Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

CTA-Based 3D Thrombus Radiomics and Machine Learning for Differentiating Stroke Etiology in Acute Middle Cerebral Artery Occlusion

Yijie Huang
Lishui Central Hospital, the Fifth Affiliated Hospital of Wenzhou Medical University
Xing Chen
Lishui Central Hospital, the Fifth Affiliated Hospital of Wenzhou Medical University
Bailong Xin
Lishui Central Hospital, the Fifth Affiliated Hospital of Wenzhou Medical University
Zekang Ye
Lishui Central Hospital, the Fifth Affiliated Hospital of Wenzhou Medical University
Hongyu Ye
Lishui Hospital of Wenzhou Medical University, The First Affiliated Hospital of Lishui University, Lishui People's Hospital
Xuejing Kong
Lishui Central Hospital, the Fifth Affiliated Hospital of Wenzhou Medical University
Xueli Cai
Lishui Central Hospital, the Fifth Affiliated Hospital of Wenzhou Medical University
Jingping Sun
lszxyysunjingping@yeah.net

Lishui Central Hospital, the Fifth Affiliated Hospital of Wenzhou Medical University

Research Article

Keywords: Middle cerebral artery occlusion, Thrombus radiomics, Computed tomography angiography, Machine learning, Stroke etiology

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11228385/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose:

To develop and evaluate machine learning models integrating three-dimensional (3D) thrombus radiomic features from preoperative computed tomography angiography (CTA) with clinical variables to differentiate large-artery atherosclerosis (LAA) from cardioembolism (CE) in patients with acute middle cerebral artery occlusion (MCAO).

Materials and Methods:

This retrospective two-center study included 453 patients with acute MCAO who underwent endovascular treatment (EVT). The 3D thrombus region was segmented on preoperative CTA, and radiomic features reflecting intensity, texture, and morphology were extracted. Multiple machine learning models integrating radiomic and clinical variables were developed using the training cohort and evaluated in a held-out test cohort. Discrimination, calibration, SHAP-based interpretability, feature importance, and decision curve analysis were assessed.

Results:

Among the evaluated models, the random forest (RF) model showed competitive performance in the held-out test cohort, with an area under the receiver operating characteristic curve of 0.822, sensitivity of 0.867, and the lowest Brier score (0.171) among the evaluated models. Both clinical and 3D thrombus radiomic features contributed to model predictions. Intraprocedural tirofiban use, admission LDL-C level, sex, admission NIHSS score, and selected thrombus texture and morphological features were important predictors. Decision curve analysis indicated positive net benefit over treat-all and treat-none strategies across threshold probabilities of approximately 0.05–0.60.

Conclusion:

Three-dimensional thrombus radiomics from routine preoperative CTA combined with clinical variables may help differentiate LAA from CE in acute MCAO. The interpretable RF model may serve as a complementary tool for acute-phase stroke etiological assessment.

Introduction

Acute ischemic stroke is one of the leading causes of death and long-term disability among adults worldwide, with large vessel occlusion (LVO) being associated with more severe neurological deficits and poorer clinical outcomes[1, 2].For patients with middle cerebral artery occlusion (MCAO), endovascular treatment (EVT) has become an important therapeutic approach for achieving vascular

recanalization, restoring cerebral tissue perfusion, and improving functional outcomes[3, 4].However, the etiology of stroke in patients with MCAO is highly heterogeneous, with large-artery atherosclerosis (LAA) and cardioembolism (CE) representing two important etiological subtypes[5, 6]. Differentiating these etiological subtypes is clinically relevant because they may differ in acute management considerations, secondary prevention strategies, and prognostic implications[7].

The determination of stroke etiology is generally based on etiological classification systems such as the TOAST classification, with a comprehensive assessment of clinical characteristics, brain imaging, cardiac imaging, vascular imaging, and laboratory findings[5, 8].Previous studies have demonstrated an association between different stroke etiologies and the histological composition of retrieved thrombi, suggesting that the thrombus itself may contain information related to stroke etiology[9–11]. However, determining stroke etiology requires the integration of vascular, cardiac, and other relevant examinations, and some etiological assessments may need to be completed based on the results of initial evaluations. Therefore, the etiological information available during the acute phase may be incomplete[8, 12].Meanwhile, differences in embolic sources and thrombus formation mechanisms may result in variations in thrombus composition, structure, and imaging characteristics, and these potential differences may be reflected in the imaging features of the thrombus itself[10, 11].Therefore, further mining of stroke etiology-related information from thrombus imaging obtained during the emergency phase may provide a novel approach for differentiating between LAA and CE[13, 14].

Conventional assessment of thrombi relies mainly on qualitative or limited quantitative imaging markers and may not fully capture their complex internal heterogeneity and three-dimensional characteristics. Quantitative CT-based thrombus characteristics, such as thrombus density, have been investigated in acute ischemic stroke and may provide imaging information beyond visual assessment.[15] Radiomics enables the quantitative extraction of intensity, shape, and texture features from thrombus regions, thereby providing a more comprehensive representation of thrombus characteristics that may be difficult to capture by conventional assessment[16, 17].Previous studies have suggested that CT-based thrombus radiomic features are associated with thrombus composition and may provide additional information for differentiating stroke etiologies[14, 18–20].

Therefore, this study aimed to develop and evaluate machine-learning models integrating preoperative three-dimensional CTA thrombus radiomics with clinical features to differentiate LAA from CE in patients with acute MCAO undergoing EVT.

Methods

Patient selection

This retrospective study consecutively enrolled 453 adult patients who underwent EVT for anterior circulation large-vessel occlusion at two participating centers between January 2022 and December 2025. This study was conducted in accordance with the Declaration of Helsinki and was approved by the

institutional ethics committee [blinded for peer review]. Written informed consent was obtained from all patients or their legal representatives.

The inclusion criteria were: (1) age ≥ 18 years; (2) time from symptom onset to EVT initiation ≤ 24 hours; (3) definitive confirmation of intracranial MCA M1 or M2 occlusion by digital subtraction angiography (DSA); (4) pre-endovascular treatment (pre-EVT) NCCT and CTA (slice thickness ≤ 5.0 mm) completed within 1 hour prior to EVT; and (5) complete clinical and follow-up data. Exclusion criteria comprised: (1) concomitant anterior cerebral artery occlusion or posterior circulation stroke; (2) inability to evaluate collateral circulation or identify the thrombus location on CTA; (3) severe NCCT/CTA artifacts, poor image quality, or CT attenuation > 100 HU (calcification or calcified thrombus); (4) pre-existing intracranial hemorrhage, tumor, or other non-ischemic brain diseases; and (5) stroke etiology classified under TOAST criteria as other than LAA or CE.

All clinical and imaging data were retrospectively retrieved from the electronic medical record system and the picture archiving and communication system. Stroke etiology was determined by attending neurologists using the TOAST criteria and dichotomized into LAA versus CE, which served as the primary endpoint for model prediction. The entire cohort was randomly partitioned into a training cohort (80%) and a held-out test cohort (20%) using stratified random sampling. The training cohort was employed for feature preprocessing, feature selection, hyperparameter tuning, and model development, whereas the held-out test cohort was reserved for final performance evaluation and comparative assessment of the developed models. The complete enrollment pathway, including exclusion/inclusion filters, is presented in Fig. 1.

CT Image Acquisition and Thrombus ROI Segmentation

All patients underwent pre-treatment NCCT and CTA using a 192-slice dual-source CT scanner (SOMATOM Force; Siemens Healthineers, Forchheim, Germany). Before scanning, an 18-gauge intravenous catheter was placed in the antecubital vein. Contrast medium was administered through a dual-channel power injector at a rate of 5.0 mL/s, using 40 mL of iohexol followed by 30 mL of normal saline. For NCCT, the acquisition parameters included a slice thickness of 5.0 mm, a tube voltage of 120 kV, and a tube current of 300 mAs. CTA was acquired with a slice thickness of 1.0 mm, a tube voltage of 80 kV, and a tube current of 150 mAs.

Pre-EVT CTA imaging acquired within 1 hour before treatment was analyzed for all patients. The occlusion segment and thrombus location were determined based on CTA findings. 3D regions of interest (ROIs) encompassing the thrombus were manually delineated on CTA images using ITK-SNAP software (version 4.2.2) (Fig. 2). The ROI included the complete 3D thrombus causing MCA M1 or M2 segment occlusion, with prominent calcified areas and adjacent normal vessel walls excluded. Thrombus segmentation was performed by two radiologists blinded to clinical information. Interobserver reproducibility was assessed using the intraclass correlation coefficient, and radiomic features with an ICC > 0.95 were considered reproducible and retained for subsequent analysis.

Radiomics Feature Extraction

Preoperative CTA images were analyzed to extract quantitative radiomic features from the 3D thrombus ROIs using the PyRadiomics platform (version 2.2.0). The extracted radiomic features were combined with clinical and treatment-related variables for machine-learning model development.

Statistical Analysis

Model Development and Validation

The total cohort was randomly partitioned into a training cohort (80%) and a held-out test cohort (20%) using stratified sampling. The held-out test cohort was not used during preprocessing, feature selection, model training, or hyperparameter optimization. Within the training cohort, the optimal number of radiomic features and model hyperparameters were jointly optimized for each algorithm using randomized search with stratified five-fold cross-validation, with ROC AUC as the optimization metric. Within each cross-validation split, continuous variables were median-imputed and standardized using Z-score normalization, whereas missing categorical variables were imputed using the most frequent category and encoded using one-hot encoding. The preprocessing steps and ANOVA-based feature selector were fitted exclusively on the training fold and subsequently applied to the corresponding validation fold. ANOVA-based SelectKBest was applied only to the radiomic features, whereas clinical and treatment-related variables were retained without further feature selection. For each algorithm, the optimal number of retained radiomic features was selected from (k) = 3, 5, 8, 10, 15, 20, 30, and 50. Six machine-learning algorithms—logistic regression (LR), random forest (RF), support vector machine (SVM), XGBoost, multilayer perceptron (MLP), and ExtraTrees were evaluated. After the optimal number of radiomic features and model hyperparameters were determined, the selected pipeline was refitted using the entire training cohort and subsequently applied to the held-out test cohort for final evaluation. For LR, both L1- and L2-regularized models were considered.

Model Evaluation

For each model, the optimal classification threshold was determined using the Youden index based on out-of-fold predictions from the training cohort. Confusion matrices were generated in the held-out test cohort using the model-specific optimal classification thresholds. Model performance in the held-out test cohort was assessed using the area under the receiver operating characteristic curve (AUC), area under the precision-recall curve (AUPRC), accuracy, F1 score, sensitivity, specificity, and Brier score. Calibration curves were generated to assess the agreement between predicted probabilities and observed outcomes. Decision curve analysis (DCA) was performed across threshold probabilities ranging from 0.01 to 0.80 to evaluate net clinical benefit relative to the treat-all and treat-none strategies. Pairwise comparisons of AUCs in the held-out test cohort were performed using two-sided DeLong tests, with unadjusted P values reported.

After comparative evaluation of the candidate models in the held-out test cohort, the RF model was selected for subsequent interpretability analyses because it provided a favorable balance of discriminative, classification, and calibration performance. Gini-based feature importance (mean decrease in impurity) was extracted from the RF model refitted on the entire training cohort. SHAP values were calculated for the held-out test cohort using TreeExplainer based on the final RF model. SHAP beeswarm and mean absolute SHAP plots were generated to characterize global feature contributions, while SHAP waterfall plots were used to illustrate local prediction patterns in representative cases. A prediction probability waterfall plot was also generated to visualize the distribution of predicted probabilities across the held-out test cohort. For logistic regression, an L1 regularization path was generated to visualize changes in feature coefficients across different regularization strengths.

All statistical analyses were performed using Python (version 3.7) and R (version 4.1.2). A two-sided P value < 0.05 was considered statistically significant.

Results

Model performance

Among the six candidate machine-learning models, ExtraTrees achieved the highest cross-validation AUC during model development (0.803). In the held-out test cohort, Test AUCs ranged from 0.733 to 0.827 (Table 1). LR achieved the highest Test AUC (0.827), followed by RF (0.822), while SVM and ExtraTrees both achieved a Test AUC of 0.811. RF yielded an accuracy of 0.714, an F1-score of 0.667, a sensitivity of 0.867, and a specificity of 0.639. RF also achieved the lowest Brier score (0.171), whereas XGBoost achieved the highest accuracy (0.725) and specificity (0.689). LR achieved the highest sensitivity (0.900) and AUPRC (0.662). Pairwise comparisons of ROC AUCs showed no statistically significant differences between any pair of models in the held-out test cohort (all two-sided DeLong P > 0.05). After comparative evaluation, RF was selected for subsequent interpretability analyses because it provided a favorable balance of discriminative, classification, and calibration performance.

Table 1
Diagnostic performance of different models for prediction of TOAST subtype (large-artery vs cardioembolic) in the held-out test cohort

Model	CV AUC	Test AUC	AUPRC	ACC	F1	SEN	SPE	Brier
LR	0.786	0.827	0.662	0.670	0.643	0.900	0.557	0.176
RF*	0.767	0.822	0.635	0.714	0.667	0.867	0.639	0.171
SVM	0.782	0.811	0.632	0.714	0.667	0.867	0.639	0.188
XGBoost	0.771	0.800	0.623	0.725	0.658	0.800	0.689	0.187
MLP	0.713	0.733	0.595	0.659	0.597	0.767	0.607	0.188
ExtraTrees	0.803	0.811	0.609	0.692	0.641	0.833	0.623	0.187

CV, cross-validation; **AUC**, area under the curve; **AUPRC**, area under the precision-recall curve; **ACC**, accuracy; **F1**, F1-score; **SEN**, sensitivity; **SPE**, specificity; **Brier**, Brier score. *Primary model selected for interpretability analysis.

Feature contribution and model interpretation

Following feature selection, the retained clinical and 3D thrombus radiomic features were used for machine-learning model development. The L1-regularized logistic regression coefficient paths showed that the absolute values of feature coefficients generally increased with increasing inverse regularization strength (C), indicating reduced shrinkage under weaker regularization (Fig. 3a).

In the primary RF model, both clinical and 3D thrombus radiomic features contributed to the discrimination between LAA and CE (Fig. 3b–d). Intraprocedural tirofiban use showed the largest mean absolute SHAP value (0.17), followed by admission LDL-C level (0.07), sex (0.04), and admission NIHSS score (0.04) (Fig. 3c). Several 3D thrombus radiomic features also ranked among the leading contributors, including the texture features original_glcm_MaximumProbability (0.03) and original_gldm_DependenceVariance (0.02), and the morphological feature original_shape_LeastAxisLength (0.02). Age also showed a mean absolute SHAP value of 0.02 (Fig. 3c). The Gini-based feature-importance analysis showed a broadly consistent ranking pattern, with both clinical variables and thrombus radiomic features among the top-ranked predictors (Fig. 3d).

Individual-level prediction explanation

Local SHAP waterfall analysis illustrated patient-specific prediction patterns in representative high- and low-risk cases for CE stroke in the primary RF model, starting from a base expected probability of 0.499 (Fig. 4). In the representative high-risk case (f(x) = 0.858), the prediction was shifted toward CE primarily by non-use of tirofiban (SHAP value, + 0.17), followed by admission NIHSS score (+ 0.04), sex (+ 0.04), admission LDL-C level (+ 0.04), and the radiomic features original_gldm_Dependence Variance (+ 0.02) and original_glcm_Maximum Probability (+ 0.02) (Fig. 4a). Conversely, in the representative low-risk case

(f(x) = 0.020), the prediction was shifted toward LAA primarily by tirofiban administration (− 0.16), followed by admission LDL-C level (− 0.07), original_glcm_Maximum Probability (− 0.05), admission NIHSS score (− 0.05), and sex (− 0.04) (Fig. 4b).

Model discrimination, calibration, and clinical utility

The test-set analyses demonstrated moderate discriminative performance across the evaluated models (Fig. 5). LR achieved the highest Test AUC (0.827), followed by RF (0.822), while SVM and ExtraTrees both achieved a Test AUC of 0.811 (Fig. 5a). Calibration performance varied across models without a consistent advantage of a single algorithm (Fig. 5b). Decision curve analysis showed that the prediction models provided positive net benefit over the treat-all and treat-none strategies across threshold probabilities of approximately 0.05–0.60 (Fig. 5c). For the RF model, a probability cutoff of 0.44 yielded 39 true-negative, 22 false-positive, 4 false-negative, and 26 true-positive classifications, corresponding to a sensitivity of 86.7% and a specificity of 63.9% (Fig. 5d).

Discussion

In this study, a machine learning model integrating preoperative 3D thrombus radiomics from CTA with clinical features was developed to differentiate LAA from CE in patients with acute MCAO. The RF model demonstrated moderate discrimination and favorable calibration in the held-out test cohort, while model interpretability analyses showed that both clinical variables and quantitative thrombus radiomic features contributed to the predictions. Notably, admission LDL-C, sex, admission NIHSS, and several thrombus texture and morphological features were among the leading contributors, suggesting that information related to both the patient and the three-dimensional thrombus phenotype may be relevant to stroke etiology. The positive net benefit observed across a relatively broad range of threshold probabilities further supports the potential clinical utility of this approach. Collectively, these findings indicate that preoperative CTA-based 3D thrombus radiomics may provide complementary information to conventional clinical assessment for the early differentiation of LAA and CE in acute MCAO.

The findings of the present study are broadly consistent with previous studies suggesting that CT-based thrombus imaging and radiomic analysis may provide information relevant to stroke etiology beyond conventional imaging metrics[14, 18–22]. Jiang et al. developed a multicenter thrombus radiomics model and demonstrated that pretreatment CT thrombus features could predict cardioembolic stroke[18], while Ma et al. reported that thrombus radiomics based on dual-energy CT could predict cardioembolism in patients with MCAO[14]. Ding et al. developed radiomics models using thrombus features derived from both NCCT and CTA and further integrated radiomic and clinical features into a combined model for differentiating large-artery atherosclerosis from cardioembolism [23].Other recent studies have further explored the associations of thrombus radiomic features with stroke etiology and thrombus composition[19–21], and dual-energy CT-based assessment of intrathrombus heterogeneity has suggested that such quantitative features may provide information related to stroke source and thrombus composition[22]. Collectively, these findings support the potential utility of quantitative

thrombus imaging for characterizing thrombus phenotypes associated with different stroke etiologies. In the context of these previous studies, the present study focused specifically on acute MCAO and evaluated three-dimensional thrombus radiomics derived from routinely acquired preoperative CTA as the imaging source.The present study also included a larger two-center cohort of 453 patients, compared with the 263 patients included in the study by Ding et al. [23], which may contribute to more robust estimation of model performance and potentially improve the generalizability of the findings. Compared with conventional attenuation measurements or selected thrombus imaging parameters, 3D radiomic analysis enables quantitative characterization of the spatial phenotype of the entire occlusive thrombus and may provide additional information on intrathrombus heterogeneity[14, 18, 19, 22]. The prominent contribution of texture and morphological features in the present model further supports the potential relevance of three-dimensional thrombus characteristics to etiological differentiation. At the modeling level, multiple machine-learning algorithms were compared within a unified framework integrating thrombus radiomic and clinical features. In addition, SHAP analysis was used to characterize the contributions of individual clinical and thrombus radiomic features to model predictions, providing further insight into the information underlying the model. Together, these findings build on previous thrombus imaging studies by focusing on a specific MCAO population, using routinely acquired preoperative CTA as the imaging source, evaluating a relatively larger multicenter cohort, and incorporating model interpretability to explore the clinical and thrombus radiomic information underlying etiological classification.

The potential clinical value of this approach may lie in its ability to derive quantitative thrombus information from routine preoperative CTA without requiring an additional imaging workflow, while clinical variables may provide complementary information for etiological assessment. Compared with thrombus analysis based on dual-energy CT, CTA is routinely performed in patients with acute LVO and may therefore be more readily incorporated into the existing emergency imaging pathway[12, 14, 22]. Rather than replacing established etiological evaluation, the proposed model may provide complementary information from the occlusive thrombus that is not fully captured by conventional visual assessment. This may be particularly relevant during the acute phase, when the complete etiological workup may not yet be available[8]. Early identification of a potential atherosclerotic mechanism may facilitate further assessment of the underlying arterial disease and consideration of reocclusion risk during EVT[8, 24, 25].

The biological plausibility of the imaging findings may be understood from three complementary aspects. First, differences in thrombus composition across stroke etiologies may contribute to distinct imaging phenotypes. Previous histopathological studies have suggested that thrombus composition and the proportions of its major components may differ between LAA- and CE-associated thrombi, with some studies reporting higher proportions of RBCs in LAA-associated thrombi and greater fibrin content in cardioembolic thrombi, although findings regarding platelet proportions have been inconsistent[26, 27]. Such differences in composition may influence CT attenuation and texture characteristics and thereby contribute to differences in radiomic features[18, 21, 27]. Second, differences in the spatial organization and three-dimensional morphology of thrombi may provide additional imaging information.

The spatial distribution of thrombus components and the structural characteristics related to thrombus formation may affect intrathrombus heterogeneity and morphological features, potentially explaining the contribution of texture and morphological radiomic features observed in the present model. However, these interpretations remain hypothetical because thrombus composition and internal organization were not directly assessed in the present study. Third, clinical variables may provide complementary information reflecting the vascular and clinical characteristics associated with different stroke etiologies. The contributions of LDL-C, sex, and admission NIHSS provide complementary clinical information relevant to stroke etiology.

This study has several limitations. First, its retrospective, two-center design and relatively limited sample size may have constrained the generalizability of the findings. The sample size was also influenced by case availability and the completeness of imaging data. Second, tirofiban administration occurred during EVT rather than before treatment initiation; therefore, its inclusion may introduce information related to intraprocedural treatment decisions and may limit the interpretation of the model as a purely pre-EVT prediction tool. Prospective, multicenter studies with larger sample sizes are needed to further validate the robustness and clinical utility of the model.

Conclusion

This study showed that integrating preoperative 3D CTA thrombus radiomics with clinical variables may provide useful information for differentiating LAA from CE in patients with acute MCAO. The RF model demonstrated competitive performance in the held-out test cohort, with an AUC of 0.822 and a sensitivity of 0.867. Among the leading contributors were admission LDL-C, sex, admission NIHSS, and the thrombus radiomic features original_glcm_MaximumProbability, original_gldm_DependenceVariance, and original_shape_LeastAxisLength. Decision curve analysis indicated a positive net benefit across threshold probabilities of approximately 0.05–0.60. These findings suggest that preoperative 3D thrombus radiomics combined with clinical information may provide complementary information to support the early differentiation of LAA and CE in acute MCAO.

Declarations

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

[Funding information blinded for peer review.]

Author Contribution

Yijie Huang and Xing Chen contributed equally to this work. Yijie Huang and Xing Chen designed the study and wrote the primary manuscript. Xing Chen analyzed the data. Xuejing Kong and Hongyu Ye were responsible for clinical and imaging data collection and curation. Bailong Xin, Zekang Ye, and Yijie Huang performed the regions of interest (ROI) segmentation and annotations. Xueli Cai and Jingping Sun reviewed and revised the manuscript. All authors approved the manuscript for publication.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author on reasonable request.

References

1. GBD 2021 Stroke Risk Factor Collaborators (2024) 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 23:973–1003. https://doi.org/10.1016/S1474-4422(24)00369-7
2. Malhotra K, Gornbein J, Saver JL (2017) Ischemic strokes due to large-vessel occlusions contribute disproportionately to stroke-related dependence and death: a review. Front Neurol 8:651. https://doi.org/10.3389/fneur.2017.00651
3. Turc G, Bhogal P, Fischer U, et al (2023) European stroke organisation (ESO) - european society for minimally invasive neurological therapy (ESMINT) guidelines on mechanical thrombectomy in acute ischemic stroke. J Neurointerventional Surg 15:e8. https://doi.org/10.1136/neurintsurg-2018-014569
4. Warner JJ, Harrington RA, Sacco RL, Elkind MSV (2019) Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke. Stroke 50:3331–3332. https://doi.org/10.1161/STROKEAHA.119.027708
5. Adams HP, Bendixen BH, Kappelle LJ, et al (1993) Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of org 10172 in acute stroke treatment. Stroke 24:35–41. https://doi.org/10.1161/01.str.24.1.35
6. Xiao H, He Z-Y, Li X-M, et al (2025) Large artery atherosclerotic versus cardioembolism subtypes of large hemispheric infarction in the middle cerebral artery. Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol 46:5173–5181. https://doi.org/10.1007/s10072-025-08347-9
7. Kleindorfer DO, Towfighi A, Chaturvedi S, et al (2021) 2021 guideline for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline from the American heart association/american stroke association. Stroke 52:e364–e467. https://doi.org/10.1161/STR.0000000000000375

8. Meschia JF (2023) Diagnostic evaluation of stroke etiology. Continuum 29:412–424. https://doi.org/10.1212/CON.0000000000001206
9. Brinjikji W, Nogueira RG, Kvamme P, et al (2021) Association between clot composition and stroke origin in mechanical thrombectomy patients: analysis of the stroke thromboembolism registry of imaging and pathology. J Neurointerventional Surg 13:594–598. https://doi.org/10.1136/neurintsurg-2020-017167
10. Brinjikji W, Duffy S, Burrows A, et al (2017) Correlation of imaging and histopathology of thrombi in acute ischemic stroke with etiology and outcome: a systematic review. J Neurointerventional Surg 9:529–534. https://doi.org/10.1136/neurintsurg-2016-012391
11. Hund HM, Boodt N, Hansen D, et al (2023) Association between thrombus composition and stroke etiology in the MR CLEAN registry biobank. Neuroradiology 65:933–943. https://doi.org/10.1007/s00234-023-03115-y
12. Mead GE, Sposato LA, Sampaio Silva G, et al (2023) A systematic review and synthesis of global stroke guidelines on behalf of the world stroke organization. Int J Stroke Off J Int Stroke Soc 18:499–531. https://doi.org/10.1177/17474930231156753
13. Hofmeister J, Bernava G, Rosi A, et al (2020) Clot-based radiomics predict a mechanical thrombectomy strategy for successful recanalization in acute ischemic stroke. Stroke 51:2488–2494. https://doi.org/10.1161/STROKEAHA.120.030334
14. Ma Y, Dai Y, Zhao Y, et al (2024) Radiomics model based on dual-energy CT can determine the source of thrombus in strokes with middle cerebral artery occlusion. Neuroradiology 66:1681–1691. https://doi.org/10.1007/s00234-024-03422-y
15. Bruggeman AAE, Aberson N, Kappelhof M, et al (2022) Association of thrombus density and endovascular treatment outcomes in patients with acute ischemic stroke due to M1 occlusions. Neuroradiology 64:1857–1867. https://doi.org/10.1007/s00234-022-02971-4
16. Zwanenburg A, Vallières M, Abdalah MA, et al (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 295:328–338. https://doi.org/10.1148/radiol.2020191145
17. van Timmeren JE, Cester D, Tanadini-Lang S, et al (2020) Radiomics in medical imaging-"how-to" guide and critical reflection. Insights Imaging 11:91. https://doi.org/10.1186/s13244-020-00887-2
18. Jiang J, Wei J, Zhu Y, et al (2023) Clot-based radiomics model for cardioembolic stroke prediction with CT imaging before recanalization: a multicenter study. Eur Radiol 33:970–980. https://doi.org/10.1007/s00330-022-09116-4
19. Porto-Álvarez J, Mosqueira Martínez AJ, Martínez Fernández J, et al (2025) Predicting stroke etiology with radiomics: a retrospective study. Med Sci 13:98. https://doi.org/10.3390/medsci13030098
20. Santo BA, Janbeh Sarayi SMM, McCall AD, et al (2023) Multimodal CT imaging of ischemic stroke thrombi identifies scale-invariant radiomic features that reflect clot biology. J Neurointerventional Surg 15:1257–1263. https://doi.org/10.1136/jnis-2022-019967

21. Gudino A, Sagues E, Dier C, et al (2026) Clot composition profiling in large vessel occlusion stroke via radiomics. Ann Neurol 99:1179–1188. https://doi.org/10.1002/ana.78160
22. Jiang J, Wang S, Xiao F, et al (2025) Dual-Energy CT-Based Assessment of Thrombotic Heterogeneity for Predicting Stroke Source and Response to Machine Thrombectomy: A Step Toward Visualization Thrombus Treatment. Adv Sci 12:e17295. https://doi.org/10.1002/advs.202417295
23. Ding R, An K, Wang F, et al (2025) Analysis of a clot-based combined radiomics model for predicting embolic etiology in acute ischemic stroke patients. Eur J Radiol OPEN 15:100709. https://doi.org/10.1016/j.ejro.2025.100709
24. Baek BH, Yoon W, Lee YY, et al (2021) Intravenous Tirofiban Infusion After Angioplasty and Stenting in Intracranial Atherosclerotic Stenosis-Related Stroke. Stroke 52:1601–1608. https://doi.org/10.1161/strokeaha.120.033551
25. de Havenon A, Zaidat OO, Amin-Hanjani S, et al (2023) Large Vessel Occlusion Stroke due to Intracranial Atherosclerotic Disease: Identification, Medical and Interventional Treatment, and Outcomes. Stroke 54:1695–1705. https://doi.org/10.1161/STROKEAHA.122.040008
26. Sujijantarat N, Templeton KA, Antonios JP, et al (2024) Is clot composition associated with cause of stroke? A systematic review and meta-analysis. Stroke 4:e001426. https://doi.org/10.1161/SVIN.124.001426
27. Patel TR, Santo BA, Jenkins TD, et al (2023) Biologically informed clot histomics are predictive of acute ischemic stroke etiology. Stroke 3:e000536. https://doi.org/10.1161/SVIN.122.000536

Figures

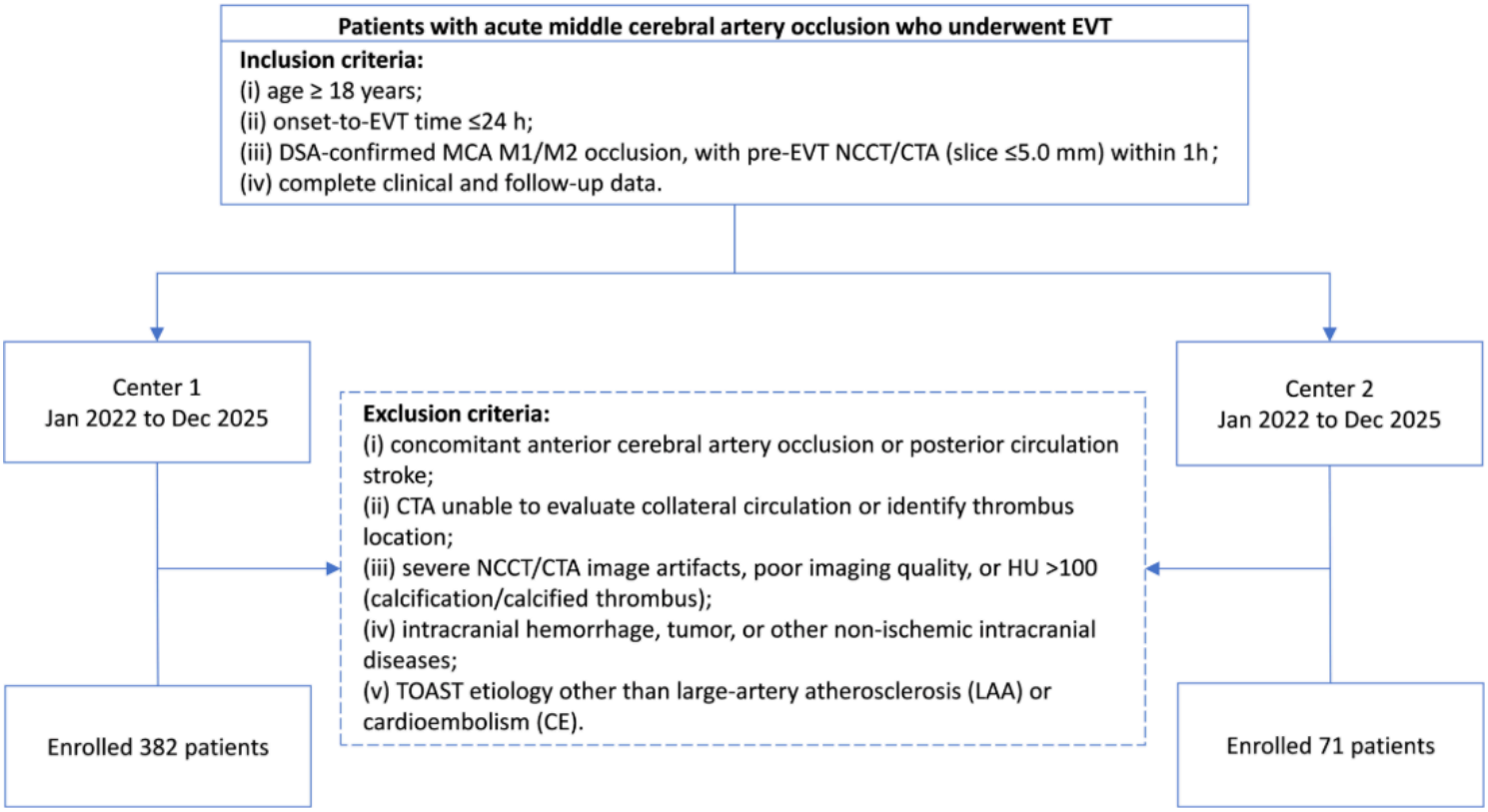


Figure 1

Flowchart of patient selection. EVT: endovascular treatment; MCA: middle cerebral artery; NCCT: non-contrast computed tomography; CTA: computed tomography angiography; HU: Hounsfield unit; TOAST: Trial of Org 10172 in Acute Stroke Treatment criteria; LAA: large-artery atherosclerosis; CE: cardioembolism.

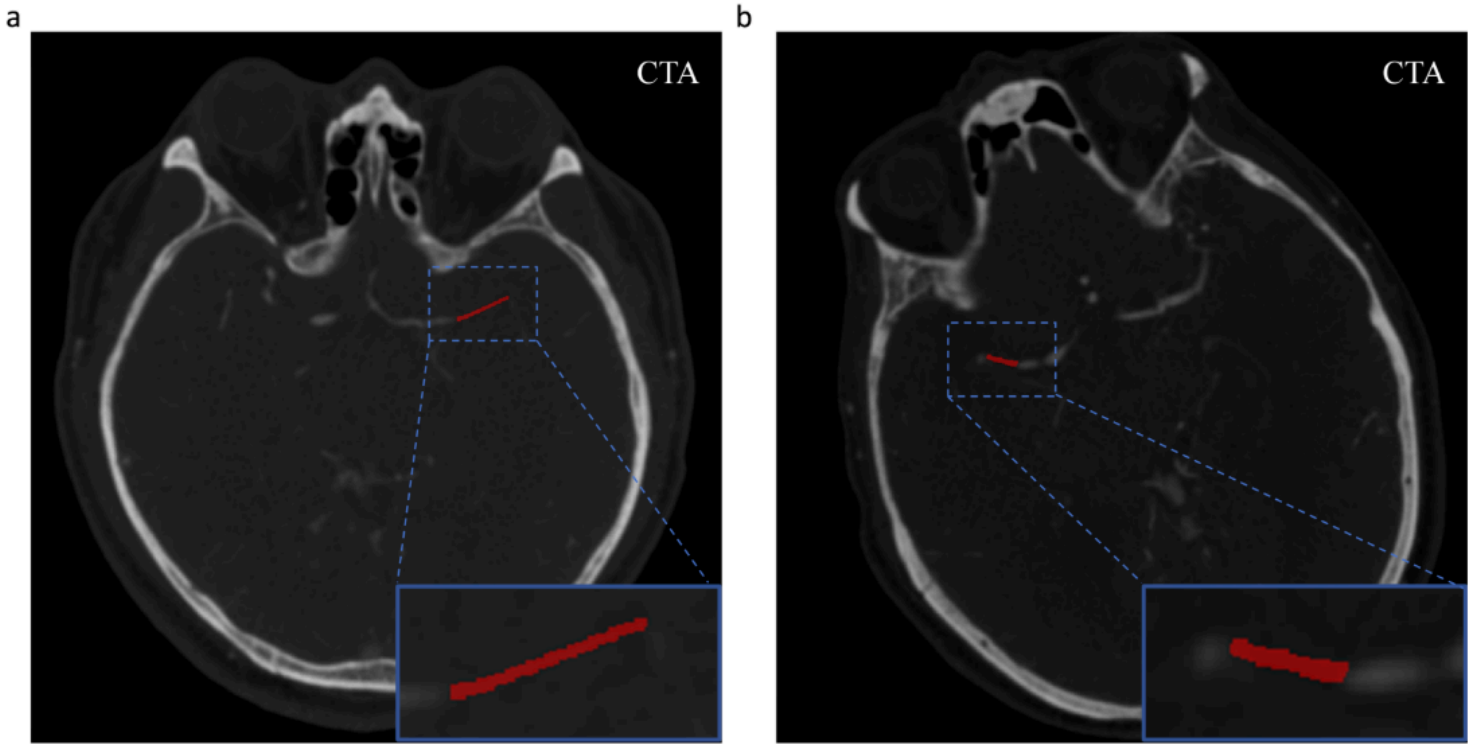


Figure 2

Representative examples of thrombus segmentation on preoperative CTA.

Axial CTA images from two representative patients (a, b) show the manually delineated thrombus regions of interest (red overlay). The bottom-right insets display the magnified views detailing the thrombus contours.

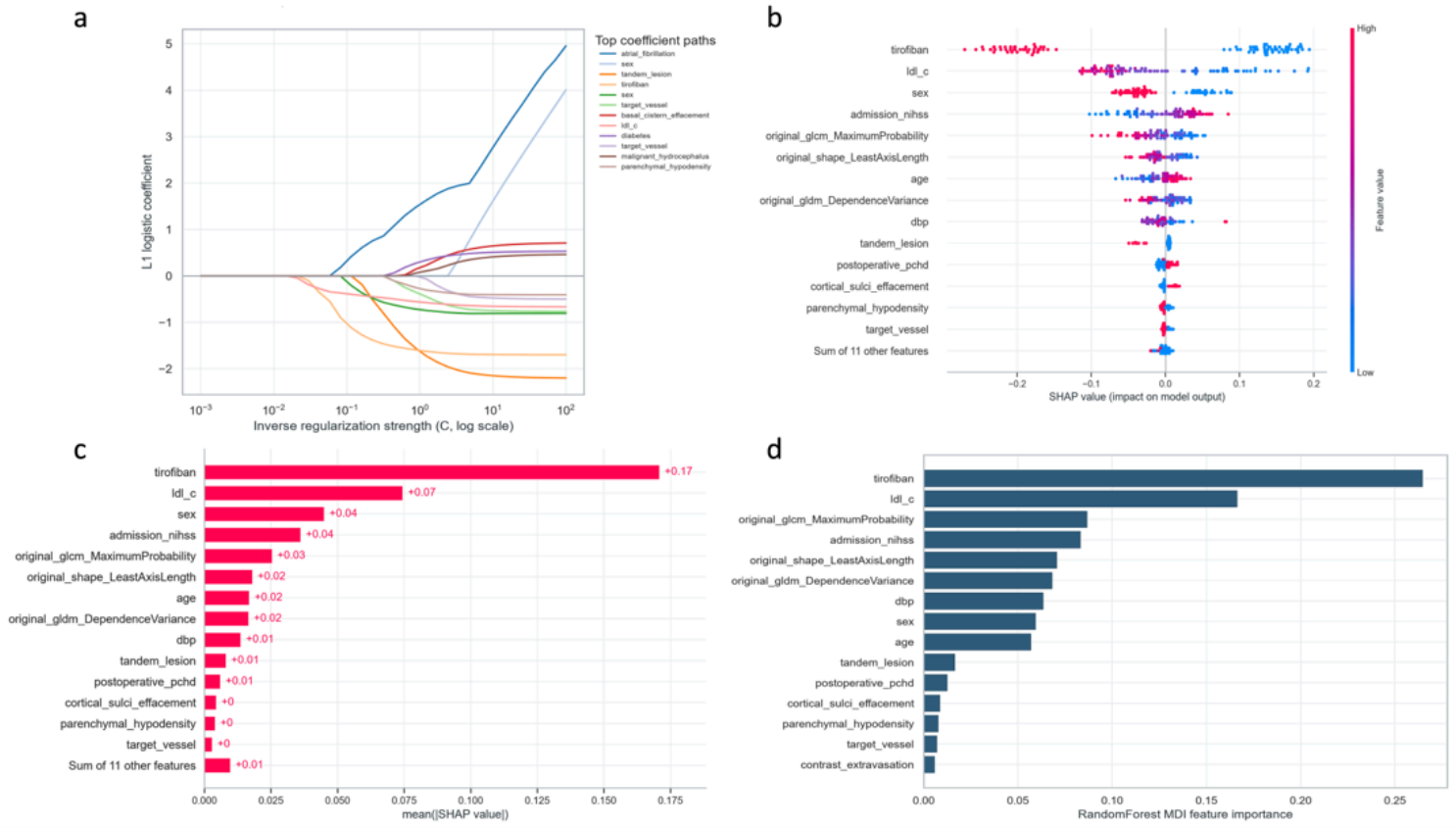


Figure 3

Feature selection and interpretability analysis for differentiating large-artery atherosclerosis and cardioembolism. a L1 logistic regression coefficient path plot illustrating changes in feature coefficients across varying inverse regularization strengths. b SHAP beeswarm plot showing the distribution of feature effects on model output for the RF model. c Ranking of features based on mean absolute SHAP values. d Feature importance ranking based on the Gini importance of the RF model.

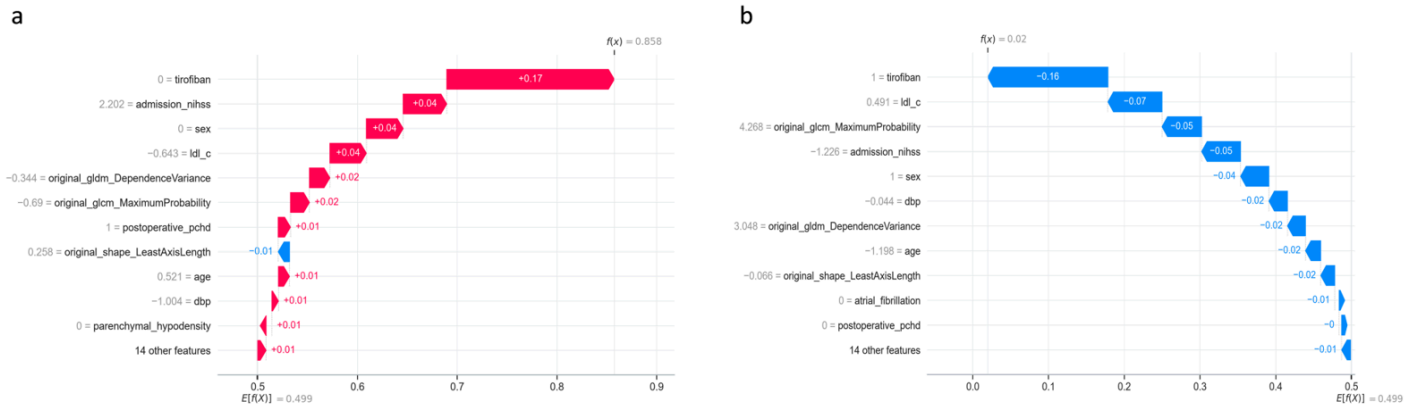


Figure 4

SHAP waterfall plots illustrating individual predictions for TOAST subtypes (large-artery vs. cardioembolic). a SHAP waterfall plot detailing feature contributions for the patient with the highest predicted risk (f(x) = 0.858). b SHAP waterfall plot detailing feature contributions for the patient with the lowest predicted risk (f(x) = 0.020). Each plot demonstrates how individual features push the Random Forest model prediction from the expected base value E[f(x)] = 0.499 to the final output probability f(x).

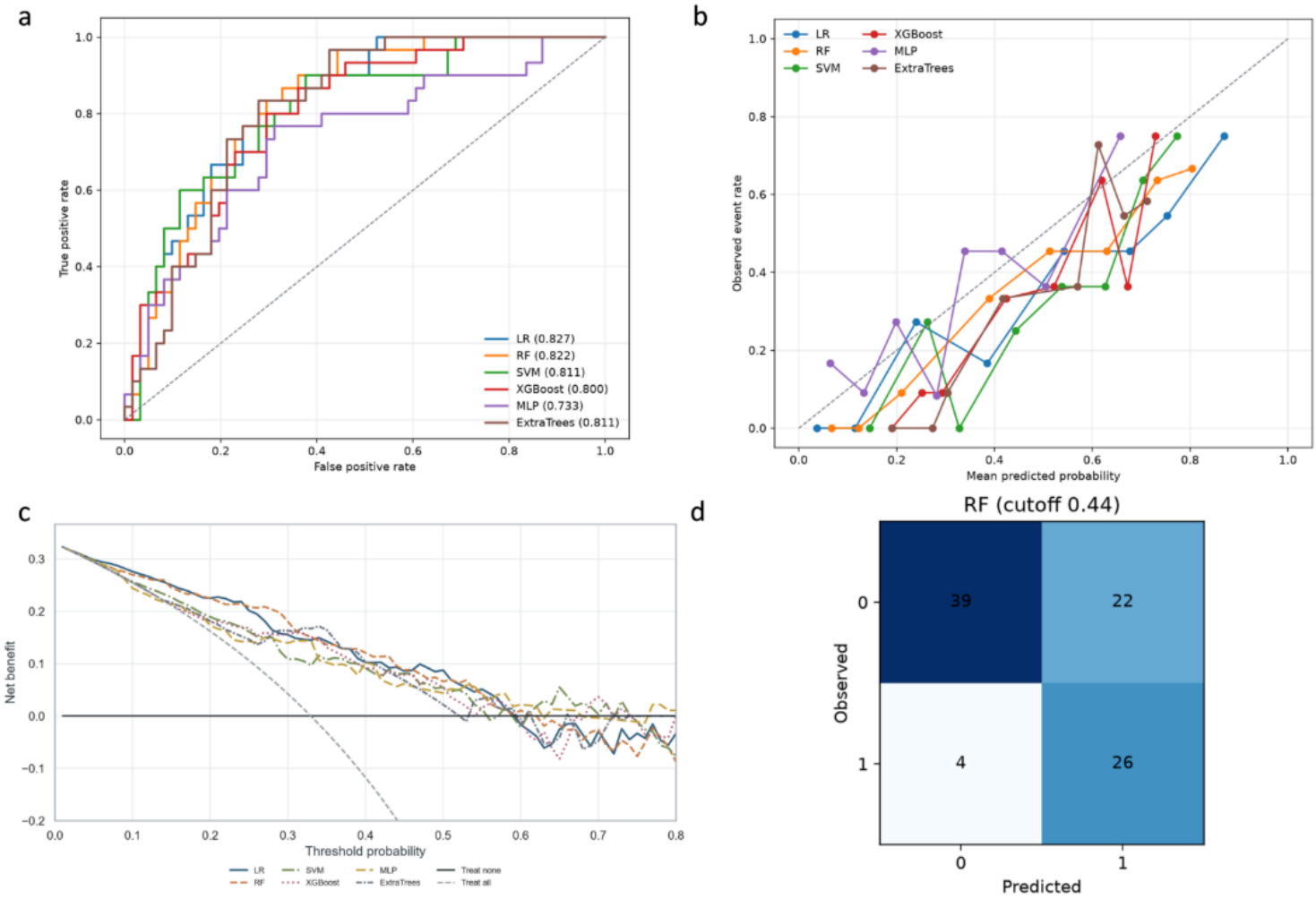


Figure 5

Performance and evaluation of machine-learning models for predicting TOAST subtypes (large-artery vs. cardioembolic). a Receiver operating characteristic (ROC) curves of candidate models on the held-out test cohort. b Calibration curves demonstrating the agreement between predicted probabilities and observed event rates across models. c Decision curve analysis (DCA) assessing net clinical benefit across a spectrum of threshold probabilities. d Confusion matrix of the Random Forest (RF) model at a probability cutoff of 0.44.