

Ensemble Machine Learning Classifiers Combining CT Radiomics and Clinical-Radiological Features for Preoperative Prediction of Pathological Invasiveness in Lung Adenocarcinoma Presenting as Part-Solid Nodules: A Multicenter Retrospective Study

Yunhua Li, BS^{1,†}, Jianbang Ding, MS^{2,†}, Kun Wu, MS^{3,†}, Wanyin Qi, BS⁴, Shanyue Lin, BS⁵, Gangwen Chen, BS¹ , and Zhichao Zuo, PhD⁶ 

Abstract

Background: Lung adenocarcinomas manifesting as part-solid nodules (PSNs) represent a distinct clinical subtype where accurate preoperative determination of pathological invasiveness critically influences both prognosis and surgical decision-making. This multicenter study aims to develop an ensemble machine learning classifier that integrates computed tomography (CT) radiomic signatures with clinical-radiological features to enhance the preoperative prediction of invasive status.

Methods: We retrospectively analyzed 344 patients with pathologically confirmed lung adenocarcinoma presenting as PSNs across three medical centers. Following random allocation into training ($n = 240$) and validation ($n = 104$) sets (7:3 ratio), we extracted 1239 quantitative radiomic features from preoperative thin-section CT scans. Through rigorous feature engineering, we constructed a radiomic score using least absolute shrinkage and selection operator regression. We systematically evaluated both single-algorithm classifiers and ensemble approaches (including hard/soft voting and stacking), incorporating both the radiomic score and clinical-radiological features.

Results: Among the various evaluated machine learning models, the stacking classifier, which combines radiomic scores and clinical-radiological features, performed the best, achieving an AUC of 0.84, an accuracy of 0.817, an FI score of 0.869, a precision of 0.818, and a recall of 0.926.

Conclusion: Our stacking ensemble learning classifier, which synergistically combines CT radiomics signatures with clinical-radiological features, provides a clinically actionable tool for the preoperative prediction of pathological invasiveness in PSN-type lung adenocarcinoma, thereby enhancing individualized surgical planning.

Keywords

ensemble machine learning classifiers, radiomic analysis, pathological invasiveness, lung adenocarcinomas, part-solid nodules

Received: 19 March 2025; revised: 2 May 2025; accepted: 28 May 2025

¹Department of Radiology, the Jintang First People's Hospital, West China Sichuan University Jintang Hospital, Chengdu, China

²Department of Stomatology, Xiangtan Central Hospital, Xiangtan, China

³Department of Cardiothoracic Surgery, Xiangtan Central Hospital, Xiangtan, China

⁴Department of Radiology, the Affiliated Hospital of Southwest Medical University, Luzhou, China

⁵Department of Radiology, Affiliated Hospital of Guilin Medical University, Guilin, China

⁶Department of Radiology, Xiangtan Central Hospital, Xiangtan, China

[†]These authors contributed equally to this work.

Corresponding Authors:

Gangwen Chen, Department of Radiology, the Jintang First People's Hospital, West China Sichuan University Jintang Hospital, Chengdu, 610400, Sichuan, China. Email: 3979232755@qq.com

Zhichao Zuo, Department of Radiology, Xiangtan Central Hospital, Xiangtan, 411100, Hunan, China.

Email: zuo_z@smail.xtu.edu.cn



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and

distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Introduction

The global prevalence of lung cancer makes it one of the most lethal malignancies, with adenocarcinoma having the highest incidence among lung cancers.^{1–3} According to the 2017 guidelines issued by the Fleischner Society, lung cancer observed in computed tomography (CT) imaging is categorized into three distinct types based on the proportion of solid to ground-glass components. These categories include pure ground-glass nodules (pGGNs), characterized exclusively by ground-glass opacity; solid nodules (SNs), which consist entirely of solid tissue; and part-solid nodules (PSNs), which comprise both ground-glass and solid components.⁴ Among these, PSNs represent a distinct clinical subtype characterized by a lower prevalence of lymphatic metastasis, visceral pleural invasion, and spread through air spaces compared to SNs. Furthermore, patients with PSNs tend to have better recurrence-free survival (RFS) and overall survival rates than those with SNs, although these rates are worse than in patients with pGGNs.^{5–7}

Lung adenocarcinomas presenting as PSNs can be classified pathologically as adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), or invasive adenocarcinoma (IAC), depending on their invasive characteristics. The 5-year recurrence-free survival (RFS) after complete resection approaches 100% for AIS and MIA, whereas it drops to approximately 74.1% for IAC.^{8,9} Therefore, wedge resection or segmentectomy is generally recommended for AIS and MIA because these procedures can preserve lung tissue and significantly reduce surgery-related complications. In contrast, lobectomy is typically necessary for IAC to achieve a lower tumor recurrence rate, despite its higher risk of adverse surgical outcomes.^{10,11} Consequently, preoperative prediction of IAC in lung adenocarcinomas manifesting as PSNs is vital for guiding treatment decisions, especially regarding the choice of surgical extent.

Radiomics has shown considerable potential in elucidating the biological characteristics of lung cancer through quantitative imaging analysis. Recent studies highlight its clinical effectiveness in predicting high-grade patterns in stage I lung adenocarcinoma,¹² enhancing preoperative tumor grading accuracy,¹³ and differentiating immune-related pneumonitis from tumor progression in stage IV non-small cell lung cancer during immunotherapy,¹⁴ thereby impacting clinical decisions. Furthermore, hybrid models that combine radiomics with clinical and radiological features demonstrate promising potential for preoperative evaluation of the pathological invasiveness of lung adenocarcinoma presenting as PSNs.^{15–18} However, in practical applications, selecting the most appropriate statistical method to integrate radiomics with clinical-radiological data can be challenging. Traditional methods relying on single-model frameworks, such as logistic regression with fixed parameters, often fail to capture the complex patterns and relationships within the data.^{15–18} Machine learning algorithms, on the other hand, can better model nonlinear relationships and high-dimensional interactions, thereby reducing the risks of overfitting and underfitting while enhancing generalizability. Future research in this

field would particularly benefit from leveraging machine learning approaches.¹⁹

To address this, we used both single-algorithm classifiers and ensemble machine learning (including hard/soft voting and stacking) to combine multiple models for final predictions,^{19–22} integrating radiomics with clinical-radiological features. Ensemble classifiers merge different models by combining various algorithms or data perspectives, which help balance their strengths and weaknesses. Consequently, they enhance overall performance in complex medical decision-making, leading to more accurate and reliable predictions than any single model alone. This is the first study to systematically compare and combine these techniques for predicting lung adenocarcinoma invasiveness in PSNs. Our goal is to develop robust prediction models by reducing complexity and maximizing data utilization, thereby guiding future efforts in selecting effective diagnostic tools for personalized treatment planning.

Materials and Methods

Study Population

In this retrospective three-center study, we collected CT images and pathological diagnoses from patients who underwent surgical resection for lung adenocarcinoma between January 2021 and June 2023. The inclusion criteria were as follows: (a) PSNs with a maximum diameter of 5–30 mm on preoperative CT images; (b) postoperative pathological examination confirming lung adenocarcinoma; and (c) a chest CT scan performed within 14 days prior to surgery. The exclusion criteria included: (a) patients who had received preoperative chemoradiotherapy; (b) patients with multiple primary lung cancers or other malignancies; and (c) unqualified images containing artifacts, atelectasis, or pneumonia, which complicated segmentation. The screening flowchart of enrolled patients is provided in Supplement Figure 1. This retrospective study enrolled 344 patients with pathologically confirmed lung adenocarcinoma presenting as PSNs. Among these patients, 235 (68.3%) were diagnosed with IAC, while 109 (31.7%) were classified as non-invasive adenocarcinoma. Within the non-IAC cases, 98 (28.5%) were classified as MIA, and 11 (3.2%) as AIS. Of the 344 enrolled lung adenocarcinoma patients, 223 (64.8%) were male, and 121 (35.2%) were female, with a median age of 59 years (interquartile range [IQR]: 52–66). The participants were divided into training ($n = 240$, 69.8%) and validation ($n = 104$, 30.2%) sets in a 7:3 ratio without stratification. The training set included 156 males (65.0%) with a median age of 59 years (IQR: 52–66.2), while the validation set comprised 67 males (64.4%) with a median age of 59.5 years (IQR: 52–66).

The institutional review boards of all participating institutions approved the study (2021-07-009 from Xiangtan Central Hospital, KY2020147 from the Affiliated Hospital of Southwest Medical University, and 2023YJSLL-121 from the Affiliated Hospital of Guilin Medical University). Due to the retrospective nature of the study and the use of anonymized data, the boards waived the requirement to obtain written informed consent.

CT Radiological Feature Evaluation

The CT scanning parameters used in this study were consistent with those from our previous research.^{12,13} For further details on the acquisition of CT images, please refer to Supplemental Digital Content 1.

Eligible CT images collected from the three centers were co-registered and stored in DICOM format. These images were subsequently imported into the Picture Archiving and Communication System to visualize all enrolled PSNs. The CT radiological features of these PSNs on high-resolution CT images were interpreted by senior cardiothoracic radiologists, who were not informed of the pathological diagnoses. Tumor size was defined as the maximum diameter, specifically the longest perpendicular diameter in the axial plane, according to the eighth edition of the TNM classification of lung cancer.²³

Additionally, the consolidation-to-tumor ratio (CTR) was defined as the ratio of the maximum size of the solid component to the maximum tumor size, as illustrated in the schematic diagram (Supplement Figure 2). CTR measurements were taken for the maximum diameter of both the entire tumor and the solid component within the lung window, which has a width of 1500 Hounsfield units (HU) and a level of -700 HU.^{24,25} The solid component was characterized as a region of heightened opacification that fully concealed the vascular markings beneath it. In contrast, ground-glass opacity is defined as an area with a minor, homogeneous increase in density that does not obscure the vascular markings below. Finally, the other CT radiological features included location, margin, shape, lobulation sign, spiculation sign, vascular convergence sign, vacuole sign, and pleural indentation sign. Any disagreements in the CT image evaluation were discussed collaboratively until a consensus diagnosis was reached.

Lesion Delineation and Radiomics Workflow

All CT images underwent semiautomatic segmentation of PSNs using 3D Slicer software (version 5.2.2; www.slicer.org). The segmentation process began with an initial contouring protocol executed by a radiologist-in-training who specialized in thoracic oncology imaging and had five years of relevant experience. This initial contouring was subsequently refined through an iterative process by an attending thoracic radiologist with 15 years of subspecialty expertise, ensuring the accuracy of the volumetric measurements.

The radiomic analysis followed a structured pipeline consisting of four main phases: image processing, feature quantification, dimensionality optimization, and predictive modeling. First, we performed image resampling using the PyRadiomics package within the Python programming environment to standardize voxel spacing to $1 \times 1 \times 1$ mm³ (x, y, z). The resampling algorithm was based on the B-spline curve, with the interpolator set to `sitkBSpline` (degree = 3), ensuring smooth and accurate interpolation of the image data. Next, we adjusted the bin width to partition the pixel intensity range of 0–255 into five intervals. This binning approach effectively discretized the intensity

values, reducing the dimensionality of the feature space and facilitating subsequent analysis. Second, we extracted 1239 quantitative radiomics features per PSN, including shape, texture, and wavelet-filtered attributes. To standardize these features, normalization was performed using a z-score transformation (mean $\mu=0$, standard deviation $\sigma=1$). Following feature extraction, univariate statistical testing (two-tailed t-test, $\alpha=0.05$) was applied to exclude non-discriminative features. A pairwise collinearity assessment was then conducted to identify redundant variables using Pearson's correlation coefficient ($|r| \geq 0.9$). In cases where correlation was detected, one feature from each cluster of correlated features was iteratively eliminated through a stochastic selection process.

The final derivation of the radiomic signature was accomplished using least absolute shrinkage and selection operator (LASSO), incorporating nested 10-fold cross-validation to minimize the risk of overfitting (Supplement Figure 3). Features that retained non-zero coefficients following the regression analysis were integrated into a composite radiomic score (Radscore) through a weighted linear combination of the selected features.

$$\text{Radscore} = \sum_{j=1}^m \omega_j \cdot F_j, \quad (1)$$

where ω_j denotes the LASSO-derived coefficient for feature F_j .

Machine Learning Framework

The Radscore was synergistically integrated with clinical-radiological features to act as composite predictors in our computational modeling pipeline. We employed dual-layer analytical protocols that encompass both foundational and sophisticated statistical constructs, facilitating the preoperative discrimination of invasive adenocarcinoma manifesting as PSNs.

Single-Algorithm Classifiers

Fundamental predictive architectures implemented monolithic machine intelligence frameworks, including random forest (RF), adaptive boosting (AdaBoost), light gradient boosting machine (LightGBM), gradient boosting decision tree (GBDT), extreme gradient boosting (XGBoost), and categorical boosting (CatBoost). These methodologies, which employ arboreal integration through stratified data subspace allocation, wherein individual decision trees undergo parallel optimization across partitioned feature-phenotype cohorts, yield promising results in the analysis of biomedical data, particularly in the diagnosis of lung cancer.^{26–29}

Ensemble Machine Learning Classifiers

Ensemble machine learning classifiers utilize advanced integration strategies, including stacking classifiers and voting mechanisms. The stacking classifier architecture adopts a hierarchical two-level structure: the base level integrates predictions from six primary classifiers (RF, GBDT, XGBoost, LightGBM,

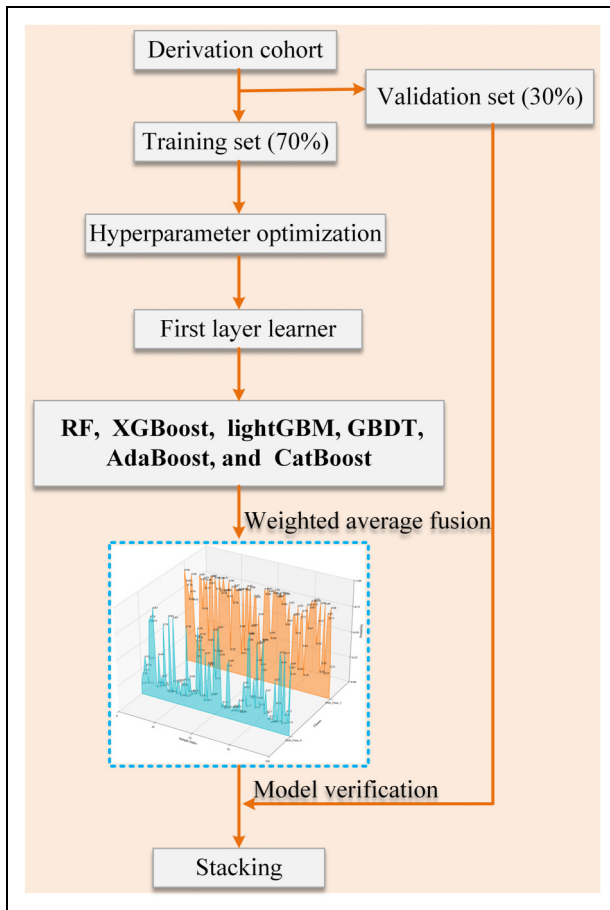


Figure 1. Architecture of the Stacking Classifier.

AdaBoost, CatBoost), while the meta-level employs a logistic regression meta-classifier to optimally combine these outputs through cross-validated probability calibration, as depicted in Figure 1. Parallel voting strategies operate through two distinct consensus protocols: hard voting, which determines class membership by majority rule across base classifier decisions, and soft voting, which calculates class probabilities via the weighted average of individual classifier confidence scores. Additionally, stacking classifiers—built upon a dual integration framework—enhance predictive robustness by simultaneously mitigating model bias through stacked generalization and minimizing variance via consensus-based decision aggregation. This framework maintains strict separation between base learner training spaces and meta-learner optimization domains, ensuring modular architecture and preventing information leakage.³⁰

Model Selection Protocol

The experimental design schematic outlines our multi-criteria optimization framework (Figure 2). Diagnostic performance quantification utilizes the area under the receiver operating characteristic (ROC) curve (AUC) as the principal criterion, supplemented by an extensive evaluation matrix that encompasses accuracy, precision, recall, and F1 score. Stratified

partitioning allowed the development cohort to facilitate model parameterization, while the independent validation cohort enabled unbiased performance benchmarking. The classifier demonstrating maximal AUC robustness during validation analyses was designated as the optimal configuration.³¹

Interpretability Analysis Framework

Model explainability was enhanced through the implementation of Shapley Additive Explanations (SHAP) to quantify feature contribution hierarchies.^{32,33} A sequential optimization protocol was initiated through ordered feature integration based on SHAP-derived importance rankings. The algorithmic workflow involved cyclical retraining of classifier architectures with progressive feature inclusion, coupled with validation cohort AUC recalibration at each augmentation stage. This dynamic feature integration paradigm enabled simultaneous monitoring of incremental predictive gain and the evolution of SHAP value trajectories. Through incremental learning frameworks, the parametric impact of individual predictors on global model performance was visualized via multivariate gradient mapping, establishing explicit correlations between feature-space perturbations and performance metric fluctuations across sequential learning iterations.

Statistical Analyses

All statistical analyses were performed using R software (version 3.6.3). Categorical variables were expressed as frequencies and percentages (%), whereas continuous variables were summarized as means with standard deviations for normally distributed data or as medians with interquartile ranges for non-normally distributed data. To compare groups, chi-square tests were used for categorical variables, and either independent samples t-tests or Mann-Whitney U tests were applied for continuous variables, depending on their distribution. A two-tailed p-value of less than 0.05 was considered statistically significant.

Result

Baseline Characteristics

This retrospective study enrolled 344 patients with pathologically confirmed lung adenocarcinoma presenting as PSNs. As shown in Table 1, a comparative analysis of the demographic and radiological characteristics between the training and validation sets revealed no statistically significant differences (all $p > 0.05$). This lack of significant disparities indicates a high level of homogeneity, ensuring balanced feature distributions for subsequent model development and evaluation.

The diagnosis of lung adenocarcinoma presenting as PSNs highlighted that pathological invasiveness was more common among male patients. This invasiveness was associated with advanced age, larger tumor sizes, the presence of vacuole signs, and higher CTR (see Supplement Table 1).

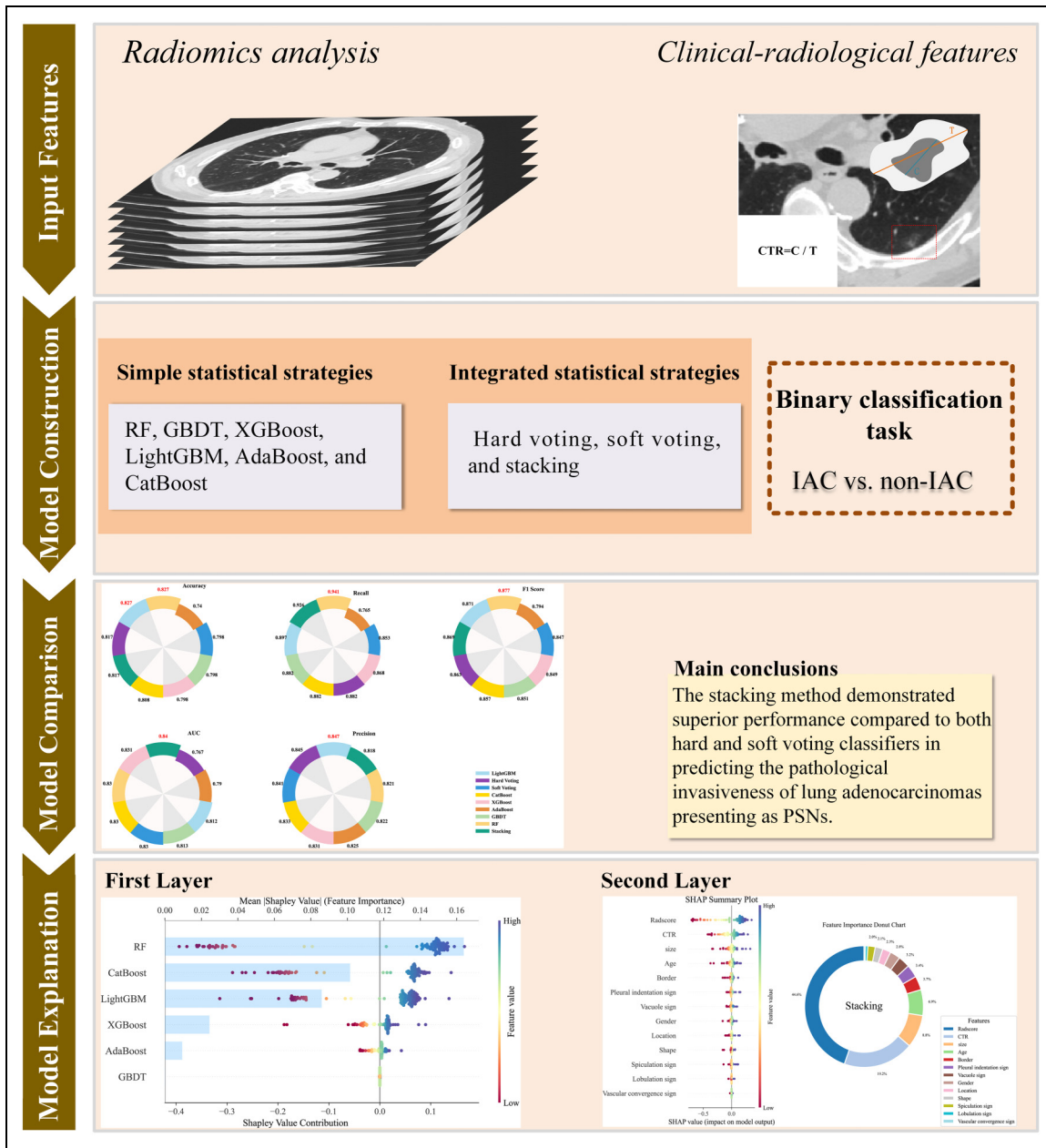


Figure 2. Schematic Representation of the Experimental Design Utilized in this Study.

Radiomic Analysis

Quantitative radiomic profiling of preoperative thin-section CT scans resulted in the identification of 1239 candidate radiomics features for each PSN. A rigorous dimensionality reduction protocol was employed to systematically refine these features in two phases. First, 639 radiomic features were selected using batch t-tests. This was followed by the retention of 188 radiomic features through a pairwise collinearity assessment, where Pearson's correlation coefficient ($|r|$) was set at ≥ 0.9 . Subsequently, this optimized set of radiomic features underwent LASSO regression with 10-fold cross-validation. This process led to the identification of 17 non-collinear features

that maintained predictive significance ($\lambda=0.027$; see Supplemental Figure 4). The details of these features are provided in Supplemental Digital Content 2. Additionally, inter-feature dependency analysis using Spearman's rank correlation confirmed minimal multicollinearity across the final signature components (see Supplement Figure 5), with absolute coefficient magnitudes consistently maintained below critical multicollinearity thresholds.

The Radscore is used for building radiomics models. In this binary classification task, we aim to distinguish IAC from non-IAC in lung adenocarcinoma presenting as PSNs. The radiomics model achieved an AUC of 0.799, an accuracy of 0.798, an F1 score of 0.840, a precision of 0.873, and a recall of 0.809.

Table 1. Comparative Analysis of Clinical-Radiological Features Between the Training and Validation Sets.

Variables	Total (n = 344)	Training set (n = 240)	Validation set (n = 104)	P value
Gender, n (%)				1
- Male	223 (64.8)	156 (65)	67 (64.4)	
- Female	121 (35.2)	84 (35)	37 (35.6)	
Age(y), Median (Q1, Q3)	59 (52, 66)	59 (52, 66.2)	59.5 (52, 66)	0.771
Size(mm), Median (Q1, Q3)	19.6 (14.9, 25.2)	19.8 (14.8, 25.2)	19.6 (15.2, 24.8)	0.804
Location, n (%)				0.52
- RUL	134 (39)	100 (41.7)	34 (32.7)	
- RLL	16 (4.7)	10 (4.2)	6 (5.8)	
- RML	54 (15.7)	38 (15.8)	16 (15.4)	
- LUL	97 (28.2)	63 (26.2)	34 (32.7)	
- LLL	43 (12.5)	29 (12.1)	14 (13.5)	
Border, n (%)				0.662
- Well- defined	193 (56.1)	137 (57.1)	56 (53.8)	
- Ill-defined	151 (43.9)	103 (42.9)	48 (46.2)	
Lobulation sign, n (%)				0.27
- Absent	100 (29.1)	65 (27.1)	35 (33.7)	
- Present	244 (70.9)	175 (72.9)	69 (66.3)	
Spiculation sign, n (%)				0.301
- Absent	133 (38.7)	88 (36.7)	45 (43.3)	
- Present	211 (61.3)	152 (63.3)	59 (56.7)	
Vascular convergence sign, n (%)				0.69
- Absent	69 (20.1)	50 (20.8)	19 (18.3)	
- Present	275 (79.9)	190 (79.2)	85 (81.7)	
Vacuole sign, n (%)				0.487
- Absent	255 (74.1)	181 (75.4)	74 (71.2)	
- Present	89 (25.9)	59 (24.6)	30 (28.8)	
Pleural indentation sign, n (%)				0.155
- Absent	93 (27)	59 (24.6)	34 (32.7)	
- Present	251 (73)	181 (75.4)	70 (67.3)	
Shape, n (%)				0.087
- Regular	102 (29.7)	64 (26.7)	38 (36.5)	
- Irregular	242 (70.3)	176 (73.3)	66 (63.5)	
CTR, n (%)				0.05
<1/4	93 (27)	63 (26.2)	30 (28.8)	
1/4–1/2	115 (33.4)	71 (29.6)	44 (42.3)	
1/2–3/4	86 (25)	68 (28.3)	18 (17.3)	
>3/4	50 (14.5)	38 (15.8)	12 (11.5)	
Pathological diagnosis, n (%)				0.521
Non-IAC	109 (31.7)	73 (30.4)	36 (34.6)	
IAC	235 (68.3)	167 (69.6)	68 (65.4)	

Abbreviations: IAC, invasive adenocarcinoma; CTR, consolidation-to-tumor ratio; RUL, right upper lobe; RLL, right lower lobe; RML, right middle lobe; LUL, left upper lobe; LLL, left lower lobe.

Machine Learning Performance Comparison

Within our predictive modeling framework, radiomics-derived Radscore and clinical-radiological features served as dual-channel inputs for evaluating pathological invasiveness in lung adenocarcinomas presenting as PSNs. Based on ROC analysis (Figure 3), a comparative evaluation of diverse algorithmic architectures revealed substantial performance disparities among the models, with clear stratification in AUC values. Notably, the XGBoost implementation demonstrated the optimal discriminative capacity with an AUC of 0.831 when evaluated as a base learner. In contrast, ensemble strategies exhibited varying levels of efficacy: conventional voting architectures, comprising hard voting (AUC = 0.770) and soft voting (AUC = 0.830), displayed suboptimal performance relative

to single-algorithm benchmarks. However, the stacking classifier achieved the peak diagnostic accuracy of AUC = 0.840 through meta-learner optimization (Table 2). Furthermore, a multidimensional performance evaluation (Figure 4) quantified classifier effectiveness across key metrics, revealing that the stacking classifier attained a balanced performance with an accuracy of 0.817, an F1 score of 0.869, a precision of 0.818, and a recall of 0.926.

Model Interpretation

When assessing the contribution of the first layer to the stacking classifier, the RF algorithm exerted the most significant influence, followed by CatBoost, LightGBM, XGBoost, AdaBoost, and GBDT, as shown in Figure 5. A global interpretation of

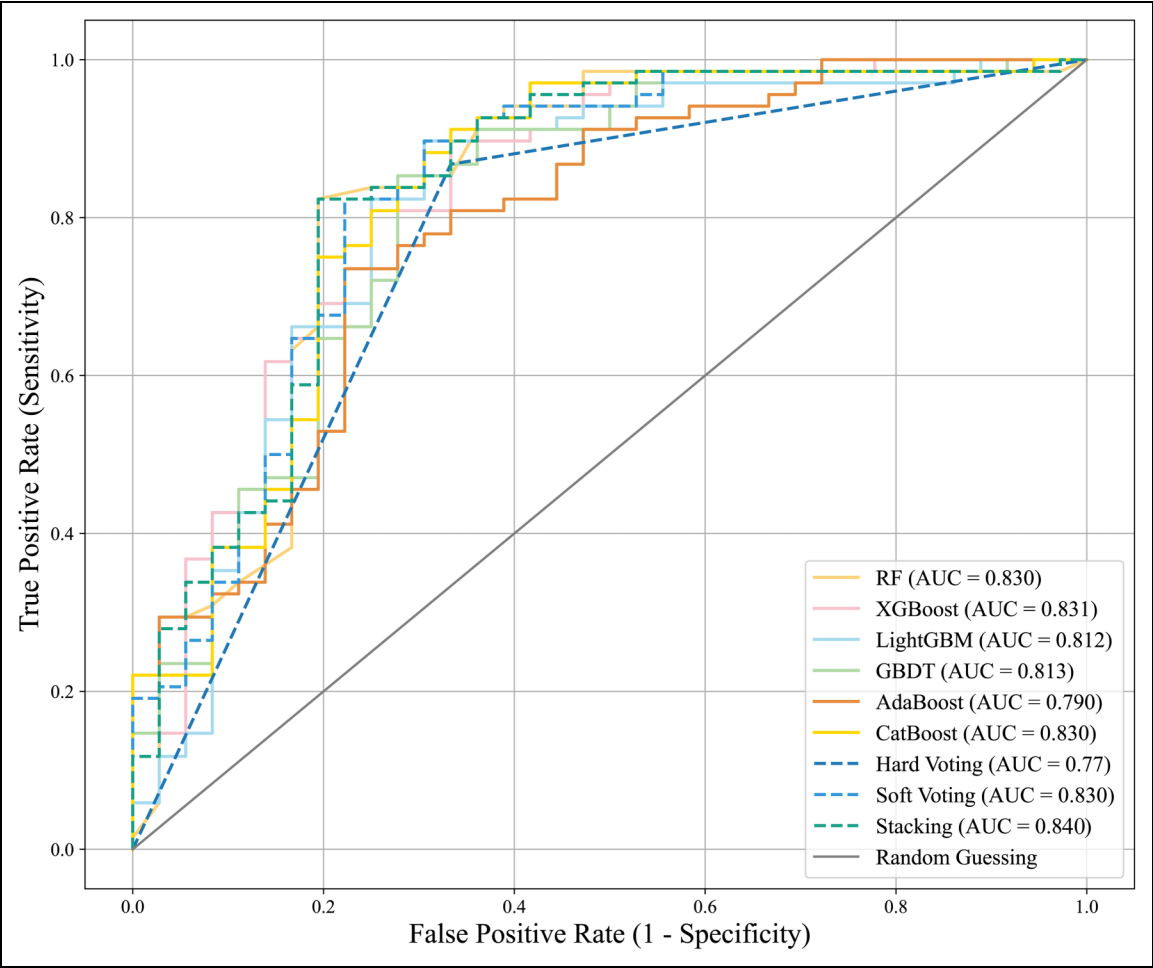


Figure 3. ROC Analysis Comparative Evaluation Showcasing Significant Performance Discrepancies among Various Algorithmic Architectures.

Table 2. Assessment of Diagnostic Efficacy Across Diverse Machine Learning Approaches.

Model	Accuracy	AUC	FI Score	Precision	Recall
RF	0.827	0.83	0.877	0.821	0.941
XGBoost	0.798	0.831	0.849	0.831	0.868
LightGBM	0.827	0.812	0.871	0.847	0.897
GBDT	0.798	0.813	0.851	0.822	0.882
AdaBoost	0.74	0.79	0.794	0.825	0.765
CatBoost	0.808	0.83	0.857	0.833	0.882
Hard Voting	0.817	0.767	0.863	0.845	0.882
Soft Voting	0.798	0.83	0.847	0.841	0.853
Stacking	0.817	0.84	0.869	0.818	0.926

Abbreviations: AUC, area under curve; GBDT, gradient boosting decision tree; AdaBoost, adaptive boosting; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; CatBoost, categorical boosting; RF, random forest.

the stacking classifier for the second layer, based on SHAP values, is presented in Figure 6. Among the features analyzed, Radscore had the most significant contribution, followed by CTR, size, age, border, pleural indentation sign, vacuole sign, gender, location, shape, spiculation sign, lobulation sign, and vascular convergence sign. Additionally, Figure 7 illustrates

the decoding of gender-specific feature contributions via the mean absolute SHAP values to address potential differences between male and female patients regarding both radiomic and clinical-radiological features. In this analysis, males contributed 53.7%, while females contributed 46.3%.

To enhance the clinical translatability of our SHAP analysis, we have included a real-world case example in the force plot (Figure 8), which illustrates the prediction explanation for a specific instance. In this plot, positive impacts are shown in red, while negative impacts are displayed in blue. The length of the arrows signifies the magnitude of each feature’s impact on the output.

Figure 9 demonstrates the relative contribution of different features to the model’s predictive performance, along with their corresponding AUC scores. The top four features (Radscore, CTR, size, age) exhibit significantly higher contribution values compared to the other variables. Notably, CTR shows the strongest contribution, indicating its substantial discriminative power in classification tasks. The next three highlighted features reveal a gradual increase in both contribution magnitude and associated AUC values. However, after incorporating additional features, the AUC began to decline slowly. Despite this, the AUC metric

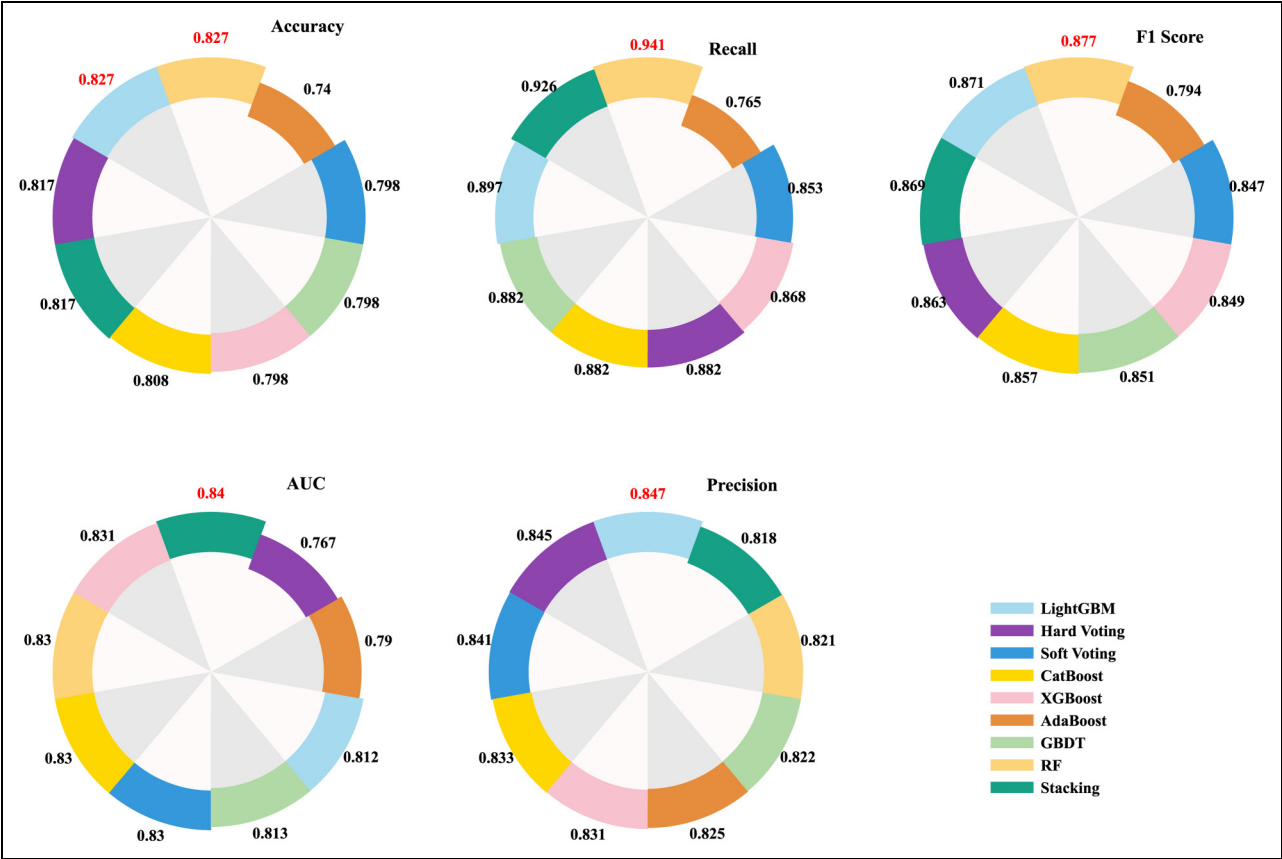


Figure 4. Multidimensional Performance Evaluation Quantitatively Assessing Classifier Efficacy Across Key Metrics, Including Accuracy, Precision, Recall, and F1 Score.

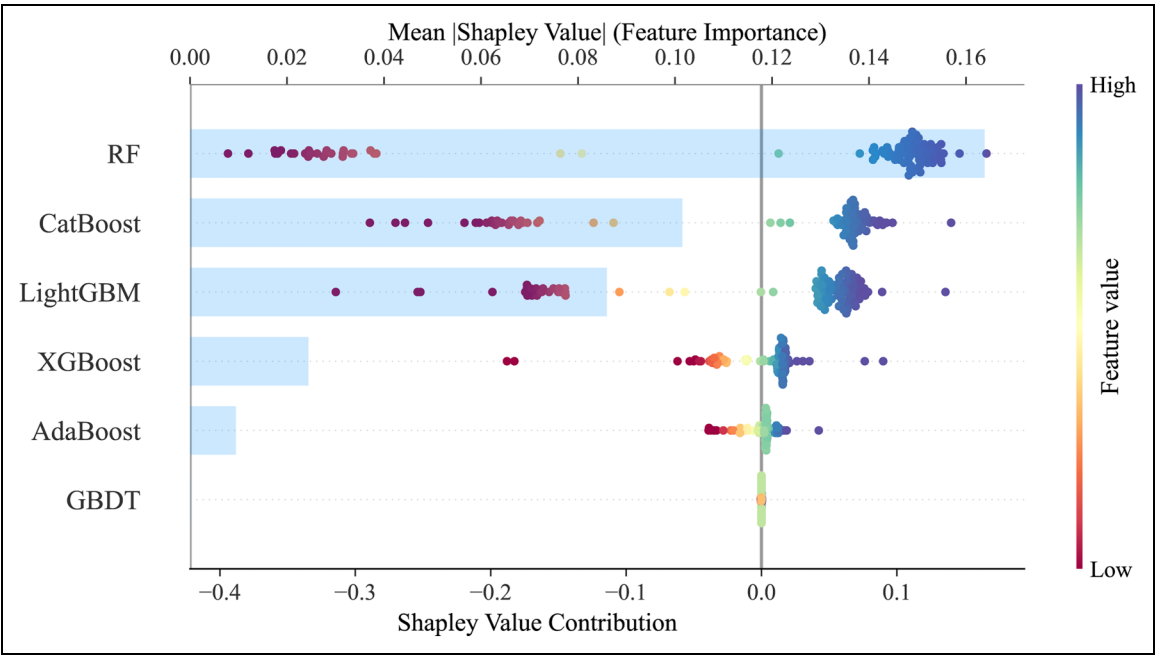


Figure 5. Bee Diagram Illustrating the Contribution of the First Layer Within the Stacking Classifier Framework.

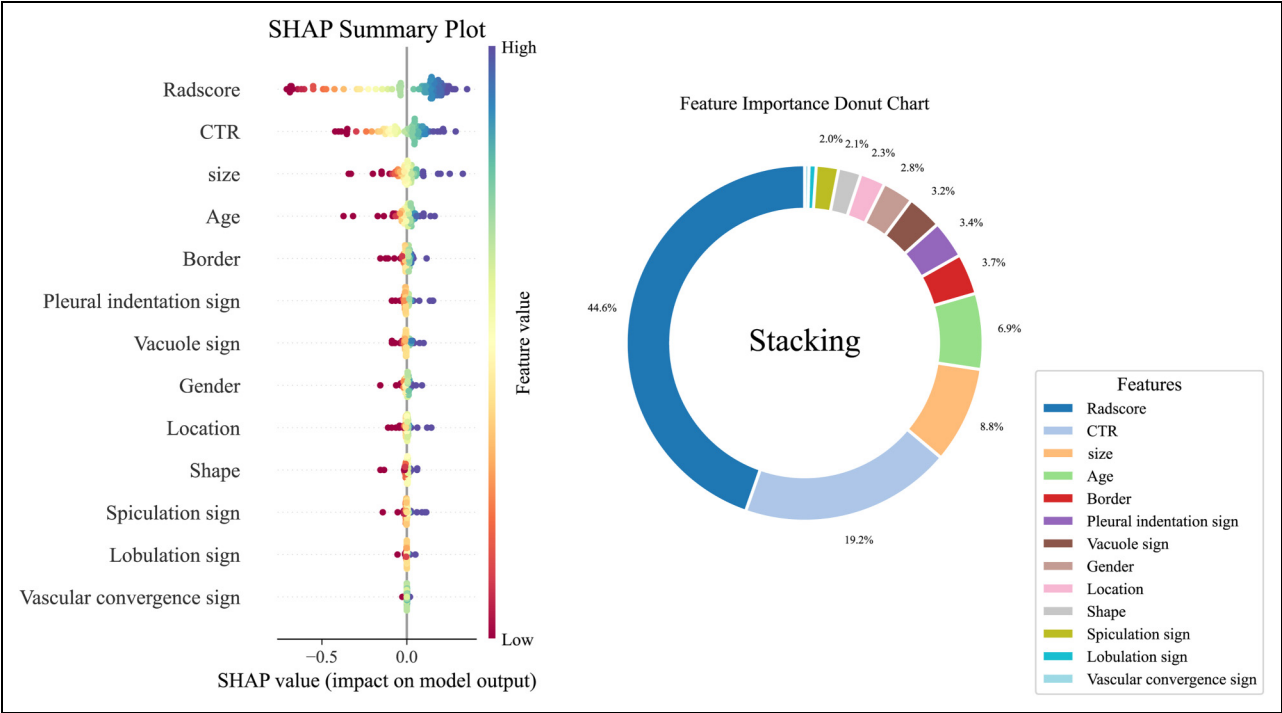


Figure 6. Global Interpretation of the Stacking Classifier for the Second Layer, Informed by SHAP Values.

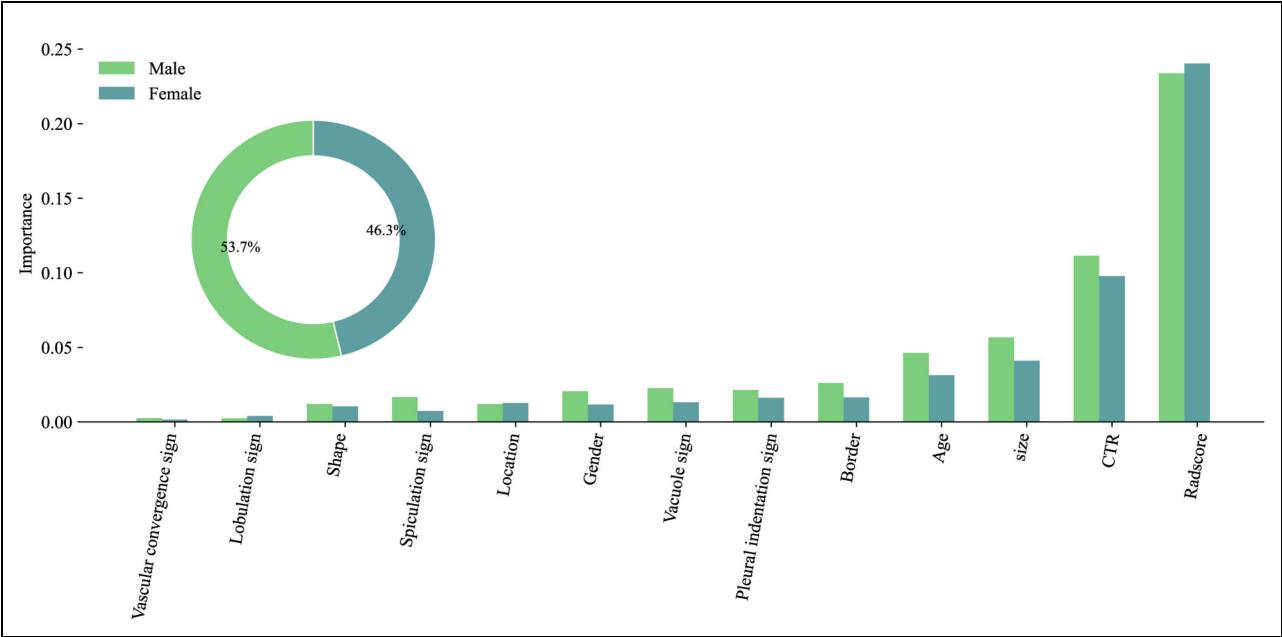


Figure 7. Male & Female Feature Contributions: Decoding with Mean Absolute SHAP.

remains consistently above 0.7 across all top features, suggesting that the model maintains adequate discrimination capability.

Discussion

This multicenter study demonstrates that ensemble machine learning classifiers, which integrate CT radiomic signatures

with clinical-radiological features, achieve robust performance in the preoperative prediction of pathological invasiveness for lung adenocarcinoma presenting as PSNs. The stacking classifier, leveraging meta-learning to harmonize heterogeneous base models, outperformed both conventional voting strategies and single-algorithm approaches.

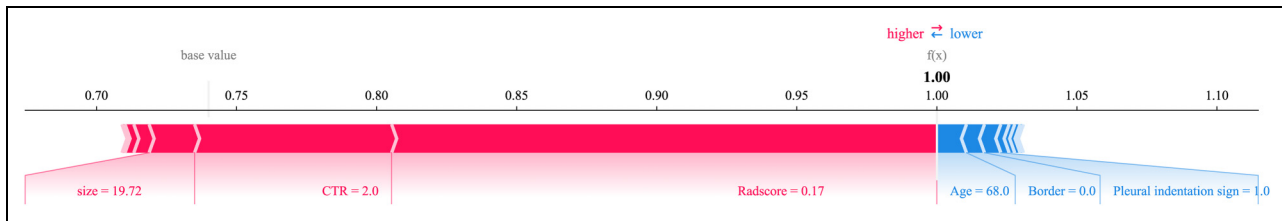


Figure 8. In this Force Plot, it is Evident that the Highest Rad Score (0.17), Followed by CTR (1/4–1/2) and Size (19.72 mm), Positively Influences the Risk of IAC in Lung Adenocarcinomas Presenting as PSNs.

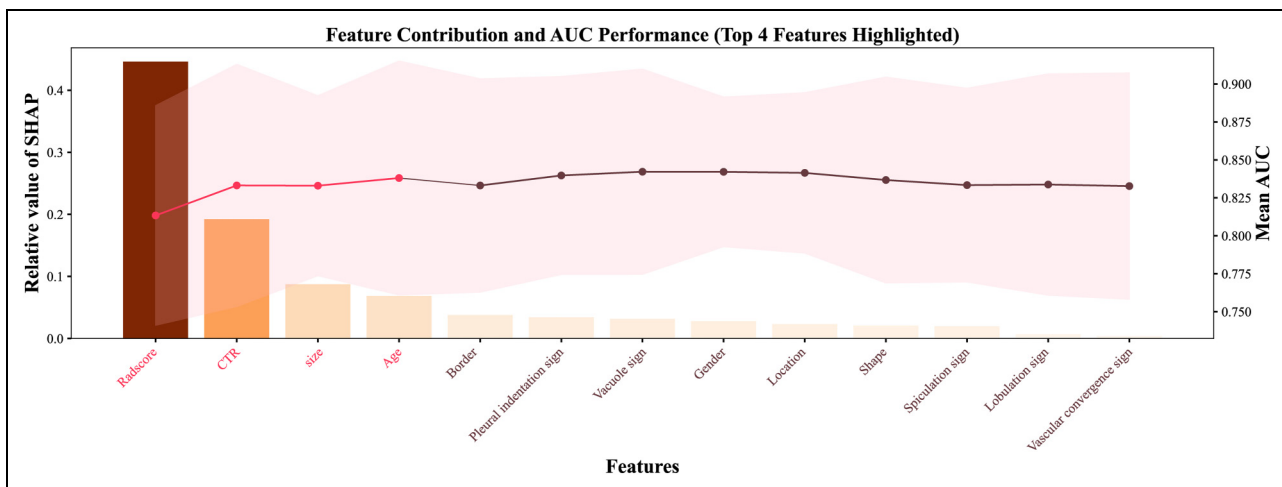


Figure 9. Relative Contribution of Distinct Features to the Model's Predictive Performance, Accompanied by Their Respective AUC Scores.

The stacking classifier effectively combines diverse features from base models like RF, CatBoost, LightGBM, XGBoost, AdaBoost, and GBDT by utilizing a meta-learner. Unlike fixed voting methods, stacking classifier employs a trainable meta-learner, such as logistic regression, to capture nonlinear interactions between base model outputs. This approach is particularly valuable in medical imaging tasks, where interpretability and robustness are crucial. By leveraging the strengths of individual base models while mitigating their weaknesses, the stacking classifier provides a robust framework for integrating multi-modal biomarkers. Consequently, stacked ensemble methods are especially advantageous in this specific context.^{20–22,30}

To further explore the interpretability analysis framework using SHAP,^{32,33} the top four features (Radscore, CTR, size, age) exhibit significantly higher contribution values compared to other variables. The Radscore contributes the most, aligning with its validated role in quantifying tumor heterogeneity, which underlies biological processes such as chaotic angiogenesis and heterogeneous cellular proliferation patterns within solid components.^{15–18} CTR serves as a surrogate marker for tumor invasiveness in lung adenocarcinoma presenting as PSNs,^{15–18,34–36} with higher CTR values reflecting aggressive pathological features. Furthermore, the semiquantitative assessment of CTR, divided into four equal parts, is regarded as a valuable tool for the prognostic evaluation of PSNs.³⁶

Tumor size, defined as the maximum diameter according to the eighth edition of the TNM classification of lung cancer,²³ reflects an infiltrative invasive growth pattern. Additionally, advanced age is associated with tumor invasiveness in lung adenocarcinoma presenting as PSNs, consistent with previous studies.^{15–18} Notably, earlier research has shown that tumor size becomes a more critical determinant of outcomes.^{37,38} This emphasizes the importance of integrating clinical-radiological features (CTR, tumor size, age) with quantifications of tumor heterogeneity for precise risk stratification, highlighting comprehensive integration within the current stacking classifier.

The present study has some limitations. First, the retrospective nature of the study introduced variability in CT scanners and imaging parameters across and within centers, which may have influenced texture feature extraction and the generalizability of the model. Second, while this stacking classifier effectively aids in the preoperative prediction of pathological invasiveness for lung adenocarcinoma presenting as PSNs, training each layer required separate datasets. This increases computational demands and algorithm complexity, and it did not incorporate molecular or genetic biomarkers known to influence tumor invasiveness, thereby limiting its biological comprehensiveness. Finally, the inclusion of only surgically resected cases may introduce selection bias, as nodules managed non-surgically were excluded, and there was a lack of data on longer

postoperative follow-up times, which affects the applicability of results to broader clinical populations.

Conclusion

In conclusion, the stacking ensemble model establishes a new benchmark for preoperative risk assessment in PSNs by synergizing multi-dimensional data streams. Radscore, tumor size, CTR, and age collectively provide complementary insights into morphological features, demographic factors, and tumor heterogeneity. These advances hold significant potential for personalizing therapeutic strategies, particularly in guiding decisions between limited resection and lobectomy for early-stage lung cancer.

Acknowledgments

We would like to extend our sincere appreciation to Editage for their meticulous language editing services, which have significantly enhanced the quality and readability of this manuscript.

ORCID iDs

Gangwen Chen  <https://orcid.org/0009-0007-2875-7311>

Zhichao Zuo  <https://orcid.org/0009-0004-4824-6138>

Compliance with Ethical Requirements

The institutional review boards of all participating institutions approved the study (2021-07-009 from Xiangtan Central Hospital, KY2020147 from the Affiliated Hospital of Southwest Medical University, and 2023YJSLL-121 from the Affiliated Hospital of Guilin Medical University).

Informed Consent Statement

The study protocol was approved by the institutional review board and due to the retrospective nature of the study, the informed consent requirement was waived.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The data for this study are available by contacting the corresponding author upon reasonable request.

Supplemental material

Supplemental material for this article is available online.

References

1. Zuo ZC, Wang LD, Peng K, et al. Development and validation of a nomogram for predicting the 1-, 3-, and 5-year survival in patients with acinar-predominant lung adenocarcinoma. *Curr Med Sci.* 2022;42(6):1178–1185. doi:10.1007/s11596-022-2672-7
2. Barta JA, Farjah F, Thomson CC, et al. The American Cancer Society National Lung Cancer Roundtable strategic plan: Optimizing strategies for lung nodule evaluation and management. *Cancer.* 2024;130(24):4177–4187. doi:10.1002/cncr.35181
3. Binson VA, Subramoniam M, Sunny Y, et al. Prediction of pulmonary diseases with electronic nose using SVM and XGBoost. *IEEE Sensors J.* 2021;21(18):20886–20895. doi:10.1109/JSEN.2021.3100390
4. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: From the Fleischner society 2017. *Radiology.* 2017;284(1):228–243. doi:10.1148/radiol.2017161659
5. Ye T, Deng L, Wang S, et al. Lung adenocarcinomas manifesting as radiological part-solid nodules define a special clinical subtype. *J Thorac Oncol.* 2019;14(4):617–627. doi:10.1016/j.jtho.2018.12.030
6. Liu M, Yang L, Sun X, et al. Evaluation of prognosis in patients with lung adenocarcinoma with atypical solid nodules on thin-section CT images. *Radiol Cardiothorac Imaging.* 2024;6(1):e220234. doi:10.1148/ryct.220234
7. Okamoto T, Miyawaki M, Toyokawa G, et al. Clinical significance of part-solid lung cancer in the eighth edition TNM staging system. *Interact Cardiovasc Thorac Surg.* 2022;34(2):219–226. doi:10.1093/icvts/ivab255
8. Yotsukura M, Asamura H, Motoi N, et al. Long-Term prognosis of patients with resected adenocarcinoma in situ and minimally invasive adenocarcinoma of the lung. *J Thorac Oncol.* Aug 2021;16(8):1312–1320. doi:10.1016/j.jtho.2021.04.007
9. Watanabe Y, Hattori A, Nojiri S, et al. Clinical impact of a small component of ground-glass opacity in solid-dominant clinical stage IA non-small cell lung cancer. *J Thorac Cardiovasc Surg.* 2022;163(3):791–801.e4. doi:10.1016/j.jtcvs.2020.12.089
10. Shi Y, Wu S, Ma S, et al. Comparison between wedge resection and lobectomy/segmentectomy for early-stage non-small cell lung cancer: A bayesian meta-analysis and systematic review. *Ann Surg Oncol.* 2022;29(3):1868–1879. doi:10.1245/s10434-021-10857-7
11. Dolan D, Swanson SJ, Gill R, et al. Survival and recurrence following wedge resection versus lobectomy for early-stage non-small cell lung cancer. *Semin Thorac Cardiovasc Surg.* 2022;34(2):712–723. doi:10.1053/j.semtcvs.2021.04.056
12. Zuo Z, Deng J, Ge W, et al. Quantifying intratumoral heterogeneity within sub-regions to predict high-grade patterns in clinical stage I solid lung adenocarcinoma. *BMC Cancer.* 2025;25(1):51. doi:10.1186/s12885-025-13445-0
13. Zuo Z, Zhang G, Lin S, et al. Radiomics nomogram based on optimal volume of interest derived from high-resolution CT for preoperative prediction of IASLC grading in clinical IA lung adenocarcinomas: A multi-center, large-population study. *Technol Cancer Res Treat.* 2024;23:1-16. doi:10.1177/15330338241300734

14. Tohidinezhad F, Bontempi D, Zhang Z, et al. Computed tomography-based radiomics for the differential diagnosis of pneumonitis in stage IV non-small cell lung cancer patients treated with immune checkpoint inhibitors. *Eur J Cancer*. 2023;183:142–151. doi:10.1016/j.ejca.2023.01.027
15. Wu G, Woodruff HC, Shen J, et al. Diagnosis of invasive lung adenocarcinoma based on chest CT radiomic features of part-solid pulmonary nodules: A multicenter study. *Radiology*. 2020;297(2):E282. doi:10.1148/radiol.202009019
16. Weng Q, Zhou L, Wang H, et al. A radiomics model for determining the invasiveness of solitary pulmonary nodules that manifest as part-solid nodules. *Clin Radiol*. 2019;74(12):933–943. doi:10.1016/j.crad.2019.07.026
17. Luo T, Xu K, Zhang Z, Zhang L, Wu S. Radiomic features from computed tomography to differentiate invasive pulmonary adenocarcinomas from non-invasive pulmonary adenocarcinomas appearing as part-solid ground-glass nodules. *Chin J Cancer Res*. 2019;31(2):329–338. doi:10.21147/j.issn.1000-9604.2019.02.07
18. Zuo Z, Zeng W, Peng K, et al. Development of a novel combined nomogram integrating deep-learning-assisted CT texture and clinical-radiological features to predict the invasiveness of clinical stage IA part-solid lung adenocarcinoma: A multicentre study. *Clin Radiol*. 2023;78(10):e698–e706. doi:10.1016/j.crad.2023.07.002
19. Binson VA, Subramoniam M, Mathew L. Prediction of lung cancer with a sensor array based e-nose system using machine learning methods. *Microsyst Technol*. 2024;30(11):1421–1434. doi:10.1007/s00542-024-05656-5
20. Binson VA, Subramoniam M, Ragesh GK, Kumar A. Early detection of lung cancer through breath analysis using adaboost ensemble learning method. In *2021 2nd International conference on advances in computing, communication, embedded and secure systems (ACCESS), Ernakulam, India, 2021*;183–187.
21. Ren Y, Zhang L, Suganthan PN. Ensemble classification and regression - recent developments, applications and future directions. *IEEE Comput Intell Mag*. 2016;11(1):41–53.
22. Zhang M, Zheng Y, Maidaiti X, et al. Integrating machine learning into statistical methods in disease risk prediction modeling: A systematic review. *Health Data Sci*. 2024;4:0165. doi:10.34133/hds.0165
23. Rami-Porta R, Nishimura KK, Giroux DJ, et al. The international association for the study of lung cancer lung cancer staging project: Proposals for revision of the TNM stage groups in the forthcoming (ninth) edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2024;19(7):1007–1027. doi:10.1016/j.jtho.2024.02.011
24. Suzuki K, Watanabe SI, Wakabayashi M, et al. A single-arm study of sublobar resection for ground-glass opacity dominant peripheral lung cancer. *J Thorac Cardiovasc Surg*. 2022;163(1):289–301.e2. doi:10.1016/j.jtcvs.2020.09.146
25. Aokage K, Saji H, Suzuki K, et al. A non-randomized confirmatory trial of segmentectomy for clinical T1N0 lung cancer with dominant ground glass opacity based on thin-section computed tomography (JCOG1211). *Gen Thorac Cardiovasc Surg*. 2017;65(5):267–272. doi:10.1007/s11748-016-0741-1
26. Shuwen H, Xi Y, Qing Z, et al. Predicting biomarkers from classifier for liver metastasis of colorectal adenocarcinomas using machine learning models. *Cancer Med*. 2020;9(18):6667–6678.
27. Va B, Mathew P, Thomas S, et al. Detection of lung cancer and stages via breath analysis using a self-made electronic nose device. *Expert Rev Mol Diagn*. 2024;24(4):341–353. doi:10.1080/14737159.2024.2316755
28. Jovel J, Greiner R. An Introduction to machine learning approaches for biomedical research. *Front Med (Lausanne)*. 2021;8:771607. doi:10.3389/fmed.2021.771607
29. Binson VA, Subramoniam M, Mathew L. Detection of COPD and lung cancer with electronic nose using ensemble learning methods. *Clin Chim Acta*. 2021;523:231–238. doi:10.1016/j.cca.2021.10.005
30. Peppes N, Daskalakis E, Alexakis T, Adamopoulou E, Demestichas K. Performance of machine learning-based multi-model voting ensemble methods for network threat detection in agriculture 4.0. *Sensors (Basel)*. 2021;21(22):7475. doi:10.3390/s21227475
31. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: Review of methods with applications in diagnostic medicine. *Phys Med Biol*. 2018;63(7):07tr01. doi:10.1088/1361-6560/aab4b1
32. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30:4768–4777.
33. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56–67. doi:10.1038/s42256-019-0138-9
34. Shimomura M, Iwasaki M, Ishihara S, Inoue M. Volume-Based consolidation-to-tumor ratio is a useful predictor for postoperative upstaging in stage I and II lung adenocarcinomas. *Thorac Cardiovasc Surg*. 2022;70(3):265–272. doi:10.1055/s-0039-1694061
35. Sun F, Huang Y, Yang X, et al. Solid component ratio influences prognosis of GGO-featured IA stage invasive lung adenocarcinoma. *Cancer Imaging*. 2020;20(1):87. doi:10.1186/s40644-020-00363-6
36. Jing W, Liu M, Li W, Li D, Wu Y, Lv F. Prognostic implication of consolidation-to-tumor ratio in early lung adenocarcinoma: A retrospective cross-sectional study. *Quant Imaging Med Surg*. 2024;14(5):3366–3380. doi:10.21037/qims-23-1438
37. Li Z, Xu W, Gu T, Cao X, Wu W, Chen L. Tumor size, but not consolidation-to-tumor ratio, is an independent prognostic factor for part-solid clinical T1 non-small cell lung cancer. *Thorac Cancer*. 2023;14(6):602–611. doi:10.1111/1759-7714.14788
38. Zhang S, Lin D, Yu Y, et al. Which will carry more weight when CTR > 0.5, solid component size, CTR, tumor size or SUVmax? *Lung Cancer*. 2022;164:14–22. doi:10.1016/j.lungcan.2021.12.007