












Radiomics-Enhanced 3D ResNet for Multi-Feature Fusion in Adrenal Tumor Diagnosis

Yuchen Liu¹, Shichu Jia², Haoxiang Wang³, Yue Ding³, Mengfei Yang³, Linyan Xue^{3,4,5},
Shuang Liu^{3,4,5}, Chunhui Liu^{6*}, Guojie Yang^{6,7*}

¹ International College, Hebei University, Baoding 071002, China

² School of Electronic and Information Engineering, Hebei University, Baoding 071002, China

³ College of Quality and Technical Supervision, Hebei University, Baoding 071002, China

⁴ Hebei Technology Innovation Center for Lightweight of New Energy Vehicle Power System, Baoding 071002, China

⁵ National & Local Joint Engineering Research Center of Metrology Instrument and System, Hebei University, Baoding 071002, China

⁶ Affiliated Hospital of Hebei University, Baoding 071002 China

⁷ Baoding Key Laboratory of Intelligent Diagnosis of Cardiovascular and Cerebrovascular Diseases, Baoding 071002 China

Corresponding Author Email: liuchs@hbu.edu.cn; fly.god@163.com

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420630>

ABSTRACT

Received: 17 May 2025

Revised: 29 October 2025

Accepted: 16 November 2025

Available online: 31 December 2025

Keywords:

adrenal tumors diagnosis, radiomics, 3D ResNet, feature fusion

Accurate discrimination between adrenal adenoma and metastasis is pivotal for precision treatment planning. However, traditional diagnostic protocols and single-modality feature extraction methods often fail to capture the full spectrum of discriminative information inherent in tumor imaging. To address this issue, we propose a multimodal fusion framework that synergistically integrates handcrafted radiomics with deep 3D representations to maximize lesion separability on contrast-enhanced CT images. Specifically, radiomics techniques are utilized to extract low-dimensional features (e.g., shape, texture, and first-order statistics) to capture the structural details of lesions. Simultaneously, a 3D ResNet combined with a Spatial-Shift MLP is employed to extract high-dimensional features, capturing deep semantic information. By fusing representations from these heterogeneous feature spaces and applying LASSO regression for robust feature selection, we identified a signature of 76 highly discriminative features. Experimental results on a clinical adrenal tumor dataset demonstrate that this method achieves superior diagnostic performance, with an AUC of 96.9%, an accuracy of 91.4%, and a specificity of 98.4%, significantly outperforming models based on radiomics or deep learning alone.

1. INTRODUCTION

Adrenal tumors are common lesions frequently identified incidentally during abdominal imaging for unrelated conditions, a clinical scenario known as 'incidentaloma'. Recent epidemiological data and large-scale retrospective studies indicate that the prevalence of adrenal masses in abdominal Computed Tomography (CT) scans is ranges from 3% to 7% in the adult population, with higher rates observed in the elder [1, 2]. While most of these lesions are benign adrenal adenomas requiring only periodic surveillance, a critical subset represents malignant metastases, particularly in patients with a known history of extra-adrenal malignancy. Consequently, the accurate differentiation between benign adenomas and malignant metastases is of paramount clinical importance, as it dictates divergent therapeutic trajectories—ranging from conservative management for benign lesions to aggressive systemic therapy or surgical intervention for malignancies [3].

Currently, the clinical diagnosis of adrenal lesions relies heavily on unenhanced and contrast-enhanced CT imaging, utilizing parameters such as tumor attenuation values (Hounsfield Units) and contrast washout rates. Although

effective for typical lipid-rich adenomas (<10 HU), these traditional radiological methods often struggle to distinguish lipid-poor adenomas from metastases due to significant overlap in their radiological features [4]. Recent research utilizing large cohorts has highlighted that relying solely on attenuation thresholds leads to suboptimal specificity for lipid-poor lesions, necessitating additional invasive workups or nuclear medicine imaging [5]. Furthermore, manual interpretation of these images is inherently subjective, prone to inter-observer variability, and heavily dependent on the radiologist's expertise. To address these diagnostic hurdles, Artificial Intelligence (AI) techniques have emerged as transformative tools, offering quantitative and reproducible assessments of tumor heterogeneity [1, 6].

Two primary AI paradigms have gained prominence in oncological imaging: Radiomics and Deep Learning (DL) [7, 8]. Radiomics involves the high-throughput extraction of handcrafted quantitative features—such as texture, shape, and first-order statistics—providing a detailed "genotype-phenotype" description often imperceptible to the human eye. This approach has demonstrated significant potential in differentiating lipid-poor adrenal adenomas from metastases and predicting metastatic potential in pheochromocytoma [9,

10]. Conversely, Deep Learning, particularly Convolutional Neural Networks (CNNs) like the 3D ResNet, excels at automatically learning hierarchical, high-level semantic features and spatial contexts directly from volumetric data [11, 12]. Recent studies have explicitly demonstrated the efficacy of two-stage 3D CNNs in automated adrenal nodule detection and segmentation, significantly reducing radiologist workload and improving incidentaloma management [13]. However, single-modality approaches exhibit distinct limitations: radiomics relies on predefined mathematical descriptors that may fail to capture abstract semantic information, while pure deep learning models can act as "black boxes" that sometimes overlook fine-grained, low-level textural details essential for tissue characterization [1, 14]. To address these interpretability challenges and feature limitations, recent research emphasizes multimodal fusion and explainable AI techniques (e.g., SHAP) [15, 16].

To address these limitations and leverage the complementary strengths of both approaches, this study proposes a novel multi-dimensional feature fusion framework for the precise diagnosis of adrenal tumors. This approach aligns with the emerging trend of multi-modal medical image fusion, which has shown superior performance in oncology by synergizing structural and semantic information. Specifically, our framework utilizes PyRadiomics to capture structural details and employs a 3D ResNet integrated with a Spatial-Shift MLP (S²-MLP) module to extract deep semantic information from volumetric CT sequences. The integration of MLP-based spatial mixing has recently proven effective in enhancing global receptive fields in medical imaging tasks. By fusing these diverse feature spaces and applying LASSO regression for rigorous feature selection to eliminate redundancy, our method aims to construct a comprehensive

diagnostic tool that significantly outperforms models based on radiomics or deep learning alone.

2. MULTI-DIMENSIONAL FEATURE EXTRACTION AND SELECTION

2.1 Model architecture

This study proposes an integrated framework combining radiomics and deep learning to enhance diagnostic accuracy for adrenal gland patients [17]. The schematic diagram of the proposed framework is illustrated in Figure 1. The overall architecture consists of three key modules: (1) Radiomics-based low-level feature extraction, (2) 3D ResNet-based high-level feature extraction, and (3) LASSO-based feature selection. Traditional methods often rely solely on deep learning, which tends to prioritize high-level semantic representations while overlooking vital low-level structural details [18]. Our multi-dimensional framework addresses this by capturing tumor characteristics across distinct feature dimensions. In the radiomics phase, CT images are processed using the pyradiomics library, where various filters are applied to enhance diverse image attributes. In the deep learning phase, a 3D ResNet network is employed to exploit the volumetric spatial information inherent in CT scans [19]. Patient slices are preprocessed via localization to standardize sequence lengths. Finally, to mitigate redundancy and eliminate irrelevant information within the high-dimensional feature space, LASSO regression is applied to shrink non-informative coefficients to zero. The resulting subset of filtered features is then fed into a classifier for precise lesion diagnosis.

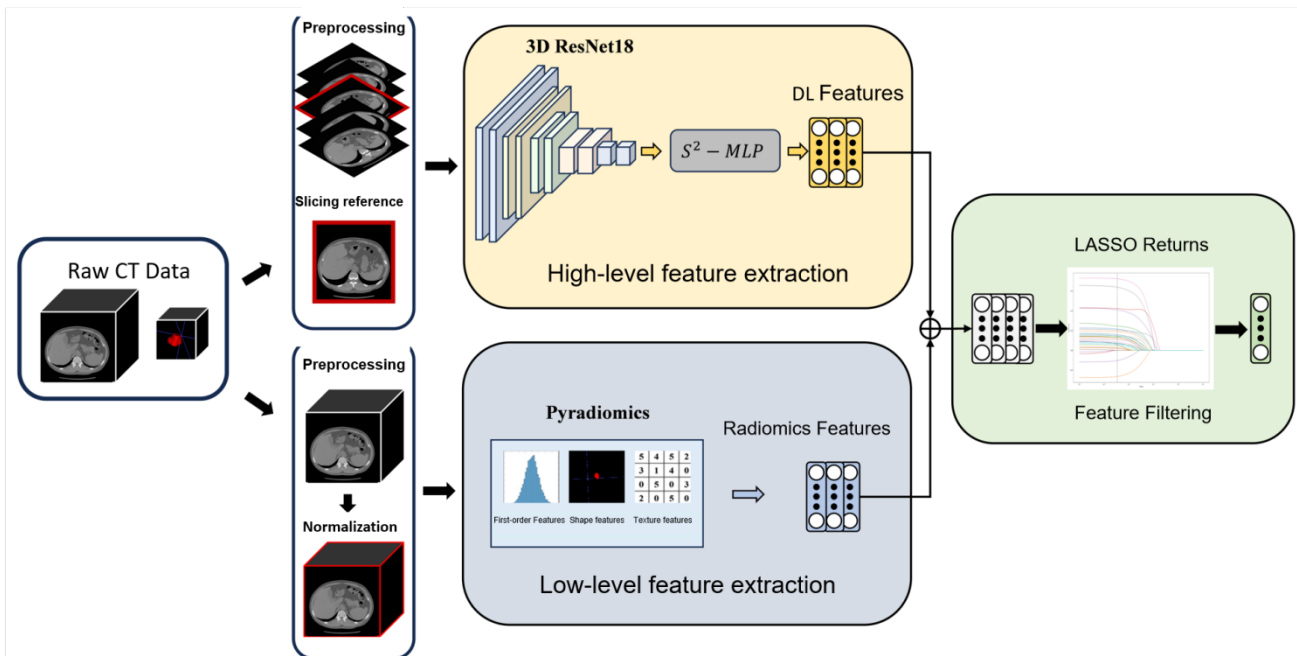


Figure 1. The overall model of multi-dimensional feature extraction

2.2 Radiomics feature extraction

Radiomics enables the high-throughput extraction of quantitative features from medical images, converting visual data into mineable datasets [20, 21]. We utilized the pyradiomics library to extract features from the Region of

Interest (ROI) in CT images. To fully extract the features of the patient's lesions, we set up a variety of filters. The filters are mainly used to enhance specific features of the image to help extract richer texture features. Different filters have different feature extraction effects. The specific screening process is shown in Figure 2. We applied nine types of filters:

Original, Wavelet (for local texture), LoG (Laplacian of Gaussian, for edge enhancement), Square, SquareRoot, Logarithm, Exponential, Gradient (for structural changes), and LBP (Local Binary Pattern) [17, 21]. These filters generated seven categories of radiomics features, totaling 1,345 features (298 first-order, 17 shape, and 1,030 texture features). The texture features include Gray Level Co-occurrence Matrix (GLCM), Gray Level Size Zone Matrix (GLSZM), Gray Level Run Length Matrix (GLRLM), Neighboring Gray Tone Difference Matrix (NGTDM), and Gray Level Dependence Matrix (GLDM) [8, 16].

First-order features describe the voxel intensity distribution. For instance, Entropy is defined as:

$$entropy = -\sum_{i=1}^{N_g} p(i) \log_2(p(i) + \epsilon) \quad (1)$$

where, $p(i)$ denotes the normalized first-order histogram, N_g represents the number of voxels, and $\epsilon \approx 2.2 \times 10^{-16}$. The entropy metric primarily measures average information, reflecting the uncertainty and randomness of image values.

Shape features describe the 3D geometry of the ROI, independent of gray intensity [16]. The surface area is calculated using a triangular mesh representation, where the total surface area A is the sum of the areas of all triangles in the mesh, indicating the model's complexity.

$$A_i = \frac{1}{2} |a_i b_i \times a_i c_i| \quad (2)$$

$$A = \sum_{i=1}^{N_f} A_i \quad (3)$$

In a triangular mesh, $a_i b_i$ and $a_i c_i$ represent the two edges of the i triangle, with its vertices defined by a_i , b_i , and c_i . The total number of triangles in the mesh, N_f , indicates the mesh's complexity. A denser mesh corresponds to a more detailed 3D model, and the total surface area A is calculated by summing the areas of all triangles.

Finally, we extract 1345 features including 298 first-order features, 17 shape features and 1030 texture features, which are used as the low-level features of the model [17].

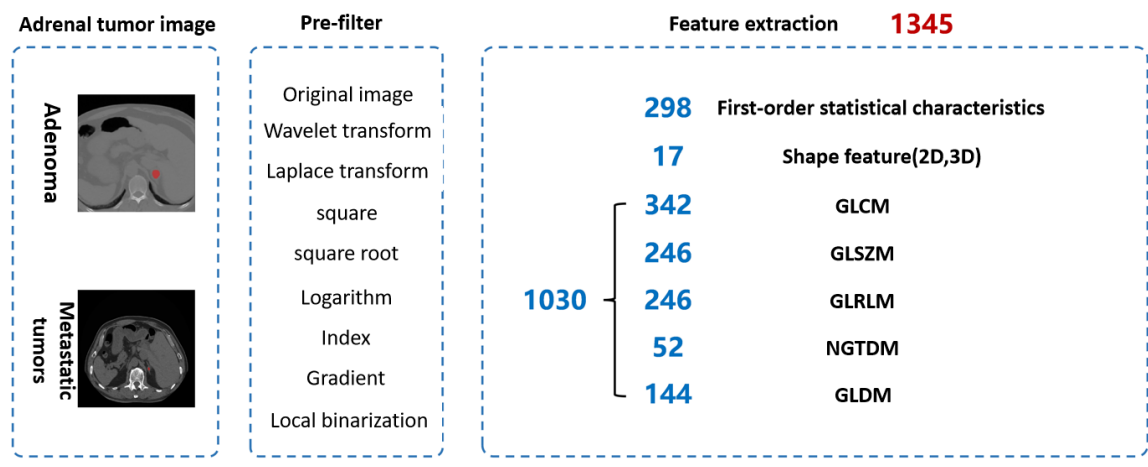


Figure 2. Low-dimensional feature extraction process

2.3 3D ResNet combined with spatially displaced MLP for high-dimensional feature extraction

3D ResNet is a powerful feature extraction neural network that enhances traditional ResNet by incorporating temporal dimensions to better capture spatial information in 3D data [22]. To ensure consistency, we preprocessed CT images of adrenal patients by precisely locating lesion regions in all cases. Using the lesion center slice as a reference point, we expanded 20 slices vertically. This preprocessing standardized the final CT sequence length to 20 slices per patient. Our study innovates in feature extraction by integrating deep learning with radiomics techniques, enabling multi-dimensional feature extraction to allow the model to learn richer feature representations [17, 21]. Adopting a preprocessing strategy similar to the radiomics pipeline, we converted the 3D CT sequences $V \in \mathbb{R}^{D \times W \times H \times C}$ (where D denotes image depth, W width, H length, and C channels) into uniformly sized voxel blocks P dimensions $d \times w \times h \times c$, which were input into the network for feature extraction. The voxel block segmentation formula is shown in Eq. (4):

$$P = \{P_{ijk}\}_{i=1,j=1,k=1}^{d,w,h} \quad (4)$$

The range of i is from 1 to d , j from 1 to w , and k from 1 to

h .

To overcome the limited tissue, contrast inherent in CT imaging, we incorporate a Spatial-Shift MLP $S^2 - MLP$ module for spatial channel mixing. $S^2 - MLP$ primarily adopts a spatial feature reorganization strategy, designing channels into four groups including the image space directions of up, down, left, and right. We expand this to six groups by adding two sagittal plane directions to the original four. Specifically, first, the extracted 3D feature tensor is arranged into four groups along the channel dimension, performing right and left shifts in the coronal plane, as well as upward and downward shifts in the sagittal plane, forming feature offsets across anatomical levels to achieve spatial dimensionality reorganization of feature vectors [23]. Subsequently, a Multi-Layer Perceptron (MLP) with quadruple expanded hidden layers is employed to construct high-order feature interactions, expanding into vectors and projecting to vectors e_i , as shown in formulas (5) and (6):

$$p_{ijk} = Flatten(P_{ijk}) \in \mathbb{R}^{C \cdot tp^2} \quad (5)$$

$$e_{ijk} = LN(W_0 p_{ijk} + b_0) \in \mathbb{R}^C \quad (6)$$

In this context, $\mathbb{R}^{C \cdot tp^2}$ denotes the dimensionality of the deep learning block, where t represents the block size and

p indicates the spatial direction block size. The learnable projection matrix W_0 and the bias vector b_0 are defined accordingly. The specific LN -layer normalization formula is given in Eq. (7):

$$\hat{e}_i = \gamma_i \frac{e_i - \mu}{\sigma} + \beta_i, \forall i \in [1, c] \quad (7)$$

where, $\mu = \frac{1}{c} \sum_{i=1}^c e_i$, $\sigma = \sqrt{\frac{1}{c} \sum_{i=1}^c (e_i - \mu)^2 + \varepsilon}$, where ε is a constant, and γ and β are learnable parameters.

The architecture of the high-dimensional feature extraction module is illustrated in Figure 3. We first apply conventional 3D convolution in the first two residual stages of the 3D ResNet network, while replacing some convolution blocks with $S^2 - MLP$ blocks in the third residual stage. This spatial communication mechanism enhances diagnostic accuracy.

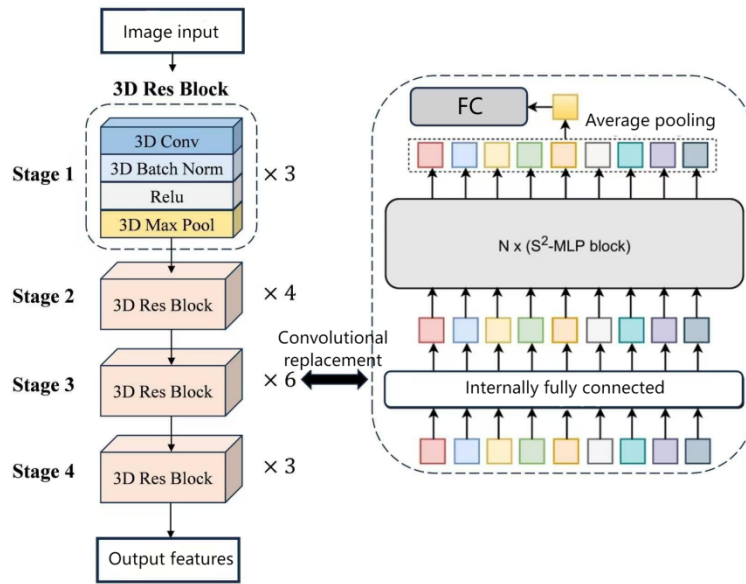


Figure 3. Network diagram of the high-dimensional feature extraction module

In stark contrast to radiomics features, deep learning features undergo nonlinear mapping transformations throughout the entire process of connecting neural network input and output spaces. Given that radiomics extracts low-level features with a dimensionality of 1345, we implemented a full connection layer (FC) to reduce the dimensionality of deep learning features extracted by 3D ResNet to 1024 through linear transformations. Finally, we concatenated the radiomics features and deep learning features obtained for each patient, resulting in a combined feature dimensionality of 2369 per patient for subsequent feature screening [17].

2.4 LASSO feature selection strategy

Through a strategy combining high-level and low-level features, we extracted multidimensional features including 1346-dimensional radiomics features and 1024-dimensional deep learning features [17]. Given potential redundancy or interference among features, we implemented feature screening to retain those critical for lesion diagnosis. The process utilized LASSO regression, a linear regression method with L1 norm regularization that introduces penalty terms in the loss function to shrink regression coefficients to zero [11, 21]. Since LASSO demonstrates limited effectiveness in

The computational formula for the 3D convolution kernel is given in Eq. (8):

$$O(x, y, z) = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \sum_{m=0}^{k-1} I_{padded}(x+i, y+i, z+m) \cdot K(i, j, m) \quad (8)$$

The variables $O(x, y, z)$ represent the output feature map, $I_{padded}(x+i, y+i, z+m)$ denote the padded input data, $K(i, j, m)$ denote the 3D convolution kernels, and k represents the kernel size. The 3D residual block primarily consists of 3D convolution, BN, ReLU, and skip connections [22]. Finally, the feature map undergoes global pooling (GAP) to compress the $2048 \times D \times H \times W$ feature map into a fixed-length 2048-dimensional feature vector, which can be further utilized for classification.

identifying highly repetitive features, we combined it with T-test statistical methods to compare differences between experimental and control groups, ensuring retention of statistically significant variations [21, 24].

Specifically, we first apply Z-score normalization to the extracted multi-dimensional features f as part of data preprocessing. This transforms features from different dimensions into a unified standardized scale, setting the mean to 0 and the standard deviation to 1. The formula for Z-score normalization is given in Eq. (9):

$$f' = \frac{f - \mu}{\sigma} \quad (9)$$

where, f represents the original multi-dimensional feature, μ denotes its mean, σ is the standard deviation, and f' stands for the standardized feature value. We then perform a t-test on the standardized feature f' . The null hypothesis H_0 states that there is no significant difference in the mean of this feature between the experimental and control groups, i.e., $\mu_1 = \mu_2$. The alternative hypothesis H_1 asserts that the means of the two groups differ significantly, i.e., $\mu_1 \neq \mu_2$. The t-test calculation formula is provided in Eq. (10):

$$t = \frac{\bar{f}'_1 - \bar{f}'_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (10)$$

where, n_1 and n_2 denote the sample sizes of the experimental and control groups respectively, while \bar{f}'_1 and \bar{f}'_2 represent their standardized features. The pooled standard deviation S_p is calculated as shown in Eq. (11):

$$S_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \quad (11)$$

where, s_1 and s_2 denote the standard deviations of the experimental and control groups, respectively. After obtaining the T-statistic, we directly compute the p-value, retaining features with p-values below the predefined significance level of 0.05. Through these screening methods, we ultimately identified 76 independent features of significant importance for classification, including 40 radiomics features and 36 deep learning features, which serve as input features for the subsequent classifier [17].

Table 1. Distribution of the adrenal tumor dataset used in this study

Affiliated Hospital	Patients with Adenomas	Patients with Metastatic Tumors	Total Number of Cases	DICOM	PNG
Participating Center 1	101	92	193	193	1938
Participating Center 2	31	40	71	71	1052
Total	132	132	264	264	2990

4. EXPERIMENTAL RESULTS AND ANALYSIS

The comparative and ablation experiments were conducted exclusively on the dataset from Participating Center 1, whereas the external validation required data from two collaborating hospitals. Specifically, Participating Center 1 served as the internal validation dataset, and Participating Center 2 was designated as the external validation dataset.

4.1 Comparison of experimental results and analysis

To validate the effectiveness of the multi-dimensional feature extraction method proposed in this study, we conducted comparative experiments between the classification model developed in this chapter and multiple deep learning and machine learning classification algorithms. The deep learning algorithms included 3D ResNet34, 3D ResNet50, 3D DenseNet121, 3D VGG-11, 3D VGG-16, C3D, and 3D EfficientNet [25, 26], while the machine learning algorithms comprised SVM, MLP, LR, and RF [20]. All comparison algorithms were trained on the dataset constructed in this study, with hyperparameters such as training epochs and learning rate kept consistent to ensure the validity and fairness of the experiments.

In Chapter 2, we developed classification evaluation metrics to comprehensively assess the performance of multi-dimensional feature extraction methods against classical classification models. The evaluation covered five key metrics: AUC, accuracy, sensitivity, precision, and specificity [27]. Optimal values are highlighted in bold, with comparative results presented in Table 2.

As shown in Table 2, the multi-dimensional feature extraction network proposed in this study achieves optimal performance across nearly all metrics. First, in terms of AUC, 3D ResNet34 achieves the highest AUC of 92.0 among

3. DATA SET INTRODUCTION

The raw data include retrospective adrenal CT scan data obtained from affiliated hospitals (Participating Center 1) and the central hospital (Participating Center 2) between January 2019 and November 2022. Participating Center 1 used a GE Discovery HD750 64-slice spiral CT scanner with the following parameters: slice thickness 5mm, pitch 0.992, field of view 350 mm × 350 mm, matrix 512 × 512, tube voltage 100-120kV, and tube current 160-300mA, totaling 1,938 CT slices with regions of interest (ROI) from 101 patients with lipophilic adrenal adenoma and 92 patients with adrenal metastatic tumors. Participating Center 2 used a Philips 256-slice Brilliance iCT scanner (Philips Healthcare; Netherlands) for abdominal CT scans, both plain and enhanced, with the following parameters: tube voltage 120kV, automatic tube current modulation, collimated 128 × 0.625 mm, pitch 0.914, matrix 512 × 512, slice thickness 5mm, and slice interval 5mm, totaling 1,052 CT clinical slices with ROI from 31 patients with lipophilic adrenal adenoma and 40 patients with adrenal metastatic tumors. The data is shown in Table 1.

classical classification networks, while our network improves it by 4.9%. Regarding accuracy, 3D ResNet34 also reaches the highest accuracy of 88.8%, but our network enhances it by 2.6%. For sensitivity, Random Forest RF achieves the highest value of 96.0% among classical models, significantly outperforming our network. This is primarily due to Random Forest RF's overemphasis on positive sample features, resulting in strong recognition capability for positive samples but only 59.2% specificity for negative samples, leading to unstable performance [25]. In precision, 3D ResNet34 maintains stable classification performance with a peak of 88.4%, while our network improves it by 4.4%. Finally, in specificity, our network achieves a peak value of 98.4% across all experiments. Across all evaluated metrics, our network demonstrates superior classification performance, with stable recognition capabilities for both positive and negative samples, and effectively balances feature learning between them.

Table 2. Comparison of different network experiment results

Model	AUC (%)	Accuracy (%)	Sensitivity (%)	Precision (%)	Specificity (%)
SVM	88.7	83.8	88.8	79.9	78.4
MLP	87.3	84.9	87.4	83.2	84.3
LR	87.3	84.9	87.4	83.2	84.3
RF	84.1	77.3	96.0	69.6	59.2
3D ResNet34	92.0	88.8	92.4	88.4	85.0
3D ResNet50	87.6	71.8	68.5	77.3	68.5
3D DenseNet121	88.2	80.3	81.2	81.0	81.2
3D VGG-11	88.7	82.0	82.7	82.3	82.7
3D VGG-16	86.0	82.0	83.1	83.1	83.1
3D EfficientNet	86.9	85.2	88.8	83.3	81.6
C3D	83.5	78.8	87.3	75.5	72.0
Ours	96.9	91.4	90.3	92.8	98.4

In conclusion, the comparative experiment results show the effectiveness of our proposed method. Compared with other classification networks, our method can learn on more feature dimensions to enrich the diversity of features, including high-level and low-level features, and effectively make the model learn more discriminative features [21].

To visually assess the model's performance, we plotted the ROC curves of both our proposed multi-dimensional feature extraction network and the best-performing 3D ResNet34 in the classical network. The ROC curve, a receiver operating characteristic curve, is primarily used to calculate the area under the curve (AUC). As shown in Figure 4, our model demonstrates a higher AUC compared to 3D ResNet34, indicating that the multi-dimensional feature extraction network outperforms 3D ResNet34 in distinguishing positive and negative samples across different thresholds [27].

Furthermore, we generated confusion matrix diagrams for all comparison networks. These matrices visually demonstrate the model's prediction outcomes across different categories, including true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). As shown in Figure 5, our proposed multi-dimensional feature extraction network achieves higher recall rates for negative samples, enabling more comprehensive category identification while

maintaining balanced recognition between positive and negative samples. Specifically, this network demonstrates superior accuracy in distinguishing positive and negative samples, with reduced error diagnosis probabilities and lower missed detection rates, thereby validating the effectiveness of our proposed methodology.

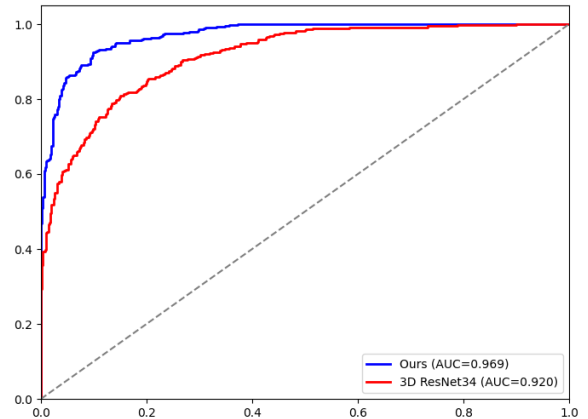


Figure 4. ROC curve of multi-dimensional feature extraction network and 3D ResNet34

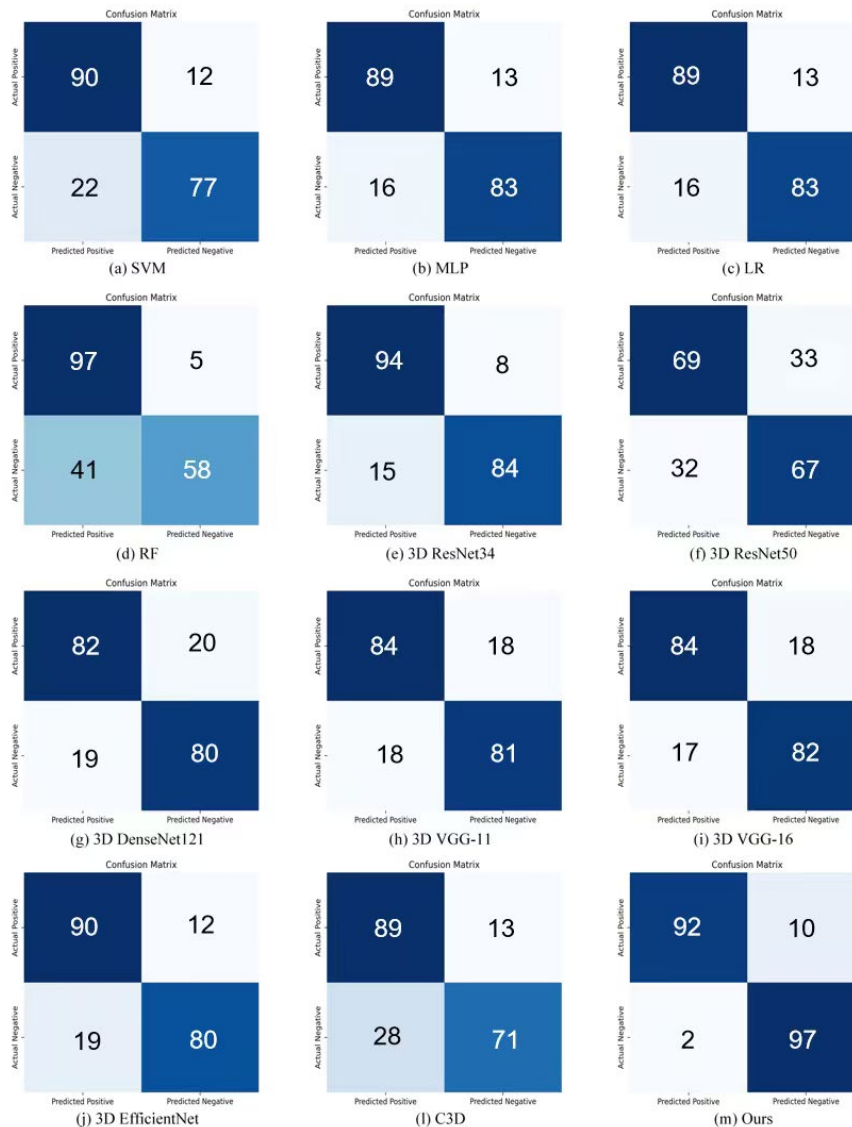


Figure 5. Confusion matrix diagram of network comparison

4.2 Results and analysis of the ablation experiment

To evaluate the effectiveness of combining high-level and low-level feature extraction modules in multi-dimensional feature extraction networks, we conducted an ablation experiment on adrenal adenoma versus adrenal metastasis classification [3, 20]. Table 3 presents comparative results of three variants to explore the synergistic effects of radiomics feature extraction and deep learning feature extraction. Specifically: (1) The first row shows models using only radiomics features combined with machine learning classification algorithms. We compared four machine learning models (SVM, MLP, LR, and RF from Table 2) while employing radiomics features. The SVM-based approach demonstrated optimal performance, which we adopted as the experimental benchmark [6, 7]. (2) The second row utilizes 3D ResNet exclusively as the high-level feature extraction network without incorporating low-level features [1]. (3) The third row represents the complete multi-dimensional feature extraction network that integrates low-dimensional radiomics

features with high-dimensional 3D ResNet features. Results clearly demonstrate the critical role of both low-level and high-level feature extraction modules. Specifically, using only radiomics for low-level feature extraction achieved an AUC of 88.7%, while employing 3D ResNet for high-level feature extraction alone yielded an AUC of 92.0%. However, directly combining high-dimensional and low-dimensional features resulted in reduced model performance due to excessive overall feature dimensions and information redundancy, achieving only an 89.7% AUC. By integrating both modules with feature selection, the model attained peak performance across all classification metrics [14]. Although the fused model's sensitivity marginally underperformed high-level feature extraction, its overall performance reached optimal levels. We hypothesize this stems from 3D ResNet's overemphasis on positive sample features, while its negative sample recognition specificity remained at 85.0%. This discrepancy explains why the high-level feature extraction module demonstrated slightly higher sensitivity compared to multi-level feature extraction.

Table 3. Results of the ablation experiment analysis

Module			Evaluating Indicator				
Low Dimensional Feature Extraction	High Dimensional Feature Extraction	Feature Screening	AUC (%)	Accuracy (%)	Sensitivity (%)	Precision (%)	Specificity (%)
√			88.7	83.8	88.8	79.9	78.4
	√		92.0	88.8	92.4	88.4	85.0
√	√		89.7	84.9	75.6	77.6	94.3
√	√	√	96.9	91.4	90.3	92.8	98.4

4.3 External validation results and analysis

To evaluate the generalization capability of the multidimensional feature extraction network proposed in this study and verify its independence from the distribution of single-center medical data, we conducted external validation at Collaborative Hospital 2 [11, 28]. Using Collaborative Hospital 1's dataset as the training set and Hospital 2's dataset as an independent test set, we performed external validation using the same network from the previous comparative experiment. The results were analyzed to compare the performance of different classification networks on the external validation set. The experimental results are presented in Table 4.

Table 4. Comparison of different network external validation experiments

Model	AUC (%)	Accuracy (%)	Sensitivity (%)	Precision (%)	Specificity (%)
SVM	59.3	56.9	42.5	70.2	76.0
MLP	64.5	46.8	43.2	64.5	50.4
LR	63.5	61.0	66.8	66.8	55.3
RF	66.0	47.4	46.6	57.3	48.2
3D ResNet34	63.3	61.2	49.1	74.4	77.5
3D ResNet50	66.0	68.7	84.6	68.2	47.4
3D	69.2	59.5	74.3	57.9	46.4
DenseNet121					
3D VGG-11	63.6	53.5	40.5	47.5	66.6
3D VGG-16	53.7	56.2	49.3	57.5	63.2
3D	58.9	64.7	60.8	53.6	68.5
EfficientNet					
C3D	68.6	42.2	42.6	58.4	41.9
Ours	69.5	74.6	71.5	70.1	77.6

Table 4 experimental results demonstrate that our proposed multi-dimensional feature extraction network achieved optimal performance across nearly all classification metrics. The network attained peak values of 69.5% AUC, 74.6% accuracy, and 77.6% specificity. While the network exhibited a reduction in sensitivity and precision, its overall performance across key classification metrics confirmed superior generalization capabilities compared to baseline models when applied to multi-center datasets. Notably, all classification networks exhibited mild to severe performance degradation during external validation, indicating overfitting to Participating Center 1 data [1, 21]. When encountering Participating Center 2 data with different distribution patterns, the model failed to acquire stable generalizable features, resulting in significant performance deterioration. Consequently, future research will prioritize enhancing the model's generalization capacity to mitigate performance degradation caused by cross-center data variations [22].

5. CONCLUSION

This study presented a multi-dimensional feature extraction framework for adrenal tumor diagnosis. By synergizing low-level radiomics features with high-level 3D deep learning representations, the model effectively captures comprehensive tumor characteristics. Extensive experiments validated that this fusion strategy, combined with rigorous feature selection, significantly outperforms conventional methods in both accuracy and specificity. Future work will prioritize the integration of unsupervised domain adaptation techniques to mitigate distribution shifts and further enhance model robustness across multi-center cohorts.

ACKNOWLEDGMENT

This work was supported by the Xiong'an New Area Science and Technology Innovation Special Project of the Ministry of Science and Technology (2023XAGG0085), the Natural Science Foundation of Hebei Province (F2024201052), Baoding Science and Technology Plan Project (2541ZF267), and the Hebei University Affiliated Hospital/Clinical College Internal Fund Project (2022QC54).

REFERENCES

- [1] Rojak, J.A., Khayru, R.K. (2022). Disparities in access to education in developing countries: Determinants, impacts, and solution strategies. *Journal of Social Science Studies*, 2(1): 31-38.
- [2] Ahmed, T.M., Rowe, S.P., Fishman, E.K., Soyer, P., Chu, L.C. (2024). Three-dimensional CT cinematic rendering of adrenal masses: Role in tumor analysis and management. *Diagnostic and Interventional Imaging*, 105(1): 5-14. <https://doi.org/10.1016/j.diii.2023.09.004>
- [3] Ren, S., Zheng, Y., Feng, Y., Wang, D. (2024). 693P Causes of death in patients with malignant adrenal tumors: A population-based analysis. *Annals of Oncology*, 35(s2): S537. <https://doi.org/10.1016/j.annonc.2024.08.756>
- [4] Pacak, K., Blake, M.A., Sweeney, A.T., Jha, A., Imperiale, A., Zaidi, H., Taïeb, D. (2025). Adrenal tumour imaging: Clinical, molecular, and radiomics perspectives. *The Lancet Diabetes & Endocrinology*, 14(1): 62-81. [https://doi.org/10.1016/S2213-8587\(25\)00300-6](https://doi.org/10.1016/S2213-8587(25)00300-6)
- [5] McPhedran, R.L., Navin, P.J. (2025). Nuclear imaging in the assessment of adrenal tumors. *Radiologic Clinics*, 63(6): 861-871. <https://doi.org/10.1016/j.rcl.2025.03.014>
- [6] Zhao, K., Sun, Z., Jiang, H., Zou, Z., Wu, Q., Xin, Y., Jiang, H. (2025). Transformer-based integration of radiomics and deep learning for differentiating lipid-poor adrenal adenomas from malignant tumors. *Meta-Radiology*, 3(4): 100183. <https://doi.org/10.1016/j.metrad.2025.100183>
- [7] Ahn, C.H., Kim, T., Jo, K., Park, S.S., Kim, M.J., Yoon, J.W., Choo, J. (2025). Two-stage deep learning model for adrenal nodule detection on CT images: A retrospective study. *Radiology*, 314(3): e231650. <https://doi.org/10.1148/radiol.231650>
- [8] Gairola, A.K., Kumar, V., Singh, G.D., Bajaj, M., Rathore, R.S., Mahmoud, M.H., El-Shafai, W. (2024). Efficient deep learning fusion-based approach for brain tumor diagnosis. *Traitement du Signal*, 41(5): 2573-2584. <https://doi.org/10.18280/ts.410530>
- [9] Wang, Y., Su, Y., Li, J., Xie, D., Liu, Z., Cai, Y., Yuan, Y. (2025). Machine learning-based differentiation of benign and malignant adrenal lesions using 18F-FDG PET/CT: a two-stage classification and SHAP interpretation study. *BMC Cancer*, 25(1): 1726. <https://doi.org/10.1186/s12885-025-15243-0>
- [10] Chen, Q., Wang, L., Deng, Z., Wang, R., Wang, L., Jian, C., Zhu, Y.M. (2025). Cooperative multi-task learning and interpretable image biomarkers for glioma grading and molecular subtyping. *Medical Image Analysis*, 101: 103435. <https://doi.org/10.1016/j.media.2024.103435>
- [11] Ma, M., Gu, W., Liang, Y., Han, X., Zhang, M., Xu, M., Huang, D. (2024). A novel model for predicting postoperative liver metastasis in R0 resected pancreatic neuroendocrine tumors: integrating computational pathology and deep learning-radiomics. *Journal of Translational Medicine*, 22(1): 768. <https://doi.org/10.1186/s12967-024-05449-4>
- [12] Bhasin, M., Jain, S., Hoda, F., Dureja, A., Dureja, A., Rathor, R.S., Aldosary, S., El-Shafai, W. (2024). Unveiling the hidden: Leveraging medical imaging data for enhanced brain tumor detection using CNN architectures. *Traitement du Signal*, 41(3): 1575-1582. <https://doi.org/10.18280/ts.410345>
- [13] Chen, Y., Liu, A., Liu, Y., He, Z., Liu, C., Chen, X. (2024). Multi-dimensional medical image fusion with complex sparse representation. *IEEE Transactions on Biomedical Engineering*, 71(9): 2728-2739. <https://doi.org/10.1109/TBME.2024.3391314>
- [14] Bai, Z., Osman, M., Brendel, M., Tangen, C.M., Flaig, T.W., Thompson, I.M., Wang, F. (2025). Predicting response to neoadjuvant chemotherapy in muscle-invasive bladder cancer via interpretable multimodal deep learning. *npj Digital Medicine*, 8(1): 174. <https://doi.org/10.1038/s41746-025-01560-y>
- [15] Meng, M., Gu, B., Fulham, M., Song, S., Feng, D., Bi, L., Kim, J. (2024). Adaptive segmentation-to-survival learning for survival prediction from multi-modality medical images. *NPJ Precision Oncology*, 8(1): 232. <https://doi.org/10.1038/s41698-024-00690-y>
- [16] Gong, J., Lu, J., Zhang, W., Huang, W., Li, J., Yang, Z., Zhao, L. (2024). A CT-based subregional radiomics nomogram for predicting local recurrence-free survival in esophageal squamous cell cancer patients treated by definitive chemoradiotherapy: A multicenter study. *Journal of Translational Medicine*, 22(1): 1108. <https://doi.org/10.1186/s12967-024-05897-y>
- [17] Jiang, C., Qian, C., Jiang, Q., Zhou, H., Jiang, Z., Teng, Y., Tian, R. (2025). Virtual biopsy for non-invasive identification of follicular lymphoma histologic transformation using radiomics-based imaging biomarker from PET/CT. *BMC Medicine*, 23(1): 49. <https://doi.org/10.1186/s12916-025-03893-7>
- [18] Huang, H., Pedrycz, W., Hirota, K., Yan, F. (2025). A multiview-slice feature fusion network for early diagnosis of Alzheimer's disease with structural MRI images. *Information Fusion*, 119: 103010. <https://doi.org/10.1016/j.inffus.2025.103010>
- [19] Kang, X., Yao, L., Huang, Y., Li, S., Chen, Z., Lin, Z., Chen, X. (2025). Multi-phase feature-aligned fusion model for automated colorectal cancer segmentation in contrast-enhanced CT scans. *Expert Systems with Applications*, 284: 127727. <https://doi.org/10.1016/j.eswa.2025.127727>
- [20] Lopez-Ramirez, F., Soleimani, S., Azadi, J.R., Sheth, S., Kawamoto, S., Javed, A.A., Chu, L.C. (2025). Radiomics machine learning algorithm facilitates detection of small pancreatic neuroendocrine tumors on CT. *Diagnostic and Interventional Imaging*, 106(1): 28-40. <https://doi.org/10.1016/j.diii.2024.08.003>
- [21] Li, J., Lv, D., Guo, Z., Zhou, H., Yao, X., Rong, Y., Shuang, W. (2025). Integration of multi-scale radiomics and deep learning for Ki-67 prediction in clear cell renal carcinoma. *NPJ Precision Oncology*. <https://doi.org/10.1038/s41698-025-01214-y>

- [22] Maqsood, H., Khan, S.U.R. (2025). MeD-3D: A multimodal deep learning framework for precise recurrence prediction in clear cell renal cell carcinoma (ccRCC). *Expert Systems with Applications*, 299: 130174. <https://doi.org/10.1016/j.eswa.2025.130174>
- [23] Guo, Q., Wang, Y., Zhang, Y., Qi, H., Hu, Y., Jiang, Y. (2026). Hyper-BTS: Brain tumor segmentation based on hypergraph guidance. *Pattern Recognition*, 169: 111926. <https://doi.org/10.1016/j.patcog.2025.111926>
- [24] Mahootiha, M., Tak, D., Ye, Z., Zapaishchykova, A., Likitlersuang, J., Climent Pardo, J.C., Kann, B.H. (2025). Multimodal deep learning improves recurrence risk prediction in pediatric low-grade gliomas. *Neuro-oncology*, 27(1): 277-290. <https://doi.org/10.1093/neuonc/noae173>
- [25] Pande, Y., Chaki, J. (2025). Brain tumor detection across diverse MR images: An automated triple-module approach integrating reduced fused deep features and machine learning. *Results in Engineering*, 25: 103832. <https://doi.org/10.1016/j.rineng.2024.103832>
- [26] Mohsen, S., Oraby, S., Abdel-Aziz, M. (2025). Deep learning and machine learning for brain tumor detection: A review, challenges, and future directions. *Archives of Computational Methods in Engineering*, 1-25. <https://doi.org/10.1007/s11831-025-10416-3>
- [27] Wang, S.C., Yin, S.N., Wang, Z.Y., Ding, N., Ji, Y.D., Jin, L. (2025). 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*, 25(1): 219. <https://doi.org/10.1186/s12880-025-01798-8>
- [28] Wang, W., Yang, G., Liu, Y., Wei, L., Xu, X., Zhang, C., Liang, X. (2025). Multimodal deep learning model for prognostic prediction in cervical cancer receiving definitive radiotherapy: A multi-center study. *NPJ Digital Medicine*, 8(1): 503. <https://doi.org/10.1038/s41746-025-01903-9>