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.

Dual-Phase CT Radiomics with Deep Learning Segmentation for Classifying Lipid-Poor Adrenal Adenomas from malignant lesions: A multi-center study

Jianing Liu
Beijing Anzhen Hospital, Capital Medical University
Xuelian Gao
Beijing Anzhen Nanchong Hospital of Capital Medical University & Nanchong Central Hospital
Wentao Zhang
The First Hospital
Hongjian Li
Affiliated Hospital of North Sichuan Medical College
Zhen Zhou
Beijing Anzhen Hospital, Capital Medical University
Kun Jiang
Beijing Anzhen Hospital, Capital Medical University
Yiming Sang
Beijing Anzhen Hospital, Capital Medical University
Lei Chen
ShangHai United Imaging Intelligence Co., Ltd
Xianpan Pan
ShangHai United Imaging Intelligence Co., Ltd
Rui Li
ShangHai United Imaging Intelligence Co., Ltd
Lei Xu
leixu2001@hotmail.com

Beijing Anzhen Hospital, Capital Medical University

Research Article

Keywords: Lipid-poor adenoma, Adrenal gland neoplasms, Radiomics, Deep learning, Diagnosis

Posted Date: June 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9788821/v1

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

Additional Declarations: No competing interests reported.

Dual-Phase CT Radiomics with Deep Learning Segmentation for Classifying Lipid-Poor Adrenal Adenomas from malignant lesions: A multi-center study

Abstract

Purpose: To evaluate deep learning-based radiomics for differentiating lipid-poor adrenal adenomas (LP-AAs) from pheochromocytomas and malignant lesions using multiphase CT.

Methods: This retrospective multi-center study included 197 patients (116 LP-AAs, 81 non-adenomas) from four clinical centers. An automated segmentation framework was developed by integrating an augmented nnU-Net with a non-local attention mechanism to capture long-range spatial dependencies. Radiomic features were extracted from non-contrast, arterial, and venous phase CT images, and modeled using logistic regression (LR), support vector machine, and random forest. The optimal radiomics signature was integrated with clinical features to establish a clinical-radiomic model. Diagnostic performance of models was evaluated using receiver operating characteristic (ROC) curves and the DeLong test. Calibration and decision curve analysis (DCA) were performed to assess accuracy and clinical utility.

Results: The proposed deep learning framework achieved high-precision automated segmentation for all lesions. The combined multiphase radiomic models outperformed single-phase models across classifiers, with AUCs exceeding 0.84. The LR combined model achieved an AUC of 0.863 (95% CI 0.769-0.957). The arterial phase model demonstrated robust performance (AUC 0.829, 95%CI 0.711-0.947), with no significant difference compared with the combined model ($p>0.05$). The clinical-radiomic model achieved an AUC of 0.892 (95% CI 0.809-0.975), with a specificity of 0.913, though improvement over the combined radiomic model was not significant ($p=0.2292$).

Conclusion: Multiphase CT radiomics with automated deep learning segmentation robustly discriminates LP-AAs from malignant and high-risk adrenal tumors. The arterial phase contributes the predominant discriminatory signal, while integration of clinical features enhances specificity.

Key words: Lipid-poor adenoma; Adrenal gland neoplasms; Radiomics; Deep learning; Diagnosis

Introduction

Adrenal incidentaloma is defined as an asymptomatic adrenal mass detected incidentally during imaging studies performed for indications unrelated to suspected adrenal pathology[1]. A tenfold rise in the detection of adrenal tumors has been documented over the last three decades[2]. Current estimates suggest a prevalence of 1-10% in the general population, peaking between the ages of 70 and 80 years[3]. Although solitary, benign, non-functional adrenocortical adenomas account for most incidental adrenal masses (approximately 75%), a clinically critical minority—including adrenocortical carcinoma (8%), pheochromocytoma (PHEO) (7%), and metastases (5%)—are malignant or high-risk[3-5].

Adrenal adenomas and malignant (or high-risk) adrenal tumors have distinct clinical implications for management and prognosis. Benign adenomas rarely require active intervention[6]. However, adrenal cortical carcinoma carries a poor prognosis, with a 5-year overall survival of 37–59.8% in adults[7]. For localized tumors, complete surgical resection is the primary curative treatment[6, 8]. For patients with adrenal metastasis, adrenalectomy is an effective treatment option that can provide durable survival benefits[9]. Stereotactic body radiotherapy serves as a well-tolerated and effective treatment method, especially for patients with oligometastatic disease[10]. Preoperative prediction of PHEO is crucial for

mitigating perioperative morbidity and mortality. Improper management can trigger life-threatening hypertension, arrhythmia, and stroke[11]. Therefore, accurately differentiating adenomas from malignant or high-risk tumors is critical to optimizing individualized therapy and patient outcomes.

Computed Tomography (CT), with its high spatial resolution and rapid acquisition, has become the preferred imaging method for the initial characterization and diagnosis of adrenal tumors. While an unenhanced attenuation of <10 HU is pathognomonic for lipid-rich adenomas [1, 6], differentiating lipid-poor adenomas (LP-AAs) remains challenging. Their unenhanced attenuation (>10 HU) overlaps with non-adenomatous lesions—such as adrenocortical carcinomas, metastases, and pheochromocytomas—with some malignancies exhibiting HU values (10-30 HU) indistinguishable from LP-AAs[12-14]. Furthermore, rapid contrast washout (>60%) can cause some small malignancies or pheochromocytomas to be misclassified as adenomas [6]. Misdiagnosis of lipid-poor adrenal adenomas as malignant carries the risk of unnecessary surgery with its attendant risks of bleeding and infection, along with the potential need for long-term hormone replacement therapy and a poorer quality of life[9, 15, 16]. More importantly, delayed diagnosis of malignant tumors worsens prognosis [17]. This underscores the urgent need for

efficient and accurate diagnostic methods to differentiate lipid-poor adrenal adenomas from malignant and high-risk tumors.

Advances in artificial intelligence (AI), particularly radiomics and deep learning (DL), have facilitated automated segmentation and classification of adrenal lesions. Accurate segmentation of adrenal lesions is essential for characterization, volumetric quantification, and therapeutic assessment. Nevertheless, diverse morphology, subtle tissue contrast, and high scale-variability pose significant challenges to automated methods. To address these, we present a deep learning framework that integrates the robust system engineering of nnU-Net with the global modeling strengths of non-local attention. By embedding non-local blocks into the deep encoder layers of the self-configuring nnU-Net, our model captures long-range spatial dependencies while maintaining structural stability. This approach achieves high-precision, end-to-end segmentation with optimized computational efficiency. Besides, radiomics facilitates a robust evaluation of tumor heterogeneity and the microenvironment through its capacity for high-throughput, quantitative feature analysis[18]. While previous investigations have utilized radiomics to discriminate LP-AAs from non-adenomas or isolated high-risk entities such as metastases or pheochromocytomas[1, 14, 19, 20], few studies have specifically addressed the differentiation between LP-AAs and a diverse spectrum of

common high-risk tumors. This study aims to explore the value of automatic segmentation-based radiomics in differentiating LPAs from malignant and high-risk tumors, ultimately assisting clinicians in guiding management strategies and improving patient outcomes.

Materials and methods

Study population

This study was approved by the institutional review board (approval No. KS2025267), which waived the requirement for informed consent due to its retrospective nature. This study retrospectively reviewed the data of consecutive patients that underwent adrenalectomy in our hospital (Center One) between September 2015 and July 2025. The inclusion criteria included: (1) Patients who underwent adrenalectomy for adrenal masses and had definitive pathological diagnoses. (2) Surgical pathology confirmed diagnosis of adrenal adenomas, pheochromocytomas and malignant adrenal neoplasms (including primary tumors and metastases). (3) Patients who underwent contrast-enhanced CT imaging within one month before treatment and obtained multiphase (unenhanced, arterial and venous) images. The exclusion criteria were as follows: (1) Lipid rich adenomas with mean attenuation values <10 HU on non-contrast CT. (2) Maximum diameter of the tumor <10 mm (smaller volume may lead to

significant errors in lesion delineation). (3) CT images with slice thickness ⩾1.25 mm. (4) Poor image quality. (5) Patients with incomplete clinical data. The same criteria were applied to collect data from three additional hospitals: Center Two from June 2020 to July 2025, Center Three from January 2021 to December 2024, and Center Four from January 2022 to December 2024. Ultimately, 197 patients were included in the final analysis, with 116 cases of LP-AA, 35 cases of PHEO, and 46 cases of malignant lesions. The data were divided as follows: Centers One and Two constituted the internal cohort, which was split into training and validation sets (8:2 ratio), while Centers Three and Four served as the external testing set. The flowchart of patient enrollment and allocation is shown in Figure 1.

CT image acquisition

CT images were obtained from a variety of multidetector CT scanners, including a TOSHIBA Aquilion ONE (Canon Medical Systems), Somatom Definition Flash, Somatom Force (both from Siemens Healthcare), and GE Medical System Revolution CT (GE Healthcare). All these scanners were located at the internal cohort. The external testing cohort provided scans from a Somatom Definition Flash and Somatom Definition AS+128 (both from Siemens Healthcare).

All CT examinations were performed using a standardized contrast-

enhanced protocol. Following unenhanced image acquisition, 1.5 mL/kg of the contrast agent iodixanol (Iopromide; Ultravist 370; Bayer) was injected intravenously at a rate of 3 – 4 mL/s. Using a bolus-tracking technique, the arterial phase was acquired 25 – 35 s after injection, followed by the portal-venous phase at 60 – 80 s.

Clinical and CT image features analysis

The patients' clinical data and images were collected from the image archives and communication systems. Image features were analyzed by a junior radiologist and reviewed by a senior radiologist with over 10 years of experience. Both radiologists performed the image analysis independently while being blinded to the clinical information and histological diagnosis. Disagreements between the two radiologists were resolved through discussion. The window settings were adjusted using a window level of 40−80 HU and a window width of 200–400 HU. The following clinical and CT imaging features of patients were recorded: sex, age, lesion location (left and right), diameter (average of the maximal long-axis and maximal short-axis perpendicular to it), attenuation values on non-contrast (NC) CT, arterial phase (AP) CT, and venous phase (VP) CT, and morphological characteristics (shape, boundary, and homogeneity).

Deep learning based-automatic segmentation

As illustrated in the integrated workflow (Figure 2), the pipeline begins

with automated adrenal gland segmentation from input 3D CT volumes to generate an anatomical mask[21]. The adrenal region, expanded by an outward padding, is subsequently cropped to define the region of interest (ROI) for subsequent lesion segmentation. To refine the results, candidate lesions outside the adrenal mask are excluded, and a size-based filter is applied to retain only those with a long-axis diameter $\geqslant$ 5 mm, thereby minimizing potential false positives. Our segmentation architecture is built upon the self-configuring nnU-Net framework, further enhanced by a non-local attention mechanism to capture long-range spatial dependencies.

The nnUNet's strength lies in automating medical image segmentation best practices, adapting to datasets rather than fixed structures: data-driven preprocessing (unifying resolution, normalizing intensity) and dynamic network configuration (optimizing patch size, depth and batch size based on image and GPU constraints). The non-local module, based on self-attention, can calculate dependencies between any two feature map positions directly. For the input feature map, the module generates three feature projections, namely Query, Key and Value, through three 1×1 convolutions. For the feature at a certain position *i* on the output feature map, its value is obtained by the weighted summation of features at all positions on the input feature map, as simplified in Formula:

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j) \tag{1}$$

where $f(x_i, x_j)$ calculates the correlation strength between position i and any position j, $g(x_j)$ transforms the feature at position j, and $C(x)$ is a normalization factor. y_i is the weighted sum of features at all positions, with weights determined by the correlation strength. The DL model adopted the Adam optimizer, Dice-cross-entropy loss, initial learning rate 0.01 (polynomial decay), 2000 epochs.

Feature selection and dimensionality reduction

Following expert review and refinement by a senior radiologist, deep learning-derived segmentation masks were processed for radiomic analysis via the uAI Research Portal (https://urp.united-imaging.com, United Imaging Intelligence, China). This end-to-end workflow, encompassing feature extraction, selection, and model construction, was executed using Python (v3.7.3) and the PyRadiomics open-source framework (https://pyradiomics.readthedocs.io/en/latest/index.html). The multi-center imaging data were normalized to a standardized abdominal window setting (window width: 300; window level: 40). High-throughput radiomics features of the lesions extracted from NCCT, APCT, and VPCT images in this study included: first-order features, shape features, high order feature, and texture features including gray level co-occurrence matrix (GLCM), gray level size zone matrix (GLSZM), gray level dependence matrix (GLDM), and neighboring gray difference matrix

(NGTDM).

After Z-score normalization, feature selection comprised sequential filtering stages. Firstly, features significantly correlated with the outcome ($p < 0.05$) were identified using correlation analysis. Secondary selection applied the maximum relevance minimum redundancy (mRMR) algorithm to identify the top 100 features. A tertiary filter using the select K-best method reduced these to 35 features. Ultimately, the least absolute shrinkage and selection operator (LASSO) regression yielded the final set of 8–10 most predictive features.

Construction of radiomics machine learning models

Risk prediction models were constructed using the selected features. To identify the optimal model for tumor data classification, this study employed three main machine learning algorithms: logistic regression (LR), support vector machines (SVM), and random forests (RF). Patients were randomly allocated into training and validation sets at an 8:2 ratio in the internal cohort. The models were subsequently evaluated on the independent external testing set. The clinical and CT characteristics in LP-AA and non-adenoma cohorts (pheochromocytomas and malignant neoplasms) were firstly analyzed using univariate analysis, and characteristics with $p < 0.05$ in univariate analysis were further included in multivariate logistic regression. The final predictor variables were selected,

and odds ratios (ORs) with 95% confidence intervals (CIs) were calculated for each factor were calculated. After evaluation, the optimal machine learning model was integrated with selected clinical and CT features to construct a radiomics-clinical combined model.

Statistical analysis

Statistical analyses of clinical and CT features were conducted using the SPSS software version 25.0 (IBM Corp., Armonk, NY, USA). Kolmogorov–Smirnov normality test was used to evaluate whether quantitative variables obey normal distribution. Continuous variables conforming to a normal distribution were expressed as the mean ± standard deviation; otherwise, they were expressed as the median (the first quantile; the third quantile). Differences between groups were assessed using the Mann–Whitney U test or independent samples t-test for continuous variables, as appropriate, and Pearson's chi-square test for categorical variables. Model diagnostic performance was evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, F1 score, positive predictive value, and negative predictive value. The AUCs of different models were compared using the Delong test. All statistical tests were two-sided, and values of $p < 0.05$ were considered significant. Calibration curves were used to evaluate the consistency between predicted and actual outcomes. In addition, the

clinical utility of the models was evaluated using decision curve analysis (DCA).

Results

General information and CT characteristics

The internal dataset comprised 141 patients, 83 of which were diagnosed with LP-AA and 58 with non-adenoma (pheochromocytomas and malignant neoplasms). The external dataset included 33 patients with LP-AA and 23 with non-adenoma. In the internal set, adrenal lipid-poor adenomas exhibited lower diameters and CTu compared to non-adenomas (medians: 22.0 HU vs 35.5 HU, 20.0 mm vs 44.5 mm, *p* values all<0.001). In addition, there were significant differences among the LP-AA and non-adenoma in shape, boundary, and homogeneity. Detailed information and CT characteristics of the internal cohort and external testing cohort are presented in Table 1. Histopathological findings were available for patients and are summarized in Table 2.

Extraction of radiomic features and optimal feature selection

All lesions were successfully identified by the DL segmentation model, and the representative results of the model are shown in Figure 3. A total of 6,792 high-throughput radiomic features were extracted, comprising 2,264

features from each of the three CT phases. For each CT phase dataset, an initial set of 100 features was selected through correlation analysis and mRMR, which was then reduced to 35 features using the select K-best method. The final features for the single-phase radiomic models were determined by LASSO regression, with 8, 8, and 9 features from the NCCT, APCT, and VPCT datasets, respectively. These features retained from the three phases were integrated into a combined feature set. After further screening using the LASSO algorithm, 10 key features were identified and utilized to build the final combined radiomics model. Univariate and multivariate logistic regression analyses conducted on the internal cohort indicated that diameter (OR = 0.895, 95% confidence interval [CI] 0.847–0.946, $p < 0.001$) and CTu (OR = 0.852, 95% CI 0.789–0.920, $p < 0.001$) are independent predictive factors. These clinical features were utilized to construct a clinical-radiomic model in conjunction with the 10 key features from the combined feature set. Figure 4 presents the Pearson's ρ correlation coefficients for the selected features.

Construction and evaluation of models

Table 3 shows the diagnostic performance of the LR, RF, and SVM models in the training, validation and external testing sets. Compared to single-phase models, all combined models demonstrated more robust performance. In the external validation sets, the AUC values for the LR,

RF, and SVM combined models were 0.863 (95% CI 0.769-0.957), 0.893 (95% CI 0.812-0.974), and 0.862 (95% CI 0.767-0.956), respectively. DeLong's test revealed no statistically significant differences in performance between any pair of combined models (all $p > 0.05$).

In this study, the radiomics models based on the LR classifier was selected for further analysis and constructing a final clinical-radiomic model. The combined model demonstrated the best diagnostic performance in the training (AUC=0.965, 95% CI 0.935-0.995), validation (AUC= 0.853, 95% CI 0.687-1), and external testing (AUC= 0.863, 95% CI 0.769-0.957) groups. According to DeLong's test, the AUC of the combined model was significantly superior to those of the NCCT and VPCT models (all $p < 0.05$) in external testing set, whereas no significant statistical difference was observed compared with the APCT model ($p > 0.05$). The clinical-radiomic model developed by integrating clinical features showed a performance boost in external testing set (AUC=0.892, 95% CI 0.809-0.975), yet no statistically significant difference was observed relative to the combined model ($p > 0.05$). Table 4 summarizes the performance evaluation results of all the LR models in the internal and external groups. Detailed comparisons of the diagnostic performance across these models within the external testing cohort are presented in Table 5 and Figure 5. The DCA curves demonstrated that the LR models yielded clinical utility

(Appendix Figure 6).

Discussion

Differentiating LP-AAs from malignant lesions on CT is challenging. This multicenter study developed robust machine learning models using multi-phase CT radiomics. Logistic regression showed that both combined and radiomic-clinical models reliably classified lesions, achieving a peak AUC of 0.892 (95% CI: 0.809–0.975). Additionally, an augmented nnU-Net with deep non-local attention was developed, significantly improving automatic segmentation accuracy for complex adrenal tumor boundaries by capturing global spatial dependencies.

While supplementary imaging or biochemical assays may facilitate the diagnosis of adrenal tumors, they entail higher costs and extended pre-treatment waiting times without entirely eliminating the risk of misdiagnosis[22]. Furthermore, diagnostic delays may adversely affect patient prognosis, highlighting the need for more efficient diagnostic tools. Recently, deep learning and radiomics have emerged as non-invasive, efficient options for tumor differentiation[1, 20, 22]. However, previous studies have predominantly focused on classification between LP-AAs and a single malignant entity. In clinical practice, radiologists need distinguish benign adenomas from a heterogeneous spectrum of malignant or high-risk

lesions requiring intervention. To better align with real-world diagnostic reasoning, our multicenter radiomics models were challenged with a heterogeneous malignant cohort from four hospitals, encompassing common metastases and rare malignancies (e.g., adrenal schwannoma, lymphoma, squamous cell carcinoma, adrenocortical carcinoma, and sarcomatoid carcinoma). Maintaining robust performance under these rigorous conditions underscores the model's generalizability and resistance to tumor-specific overfitting. Notably, according to WHO criteria, PHEO is defined as a neoplasm with malignant potential and heterogeneous behavior[23]. Given its risk for life-threatening perioperative hypertensive crises, PHEO was conservatively grouped into the malignant cohort to ensure maximum diagnostic vigilance.

Automated segmentation enhances radiomic feature extraction efficiency, delineation accuracy, and reproducibility—a critical foundation for clinical translation[24-28]. To achieve high-precision results, we augmented the nnU-Net framework with a non-local attention mechanism to capture long-range spatial dependencies. This algorithm integrates nnU-Net’s self-configuring, automated workflow with the non-local module’s global contextual perception, synergistically enhancing segmentation performance. The advantages of nnU-Net are two-fold. First, its automated, standardized preprocessing (including anisotropic resampling and

modality-specific intensity normalization) eliminates non-biological variability, letting the model focus on essential features. Second, its dynamic topology configuration autonomously optimizes patch size, network depth, and batch size based on image dimensions and GPU memory, maximizing the contextual receptive field under computational constraints. Conversely, conventional U-Nets rely on local convolutions, which sacrifice spatial resolution to expand the receptive field and fail to capture long-range dependencies efficiently. Furthermore, successive downsampling degrades global information transferred via skip connections, hindering boundary recovery in the decoder. To overcome this, we introduced a plug-and-play non-local attention module that captures long-range dependencies across arbitrary image positions. Its structural flexibility allows seamless integration across U-Net stages—either to augment local features or aggregate global context.Based on nnU-Net's automated preprocessing, we derived optimized parameters: an input patch size of [64, 160, 192], uniform 3×3×3 convolutions across six blocks, and downsampling steps of [4, 5, 5] per axis. Non-local modules were strategically inserted into deep encoder layers; here, the reduced spatial dimensions minimize computational overhead while the rich semantic information enables efficient global association. This supplies the decoder with globally enriched features, substantially improving segmentation accuracy for boundaries and complex morphologies. Trained on 871 CT

scans and evaluated on an independent test set of 164 scans, the model achieved satisfactory multi-phase segmentation results, yielding a Dice coefficient of 0.89, recall of 0.9806, and accuracy of 0.9917 (Appendix Table 6).

Based on automated segmentation, radiomics models in this study were constructed using three machine learning methods (LR, SVM and RF) across various CT phases. A primary finding of our analysis is the consistent diagnostic superiority of combined-phase models over single-phase configurations. Regardless of the machine learning classifier employed, the combined models consistently achieved the highest diagnostic efficacy. Specifically, the RF-based combined model reached a peak AUC of 0.893 (95% CI 0.812–0.974) in the external test set, while LR and SVM combined models yielded comparable AUCs of 0.863 and 0.862, respectively. This inter-algorithm consistency suggests that the integrated radiomic features from multiple temporal phases capture essential biological information that is robust across different mathematical frameworks. Furthermore, models based on NC and VP phases showed relatively low AUC values and modest diagnostic performance, whereas AP models demonstrated robust diagnostic efficacy. For instance, within the RF framework, the AP model achieved an AUC of 0.879 in the external test set, whereas the NC and VP models dropped to 0.640 and 0.535,

respectively. This result has preliminarily revealed the clinical importance of integrating multiphase radiomic features and the pivotal role of the AP in adrenal tumor discrimination.

The LR framework was chosen for its interpretability and compatibility with analytical tools. Results show that integrating multiphase CT radiomics significantly enhances preoperative adrenal tumor characterization compared to single-phase models. The combined model achieved a robust external test AUC of 0.863, outperforming the NC (0.677) and VP (0.672) models. Significant IDI metrics confirmed this improvement against NC (p = 0.0092) and VP (p = 0.0022), proving that temporal enhancement patterns reflect tumor heterogeneity more comprehensively than single-phase "snapshots." Furthermore, the AP emerged as the pivotal single-phase surrogate for adrenal lesion differentiation, yielding the highest single-phase AUC of 0.829 in the external validation cohort. Although the combined model achieved a higher numerical AUC, the difference was not statistically significant (p = 0.2416). Consistently, Shan et al. [22] showed that tri-phase radiomics outperformed other models in differentiating PHOEs from LP-AAs (AUC 0.911 vs. arterial-phase AUC 0.907). While multiphase CT characterizes dynamic enhancement and tumor heterogeneity based on perfusion and angiogenesis differences, the AP specifically captures initial bolus passage,

reflecting micro-vessel density and malignancy-associated hypervascularity[29]. Clinically, the AP retains strong independent diagnostic value, making a targeted arterial-phase scan a high-efficiency, low-radiation alternative. Finally, the clinical-radiomic fusion model critically enhanced diagnostic specificity. Though its AUC improvement over the combined radiomics model lacked statistical significance (p = 0.2292), the fusion model reached a peak external test specificity of 0.913 versus 0.696 for the combined model. This substantial false-positive reduction is vital in adrenal oncology, directly minimizing unnecessary biopsies and surgeries for benign incidentalomas, thus optimizing patient management and healthcare costs.

This study identified tumor diameter and CTu as independent predictors for differentiating LP-AAs from non-adenomas. In the internal cohort, non-adenomas exhibited a significantly larger median diameter (44.5 mm) than LP-AA (20.0 mm), a disparity that persisted in the external validation set ($p < 0.001$). This aligns with the clinical consensus that larger adrenal masses possess higher malignant potential [30]. Furthermore, while LP-AA typically shows higher attenuation than lipid-rich variants, its median CTu remained significantly lower than that of non-adenomas (22.0 HU vs. 35.5 HU, $p < 0.001$). This suggests that even among "lipid-poor" lesions, a lower baseline density is a robust indicator favoring

adenoma over aggressive malignancies or PHEOs. Qualitative CT features also provided further diagnostic insight, as non-adenomas more frequently exhibited irregular shapes, ill-defined margins, and internal heterogeneity (all $p < 0.001$). These findings likely reflect invasive growth, necrosis, or hemorrhage characteristic of adrenal malignancies. Notably, absolute attenuation values in the arterial and venous phases were comparable between groups ($p > 0.05$). This lack of differentiation underscores the insufficiency of conventional contrast-enhanced CT values for definitive diagnosis and justifies the application of advanced radiomics to capture subtle, non-visual textural patterns.

This study has several limitations. First, its retrospective nature may introduce selection bias. Second, although diverse non-adenomatous lesions simulated real-world scenarios, the small sample size of rare malignant subtypes limited the model's ability to capture their specific radiomic signatures. Third, the clinical-radiomic model excluded longitudinal imaging and biochemical markers (e.g., hormone levels) crucial for adrenal incidentaloma management. Future prospective, larger multicenter studies integrating multimodal data are needed to enhance clinical utility. Finally, our deep learning framework only performs high-precision segmentation; future work will optimize this technology into an integrated pipeline coupling automated segmentation with lesion

classification.

Conclusion

Integrating CT and radiomic features, our logistic regression classifier effectively differentiates lipid-poor adrenal adenomas from high-risk tumors. High-precision, reproducible radiomic feature extraction was ensured via automated segmentation using an enhanced nnU-Net with non-local attention. This joint approach provides a reliable, non-invasive preoperative tool for quantitative adrenal disease diagnosis, ultimately optimizing clinical management strategies for adrenal incidentalomas.

Abbreviations:

AP, arterial phase; AI, artificial intelligence; AUC, the area under the receiver operating characteristic curve; CT, computed tomography; DCA, decision curve analysis; DL, deep learning; LP-AA, lipid-poor adrenal adenoma; LR, logistic regression; NC, non-contrast; NPV, negative predictive value; PPV, positive predictive value; PHEO, pheochromocytoma; SVM, support vector machines; RF, random forests; VP, venous phase.

References:

[1]. Feliciani, G., et al., Radiomics in the characterization of lipid-poor adrenal adenomas at unenhanced CT: time to look beyond usual density metrics. European Radiology, 2024. 34(1): p. 422-432.

[2]. Ebbehoj, A., et al., Epidemiology of adrenal tumours in Olmsted County, Minnesota, USA: a population-based cohort study. The Lancet Diabetes & Endocrinology, 2020. 8(11): p. 894-902.

[3]. Pacak, K., et al., Adrenal tumour imaging: clinical, molecular, and radiomics perspectives. The lancet. Diabetes & endocrinology, 2026. 14(1): p. 62-81.

[4]. Fassnacht, M., et al., European Society of Endocrinology clinical practice guidelines on the management of adrenal incidentalomas, in collaboration with the European Network for the Study of Adrenal Tumors. European Journal of Endocrinology, 2023. 189(1): p. G1-G42.

[5]. Jing, Y., et al., Prevalence and Characteristics of Adrenal Tumors in an Unselected Screening Population: A Cross-Sectional Study. Annals of internal medicine, 2022. 175(10): p. 1383-1391.

[6]. Chung, R., et al., Adrenal Neoplasms: Lessons from Adrenal Multidisciplinary Tumor Boards. Radiographics : a review publication of the Radiological Society of North America, Inc, 2023. 43(7): p. e220191.

[7]. Ghosh, C., J. Hu and E. Kebebew, Advances in translational research of the rare cancer type adrenocortical carcinoma. Nature Reviews Cancer, 2023. 23(12): p. 805-824.

[8]. Flauto F, et al., The management of adrenocortical carcinoma in the era of precision medicine. Crit Rev Oncol Hematol, 2025. 214(Oct): p. 104839.

[9]. Rémond, A., et al., Survival and Prognostic Factors After Adrenalectomy for Secondary Malignancy: A Combined Analysis of a French University Center Registry (Eurocrine) of 307 Patients and a French Nationwide Study of 2515 Patients. Annals of Surgery, 2024. 280(5): p. 825-834.

[10]. Stumpf, P.K., et al., Modeling of Tumor Control Probability in Stereotactic Body Radiation Therapy for Adrenal Tumors. International Journal of Radiation Oncology*Biology*Physics, 2021. 110(1): p. 217-226.

[11]. Kong, J., et al., Development of a radiomics model to diagnose pheochromocytoma preoperatively: a multicenter study with prospective validation. Journal of Translational Medicine, 2022. 20(1): p. 31.

[12]. Nagayama, Y., et al., Relative Enhancement Ratio of Portal Venous Phase to Unenhanced CT in the Diagnosis of Lipid-poor Adrenal Adenomas. Radiology, 2021. 301(2): p. 360-368.

[13]. Yip, L., et al., American Association of Endocrine Surgeons Guidelines for Adrenalectomy: Executive Summary. JAMA Surgery, 2022. 157(10): p. 870.

[14]. Wang, S., et al., Evaluation of a fusion model combining deep learning models based on enhanced CT images with radiological and clinical features in distinguishing lipid-poor adrenal adenoma from metastatic lesions. BMC Medical Imaging, 2025. 25(1): p. 219.

[15]. Fiore, M., et al., Residual Adrenal Function After Multivisceral Resection With Adrenalectomy in Adult Patients. JAMA Surgery, 2022. 157(5): p. 415.

[16]. Guarnotta, V., et al., Conventional steroids vs. dual-release hydrocortisone on metabolic, cardiovascular, and bone outcomes in adrenal insufficiency: a 10-year study. European Journal of Endocrinology, 2024. 191(3): p. 300-311.

[17]. Fassnacht, M., et al., Adrenocortical carcinoma: a practical guide for clinicians. The lancet. Diabetes & endocrinology, 2025. 13(5): p. 438-452.

[18]. Kang, W., et al., Application of radiomics-based multiomics combinations in the tumor

microenvironment and cancer prognosis. Journal of Translational Medicine, 2023. 21(1): p. 598.
[19]. Gündoğdu, E., B.K. Aşılıoğlu and C. Yazıcı, Whole-lesion CT histogram analysis as an advanced technique in the portal venous phase: differentiating lipid poor adrenal adenomas from pheochromocytomas. Abdominal Radiology, 2025. 50(3): p. 1219-1227.
[20]. Cao, L., et al., Clinical-imaging-radiomic nomogram based on unenhanced CT effectively predicts adrenal metastases in patients with lung cancer with small hyperattenuating adrenal incidentalomas. Oncol Lett, 2024. 28(2): p. 340.
[21]. Z Wang, et al., Non-Local U-Nets for Biomedical Image Segmentation. Proceedings of the AAAI Conference on Artificial Intelligence, 2020. 34(4): p. 6315-6322.
[22]. Shan, Z., et al., Combined nomogram for differentiating adrenal pheochromocytoma from large-diameter lipid-poor adenoma using multiphase CT radiomics and clinico-radiological features. BMC Medical Imaging, 2025. 25(1): p. 313.
[23]. Mete, O., et al., Overview of the 2022 WHO Classification of Paragangliomas and Pheochromocytomas. Endocrine pathology, 2022. 33(1): p. 90-114.
[24]. Currie, G.M. and K.E. Hawk, Artificial Intelligence Augmented Cerebral Nuclear Imaging. Seminars in Nuclear Medicine, 2025. 55(4): p. 616-628.
[25]. Veiga-Canuto, D., et al., Reproducibility Analysis of Radiomic Features on T2-weighted MR Images after Processing and Segmentation Alterations in Neuroblastoma Tumors. Radiology. Artificial intelligence, 2024. 6(4): p. e230208.
[26]. Chen, X., et al., Gastric Neoplasm Detection at Contrast-enhanced CT with Deep Learning. Radiology. Artificial intelligence, 2026. 8(1): p. e250145.
[27]. Feng, X., et al., End-to-end deep learning model for the diagnosis and segmentation of primary retroperitoneal neoplasm: a multicenter cohort study. EClinicalMedicine, 2025. 89: p. 103498.
[28]. Boyd, A., et al., Stepwise Transfer Learning for Expert-level Pediatric Brain Tumor MRI Segmentation in a Limited Data Scenario. Radiology. Artificial intelligence, 2024. 6(4): p. e230254.
[29]. Nagayama, Y., et al., Relative Enhancement Ratio of Portal Venous Phase to Unenhanced CT in the Diagnosis of Lipid-poor Adrenal Adenomas. Radiology, 2021. 301(2): p. 360-368.
[30]. Chung, R., et al., Adrenal Neoplasms: Lessons from Adrenal Multidisciplinary Tumor Boards. Radiographics : a review publication of the Radiological Society of North America, Inc, 2023. 43(7): p. e220191.

Table 1 Demographic information and CT characteristics of the internal cohort and external testing cohort

Characteristics	Internal cohort (n=141)			External testing cohort (n=56)		
	LP-AA (n=83)	Non-adenoma (n=58)	*p*	LP-AA (n=33)	Non-adenoma (n=23)	*p*
Gender			0.309			0.258
Female	43 (51.8%)	25 (43.1%)		15 (45.5%)	7 (30.4%)	
Male	40 (48.2%)	33 (56.9%)		18 (54.5%)	16 (69.6%)	
Age (years)	53.4 ± 11.5	55.9 ± 14.4	0.249	54.0 (48.5; 59.5)	56.0 (47.5; 64.5)	0.041
Location			0.404			0.159
Left	53 (63.9%)	33 (56.9%)		22 (66.7%)	11 (47.8%)	
Right	30 (36.1%)	25 (43.1%)		11 (33.3%)	12 (52.2%)	
Diameter (mm)	20.0 (14.0; 26.0)	44.5 (27.6; 61.4)	<0.001	24.0 (18.0; 30.0)	38.0 (27.5; 48.5)	<0.001
CTu (HU)	22.0 (13.5; 30.5)	35.5 (32.5; 38.5)	<0.001	27.21 ± 8.989	32.17 ± 8.462	0.042
CTa (HU)	59 (41.5; 76.5)	53.0 (32.5; 73.5)	0.608	63.0 (39.0; 87.0)	51.0 (29.0; 73.0)	0.364
CTv (HU)	76.0 (62.5; 89.5)	64.5 (24.5; 104.5)	0.226	64.58 ± 21.062	70.96 ± 23.895	0.296
Shape			0.001			<0.001
Regular	79 (95.2%)	44 (75.9%)		33 (100.0%)	14 (60.9%)	
Irregular	4 (4.8%)	14 (24.1%)		0 (0.0%)	9 (39.1%)	
Boundary			<0.001			<0.001
Sharp	82 (98.8%)	46 (79.3%)		33 (100.0%)	14 (60.9%)	
Blotted	1 (1.2%)	12 (20.7%)		0 (0.0%)	9 (39.1%)	
Homogeneity			<0.001			0.034
Homogenous	65 (78.3%)	17 (29.3%)		21 (63.6%)	8 (34.8%)	
Heterogenous	18 (21.7%)	41 (70.7%)		12 (36.3)	15 (65.2%)	

Data are expressed as number (%), or mean ± standard deviation

CTu Mean attenuation value on non-contrast CT, *CTa* Mean attenuation value on arterial phase CT, *CTv* Mean attenuation value on venous phase CT

Table 2 Histopathological findings of the internal cohort and external testing cohort

	Histopathological findings	
Internal cohort		
	Adenoma	83 (58.9%)
	Metastasis	19 (13.5%)
	Adrenal schwannoma	4 (2.8%)
	Lymphoma	3 (2.1%)
	Adrenal squamous cell carcinoma	1 (0.7%)
	Adrenocortical carcinoma	5 (3.5%)
	Adrenal sarcomatoid carcinoma	1 (0.7%)
	Pheocromocytoma	25 (17.7%)
	Total	141 (100%)
External testing cohort		
	Adenoma	33 (58.9%)
	Metastasis	11 (19.6%)
	Lymphoma	1 (1.8%)
	Adrenocortical carcinoma	1 (1.8%)
	Pheocromocytoma	10 (17.9%)
	Total	56 (100%)

Data are expressed as number (%)

Table 3 The diagnostic performance of different radiomic models in training set, validation set, and external test set

Classification	Model	AUC (95% CI)			Sensitivity			Specificity			Accuracy		
		Train	Validation	Test	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test
LR	NC	0.823 (0.738-0.907)	0.789 (0.613-0.965)	0.677 (0.532-0.823)	0.849	0.824	0.758	0.739	0.667	0.565	0.759	0.679	0.836
	AP	0.955 (0.919-0.990)	0.819 (0.632-1)	0.829 (0.711-0.947)	0.894	0.824	0.788	0.870	0.750	0.652	0.793	0.732	0.901
	VP	0.817 (0.740-0.894)	0.716 (0.512-0.920)	0.672 (0.520-0.824)	0.758	0.824	0.576	0.652	0.667	0.696	0.759	0.625	0.758
	combine	0.965 (0.935-0.995)	0.853 (0.687-1)	0.863 (0.769-0.957)	0.909	0.882	0.788	0.848	0.750	0.696	0.828	0.750	0.902
RF	NC	0.943 (0.903-0.983)	0.726 (0.515-0.937)	0.640 (0.492-0.789)	0.909	0.882	0.636	0.761	0.583	0.609	0.848	0.759	0.625
	AP	0.981 (0.963-1)	0.828 (0.648-1)	0.879 (0.791-0.967)	0.924	0.882	0.849	0.891	0.750	0.739	0.911	0.828	0.804
	VP	0.891 (0.835-0.948)	0.716 (0.523-0.909)	0.535 (0.380-0.690)	0.894	0.765	0.455	0.652	0.417	0.565	0.795	0.621	0.500
	combine	0.999 (0.996-1)	0.860 (0.707-1)	0.893 (0.812-0.974)	0.985	0.941	0.849	0.957	0.750	0.696	0.973	0.862	0.786
SVM	NC	0.821 (0.735-0.906)	0.789 (0.614-0.964)	0.677 (0.531-0.824)	0.788	0.824	0.697	0.761	0.667	0.696	0.777	0.759	0.696
	AP	0.954 (0.919-0.990)	0.824 (0.643-1)	0.859 (0.760-0.958)	0.924	0.882	0.788	0.870	0.750	0.609	0.902	0.828	0.714
	VP	0.820 (0.743-0.896)	0.735 (0.541-0.929)	0.679 (0.529-0.828)	0.818	0.824	0.455	0.696	0.583	0.696	0.768	0.724	0.554
	combine	0.963 (0.931-0.994)	0.843 (0.677-1)	0.862 (0.767-0.956)	0.909	0.824	0.788	0.870	0.750	0.739	0.893	0.793	0.768

AUC Area under the receiver operating characteristic curve, *CI* Confidence interval

Table 4 The diagnostic performance of the LR models in training set, validation set, and external test set

Group	Model	AUC (95% CI)	Sensitivity	Specificity	Accuracy	F1 score	PPV	NPV
Train	NC	0.823 (0.738-0.907)	0.849	0.739	0.759	0.836	0.824	0.773
	AP	0.955 (0.919-0.990)	0.894	0.870	0.793	0.901	0.908	0.851
	VP	0.817 (0.740-0.894)	0.758	0.652	0.759	0.758	0.758	0.652
	Combined	0.965 (0.935-0.995)	0.909	0.848	0.828	0.902	0.896	0.867
	Clinical-radiomic	0.966 (0.938-0.994)	0.864	0.957	0.902	0.912	0.966	0.830
Validation	NC	0.789 (0.613-0.965)	0.824	0.667	0.679	0.800	0.778	0.727
	AP	0.819 (0.632-1)	0.824	0.750	0.732	0.824	0.824	0.750
	VP	0.716 (0.512-0.920)	0.824	0.667	0.625	0.800	0.778	0.727
	Combined	0.853 (0.687-1)	0.882	0.750	0.750	0.857	0.833	0.818
	Clinical-radiomic	0.892 (0.772-1)	0.824	0.833	0.828	0.849	0.875	0.769
External	NC	0.677 (0.532-0.823)	0.758	0.565	0.836	0.735	0.714	0.619
test	AP	0.829 (0.711-0.947)	0.788	0.652	0.901	0.776	0.765	0.682
	VP	0.672 (0.520-0.824)	0.576	0.696	0.758	0.644	0.731	0.533
	Combined	0.863 (0.769-0.957)	0.788	0.696	0.902	0.788	0.788	0.696
	Clinical-radiomic	0.892(0.809-0.975)	0.788	0.913	0.839	0.853	0.929	0.750

AUC Area under the receiver operating characteristic curve, *CI* Confidence interval, *NPV* Negative predictive value, *PPV* Positive predictive value

Table 5 Comparison of the LR models at different phases in external test set

Comparison	**Z statistic (95% CI)**	***p***	**NRI(Categorical) (95% CI)**	***p***	**NRI(Continuous) (95% CI)**	***p***	**IDI (95% CI)**	***p***
Combined VS. NC	2.1508 (0.0165-0.3551)	0.0315	0.2609 (-0.0724-0.5941)	0.1250	0.6298 (0.1371-1.1223)	0.0122	0.1956 (0.0485-0.3427)	0.0092
Combined VS. AP	1.1709 (-0.0231,0.0916)	0.2416	0.1436 (-0.0220-0.3092)	0.0891	1.0356 (0.591-1.4801)	<0.001	0.0626 (0.0107-0.1146)	0.0181
Combined VS. VP	2.1525 (0.0171-0.3650)	0.0314	0.3123 (-0.0270-0.6515)	0.0713	0.7589 (0.2650-1.2528)	0.0026	0.2120 (0.0761-0.3479)	0.0022
Clinical-radiomic VS. Combined	1.2025 (-0.0183-0.0762)	0.2292	0.1173 (-0.0368-0.2713)	0.1358	0.2477 (-0.2771-0.7725)	0.3550	0.0268 (-0.0215-0.0751)	0.2764

Figure legends

Figure 1 The flowchart of the patient recruitment process

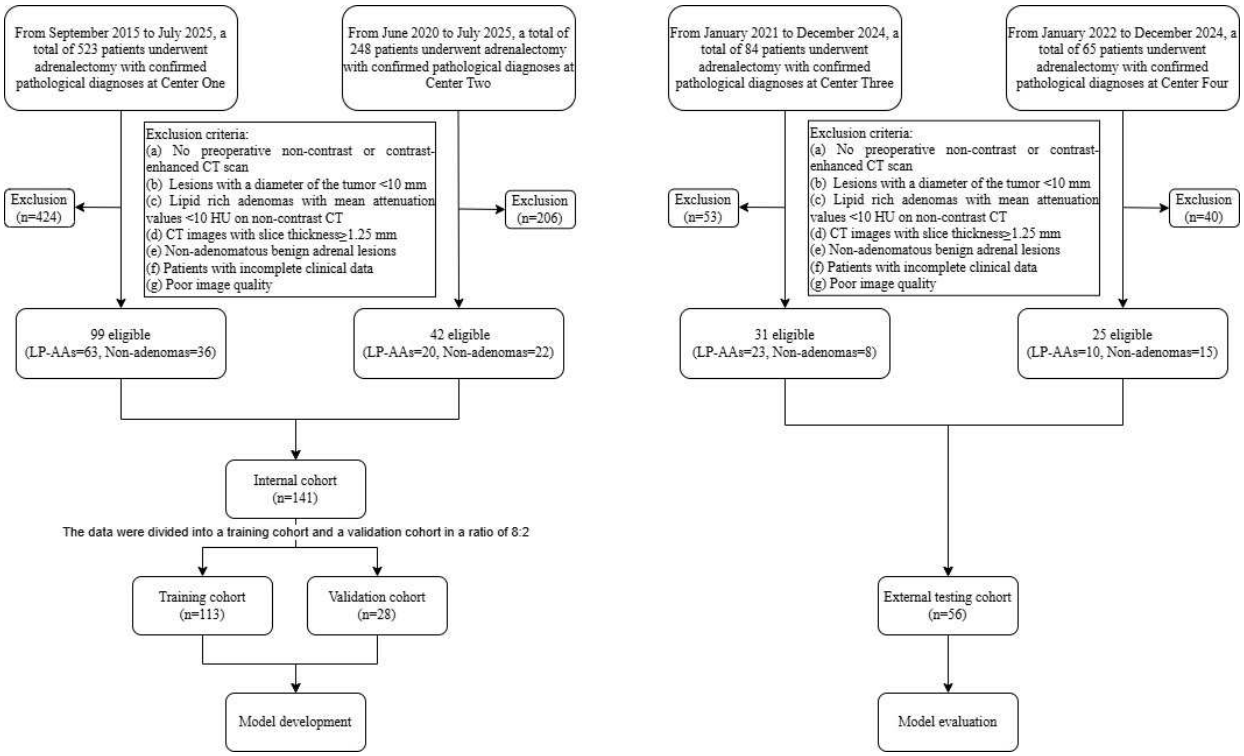


Figure 2 The workflow of the deep learning segmentation model

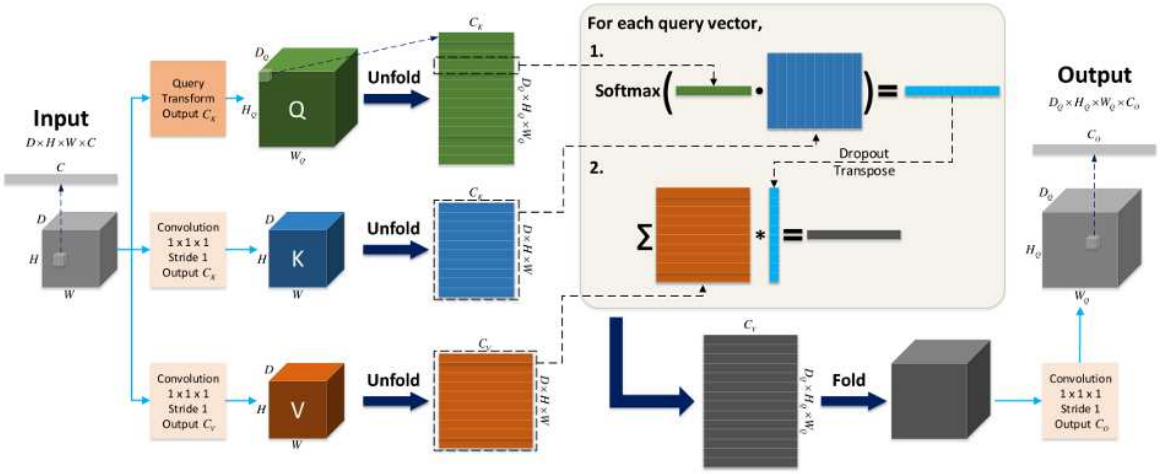
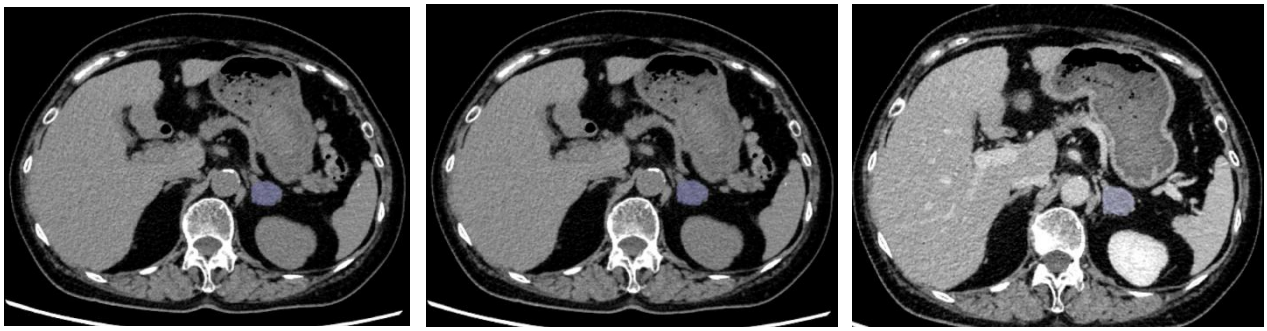


Figure 3 Representative segmentation results of the DL model. (a), (b), and (c) correspond to the unenhanced, arterial, and venous phase CT images of adrenal lipid-poor adenoma; (d), (f), (e) correspond to the unenhanced, arterial, and venous phase CT images of adrenal metastasis from clear cell renal cell carcinoma

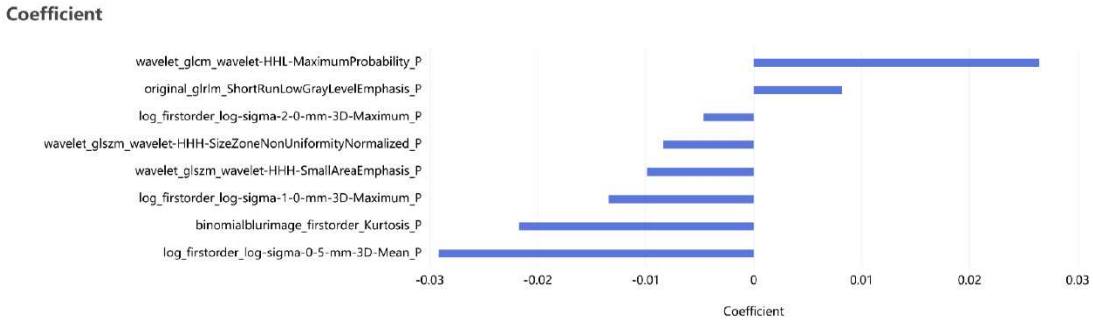


(a) (b) (c)

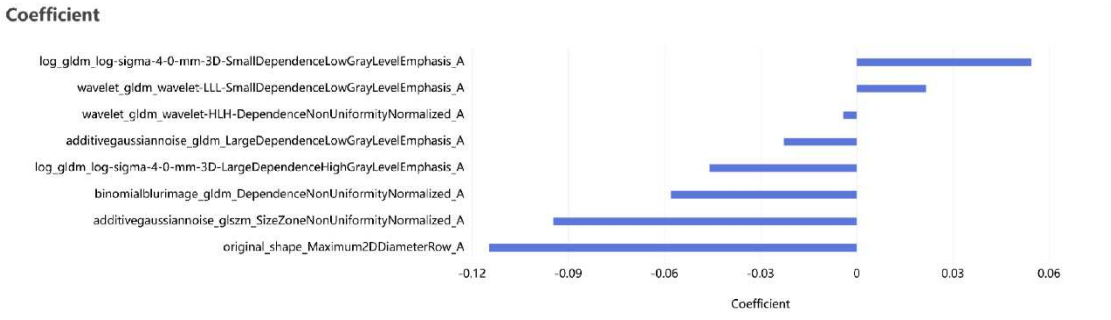


(d) (e) (f)

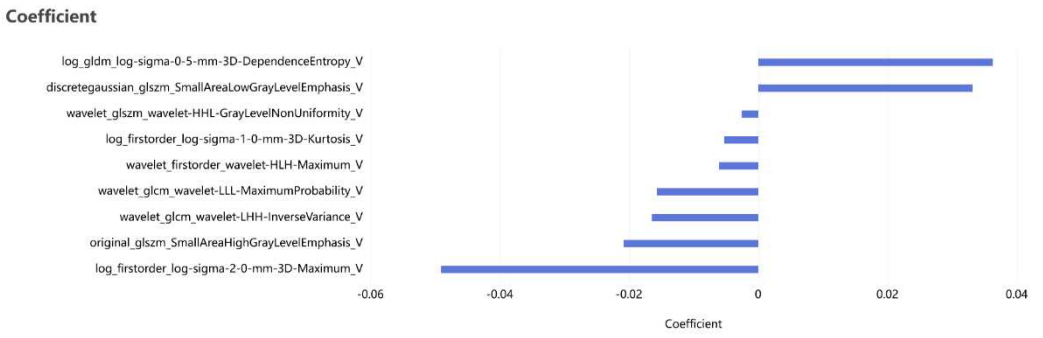
Figure 4 Feature coefficients of the radiomics models (a, b, c, and d correspond to the NC, AP, VP and combined models, respectively)



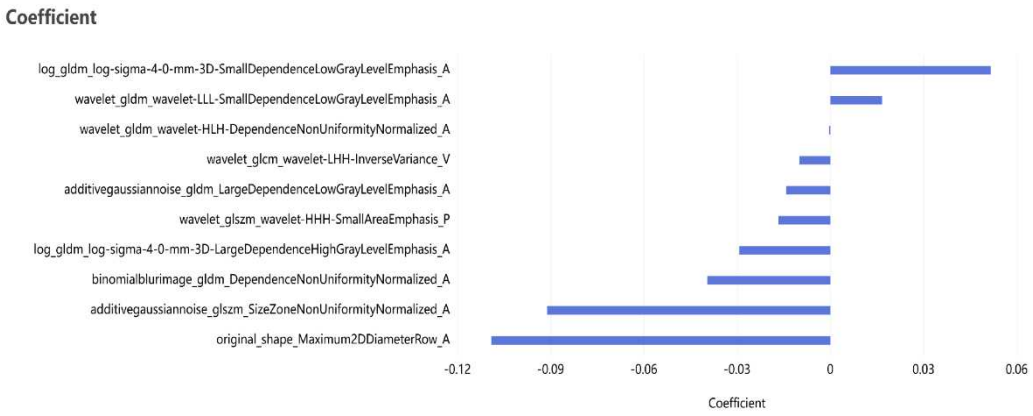
(a)



(b)



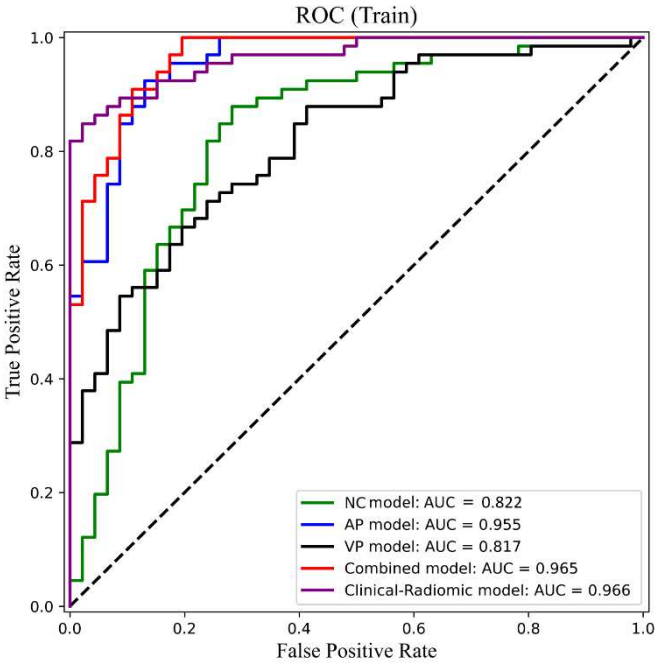
(c)


Coefficient
log_gldm_log-sigma-4-0-mm-3D-SmallDependenceLowGrayLevelEmphasis_A
wavelet_gldm_wavelet-LLL-SmallDependenceLowGrayLevelEmphasis_A
wavelet_gldm_wavelet-HLH-DependenceNonUniformityNormalized_A
wavelet_glcm_wavelet-LHH-InverseVariance_V
additivegaussiannoise_gldm_LargeDependenceLowGrayLevelEmphasis_A
wavelet_glszm_wavelet-HHH-SmallAreaEmphasis_P
log_gldm_log-sigma-4-0-mm-3D-LargeDependenceHighGrayLevelEmphasis_A
binomialblurimage_gldm_DependenceNonUniformityNormalized_A
additivegaussiannoise_glszm_SizeZoneNonUniformityNormalized_A
original_shape_Maximum2DDiameterRow_A
-0.12
-0.09
-0.06
-0.03
0
0.03
0.06
Coefficient

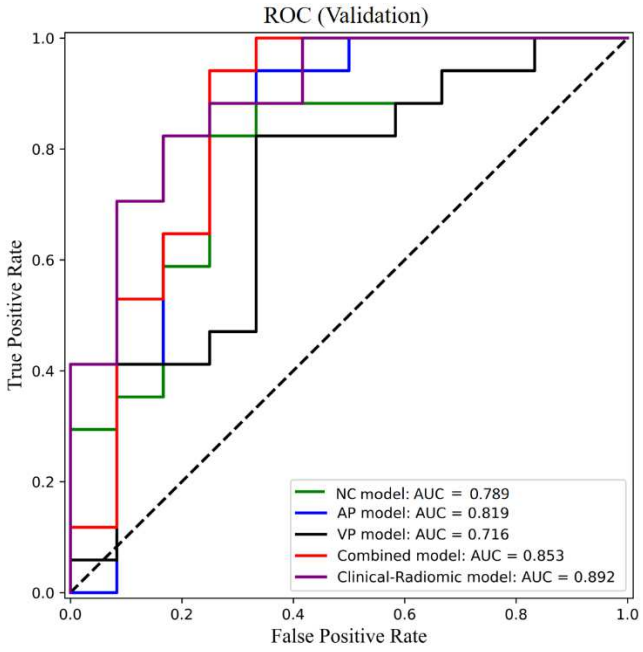
(d)

Figure 5 The ROC curves of the five LR models for the classification of LP-AA and non- adenoma in training group (a), validation (b), and external testing group (c).

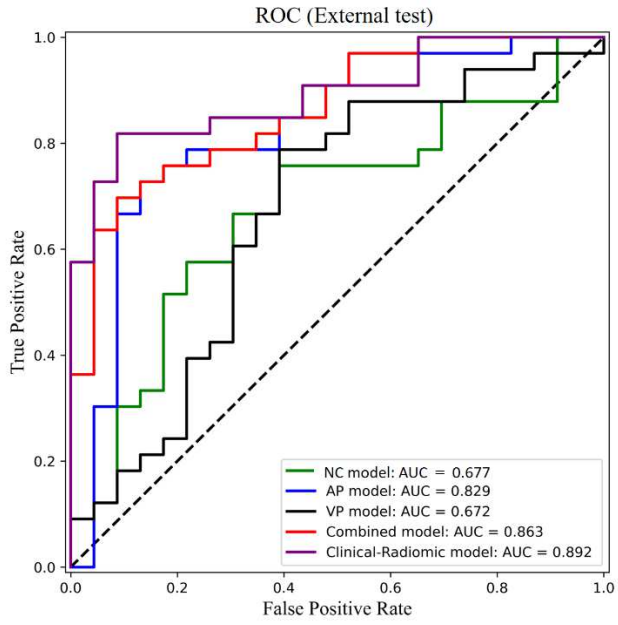
ROC, Receiver operating characteristic



(a)


ROC (Validation)
True Positive Rate
False Positive Rate
NC model: AUC = 0.789
AP model: AUC = 0.819
VP model: AUC = 0.716
Combined model: AUC = 0.853
Clinical-Radiomic model: AUC = 0.892

(b)


ROC (External test)
True Positive Rate
False Positive Rate
NC model: AUC = 0.677
AP model: AUC = 0.829
VP model: AUC = 0.672
Combined model: AUC = 0.863
Clinical-Radiomic model: AUC = 0.892

(c)

Supplementary Files

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

- Appendix.docx