Integrate Semantic Radiomics as Prior Evidence into Evidential Deep Learning for Pelvic Lipomatosis Diagnosis

Zheran Zhang[1], Xiaodong Yue[1(✉)], Maoyu Wang[2], Zhikang Xu[3], Yufei Chen[4], and Zhipeng Wei[1]

[1] Shanghai University, Shanghai, China
yswantfly@shu.edu.cn
[2] Naval Medical University, Shanghai, China
[3] Shanxi University, Taiyuan, China
[4] Tongji University, Shanghai, China

Abstract. Pelvic Lipomatosis (PL) is a rare disorder characterized by abnormal fat proliferation in the pelvic region, where subtle imaging differences between pathological and normal fat pose significant diagnostic challenges. Existing deep-learning-based computer-aided diagnosis methods struggle to integrate high-level clinical semantics, which limits the diagnosis accuracy. This paper proposes a novel Evidential Deep Learning (EDL) method that synergistically fuses multi-type semantic radiomics priors derived from clinical expertise to enhance PL diagnosis. First, referring to clinical experiences, the critical PL semantic radiomics including bladder-rectal fat distance, rectal circularity, bladder-seminal vesicle angle, and relative pelvic fat volume are extracted from 3D abdominal CT images. Second, these semantic radiomics are probabilistically formulated as prior evidences to quantify their diagnostic relevance. Finally, the prior evidences are fused into the EDL backbone to implement PL diagnosis. Comparing with the pure deep learning methods, the EDL method with prior evidences not only reduces overconfident predictions but also enables interpretable decision-making by involving clinical knowledge. Experiments demonstrate the state-of-the-art performance of the proposed method, which achieves great improvements over conventional deep learning baselines. Ablation studies also validate the necessity of integrating the semantic features. Theoretical proofs further confirm that clinically consistent priors minimize prediction loss and enhance model stability. This work advances the diagnosis by bridging clinical radiomics with data-driven deep learning and provides a paradigm for interpretable PL medical image analysis.

Keywords: Pelvic Lipomatosis Diagnosis · Semantic Radiomics · Evidential Deep Learning.

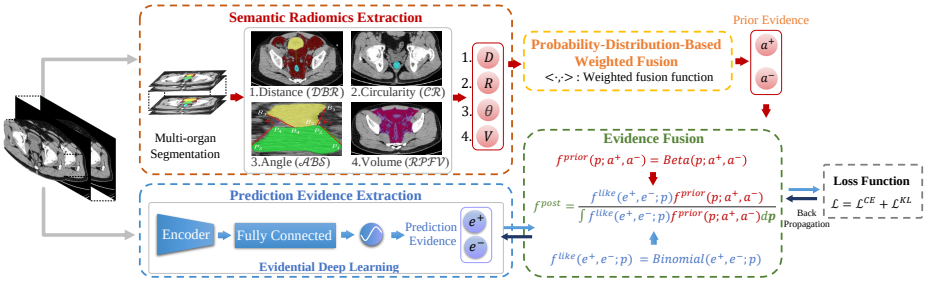

Fig. 1: Framework of the proposed method that integrates multiple clinical semantic radiomics to optimize PL diagnosis.

1 Introduction

Pelvic lipomatosis (PL) is a rare benign proliferative condition characterized by the overgrowth of non-encapsulated fatty tissue in the perirectal and perivesical spaces of the pelvis [1], causes organ compression and deformation [2]. Its low incidence (0.6-1.7/100,000) [3] and rarity complicate data acquisition while similarity with fat-related conditions (e.g., obesity) leads to frequent misdiagnosis.

While traditional Deep Neural Networks (DNNs) show potential in computer-aided diagnosis for common diseases [4–10], their direct application to PL is challenged by data scarcity. Consequently, there is a recent inclination to incorporate additional information to address data limitations [11, 12]. However, current knowledge-integrated methods often rely on shallow fusion strategies (e.g., transfer learning [13–15] or simplistic clinical feature concatenation [16–19]), which are insufficient for the complexity of PL, requiring integrating multiple, high-level clinical semantics to distinguish it from common fat diseases.

To tackle the problems, we construct *clinically semantic radiomics* based on *clinical domain knowledge* and integrate them as prior evidences into Evidential Deep Learning (EDL) [20] backbone for PL diagnosis. This integration *reduces dependence on labeled data*, while maintaining diagnostic accuracy with limited dataset. By guiding the diagnosis process with semantic radiomics, the decision of model is more consistent with *clinical diagnostic expertise*, distinguishing PL from other fat diseases. The contributions of this paper are summarized below.

- *Construct semantic radiomics based on clinical domain knowledge of PL.* These radiomics include bladder-rectal fat distance, rectal circularity, bladder-seminal vesicle angle, and relative pelvic fat volume, which cover fat-induced compression and displacement of pelvic organs.
- *Propose strategy to integrate clinical semantic radiomics into DNN model.* The extracted radiomics are transformed into probabilistic distributions, which are then fused as prior evidences into EDL backbone model to enhance PL diagnosis.

The illustration of our proposed framework can be seen in Fig. 1. First, based on the multi-organ segmentation results of abdominal CT images, morphological

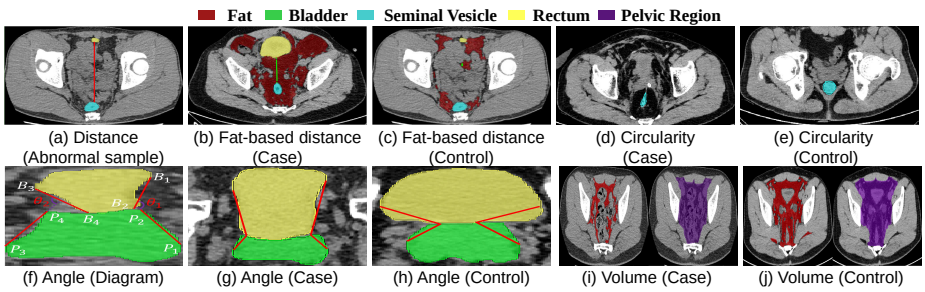

Fig. 2: Four semantic radiomics of PL. The red line in (a) is the absolute distance from bladder to rectum, the green line in (b)(c) represents the fat-based distance.

changes of pelvic organs are quantified. Then, these radiomics are converted into class probability distributions and integrated as weighted prior evidences. Finally, the prior evidences are fused into EDL backbone to enhance diagnostic accuracy under data scarcity.

2 Methodology

2.1 Semantic Radiomics of PL

PL is characterized by overgrowth of abnormally proliferated fat tissue in the pelvis causing organ compression and deformation. Guided by clinical expertise and prior studies [21] [22], we focus on the imaging parameters with minimal sensitivity to bladder filling and high reproducibility, including:

1. Distance from Bladder to Rectum ($\mathcal{DBR}$)
2. Circularity of Rectum ($\mathcal{CR}$)
3. Angle between Bladder Wall and Seminal Vesicle ($\mathcal{ABS}$)
4. Relative Pelvic Fat Volume ($\mathcal{RPFV}$)

These semantic radiomics are extracted from 3D abdominal CT images.

Distance from Bladder to Rectum. To accurately measure true organ displacement caused by pathological fat proliferation (Fig. 2(b)(c)), excluding interference from intervening organs (Fig. 2(a)), we define the fat-based distance: $D = n \times d$, where d is the pixel spacing, and $n = \sum_{t \in l(x)} I(t)$ counts fat pixels along the line $l(x)$ between bladder base $\hat{a}$ and rectum apex $\hat{b}$: $\hat{a}, \hat{b} = \arg\min_{a \in A} \left(\arg\min_{b \in B} \|a - b\| \right)$, where a is bladder pixel points, b is rectum pixel points, A denotes bladder pixel sets, B denotes rectum pixel sets, and $\|\cdot\|$ denotes the Euclidean distance. $I(t)$ takes the value of 1 if pixel point t is fat pixel point, and 0 otherwise.

Circularity of Rectum. As shown in Fig. 2(d)(e), we quantify fat-induced rectal deformation: $R = 4\pi \cdot S/C^2$, where S represents the area enclosed by rectum contour and C denotes the circumference of contour.

Angle between Bladder Wall and Seminal Vesicle. Pathological fat proliferation alters the relative position of bladder and seminal vesicle (Fig. 2(g)(h)),

quantified through bilateral angular measurements. The right (θ_1) and left (θ_2) angles are computed via vector dot products between anatomical landmarks: $\theta_1 = \arccos \frac{\overrightarrow{B_2B_1} \cdot \overrightarrow{P_2P_1}}{|\overrightarrow{B_2B_1}| \cdot |\overrightarrow{P_2P_1}|}, \theta_2 = \arccos \frac{\overrightarrow{B_4B_3} \cdot \overrightarrow{P_4P_3}}{|\overrightarrow{B_4B_3}| \cdot |\overrightarrow{P_4P_3}|}$, where B_1, B_2, B_3, B_4 and P_1, P_2, P_3, P_4 denote bladder and seminal vesicle boundary points (see in Fig. 2(f)). The average angle θ is then calculated as: $\theta = (\theta_1 + \theta_2)/2$.

Relative Pelvic Fat Volume. As illustrated in Fig. 2(i)(j), the relative fat volume V is the ratio of fat volume V_{fat} to pelvic volume V_{pelvic}: $V = V_{fat}/V_{pelvic}$.

2.2 Fusion of Semantic Radiomics as Prior Evidence into EDL

Semantic radiomics are translated into prior evidences through class conditional modeling. Specifically, for each feature $k \in \{D, R, \theta, V\}$ (as mentioned in Section 2.1), we compute Gaussian parameters for the PL class (μ_k^+, σ_k^+) and the control class (μ_k^-, σ_k^-). We calculate the probability and confidence for each sample x_i: the likelihood is given by the probability density function $a_{i,k}^+ = \frac{1}{\sigma_k^+\sqrt{2\pi}} \exp\left(-(k-\mu_k^+)^2/2(\sigma_k^+)^2\right)$, which quantify the relative likelihood, and the confidence is determined by the cumulative distribution function $c_{i,k}^+ = \frac{1}{\sigma_k^+\sqrt{2\pi}} \int_{-\infty}^{k} \exp\left(-(k-\mu_k^+)^2/2(\sigma_k^+)^2\right) d\boldsymbol{x_i}$, amplifying features where samples occupy statistically significant positions. Based on this, the prior evidences that x_i belongs to positive and negative classes are determined by

$$a_i^+ = \left\langle c_{i,k}^+, a_{i,k}^+ \right\rangle = \frac{\sum_k c_{i,k}^+ \cdot a_{i,k}^+}{\sum_k c_{i,k}^+}, \quad a_i^- = \left\langle c_{i,k}^-, a_{i,k}^- \right\rangle = \frac{\sum_k c_{i,k}^- \cdot a_{i,k}^-}{\sum_k c_{i,k}^-}, \tag{1}$$

where $a_i^+ + a_i^- = 2$ is achieved through normalization.

Traditional DNNs suffer from overconfidence due to softmax's exponential amplification of logit differences. In contrast, EDL [20] treats class probabilities as multivariate random variables following a distribution. The mathematic expectation of this distribution reduces model dependence on labeled data while improving classification accuracy. Based on Bayesian theorem, the posterior distribution of a multivariate random variable can be expressed as

$$f^{post}(\boldsymbol{p} \mid x) = \frac{f^{like}(x \mid \boldsymbol{p}) f^{prior}(\boldsymbol{p})}{\int_{\boldsymbol{p}} f^{like}(x \mid \boldsymbol{p}) f^{prior}(\boldsymbol{p}) d\boldsymbol{p}}, \tag{2}$$

where $\boldsymbol{p}$ denotes the class probabilities output from the neural network, $f^{like}(x \mid \boldsymbol{p})$ denotes the likelihood function and $f^{prior}(\boldsymbol{p})$ denotes the prior distribution. Based on the conjugate relationship between binomial likelihood ($f^{like}(e^+, e^-; p_1, p_0) = \boldsymbol{Binomial}(e^+, e^-; p_1, p_0) = C_{e^+ + e^-}^{e^+} p_1^{e^+} p_0^{e^-}$) and Beta prior ($f^{prior}(p_1, p_0; a^+, a^-) = \boldsymbol{Beta}(p_1, p_0; a^+, a^-) = \frac{p_1^{a^+-1} p_0^{a^--1}}{\mathcal{B}\left(a^+, a^-\right)}$) [23], the posterior can be com-

puted as

$$f^{post}(p_1,p_0;e^+,e^-,a^+,a^-)=\frac{f^{like}(e^+,e^-;p_1,p_0)f^{prior}(p_1,p_0;a^+,a^-)}{\int_0^1 f^{like}(e^+,e^-;p_1,p_0)f^{prior}(p_1,p_0;a^+,a^-)d\boldsymbol{p}}$$
$$=\frac{C_{e^+\!+e^-}^{e^+}p_1^{e^+}p_0^{e^-}\cdot\frac{1}{\mathcal{B}\left(a^+,a^-\right)}p_1^{a^+-1}p_0^{a^--1}}{\int_0^1 C_{e^+\!+e^-}^{e^+}p_1^{e^+}p_0^{e^-}\cdot\frac{1}{\mathcal{B}\left(a^+,a^-\right)}p_1^{a^+-1}p_0^{a^--1}d\boldsymbol{p}}=\frac{p_1^{e^+\!+a^+-1}p_0^{e^-\!+a^--1}}{\mathcal{B}\left(e^++a^+,e^-+a^-\right)},\quad(3)$$

where the probabilities for the PL and control classes are denoted by p_1 and p_0, respectively, e^+ and e^- represent the observational evidences gathered by DNN, a^+ and a^- are the prior evidences based on semantic radiomics (Eqs. (1)), and $\mathcal{B}(\cdot)$ is *beta* function.

2.3 PL Learning with Semantic Radiomics Prior Evidence

Given a dataset $\mathcal{D}=\{x_i,\boldsymbol{y}_i\}_{i=1}^N$ which comprises N labelled 3D CT images. Each image is encoded with a one-hot label vector $\boldsymbol{y}_i=(y_i^+,y_i^-)$. The loss function for PL diagnosis is defined as the combination of the cross-entropy term $\mathcal{L}_i^{CE}$ and the regularizing term $\mathcal{L}_i^{KL}$, which can be expressed as

$$\mathcal{L}=\frac{1}{N}\sum_{i=1}^{N}(\mathcal{L}_i^{CE}+\lambda\cdot\mathcal{L}_i^{KL}),\quad(4)$$

where $\lambda=min(1.0,\frac{\tau}{1000})\in[0,1]$ is the annealing coefficient and τ is the index of the current training epoch. The cross-entropy loss can be calculated as follows

$$\begin{aligned}\mathcal{L}_i^{CE}&=-\int[y_i^+\log p_{i1}+y_i^-\log p_{i0}]f^{post}(p_{i1},p_{i0};e_i^+,e_i^-,a_i^+,a_i^-)d\boldsymbol{p}\\&=y_i^+\left(\psi(\mathcal{S}_i)-\psi(e_i^++a_i^+)\right)+y_i^-\left(\psi(\mathcal{S}_i)-\psi(e_i^-+a_i^-)\right),\end{aligned}\quad(5)$$

where $\Gamma(\cdot)$ is *gamma* function, $\mathcal{S}_i=e_i^++e_i^-+2$, and $\psi(\cdot)$ is *digamma* function.

In order to minimize the evidence for the incorrect category while maintaining the evidence for the correct category at a constant level, we introduce a regularization loss term, which is defined as the KL divergence between $f\left(p_{i1},p_{i0};\tilde{\alpha}_i,\tilde{\beta}_i\right)$ and $f^{prior}\left(p_{i1},p_{i0};a_i^+,a_i^-\right)$

$$\begin{aligned}\mathcal{L}_i^{KL}&=KL\left[f\left(p_{i1},p_{i0};\tilde{\alpha}_i,\tilde{\beta}_i\right)\parallel f^{prior}\left(p_{i1},p_{i0};a_i^+,a_i^-\right)\right]\\&=\left(\tilde{\alpha}_i-a_i^+\right)\psi\left(\tilde{\alpha}_i\right)+\left(\tilde{\beta}_i-a_i^-\right)\psi\left(\tilde{\beta}_i\right)+\left(2-\tilde{\alpha}_i-\tilde{\beta}_i\right)\psi\left(\tilde{\alpha}_i+\tilde{\beta}_i\right)\\&+\log\frac{\mathcal{B}(a_i^+,a_i^-)}{\mathcal{B}(\tilde{\alpha}_i,\tilde{\beta}_i)},\end{aligned}\quad(6)$$

where $\tilde{\alpha}_i=y_i^-(e_i^++a_i^+)+y_i^+a_i^+$, $\tilde{\beta}_i=y_i^+(e_i^-+a_i^-)+y_i^-a_i^-$.

The following theorem can be deduced in order to analyse the enhancement of prediction that arises from by integrating clinical knowledge as semantic priors.
Theorem *In the case of PL sample, if prior evidence $a^+ > a^-$, or in the case of the sample in the control group, if prior evidence $a^+ < a^-$, it is possible to incur a greater loss of prediction without prior evidence, $\mathcal{L}^{ENN} > \mathcal{L}$.*
Proof: The prediction loss $\mathcal{L}^{ENN}$ of EDL backbone can be calculated as follows with uniform prior distribution ($a^+ = a^- = 1$)

$$\mathcal{L}^{ENN} = y^+ \left(\psi(\mathcal{S}) - \psi(e^+ + 1)\right) + y^- \left(\psi(\mathcal{S}) - \psi(e^- + 1)\right). \tag{7}$$

The prediction loss $\mathcal{L}$ of EDL which integrates prior evidences a^+ and a^- is

$$\mathcal{L} = y^+ \left(\psi(\mathcal{S}) - \psi(e^+ + a^+)\right) + y^- \left(\psi(\mathcal{S}) - \psi(e^- + a^-)\right). \tag{8}$$

Because $a^+ > a^-$, $a^+ + a^- = 2$, we have $e^+ + a^+ > e^+ + 1$ and $e^- + a^- < e^- + 1$. Furthermore, given that the sample is PL, we have $y^+ = 1$ and $y^- = 0$. According to $\psi(\cdot)$ increases monotonically on the interval $(0, +\infty)$, we can infer that

$$y^+ \left(\psi(\mathcal{S}) - \psi(e^+ + 1)\right) > y^+ \left(\psi(\mathcal{S}) - \psi(e^+ + a^+)\right),$$
$$\mathcal{L}^{ENN} > \mathcal{L}. \tag{9}$$

For the sample in the control group, we have similar proof. □

The derivation proof of the theorem enables the selection of suitable prior evidence that is consistent with the target labels. This helps to minimise prediction loss to enhance classification accuracy and stability while reducing data dependency and improving model interpretability.

3 Experiments

Our dataset comprises 126 3D CT images collected from patients who underwent CT urography at Changhai hospital over eight years, reflecting PL's rarity (incidence: 0.6-1.7/100,000) [3] and male predominance (male-to-female ratio 18:1) [24], with 62 male PL cases and 64 age and BMI-matched male controls.
Implementation Details. The evidential ViT backbone (input: 224×224×64, downsampled for GPU constraint) is trained on 80% of the data with SGD (learning rate is 0.005), 8×8×8 patch size, and batch size 1.

The experiments consist of ablation studies to verify the impact of integrating prior knowledge and comparisons with state-of-the-art (SOTA) methods.

Table 1: Ablation studies of integrating various amounts of semantic radiomics.

Features	**Number of Features**														
	1				**2**						**3**				**4**
$\mathcal{DBR}$	✓				✓	✓	✓				✓	✓	✓		✓
$\mathcal{CR}$		✓			✓			✓	✓		✓	✓		✓	✓
$\mathcal{ABS}$			✓			✓		✓		✓	✓		✓	✓	✓
$\mathcal{RPFV}$				✓			✓		✓	✓		✓	✓	✓	✓
ACC(%)	75.0	75.0	87.0	70.8	75.0	75.0	75.0	79.2	83.3	87.0	83.3	79.2	70.8	79.2	**87.5**

Table 2: Ablation studies of integrating semantic radiomics. (W/O: Without, W/: With)

Methods	W/O prior	W/ prior
Evi-ResNet [25]	50.00	**75.00**
Evi-DenseNet [26]	62.50	**70.83**
Evi-ViT [27]	66.67	**87.50**

Table 3: Ablation studies of integrating semantic radiomics under different conditions.

Method	Train	Test	Accuracy(%)
Evi-ViT	-	-	66.67
	✓	-	70.83
	-	✓	83.33
	✓	✓	**87.50**

3.1 Ablation study

In the following experiments, we validate the effectiveness of EDL backbone, which incorporates expert knowledge, aiming to show its ability to improve diagnostic accuracy and reduce the dependency of model on large datasets.

Validating the Improvement of Model Accuracy. Comparative experiments on multi-prior integration (1-4 features) into evidential ViT demonstrate that fusing all four semantic radiomics yields optimal accuracy (87.5%), outperforming individual or partial combinations and highlighting feature fusion in enhancing model performance (Table 1).

Additionally, ablation studies across three EDL-based models confirm the necessity of semantic radiomics integration, with prior knowledge consistently improving accuracy (Table 2). We further validate three fusion strategies for evidential ViT: 1) priors in training only, 2) priors in testing only, and 3) joint training-testing fusion (our method). Table 3 shows that our method (train + test) achieves the best performance.

In addition, Fig. 3 demonstrates the limitations of pure image-driven evidential ViT: misclassifying high-fat cases with deformation as controls (Fig. 3(a)) and low-fat cases without deformation as PL (Fig. 3(b)). It also fail in confusing situations: low-fat cases with deformation (Fig. 3(c)) or high-fat cases

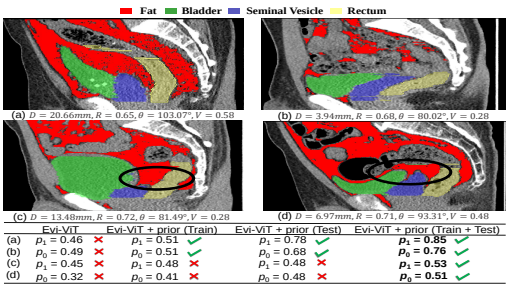

Fig. 3: Cases of PL prediction improvements brought by prior evidences. "Green ticks" indicate correct predictions and "red crosses" denote errors. ((a)(c): PL, (b)(d): Control)

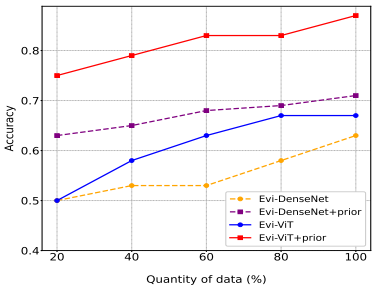

Fig. 4: Classification accuracy of comparing methods at different quantity of data.

without deformation (Fig. 3(d)), which other models except our method also misdiagnose. This demonstrates that relying only on image data or single-stage fusion (introducing priors separately in training or testing) fails to capture such nuanced features. In contrast, our cross-stage fusion strategy, which integrates semantic radiomics in both training and testing, effectively distinguishes organ deformation, particularly in the context of diagnosing intricate samples.

Validation of Data Dependency. The integration of semantic radiomics significantly enhances model robustness under data scarcity by encoding clinical expertise as prior knowledge (Fig. 4). While EDL-based models exhibit catastrophic performance degradation with limited data, our framework maintains diagnostic reliability by fusing prior aligned with clinical diagnostic experience.

3.2 Comparison with SOTA Methods

Our method outperforms against three categories of approaches (Table 4):
1) Manual feature-based machine learning (Logistic Regression, Shallow Neural Networks (ShallowNN)); 2) Adaptive DNNs (ResNet, DenseNet, ViT, evidential DenseNet (Evi-DenseNet), evidential ViT (Evi-ViT)); 3) Knowledge-integrated DNNs (SVD-Net, ViT-AD). While DNNs struggle with data scarcity of PL and have difficulty capturing subtle fat morphology, our EDL-based framework overcomes these limitations by fusing multi-semantic priors, surpassing both pure imaging-based DNNs and traditional machine learning through guidance of clinical diagnosis experience. Rigorous paired t-tests on 5-fold cross-validation splits confirm statistically significant superiority over all baselines ($p < 0.01$).

Table 4: Accuracy results of various PL diagnosis methods. Results were reported as mean ± SDs using 5-fold cross validation. The p-value indicates statistical significance.

	Methods	**Accuracy(%)**
Traditional radiomics methods	LogisticRegression [28]	81.51±7.87
	ShallowNN	81.57±7.14
Adaptive learning methods	ResNet [25]	50.71±3.18
	DenseNet [26]	62.59±8.41
	ViT [27]	61.83±8.03
	Evi-DenseNet	58.75±4.87
	Evi-ViT	62.79±6.45
Knowledge-integrated methods	SVD-Net [29]	64.39±6.80
	ViT-AD [30]	68.45±8.67
	Our Method	**83.51±6.89**
	p-value	2.04e-05

4 Conclusion

Traditional DNNs frequently misclassify PL due to limited data and adaptive-learned representations that do not match clinical diagnostic standards. Our evidential DNN framework overcomes this by integrating clinically validated morphological priors, enhancing diagnostic accuracy and interpretability while demonstrating robust performance in our partner hospital clinical deployment.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Serial Nos. No. 62476165,62472315,62173252), OpenProject Foundation of Intelligent Information Processing Key Laboratory of Shanxi Province, China (No. CICIP2021001), Natural Science Foundation of Shanghai (No. 21ZR1423900), and Shanghai Science and Technology Innovation Action Plan (22511101903).

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

References

1. Bååth L, Nyman U, Aspelin P, et al. Computed tomography of pelvic lipomatosis. *Acta Radiologica. Diagnosis*, 27(3):311–314, May 1986.
2. Cruz J, Vieira F, Mendes L, et al. Pelvic lipomatosis: a case report and literature review. *Radiologia Brasileira*, page 175–177, Jun 2012.
3. Xia M, Xiong S, Li Z, et al. Surgical treatment of pelvic lipomatosis: a systematic review of 231 cases. *Therapeutic Advances in Urology*, 15, 2023.
4. Kumar A, Tripathi A, Satapathy S, et al. Sars-net: Covid-19 detection from chest x-rays by combining graph convolutional network and convolutional neural network. *Pattern Recognition*, page 108255, Feb 2022.
5. Oghbaie M, Araújo T, Emre T, et al. Transformer-based end-to-end classification of variable-length volumetric data. *Medical Image Computing and Computer-Assisted Intervention*, pages 358–367, 2023.
6. Cheng J, Wu Z, Yuan X, et al. Disentangled hybrid transformer for identification of infants with prenatal drug exposure. *Medical Image Computing and Computer-Assisted Intervention*, pages 67–76, 2024.
7. Yi L, Zhang L, Xu X, et al. Multi-label softmax networks for pulmonary nodule classification using unbalanced and dependent categories. *IEEE Transactions on Medical Imaging*, 42(1):317–328, 2023.
8. Yang Y, Ye C, Su G, et al. Brainmass: Advancing brain network analysis for diagnosis with large-scale self-supervised learning. *IEEE Transactions on Medical Imaging*, 43(11):4004–4016, 2024.
9. Zhang J, Guo Y, Zhou L, et al. Constructing hierarchical attentive functional brain networks for early ad diagnosis. *Medical Image Analysis*, 94:103137, 2024.
10. Wang C, Liu T, Lai P, et al. Ensemble transformer-based multiple instance learning to predict pathological subtypes and tumor mutational burden from histopathological whole slide images of endometrial and colorectal cancer. *Medical Image Analysis*, 99:103372, 2025.
11. Xie X, Niu J, Liu X, et al. A survey on incorporating domain knowledge into deep learning for medical image analysis. *Medical Image Analysis*, 69:101985, 2021.

12. Jin W, Li X, Fatehi M, et al. Guidelines and evaluation of clinical explainable ai in medical image analysis. *Medical image analysis*, 84:102684, 2023.
13. Rajput I, Gupta A, Jain V, et al. A transfer learning-based brain tumor classification using magnetic resonance images. *Multimedia Tools & Applications*, 83(7), 2024.
14. Shamshiri M, Krzyżak A, Korbicz M, et al. Compatible-domain transfer learning for breast cancer classification with limited annotated data. *Computers in Biology and Medicine*, page 106575, Mar 2023.
15. Ajilisa O, Jagathy R, and Sabu M. A deep learning framework for the characterization of thyroid nodules from ultrasound images using improved inception network and multi-level transfer learning. *Diagnostics*, 13(14):2463, 2023.
16. Zhang L, Wang L, Yu M, et al. Hybrid representation learning for cognitive diagnosis in late-life depression over 5 years with structural mri. *Medical Image Analysis*, Jan 2024.
17. Mei L, Deng K, Cui Z, et al. Clinical knowledge-guided hybrid classification network for automatic periodontal disease diagnosis in x-ray image. *Medical Image Analysis*, 99:103376, 2025.
18. Liu M, Liu Y, Xu P, et al. Exploiting geometric features via hierarchical graph pyramid transformer for cancer diagnosis using histopathological images. *IEEE Transactions on Medical Imaging*, 43(8):2888–2900, 2024.
19. Zhou H, He L, Chen B, et al. Multi-modal diagnosis of alzheimer's disease using interpretable graph convolutional networks. *IEEE Transactions on Medical Imaging*, 44(1), 2025.
20. Sensoy M, Kaplan L, and Kandemir M. Evidential deep learning to quantify classification uncertainty. *Neural Information Processing Systems,Neural Information Processing Systems*, Jan 2018.
21. Zhang Y, Wu S, Xi Z, et al. Measuring diagnostic accuracy of imaging parameters in pelvic lipomatosis. *European Journal of Radiology*, page 3107–3114, Nov 2012.
22. Bai X, Zhang G, Xu L, et al. Diagnostic accuracy of ct imaging parameters in pelvic lipomatosis. *Abdominal Radiology*, page 2779–2788, Jun 2021.
23. Audun J. Subjective logic. 3, 2016.
24. Chris F. Pelvic lipomatosis: A review of its diagnosis and management. *Journal of Urology*, page 267–273, Jul 1991.
25. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Jun 2016.
26. Huang G, Liu Z, Van D, et al. Densely connected convolutional networks. Jul 2017.
27. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*, 2021.
28. Andrew J, David W, and Stanley L. Applied logistic regression. *Technometrics*, 34:358–359, 1990.
29. Zhang S, Li Z, Zhou H, et al. Advancing 3d medical image analysis with variable dimension transform based supervised 3d pre-training. *Neurocomputing*, page 11–22, Apr 2023.
30. Kunanbayev K, Shen V, and Kim D. Training vit with limited data for alzheimer's disease classification: an empirical study. *Medical Image Computing and Computer Assisted Intervention*, LNCS 15012:334 – 343, October 2024.