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.

Image-Level Weakly Supervised Classification and Localisation of Endometrial Lesions in Hysteroscopic Images

Xin Li
Fudan University
Wenhao Sun
Fudan University
Jiayi Zhang
Ningxia Medical University
Junxian Yang
Ningxia Medical University
Xiaoran Li
Fudan University
Dan Liu
Ningxia Medical University General Hospital
Biqin Dong
dongbq@fudan.edu.cn
Fudan University

Research Article

Keywords: Hysteroscopy, Endometrial lesion, Weakly-supervised learning, Layer-CAM, EfficientNet, Lesion localisation, Image processing

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11630912/v1

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

Additional Declarations: Competing interest reported. Biqin Dong is a founder and equity holder of Feimiao Intelligent Science & Technology (Taizhou) Co., Ltd.

Image-Level Weakly Supervised Classification and Localisation of Endometrial Lesions in Hysteroscopic Images

Xin Li,[1] Wenhao Sun,[1] Jiayi Zhang,[2] Junxian Yang,[2] Xiaoran Li,[1]

Dan Liu,[2,3,*] and Biqin Dong[1,*]

[1]Yiwu Research Institute, Fudan University, Shanghai, China

[2]Institute of Medical Sciences, General Hospital of Ningxia Medical University, Yinchuan, China

[3]Department of Gynecology, General Hospital of Ningxia Medical University, Yinchuan, China

*Corresponding authors: Dan Liu (nxld@nyfy.com.cn) and Biqin Dong (dongbq@fudan.edu.cn)

Abstract

Background: Endometrial lesions are a leading cause of abnormal uterine bleeding, and hysteroscopy combined with targeted biopsy is the clinical gold standard for diagnosis. However, traditional hysteroscopic assessment is subjective and experience-dependent. Deep learning offers an automated alternative, yet most existing lesion localisation methods require expensive pixel-level annotations.

Methods: A weakly-supervised framework combining an EfficientNet-B4 classifier with Layer-CAM was developed for four-class classification (endometrial cancer, EC; endometrial hyperplasia, EH; endometrial polyp, EP; and normal) and lesion localisation using only image-level labels. The model was trained on 6,487 hysteroscopy images, comprising 3,276 clinical images from 686 cases and 3,211 images from a public dataset. Layer-CAM heat maps were refined via thresholding and morphological post-processing into lesion contours. Comparative experiments with ResNet-50, ShuffleNetV2, and EfficientNet-B0, and an ablation study on public data augmentation, were conducted.

Results: On a held-out test set of 447 images, the proposed model achieved a macro-averaged sensitivity of 0.847, specificity of 0.958, F1 score of 0.831, and AUC of 0.965. EfficientNet-B4 outperformed all compared backbones (ResNet-50: 0.934; EfficientNet-B0: 0.930; ShuffleNetV2: 0.870). Public data augmentation improved macro AUC from 0.943 to 0.965, with the largest gains in EC (+0.116 sensitivity) and EH (+0.128). Layer-CAM produced more focused activation maps with sharper boundaries than Grad-CAM.

Conclusions: The proposed framework shows promise as an efficient, low-cost computer-aided diagnosis tool for hysteroscopic practice, achieving high diagnostic performance on the reported test set and providing interpretable lesion localisation without pixel-level annotations. Independent external testing and expert-annotated validation of the localisation outputs are required before clinical use. The combination of Layer-CAM and morphological post-processing offers a practical image processing pipeline for clinical visualisation of endometrial lesions.

Keywords: Hysteroscopy; Endometrial lesion; Weakly-supervised learning; Layer-CAM; EfficientNet; Lesion localisation; Image processing

Introduction

Endometrial cancer (EC) is one of the most common gynaecological malignancies worldwide, and its incidence has been rising in younger populations in recent years [1]. Endometrial polyp (EP) and endometrial hyperplasia (EH) are benign lesions that share similar early symptoms with EC, such as irregular vaginal bleeding and prolonged menstruation, making clinical differentiation challenging [2]. The five-year survival rate for early-stage EC can reach 95%, but drops sharply to approximately 15% for advanced disease, underscoring the critical importance of early and accurate diagnosis [3].

Hysteroscopy is the core minimally invasive technique for evaluating endometrial lesions, allowing direct visualisation of lesion morphology, colour, vascular patterns, and boundaries. Combined with targeted biopsy, it is regarded as the diagnostic gold standard [4]. Nevertheless, hysteroscopic assessment remains operator-dependent, and image interpretation may be complicated by uneven lighting, mucus, and instrument reflection

[5]. Targeted biopsy is also subject to sampling limitations, reinforcing the importance of appropriate histopathological assessment [4].

Deep learning has become an important approach to medical image analysis [6], with early studies demonstrating the feasibility of convolutional neural networks for endometrial lesion classification [7]. Architectures such as EfficientNet, ResNet, and MobileNetV2 provide alternative backbones for learning discriminative image features [8–10]. However, classification outputs alone do not reveal which image regions drive model predictions. Class activation mapping offers a means of visual interpretation: Grad-CAM highlights class-discriminative regions using target-class gradients [11], whereas Layer-CAM generates hierarchical localisation maps by integrating spatially resolved, gradient-weighted activations across convolutional layers [12]. In parallel, hysteroscopy-specific studies have investigated automated image detection and classification, including computer-aided diagnosis of atypical endometrial hyperplasia and endometrial cancer [13,14]. Nevertheless, activation maps do not directly provide validated lesion boundaries, and translating image-level predictions into reliable lesion localisation remains an important methodological challenge.

Despite these advances, several challenges remain. AI applications in minimally invasive gynecological surgery, including hysteroscopy, are supported predominantly by observational evidence, with limited external validation and concerns regarding risk of bias, which constrain clinical translation [15]. In addition, many fully supervised localisation methods require pixel-level annotations, which are costly and time-consuming to obtain [16]. Integrated frameworks that combine multiclass endometrial lesion classification with interpretable localisation using only image-level labels remain underexplored. To address these gaps, we propose a weakly supervised framework based on EfficientNet-B4 and Layer-CAM for four-class classification (normal, EH, EP, and EC) and lesion localisation using image-level labels.

The main contributions are: (1) A weakly supervised pipeline that combines EfficientNet-B4 classification with Layer-CAM activation mapping, requiring only image-level labels and eliminating the need for pixel-level annotations. (2) Comprehensive comparison with three backbone networks (ResNet-50, ShuffleNetV2, and EfficientNet-B0) [8, 9, 17], and ablation studies on public data augmentation. (3) A morphological post-processing pipeline that converts Layer-CAM heat maps into clean lesion contours, providing clinically interpretable visualisation for hysteroscopic image analysis.

Methods

The proposed framework comprises three stages: classification feature learning, attention heat map generation, and lesion region extraction. Given only a hysteroscopy image and its image-level class label, the framework outputs both a classification result and a localised lesion region. The overall pipeline is illustrated in Figure 1.

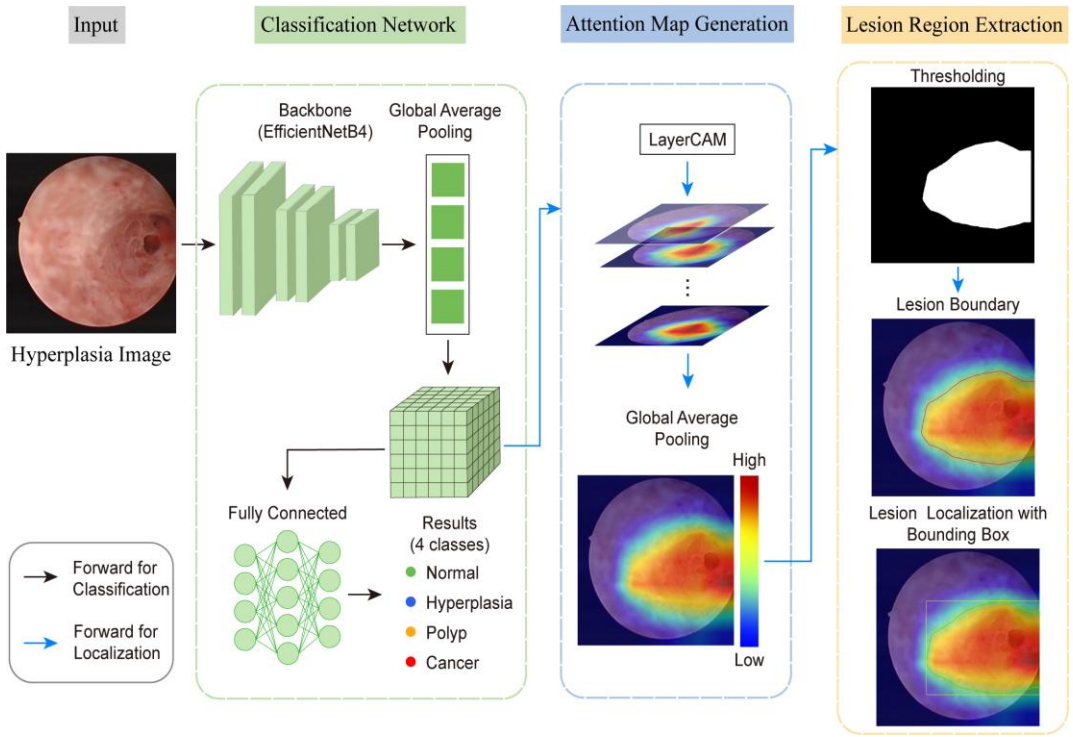

Figure 1. Overview of the proposed framework. The EfficientNet-B4 backbone feeds two branches: (A) a classification head producing four-class probabilities, and (B) a Layer-CAM module generating attention heat maps that are refined via thresholding and morphological post-processing into lesion contours with bounding boxes.

Dataset

A total of 686 clinical hysteroscopy cases were collected, comprising 151 EC, 111 EH, 330 EP, and 94 Normal cases (Table 1). From these cases, 3,276 images were extracted after deduplication and quality screening. To enhance training diversity and improve model generalisation, an additional 3,211 images from the public HS-CMU hysteroscopy dataset (https://opendatalab.org.cn/jiyuanmedical/HS-CMU) were incorporated into the training set. The combined dataset totals 6,487 images, distributed as follows: 701 EC (10.8%), 1,304 EH (20.1%), 3,281 EP (50.6%), and 1,201 Normal (18.5%). The clinical images were divided at the case (patient) level into training, validation, and test sets so that images from the same case did not appear in more than one split, and all 3,211 public images (which contain no normal cases) were added to the training set. The final splits comprised 5,452 training, 588 validation, and 447 test images (test: 69 EC, 39 EH, 237 EP, 102 Normal). All images were resized to 380 × 380 380×380pixels and normalised using ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]).

Table 1. Distribution of clinical cases and images across diagnostic categories.

Data level	Dataset / source	Diagnostic category (n)				Total
		EC	EH	EP	Normal	
Clinical cases	**Overall clinical cohort**	151	111	330	94	**686**
Images	Internal dataset	453	403	1,219	1,201	3,276
	Public dataset (HS-CMU)	248	901	2,062	0	3,211
	Combined dataset	701	1,304	3,281	1,201	**6,487**
	Test set (internal)	69	39	237	102	**447**

Network architecture

EfficientNet-B4 was adopted as the feature extraction backbone. EfficientNet employs a compound scaling strategy that jointly optimises network depth, width, and input resolution, achieving high representational capacity with relatively few parameters [8]. The input image is passed through a sequence of mobile inverted bottleneck blocks (MBConv) with squeeze-and-excitation optimisation, producing high-dimensional

semantic feature maps. A global average pooling layer reduces the spatial dimensions to a feature vector, which is fed to a fully connected classifier producing four-class probabilities via softmax.

The model was trained using the standard cross-entropy loss, with the Adam optimiser (learning rate of $1 \times 10^{-4}1\times10^{-4}$) and a batch size of 32 for 100 epochs. Images were augmented with random resized cropping, horizontal flipping, colour jitter, CLAHE contrast enhancement, and random erasing. Training was performed on a single NVIDIA GPU with CUDA 11.8 support.

Layer-CAM for weakly-supervised localisation

Layer-CAM extends the Grad-CAM paradigm by computing element-wise gradient-weighted activations across multiple convolutional layers rather than only the final layer [12]. For a given target class c, the forward activations A^k from the last three feature blocks of EfficientNet-B4 are recorded, and the gradients $\frac{\partial y^c}{\partial A^k}$ are obtained via backpropagation. The per-layer contribution is computed as:

$$\mathrm{CAM}^k=\mathrm{ReLU}(\alpha^k \odot A^k),\text{where}\alpha^k=\mathrm{ReLU}(\mathrm{mean}_c(\frac{\partial y^c}{\partial A^k}))$$

where $\odot$ denotes element-wise multiplication and the ReLU on gradients retains only positive contributions. The final Layer-CAM heat map is obtained by averaging the per-layer CAMs and resizing to the input resolution. By fusing shallow features (which capture fine-grained boundary information) with deep features (which encode overall lesion location), Layer-CAM produces more focused and boundary-preserving activation maps than Grad-CAM, effectively suppressing background noise.

Lesion contour extraction

The continuous Layer-CAM heat map is converted into a binary lesion mask through the following pipeline:

1. Thresholding: pixels with activation above the 75th percentile are retained as candidate lesion regions.

2. Morphological opening: a 5 × 5 kernel erosion followed by dilation removes isolated noise points and smooths region boundaries.

3. Dilation: a subsequent dilation operation expands the cleaned mask to recover regions potentially eroded by the opening step.

4. Contour extraction: connected components in the dilated mask are identified as candidate lesion regions, and their outer contours are extracted and overlaid on the original hysteroscopy image. This enables the detection of multiple lesions when present.

Evaluation metrics

Classification performance was assessed using overall accuracy, macro-averaged sensitivity (recall), specificity, F1 score, and the area under the receiver operating characteristic curve (AUC). Sensitivity measures the proportion of true positives correctly identified, specificity measures the proportion of true negatives correctly rejected, and the F1 score is the harmonic mean of precision and recall. The AUC was computed via one-vs-rest binarisation for each class. All metrics were evaluated on the held-out test set of 447 images. Confusion matrices and per-class ROC curves were also generated for detailed analysis.

Comparative experiments and ablation design

To evaluate the contribution of the backbone architecture, the proposed EfficientNet-B4 model was compared against three alternative backbones: EfficientNet-B0, ResNet-50, and ShuffleNetV2 [8, 9, 17]. All models were trained on the same combined dataset with identical hyperparameters and data augmentation. An ablation study was conducted by training EfficientNet-B4 on the internal dataset only (3,276 images) versus the combined dataset (6,487 images), to assess the effect of public data augmentation on classification performance.

Results

Classification performance of the proposed model

The proposed EfficientNet-B4 model achieved an overall accuracy of 87.7% on the 447-image test set. Table 2 summarises the per-class metrics, and Figure 2 visualises the dataset distribution, feature separability, confusion patterns, and ROC curves. The model performed excellently for Normal (sensitivity = 1.000, AUC = 0.996) and EP (sensitivity = 0.869, AUC = 0.964), while EH was the most challenging class (sensitivity = 0.692), reflecting the morphological overlap between hyperplasia and cancer. The macro-averaged sensitivity, specificity, F1 score, and AUC were 0.847, 0.958, 0.831, and 0.965, respectively.

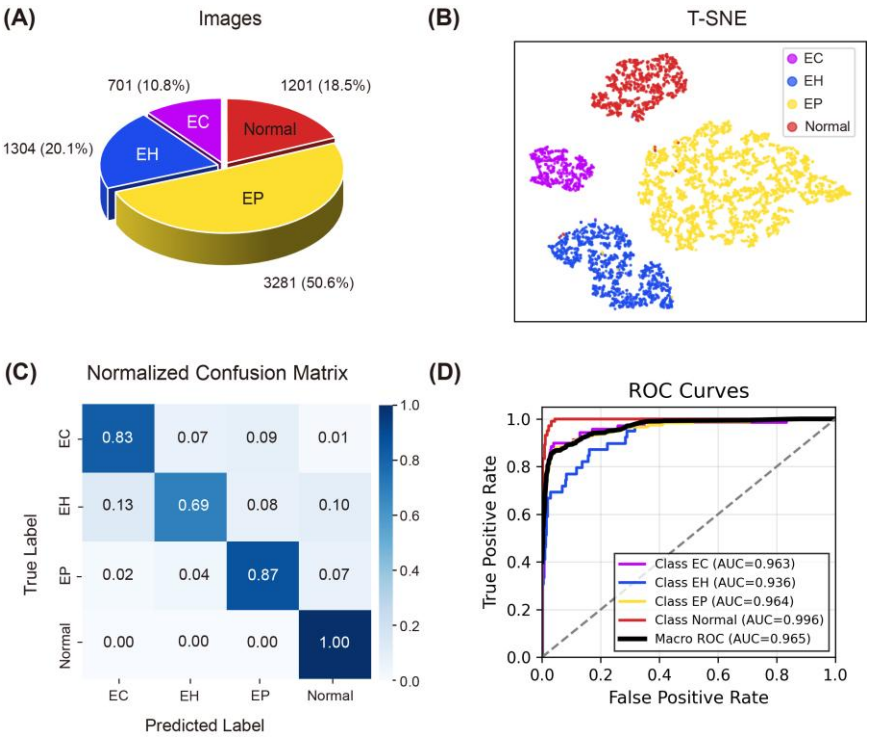

Figure 2. Dataset distribution and classification performance. (A) Pie chart of image distribution across four classes (EC 701, EH 1,304, EP 3,281, Normal 1,201). (B) t-SNE visualisation of deep features in the test set, showing class separability. (C) Normalised confusion matrix (rows = ground truth, columns = predictions). (D) Per-class ROC curves with AUC values (macro AUC = 0.965).

Table 2. Per-class classification performance of the proposed EfficientNet-B4 model on the test set (447 images).

Class	Sensitivity	Specificity	F1 Score	AUC
EC	0.826	0.976	0.844	0.963
EH	0.692	0.963	0.667	0.936
EP	0.869	0.957	0.912	0.964
Normal	1.000	0.936	0.903	0.996
Macro avg	0.847	0.958	0.831	0.965

The confusion matrix (Figure 2C) shows that all 102 Normal images were correctly classified, while the main confusions occurred between clinically similar categories: 5 EH were misclassified as EC and 10 EP were misclassified as EH. Crucially, only 1 EC case was misclassified as Normal, while no Normal case was classified as EC, indicating that the model reliably separates malignant from normal endometrium. The t-SNE plot (Figure 2B) is consistent with these results, showing compact Normal and EP clusters but a more dispersed EH cluster.

Grad-CAM versus Layer-CAM: qualitative comparison

Figure 3 compares Grad-CAM and Layer-CAM activation maps across representative EH, EC, and EP samples. Both methods highlight the lesion region, but Layer-CAM produces more focused and spatially coherent activations with sharper boundaries, particularly for EC and EP. Grad-CAM tends to disperse attention across a broader area including background tissue, whereas Layer-CAM concentrates on the diagnostically relevant region. The contrast is most evident in the EH samples, where Grad-CAM activates diffusely across the cavity while Layer-CAM tightly localises the hyperplastic region.

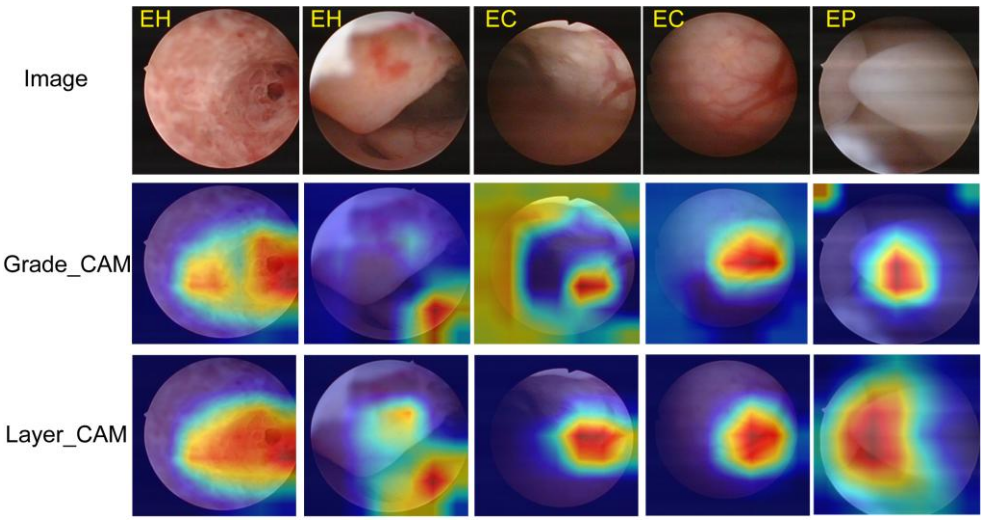

Figure 3. Qualitative comparison of Grad-CAM and Layer-CAM overlays. Top row: original images; middle row: Grad-CAM heat maps; bottom row: Layer-CAM heat maps. Columns show representative samples from EH (cols 1-2), EC (cols 3-4), and EP (col 5). Layer-CAM produces more focused activations with sharper lesion boundaries.

Lesion localisation results

Figure 4 demonstrates the final lesion localisation output of the proposed pipeline. Starting from the Layer-CAM heat map, thresholding and morphological post-processing (opening, dilation, contour extraction) yield precise lesion boundaries, with the green contours delineating lesions and the yellow bounding boxes providing localisation rectangles for clinical reference. Across all shown cases, the contours accurately enclose the pathological regions, including an EH case where two independent lesions are successfully localised.

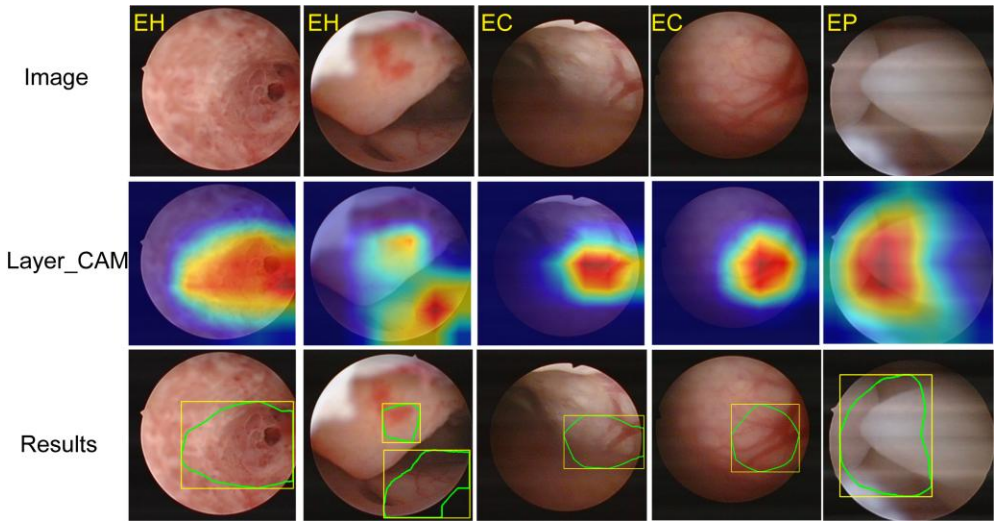

Figure 4. Final lesion localisation results. Top row: original images; middle row: Layer-CAM heat maps; bottom row: localised lesion contours (green) with bounding boxes (yellow). Columns correspond to the same samples as in Figure 3.

Comparison with backbones and ablation study

To assess the contribution of the backbone architecture and the training data, we compared EfficientNet-B4 against ResNet-50, ShuffleNetV2, and EfficientNet-B0, and evaluated the effect of supplementing the internal dataset with public images. All four backbones plus the private-data EfficientNet-B4 ablation variant (five model configurations in total) were evaluated on the same 447-image test set; the results are summarised in Figure 5.

Backbone comparison (Figure 5A). EfficientNet-B4 achieved the highest macro AUC (0.965), outperforming ResNet-50 (0.934), EfficientNet-B0 (0.930), and ShuffleNetV2 (0.870). The superior performance of B4 over B0 highlights the benefit of the larger model's representational capacity, while ResNet-50 achieved a competitive AUC but struggled with the EH class.

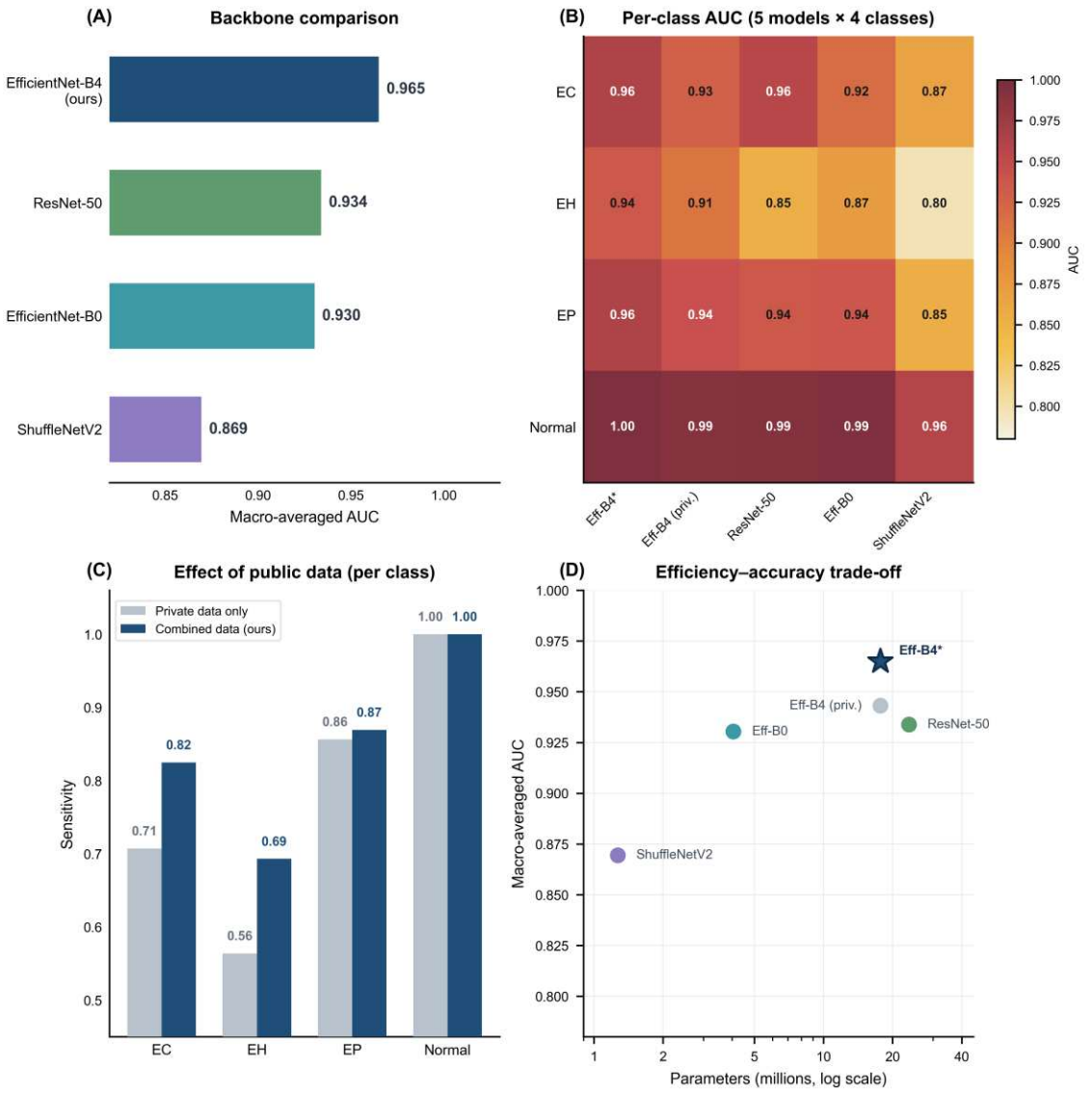

Figure 5. Comprehensive comparison and ablation analysis. (A) Backbone comparison measured by macro-averaged AUC. (B) Per-class AUC heatmap for the five model configurations (four backbones plus the private-data EfficientNet-B4 variant). (C) Ablation of public data: per-class sensitivity with and without the 3,211 public images. (D) Parameter–AUC trade-off: macro-averaged AUC versus model parameters. The proposed model is marked with an asterisk.

Per-class analysis (Figure 5B). The per-class AUC heatmap reveals that all models recognise the Normal class almost perfectly (AUC > 0.95), whereas EH is consistently the most difficult class across every backbone. The advantage of EfficientNet-B4 is most pronounced for EC and EH.

Ablation of public data (Figure 5C). Training on the internal dataset only (3,276 images) yielded a macro AUC of 0.943, whereas incorporating 3,211 public images raised it to 0.965. The gain is concentrated in the underrepresented classes: public data improved sensitivity for EC from 0.710 to 0.826 (+0.116) and for EH from 0.564 to 0.692 (+0.128), while EP and Normal were barely affected. This confirms that public-data augmentation mainly strengthens the model's ability to recognise the rarer malignant and hyperplastic patterns.

Parameter–AUC trade-off (Figure 5D). EfficientNet-B4 attains the best AUC with a moderate parameter count (17.7 million), outperforming the much larger ResNet-50 (23.6 million) while avoiding the accuracy penalty of the lightweight ShuffleNetV2 (1.3 million).

Discussion

This study presents a weakly-supervised framework for hysteroscopic lesion classification and localisation that requires only image-level labels. The proposed EfficientNet-B4 + Layer-CAM pipeline achieved a macro AUC of 0.965 on a four-class task spanning normal, hyperplasia, polyp, and cancer categories. Early work by Zhang et al. [7] demonstrated the feasibility of CNNs for endometrial lesion classification, and subsequent hysteroscopy-specific studies have focused on detecting endometrial cancer or differentiating atypical hyperplasia from cancer [13]. Our work extends this line by addressing the broader four-class differentiation that more closely mirrors routine clinical workflow, where distinguishing normal, hyperplasia, polyp, and cancer is essential for triage decisions [4]. These results demonstrate that high diagnostic performance can be obtained without pixel-level annotations, a point of particular relevance given that expert annotation remains a major bottleneck in medical image classification [18].

The backbone comparison revealed a clear advantage for EfficientNet-B4 over smaller or less optimised architectures. Although ResNet-derived models remain a common default in medical image classification [9], the compound scaling strategy of EfficientNet [8], which jointly balances depth, width, and resolution, proved particularly suited to the fine-grained texture and morphology discrimination required for endometrial lesion

classification. The improvement over EfficientNet-B0 (AUC 0.965 vs. 0.930) confirms that the additional capacity of B4 is beneficial, while the moderate parameter count keeps training feasible on a single GPU. This finding suggests that backbone selection deserves the same scrutiny as data curation in medical AI pipelines, an aspect that has received limited attention in prior hysteroscopic studies [5].

The ablation study demonstrated that supplementing the internal dataset with 3,211 public images improved macro AUC by 2.2 percentage points and macro sensitivity by 6.4 percentage points, with the largest gains concentrated in the underrepresented malignant and hyperplastic classes. This finding aligns with systematic reviews noting that data augmentation strategies are critical for addressing class imbalance and limited sample sizes in medical imaging [19]. Our results further indicate that enlarging data diversity is more effective than architectural changes alone for boosting minority-class recognition, and that the standard cross-entropy loss with the combined dataset already provides a well-balanced training signal across all classes. This observation supports a practical strategy of combining institutional with public data when the latter is available, although clinical translation requires careful consideration of data provenance and potential distribution shifts [15].

A key finding for medical image processing is the superiority of Layer-CAM [12] over Grad-CAM [11] for lesion visualisation. Grad-CAM averages gradients across spatial dimensions, which can blur fine structures, whereas Layer-CAM preserves spatially resolved gradient information across hierarchical feature levels, yielding sharper and more localised activation maps. Weakly-supervised localisation has been explored in various medical imaging contexts, including breast lesion detection in MRI [20], but remains underexplored in endoscopy. Our study extends this paradigm to hysteroscopic images and shows that aggregating positive gradient-weighted activations across the last three feature blocks, together with simple morphological post-processing, yields tight and clinically interpretable lesion contours without ground-truth segmentations. This pipeline is broadly applicable to other endoscopic modalities where weakly-supervised localisation is needed.

The most challenging class was EH (sensitivity 0.692), which is morphologically intermediate between normal endometrium and EC. This difficulty mirrors the well-documented clinical challenge of distinguishing atypical hyperplasia from early cancer [2], a limitation that has also constrained prior computer-aided diagnosis systems targeting these two classes [14]. Sun et al. [21] similarly applied CNN-based CAD to endometrial histopathology, suggesting that subclass discrimination difficulty is a shared challenge across imaging modalities. Hyperplasia may present with subtle thickening that overlaps with both normal and early malignant features, and the clinical stakes of misclassification are high because EH can progress to EC if untreated [3]. Future work could explore auxiliary clinical information (e.g., patient age, menopausal status) or temporal sequences to improve EH discrimination.

Several limitations should be acknowledged. First, the test set is relatively small (447 images), and external validation on independent multi-centre cohorts is needed to confirm generalisability; systematic evaluations of medical AI models have identified insufficient external validation as a pervasive limitation across the field [22]. Second, the lesion localisation is weakly supervised and has not been quantitatively evaluated against ground-truth segmentation masks [16]; the current contours therefore support qualitative visualisation rather than proof of segmentation accuracy. Third, the EfficientNet-B7 model could not be evaluated due to GPU memory constraints, and its potential benefit remains unexplored. Despite these limitations, the framework offers a practical approach to reducing operator-dependent variability in hysteroscopic assessment, and future work should prioritise multi-centre validation and integration with biopsy guidance to assess clinical utility.

Conclusions

We proposed a weakly-supervised framework combining EfficientNet-B4 classification with Layer-CAM activation mapping for hysteroscopic lesion diagnosis. Using only image-level labels, the model achieved a macro AUC of 0.965 for four-class classification (EC, EH, EP, Normal) and produced focused, clinically interpretable lesion localisation via morphological post-processing. Comparative and ablation experiments confirmed the superiority of the EfficientNet-B4 backbone and the benefit of public data

augmentation. The framework shows promise as an efficient, low-cost computer-aided diagnosis tool, reducing the annotation burden while maintaining high diagnostic accuracy and interpretability. However, the current results are limited to a single-centre test set with qualitative localisation; independent external validation and quantitative assessment against expert-annotated lesion masks are needed before clinical deployment. The Layer-CAM and morphological post-processing pipeline provides a practical image processing approach that can be extended to other medical imaging applications.

List of abbreviations

AUC: Area under the receiver operating characteristic curve

CAM: Class activation mapping

CLAHE: Contrast-limited adaptive histogram equalisation

EC: Endometrial cancer

EH: Endometrial hyperplasia

EP: Endometrial polyp

F1: F1 score (harmonic mean of precision and recall)

Grad-CAM: Gradient-weighted class activation mapping

Layer-CAM: Layer-wise class activation mapping

MBConv: Mobile inverted bottleneck block

ROC: Receiver operating characteristic

t-SNE: t-distributed stochastic neighbour embedding

Declarations

Ethics approval and consent to participate

This study was conducted with approval from the Medical Scientific Research Ethics Review Committee of General Hospital of Ningxia Medical University (Approval No.

KYLL-2022-1237). Written informed consent was obtained from all participants. All procedures were performed in accordance with the Declaration of Helsinki.

Availability of data and materials

The public HS-CMU hysteroscopy dataset used in this study is available at https://opendatalab.org.cn/jiyuanmedical/HS-CMU. The internal clinical dataset used and/or analysed during the current study is available from the corresponding author on reasonable request, subject to approval by the Institutional Ethics Committee. The source code for model training, evaluation, and Layer-CAM generation is available from the corresponding author on reasonable request.

Competing interests

Biqin Dong is a founder and equity holder of Feimiao Intelligent Science & Technology (Taizhou) Co., Ltd. The remaining authors declare no competing interests.

Funding

This work was supported by the Shanghai Pilot Program for Basic Research (22TQ020), Shanghai Science and Technology Innovation Action Plan (22S31905500), Yinchuan Science and Technology Tackling Project (2023SFZD02), and Ningxia Science and Technology Innovation Leading Talent Project (No. 2022GKLRLX010).

Authors' contributions

Xin Li and Wenhao Sun developed the deep learning framework and performed the analysis. Jiayi Zhang and Junxian Yang collected and preprocessed the clinical data and contributed to data analysis. Xin Li, Wenhao Sun, and Xiaoran Li drafted the manuscript. Dan Liu and Biqin Dong designed the study, supervised the research, reviewed and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors acknowledge the creators of the HS-CMU public dataset for making their data openly available.

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023 Jan;73(1):17–48. doi:10.3322/caac.21763 PubMed PMID: 36633525.

2. Fatahichegeni M, Haji Ghasem Sharbatdar Z, Ansarian MA. Advances in endometrial cancer screening: a comprehensive review of current methods and emerging technologies. Int J Clin Oncol. 2026 Feb 1;31(2):254–70. doi:10.1007/s10147-025-02941-9

3. Creasman WT, Odicino F, Maisonneuve P, Quinn MA, Beller U, Benedet JL, et al. Carcinoma of the corpus uteri. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. Int J Gynaecol Obstet. 2006 Nov 1;95 Suppl 1:S105-43. doi:10.1016/s0020-7292(06)60031-3 PubMed PMID: 17161155.

4. Visually Directed Hysteroscopic Biopsy in the Evaluation of Abnormal Uterine Bleeding and Postmenopausal Bleeding: A Joint Society Practice Guideline. Journal of Minimally Invasive Gynecology. 2026 Aug 1;33(8):1027–40. doi:10.1016/j.jmig.2026.06.018

5. Watrowski R, Di Spiezio Sardo A, Török P, Rosati A, Kostov S, Alkatout I, et al. Artificial Intelligence, Deep Learning, and Computer Vision in Hysteroscopy: A Systematic Review. Diagnostics. 2026 Jun 18;16(12):1899. doi:10.3390/diagnostics16121899

6. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019 Jan;25(1):24–9. doi:10.1038/s41591-018-0316-z PubMed PMID: 30617335.

7. Zhang Y, Wang Z, Zhang J, Wang C, Wang Y, Chen H, et al. Deep learning model for classifying endometrial lesions. J Transl Med. 2021 Jan 6;19(1):10. doi:10.1186/s12967-020-02660-x

8. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [Internet]. arXiv; 2020 [cited 2026 Oct 12]. Available from: http://arxiv.org/abs/1905.11946 doi:10.48550/arXiv.1905.11946

9. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition [Internet]. arXiv; 2015 [cited 2026 Oct 12]. Available from: http://arxiv.org/abs/1512.03385 doi:10.48550/arXiv.1512.03385

10. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted Residuals and Linear Bottlenecks [Internet]. arXiv; 2019 [cited 2026 Oct 12]. Available from: http://arxiv.org/abs/1801.04381 doi:10.48550/arXiv.1801.04381
11. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Int J Comput Vis. 2020 Feb 1;128(2):336–59. doi:10.1007/s11263-019-01228-7
12. Jiang PT, Zhang CB, Hou Q, Cheng MM, Wei Y. LayerCAM: Exploring Hierarchical Class Activation Maps for Localization. IEEE Transactions on Image Processing. 2021;30:5875–5888. doi:10.1109/TIP.2021.3089943
13. Raimondo D, Raffone A, Salucci P, Raimondo I, Capobianco G, Galatolo FA, et al. Detection and Classification of Hysteroscopic Images Using Deep Learning. Cancers (Basel). 2024 Mar 28;16(7):1315. doi:10.3390/cancers16071315 PubMed PMID: 38610993; PubMed Central PMCID: PMC11011142.
14. Wang W, Cai Y, Guo Z, Zhao A, Ma W, Wang W, et al. A deep learning-based computer-aided diagnosis system for detecting atypical endometrial hyperplasia and endometrial cancer through hysteroscopy. iScience. 2025 Aug 15;28(8):113045. doi:10.1016/j.isci.2025.113045
15. Alnajjar E, Hajeri FA, Alsaad L, Harrity C, Chummun K, Rajab H. Artificial Intelligence in Minimally Invasive Gynecological Surgery: A Systematic Review of Task-Specific Performance and Clinical Translational Readiness. Journal of Minimally Invasive Gynecology. 2026 Aug 1;33(8):1110–8. doi:10.1016/j.jmig.2026.04.015
16. Rony J, Belharbi S, Dolz J, Ayed IB, McCaffrey L, Granger E. Deep Weakly-Supervised Learning Methods for Classification and Localization in Histology Images: A Survey. Melba. 2023 Mar 6;2023(March 2023):96–150. doi:10.59275/j.melba.2023-5g54
17. Ma N, Zhang X, Zheng HT, Sun J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In: Proceedings of the European Conference on Computer Vision (ECCV). 2018:116–131. doi:10.1007/978-3-030-01264-9_8
18. Chen C, Mat Isa NA, Liu X. A review of convolutional neural network based methods for medical image classification. Computers in Biology and Medicine. 2025 Feb 1;185:109507. doi:10.1016/j.compbiomed.2024.109507

19. Garcea F, Serra A, Lamberti F, Morra L. Data augmentation for medical imaging: A systematic literature review. Comput Biol Med. 2023 Jan;152:106391. doi:10.1016/j.compbiomed.2022.106391 PubMed PMID: 36549032.
20. Sun R, Wei C, Jiang Z, Huang G, Xie Y, Nie S. Weakly Supervised Breast Lesion Detection in Dynamic Contrast-Enhanced MRI. J Digit Imaging. 2023 Aug 1;36(4):1553–64. doi:10.1007/s10278-023-00846-5
21. Sun H, Zeng X, Xu T, Peng G, Ma Y. Computer-aided diagnosis in histopathological images of the endometrium using a convolutional neural network and attention mechanisms. IEEE J Biomed Health Inform. 2020 Jun;24(6):1664–76. doi:10.1109/JBHI.2019.2944977
22. Kelly BS, Judge C, Bollard SM, Clifford SM, Healy GM, Aziz A, et al. Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE). Eur Radiol. 2022 Nov 1;32(11):7998–8007. doi:10.1007/s00330-022-08784-6

Figure legends

Figure 1. Overview of the proposed framework. The EfficientNet-B4 backbone feeds two branches: (A) a classification head producing four-class probabilities, and (B) a Layer-CAM module generating attention heat maps that are refined via thresholding and morphological post-processing into lesion contours with bounding boxes.

Figure 2. Dataset distribution and classification performance. (A) Pie chart of image distribution across four classes (EC 701, EH 1,304, EP 3,281, Normal 1,201). (B) t-SNE visualisation of deep features in the test set, showing class separability. (C) Normalised confusion matrix (rows = ground truth, columns = predictions). (D) Per-class ROC curves with AUC values (macro AUC = 0.965).

Figure 3. Qualitative comparison of Grad-CAM and Layer-CAM overlays. Top row: original images; middle row: Grad-CAM heat maps; bottom row: Layer-CAM heat maps. Columns show representative samples from EH (cols 1-2), EC (cols 3-4), and EP (col 5). Layer-CAM produces more focused activations with sharper lesion boundaries.

Figure 4. Final lesion localisation results. Top row: original images; middle row: Layer-CAM heat maps; bottom row: localised lesion contours (green) with bounding boxes (yellow). Columns correspond to the same samples as in Figure 3.

Figure 5. Comprehensive comparison and ablation analysis. (A) Backbone comparison measured by macro-averaged AUC. (B) Per-class AUC heatmap for the five model configurations (four backbones plus the private-data EfficientNet-B4 variant). (C) Ablation of public data: per-class sensitivity with and without the 3,211 public images. (D) Parameter–AUC trade-off: macro-averaged AUC versus model parameters. The proposed model is marked with an asterisk.