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.

A Stacking Fusion Model Integrating MRI Radiomics and Deep Learning for Prediction of Occult Lymph Node Metastasis in Cervical Cancer

Dongqi Cui
First Hospital of Shanxi Medical University
Lingyu Zhang
First Hospital of Shanxi Medical University
Fangmin Tao
First Hospital of Shanxi Medical University
Juntao Zhou
First Hospital of Shanxi Medical University
Xuhui Zhao
Shanxi Provincial People's Hospital
Da Li
First Hospital of Shanxi Medical University
Yanjie Hou
First Hospital of Shanxi Medical University
Xiaofeng Wu
First Hospital of Shanxi Medical University
Lingjie Wang
First Hospital of Shanxi Medical University
Xianfeng Li
lixianfeng-lxf@263.net
First Hospital of Shanxi Medical University

Research Article

Keywords: Cervical cancer, Occult lymph node metastasis, Radiomics, Deep learning, MRI

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11242858/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

To develop and validate a stacking model integrating MRI-based radiomics and deep learning model outputs for the preoperative prediction of occult lymph node metastasis (OLNM) in early-stage cervical cancer (CC), and to explore its potential complementary value in preoperative risk stratification.

Methods

This dual-center retrospective study included 392 patients with early-stage CC. Radiomics features were extracted from preoperative T2WI-FS and DWI sequences, and feature selection was conducted using intraclass correlation coefficient (ICC), Spearman correlation analysis, and least absolute shrinkage and selection operator (LASSO) regression to construct the radiomics model. Deep learning models, including 2D-DL and multichannel 2.5D-DL, were developed separately for T2WI-FS and DWI sequences based on the ResNet50 architecture. A stacking model was subsequently established by integrating the output probabilities of the radiomics, 2D-DL, and multichannel 2.5D-DL models using an ensemble learning strategy. Model performance was evaluated using the AUC, and differences in AUCs between models were compared using the DeLong test. Calibration was assessed using calibration curves.

Results

The incidence of OLNM was 14% in the training cohort and 16% in the external validation cohort. The stacking model achieved the highest AUCs in the training, internal validation, and external validation cohorts (0.983, 0.932, and 0.868, respectively), and was numerically superior to multiple single-sequence models. The calibration curve demonstrated good agreement between the predicted probabilities and actual outcomes, and the decision curve analysis indicated a favorable net clinical benefit.

Conclusions

The stacking model proposed in this study demonstrated good discriminative ability and stability for the preoperative assessment of OLNM in patients with early-stage CC. By integrating outputs from radiomics and deep learning models, this approach may, to some extent, complement conventional imaging assessment and provide additional information for preoperative OLNM risk stratification.

1. Introduction

Cervical cancer (CC) is one of the leading causes of cancer-related deaths among women [1]. According to the latest global cancer statistics, there were approximately 662,000 new cases and 349,000 related deaths in 2022, showing an increasing trend compared with 2020 [2]. Despite advancements in screening programs and the widespread implementation of human papillomavirus (HPV) vaccination, CC continues to pose a significant public health challenge, particularly in low- and middle-income countries [3].

Lymph node metastasis (LNM) has been established as a prognostic factor in CC and is closely associated with local recurrence, distant metastasis, and long-term survival. In 2018, the International Federation of Gynecology and Obstetrics (FIGO) revised the staging system for CC to incorporate imaging-based lymph node assessment into the staging criteria, underscoring the importance of accurate preoperative identification of LNM [4].the National Comprehensive Cancer Network (NCCN)[5] guidelines recommend stratified treatment strategies based on nodal status: For patients with early-stage cervical cancer (stages IB1, IB2, and IIA1), radical hysterectomy with pelvic lymph node dissection is a preferred treatment option, while para-aortic nodal assessment may be selectively performed based on intraoperative nodal findings. In contrast, definitive concurrent chemoradiotherapy rather than surgery is recommended for patients with stage IIIC cervical cancer due to nodal metastasis, potentially reducing the treatment-related morbidity associated with radical surgery followed by postoperative chemoradiotherapy. These recommendations underscore the importance of accurate preoperative lymph node assessment for individualized, guideline-concordant treatment planning.

Occult lymph node metastasis (OLNM) refers to nodal involvement that is not detected on preoperative imaging but is confirmed by postoperative pathological examination [6]. In early-stage CC, approximately 10%–30% of patients are found to have pathologically confirmed pelvic lymph node metastasis after surgery, which may lead to upstaging to FIGO 2018 stage IIIC and require adjuvant chemoradiotherapy. Consistent evidence and guidelines indicate that, for early-stage CC patients at risk of LNM, radical surgery followed by adjuvant therapy does not improve recurrence-free survival compared with definitive chemoradiotherapy, but may increase long-term complications (e.g., lower-limb lymphedema, lymphocele, and pelvic dysfunction) and impose an additional financial burden [7–11]. Accurate preoperative identification of OLNM may facilitate appropriate treatment selection and reduce the likelihood of radical surgery followed by postoperative chemoradiotherapy due to previously undetected nodal metastases, thereby potentially decreasing cumulative treatment-related toxicity and improving quality of life.[7, 10, 12].

Pelvic magnetic resonance imaging (MRI) and abdominopelvic computed tomography (CT) are the most widely used imaging modalities for preoperative staging of CC; however, their diagnostic performance for LNM remains suboptimal. Meta-analyses have reported that MRI yields a sensitivity of 52–75% and a specificity of 75–91% [13, 14], while CT demonstrates a sensitivity of 40–67% and a specificity of 81–92% [13, 15]. The relatively modest sensitivity indicates that a considerable proportion of true-positive lymph node metastases—especially OLNM—are misclassified as N0, leading to inappropriate

management. Although positron emission tomography–computed tomography (PET/CT) exhibits higher overall sensitivity, it performs poorly for micrometastases (< 10 mm), with reported false-negative rates of 10–30% [13, 16], and is limited by cost and availability in many regions.

In response to these limitations, a growing number of studies have investigated the use of MRI- or CT-based radiomics to predict LNM in CC, yielding promising preliminary results [17–19]. However, most existing studies are single-center, involve small sample sizes, and lack a specific focus on OLNM. Identifying radiologically node-negative but pathologically node-positive patients remains a critical unmet need in clinical practice. Therefore, the present study aims to develop and validate an MRI-based stacking model that integrates radiomics and deep learning outputs to accurately identify OLNM in CC patients preoperatively, providing a reliable basis for personalized treatment decision-making.

2. Materials and Methods

2.1. Ethics statement

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Institutional Review Board of the participating institutions (IRB approval number: KYLL-2025-052). Given the retrospective design and the use of fully anonymized data, the requirement for informed consent was waived by the ethics committee.

2.2. Patient population

The patient selection flowchart is shown in Fig. 1. A total of 392 surgically treated patients with pathologically confirmed CC from two tertiary hospitals were retrospectively included. All cases were screened according to predefined inclusion and exclusion criteria. The inclusion criteria were: (1) pathologically confirmed CC; (2) underwent radical hysterectomy with pelvic lymph node dissection, with para-aortic lymph node dissection as needed; (3) abdominopelvic MRI completed within 2 weeks before surgery; (4) preoperative FIGO 2018 stage IB1, IB2, or IIA1; and (5) no evidence of multiple primary malignancies. The exclusion criteria were: (1) receipt of radiotherapy, chemotherapy, or any antitumor treatment before surgery; (2) incomplete clinical or imaging data, including missing key MRI sequences (e.g., axial T2WI-FS or DWI); and (3) poor image quality or severe artifacts compromising image evaluation; (4) patients who did not undergo lymph node dissection or had clear LNM on MRI(short-axis diameter ≥ 10 mm, rounded morphology, irregular nodal shape, indistinct/ill-defined margins, heterogeneous enhancement, or central necrosis). All pathological results were independently reviewed by two senior pathologists, and MRI images were independently assessed by two experienced radiologists. Any discrepancies were resolved by consensus with a third expert in the respective field.

The research workflow is illustrated in Fig. 2. A total of 292 patients from the First Hospital of Shanxi Medical University (Center 1), enrolled between August 2019 and August 2025, were randomly allocated to the training and internal validation cohorts at a 7:3 ratio. The external validation cohort comprised 100 patients from Shanxi Provincial People's Hospital (Center 2), enrolled between January 2021 and August 2025. To ensure inter-center consistency, standardized protocols were implemented for data acquisition and processing.

2.3. Clinical data collection

Clinical baseline characteristics and preoperative MRI data were collected for all enrolled patients. Baseline variables included age, body mass index (BMI), laboratory parameters, squamous cell carcinoma antigen (SCC), carbohydrate antigen 125 (CA125), menopausal status, histologic type, HPV status, and preoperative FIGO stage.

2.4. MRI image processing and segmentation

All patients underwent preoperative abdominopelvic MRI, and detailed scanning parameters are provided in Supplementary Material (1). To minimize resolution variability across scanners, all images were resampled using cubic B-spline interpolation to an isotropic voxel size of 1 × 1 × 1 mm^3. Subsequently, N4 bias field correction was applied using the N4BiasFieldCorrection algorithm in the ANTs toolbox to mitigate intensity inhomogeneity. In accordance with the Image Biomarker Standardization Initiative (IBSI) guidelines, image intensities were normalized using z-score transformation, and gray-level intensities were discretized into 25 bins to ensure cross-center comparability. T2WI-FS images were used as the reference for rigid registration of DWI sequences using the Elastix module in 3D Slicer (version 5.8.0) [20].

Tumor regions of interest (ROIs) were manually delineated slice-by-slice on axial T2WI-FS images using ITK-SNAP (version 3.8.0) [21] by two radiologists, each with over 10 years of experience. Radiologist 1 performed the initial segmentation for all 392 patients. To evaluate segmentation reproducibility, 30 cases were randomly selected, and radiologist 2 independently resegmented the ROIs after a 2-week interval. Interobserver consistency was quantified using the intraclass correlation coefficient (ICC).

2.5. Radiomics analysis and model development

We employed the PyRadiomics module for feature extraction. During feature extraction, wavelet transforms (eight default decompositions) and Laplacian of Gaussian (LoG) filtering (σ values of 1.0, 2.0, and 3.0) were applied. Extracted features included 14 shape-based features, 234 first-order features, 182 gray-level dependence matrix (GLDM) features, 208 gray-level size zone matrix (GLSZM) features, 65 neighborhood gray-level difference matrix (NGTDM) features, 208 gray-level run-length matrix (GLRLM) features, and 286 gray-level co-occurrence matrix (GLCM) features. In total, 1,197 radiomics features were extracted from each ROI for each sequence, yielding 2,394 radiomics features per patient. The feature distribution map is shown in Supplementary Fig. S1.

All data-driven preprocessing and feature selection steps were performed exclusively within the training cohort to minimize information leakage. Radiomics features were standardized using z-score normalization, with normalization parameters derived from the training data and subsequently applied unchanged to the validation cohorts.

Feature selection was conducted using a three-step approach. First, interobserver reproducibility was assessed using intraclass correlation coefficients (ICCs), and features with ICC values greater than 0.80 were retained. Second, redundant features were removed based on pairwise Spearman correlation analysis ($|r| > 0.90$). Third, least absolute shrinkage and selection operator (LASSO) regression was applied, with the regularization parameter determined via stratified five-fold cross-validation, to identify the most informative features.

Based on the selected radiomics features, six machine-learning classifiers were developed, including ExtraTrees, Light Gradient Boosting Machine (LightGBM), Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). To address class imbalance in OLNM, the synthetic minority over-sampling technique (SMOTE) was applied during model training using the training data only and was not applied to the validation cohorts. Model hyperparameters for all classifiers were tuned using grid search within stratified five-fold cross-validation loops conducted exclusively on the training cohort. Using the optimal hyperparameters, each classifier was subsequently retrained on the entire training dataset to obtain the final radiomics-based models.

2.6. 2D-DL model and multichannel 2.5D-DL model development and feature extraction

In this study, both 2D-DL and multichannel 2.5D-DL models were developed for occult lymph node metastasis (OLNM) prediction. For the 2D-DL model, a rectangular bounding box was used to crop the single axial slice containing the largest tumor cross-sectional area, which was subsequently used as a single-channel input. In contrast, the multichannel 2.5D-DL model incorporated limited volumetric contextual information by extracting the axial slice with the largest tumor region together with its adjacent superior and inferior slices, forming a three-channel input representation.

T2WI-FS and DWI sequences were independently fed into both models after z-score intensity normalization. To enhance model robustness and mitigate overfitting, data augmentation strategies—including random flipping, rotation, and center cropping to a fixed resolution of 224 × 224 pixels—were applied during training. Both models were implemented using a pre-trained ResNet50 backbone with transfer learning and were optimized using the Adam optimizer (learning rate = 1×10^{-3}, batch size = 32, maximum epochs = 200, weight decay = 1×10^{-4}). To address class imbalance, class-weighted loss functions were employed during training. Finally, a 2,048-dimensional deep feature vector was extracted from the global average pooling layer of ResNet50, and the softmax layer was used to generate the probability of OLNM. The model structure is shown in Fig. 2.

2.7. Stacking model construction and model assessment

A stacking ensemble strategy was employed to integrate complementary information across different models and MRI sequences. Specifically, the out-of-fold (OOF) predicted probabilities generated by the single-sequence base models within the training cohort were used as input features to train second-level meta-classifiers, which were implemented using the same six machine-learning algorithms as those applied in the radiomics models. The hyperparameters of the meta-classifiers were optimized using grid search conducted exclusively within the training cohort. To prevent information leakage, OOF predictions were generated using stratified five-fold cross-validation conducted exclusively within the training cohort. In each fold, base models were trained on four folds and generated predictions for the remaining hold-out fold. After completing all five folds, the concatenated OOF predicted probabilities for all training samples were used to construct the stacking ensemble. The stacking model was trained entirely within the training cohort and subsequently applied unchanged to the internal validation cohort and the external validation cohort for independent performance evaluation.

2.8. Performance Evaluation

Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Calibration was assessed using calibration curves, and decision curve analysis (DCA) was performed to evaluate net clinical benefit across threshold probabilities. Differences in AUCs between models were compared using the DeLong test.

2.9. Statistical analysis

Statistical analyses were conducted using IBM SPSS Statistics (version 27.0) and Python (version 3.7.12). Normality of continuous variables was assessed using the Shapiro–Wilk test. Normally distributed variables are presented as mean ± standard deviation, and non-normally distributed variables as median (interquartile range, IQR). Categorical variables are presented as frequencies and percentages. For comparisons between two groups (e.g., OLNM-positive vs OLNM-negative), categorical variables were compared using the chi-square test ($n \geq 5$) or Fisher's exact test ($n < 5$), as appropriate. Continuous variables were compared using the independent-samples t test or the Mann–Whitney U test, depending on distribution. For comparisons among three cohorts (training, internal validation, and external validation), categorical variables were compared using the chi-square test ($n \geq 5$) or Fisher's exact test ($n < 5$), as appropriate. Normally distributed continuous variables were compared using one-way analysis of variance (ANOVA), and non-normally distributed continuous variables were compared using the Kruskal–Wallis H test.

3.Results

3.1.Patient characteristics

A total of 392 patients with surgically and pathologically confirmed early-stage CC were enrolled in this study. From Center 1, 292 patients were randomly allocated to the training cohort (n = 205) and the internal validation cohort (n = 87) in a 7:3 ratio, while 100 patients from Center 2 constituted the external

validation cohort. Based on postoperative pathology, patients were categorized into the OLNM-negative and OLNM-positive groups. Specifically, the training cohort comprised 176 (85.85%) OLNM-negative and 29 (14.15%) OLNM-positive patients; the internal validation cohort included 75 (86.21%) OLNM-negative and 12 (13.79%) OLNM-positive patients; and the external validation cohort consisted of 84 (84.00%) OLNM-negative and 16 (16.00%) OLNM-positive patients. Significant differences were observed in age, SCC level, and menopausal status across the three cohorts ($P < 0.05$), whereas no significant differences were found in the remaining variables, supporting the overall comparability of the cohorts (Table 1). In the training cohort, no significant differences were observed in clinical characteristics between the OLNM-positive and OLNM-negative groups (all $P > 0.05$), except for FIGO stage (Table 2).

3.2. Radiomics model development and validation

Radiomics features were extracted from ROIs delineated on axial T2WI-FS and DWI sequences. Following a multi-step feature selection procedure, 12 features from T2WI-FS and 7 features from DWI were retained for model development (Figs. 3 and 4). Among the six evaluated machine-learning classifiers, LR demonstrated the best predictive performance and was therefore selected as the final classifier. The T2WI-FS-based model achieved AUCs of 0.898, 0.873, and 0.728 in the training, internal validation, and external validation cohorts, respectively, whereas the DWI-based model achieved AUCs of 0.851, 0.796, and 0.797 in the corresponding cohorts (Table 3). Detailed performance metrics are provided in Supplementary Fig. S2 and S3.

3.3.2D-DL model and multichannel 2.5D-DL model development and validation

In this study, 2D-DL and multichannel 2.5D-DL models were developed separately using axial T2WI-FS and DWI sequences. Both models used ResNet50 as the backbone network. A 2048-dimensional deep feature vector was extracted from the global average pooling layer, and the final probability of OLNM was obtained using a softmax layer. For the T2WI-FS-based 2D-DL model, the AUCs in the training, internal validation, and external validation cohorts were 0.784, 0.716, and 0.758, respectively. The DWI-based 2D-DL model yielded AUCs of 0.722, 0.673, and 0.618 in the corresponding cohorts. For the T2WI-FS-based multichannel 2.5D-DL model, the AUCs were 0.934, 0.827, and 0.789 in the training, internal validation, and external validation cohorts, respectively. The DWI-based 2.5D-DL model achieved AUCs of 0.838, 0.726, and 0.713 in the corresponding cohorts (Table 3).Detailed performance metrics are provided in Supplementary Fig. S2.

3.4.Combined model construction and validation

To further improve predictive performance, the output probabilities from the radiomics model, the 2D-DL model, and the multichannel 2.5D-DL model were used as input features to train six classical machine-learning classifiers as candidate meta-classifiers for the stacking model. LR showed the best overall performance and was therefore selected as the final meta-classifier (Supplementary Fig. S3). The stacking model achieved AUCs of 0.983, 0.932, and 0.868 in the training, internal validation, and external validation cohorts (Table 3), respectively, which were the highest among the evaluated models. The calibration curve showed good agreement between predicted probabilities and observed outcomes, and DCA suggested that the model could provide net clinical benefit across a clinically relevant range of threshold probabilities (Fig. 5). Overall, the stacking model maintained good discrimination across all cohorts, with favorable calibration and potential clinical benefit.The DeLong test showed that, in the training cohort, the fusion model achieved a significantly higher AUC than all single-sequence Rad, 2D-DL, and multichannel 2.5D-DL models (all $P < 0.05$). In the internal validation cohort, the AUC of the fusion model was significantly higher than those of the DWI-based Rad model, DWI-based 2D-DL model, DWI-based multichannel 2.5D-DL model, and T2WI-FS-based 2D-DL model (all $P < 0.05$). In the external validation cohort, the fusion model also achieved a significantly higher AUC than the DWI-based 2D-DL model and DWI-based multichannel 2.5D-DL model (all $P < 0.05$). Detailed results are presented in Supplementary Fig. S4.

4.Discussion

In this dual-center study of 392 patients with early-stage CC, we developed and validated a stacking model that integrates the output probabilities from a radiomics model, a 2D deep learning (2D-DL) model, and a multichannel 2.5D deep learning (2.5D-DL) model to preoperatively predict OLNM. The stacking model showed excellent discrimination, with AUCs of 0.983, 0.932, and 0.868 in the training, internal validation, and external validation cohorts, respectively, and achieved the highest AUCs among the evaluated models. These results suggest that the stacking model may be a useful tool for preoperative OLNM risk stratification and could support clinical decision-making.

The clinical characteristics of the study cohort were broadly consistent with those reported in large surgical series. The distribution of FIGO stages was predominantly IB2 (47.19%), followed by IB1 (31.38%) and IIA1 (21.43%) [22]. Squamous cell carcinoma accounted for most cases (89.80%), consistent with the reported prevalence of 82%–88% in the literature [22, 23]. The HPV infection rate was also comparable to prior studies, supporting the clinical representativeness of the cohort [24].

Regarding baseline characteristics in the training cohort, no significant differences were observed between the OLNM-positive and OLNM-negative groups in age, menopausal status, BMI, histopathological subtype, or tumor marker levels, except for FIGO stage. This is in line with previous evidence suggesting that, in early-stage CC, nodal status may not be strongly reflected by routine clinical variables [18, 25, 26].

Currently, preoperative assessment of LNM in CC mainly relies on abdominopelvic CT and MRI. Most guidelines and studies use a short-axis diameter ≥ 10 mm as the primary criterion for metastatic lymph nodes, with additional morphological features such as a round shape, indistinct margins, heterogeneous enhancement, and central necrosis [14, 27, 28]. However, OLNM often lacks these typical imaging features: involved nodes are usually small, without obvious necrosis or abnormal enhancement, and may appear imaging-negative, making accurate detection with conventional criteria particularly challenging. In recent years, radiomics has been shown to capture microstructural and phenotypic alterations that are not visually appreciable on routine

imaging and has demonstrated improved performance for predicting LNM in CC [18, 29–31]. Nevertheless, most prior studies have focused on overall LNM, with relatively limited dedicated evaluation of OLNM, a clinically high-risk subgroup.

In this study, we selected 12 and 7 radiomics features with predictive value from the MRI T2WI-FS and DWI sequences, respectively. In the T2WI-FS sequence, the most contributory features were GLSZM texture features (e.g., wavelet_LLH_glszm_GrayLevelNonUniformity and wavelet_HHL_glszm_LargeAreaHighGrayLevelEmphasis) and shape features (e.g., original_shape_LeastAxisLength). These features reflect non-uniformity in intratumoral gray-level distribution, aggregation of large high-intensity areas, and geometric complexity of the lesion [32, 33], which may indicate more pronounced heterogeneity in cellular density, stromal remodeling, and potentially increased invasive and proliferative activity [34, 35]. In the DWI sequence, key features were primarily derived from GLSZM and GLRLM texture features (e.g., wavelet_LLH_glszm_LargeAreaHighGrayLevelEmphasis and wavelet_LLH_glrlm_RunLengthNonUniformity). These features characterize signal-intensity differences and spatial distribution complexity within diffusion-restricted regions [32, 36] and may be related to deeper histological patterns of spatial organization[36].

In MRI-based radiomics studies by Shi et al. [37] and Wang et al. [18] that developed Rad models to predict LNM in cervical cancer, the selected radiomic features were predominantly GLCM features, which mainly characterize local texture variations. In contrast, our model was dominated by GLSZM features, which quantify large-scale heterogeneity[32]. This shift in feature importance may be clinically relevant: since OLNM is often radiologically negative, its microstructural changes might manifest as larger-scale tissue disorganization rather than localized texture changes. Notably, glszm_LargeAreaHighGrayLevelEmphasis was a high-weight feature in both T2WI-FS and DWI, suggesting that OLNM-positive tumors may exhibit more prominent aggregation of large high gray-level regions. This observation warrants further validation using spatial imaging–pathology correlation in future work. In addition, most high-contribution features in our study were extracted from wavelet-transformed images. Wavelet transforms can capture subtle texture changes and deeper structural heterogeneity across multiple scales and frequencies [38], and may therefore be more sensitive to microscopic abnormalities associated with OLNM that are difficult to perceive visually on routine imaging.

The single-sequence radiomics models constructed using these features showed good predictive performance. The T2WI-FS radiomics model achieved AUCs of 0.898 and 0.728 in the training and external validation cohorts, respectively. The DWI model achieved AUCs of 0.851 and 0.797 in the corresponding cohorts, suggesting acceptable cross-center stability and potential for clinical translation. However, radiomics still relies largely on handcrafted feature engineering [39], which may not fully capture higher-order spatial structure and advanced phenotypic patterns in tumor imaging.

Deep learning can automatically learn latent spatial textures and biological heterogeneity from medical images [40, 41], but its performance depends on network architecture and sample size [42, 43]. Previous studies have shown that 3D CNNs, due to their large number of parameters and high-dimensional inputs, are more sensitive to training sample size than 2D or multichannel 2.5D CNNs and are more prone to overfitting and unstable performance in small-sample medical imaging tasks [42, 44, 45]. Multiple studies indicate that multichannel 2.5D deep learning may offer better stability and generalizability under limited-sample settings [45, 46]. To date, no studies have evaluated 2D-DL or multichannel 2.5D-DL models specifically for OLNM prediction in CC. In this study, we developed both architectures. The 2.5D-DL model stacks adjacent slices along the channel dimension, preserving 3D contextual information while limiting model complexity. Combined with transfer learning and class weighting, this architecture enhances minority-class recognition and improves training stability. Across all cohorts and both sequences, the multichannel 2.5D-DL models consistently achieved higher AUCs than the corresponding 2D-DL models. For T2WI-FS, the 2.5D-DL model reached AUCs of 0.934 (training) and 0.789 (external validation), versus 0.784 and 0.758 for the 2D-DL model. For DWI, the 2.5D-DL model achieved AUCs of 0.838 and 0.713, compared with 0.722 and 0.618 for the 2D-DL model.Although the multichannel 2.5D-DL models yielded numerically higher AUCs than the corresponding 2D-DL models for both imaging sequences, most pairwise DeLong comparisons across the three datasets did not reach statistical significance. This may be attributable to the limited sample size, class imbalance, and differences in tumor characteristics. These findings are consistent with those of Niu et al. [46], who showed that a 2.5D MRI transformer–based deep learning model combined with radiomics achieved higher AUCs than a 2D model for predicting IDH-wildtype, TERT promoter–mutant gliomas.

To further improve model performance, we integrated the output scores from the single-sequence radiomics, 2D-DL, and multichannel 2.5D-DL models and performed ensemble fusion using six classical machine-learning classifiers. After comparing different classifiers (Supplementary Fig. S3), logistic regression (LR) showed relatively stable performance across fusion settings and better generalizability in the external validation cohort. Therefore, we selected LR as the meta-classifier for the stacking model and ensembled the outputs of the radiomics, 2D-DL, and multichannel 2.5D-DL models. This approach maximized the use of complementary information across sequences and model types [47–49], and achieved an AUC of 0.868 in the external validation cohort.However, its external validation performance remained lower than that in the training cohort (AUC, 0.868 vs 0.983), possibly reflecting inter-center differences in MRI scanners and acquisition parameters that may induce feature distribution shifts. Future studies should investigate harmonization methods, such as ComBat[50, 51], to mitigate site- and scanner-related batch effects and improve cross-center generalizability.

The fusion model demonstrated good discriminative ability for OLNM risk stratification in patients considered lymph node–negative on preoperative MRI and may help identify those at increased risk of occult nodal metastasis. A high-risk prediction could therefore serve as an adjunctive indicator of potential OLNM. When the model indicates an elevated risk, multidisciplinary evaluation may be considered to further refine staging, and to support more individualized treatment planning. Such an approach may help achieve a more appropriate balance between oncologic control and treatment-related morbidity while potentially improving quality of life (Fig. 6).

Several limitations should be acknowledged. First, this was a retrospective study with a relatively limited sample size; therefore, the robustness and clinical utility of the model require further validation in prospective, multicenter cohorts. Second, only the primary cervical tumors were manually segmented, whereas lymph node regions were not directly analyzed. The model therefore predicts OLNM indirectly by capturing imaging characteristics of the primary tumor that may be associated with occult nodal spread, which may limit its predictive scope. Manual segmentation may also introduce

observer-dependent variability. Moreover, the present model classified only the presence or absence of lymph node metastasis and did not localize the specific nodal regions involved. Future studies integrating imaging–pathology correlation of lymph nodes and molecular analyses may help clarify the underlying biological mechanisms, improve model interpretability, and facilitate clinical translation.

5.Conclusion

In this dual-center study, we developed and externally validated an MRI-based fusion model integrating radiomics and deep learning outputs for the preoperative prediction of OLNM in early-stage cervical cancer. The model demonstrated good discrimination across all cohorts and benefited from complementary information across imaging sequences and model types. Further validation in larger prospective cohorts and integration of imaging–pathology lymph node correspondence are needed to support more precise regional prediction. The proposed model shows potential for assisting individualized clinical decision-making and provides a basis for further investigation.

Abbreviations

OLNM Occult lymph node metastasis

CC Cervical cancer

ICC Intraclass correlation coefficient

LASSO Least absolute shrinkage and selection operator

AUC Area under the receiver operating characteristic curve

HPV Human papillomavirus

LNM Lymph node metastasis

FIGO International Federation of Gynecology and Obstetrics

NCCN National Comprehensive Cancer Network

CCRT Concurrent chemoradiotherapy

MRI Magnetic resonance imaging

CT Computed tomography

PET/CT Positron emission tomography–computed tomography

BMI Body mass index

SCC Squamous cell carcinoma

CA125 Carbohydrate antigen 125

IBSI Image Biomarker Standardization Initiative

ROIs Regions of interest

LightGBM Light Gradient Boosting Machine

LR Logistic Regression

RF Random Forest

SVM Support Vector Machine

XGBoost Extreme Gradient Boosting

SMOTE Synthetic minority over-sampling technique

OOF Out-of-fold

PPV Positive predictive value

NPV Negative predictive value

DCA Decision curve analysis

ANOVA One-way analysis of variance

DL Deep learning

Declarations

Ethical Approval and Consent to participate This retrospective study was approved by the Ethics Review Board of the First Hospital of Shanxi Medical University (IRB approval number: KYLL-2025-052); the requirement for written informed consent was waived.

Consent for publication Not applicable.

Availability of data and materials The data that support the findings of this study are not publicly available due to privacy or ethical restrictions, but are available on request from the corresponding author.

Competing interests The authors declare no competing interests.

Funding This work was supported by the Shanxi Province Basic Research Program (Grant No. 20210302124667).

Author contributions D.Q.C. conceptualized the study, developed the methodology, performed the formal analysis, conducted the investigation, curated the data, and prepared the original draft of the manuscript. L.Y.Z. curated the data, performed the formal analysis, and contributed to visualization. F.M.T. conducted the investigation and reviewed and edited the manuscript. J.T.Z., X.H.Z., D.L., Y.J.H., and X.F.W. contributed to investigation. X.F.L. and L.J.W. contributed to conceptualization and supervision, and critically reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements Not applicable.

References

1. Filho AM, Laversanne M, Ferlay J, Colombet M, Piñeros M, Znaor A, Parkin DM, Soerjomataram I, Bray F. The GLOBOCAN 2022 cancer estimates: Data sources, methods, and a snapshot of the cancer burden worldwide. Int J Cancer. 2025;156:1336–46.
2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians 2024, 74:229–63.
3. Li Y, Song W, Gao P, Guan X, Wang B, Zhang L, Yao Y, Guo Y, Wang Y, Jiang S, Sun S. Global, regional, and national burden of breast, cervical, uterine, and ovarian cancer and their risk factors among women from 1990 to 2021, and projections to 2050: findings from the global burden of disease study 2021. BMC Cancer. 2025;25:330.
4. Bhatla N, Berek JS, Cuello Fredes M, Denny LA, Grenman S, Karunaratne K, Kehoe ST, Konishi I, Olawaiye AB, Prat J, Sankaranarayanan R. Revised FIGO staging for carcinoma of the cervix uteri. Int J Gynecol Obstet. 2019;145:129–35.
5. NCCN Clinical Practice Guidelines in Oncology. Cervical Cancer [Available at: https://www.nccn.org/guidelines (accessed December 25, 2025)].
6. Wang W, Liang H, Zhang Z, Xu C, Wei D, Li W, Qian Y, Zhang L, Liu J, Lei D. Comparing three-dimensional and two-dimensional deep-learning, radiomics, and fusion models for predicting occult lymph node metastasis in laryngeal squamous cell carcinoma based on CT imaging: a multicentre, retrospective, diagnostic study. EClinicalMedicine. 2024;67:102385.
7. Machida H, Matsuo K, Furusawa A, Kita T, Kitagawa R, Mikami M. Profile of treatment-related complications in women with clinical stage IB-IIB cervical cancer: A nationwide cohort study in Japan. PLoS ONE. 2019;14:e0210125.
8. Cibula D, Dostalek L, Hillemanns P, Scambia G, Jarkovsky J, Persson J, Raspagliesi F, Novak Z, Jaeger A, Capilna ME, et al. Completion of radical hysterectomy does not improve survival of patients with cervical cancer and intraoperatively detected lymph node involvement: ABRAX international retrospective cohort study. Eur J Cancer. 2021;143:88–100.
9. Zhong ML, Liu YC, Yang JT, Wang YN, Ao MH, Xiao Y, Zeng SY, Liang MR. Treatment for locally resectable stage IIIC1r cervical cancer: surgery or chemoradiotherapy? BMC Cancer. 2024;24:217.
10. Li YX, Cao SY, Fan Y, Zhang YF, Li JK. Survival outcomes in IIIC cervical cancer by treatment strategies: a systematic review and meta-analysis. BMC Cancer. 2025;25:1340.
11. Derks M, Groenman FA, van Lonkhuijzen L, Schut PC, van der Westerveld H, Kenter GG. Completing or Abandoning Radical Hysterectomy in Early-Stage Lymph Node-Positive Cervical Cancer: Impact on Disease-Free Survival and Treatment-Related Toxicity. Int J Gynecol Cancer. 2017;27:1015–20.
12. Cibula D, Raspollini MR, Planchamp F, Centeno C, Chargari C, Felix A, Fischerova D, Jahnn-Kuch D, Joly F, Kohler C, et al. ESGO/ESTRO/ESP Guidelines for the management of patients with cervical cancer - Update 2023. Virchows Arch. 2023;482:935–66.
13. Olthof EP, Bergink-Voorthuis BJ, Wenzel HHB, van der Mongula J, Spijkerboer AM, Adam JA, Bekkers RLM, Beltman JJ, Slangen BFM, et al. Diagnostic accuracy of MRI, CT, and [(18)F]FDG-PET-CT in detecting lymph node metastases in clinically early-stage cervical cancer - a nationwide Dutch cohort study. Insights Imaging. 2024;15:36.
14. Chen J, Kitzing YX, Lo G. Systematic Review-Role of MRI in Cervical Cancer Staging. Cancers (Basel) 2024, 16.
15. Jung W, Park KR, Lee KJ, Kim K, Lee J, Jeong S, Kim YJ, Kim J, Yoon HJ, Kang BC, et al. Value of imaging study in predicting pelvic lymph node metastases of uterine cervical cancer. Radiat Oncol J. 2017;35:340–8.

16. Haldorsen IS, Lura N, Blaakaer J, Fischerova D, Werner HMJ. What Is the Role of Imaging at Primary Diagnostic Work-Up in Uterine Cervical Cancer? Curr Oncol Rep. 2019;21:77.
17. Monthatip K, Boonnag C, Muangmool T, Charoenkwan K. A machine learning-based prediction model of pelvic lymph node metastasis in women with early-stage cervical cancer. J Gynecol Oncol. 2024;35:e17.
18. Wang T, Li YY, Ma NN, Wang PA, Zhang B. A MRI radiomics-based model for prediction of pelvic lymph node metastasis in cervical cancer. World J Surg Oncol. 2024;22:55.
19. Zhang Z, Wan X, Lei X, Wu Y, Zhang J, Ai Y, Yu B, Liu X, Jin J, Xie C, Jin X. Intra- and peri-tumoral MRI radiomics features for preoperative lymph node metastasis prediction in early-stage cervical cancer. Insights Imaging. 2023;14:65.
20. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging. 2012;30:1323–41.
21. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage. 2006;31:1116–28.
22. Chen X, Duan H, Liu P, Lin L, Ni Y, Li D, Dai E, Zhan X, Li P, Huo Z, et al. Development and validation of a prognostic nomogram for 2018 FIGO stages IB1, IB2, and IIA1 cervical cancer: a large multicenter study. Ann Transl Med. 2022;10:121.
23. Wang M, Huang K, Wong MCS, Huang J, Jin Y, Zheng ZJ. Global Cervical Cancer Incidence by Histological Subtype and Implications for Screening Methods. J Epidemiol Glob Health. 2024;14:94–101.
24. Xu X, Feng T, Li D, Lou H, Lan H. Prevalent distribution and survival outcome of HPV infection in patients with early-stage cervical cancer in Hangzhou, China. BMC Infect Dis. 2022;22:941.
25. Qin F, Sun X, Tian M, Jin S, Yu J, Song J, Wen F, Xu H, Yu T, Dong Y. Prediction of lymph node metastasis in operable cervical cancer using clinical parameters and deep learning with MRI data: a multicentre study. Insights Imaging. 2024;15:56.
26. Isikhuemen M, Ezeanochie M, Okonkwo C, Okonofua F. Squamous cell carcinoma antigen levels in women with cervical cancer and those without cervical cancer: a cross-sectional study among women of African descent. Eur J Med Res. 2025;30:573.
27. Lakhman Y, Aherne EA, Jayaprakasam VS, Nougaret S, Reinhold C. Staging of Cervical Cancer: A Practical Approach Using MRI and FDG PET. AJR Am J Roentgenol. 2023;221:633–48.
28. Manganaro L, Lakhman Y, Bharwani N, Gui B, Gigli S, Vinci V, Rizzo S, Kido A, Cunha TM, Sala E, et al. Correction to: Staging, recurrence and follow-up of uterine cervical cancer using MRI: Updated Guidelines of the European Society of Urogenital Radiology after revised FIGO staging 2018. Eur Radiol. 2022;32:738.
29. Luo S, Guo Y, Ye Y, Mu Q, Huang W, Tang G. Prediction of cervical cancer lymph node metastasis based on multisequence magnetic resonance imaging radiomics and deep learning features: a dual-center study. Sci Rep. 2025;15:29259.
30. Dong S, Peng YQ, Feng YN, Li XY, Gong LP, Zhang S, Du XS, Sun LT. Based on 3D-PDU and clinical characteristics nomogram for prediction of lymph node metastasis and lymph-vascular space invasion of early cervical cancer preoperatively. BMC Womens Health. 2024;24:438.
31. Xia X, Li D, Du W, Wang Y, Nie S, Tan Q, Gou Q. Radiomics Based on Nomogram Predict Pelvic Lymphnode Metastasis in Early-Stage Cervical Cancer. Diagnostics (Basel) 2022, 12.
32. Mayerhoefer ME, Materka A, Langs G, Haggstrom I, Szczypinski P, Gibbs P, Cook G. Introduction to Radiomics. J Nucl Med. 2020;61:488–95.
33. Riberdy V, Guida A, Rioux J, Brewer K. Radiomics in preclinical imaging research: methods, challenges and opportunities. Npj Imaging. 2025;3:45.
34. Toyama Y, Hotta M, Motoi F, Takanami K, Minamimoto R, Takase K. Prognostic value of FDG-PET radiomics with machine learning in pancreatic cancer. Sci Rep. 2020;10:17024.
35. Lee SK, Joo MW, Kim JY, Kim M. MRI Radiomics for Predicting the Diffuse Type of Tenosynovial Giant Cell Tumor: An Exploratory Study. Diagnostics (Basel) 2025, 15.
36. Wang JH, Wahid KA, van Dijk LV, Farahani K, Thompson RF, Fuller CD. Radiomic biomarkers of tumor immune biology and immunotherapy response. Clin Transl Radiat Oncol. 2021;28:97–115.
37. Shi J, Dong Y, Jiang W, Qin F, Wang X, Cui L, Liu Y, Jin Y, Luo Y, Jiang X. MRI-based peritumoral radiomics analysis for preoperative prediction of lymph node metastasis in early-stage cervical cancer: A multi-center study. Magn Reson Imaging. 2022;88:1–8.
38. Song H, Yang S, Yu B, Li N, Huang Y, Sun R, Wang B, Nie P, Hou F, Huang C, et al. CT-based deep learning radiomics nomogram for the prediction of pathological grade in bladder cancer: a multicenter study. Cancer Imaging. 2023;23:89.
39. Rundo L, Militello C. Image biomarkers and explainable AI: handcrafted features versus deep learned features. Eur Radiol Exp. 2024;8:130.
40. Klyuzhin IS, Xu Y, Ortiz A, Ferres JL, Hamarneh G, Rahmim A. Testing the Ability of Convolutional Neural Networks to Learn Radiomic Features. Computer Methods and Programs in Biomedicine; 2022. p. 219.
41. Nyman J, Denize T, Bakouny Z, Labaki C, Titchen BM, Bi K, Hari SN, Rosenthal J, Mehta N, Jiang B, et al. Spatially aware deep learning reveals tumor heterogeneity patterns that encode distinct kidney cancer states. Cell Rep Med. 2023;4:101189.
42. Huang CC, Chiang HF, Hsieh CC, Zhu BR, Wu WJ, Shaw JS. Impact of Dataset Size on 3D CNN Performance in Intracranial Hemorrhage Classification. Diagnostics (Basel) 2025, 15.
43. Piffer S, Ubaldi L, Tangaro S, Retico A, Talamonti C. Tackling the small data problem in medical image classification with artificial intelligence: a systematic review. Prog Biomed Eng (Bristol) 2024, 6.

44. Cubero L, Castelli J, Simon A, de Crevoisier R, Acosta O, Pascau J. Deep Learning-Based Segmentation of Head and Neck Organs-at-Risk with Clinical Partially Labeled Data. Entropy (Basel) 2022, 24.
45. Garg A, Garg A, Duncan D. 2.5 CNN: Leveraging 2D CNNs to Pretrain 3D Models in Low-Data Regimes for COVID-19 Diagnosis. Electronics 2025, 14.
46. Niu W, Yan J, Hao M, Zhang Y, Li T, Liu C, Li Q, Liu Z, Su Y, Peng B, et al. MRI transformer deep learning and radiomics for predicting IDH wild type TERT promoter mutant gliomas. NPJ Precis Oncol. 2025;9:89.
47. Zeng Q, Jia F, Tang S, He H, Fu Y, Wang X, Zhang J, Tan Z, Tang H, Wang J, et al. Ensemble learning-based radiomics model for discriminating brain metastasis from glioblastoma. Eur J Radiol. 2025;183:111900.
48. Wang J, Hu F, Li J, Lv W, Liu Z, Wang L. Comparative performance of multiple ensemble learning models for preoperative prediction of tumor deposits in rectal cancer based on MR imaging. Sci Rep. 2025;15:4848.
49. Tandel GS, Tiwari A, Kakde OG, Gupta N, Saba L, Suri JS. Role of Ensemble Deep Learning for Brain Tumor Classification in Multiple Magnetic Resonance Imaging Sequence Data. Diagnostics 2023, 13.
50. Pei W, Wei Y, Lai P, Mao L, Huang X, Su D, Liao H, Lu S. Impact of ComBat harmonization on MRI radiomic-analysis of nasopharyngeal carcinoma: a dual-center study. Front Oncol. 2026;16:1705567.
51. Khajetash B, Hajianfar G, Talebi A, Mahdavi SR, Ghavidel B, Kalati FA, Molana SH, Lei Y, Tavakoli M. Impact of harmonization on predicting complications in head and neck cancer after radiotherapy using MRI radiomics and machine learning techniques. Med Phys. 2025;52:5091–103.

Tables

Table 1 Baseline characteristics of the study population.

Characteristics	Training cohort (n =205)	Internal validation cohort (n=87)	External validation cohort (n =100)	*P* value
Age(years)[a]	53(43-59)	51(43-58)	54.5(48-60)	0.032*
BMI(kg/m^2)[b]	23.89±3.26	24.38±3.00	23.79±3.20	0.394
RBC(×10^{12}/L)[b]	4.32±0.38	4.27±0.38	4.36±0.52	0.360
HGB(g/L)[b]	131.94±10.53	131.00±10.41	131.45±11.43	0.784
WBC(×10^9/L)[a]	5.63(4.63-6.71)	5.52(4.70-6.70)	6.11(4.90-7.61)	0.077
NEU(×10^9/L)[a]	3.30(2.45-4.20)	3.19(2.50-4.10)	3.57(2.70-4.92)	0.137
LYM(×10^9/L)[a]	1.81(1.49-2.20)	1.81(1.50-2.23)	1.65(1.31-2.06)	0.050
PLT(×10^9/L)[b]	235.55±59.23	250.26±71.37	250.94±69.69	0.072
RDW_SD(fL)[a]	42.70(41.10-44.75)	42.90(41.10-44.90)	42.75(40.03-45.48)	0.836
SCC(ng/mL)[a]	1.23(0.76-2.39)	1.23(0.72-2.19)	1.73(1.15-3.48)	<0.001*
CA125(U/mL)[a]	12.16(8.87-15.60)	11.06(7.74-15.27)	11.5(8.87-16.98)	0.657
Menopausal status[c]				0.005*
Premenopausal	101(49.27%)	44(50.57%)	31(31.00%)	
Postmenopausal	104(50.73%)	43(49.43%)	69(69.00%)	
Histologic type[c]				0.760
Squamous cell carcinoma	186(90.73%)	78(89.66%)	88(88.00%)	
Others	19(9.27%)	9(10.34%)	12(12.00%)	
HPV status[c]				0.518
Positive	187(91.22%)	76(87.36%)	88(88.00%)	
Negative	18(8.78%)	11(12.64%)	12(12.00%)	
FIGO stage[c]				0.616
IB1	68(33.17%)	30(34.48%)	25(25.00%)	
IB2	94(45.85%)	40(45.98%)	51(51.00%)	
IIA1	43(20.98%)	17(19.54%)	24(24.00%)	

Abbreviations: BMI, body mass index; RBC, red blood cell count; HGB, hemoglobin; WBC, white blood cell count; NEU, neutrophil count; LYM, lymphocyte count; PLT, platelet count; RDW_SD, red blood cell distribution width (standard deviation); SCC, squamous cell carcinoma antigen; CA125, carbohydrate antigen 125; HPV, human papillomavirus; FIGO, International Federation of Gynecology and Obstetrics 2018 staging system.

[a] Nonnormally distributed variable was shown as the Median (interquartile range).[b] Normally distributed variable was shown as Mean ± standard deviation.[c] Categorical variable was shown as n (%).

Note: Comparisons among groups were performed using one-way ANOVA or the Kruskal–Wallis H test for continuous variables, and the chi-square test (n≥5) or Fisher's exact test (n<5) for categorical variables, as appropriate.

*$P < 0.05$

Table 2 Comparisons of clinical characteristics for different lymph node statuses in three cohorts.

Characteristics	Training cohort (n =205)			Internal validation cohort (n=87)			External validation cohort (n =100)		
	OLNM-positive	OLNM-negative	*P* value	OLNM-positive	OLNM-negative	*P* value	OLNM-positive	OLNM-negative	*P* value
Age(years)[a]	49(42-56.5)	53(43.25-60)	0.226	43(40.5-51)	53(44-58)	0.058	53(45.25-55.75)	55(48-62.75)	0.119
BMI(kg/m^2)[b]	23.79±3.06	23.91±3.30	0.864	24.36±2.29	24.38±3.11	0.988	22.73±2.06	23.99±3.34	0.149
RBC(×10^{12}/L)[b]	4.38±0.30	4.31±0.39	0.346	4.17±0.37	4.29±0.39	0.316	4.37±0.45	4.36±0.53	0.929
HGB(g/L)[b]	132.17±9.42	131.90±10.72	0.897	122.67±6.76	132.34±10.30	0.002*	129.75±13.59	131.77±11.04	0.675
WBC(×10^9/L)[a]	6.1(5.20-7.05)	5.6(4.61-6.63)	0.130	5.59(4.58-7.60)	5.52(4.7-6.7)	0.796	5.97(4.98-7.76)	6.11(4.9-7.44)	0.821
NEU(×10^9/L)[a]	3.7(2.84-4.3)	3.2(2.4-4.2)	0.100	3.68(3.11-5.01)	3.15(2.46-4)	0.047*	3.72(2.60-4.92)	3.56(2.7-4.92)	0.892
LYM(×10^9/L)[a]	1.9(1.48-2.16)	1.8(1.5-2.2)	0.787	1.57(1.27-1.81)	1.9(1.55-2.32)	0.013*	1.79(1.55-2.31)	1.62(1.24-2)	0.168
PLT(×10^9/L)[b]	236.59±64.64	235.38±58.49	0.919	267.83±74.06	247.45±71.03	0.346	238.12±78.61	253.38±68.11	0.425
RDW_SD(fL)[a]	42.8(41.15-45.65)	42.5(41.1-44.6)	0.282	43.7(40.8-45)	42.9(41.2-44.9)	0.961	42.8(39.7-45.6)	42.75(40.15-45.33)	0.888
SCC(ng/mL)[a]	1.29(0.51-1.92)	1.22(0.763-2.53)	0.434	1.10(0.76-2.39)	1.28(0.71-2.19)	0.995	1.97(1.35-3.41)	1.73(1.13-3.87)	0.940
CA125(U/mL)[a]	13.24(9.95-21.30)	11.88(8.82-15.44)	0.125	12.51(8.64-19.67)	11(7.71-15.27)	0.703	11.6(8.95-18.18)	11.3(8.78-16.98)	0.672
Menopausal status[c]			0.277			0.068			0.770
Premenopausal	17(58.62%)	84(47.73%)		9(75.00%)	35(46.67%)		4(25.00%)	27(32.14%)	
Postmenopausal	12(41.38%)	92(52.27%)		3(25.00%)	40(53.33%)		12(75.00%)	57(67.86%)	
Histologic type[c]			0.157			0.350			0.402
Squamous cell carcinoma	24(82.76%)	162(92.05%)		12(100.00%)	66(88.00%)		13(81.25%)	75(89.29%)	
Others	5(17.24%)	14(7.95%)		0(0.00%)	9(12.00%)		3(18.75%)	9(10.71%)	
HPV status[c]			0.479			0.349			0.685
Positive	28(96.55%)	159(90.34%)		12(100.00%)	64(85.33%)		15(93.75%)	73(86.90%)	
Negative	1(3.45%)	17(9.66%)		0(0%)	11(14.67%)		1(6.25%)	11(13.10%)	
FIGO stage[c]			0.001*			0.005*			0.022*
IB1	2(6.90%)	66(37.50%)		0(0%)	30(40.00%)		1(6.25%)	24(28.57%)	
IB2	15(51.72%)	79(44.89%)		7(58.33%)	33(44.00%)		7(43.75%)	44(52.38%)	
IIA1	12(41.38%)	31(17.61%)		5(41.67%)	12(16.00%)		8(50.00%)	16(19.05%)	

[a] Nonnormally distributed variable was shown as the Median (interquartile range). [b] Normally distributed variable was shown as Mean ± standard deviation. [c] Categorical variable was shown as n (%).

Note: Continuous variables were compared using the independent samples t-test or the Mann-Whitney U test, as appropriate for their distribution. Categorical variables were compared using the Chi-square test (n≥5) or Fisher's exact test (n<5).

*$P < 0.05$

Table 3 Performances of the predictive models in the study cohorts.

Cohort	Models	AUC (95%CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
Training	Radiomics-T2WI-FS	0.898(0.839-0.958)	0.863	0.759	0.881	0.512	0.957
	Radiomics-DWI	0.851(0.768-0.933)	0.844	0.724	0.864	0.467	0.95
	2D-DL-T2WI-FS	0.784(0.700-0.868)	0.659	0.828	0.631	0.270	0.957
	2D-DL-DWI	0.722(0.631-0.813)	0.595	0.828	0.557	0.235	0.951
	2.5D-DL-T2WI-FS	0.934(0.889-0.979)	0.863	0.828	0.869	0.511	0.968
	2.5D-DL-DWI	0.838(0.761- 0.915)	0.854	0.690	0.881	0.488	0.945
	Stacking	0.983(0.970- 0.997)	0.927	1.000	0.915	0.659	1.000
Internal Validation	Radiomics-T2WI-FS	0.873(0.785-0.961)	0.862	0.750	0.880	0.500	0.957
	Radiomics-DWI	0.796(0.637- 0.954)	0.747	0.833	0.733	0.333	0.965
	2D-DL-T2WI-FS	0.716(0.534-0.897)	0.816	0.583	0.853	0.389	0.928
	2D-DL-DWI	0.673(0.534-0.812)	0.483	0.917	0.413	0.200	0.969
	2.5D-DL-T2WI-FS	0.827(0.683-0.970)	0.931	0.583	0.987	0.875	0.937
	2.5D-DL-DWI	0.726(0.549-0.902)	0.851	0.500	0.907	0.462	0.919
	Stacking	0.932(0.834- 1.000)	0.943	0.917	0.947	0.733	0.986
External validation	Radiomics-T2WI-FS	0.728(0.591- 0.865)	0.820	0.562	0.869	0.450	0.912
	Radiomics-DWI	0.797(0.682- 0.912)	0.800	0.875	0.786	0.437	0.971
	2D-DL-T2WI-FS	0.758(0.629-0.888)	0.78	0.625	0.81	0.385	0.919
	2D-DL-DWI	0.618(0.470- 0.766)	0.57	0.75	0.536	0.235	0.918
	2.5D-DL-T2WI-FS	0.789(0.655- 0.922)	0.76	0.75	0.762	0.375	0.941
	2.5D-DL-DWI	0.713(0.565-0.860)	0.63	0.812	0.595	0.277	0.943
	Stacking	0.868(0.787- 0.948)	0.790	0.937	0.762	0.429	0.985

Figures

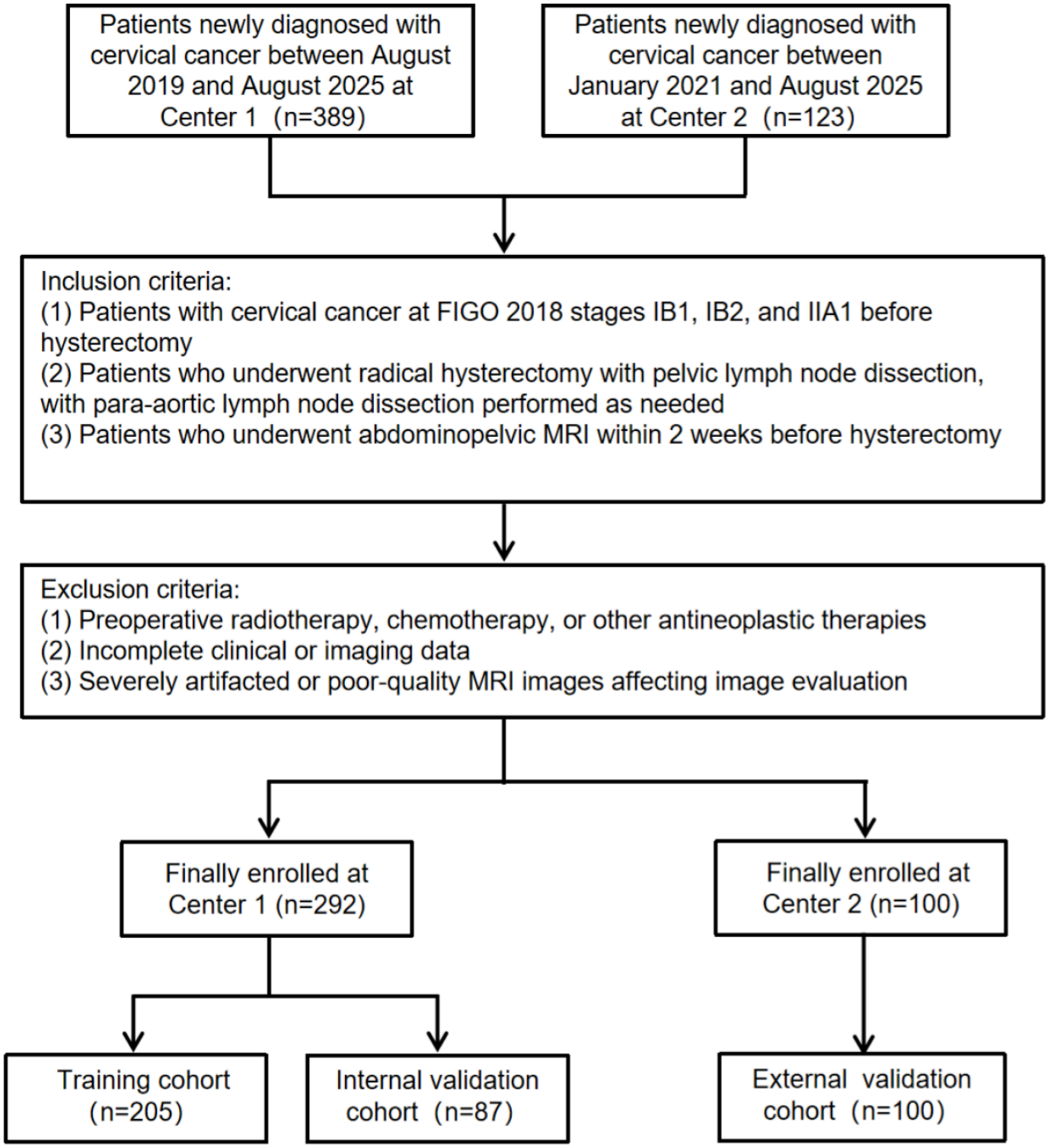


Figure 1

Flowchart for patient selection.

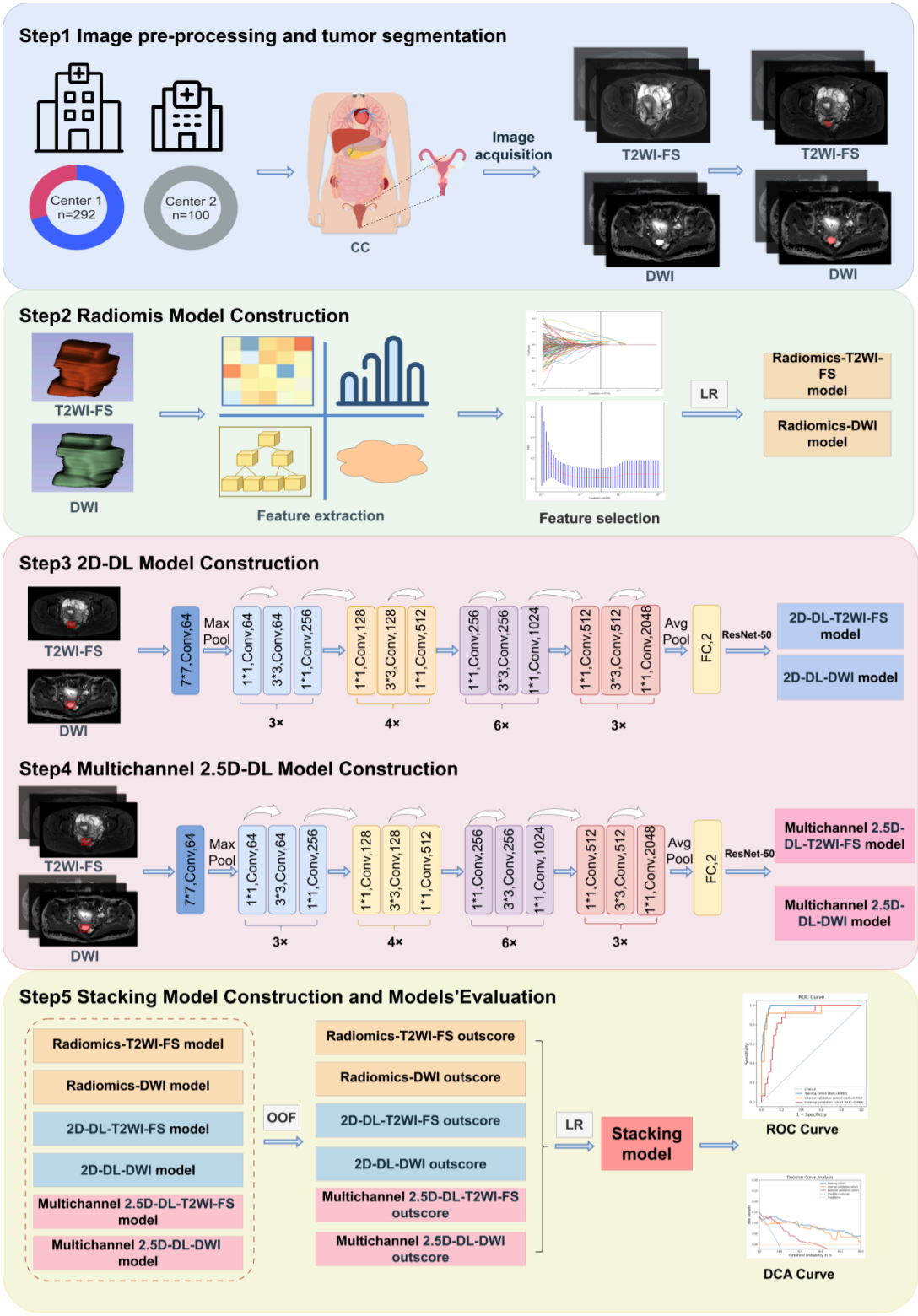


Figure 2

Flowchart of the study.

Out-of-fold (OOF) predicted probabilities from base models were used to avoid information leakage. The outscore denotes the predicted probability output of each base model and served as input features for the stacking model.

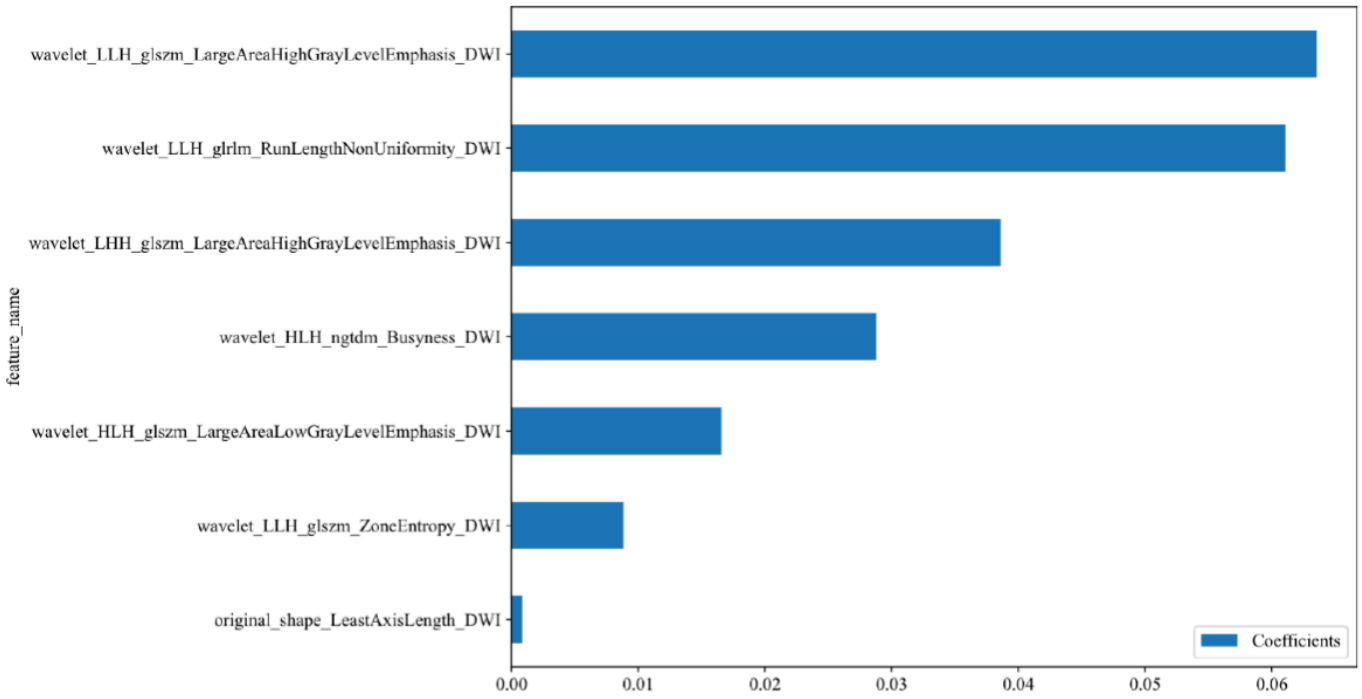


Figure 3

Key features and their coefficients in the DWI radiomics model.

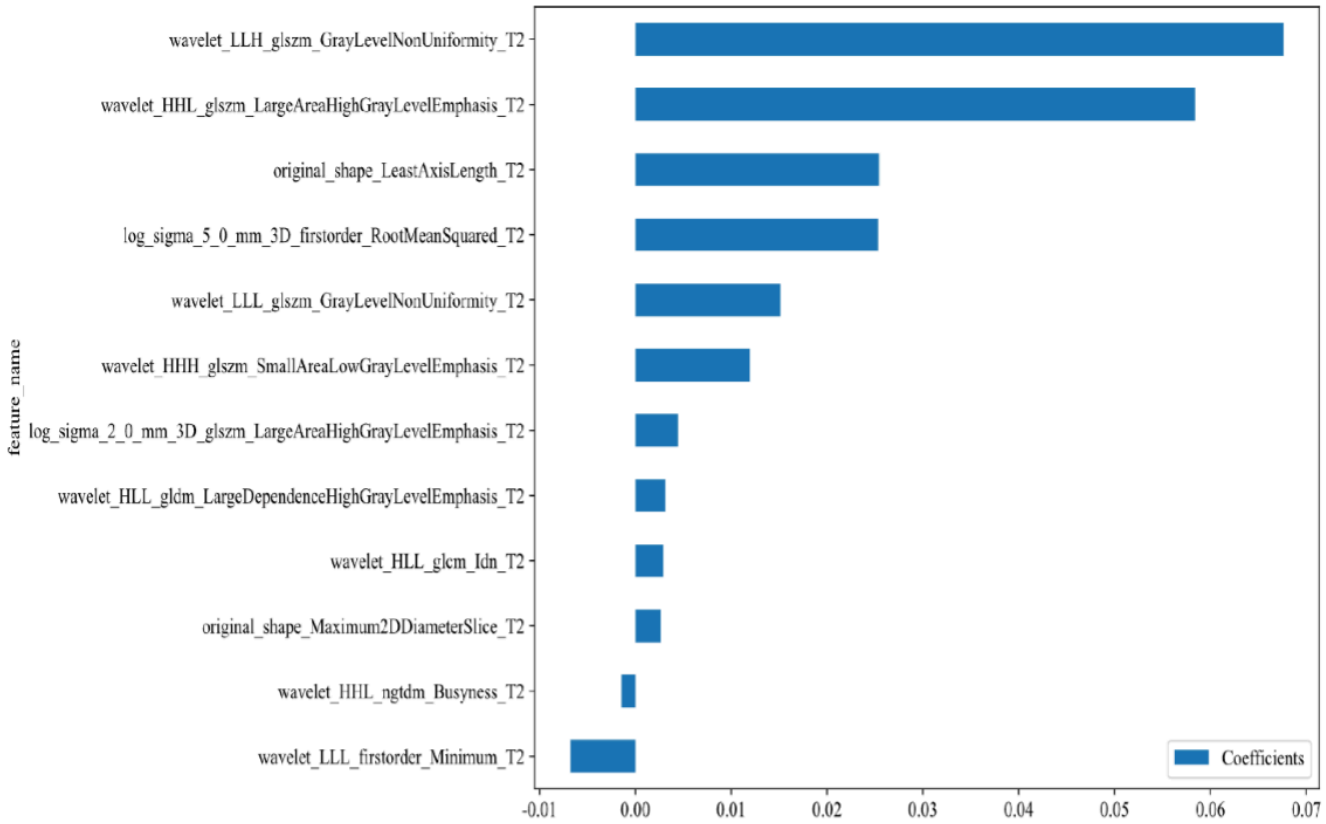


Figure 4

Key features and their coefficients in the T2WI-FS radiomics model.

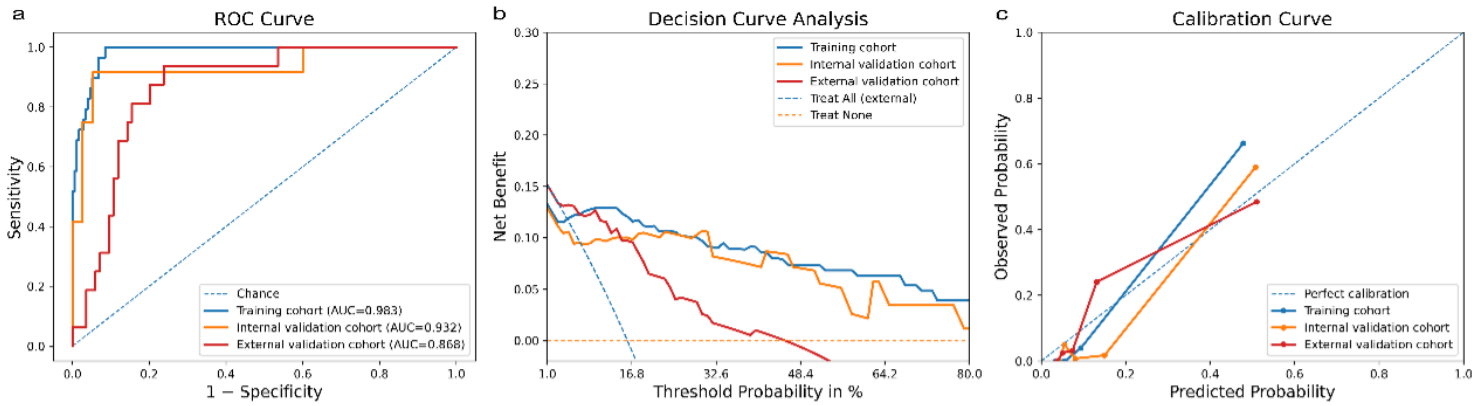


Figure 5

Performance evaluation of the stacking model in the training, internal validation, and external validation cohorts.

(a), Receiver operating characteristic (ROC) curves. (b), Decision curve analysis (DCA). (c), Calibration curves.

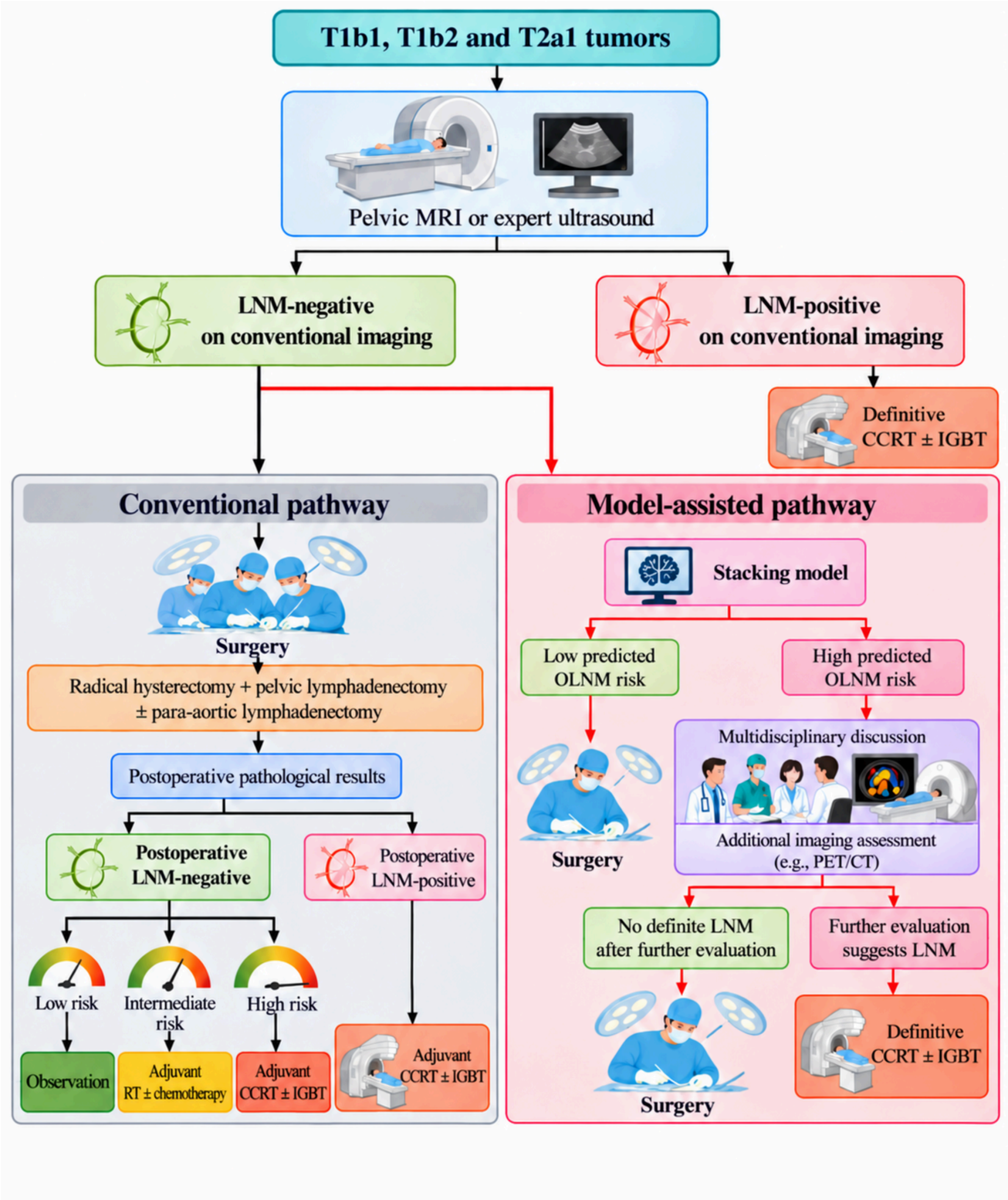


Figure 6

Clinical decision-support framework integrating the stacking model for preoperative OLNM risk stratification..

Abbreviations: T1b1, T1b2, and T2a1, tumor categories according to the UICC TNM Classification of Malignant Tumours, 9th edition; OLNM, occult lymph node metastasis; LNM, lymph node metastasis; MRI, magnetic resonance imaging; PET/CT, positron emission tomography/computed tomography; RT, radiotherapy; CCRT, concurrent chemoradiotherapy; IGBT, image-guided brachytherapy.

Supplementary Files

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

- Supplementarymaterial.docx