyen <https://orcid.org/0000-0002-4105-2558>

Abstract. Accurate prediction of early functional outcome after acute ischemic stroke is critical for clinical decision-making. This retrospective cohort study developed and externally validated a machine learning model using routine clinical data from two settings. A total of 11,911 ischemic patients from the 2023 Stroke Care Quality (RES-Q) registry in 52 hospitals across Vietnam and 83 patients from UGM Academic Hospital, Indonesia, were included. The primary outcome was discharge modified Rankin Scale (mRS ≤ 2 =favorable; >2 = poor). The XGBoost model achieved strong internal discrimination (AUC=0.905, F1=0.836) and maintained robust external performance (AUC=0.888, F1=0.810). SHAP interpretability identified pre-stroke mRS, admission NIHSS score, first-day glucose check, and age as the strongest predictors of poor functional outcome, while early dysphagia screening, physiotherapy evaluation, and small-vessel etiology were associated with better recovery. These results demonstrate that a data-driven AI approach using routine clinical parameters can achieve reliable cross-regional generalization and may support early prognostication in diverse stroke care settings.

Keywords. machine learning, ischemic stroke, modified ranking scale, clinical data

1. Introduction

Stroke remains a leading cause of death and long-term disability worldwide, with a growing burden in low- and middle-income countries (LMICs) where access to

¹ Corresponding Author: Prof. Huan X Nguyen. Email: H.Nguyen@mdx.ac.uk

standardized rehabilitation and follow-up is limited [1]. Early evaluation of functional outcome is crucial for guiding treatment decisions, resource allocation, and benchmarking care quality [2]. Although the 90-day modified Rankin Scale (mRS) is the global standard, many LMIC hospitals face challenges in post-discharge follow-up [3]. As a result, discharge mRS, has emerged as a practical in-hospital surrogate for predicting long-term recovery and system performance.

Artificial intelligence (AI) and machine learning (ML) show promise in predicting post-stroke outcomes from routinely collected clinical data [4]. However, most models are developed in high-income settings, limiting their generalizability to LMIC contexts, where patient profiles, treatment timelines, and clinical workflows differ [3]. This study develops and externally validates an AI-based model to predict early functional outcome after acute ischemic stroke, trained on multicenter data from the 2023 Vietnam Stroke Care Quality (RES-Q) registry and validated on a 2024 cohort from UGM Academic Hospital. The model predicts discharge mRS using routine clinical variables.

2. Methods

2.1. Study Design and Data Sources

This retrospective cohort study used de-identified, real-world clinical data from two Stroke Care Quality (RES-Q) registries. The development cohort consisted of 11,911 ischemic stroke patients from 52 hospitals in Vietnam (2023), while the external validation cohort included 83 ischemic patients admitted to UGM Academic Hospital, Indonesia (Q3 2024). Both datasets captured standardized clinical information based on the RES-Q protocol. Only cases with complete discharge modified Rankin Scale (mRS) scores were included, with modified Rankin Scale (mRS) dichotomized as: (1) Favorable outcome: mRS $\leq 0-2$ and (2) Poor outcome: mRS 3–6. Patients were eligible if they were admitted with acute ischemic stroke, aged ≥ 18 years, and had a recorded discharge mRS score. The Vietnamese dataset was obtained from the UK-Vietnam-Indonesia collaboration on an AI-enabled digital twin for stroke prediction. The Indonesian dataset was derived from the UGM Academic Hospital Stroke Registry and approved by the Medical and Health Research Ethics Committee (MHREC), Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada - Dr. Sardjito General Hospital under approval number KE/FK/1408/EC/2025.

2.2. Predictor Variables

Predictors encompassed five main domains:

- Demographic and baseline characteristics (e.g., age, sex, pre-stroke mRS, vascular risk factors such as hypertension, diabetes, and atrial fibrillation)
- Clinical and neurological scores (e.g., NIHSS, GCS, ASPECTS, ABCD2)
- Workflow and timing metrics (e.g., onset-to-door, door-to-needle, door-to-imaging intervals)
- Treatment and intervention data (e.g., thrombolysis, thrombectomy, anticoagulant/antiplatelet use, dysphagia screening, physiotherapy) and
- In-hospital complications and laboratory findings (e.g., glucose, temperature, post-stroke pneumonia).

Clinical and neurological scores were recorded at admission, while treatment and monitoring variables were collected during hospitalization prior to discharge; discharge-only variables were excluded from the predictor set. The outcome was the discharge mRS.

2.3. Data Pre-processing and Feature Selection

Variables with more than 70% missing data were excluded independently in each dataset, leaving 157 candidate variables remained in the Vietnamese dataset and around 120 in the Indonesian dataset. Continuous features were standardized using z-score normalization, while categorical variables were one-hot encoded. Missing numeric and categorical values were imputed using the median and most frequent strategies, respectively. Average patient-level missingness was 14.99% (max 78.99%) in Vietnam and 65.90% (max 78.99%) in Indonesia, and no patients were excluded. To address class imbalance, Synthetic Minority Oversampling Technique (SMOTE) was applied to the training set with a sampling ratio of 0.6 and $k = 5$ neighbors. Feature selection using SelectKBest with an ANOVA F-test retained the top 40 predictors, and external validation used the same predictors available in both datasets.

2.4. Model Development

An Extreme Gradient Boosting (XGBoost) classifier was implemented within a pipeline integrating imputation, scaling, encoding, and SMOTE. The dataset was split into 80% training and 20% testing subsets using stratified random sampling. Model performance was assessed by AUC, precision, recall, and F1-score. The final model was externally validated using the Indonesian cohort. Model interpretability was assessed using SHAP (Shapley Additive Explanations) values. Global feature importance was derived from mean absolute SHAP values, while summary plots illustrated the direction and magnitude of predictor effects on outcome probability. All analyses were conducted in Python 3.12.12 using scikit-learn, XGBoost, and SHAP libraries via Google Colab.

3. Result

A total of 11,911 ischemic-stroke cases from the Vietnam Multisite Stroke Center and 83 cases from Indonesia’s UGM Academic Hospital were analysed (Table 1). Patients in both cohorts showed comparable baseline characteristics, with mean ages of 66.1 ± 11.8 and 63.5 ± 12.4 years, and similar median NIHSS and mRS scores.

Table 1. Baseline Characteristics of patients included in the Vietnamese and Indonesian cohorts

Variable	Vietnam Multisite stroke center	Indonesia (UGM Academic Hospital)
Age (years)	66.1 ± 11.8	63.5 ± 12.4
Female (%)	40.7 %	43.0 %
NIHSS (median [IQR])	10 [6–17]	8 [5–15]
mRS (median [IQR])	3 [2–5]	3 [2–4]
Sample Size (n)	11,911	83

The XGBoost model trained on the Vietnam dataset achieved strong internal validity (AUC = 0.905, accuracy = 0.836), while maintaining excellent external generalizability on the UGM dataset (AUC = 0.888, accuracy = 0.814) (Table 2, Figure 1). Macro- and weighted-average F1-scores (>0.80) confirmed balanced discrimination between

favorable ($mRS \leq 2$) and poor ($mRS > 2$) outcomes. External validation improved precision for poor outcomes (0.903) and maintained high recall for good outcomes (0.933), indicating consistent cross-population performance despite sample-size differences.

Table 2. Model performance comparison between Vietnam multicenter and UGM Academic Hospital

Data	Precision	Recall	F1-score
Vietnam Multicenter			
0= favourable outcome ($mRS \leq 2$)	0.877	0.887	0.882
1= poor outcome ($mRS > 2$)	0.744	0.726	0.735
UGM Academic Hospital			
0= favourable outcome ($mRS \leq 2$)	0.764	0.933	0.840
1= poor outcome ($mRS > 2$)	0.903	0.683	0.778

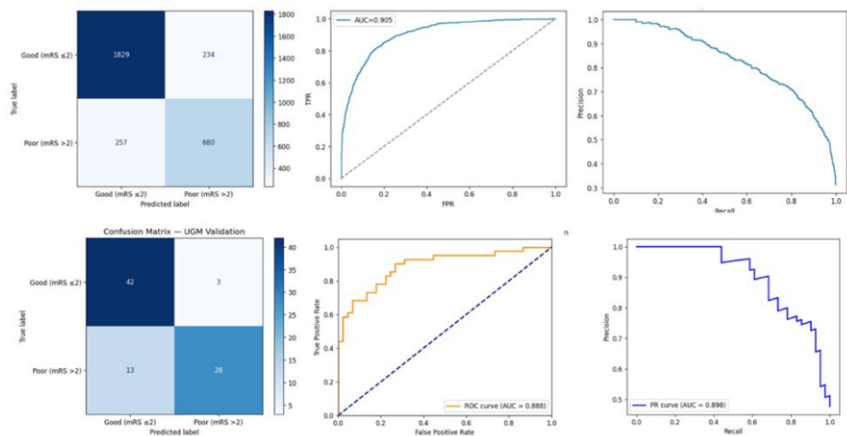


Figure 1. Model Performance Comparison a) Vietnam registry b) UGM Academic Hospital registry

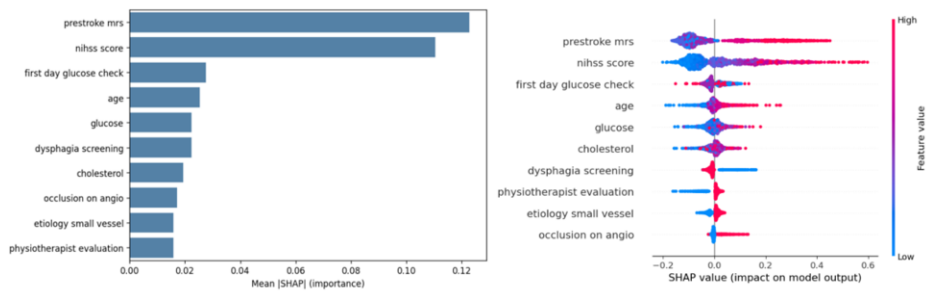


Figure 2. SHAP explainability results

The SHAP analysis (Figure 2) identified pre-stroke mRS, NIHSS score, and first-day glucose check as the most influential predictors of poor outcome. Higher pre-stroke disability, older age, elevated NIHSS, and increased glucose were associated with worse recovery, whereas early dysphagia screening, physiotherapist evaluation, and small-vessel etiology were linked to improved functional outcomes. These results demonstrate the model’s potential as a generalizable prognostic component for future digital-twin frameworks in stroke care.

4. Discussion

This study demonstrated that an AI-based model trained on multicenter Vietnamese data can predict early functional outcomes after acute ischemic stroke and generalize effectively to an Indonesian cohort (using XGBoost classifier achieved AUC=0.905 internally; AUC=0.888 externally). While prior work in high-income settings has shown that deep learning models combining imaging and clinical variables can predict 90-day outcomes with reduced subjectivity and user burden [3], such approaches require advanced imaging infrastructure. In contrast, our model relies on routinely collected clinical variables yet achieved strong cross-regional performance, supporting its feasibility for LMIC settings. SHAP analysis confirmed clinical plausibility predictors, including pre-stroke disability, admission NIHSS, age, and early glucose checks, and identified modifiable factors such as dysphagia screening and early physiotherapy, supporting integration into routine decision-support workflows. Limitation include reliance on registry data, the small external validation cohort, and the absence of imaging or long-term outcomes (e.g., 90-day mRS).

5. Conclusion

An AI-based model trained on multicenter Vietnamese registry data demonstrated strong discrimination and maintained robust performance when externally validated on an Indonesian cohort. The findings suggest that machine-learning models built from routine clinical variables can generalize across LMIC settings and support early prognostication. Future work will include multicentre external validation, integration of multimodal data, temporal updating, and prospective evaluation to enhance robustness and clinical utility.

Acknowledgements

This work was supported by The British Council ISPF Research Collaborations Programme, Grant ID #1203701580; the Academy of Medical Sciences' ISPF Networking Grant, Grant ID: NGR1\1961; the Faculty of Medicine, Public Health, and Nursing UGM VPGL Grant (No. 2649/UN.1/KU.1/AK/DI.1/2025); and the Pasca Doctoral - Community Funding Grant (No. 2075/UN1/KU/PP/PT.01.03/2025).

References

- [1] GBD Stroke Risk Factor Collaborators, Global, regional, and national burden of stroke and its risk factors, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021, *Lancet Neurol* **23** (2024), 973-1003.
- [2] J. Klug, G. Leclerc, E. Dirren, and E. Carrera, Machine learning for early dynamic prediction of functional outcome after stroke, *Commun Med (Lond)* **4** (2024), 232.
- [3] Y. Liu, Y. Yu, J. Ouyang, B. Jiang, G. Yang, S. Ostmeier, M. Wintermark, P. Michel, D.S. Liebeskind, M.G. Lansberg, G.W. Albers, and G. Zaharchuk, Functional Outcome Prediction in Acute Ischemic Stroke Using a Fused Imaging and Clinical Deep Learning Model, *Stroke* **54** (2023), 2316-2327.
- [4] Z. Liu, R. Zhang, K. Ouyang, B. Hou, Q. Cai, Y. Xie, and Y. Liu, Predicting functional outcome in acute ischemic stroke patients after endovascular treatment by machine learning, *Transl Neurosci* **14** (2023), 20220324.