ORIGINAL RESEARCH

High Resolution Ultrasound Radiomics Differences in Carotid Plaques between Patients with and without Ischemic Stroke

Shenglan Yang[a], Xiaoyu Luo[a], Wendi Zheng[a], Xiaoyu Yue[a], Wenhao Liu[a], Yanan Pan[a], Yichun Zhou[a], Xiaoyu Wan[a], Mingjun Xu[a], Mei Zhang[a,*]

[a] State Key Laboratory for Innovation and Transformation of Luobing Theory, Key Laboratory of Cardiovascular Remodeling and Function Research, Ministry of Education, National Health Commission and Chinese Academy of Medical Sciences, Department of Cardiology, Qilu Hospital of Shandong University, Jinan, China

Received August 30, 2026; revision requested September 16, 2026; accepted October 11, 2026

Background: High-resolution ultrasound (HRUS) provides improved spatial resolution for superficial vascular structures, while radiomics quantifies image intensity and texture patterns. However, the robustness of HRUS radiomics differences between carotid plaques from patients with and without ischemic stroke (IS) remains uncertain. This study aimed to identify radiomics features robustly associated with IS.

Methods: This retrospective study included a primary cohort of 1,148 patients (361 with IS and 787 without IS) and an additional population of 83 patients with IS and moderate or severe carotid plaque. We extracted 474 radiomics features in the primary cohort. Features meeting the reliability criterion were analyzed primarily using propensity score overlap weighting for eight clinical covariates. Associations were also assessed using covariate adjusted logistic regression, with false discovery rate correction applied to the overlap-weighted and covariate-adjusted analyses. LASSO selection stability was evaluated in bootstrap samples after penalty selection by the one standard error rule. A separate 1,000 iteration bootstrap of the full modeling pipeline provided internal validation. An expanded clinical model, interaction analyses, and a comparison of the two IS case cohorts were also performed.

Results: A total of 412 of 474 features (86.9%) met the ICC threshold of 0.80. After overlap weighting, all eight clinical covariates had absolute standardized mean differences below 0.01 (effective sample size = 817.2). Of the 403 estimable features, 73 were associated with IS after false discovery rate correction. Seven features in the primary cohort met all robustness criteria. Wavelet-LH_glrlm_RunEntropy showed the strongest inverse association in the weighted analysis (OR per SD, 0.60; 95% CI, 0.49 to 0.74; $q < 0.001$). In the expanded clinical model, all seven features retained the same association direction, and one remained significant after false discovery rate correction.

Conclusion: HRUS radiomics profiles differed between patients with and without IS in the overlap-weighted analysis. Seven features met the prespecified robustness criteria. These findings suggest that plaque echogenicity and spatial texture are associated with IS. Prospective multicenter validation is needed.

Key words: Carotid plaque; High resolution ultrasound; Radiomics; Ischemic stroke; Overlap weighting

Advanced Ultrasound in Diagnosis and Therapy 2026; 04: 233-244 DOI: 10.26599/AUDT.2026.260034

[*] *Corresponding author: Department of Cardiology, Qilu Hospital of Shandong University, 107 Wenhua West Road, Jinan, Shandong 250012, China. (Mei Zhang)*
e-mail: daixh@vip.sina.com (M Z)

2576-2508/© AUDT 2026 • http://www.AUDT.org
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Stroke remains a major cause of death and disability worldwide, and ischemic stroke (IS) accounts for most incident strokes [1]. Atherosclerotic plaques at the carotid bifurcation are an established source of IS [2]. In current clinical practice, the assessment of plaque characteristics relies heavily on luminal stenosis and neurological history, although plaque composition, surface disruption, intraplaque hemorrhage, neovascularization, inflammation, and hemodynamic characteristics also affect plaque vulnerability [2,3].

Ultrasound is the principal imaging method used to evaluate carotid plaque. Conventional ultrasound can show plaque thickness, surface morphology, echogenicity, and calcification. Radiomics complements these observations by quantifying the distribution and spatial organization of image intensities, which may reveal imaging patterns that are not readily summarized by visual assessment [4,5]. Recent studies have reported radiomics differences between carotid plaques associated with and without ischemic events [6-9].

High-resolution ultrasound (HRUS) uses a higher-frequency probe than conventional carotid ultrasonography, providing improved spatial resolution for delineating superficial arterial wall structures. This technique can distinguish the intimal and medial layers more clearly and permits quantitative assessment of subtle arterial wall changes that may not be adequately characterized by conventional measurements of combined intima–media thickness [10,11]. Using a 24-MHz high-resolution probe, a study demonstrated that carotid and radial intimal thickness could be measured feasibly and reliably. These measurements provided incremental value beyond traditional vascular risk factors in differentiating patients with large-artery atherosclerotic IS from control participants [11]. The improved depiction of arterial wall and plaque microstructure provides a suitable imaging basis for radiomics analysis, which can further quantify variations in image intensity, texture, and spatial organization that are not readily captured by visual assessment alone.

Recent systematic reviews found encouraging performance for carotid plaque radiomics, but also reported low overall methodological quality, substantial heterogeneity, and limited external validation [12,13]. Reproducibility is a particular concern because many radiomics features change when the plaque boundary or preprocessing parameters change [14,15]. Current guidance for data-driven imaging markers likewise stresses transparent preprocessing, robustness assessment, and validation [16].

However, the stability of a feature selected in one random split but not another is uncertain, because radiomics results can vary across data partitions [17]. Therefore, we compared HRUS radiomics features in carotid plaques from patients with IS and without IS. To account for measured baseline differences, we used propensity score overlap weighting as the primary analysis and assessed whether the associations between radiomics features and IS remained consistent in an expanded model incorporating additional clinical covariates. We assessed feature selection stability with LASSO and bootstrap resampling, and estimated optimism in the radiomics model using a separate bootstrap analysis of the radiomics modeling process. We also explored effect modification and compared the additional 83 patients with IS and moderate or severe carotid plaque with IS patients in the primary cohort in an exploratory analysis.

Materials and Methods

Study design and participants

This retrospective observational study included patients with carotid plaque examined in Qilu Hospital of Shandong University and participating communities, between September 2018 and December 2024. The institutional ethics committee approved the study and waived the requirement for informed consent. The study was conducted in accordance with the Declaration of Helsinki.

Eligible patients had at least one carotid plaque visible on HRUS and an image of sufficient quality for plaque segmentation. Exclusion criteria were previous carotid stenting, carotid occlusion, posterior circulation events, stroke attributed to causes unrelated to carotid plaque (including arterial dissection or cardioembolism), inability to complete carotid ultrasound, and image quality inadequate for radiomics analysis. Patients were divided into the IS and Without IS groups. The primary cohort included 361 patients with IS and 787 patients without IS. An additional 83 patients with IS and moderate or severe carotid plaque were included in an exploratory comparison with the 361 patients with IS in the primary cohort. The study workflow is shown in Figure 1.

Ultrasound acquisition and plaque selection

All examinations used an Aplio i900 ultrasound system (Canon Toshiba Ultrasound, Tochigi, Japan) and a PLI 2004BX 24 MHz linear array transducer. Patients were examined in the supine position with synchronous electrocardiographic recording. Both carotid arteries were surveyed from the common carotid artery through the bifurcation to the proximal internal carotid artery. The operator observed three consecutive cardiac cycles and identified end diastole at the peak of the R wave.

清華大學出版社 Tsinghua University Press | SciOpen

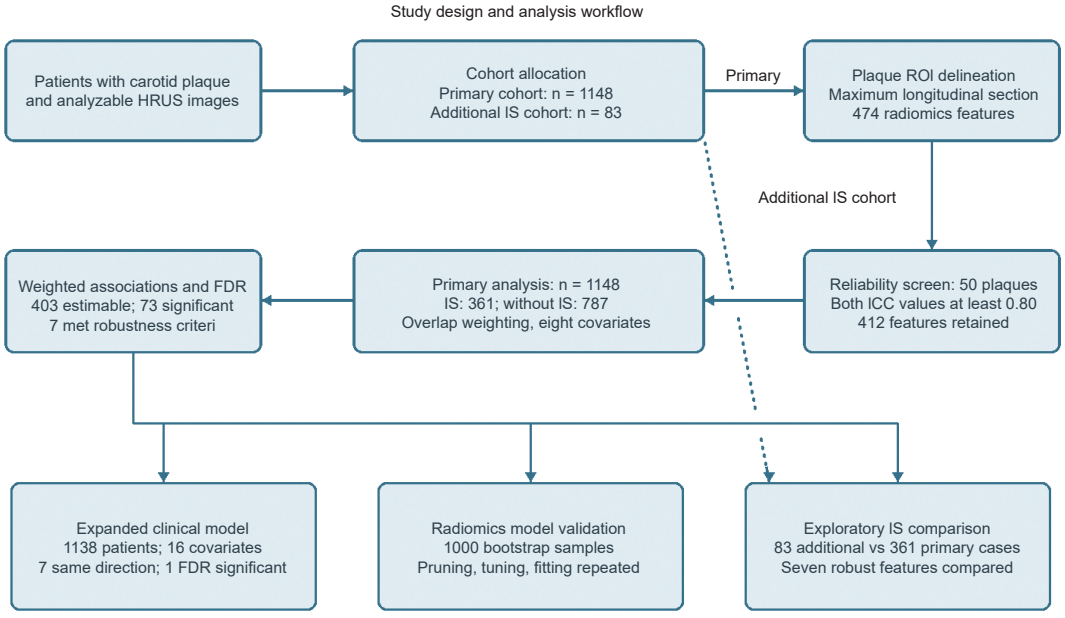


Figure 1 Study design and analysis workflow. FDR, false discovery rate; HRUS, high resolution ultrasound; IS, ischemic stroke; ICC, intraclass correlation coefficient; ROI, region of interest.

One longitudinal end diastolic frame that showed the plaque most completely and clearly and had the largest plaque area was selected. Images were exported in Digital Imaging and Communications in Medicine (DICOM) format.

Plaque segmentation

The DICOM image was loaded into 3D Slicer (version 5.6.2). The operator used the drawing tool to trace the plaque boundary manually and create a region of interest (ROI). The anterior echo defined the lumen to plaque interface, and the posterior echo defined the adventitia to plaque interface. Areas obscured by acoustic shadowing from calcification were excluded [18]. The completed ROI was converted to a mask. The original image and ROI mask were saved in nearly raw raster data (NRRD) format (Figure 2).

Intraobserver and interobserver reliability assessment

Fifty plaques were selected for reliability assessment. The primary reader and a second physician, each with approximately three years of experience in plaque imaging, independently delineated each plaque. One month later, the primary reader repeated the delineations without access to the initial results. For each radiomics feature, intraobserver reliability was assessed using the primary reader's two delineations, and interobserver reliability using the primary reader's initial delineation and that of the second physician. ICCs were calculated using a two-way random-effects model for absolute agreement based on a single measurement (ICC[2,1]) [19]. Features with both intraobserver and interobserver ICCs of at least 0.80 were included in the primary analysis.

Radiomics feature extraction

Features were extracted with PyRadiomics version 3.0.1 and SimpleITK version 2.1.1 using a frozen parameter file, following standardized radiomics definitions [4]. After image resampling, radiomic features were extracted in two dimensions from each delineated plaque region (force2D enabled). Shape features were calculated from the plaque masks, while first-order and texture features were calculated from the original and wavelet-filtered images. Texture features were derived from the gray-level co-occurrence matrix (GLCM), gray-

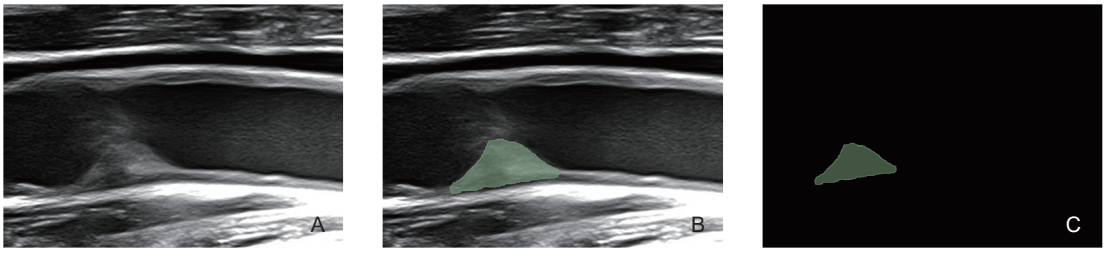


Figure 2 HRUS plaque image processing. (A) shows the selected longitudinal end diastolic image at the maximum plaque area. (B) shows the manually delineated ROI overlaid on the HRUS image. (C) shows the resulting plaque mask. HRUS, high resolution ultrasound; ROI, region of interest.

level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), gray-level dependence matrix (GLDM), and neighboring gray-tone difference matrix (NGTDM).

Statistical analysis

Continuous variables are summarized as mean (standard deviation) or median (first and third quartiles), and categorical variables as count (percentage). Continuous variables were compared using Welch's t test or the Mann–Whitney U test, and categorical variables using the chi-square test or Fisher's exact test. Baseline differences were described using standardized mean differences (SMDs). For the primary analysis, a logistic propensity score model estimated each patient's probability of belonging to the IS group within the study sample based on age, sex, hypertension, diabetes mellitus, coronary heart disease, current or former smoking, alcohol consumption, and lipid-lowering medication use. Overlap weights were defined as 1 minus the propensity score for patients with IS and as the propensity score for patients without IS. We assessed propensity-score overlap, weight distributions, effective sample size, and covariate balance before and after weighting. An absolute weighted SMD below 0.10 was considered acceptable balance [20].

Each radiomics feature meeting the reliability criteria was standardized to a mean of zero and a standard deviation of one. Its association with IS was assessed in a separate logistic regression model using overlap weights and robust standard errors. To assess the consistency of these associations under conventional adjustment, each feature was also examined in an unweighted multivariable logistic regression model adjusted for the same eight clinical covariates. Two-sided P values from the two analyses were corrected separately using the Benjamini and Hochberg false discovery rate procedure, with $q < 0.05$ considered significant. Only features with finite coefficient and standard error estimates in the weighted model were included in the subsequent robustness assessment.

To identify robust radiomics features and evaluate the internal validity of the radiomics model, highly correlated features were pruned at an absolute Spearman correlation of at least 0.90, followed by unweighted LASSO logistic regression. Ten logarithmically spaced values of λ were evaluated by stratified fivefold cross validation using mean AUC. The largest λ with a mean AUC within one standard error of the highest mean AUC was selected. Correlation pruning, penalty selection, and LASSO fitting were repeated in 1,000 patient level bootstrap samples. Selection frequency and coefficient direction were recorded [21]. Robust features required $q < 0.05$ in both the weighted and adjusted analyses, selection frequency of at least 70%, coefficient direction consistency of at least 80%, and concordant association directions in the two analyses. For internal validation, imputation was repeated within each bootstrap sample, and the fitted model was evaluated in both that sample and the original cohort using AUC, Brier score, calibration intercept, and calibration slope. Mean bootstrap optimism was subtracted from apparent performance for each measure.

An expanded propensity score model included BMI, systolic and diastolic blood pressure, fasting blood glucose, HDL-C, LDL-C, triglycerides, and creatinine in addition to the eight primary covariates. Complete cases were analyzed using the same weighting, feature association models, robust standard errors, and false discovery rate correction as in the primary analysis. Associations of the robust features were examined within clinical subgroups, and interactions with the grouping variable were tested using logistic regression. For each radiomics feature meeting the prespecified robustness criteria, an interaction term between the feature and lipid-lowering medication use was included in the model, and false discovery rate (FDR) correction was applied across the interaction tests. The additional 83 patients with IS were compared with the 361 patients with IS in the primary cohort using feature means and SMDs.

Results

Participants and covariate balance

The primary cohort included 1,148 patients (361 with IS and 787 without IS). Compared with patients without IS, those with IS were younger (median, 68 vs. 70 years), more often male (65.4% vs. 56.7%), and more frequently had hypertension (69.5% vs. 51.5%) or diabetes mellitus (38.8% vs. 28.6%). They also had lower TC, HDL-C, and LDL-C levels (Table 1).

Compared with the 361 IS cases in the primary cohort, the 83 patients in the additional IS cohort had lower BMI, TC, and fasting glucose and higher ALT and AST. Diabetes mellitus and coronary heart disease were less frequent, whereas smoking and alcohol consumption were more frequent (Table 1).

The primary propensity model included age, sex, hypertension, diabetes mellitus, coronary heart disease, current or former smoking, alcohol consumption, and lipid-lowering medication. Before weighting, the largest absolute SMD among these eight covariates was 0.498. After overlap weighting, all eight had absolute SMDs below 0.10 (maximum, 0.0046), with an effective sample size of 817.2 (Table 2; Figures 3-4).

清華大學出版社 Tsinghua University Press | SciOpen

Table 1 Clinical characteristics of patients without IS, patients with IS in the primary cohort, and patients in the additional IS cohort

Characteristic	Without IS (*n* = 787)	Primary IS (*n* = 361)	Additional IS (*n* = 83)	*P* value 1	*P* value 2
Age, years	70.0 (65.0-74.0)	68.0 (61.0-72.0)	68.0 (64.0-71.0)	< 0.001	0.893
Height, m	1.62 (0.09)	1.65 (0.08)	1.66 (0.08)	< 0.001	0.856
Weight, kg	66.8 (11.1)	69.0 (11.0)	66.6 (10.5)	0.002	0.065
BMI, kg/m2	25.4 (3.4)	25.2 (3.4)	24.2 (3.3)	0.381	0.023
SBP, mmHg	134.2 (19.0)	139.4 (20.2)	141.7 (19.5)	< 0.001	0.332
DBP, mmHg	73.9 (10.9)	77.9 (13.1)	80.2 (10.0)	< 0.001	0.079
ALT, U/L	16.0 (12.0-23.0)	18.0 (13.0-27.0)	20.0 (16.0-30.0)	0.010	0.024
AST, U/L	20.9 (17.1-25.0)	18.0 (15.0-23.0)	21.0 (18.0-25.0)	< 0.001	0.001
TC, mmol/L	4.50 (3.44-5.41)	3.63 (3.06-4.48)	3.44 (2.93-3.97)	< 0.001	0.021
HDL-C, mmol/L	1.20 (1.00-1.43)	1.08 (0.91-1.29)	1.16 (0.96-1.26)	< 0.001	0.598
LDL-C, mmol/L	2.50 (1.83-3.18)	1.99 (1.63-2.60)	1.79 (1.53-2.41)	< 0.001	0.060
TG, mmol/L	1.30 (0.96-1.79)	1.20 (0.91-1.65)	1.11 (0.78-1.43)	0.055	0.058
Cr, umol/L	70.5 (60.4-80.0)	70.0 (60.0-80.0)	68.0 (60.0–80.5)	0.305	0.928
FBG, mmol/L	5.80 (5.20-6.70)	5.71 (4.92-6.96)	5.16 (4.58-5.73)	0.332	< 0.001
Male sex	446 (56.7%)	236 (65.4%)	62 (74.7%)	0.005	0.103
Hypertension	405 (51.5%)	251 (69.5%)	49 (59.0%)	< 0.001	0.066
Diabetes mellitus	225 (28.6%)	140 (38.8%)	24 (28.9%)	< 0.001	0.043
Coronary heart disease	359 (45.6%)	121 (33.5%)	14 (16.9%)	< 0.001	0.003
Lipid-lowering medication	264 (33.5%)	208 (57.6%)	22 (26.8%)	< 0.001	< 0.001
Current or former smoking	278 (35.3%)	135 (37.4%)	44 (53.0%)	0.497	0.009
Alcohol consumption	246 (31.3%)	118 (32.7%)	51 (61.4%)	0.629	< 0.001

Values are presented as mean (SD), median (25th to 75th percentile), or n (%). P value 1 compares patients without IS with patients with IS in the primary cohort; P value 2 compares patients with IS in the primary cohort with those in the additional IS cohort. ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; Cr, creatinine; DBP, diastolic blood pressure; FBG, fasting blood glucose; HDL-C, high-density lipoprotein cholesterol; IS, ischemic stroke; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure; TC, total cholesterol; TG, triglyceride

Table 2 Overlap-weighting balance diagnostics for the eight clinical covariates in the primary cohort

Characteristic	Unweighted SMD	Weighted SMD	Weighted without IS	Weighted with IS
Age, years	−0.229	−0.001	67.71	67.71
Male sex	0.179	0.001	62.68	62.74
Hypertension	0.376	0.002	65.20	65.30
Diabetes mellitus	0.217	< 0.001	36.45	36.49
Coronary heart disease	−0.249	−0.005	42.63	42.40
Current or former smoking	0.043	< 0.001	37.87	37.86
Alcohol consumption	0.031	< 0.001	32.14	32.12
Lipid-lowering medication	0.498	0.004	45.66	45.88

Weighted values are presented as mean age in years or percentages for categorical variables. IS, ischemic stroke; SMD, standardized mean difference

Intraobserver and interobserver reliability

The median intraobserver ICC was 0.940, and 412 of 474 features (86.9%) had intraobserver ICC values of at least 0.80. The median interobserver ICC was 0.993, and all 474 features had interobserver ICC values of at least 0.80. 412 features met the requirement for both assessments and entered the primary association analysis.

Primary overlap-weighted radiomics associations

Of the 412 reliability-eligible features, 403 were estimable in the overlap-weighted feature-wise models. A total of 73 features remained associated with IS after FDR correction. In the conventional covariate-adjusted sensitivity analysis, 212 features had $q < 0.05$. 7 features met all prespecified robustness criteria (Table 3; Figure 5). Wavelet-LH_glrlm_RunEntropy showed the strongest inverse weighted association (OR per SD, 0.60; 95% CI, 0.49-0.74; $q < 0.001$). The largest positive association among the robust features was observed for wavelet-HL_glcm_MaximumProbability (OR, 1.50; 95% CI, 1.13-1.97; $q = 0.029$).

LASSO feature selection and internal validation

Fivefold cross validation using the one standard error rule selected a LASSO penalty parameter of $\lambda = 4.6416$ (Figure 6). The LASSO model fitted to the full dataset

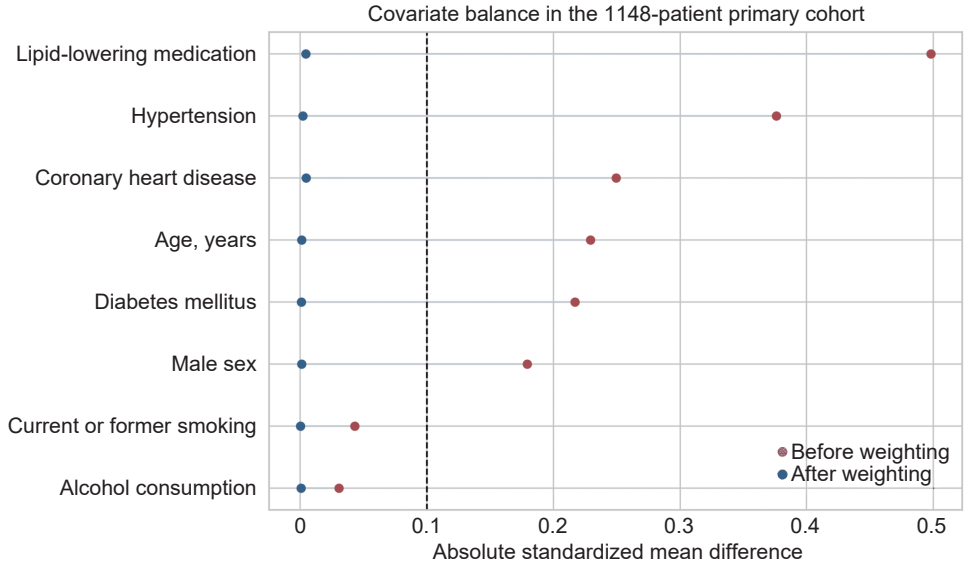


Figure 3 Absolute standardized mean differences before and after overlap weighting. The dashed reference line marks an absolute SMD of 0.10.

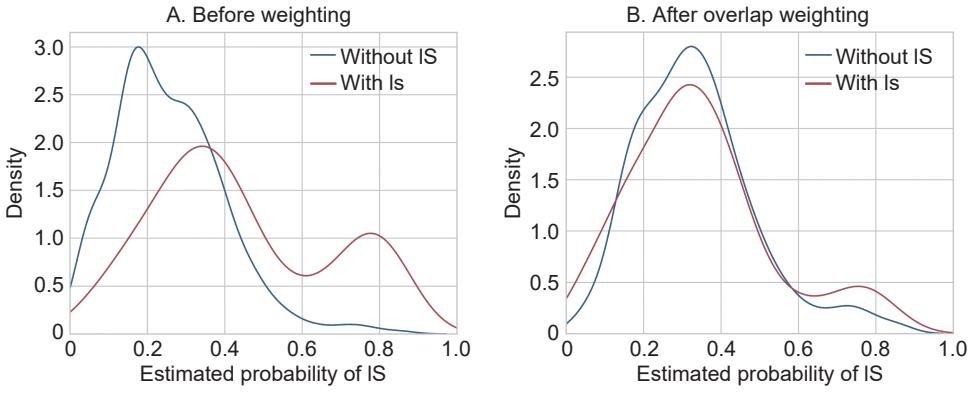


Figure 4 Propensity score distributions before and after overlap weighting in the primary cohort. Panel A shows unweighted distributions and panel B shows weighted distributions; blue and red curves represent patients without and with IS.

Table 3 Radiomics features meeting the prespecified reliability and robustness criteria

Feature	Weighted OR (95% CI)	FDR q	Selection, %	Direction, %	ICC
wavelet-LH_glcm_InverseVariance	1.35 (1.10-1.64)	0.026	93.5	96.8	0.938
original_firstorder_10Percentile	0.71 (0.58-0.88)	0.016	90.3	95.8	0.931
wavelet-HL_glcm_MaximumProbability	1.50 (1.13-1.97)	0.029	90.1	98.4	0.903
wavelet-LH_glrlm_RunEntropy	0.60 (0.49-0.74)	< 0.001	85.1	99.8	0.897
wavelet-LH_glszm_SizeZoneNonUniformityNormalized	1.37 (1.09-1.72)	0.039	77.7	100.0	0.972
wavelet-HH_glszm_GrayLevelNonUniformityNormalized	0.74 (0.60-0.90)	0.027	77.4	88.8	0.914
wavelet-HL_glszm_SizeZoneNonUniformityNormalized	1.38 (1.10-1.71)	0.029	77.2	94.3	0.877

ORs are per one standard deviation of each feature. Selection is the percentage of 1,000 bootstrap samples in which the feature was selected; direction is coefficient sign consistency among selected samples. ICC is the lower of the intraobserver and interobserver ICCs. q values are false discovery rate–adjusted P values. CI, confidence interval; FDR, false discovery rate; ICC, intraclass correlation coefficient; OR, odds ratio.

selected 37 features. 7 features met the prespecified robustness criteria; across 1,000 bootstrap samples, their selection frequencies ranged from 77.2% to 93.5%, and coefficient direction consistency ranged from 88.8% to 100% (Figures 7-8). The full data model had an apparent AUC of 0.776. After bootstrap validation of the complete modeling pipeline, the optimism corrected AUC was 0.724, with a Brier score of 0.182, calibration intercept of −0.017, and calibration slope of 0.905 (Table 4; Figure 9).

Exploratory analyses

The expanded clinical model included 1,138 patients (355 with IS and 783 without IS). After overlap weight-

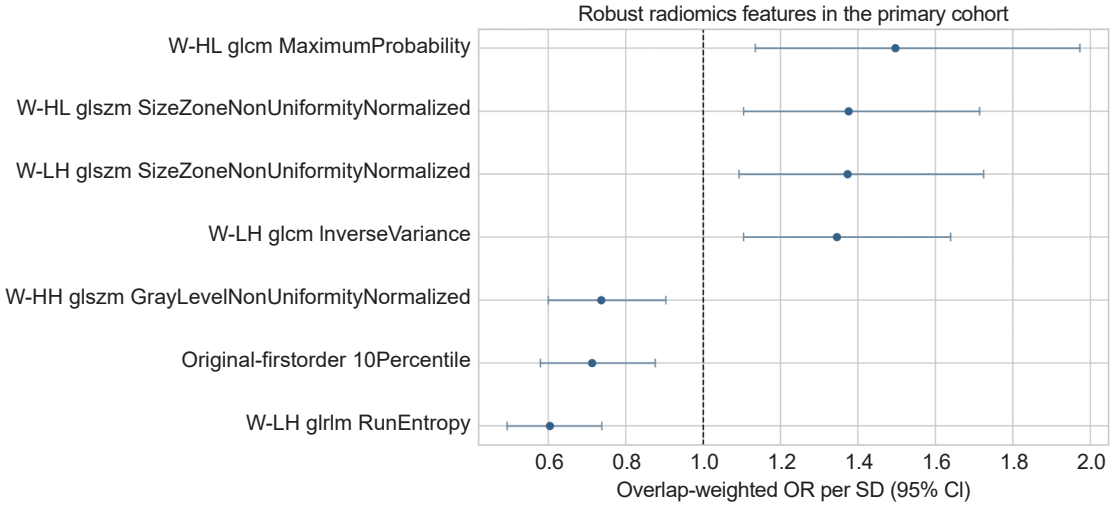


Figure 5 Overlap-weighted associations of seven robust radiomics features with ischemic stroke. Points show odds ratios per standard deviation, horizontal lines show 95% confidence intervals, and the dashed line marks an odds ratio of 1. W denotes wavelet features.

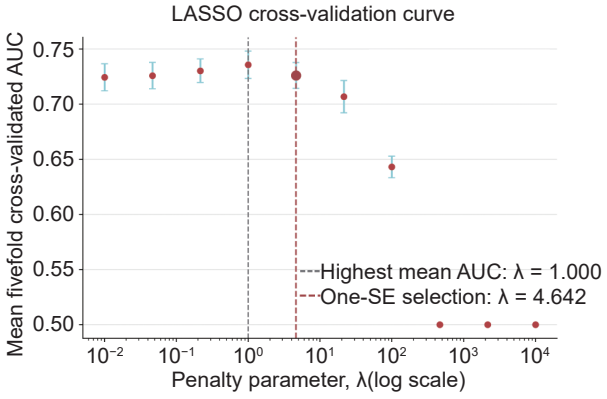


Figure 6 LASSO penalty selection by fivefold cross validation. Points show mean AUC and error bars show one standard error. Dashed lines mark the penalty with the highest mean AUC (gray, λ = 1.0000) and the penalty selected by the one standard error rule (red, λ = 4.6416).

ing, all 16 covariates had absolute SMDs below 0.10 (maximum, 0.0060; effective sample size, 735.0; Figure 10). Among 411 estimable features meeting the reliability criterion, 30 had q < 0.05. All 7 robust features from the primary analysis retained their association directions. One feature, wavelet-LH_glrlm_RunEntropy, remained significant after FDR correction (OR, 0.64; 95% CI, 0.52 to 0.79; q = 0.013; Table 5; Figure 11).

Interactions with lipid-lowering medication had q < 0.05 for 6 of the 7 robust features (Figure 12). Several robust features also had different mean values in the 83 additional IS cases and the 361 IS cases in the primary cohort; the largest absolute SMD was 0.882 (Figure 13).

Discussion

This study identified differences in HRUS-derived radiomics features between carotid plaques from patients with and without IS after propensity score overlap weighting for measured clinical covariates and correction for multiple comparisons. 7 features met the reliability, association, and stability criteria. They described gray level intensity, local probability, run length entropy and size zone organization, supporting differences in plaque echogenic composition and spatial texture.

The findings agree with recent carotid imaging studies. Huang et al. found radiomics differences between

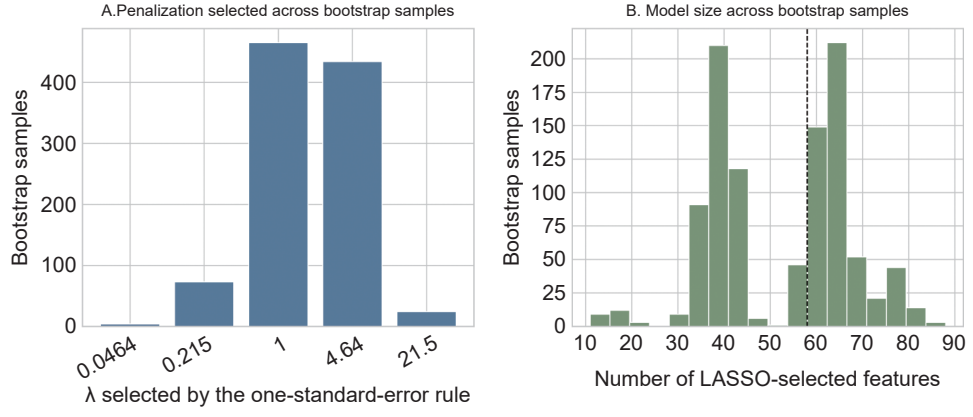


Figure 7 LASSO penalty selection and model size across 1,000 bootstrap samples. Panel A shows λ selected by fivefold cross-validation using the one-standard-error rule; panel B shows the number of selected features, with the median marked by a dashed line.

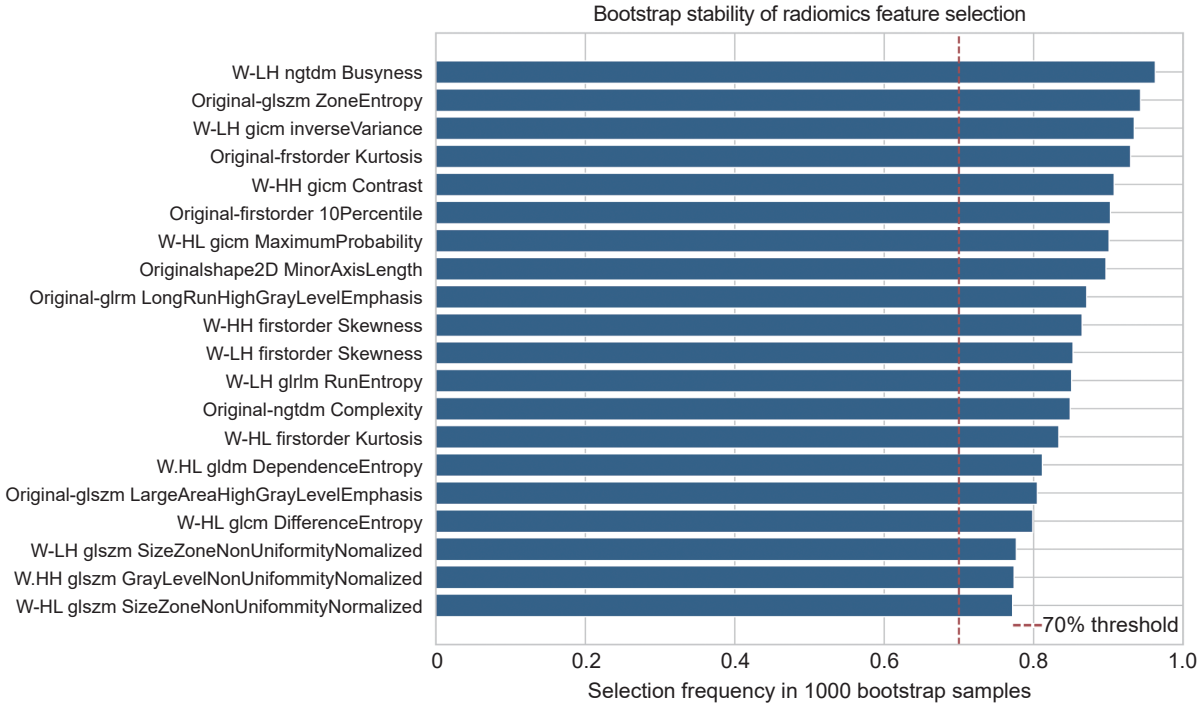


Figure 8 Selection frequencies of the 20 most frequently selected radiomics features across 1,000 bootstrap samples. Bars indicate the proportion of samples selecting each feature; the dashed line marks 70%.

Table 4 Full pipeline bootstrap internal validation with 1,000 iterations

Metric	Apparent	Optimism-corrected	Mean optimism	Optimism 95% interval
AUC	0.776	0.724	0.051	0.017 to 0.085
Brier score	0.164	0.182	−0.019	−0.037 to −0.003
Calibration intercept	−0.023	−0.017	−0.006	−0.156 to 0.142
Calibration slope	1.213	0.905	0.307	0.118 to 0.483

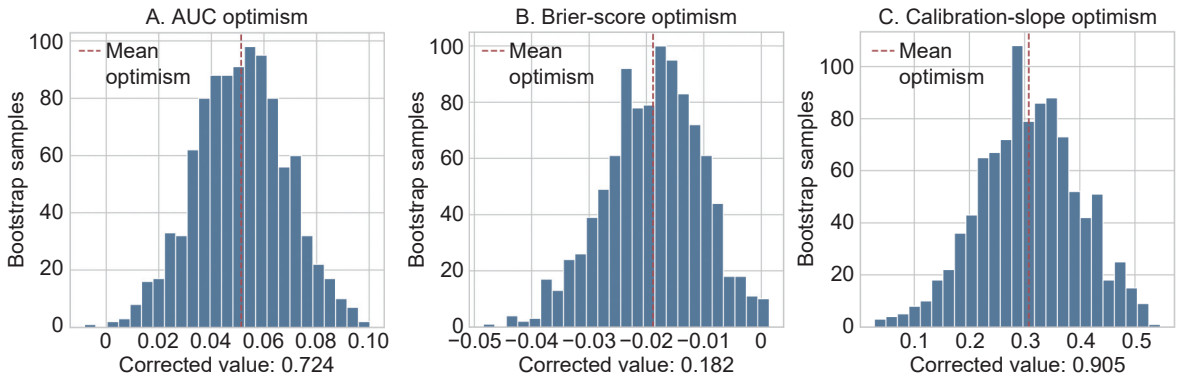


Figure 9 Bootstrap optimism in model performance. (A-C) show optimism in AUC, Brier score, and calibration slope across 1,000 bootstrap samples. Dashed lines mark mean optimism; corrected estimates appear below each panel.

symptomatic and asymptomatic plaques on grayscale and strain elastography images [6]. Ultrasound radiomics has also been used to assess plaque stability, while CTA and MRI studies have reported texture differences associated with ischemic events [7-9,22]. Systematic reviews nevertheless describe marked variation in segmentation, preprocessing, feature selection, and validation [12-15]. To reduce the influence of this methodological variability, we retained only features that demonstrated adequate measurement reliability and were consistently selected across repeated bootstrap samples.

The strongest inverse association involved Wavelet-LH_glrlm_RunEntropy, whereas several robust positive associations involved inverse variance, maximum probability, and size zone nonuniformity. Radiomics feature

清華大學出版社 Tsinghua University Press | SciOpen

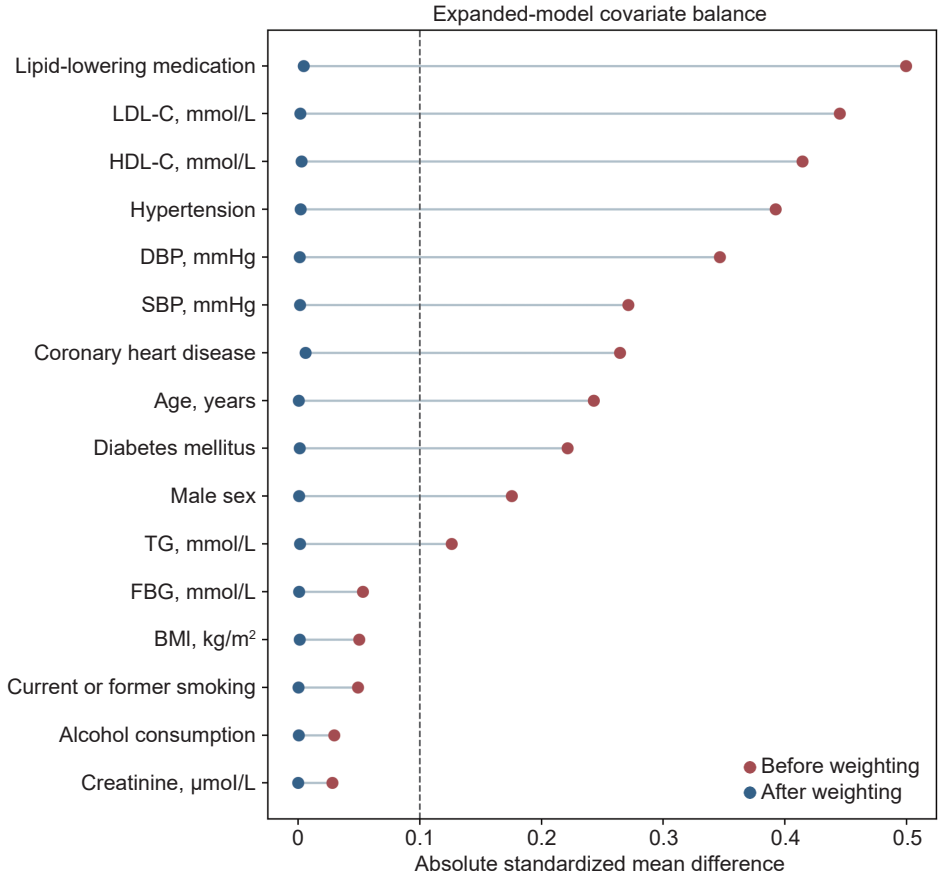


Figure 10 Balance of 16 clinical covariates in the expanded model. Points show absolute SMDs before and after overlap weighting; the dashed line marks 0.10.

Table 5 Robust features in the primary and expanded clinical models

Radiomics feature	Primary OR (95% CI)	Primary q	Expanded OR (95% CI)	Expanded q
wavelet-LH_glcm_InverseVariance	1.35 (1.10-1.64)	0.026	1.31 (1.07-1.61)	0.075
original_firstorder_10Percentile	0.71 (0.58-0.88)	0.016	0.72 (0.58-0.90)	0.054
wavelet-HL_glcm_MaximumProbability	1.50 (1.13-1.97)	0.029	1.48 (1.11-1.96)	0.060
wavelet-LH_glrlm_RunEntropy	0.60 (0.49-0.74)	< 0.001	0.64 (0.52-0.79)	0.013
wavelet-LH_glszm_SizeZoneNonUniformityNormalized	1.37 (1.09-1.72)	0.039	1.25 (1.00-1.55)	0.204
wavelet-HH_glszm_GrayLevelNonUniformityNormalized	0.74 (0.60-0.90)	0.027	0.81 (0.66-0.99)	0.189
wavelet-HL_glszm_SizeZoneNonUniformityNormalized	1.38 (1.10-1.71)	0.029	1.24 (1.00-1.55)	0.203

Odds ratios are per one-standard-deviation increase in the radiomics feature. CI, confidence interval; OR, odds ratio

names are mathematical summaries, not direct histological measurements. Lower entropy may indicate more ordered run length patterns after wavelet filtering, and higher maximum probability may reflect a more concentrated distribution of neighboring gray level pairs, but histological validation was not included in this study. These interpretations remain hypotheses and require direct pathology-referenced evaluation.

Reliability screening reduced the influence of manual segmentation on feature selection. The repeat delineation after one month showed that 62 features failed the intraobserver threshold even though interobserver reliability was high. Requiring both ICC values to reach 0.80 prevented these features from entering the main analysis and followed current recommendations for quantitative imaging markers [16,19].

Beyond measurement reliability, bootstrap-based feature selection and full-pipeline internal validation addressed two further sources of uncertainty [16,21]. The seven features recurred across bootstrap samples, while the decrease in AUC from 0.776 to 0.724 after optimism correction indicates that the apparent discrimination was somewhat optimistic. These analyses strengthen the internal assessment of the findings but

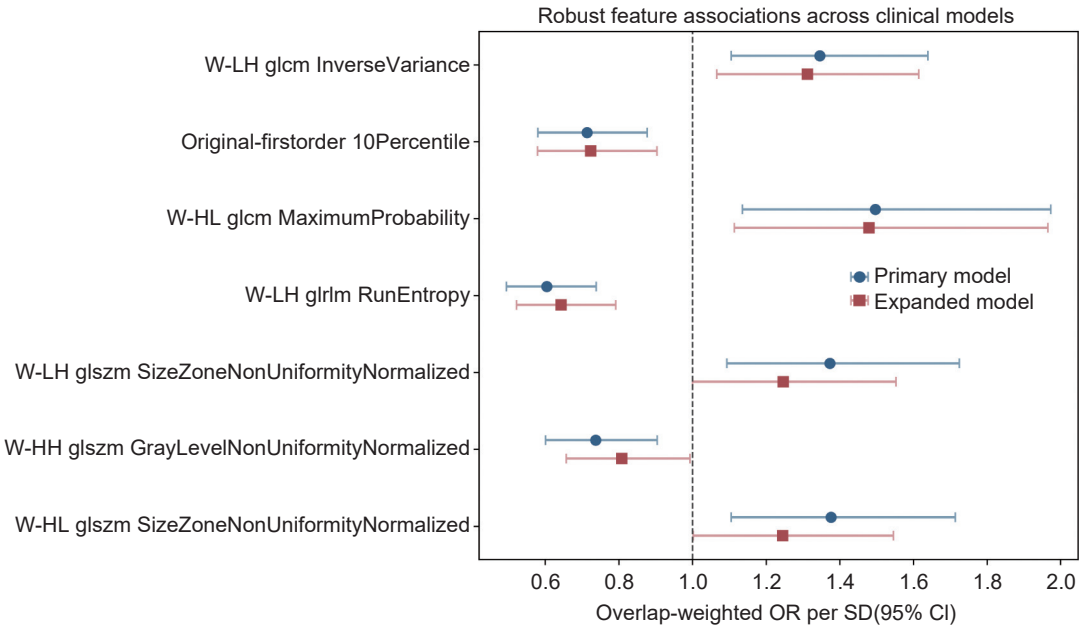


Figure 11 Overlap-weighted associations of seven robust radiomics features in the primary and expanded clinical models. Points show odds ratios per standard deviation, horizontal lines show 95% confidence intervals, and the dashed line marks an odds ratio of 1.

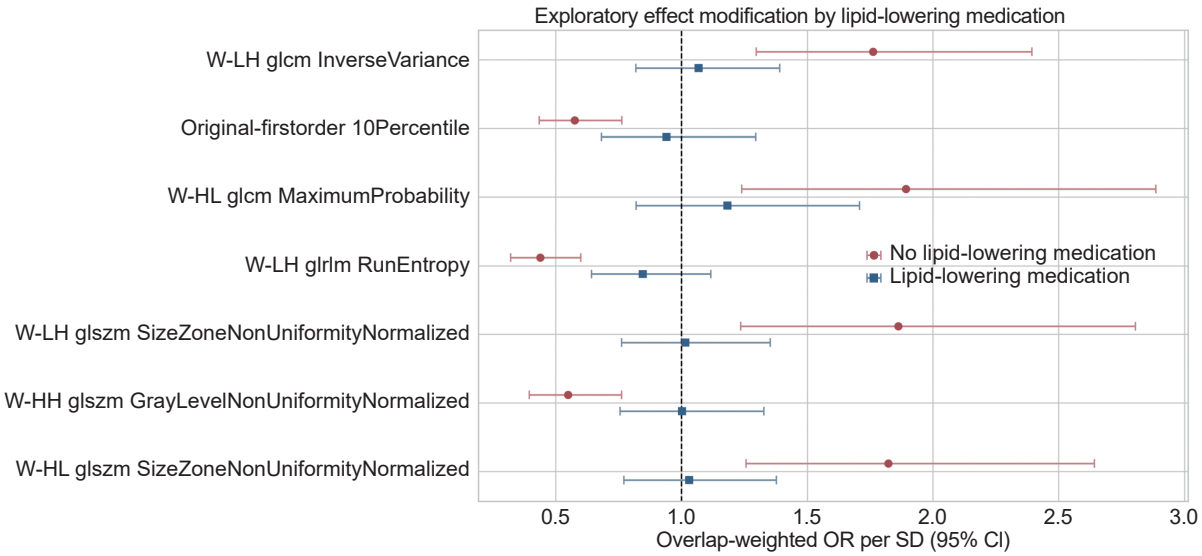


Figure 12 Exploratory associations of the seven robust radiomics features with IS by lipid-lowering medication use. Points show overlap-weighted ORs per SD with 95% CIs; the dashed line marks an OR of 1.

cannot establish how the features will perform in other clinical settings or ultrasound systems [16].

The additional IS cohort of 83 patients, all of whom had moderate or severe carotid plaque, was included in an exploratory comparison with the 361 patients with IS in the primary cohort. Several of the seven robust features differed between the two case series (largest absolute SMD, 0.882). Because the additional cohort had no patients without IS, this analysis assessed variation in feature distributions among IS cases rather than validating their associations with IS. These differences suggest that feature distributions should be examined before a fixed radiomics signature is applied to a clinically different population.

Several baseline findings require cautious interpretation. Patients with IS were younger and had lower cholesterol concentrations despite a greater burden of hypertension and diabetes. All measurements were obtained at the time of ultrasound, but prior treatment could have lowered lipid concentrations, especially because lipid lowering medication was more common in the IS group. Referral patterns and selection into the hospital and community cohorts may also have influenced age and risk factor distributions. Overlap weighting balanced the eight clinical covariates in the primary model, although cholesterol measurements remained

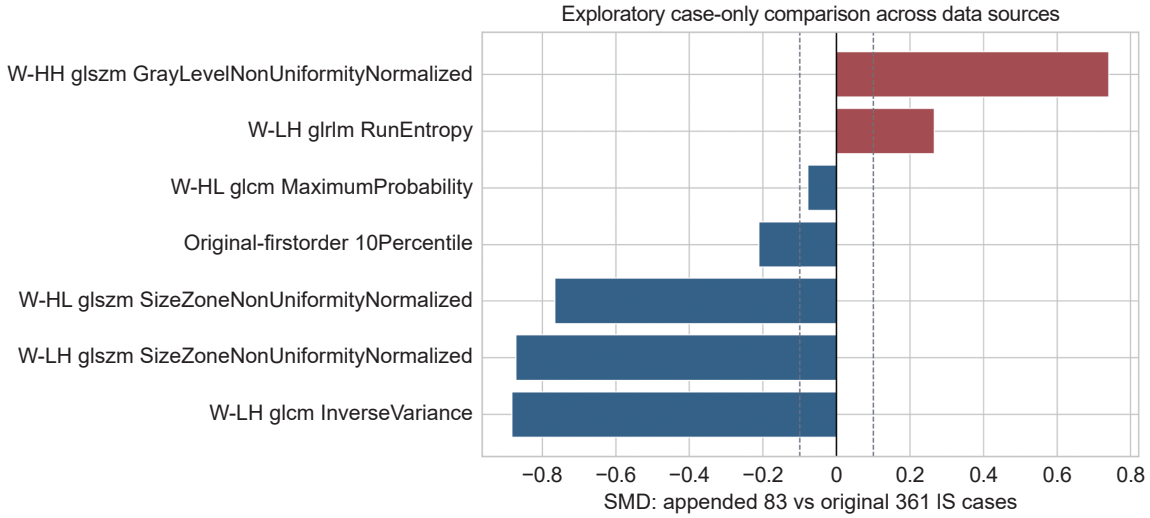


Figure 13 Differences in robust radiomics features between the two IS case series. Bars show SMDs comparing the 83 additional IS cases with the 361 IS cases in the primary cohort; positive values indicate higher feature values in the additional cohort.

imbalanced. Residual confounding from treatment intensity, disease duration, and other unmeasured factors remains possible [2,23].

Overlap weighting improved comparability between patients with and without IS across the eight clinical variables in the primary model [20]. Other clinical measurements were not necessarily balanced, and the expanded model examined how the findings changed when additional variables were considered. All seven robust features retained their association directions, although only Wavelet-LH_glrlm_RunEntropy remained significant after FDR correction. Thus, the directions were consistent, but the statistical evidence for several features was weaker under the expanded model. Blood pressure and laboratory measurements were obtained at enrollment, after the IS event in affected patients, and may reflect treatment or changes associated with the event as well as prior risk.

This study has several limitations. First, this was a retrospective study, which may limit the generalizability of the findings. Second, radiomics analysis was based on a single two-dimensional frame acquired using one ultrasound system, which may limit assessment of whole-plaque heterogeneity and generalizability to other settings. Third, the supplementary population included only 83 patients with IS and moderate or severe carotid plaque and lacked a comparison group without IS from the same source. Finally, although overlap weighting balanced the eight prespecified clinical covariates, residual confounding from unmeasured factors remains possible. The expanded analysis included measurements obtained after the IS event, which complicates their interpretation. Bootstrap resampling provided internal validation, but prospective multicenter validation is required.

Conclusions

Carotid plaques from patients with and without IS had different HRUS radiomics profiles after overlap weighting for measured clinical factors. Seven features met the prespecified reliability, association, and stability criteria in the primary cohort. These findings suggest that HRUS radiomics captures differences in plaque echogenicity and spatial texture associated with IS. Prospective multicenter studies using different ultrasound systems are needed to assess their generalizability and clinical value.

Acknowledgment

Not applicable.

Authors' Contributions

Shenglan Yang: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, and Writing – original draft. Xiaoyu Luo: Investigation, Data curation, Formal analysis, and Writing – original draft. Wendi Zheng, Xiaoyu Yue, Wenhao Liu, Yanan Pan, Yichun Zhou, and Xiaoyu Wan: Investigation, Data curation, and Formal analysis. Mingjun Xu: Conceptualization, Methodology, Resources, and Supervision. Mei Zhang: Conceptualization, Methodology, Resources, Supervision, Project administration, and Funding acquisition. All authors reviewed and approved the final manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (No. 2022YFC3602403).

Data Availability

Deidentified data supporting the findings of this study and the analysis code are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The institutional ethics committee approved the retrospective study and waived the requirement for informed consent. Identifying approval details are provided in the separate title page.

Conflict of interest

The authors have no competing interests to declare that are relevant to the content of this article.

Consent for publication

Not applicable. The manuscript contains no identifiable individual information.

References

[1] GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurology 2021;20:795-820.

[2] Naylor AR, Rantner B, Ancetti S. European society for vascular surgery 2023 clinical practice guidelines on the management of atherosclerotic carotid and vertebral artery disease. European Journal of Vascular and Endovascular Surgery 2023;65:7-111.

[3] Saba L, Brinjikji W, Spence JD, Wintermark M, Castillo M, Borst GJD de, et al. Roadmap consensus on carotid artery plaque imaging and impact on therapy strategies and guidelines: an international, multispecialty, expert review and position statement. AJNR Am J Neuroradiol 2021;42:1566-1575.

[4] Zwanenburg A, Vallieres M, Abdalah MA. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 2020;295:328-338.

[5] Griethuysen JJM van, Fedorov A, Parmar C. Computational radiomics system to decode the radiographic phenotype. Cancer Research 2017;77:e104-e107.

[6] Huang Z, Cheng X-Q, Liu H-Y, Bi X-J, Liu Y-N, Lv W-Z, et al. Relation of carotid plaque features detected with ultrasonography-based radiomics to clinical symptoms. Translational Stroke Research 2022;13:970-982.

[7] Song J, Zou L, Li Y, Wang X, Qiu J, Gong K. Combining artificial intelligence assisted image segmentation and ultrasound based radiomics for the prediction of carotid plaque stability. BMC Medical Imaging 2025;25:89.

[8] Chen S, Liu C, Chen X, Liu WV, Ma L, Zha Y. A radiomics approach to assess high risk carotid plaques: a non invasive imaging biomarker, retrospective study. Frontiers in Neurology 2022;13:788652.

[9] Shi J, Sun Y, Hou J. Radiomics signatures of carotid plaque on computed tomography angiography: an approach to identify symptomatic plaques. Clinical Neuroradiology 2023;33:931-941.

[10] Mohler ER 3rd, Sibley AA, Schultz SM, Zhang L, Sehgal CM. High-frequency ultrasound for evaluation of intimal thickness. J Am Soc Echocardiogr 2009;22:1129-1133.

[11] Xu M, Jin S, Li F, Jia G, Zhang C, Zhang M, et al. The diagnostic value of radial and carotid intima thickness measured by high-resolution ultrasound for ischemic stroke. Journal of the American Society of Echocardiography 2021;34:72-82.

[12] Vacca S, Scicolone R, Gupta A. Atherosclerotic carotid artery disease radiomics: a systematic review with meta-analysis and radiomic quality score assessment. European Journal of Radiology 2024;177:111547.

[13] Hou C, Liu XY, Du Y. Radiomics in carotid plaque: a systematic review and Radiomics Quality Score assessment. Ultrasound in Medicine and Biology 2023;49:2437-2445.

[14] Pfaehler E, Zwanenburg A, Jong JR de, Boellaard R. Repeatability and reproducibility of radiomic features: a systematic review. Radiotherapy and Oncology 2019;133:176-187.

[15] Teng X, Wang Y, Nicol AJ. Enhancing the clinical utility of radiomics: addressing the challenges of repeatability and reproducibility in CT and MRI. Diagnostics 2024;14:1835.

[16] Wang X, Pennello G, deSouza NM. Multiparametric data-driven imaging markers: guidelines for development, application and reporting of model outputs in radiomics. Academic Radiology 2023;30:215-229.

[17] An C, Park YW, Ahn SS, Han K, Kim H, Lee S-K. Radiomics machine learning study with a small sample size: Single random training-test set split may lead to unreliable results. PLoS One 2021;16:e0256152.

[18] Johri AM, Nambi V, Naqvi TZ. Recommendations for the assessment of carotid arterial plaque by ultrasound for the characterization of atherosclerosis and evaluation of cardiovascular risk. Journal of the American Society of Echocardiography 2020;33:917-933.

[19] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine 2016;15:155-163.

[20] Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. Journal of the American Statistical Association 2018;113:390-400.

[21] Meinshausen N, Buhlmann P. Stability selection. Journal of the Royal Statistical Society Series B 2010;72:417-473.

[22] Chen C, Tang W, Chen Y, Xu W, Yu N, Liu C. Computed tomography angiography based radiomics model to identify high risk carotid plaques. Quantitative Imaging in Medicine and Surgery 2023;13:6089-6104.

[23] Kleindorfer DO, Towfighi A, Chaturvedi S. 2021 guideline for the prevention of stroke in patients with stroke and transient ischemic attack. Stroke 2021;52:e364-e467.

清華大學出版社 Tsinghua University Press | SciOpen