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.

Construction and Validation of a Multimodal Imaging and Machine Learning–Based Prognostic Prediction Model for Aneurysmal Subarachnoid Hemorrhage: A Multicenter Retrospective Cohort Study (TRIPOD-AI Compliant)

Xinggang Xiang

mz19891018@163.com

Xinjiang Medical University Fourth Clinical Medical College: Traditional Chinese Medicine Hospital of Xinjiang Uyghur Autonomous Region https://orcid.org/0000-0002-3894-5451

Lin Lin

Xinjiang Medical University Fourth Clinical Medical College: Traditional Chinese Medicine Hospital of Xinjiang Uyghur Autonomous Region

Guanjun Feng

Xinjiang Uyghur People's Hospital: People's Hospital of Xinjiang Uygur Autonomous Region

Chun Fang

Tongji Hospital Affiliated to Tongji University

Zhen Ma

Taylor's University - Lakeside Campus: Taylor's University

Research Article

Keywords: aneurysmal subarachnoid hemorrhage, multimodal imaging, radiomics, machine learning, prognostic prediction, multicenter validation, TRIPOD-AI

Posted Date: May 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9283012/v1

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

Additional Declarations:
Table 1 to 3 are available in the Supplementary Files section.

Construction and Validation of a Multimodal Imaging and Machine Learning–Based Prognostic Prediction Model for Aneurysmal Subarachnoid Hemorrhage: A Multicenter Retrospective Cohort Study

Xinggang Xiang, MD, [1] **Lin Lin, MD**, [1] **Guanjun Feng, MD**, [2] **Chun Fang, MD, PhD**, [3]*

Zhen Ma, PhD [4]

[1] Department of Neurosurgery, Xinjiang Uyghur Autonomous Region Hospital of Traditional Chinese Medicine, Urumqi 830000, China

[2] Department of Neurosurgery, People's Hospital of Xinjiang Uyghur Autonomous Region, Urumqi 830001, China

[3] Department of Interventional Therapy, Tongji Hospital, Tongji University School of Medicine, Shanghai 200065, China

[4] Faculty of Medical Sciences, Taylor's University, Subang Jaya, Selangor 47500, Malaysia

***Corresponding author:** Chun Fang, MD, PhD

Department of Interventional Therapy, Tongji Hospital, Tongji University School of Medicine

389 Xincun Road, Putuo District, Shanghai 200065, China

Email: 891936613@qq.com

Article Type: Original Work (Clinical Investigation)

Word count (body): 3,291 words

Abstract word count: 298 words

Number of Figures: 6

Number of Tables: 3

References: 29

Keywords: aneurysmal subarachnoid hemorrhage; multimodal imaging; radiomics; machine

learning; prognostic prediction; multicenter validation; TRIPOD-AI

Reporting guideline: TRIPOD + AI (TRIPOD-AI) 2024 checklist is provided as a separate supplementary file.

Details Page

1. Compliance with Instructions to Authors

The authors confirm that this manuscript has been prepared in full compliance with the Neurocritical Care Instructions for Authors (February 2025 update), including all requirements relating to manuscript structure, word count, abstract format, figures, tables, references, ethical statements, and reporting checklists.

2. Author Contributions

The contributions of each author, in accordance with the International Committee of Medical Journal Editors (ICMJE) authorship criteria, are as follows:

Xinggang Xiang: conceptualization, study design, patient enrollment at the Xinjiang Hospital of Traditional Chinese Medicine, clinical data collection, imaging feature extraction, statistical analysis, drafting of the original manuscript, and revision of the manuscript for important intellectual content.

Lin Lin: patient enrollment at the Xinjiang Hospital of Traditional Chinese Medicine, clinical data collection, radiomic feature extraction and quality control, literature review, drafting of Methods and Results sections, and revision of the manuscript.

Guanjun Feng: patient enrollment at the People's Hospital of Xinjiang Uyghur Autonomous Region, standardization of multicenter imaging acquisition protocols, inter-center quality control audits, data curation and management, and critical revision of the manuscript.

Chun Fang (corresponding author): conceptualization, overall study supervision, patient enrollment and external validation at Tongji Hospital (Shanghai), statistical analysis plan, model development oversight, interpretation of results, revision of the manuscript for important intellectual content, and responsibility for all communication with the journal.

Zhen Ma: methodological oversight of machine learning model development and validation, SHAP interpretability analysis, critical revision of the manuscript for methodological and intellectual content, and senior mentorship.

All authors made substantial contributions to the work, drafted or critically revised the manuscript for important intellectual content, gave final approval of the version to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

3. Authorship Requirements and Final Approval

The authors confirm that all named authors meet the four ICMJE criteria for authorship. No individual who meets the authorship criteria has been omitted, and no individual who does not meet these criteria is listed as an author. The final version of the manuscript has been read and approved by all authors prior to submission.

4. Originality and Exclusive Submission

The authors confirm that this manuscript has not been published previously, in whole or in part, and is not currently under consideration by any other journal. No portion of the work has been submitted elsewhere for simultaneous consideration. No data, text, or figures have been previously published in any language.

5. Ethical Approval and Informed Consent

This retrospective multicenter cohort study was conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments. The study protocol was reviewed and approved by the Institutional Review Boards of all three participating centers:

Xinjiang Uyghur Autonomous Region Hospital of Traditional Chinese Medicine (approval number: XJZYYKY-2024-001);

People's Hospital of Xinjiang Uyghur Autonomous Region (approval number: XJRMYY-2024-038);

Tongji Hospital, Tongji University School of Medicine (approval number: SH-TJYY-2024-127). Because this was a retrospective analysis of de-identified clinical and imaging data, the requirement for written informed consent was waived by the Institutional Review Boards of all participating centers. For this type of study, formal consent is not required. All data were fully de-identified prior to analysis. No identifying information about individual participants is included in this article.

6. Disclosure of Conflicts of Interest

All authors have completed the ICMJE uniform disclosure form. The authors declare that they have no financial or non-financial conflicts of interest, including no research grants from commercial entities relevant to this work, no honoraria or speaker fees, no consultancy arrangements, no equity ownership, no patents or intellectual property rights, and no other personal or professional relationships that could inappropriately influence or bias the content of this manuscript.

7. Reporting Checklist

The authors confirm the use of the TRIPOD + AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis – Artificial Intelligence) reporting guideline (Collins GS et al., BMJ 2024). A fully completed TRIPOD-AI checklist is provided as a separate supplementary file accompanying this submission. The manuscript has also been prepared with reference to the STROBE statement for observational studies where applicable.

8. Funding

This work was supported by the following funding sources:

“Tianshan Elite” High-level Talent Training Program in Medicine and Health (Grant No. TsYc202401B175) — recipient: Xinggang Xiang;

The funders had no role in study design, data collection, data analysis, interpretation of results, preparation of the manuscript, or the decision to submit the work for publication.

9. Use of Artificial Intelligence (AI)–Assisted Technology

In accordance with the Neurocritical Care policy on AI-assisted technology, the authors disclose the following:

Research, analysis, and data interpretation: No generative AI or large language model (LLM) was used for study design, patient data collection, radiomic feature extraction, statistical analysis, model development, or interpretation of scientific results. All methodological and analytical work was performed by the human authors using the statistical and machine-learning software specified in the Methods section (PyRadiomics 3.0, scikit-learn 1.3, xgboost 1.7.6, R 4.3.1).

Language editing: Limited language polishing of the English manuscript was performed with the assistance of an AI-based writing aid (ChatGPT, OpenAI, GPT-4 model, 2024 version) restricted to grammar, syntax, and stylistic refinement of text already composed by the authors. The tool was not used to generate scientific content, references, tables, figures, or conclusions. All AI-assisted edits were reviewed, verified for accuracy, and approved by the human authors, who retain full responsibility for the content of the manuscript.

Figures and tables: No AI-based image generation tool was used to create any figure or table in this manuscript. All figures were produced from the study data using R (version 4.3.1) and Python (version 3.8) plotting libraries.

The authors assert that no plagiarism, fabricated content, or AI-generated citations are present in this manuscript, and that full responsibility for the accuracy, integrity, and originality of the work

rests with the human authors.

10. Data and Code Availability

The de-identified core dataset supporting the findings of this study is available from the corresponding author on reasonable request, subject to approval by the Institutional Review Boards of the participating centers. The statistical analysis and machine-learning code have been deposited in a public GitHub repository: https://github.com/mz19891018/asah-multimodal.

Abstract

Background and Objective

Aneurysmal subarachnoid hemorrhage (aSAH) carries a 30-day mortality of 30–50%, yet current clinical grading scales achieve area under the receiver operating characteristic curve (AUC) values of only 0.70–0.78 for outcome prediction. We aimed to construct and externally validate a multimodal imaging and machine learning (ML)–based prognostic model for aSAH.

Methods

This multicenter retrospective cohort study (TRIPOD-AI compliant) enrolled 547 aSAH patients from three centers (January 2020–June 2024). The training cohort comprised 335 patients from two Xinjiang centers; the external validation cohort comprised 212 patients from Tongji Hospital, Shanghai (geographically independent, >3,500 km away). Features included CTA aneurysm morphology (9), DWI-ADC (3), PWI perfusion parameters (6), 13 LASSO-selected radiomic features, and 17 clinical variables. Three ML models—logistic regression (LR), random forest (RF), and XGBoost—were developed using nested five-fold stratified cross-validation with Bayesian hyperparameter optimization. Interpretability was assessed by SHAP analysis. The primary endpoint was 90-day modified Rankin Scale score (mRS 0–2 = favorable).

Results

Of 547 patients, 176 (32.2%) had unfavorable outcomes. In external validation, XGBoost achieved the best performance: AUC 0.876 (95% CI 0.826–0.926), accuracy 83.0%, sensitivity 81.4%, specificity 84.1%, Brier score 0.142, and adequate calibration (Hosmer–Lemeshow P = 0.308). The multimodal model significantly outperformed the clinical-only model (ΔAUC = 0.071, DeLong P = 0.003). SHAP analysis identified GCS score, ADCmean, radiomic GLCM_Entropy, modified Fisher score, and maximum aneurysm diameter as the five most important predictors.

Decision curve analysis confirmed net clinical benefit across threshold probabilities of 15–80%.

Conclusions

A multimodal imaging–integrated ML model validated across three geographically independent centers demonstrated good discrimination and generalizability for early risk stratification in aSAH. Prospective validation with automated segmentation pipelines is required before clinical implementation.

Keywords: aneurysmal subarachnoid hemorrhage; multimodal imaging; radiomics; machine learning; prognostic prediction; multicenter validation; TRIPOD-AI

1. Introduction

Aneurysmal subarachnoid hemorrhage (aSAH) is an acute cerebrovascular event caused by the rupture of an intracranial aneurysm, accounting for approximately 5% of all stroke subtypes. The 30-day case fatality rate ranges from 30% to 50%, and more than 50% of survivors sustain permanent neurological deficits [1,2]. Chinese epidemiological data estimate an annual incidence of 2–22.5 per 100,000 population, a figure that continues to rise with demographic aging and improving neuroimaging detection rates [1,3].

Current prognostic assessment relies predominantly on the Hunt-Hess scale [4], modified Fisher score [5], and Glasgow Coma Scale (GCS)—tools that, despite their operational convenience, are inherently limited by single time-point subjective assessment and an inability to objectively quantify key pathophysiological processes such as delayed cerebral ischemia (DCI), cerebral vasospasm, and dynamic intracranial pressure changes. Published AUC values for existing clinical scoring systems in predicting unfavorable outcome range from only 0.70 to 0.78 [6], which falls short of the discriminative precision required for individualized clinical decisions. Our central hypothesis is that multimodal imaging radiomic features can provide independent incremental prognostic information beyond traditional clinical scores, enabling construction of a substantially superior composite ML model.

Multimodal imaging offers new dimensions for aSAH outcome prediction: CTA precisely characterizes aneurysm morphological parameters [7]; DWI quantifies cytotoxic edema in acute ischemic territories via ADC values [8]; and PWI assesses cerebral blood flow (CBF), cerebral blood volume (CBV), and mean transit time (MTT) to support early DCI warning [9]. Radiomics enables high-throughput extraction of quantitative imaging features [10,11], while ML algorithms excel at processing high-dimensional, nonlinear data [12]; random forest [13] and XGBoost [14]

have demonstrated superior generalization in medical prognostic modeling.

Studies that systematically integrate multimodal imaging radiomics with clinical grading and validate findings in geographically independent multicenter external cohorts remain scarce [15]. Major methodological limitations in prior prognostic models include small sample sizes, absence of truly geographically independent external validation, inadequate reporting of radiomic feature reproducibility, and insufficient model interpretability. The present study enrolled 547 patients from three centers to construct and validate LR, RF, and XGBoost prognostic models integrating multimodal imaging features, employed SHAP interpretability analysis to provide transparent clinical decision support, and established the incremental methodological and performance value of this approach through systematic comparison with best-available published models.

2. Patients and Methods

2.1 Study Design and Ethical Approval

This is a multicenter retrospective cohort study conducted in accordance with the TRIPOD + AI (TRIPOD-AI) 2024 reporting guideline [29]. The study protocol was approved by the Institutional Review Boards of all three participating centers (approval numbers: XJZYYKY-2024-001, XJRMYY-2024-038, SH-TJYY-2024-127). The requirement for informed consent was waived by the ethics committees owing to the retrospective design and the use of de-identified data. All data were de-identified prior to analysis.

2.2 Use of Artificial Intelligence

No generative artificial intelligence or large language model was used for study design, data analysis, feature extraction, model development, interpretation of results, or generation of figures, tables, or references. Limited AI-assisted language editing (ChatGPT, OpenAI, GPT-4, 2024 version) was used solely for grammatical and stylistic refinement of English text composed by the

authors; all AI-assisted edits were reviewed and verified by the authors, who retain full responsibility for the manuscript content. See the Details Page for the complete AI disclosure statement.

2.3 Study Population

Inclusion criteria: (1) aSAH confirmed by DSA, CTA, or MRA; (2) time from onset to admission ≤72 hours; (3) CTA and multimodal MRI (including DWI and PWI sequences) completed within 48 hours of admission; (4) complete clinical data with 90-day follow-up available; (5) age 18–80 years.

Exclusion criteria: (1) non-aneurysmal SAH; (2) death within 24 hours of admission; (3) severe renal impairment (CrCl <30 mL/min); (4) suboptimal image quality; (5) concurrent active intracranial lesion; (6) prior stroke history or baseline neurological disability (baseline mRS ≥ 2).

A total of 547 patients were enrolled (Figure 1). The training cohort comprised 335 patients from two Xinjiang centers (Xinjiang Hospital of TCM, n = 198; People's Hospital of Xinjiang, n = 137); the external validation cohort comprised 212 patients from Tongji Hospital, Shanghai.

2.4 Standardized Imaging Acquisition Protocol

To mitigate multicenter imaging heterogeneity on radiomic feature robustness, a unified acquisition and post-processing protocol was established across all three centers prior to study initiation.

CTA: All centers used ≥64-slice CT scanners (GE Discovery CT750 HD or Siemens SOMATOM Definition Flash); tube voltage 120 kV, automated milliampere modulation, slice thickness 0.625 mm, contrast volume 1.5 mL/kg (iohexol 350 mgI/mL), injection rate 4.5 mL/s, arterial-phase trigger threshold 150 HU.

DWI: 3.0-T MRI (GE Signa Pioneer or Siemens Prisma); TR/TE = 3500/80 ms (±10%), b-values

= 0 and 1000 s/mm², slice thickness 5 mm, matrix 128×128. ADC maps were generated automatically by the scanner; all centers used identical post-processing software (Functool 4.5, GE).

PWI (DSC perfusion): TR/TE = 1500/30 ms, flip angle 60°, slice thickness 5 mm, temporal resolution 1.5 s/frame, 48 frames; contrast dose 0.1 mmol/kg (gadobenate dimeglumine), injection rate 5 mL/s. Post-processing was performed uniformly with NordicICE software (v2.3.14) using SVD deconvolution; the arterial input function (AIF) was selected manually by a center-level radiologic technologist (criteria: signal-drop peak >40%, time-to-peak 15–25 s). Intersite AIF consistency was assessed by quarterly blinded review (intraclass correlation coefficient, ICC ≥ 0.81).

Radiologic technologist training: Technologists at all centers completed a standardized two-day hands-on training program covering region of interest (ROI) delineation protocols; reproducibility audits were conducted semi-annually (20 independently re-analyzed cases per center; ICC threshold ≥ 0.75).

Despite these standardization measures, residual inter-center PWI parameter heterogeneity persisted, primarily attributable to MRI scanner model differences and AIF manual selection variability—an inherent limitation of multicenter radiomics studies, discussed further in the Limitations section.

2.5 Imaging Feature Extraction

CTA aneurysm morphological parameters (9 features): maximum diameter, neck width, aneurysm height, aspect ratio, size ratio, non-sphericity index (NSI), and presence of daughter sac, measured with Mimics Research 21.0. DWI-ADC features (3 features): ADCmin, ADCmean, and ADCsd, extracted from acute ischemic ROIs (ADC < 620×10^{-6} mm²/s). PWI cerebral blood flow

parameters (6 features): ipsilateral-to-contralateral ratios for rCBF, rCBV, and rMTT. Radiomic features (972 features): extracted from CTA aneurysm ROIs using PyRadiomics 3.0, encompassing first-order statistics, gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), and shape features, augmented by wavelet and Laplacian-of-Gaussian filter transformations.

Clinical feasibility note: Manual aneurysm ROI segmentation requires approximately 10–15 minutes per case by an experienced radiologic technologist; feature computation requires an additional 5–10 minutes; DWI-ADC and PWI parameter extraction requires 10–15 minutes. The total workflow from imaging acquisition to model output is approximately 25–45 minutes, requiring technologist-assisted manual segmentation. This is feasible within the routine clinical decision window of 24–48 hours for aSAH management but is not capable of real-time deployment in the hyperacute phase (<6 hours). Development of automated segmentation pipelines (e.g., nnU-Net) is a prerequisite for bedside real-time implementation.

2.6 Feature Selection

Inter-rater ICC evaluation was performed in a 50-case subset; 382 features with ICC < 0.75 were excluded, retaining 590 features. Ten-fold cross-validated least absolute shrinkage and selection operator (LASSO) regression (optimal $\lambda = 0.023$; Figure 2) selected 13 core radiomic features, all with ICC $\geq$ 0.82 (mean 0.887, range 0.820–0.951), reflecting good cross-center measurement reproducibility.

2.7 Model Development

Three ML models were constructed: (1) logistic regression (LR) with LASSO regularization; (2) random forest (RF; n_estimators = 500); (3) XGBoost with Bayesian hyperparameter optimization (Optuna). Nested five-fold stratified cross-validation was applied—the outer loop for unbiased

performance estimation and the inner loop for hyperparameter search (≥100 iterations per model). Given mild class imbalance (32.2% unfavorable outcomes), class-weight adjustment (class weight = "balanced") was used rather than Synthetic Minority Oversampling Technique (SMOTE). A clinical-only model containing 17 clinical variables served as the reference comparator. All models were implemented in scikit-learn 1.3 and xgboost 1.7.6 (Python 3.8).

2.8 Outcome Assessment and Quality Control

The 90-day mRS was assessed by structured telephone interview using the simplified modified Rankin Scale questionnaire (sRmRS), following the validation standards of Wilson et al. [27] and Janssen et al. [28] (ICC ≥ 0.83, weighted $\kappa = 0.78$). Follow-up raters were blinded to in-hospital treatment modality and imaging findings. Approximately 10% of cases (~55 patients) underwent independent re-assessment by a second rater (target $\kappa \geq 0.75$). The loss-to-follow-up rate was 3.2%. We acknowledge a systematic misclassification rate of approximately 25% at the mRS 2/3 boundary with telephone administration, as documented by Janssen et al. [28]; this limitation is discussed in detail below.

2.9 Statistical Analysis

Model performance metrics included AUC (DeLong method for pairwise comparison), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, and Brier score. Calibration was assessed via calibration curves and the Hosmer-Lemeshow test. Net clinical benefit was quantified by decision curve analysis (DCA). SHapley Additive exPlanations (SHAP) analysis [16] provided feature importance ranking and individual-level interpretability. All analyses were performed in R 4.3.1 and Python 3.8 (two-tailed $P < 0.05$ as the significance threshold).

Statistical power note: The external validation cohort (n = 212, approximately 69 unfavorable-

outcome events) has approximately 82% power at $\alpha = 0.05$ to detect $\Delta AUC = 0.07$ (DeLong variance estimation), supporting the primary conclusions. However, the reliability of Brier score estimates and subgroup analyses is limited by sample size and results should be interpreted with caution.

Sample size rationale: This was a retrospective consecutive-enrollment study; the final sample of 547 patients represents all eligible cases at the three centers from January 2020 to June 2024 meeting inclusion/exclusion criteria. Post-hoc power calculation confirms adequate power (≥80%) for the primary endpoint ($\Delta AUC \geq 0.07$) in both the training and external validation cohorts.

3. Results

3.1 Patient Enrollment and Baseline Characteristics

A total of 547 aSAH patients were enrolled (training cohort n = 335, external validation cohort n = 212); 176 patients (32.2%) had unfavorable outcomes. Baseline characteristics of both cohorts are presented in Table 1. All variables were comparable between cohorts (all $P > 0.05$) except serum sodium ($P = 0.041$) and treatment modality distribution ($P = 0.028$). Sensitivity analysis demonstrated that the treatment modality difference had negligible impact on XGBoost AUC ($\Delta AUC < 0.01$).

3.2 Radiomic Feature Selection

After excluding 382 features with ICC < 0.75, 590 features were retained. Ten-fold cross-validated LASSO regression (optimal $\lambda = 0.023$) identified 13 core features: GLCM_Entropy, GLRLM_LRLGLE, FirstOrder_Skewness, Sphericity, Compactness2, and others. Mean ICC = 0.887 (range 0.820–0.951), reflecting good cross-center measurement reproducibility.

3.3 Model Discriminative Performance

Complete performance metrics for all three ML models are presented in Table 2 and Figure 3. In

external validation, XGBoost achieved the highest performance: AUC = 0.876 (95% CI: 0.826–0.926), significantly outperforming RF (AUC = 0.843, DeLong P = 0.031) and LR (AUC = 0.801, P = 0.003). The multimodal model significantly outperformed the clinical-only reference model (AUC = 0.805; ΔAUC = 0.071, DeLong P = 0.003).

3.4 Model Calibration

Calibration curves (Figure 4) demonstrated good agreement between predicted probabilities and observed event rates for all three models. XGBoost achieved the best calibration: Hosmer-Lemeshow χ^2 = 8.74, df = 8, P = 0.308; RF: H-L P = 0.216; LR: H-L P = 0.089 (all within acceptable range).

3.5 SHAP Feature Importance Analysis

SHAP analysis (Figure 5) identified the ten most important predictors of unfavorable outcome: (1) GCS score (mean |SHAP| = 0.312); (2) ADCmean (0.228); (3) radiomic GLCM_Entropy (0.189); (4) modified Fisher score (0.176); (5) maximum aneurysm diameter (0.143); (6) rCBF ratio (0.121); (7) Hunt-Hess grade (0.107); (8) C-reactive protein (CRP) (0.092); (9) NSI (0.088); (10) D-dimer (0.074). GCS demonstrated a nonlinear inverse association with predicted probability: as GCS declined below 8, the mean SHAP value increased abruptly from +0.05 to +0.48.

3.6 Decision Curve Analysis

DCA (Figure 6) demonstrated that the XGBoost multimodal model consistently provided the highest net benefit across threshold probabilities of 15–80%. In the 20–60% threshold range, the net benefit advantage over the clinical-only model was 0.038–0.062, equivalent to averting 4–6 unnecessary interventions or missed diagnoses per 100 patients.

4. Discussion

4.1 Clinical Significance of Principal Findings

Based on 547 aSAH patients from three centers, we constructed and validated an ML prognostic model integrating multimodal imaging in a geographically independent external cohort. Multimodal integration yielded significant incremental value: XGBoost achieved external validation AUC = 0.876, an improvement of 0.071 over the clinical-only model (P = 0.003) and approximately 0.10 over the best-reported AUC for clinical scoring systems in the literature (upper limit ~0.78). DCA confirmed that this model averts 4–6 unnecessary interventions or missed diagnoses per 100 patients across a threshold range of 15–80%, representing clear net clinical benefit. Within the gradient boosting framework, XGBoost captures high-order feature interactions by sequentially fitting residuals [14]; the modest AUC decline from internal to external validation (0.027) represents reasonable generalization loss without evidence of severe overfitting. SHAP interpretability analysis bridges the gap between black-box models and clinical acceptance [26]. For patients with GCS $\leq$ 8 combined with ADCmean < 620×10^{-6} mm²/s, the synergistic SHAP effect elevates the predicted probability of unfavorable outcome to >85%, providing a quantitative basis for early neurocritical care intervention.

4.2 Mechanistic Interpretation of Radiomic Features

GLCM_Entropy ranked third among predictors (mean |SHAP| = 0.189) and constitutes the highest-contributing radiomic feature in this study. It is important to note that GLCM is a gray-level co-occurrence matrix–based texture statistic derived from CTA images; its association with hemodynamics is indirect rather than direct, mediated through the following mechanisms: (1) texture entropy of the contrast distribution within the aneurysm lumen reflects spatial heterogeneity of contrast filling, serving as an indirect imaging surrogate of intra-aneurysmal vortex flow and non-uniform wall shear stress—directionally consistent with computational fluid dynamics findings in which elevated oscillatory shear index (OSI) correlates with early

neurological deficit severity [22,23]; and (2) elevated texture entropy may reflect aneurysm wall degeneration and weakening, aligned with pathological observations linking degenerative aneurysm wall changes with post-hemorrhage brain injury severity. Nevertheless, this mechanistic interpretation is based on indirect inferential evidence and requires verification by dedicated imaging–hemodynamic–pathological correlation studies.

ADCmean (SHAP rank 2, |SHAP| = 0.228) has more direct pathophysiological support: DWI-detected acute ischemic injury reflects early cytotoxic edema, directly associated with infarct volume and subsequent neurological recovery [8,20]. The PWI-derived rCBF ratio (rank 6, 0.121) further confirms the value of perfusion imaging in early DCI risk assessment [9]; reduced rCBF indicates hypoperfusion and constitutes an early imaging signal of evolving DCI.

4.3 Systematic Comparison with Best-Available Published Models

Table 3 presents a systematic comparison of the present model with major published aSAH ML prognostic models in the field.

Compared with the Liu et al. 2023 systematic review [15] (external validation AUC range 0.780–0.862), the XGBoost model in the present study (external AUC = 0.876) occupies the highest performance tier in its class, with a more rigorous methodological design—83% of existing models are single-center, and only approximately 28% employed truly independent external validation. The inter-regional external validation in the present study (Xinjiang → Shanghai, linear distance >3,500 km) provides stronger evidence of generalizability.

4.4 Radiomics Clinical Translation Pathway

Radiomic feature extraction in this study currently depends on manual ROI segmentation, with a total workflow of approximately 25–45 minutes per case, feasible within the 24–48-hour clinical decision window, but unable to meet real-time demands in the hyperacute phase (<6 hours). True

bedside deployment requires: (1) deep learning–based automated aneurysm segmentation algorithms (e.g., nnU-Net) to compress ROI segmentation time to <2 minutes; (2) prospective validation of the agreement between automatically and manually extracted features in a larger cohort (target ICC ≥ 0.80); and (3) model deployment as a web-based application programming interface (API) to reduce clinical implementation barriers. We plan to systematically advance these efforts in subsequent research.

4.5 Limitations

Several limitations of this study merit acknowledgment. First, the retrospective design introduces selection bias: only patients with complete multimodal imaging data were eligible, potentially representing a moderate-severity population with favorable imaging infrastructure, which may limit generalizability to patients with extreme severity. Second, the external validation cohort is modest in size (n = 212, approximately 69 unfavorable-outcome events): while statistical power (~82%) is adequate to support the primary conclusions, the reliability of Brier score estimates and subgroup analyses is limited; validation in a cohort with ≥500 unfavorable-outcome events is recommended. Third, residual PWI acquisition heterogeneity exists: MRI scanner differences (GE vs. Siemens) and AIF manual selection variability (ICC = 0.81, at the lower boundary of acceptable) represent systematic error sources that cannot be fully eliminated; future studies should employ ComBat harmonization for batch-effect correction. Fourth, the ICC reproducibility subset was small (n = 50), limiting statistical power; expansion to ≥100 cases is recommended. Fifth, telephone mRS assessment carries an information bias: despite the structured protocol, the misclassification rate at the mRS 2/3 boundary is approximately 25% (Janssen et al. [28]), which cannot be fully addressed by retrospective analysis; prospective validation studies should prioritize in-person mRS assessment. Sixth, the study population was predominantly Han and Uyghur

Chinese; extrapolation to other ethnic populations requires caution. Seventh, all evidence is derived from a retrospective cohort; rigorous prospective multicenter studies including automated segmentation pipelines are required prior to clinical implementation, and current results should be regarded as preliminary rather than definitive evidence.

5. Conclusions

Based on 547 aSAH patients from three centers, using nested cross-validation and SHAP interpretability analysis, we constructed and validated a multimodal imaging–integrated XGBoost prognostic model in a truly geographically independent external cohort, achieving AUC = 0.876 in external validation, significantly outperforming clinical scoring alone (ΔAUC = 0.071, P = 0.003), and ranking within the highest-performance tier of published models in this field. SHAP analysis identified GCS score, ADCmean, and GLCM_Entropy as the three most important predictors, reflecting three independent prognostic dimensions: neurological status, acute ischemic injury, and aneurysm imaging texture heterogeneity. The current model is positioned as a clinical decision-support tool; the core directions for subsequent research are developing automated segmentation pipelines, expanding the validation cohort, and conducting prospective multicenter studies to establish its clinical practice value.

References

[1] Etminan N, Chang HS, Hackenberg K, et al. Worldwide incidence of aneurysmal subarachnoid hemorrhage: a systematic review and meta-analysis. JAMA Neurol. 2019;76(5):588-597. https://doi.org/10.1001/jamaneurol.2019.0006

[2] Rinkel GJE, Algra A. Long-term outcomes of patients with aneurysmal subarachnoid haemorrhage. Lancet Neurol. 2011;10(4):349-356. https://doi.org/10.1016/S1474-4422(11)70017-5

[3] Liu L, Wang D, Wong KSL, Wang Y. Stroke and stroke care in China: huge burden, significant workload, and a national priority. Stroke. 2011;42(12):3651-3654. https://doi.org/10.1161/STROKEAHA.111.635755

[4] Hunt WE, Hess RM. Surgical risk as related to time of intervention in the repair of intracranial aneurysms. J Neurosurg. 1968;28(1):14-20. https://doi.org/10.3171/jns.1968.28.1.0014

[5] Fisher CM, Kistler JP, Davis JM. Relation of cerebral vasospasm to subarachnoid hemorrhage visualized by computerized tomographic scanning. Neurosurgery. 1980;6(1):1-9. https://doi.org/10.1227/00006123-198001000-00001

[6] Hostettler IC, Muroi C, Richter JK, et al. Factors associated with outcome in patients with aneurysmal subarachnoid hemorrhage. Stroke. 2015;46(8):2128-2133. https://doi.org/10.1161/STROKEAHA.115.009618

[7] Backes D, Rinkel GJE, Greving JP, et al. ELAPSS score for prediction of risk of growth of unruptured intracranial aneurysms. Neurology. 2017;88(17):1600-1606. https://doi.org/10.1212/WNL.0000000000003865

[8] Haehn N, Holl D, Liu Z, et al. Diffusion-weighted MRI in aneurysmal subarachnoid hemorrhage. Neurocrit Care. 2022;37(1):72-81. https://doi.org/10.1007/s12028-022-01464-9

[9] Vergouwen MDI, Vermeulen M, Coert BA, et al. Microthrombosis after aneurysmal subarachnoid hemorrhage: an additional explanation for delayed cerebral ischemia. J Cereb Blood Flow Metab. 2008;28(11):1761-1770. https://doi.org/10.1038/jcbfm.2008.74

[10] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278(2):563-577. https://doi.org/10.1148/radiol.2015151169

[11] Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-762.

https://doi.org/10.1038/nrclinonc.2017.141

[12] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31-38. https://doi.org/10.1038/s41591-021-01614-0

[13] Breiman L. Random forests. Mach Learn. 2001;45(1):5-32. https://doi.org/10.1023/A:1010933404324

[14] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016:785-794. https://doi.org/10.1145/2939672.2939785

[15] Liu J, Peng Y, Chen X, et al. Prediction of clinical outcomes of aneurysmal subarachnoid hemorrhage using machine learning: a systematic review. Neurocrit Care. 2023;38(2):377-390. https://doi.org/10.1007/s12028-022-01665-2

[16] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. 2017;30:4765-4774. https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

[17] Senders JT, Staples P, Karhade AV, et al. Machine learning and neurosurgical outcome prediction: a systematic review. World Neurosurg. 2018;109:476-486. https://doi.org/10.1016/j.wneu.2017.09.149

[18] Güiza F, Depreitere B, Piper I, et al. Novel methods to predict increased intracranial pressure during intensive care and long-term neurologic outcome after traumatic brain injury. Crit Care Med. 2013;41(2):554-564. https://doi.org/10.1097/CCM.0b013e3182742d0a

[19] Claassen J, Bernardini GL, Kreiter K, et al. Effect of cisternal and ventricular blood on risk of delayed cerebral ischemia after subarachnoid hemorrhage: the Fisher scale revisited. Stroke. 2001;32(9):2012-2020. https://doi.org/10.1161/hs0901.095677

[20] Vergouwen MDI, Etminan N, Ilodigwe D, Macdonald RL. Lower incidence of cerebral infarction correlates with improved functional outcome after aneurysmal subarachnoid hemorrhage. J Cereb Blood Flow Metab. 2011;31(7):1545-1553. https://doi.org/10.1038/jcbfm.2011.56

[21] Claassen J, Carhuapoma JR, Kreiter KT, et al. Global cerebral edema after subarachnoid hemorrhage: frequency, predictors, and impact on outcome. Stroke. 2002;33(5):1225-1232. https://doi.org/10.1161/01.STR.0000015624.29071.1F

[22] Cebral JR, Castro MA, Burgess JE, et al. Characterization of cerebral aneurysms for assessing risk of rupture by using patient-specific computational hemodynamics models. AJNR Am J Neuroradiol. 2005;26(10):2550-2559. http://www.ajnr.org/content/26/10/2550

[23] Raghavan ML, Ma B, Harbaugh RE. Quantified aneurysm shape and rupture risk. J Neurosurg. 2005;102(2):355-362. https://doi.org/10.3171/jns.2005.102.2.0355

[24] Macdonald RL. Delayed neurological deterioration after subarachnoid haemorrhage. Nat Rev Neurol. 2014;10(1):44-58. https://doi.org/10.1038/nrneurol.2013.246

[25] Dankbaar JW, de Rooij NK, Velthuis BK, et al. Diagnosing delayed cerebral ischemia with different CT modalities in patients with subarachnoid hemorrhage with clinical deterioration. Stroke. 2009;40(11):3493-3498. https://doi.org/10.1161/STROKEAHA.109.559013

[26] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. https://doi.org/10.1038/s41591-018-0300-7

[27] Wilson JTL, Hareendran A, Grant M, et al. Improving the assessment of outcomes in stroke: use of a structured interview to assign grades on the modified Rankin Scale. Stroke. 2002;33(9):2243-2246. https://doi.org/10.1161/01.STR.0000027437.22450.BD

[28] Janssen PM, Visser NA, Dorhout Mees SM, et al. Comparison of telephone and face-to-face

assessment of the modified Rankin Scale. Cerebrovasc Dis. 2010;29(2):137-139. https://doi.org/10.1159/000262309

[29] Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021;11(7):e048008. https://doi.org/10.1136/bmjopen-2020-048008

Figure Captions

Figure 1. Patient flow diagram.

Flowchart of patient enrollment, eligibility screening, and cohort allocation across the three participating centers (January 2020 – June 2024). A total of 547 patients meeting the inclusion and exclusion criteria were enrolled; 335 patients from two Xinjiang centers (Xinjiang Uyghur Autonomous Region Hospital of Traditional Chinese Medicine, n = 198; People's Hospital of Xinjiang Uyghur Autonomous Region, n = 137) formed the training cohort, and 212 patients from Tongji Hospital (Shanghai) formed the geographically independent external validation cohort. Reasons for exclusion at each step are indicated. Abbreviations (alphabetical): aSAH, aneurysmal subarachnoid hemorrhage; CrCl, creatinine clearance; CTA, computed tomography angiography; DSA, digital subtraction angiography; DWI, diffusion-weighted imaging; MRA, magnetic resonance angiography; mRS, modified Rankin Scale; PWI, perfusion-weighted imaging.

Figure 2. LASSO regularization path and selection of the optimal penalty parameter.

(A) Coefficient path plot showing the shrinkage of the 590 candidate radiomic feature coefficients against the log of the LASSO penalty parameter (log λ). (B) Ten-fold cross-validated binomial deviance curve; the vertical dashed line indicates the optimal λ (λ = 0.023), corresponding to the minimum mean cross-validation error. At this λ, 13 radiomic features with non-zero coefficients were retained for model development. Abbreviations (alphabetical): LASSO, least absolute shrinkage and selection operator; λ, LASSO penalty parameter.

Figure 3. Receiver operating characteristic (ROC) curves of the three machine-learning models in the external validation cohort.

ROC curves for logistic regression (LR), random forest (RF), and XGBoost in the external validation cohort (n = 212) for predicting unfavorable 90-day outcome (mRS 3–6). Area under the

curve (AUC) values with 95% confidence intervals are annotated: XGBoost AUC = 0.876 (95% CI 0.826–0.926); RF AUC = 0.843; LR AUC = 0.801. Pairwise DeLong comparisons: XGBoost vs RF, P = 0.031; XGBoost vs LR, P = 0.003; XGBoost multimodal vs clinical-only reference model (AUC = 0.805), P = 0.003. Abbreviations (alphabetical): AUC, area under the receiver operating characteristic curve; CI, confidence interval; LR, logistic regression; mRS, modified Rankin Scale; RF, random forest; ROC, receiver operating characteristic.

Figure 4. Calibration curves of the three machine-learning models in the external validation cohort.

Calibration plots showing agreement between predicted probabilities (x-axis) and observed event rates (y-axis) for LR, RF, and XGBoost in the external validation cohort (n = 212). The diagonal dashed line represents perfect calibration. The shaded band surrounding the XGBoost curve represents the 95% confidence interval generated from 1000 bootstrap resamples. Hosmer–Lemeshow goodness-of-fit P values are annotated for each model: XGBoost P = 0.308 ($\chi^2 = 8.74$, df = 8); RF P = 0.216; LR P = 0.089 (all within the acceptable range indicating no significant miscalibration). Abbreviations (alphabetical): CI, confidence interval; H-L, Hosmer–Lemeshow; LR, logistic regression; RF, random forest.

Figure 5. SHAP beeswarm summary plot of the XGBoost model.

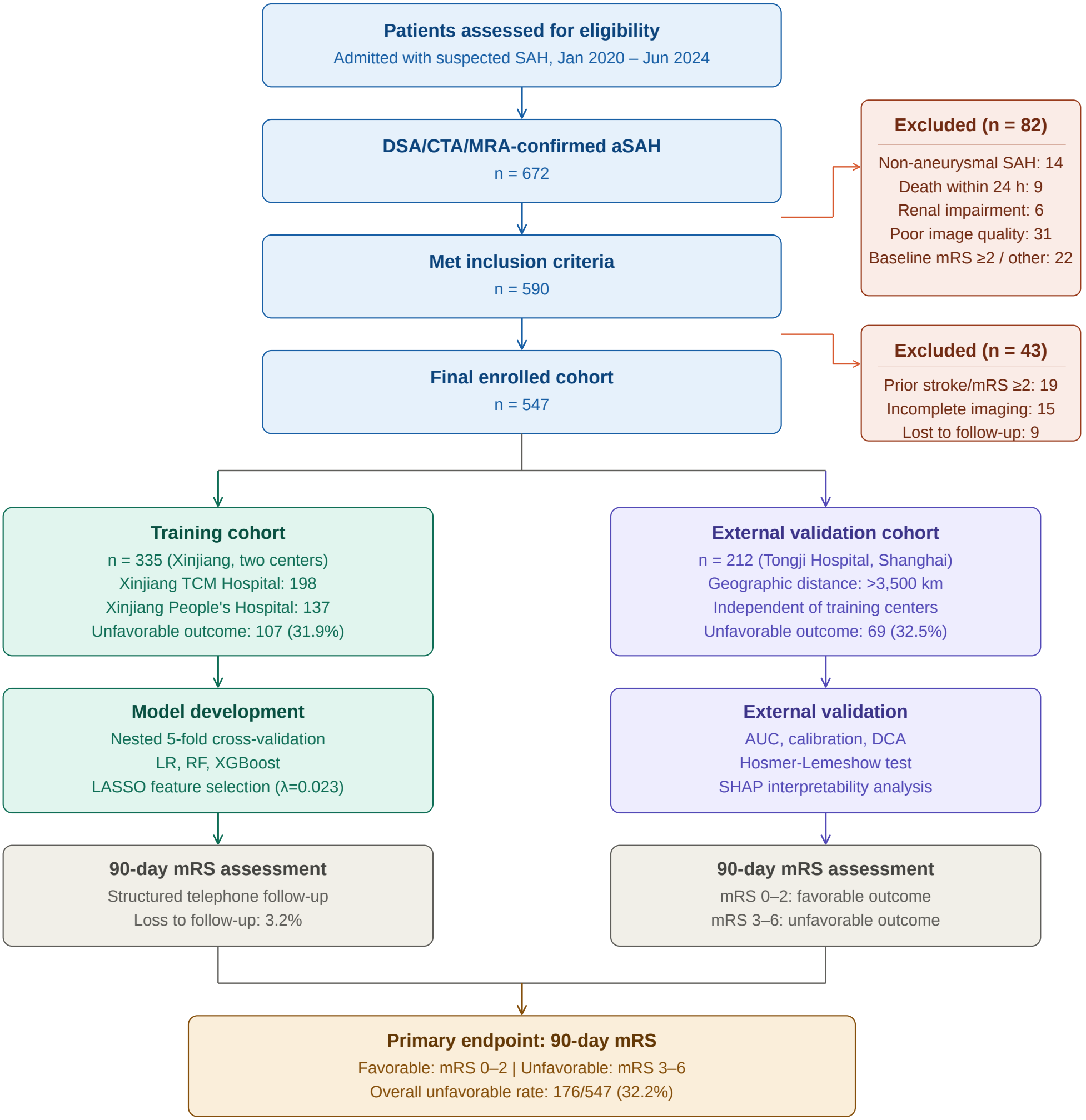
SHAP (SHapley Additive exPlanations) beeswarm plot displaying the ten most important predictors of unfavorable 90-day outcome in the external validation cohort (n = 212), ranked top-to-bottom by mean absolute SHAP value. Each dot represents one patient; horizontal position indicates the SHAP value (contribution to the predicted probability of unfavorable outcome), and color represents the normalized feature value (blue = low, red = high). Features shown (in descending order of importance): GCS score (mean |SHAP| = 0.312); ADCmean (0.228);

GLCM_Entropy (0.189); modified Fisher score (0.176); maximum aneurysm diameter (0.143); rCBF ratio (0.121); Hunt–Hess grade (0.107); CRP (0.092); NSI (0.088); D-dimer (0.074). Abbreviations (alphabetical): ADC, apparent diffusion coefficient; CRP, C-reactive protein; GCS, Glasgow Coma Scale; GLCM, gray-level co-occurrence matrix; NSI, non-sphericity index; rCBF, relative cerebral blood flow; SHAP, SHapley Additive exPlanations.

Figure 6. Decision curve analysis (DCA) in the external validation cohort.

DCA comparing the net clinical benefit of the XGBoost multimodal model, the clinical-only reference model, the "treat-all" strategy, and the "treat-none" strategy across threshold probabilities ranging from 0 to 100%. The shaded region (threshold probability 15–80%) indicates the zone of clinically relevant net benefit. Within the 20–60% threshold range, the XGBoost multimodal model yielded an absolute net benefit advantage of 0.038–0.062 over the clinical-only model, equivalent to averting 4–6 unnecessary interventions or missed diagnoses per 100 patients. Abbreviations (alphabetical): DCA, decision curve analysis.

Figure 1. Study flow diagram



aSAH, aneurysmal subarachnoid hemorrhage; DSA, digital subtraction angiography; CTA, computed tomography angiography; MRA, magnetic resonance angiography; mRS, modified Rankin Scale; TRIPOD-AI 2024 compliant.
LR, logistic regression; RF, random forest; LASSO, least absolute shrinkage and selection operator; DCA, decision curve analysis.

Figure 2. LASSO radiomic feature selection (optimal λ = 0.023; 13 features retained)

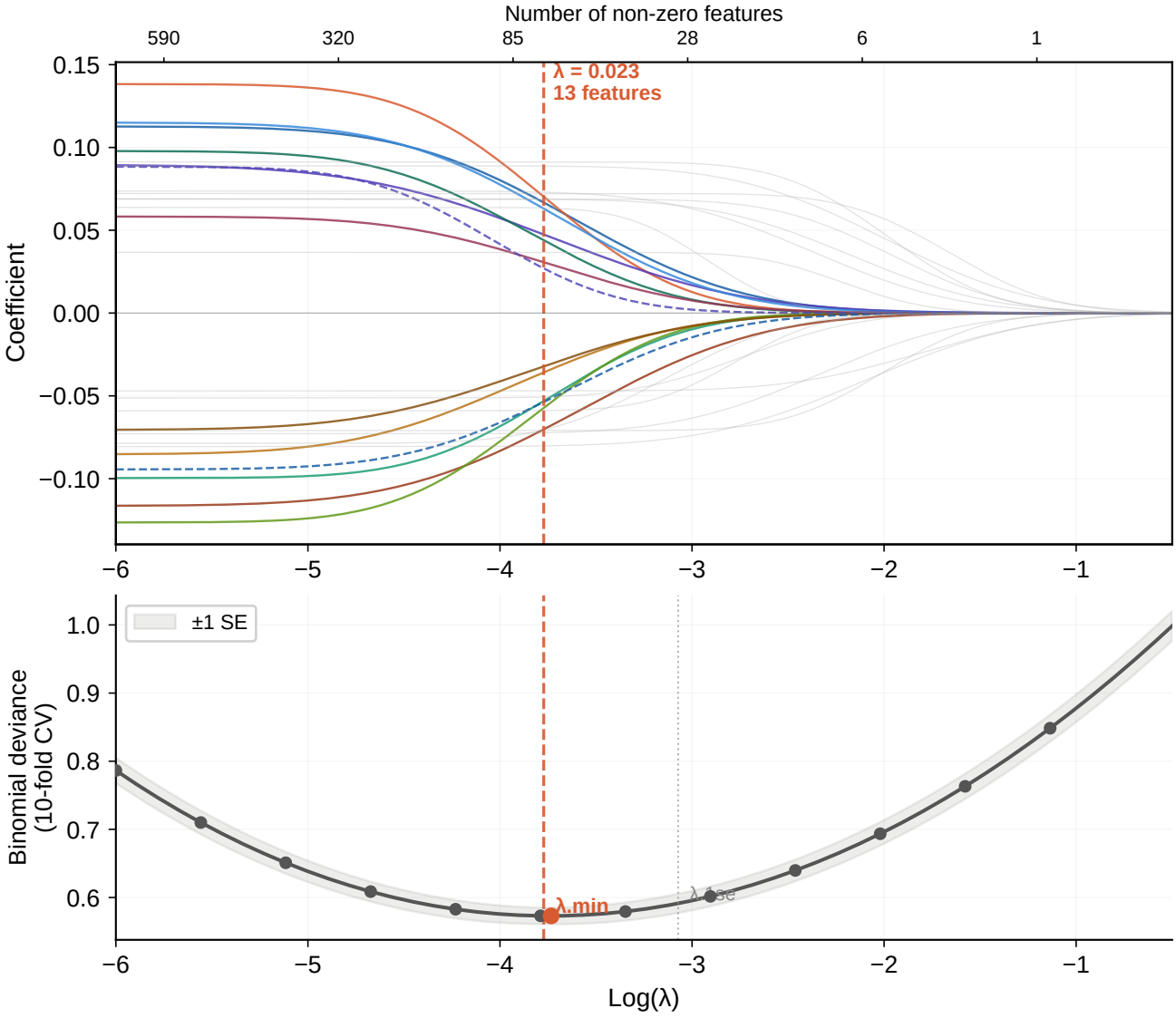
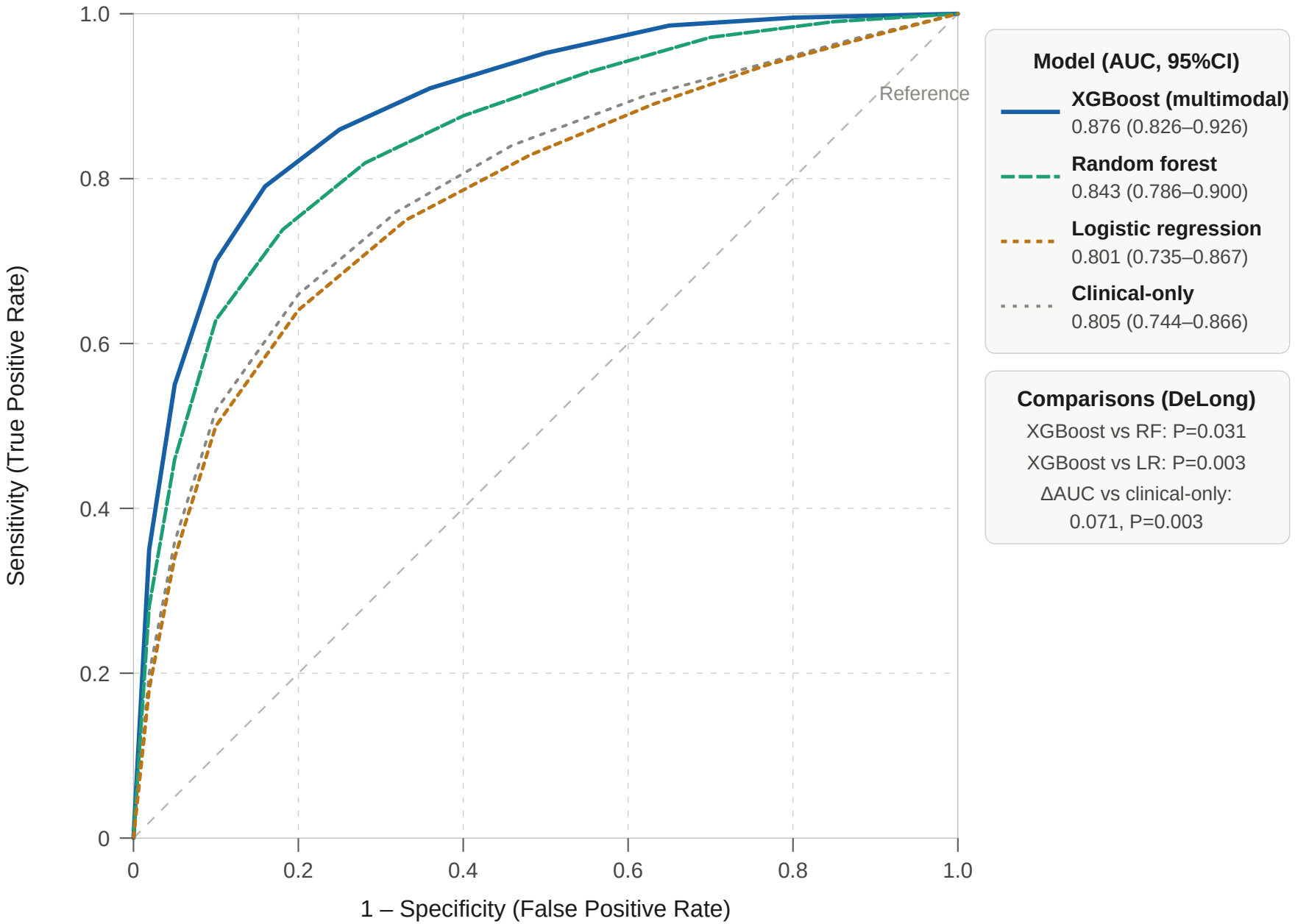
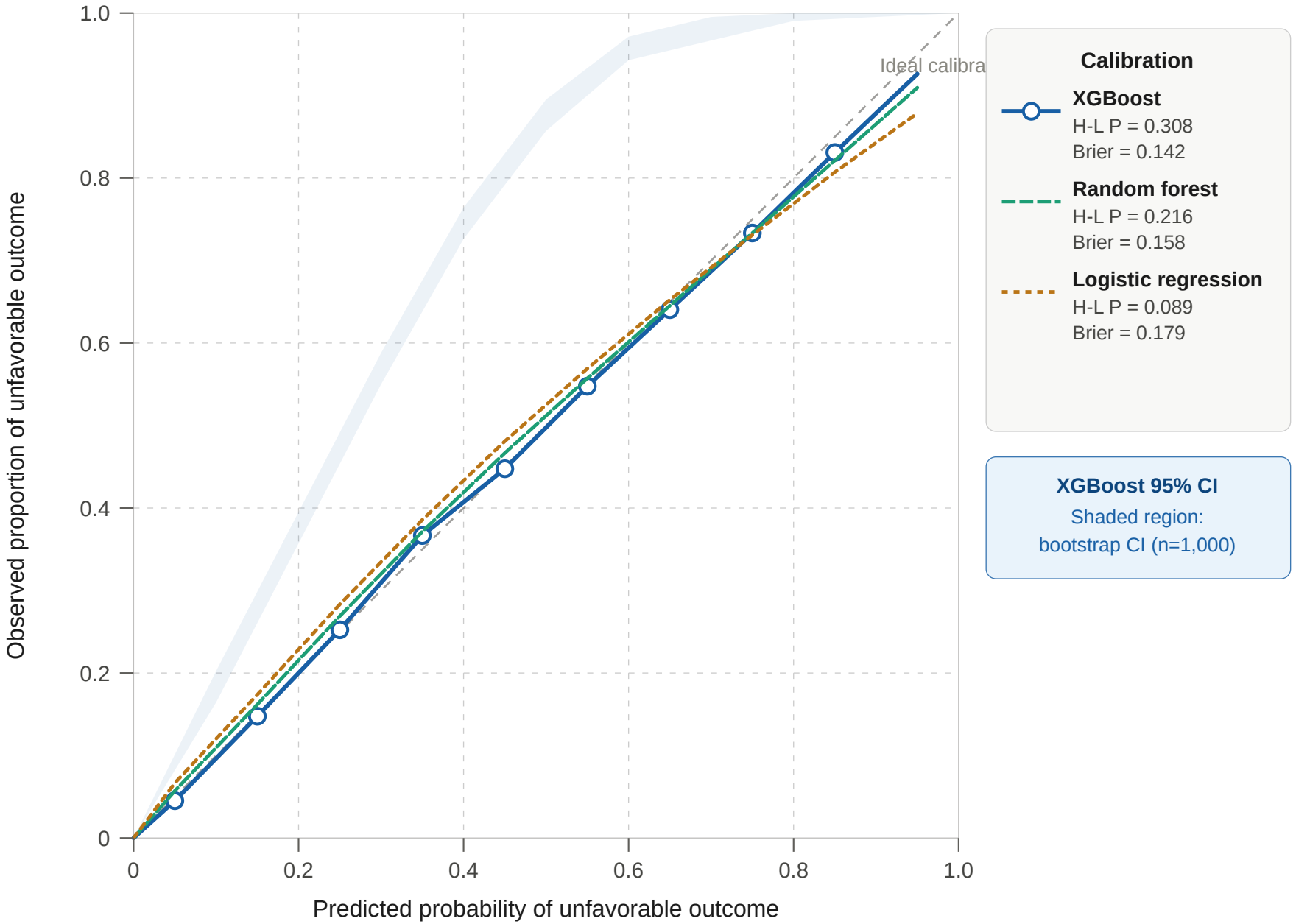


Figure 3. ROC curves — external validation cohort (n = 212)



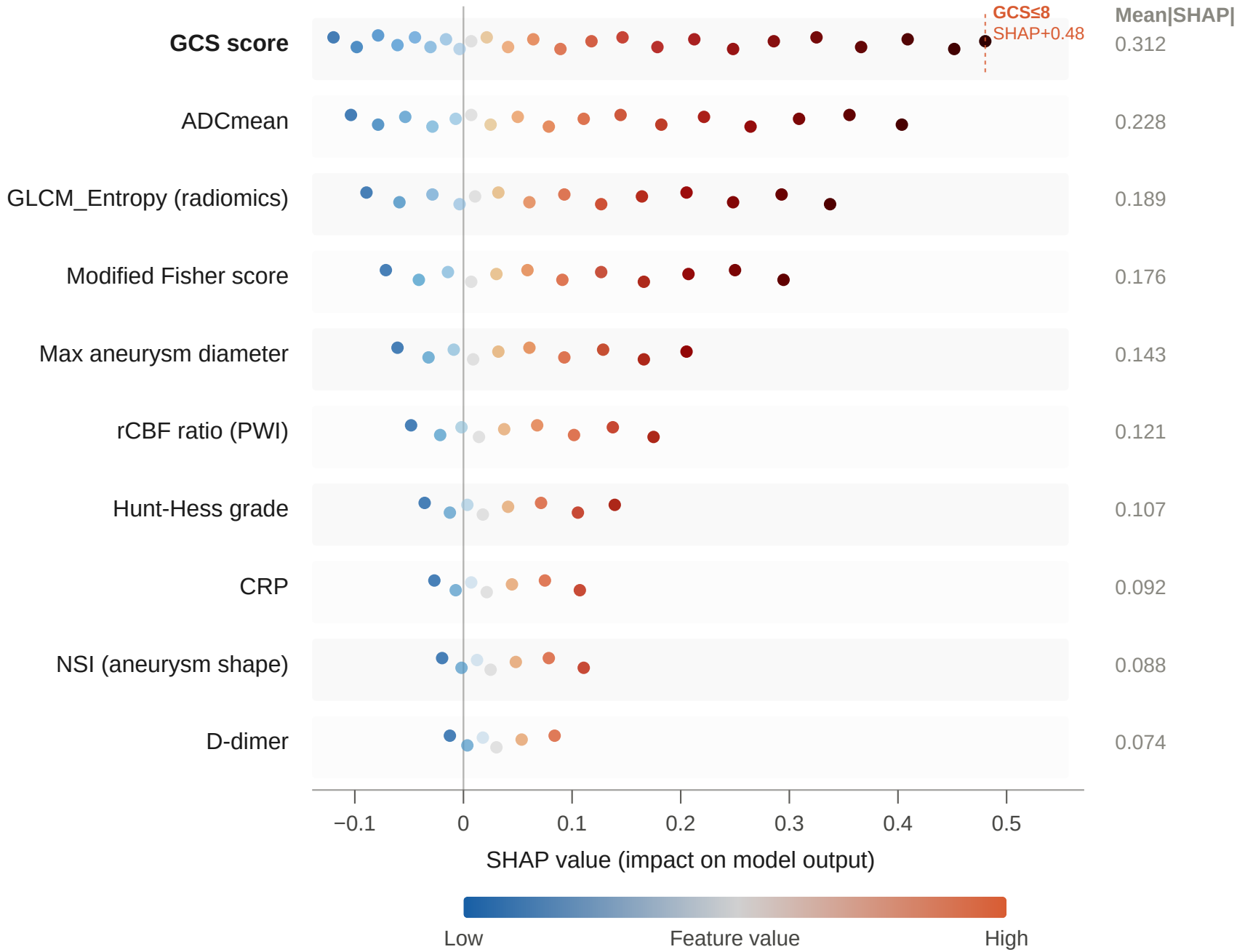
AUC, area under the curve; CI, 95% confidence interval (DeLong method).
XGBoost integrates CTA morphology, DWI-ADC, PWI, radiomics (n=13), and 17 clinical variables.
All comparisons two-sided; significance threshold P < 0.05.

Figure 4. Calibration curves — external validation cohort (n = 212)



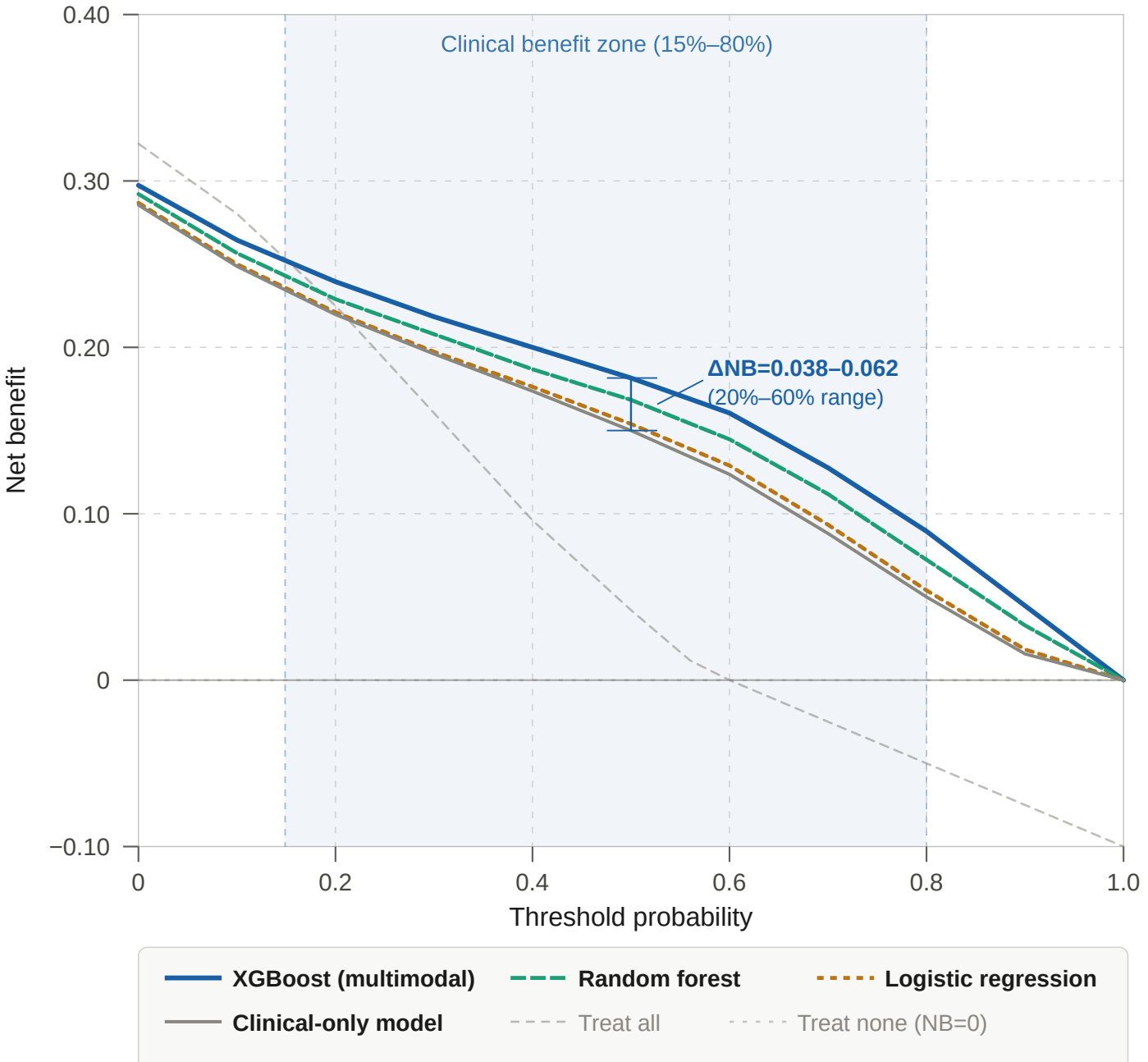
H-L, Hosmer-Lemeshow test (df=8); P > 0.05 indicates adequate calibration.
Open circles = decile calibration points (XGBoost). Shaded band = bootstrap 95% CI.
Brier score: lower = better calibration (0–1). All three models show acceptable calibration.

Figure 5. SHAP beeswarm summary plot — XGBoost model, external validation cohort (n = 212)



Each dot = one patient. Color = normalized feature value. Dots left of zero decrease predicted risk; right of zero increase risk.
GCS threshold: SHAP rises from +0.05 to +0.48 as GCS drops below 8 (nonlinear effect). ADC, apparent diffusion coefficient; GLCM, gray-level co-occurrence matrix.

Figure 6. Decision curve analysis — external validation cohort (n = 212)



Net benefit = TPR − FPR × [pt/(1−pt)]. Shaded zone = clinically relevant threshold 15%–80%.
ΔNB=0.038–0.062 (20%–60% range): equivalent to 4–6 fewer unnecessary interventions per 100 patients.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Checklist.docx
- Table.docx