Scientific Reports

https://doi.org/10.1038/s41598-026-39942-1

Article in Press

Automated detection of fetal vascular malperfusion via data augmentation and algorithm improvement

Received: 16 August 2025

Accepted: 9 February 2026

Published online: 25 February 2026

Cite this article as: Li X., Jiang Z., Chen F. *et al.* Automated detection of fetal vascular malperfusion via data augmentation and algorithm improvement. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-39942-1

Xuxuan Li, Zhifa Jiang, Fengchao Chen, Jianfeng Peng, Jingwen Liu, Ruoping Lin, Xiangyun Ye & Zhen Zhang

Automated Detection of Fetal Vascular Malperfusion via Data Augmentation and Algorithm Improvement

Xuxuan Li[1], Zhifa Jiang[2#], Fengchao Chen[1], Jianfeng Peng[2], Jingwen Liu[2], Ruoping Lin[2], Xiangyun Ye[2] and Zhen Zhang*[1]

Department of Soft Engineering, Huizhou University, Huizhou, Guangdong, China.

2. Obstetrics and Gynaecology, Huizhou First Maternal and Child Health Care Hospital, Huizhou, Guangdong, China

Contributing authors: lxxhnsf@163.com; jiangzf@hzu.edu.cn; fdwyyxhh@163.com; pjfjfp@163.com; gdmcljw@qq.com;603965361@qq.com; xiangyunye697@gmail.com;zzsjbme@sjtu.edu.cn

Corresponding author(s) E-mail(s): zzsjbme@sjtu.edu.cn;

Co-first author

Abstract: Fetal vascular malperfusion (FVM) is an important pathological factor leading to adverse pregnancy outcomes, but current manual diagnosis faces challenges such as high subjectivity and low efficiency. To address these problems, this paper proposes a joint analysis strategy based on data augmentation and deep learning model improvement. Using MONAI-based data augmentation it increases the number of FVM histopathology images while embedding a LocalWindow attention mechanism to enhance the YOLOv11 model. The experimental results show that this synergistic strategy of data augmentation and model improvement yields optimal recognition performance, with the F1 score, mAP50, and mAP50-95 increased by 7.84%, 6.53%, and 6.63%, respectively, compared with the YOLOv11 baseline model.

This study indicates that a strategy combining data augmentation with model structural improvement can effectively enhance detection performance for FVM and provides a useful reference for the development of intelligent diagnostic tools for FVM in clinical practice.

Keywords: fetal vascular malperfusion, computational pathology, automated detection, MONAI, LocalWindow_YOLOv11, entropy-weighted TOPSIS

1. Background

Fetal vascular malperfusion (FVM) refers to a spectrum of placental lesions resulting from chronic obstruction of blood flow in the fetal–placental circulation and is commonly associated with umbilical cord abnormalities, fetal thrombotic vasculopathy, maternal hypertension, and other factors[1,2]. Numerous studies have shown that FVM is a major cause of adverse pregnancy outcomes, including fetal growth restriction, preterm birth, and stillbirth, and may increase the risk of long-term neurodevelopmental impairments or metabolic diseases in the offspring[3]. In view of these risks, accurate diagnosis of FVM is of great clinical importance for improving maternal and neonatal health management. For the neonate, a definitive diagnosis of FVM can indicate potential long-term health risks, thereby providing crucial evidence to guide early follow-up and intervention strategies. For the mother, placental diagnosis in the current pregnancy is a key step in identifying risks in future pregnancies, facilitating the implementation of closer antenatal surveillance and preventive strategies for pregnant women at high risk, to effectively reduce the recurrence of adverse outcomes[4]. Therefore, placental pathological examination not only helps to elucidate the pathophysiological mechanisms underlying adverse pregnancy outcomes but also provides important decision support for optimizing maternal and neonatal health management strategies.

In clinical practice, the diagnosis of FVM mainly relies on the analysis of placental pathological images. Pathological studies have shown that FVM presents with a series of characteristic lesions, including stem vessel obliteration, villous stromal-vascular karyorrhexis, and seven other distinct categories of lesions[5]. These pathological changes are not only important histological diagnostic criteria for FVM but also key references for analyzing the causal relationships of lesions. Currently, lesion identification mainly depends on manual interpretation by pathologists, which is highly subjective and prone to missed or incorrect diagnoses, especially when lesions are extensive or unevenly distributed[6]. Resource-limited settings face additional challenges from pathologist shortages and inconsistent diagnostic expertise, further affecting diagnostic efficiency and accuracy. These limitations highlight the need for objective, efficient, and standardized diagnostic solutions.

Advances in computer vision have brought object detection algorithms into widespread use for biomedical image analysis[7,8]. In medical image analysis, these methods have also been actively applied and have achieved good progress. Chen et al. proposed a real-time detection model, HD-YOLO, to address low efficiency in

endometrial lesion detection. By introducing dual-domain fusion convolution and multi-scale channel fusion modules, the model significantly enhanced the extraction of multi-scale lesion features[9]. Li et al. proposed an improved YOLOv8-based model, SA-YOLO, to address the challenges of complex structures and subtle features in cardiac images, effectively improving accuracy and stability[10]. Our prior work introduced the GAMatrix-YOLOv8 model, which attained ~100% accuracy in detecting delayed villous maturation in gestational diabetes mellitus placentas, outperforming YOLOv7 and GoogleNet[11].

Despite these successes, FVM-focused automated detection remains underexplored. Meanwhile, FVM pathology images suffer from limited sample size, large differences in lesion size, and complex features, which impose higher requirements on model training and generalization capability. To address these challenges, we propose an automated FVM detection method combining MONAI-based augmentation and YOLOv11 structural improvements, with the main contributions as follows:

1. The MONAI generative model was used to enhance the pathology image dataset, effectively expanding the sample size and making lesion features clearer and more prominent, thus providing a richer and higher-quality data foundation for subsequent model training.

2. Based on the YOLOv11 framework, we introduced the LocalWindow attention module to strengthen the model's focus on local lesion information. The entropy-weighted TOPSIS algorithm was used to comprehensively evaluate the performance of various attention modules, demonstrating that the LocalWindow module offers significant advantages in improving overall model performance and helps achieve more precise lesion detection.

3. By combining the data augmentation strategy with model structural improvements, we built the LocalWindow_YOLOv11 model with stronger feature extraction capabilities. The entropy-weighted TOPSIS algorithm was then used to evaluate the comprehensive performance of different object detection models, confirming that our proposed model outperforms other mainstream object detection models in automated FVM detection tasks.

The paper proceeds as follows: Methods detail the dataset, augmentation process, model architecture, and evaluation methodology. Results present comparative and ablation experiments validating our contributions. Discussion addresses clinical

implications, methodological limitations, and avenues for future research.

2. Related work

To better position this study within the relevant research landscape, this section reviews current advances from three perspectives: the application of artificial intelligence in placental pathology analysis, data augmentation techniques, and attention mechanisms in medical imaging. Table 1 summarizes and compares the related studies discussed in this section.

2.1 Current status of AI-based placental pathology research

In recent years, artificial intelligence (AI) techniques have been widely applied to placental pathology image analysis, offering new solutions for the quantification of pathological features, prediction of pregnancy outcomes, and assistance in disease diagnosis. To address the need for automated histological assessment, Vanea et al. proposed a multi-stage deep learning framework that integrates the localization capability of RetinaNet with the classification capability of ResNet-50, thereby enabling a fully automated pipeline from nuclear detection to tissue classification[12].

In the domain of multimodal data analysis, Lee et al. used data from 131 preterm singleton placentas to systematically compare multiple algorithms, including linear regression and random forests, for predicting gestational age and neonatal birth weight. Among these, the ridge regression model achieved the best predictive performance[13]. In obstetric imaging diagnosis, Andreasen et al. applied Mask R-CNN to the analysis of placental ultrasound images and, on a private dataset of 7,500 images, achieved a classification accuracy of 81.42% and a mean IoU of 0.78, providing a novel tool for the clinical assessment of placenta previa[14]. Collectively, these studies demonstrate from different perspectives the feasibility and clinical value of AI in placental pathology image analysis and offer important methodological references for our further exploration of automated detection in placental histopathology.

2.2 Data augmentation techniques

In medical image analysis, large-scale datasets are a prerequisite for building high-performance models. However, medical studies usually face challenges such as difficulty in data acquisition, high annotation costs, and strict privacy protection requirements. As a commonly used data expansion strategy, data augmentation can

alleviate data scarcity and class imbalance by generating new training samples, thereby improving the model's generalization ability and robustness.

With respect to basic augmentation strategies, Fuji et al. evaluated the performance of several augmentation techniques in AI-assisted pathological diagnosis and found that not all augmentation methods can improve model generalization; among them, geometric transformations were shown to be a robust and effective strategy that helps enhance the model's ability to recognize cancer cells[15]. To address the class imbalance that is common in medical data, Hong et al., in a study on breast cancer recurrence prediction, used the SMOTE technique to oversample minority cases and effectively reduced training bias caused by uneven data distribution[16]. With the development of deep learning, data augmentation methods based on generative models have shown promising application prospects. Marouf et al. proposed a generative adversarial network variant named cscGAN, which can learn complex nonlinear gene dependencies and thereby generate high-quality single-cell sequencing data[17]. In the field of image segmentation, the GenSeg framework developed by Zhang et al. adopts an end-to-end, segmentation-performance-oriented generative approach and, in scenarios with very limited data, effectively improves the model's segmentation accuracy and data utilization efficiency[18]. Compared with traditional augmentation methods, such generative models can synthesize samples that are closer to the true data distribution and more diverse, providing a new solution to the problem of insufficient data in medical imaging.

2.3 Attention mechanisms in medical imaging

The attention mechanism originated from mimicking the human visual system and enables models to dynamically allocate computational resources, focusing on the most informative parts of the input data. Since the introduction of the Transformer architecture, attention has gradually evolved into one of the core techniques for improving performance and interpretability in medical image analysis.

In practice, researchers have integrated attention modules into network architectures in various ways to address specific challenges in medical image analysis. Mobadersany et al. incorporated bio-inspired attention and aggregation mechanisms into the GestAltNet model, enabling estimation of gestational age by assessing placental maturity and achieving performance superior to that of baseline models[19].

Parathanath et al. proposed the CBMNet model, which integrates dual attention

mechanisms (CBAM and MSAM) into a ConvNeXt-Tiny backbone to jointly enhance the model's sensitivity to local lesion details and global contextual information, achieving performance that surpasses multiple baseline models on dental caries classification[20]. For medical segmentation tasks, Abrar et al. integrated attention gates into the U-Net architecture and proposed the ACU-Net model. By strengthening the focus on features within tumor regions and suppressing irrelevant background information, ACU-Net effectively captures complex tumor morphologies[21]. These studies demonstrate the application potential of attention mechanisms in medical imaging and underscore their substantial contribution to improving both model accuracy and interpretability.

Table 1 Summary and comparison of related studies

Author	Method	Dataset	Performance	Limitation
Vanea et al.	Three-stage deep learning–based pipeline	Multicenter whole-slide images of human placental tissue	Nuclear detection F1-score: 0.884; cell classification accuracy: 84.29%	Model performance is affected by variability in tissue staining
Lee et al.	Machine-learning regression analysis based on gross placental morphological features	Gross placental images from 131 preterm singleton pregnancies	Coefficient of determination of ridge regression reached 0.77	Limited sample size
Andreasen et al.	Mask R-CNN–based placental classification, localization, and segmentation	7,500 obstetric ultrasound images	Image classification accuracy: 81.42%; mean IoU for placental segmentation: 0.78	Lower performance on images acquired in early pregnancy
Fuji et al.	Comparison of data augmentation techniques using an AI system	4,159 cytology slide images	Accuracy increased to 88.2% after geometric transformations	Lack of external validation
Hong et al.	Handling class imbalance using SMOTE	Blood microbiome data from 229 participants	Random forest AUROC: 0.94; recall: 0.81	Lack of external validation
Marouf et al.	Conditional single-cell generative adversarial network (cscGAN)	Multiple scRNA-seq datasets	Generated high-quality single-cell data	Training depends on high-performance GPUs
Zhang et al.	Multi-level optimization–based generative framework (GenSe)	Nineteen medical imaging datasets	Segmentation performance improved by 10–20%	Quality of generated data strongly depends on the original data used for initialization
Mobadersany et al.	Deep learning model combining attention mechanisms and feature aggregation	154 placental tissue sections	Coefficient of determination on the test set reached 0.9444	Lack of multicenter validation

Parathanath et al.	ConvNeXt-Tiny combined with dual attention mechanisms	1,103 periapical radiographs	Mean validation accuracy: 93.26%	Lack of multicenter validation
Abrar et al.	U-Net model incorporating attention gates	BraTS 2018 MRI dataset	Whole-tumor Dice: 94.04%; tumor core: 98.63%; enhancing tumor: 98.77%	Evaluated on a single dataset only

3. Methods and Materials

3.1 Dataset description:

Data Source: The data analyzed in this study were derived from the prospective cohort project entitled "Integrative modeling and metabolomics study of multimodal lipid metabolism characteristics in gestational diabetes mellitus." The study population comprised 310 women with singleton pregnancies complicated by gestational diabetes mellitus (GDM) who received prenatal care and delivered at Huizhou First Maternal and Child Health Care Hospital, Guangdong Province, China, with gestational age at delivery between 37 and 41 weeks.

Sample processing and annotation: After delivery, placental specimens were sampled by pathologists in accordance with the consensus statement of the Amsterdam Placental Workshop Group[5]. For each placenta, at least four standard regions were collected: one roll of placental membranes, one cross-sectional block of umbilical cord, and two full-thickness sections of placental parenchyma. All tissue blocks were fixed in 10% neutral buffered formalin and stained using a standard hematoxylin and eosin (H&E) protocol. All slides were then digitized under identical settings using the same whole-slide imaging system (Motic Digital Slide Scanning and Application System, Motic Industrial Group Co., Ltd.). During image acquisition, in view of the substantial differences in lesion size, an appropriate magnification was selected for each lesion type (detailed parameters are provided in Supplementary Table 1), while the image resolution was uniformly set to 4000×3000 pixels. The digitized histopathological images were independently diagnosed and annotated by two junior pathologists, and their annotations were subsequently compared. Regions with concordant labels were accepted directly. For discordant regions, the two pathologists jointly reviewed the original glass slides and reassessed the findings to reach a final annotation. If consensus could not be achieved after discussion, the case was further reviewed by a senior

pathologist, who made the final diagnostic decision.

Dataset partitioning: For model training and performance evaluation, the 310 patients were randomly assigned to the training, validation, and test sets at a ratio of 8:1:1. Subsequently, an equal number of images was cropped from the pathological slides of each patient to construct the three data subsets. In total, the training, validation, and test sets contained 1,892, 234, and 242 images, respectively. This dataset encompassed seven typical lesion types associated with FVM, as illustrated in Figure 1: Delayed villous maturation, Stem vessel obliteration, Villous stromal-vascular karyorrhexis, Intramural fibrin deposition, Vascular ectasia, Thrombosis, and Avascular villi.

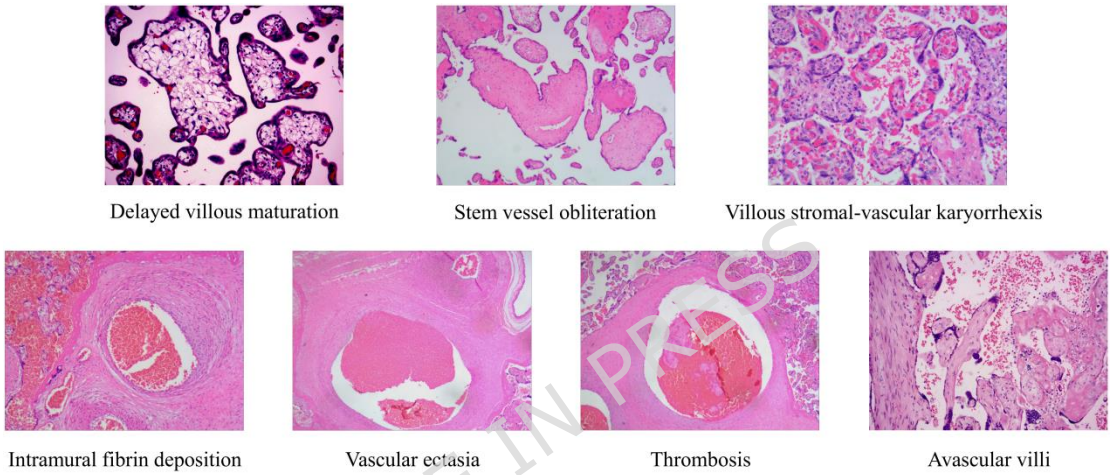

Figure 1 Seven typical lesion types in FVM

3.2 MONAI-based Data Augmentation

MONAI is an open-source medical imaging framework integrating GANs, Transformers, and other state-of-the-art architectures to support image synthesis, augmentation, and reconstruction[22]. Its hierarchical, modular design (Supplementary Figure 1) provides stable data transformation, network, and loss function modules at the core, overlaid with automated workflow control from preprocessing through training, validation, and inference. MONAI's support for federated learning enables privacy-preserving, multi-center collaborations without raw data exchange[23].

Leveraging the strengths of the MONAI framework, this study designed a comprehensive augmentation scheme tailored to the morphological characteristics of placental histopathology images and clinical requirements, encompassing three dimensions: geometric transformations, image quality, and staining properties. These

augmentation strategies can simulate complex and variable clinical scenarios, thereby enriching the diversity of pathological image types in the training set. In terms of geometric transformations, elastic deformation and random flipping were used to mimic artifacts such as folding and stretching introduced during tissue processing, as well as variations in slide orientation. With respect to image quality, Gaussian smoothing and Gaussian sharpening were applied to modulate high-frequency components, thereby emulating differences in imaging quality across various digital pathology scanners. Regarding staining properties, contrast adjustment nonlinearly altered the distribution of pixel intensities, effectively simulating variations in H&E staining intensity caused by differences in reagent concentration, incubation time, or batch effects. The parameter settings for the above data augmentation strategies are summarized in Supplementary Table 2.

3.3 LocalWindow_YOLOv11

3.3.1 YOLOv11 overview: YOLOv11 is an end-to-end object detection algorithm based on deep learning, which transforms object detection into an efficient regression problem through convolutional neural networks, significantly improving detection efficiency and real-time performance[24]. This advantage has led to its widespread application in medical image analysis. Its architecture comprises:

- Backbone: Multi-scale feature extraction via stacked convolutions with efficiency modules (C3k2, SPPF, C2PSA).

- Neck: FPN + PAN for cross-scale feature fusion.

- Head: Decoupled head for independent classification and localization.

3.3.2 LocalWindow attention module: In medical images, lesions are often small and subtle, making them easy to overlook in conventional CNN-based feature extraction, which reduces detection accuracy. To address this, we introduced the LocalWindow attention module into YOLOv11 to enhance multi-scale lesion detection. As shown in Figure 2, the module consists of local window partitioning and cascaded group attention (CGA)[25].

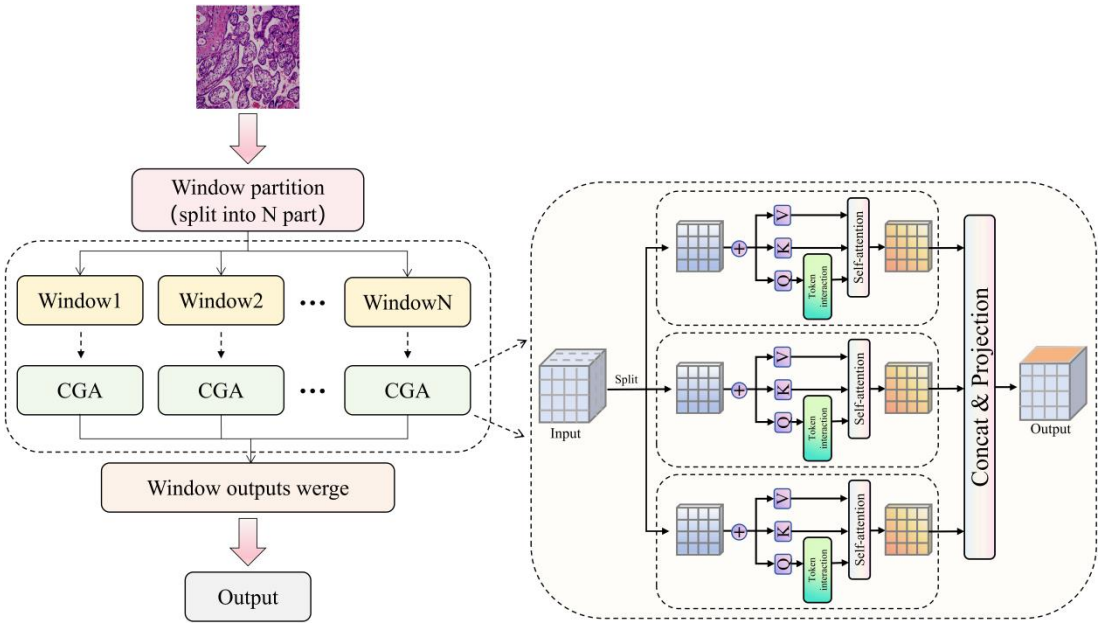

Figure 2 Structure of the LocalWindow attention module

The LocalWindow attention module is designed to overcome the limitations of conventional global attention mechanisms when processing high-resolution pathology images. Specifically, traditional global attention computes the relationships between any given region and all other regions in the image, which, for high-resolution pathology slides, not only incurs a substantial computational burden but also introduces a large amount of irrelevant background information, thereby impairing the extraction of subtle lesion features[26]. In contrast, the LocalWindow attention module partitions the feature map into multiple non-overlapping local windows and constrains attention computation within each window. This focus on local regions greatly reduces redundant computation and improves computational efficiency, while excluding interference from distant, irrelevant background areas. As a result, the model can concentrate more effectively on fine-grained features within each window, particularly subtle changes in boundary texture and cellular morphology of small lesions.

Specifically, the input feature map is divided into multiple non-overlapping 7×7 windows. Each local window's features are processed by the CGA module, where features are grouped and assigned to different attention heads. Within each head, grouped features are projected into different subspaces, and attention is computed via Q, K, and V parameters to enhance inter-feature communication. Finally, outputs from all heads are concatenated and linearly transformed back to the original channel dimension.

The specific formula is as follows:

$$\tilde{X}_{ij} = \mathrm{Attn}\left(X_{ij}W_{ij}^{Q}, X_{ij}W_{ij}^{K}, X_{ij}W_{ij}^{V}\right) \quad (1)$$

$$\tilde{X}_{i+1} = \mathrm{Concat}\left[\tilde{X}_{ij}\right]_{j=1:h} W_{i}^{P} \quad (2)$$

Here, X_{ij} represents the j-th segment of the input feature X_i, $\tilde{X}_{ij}$ represents the output feature of the j-th attention head, h is the number of attention heads and $i \le j \le h$. w_{ij}^{Q}, w_{ij}^{K} and w_{ij}^{V} represents the projection layers of Q, K, V. w_{i}^{P} represents the linear layer, which is used to project the concatenated output features back to the same channel dimension as the input features.

In addition, to further enhance the feature expression capability, the CGA module adopts a cascaded structure. The output features of each attention head are added to the input features of the next attention head, which realizes the fusion and transmission of cross-head information, thereby improving the feature utilization of the model. The specific formula is as follows:

$$X_{ij}^{'} = X_{ij} + \tilde{X}_{i(j-1)} \quad (3)$$

Here, $X_{ij}^{'}$ represents the sum of the calculation result of the j-th segmented feature and the (j-1)-the attention head, and serves as the new input feature of the j-th attention head.

LocalWindow attention module: To enhance small-lesion detection, the LocalWindow module partitions features into 7×7 non-overlapping windows processed by Cascaded Group Attention (CGA) [25]. CGA groups features per attention head, projects them into subspaces, and applies Q–K–V attention before concatenating and linearly projecting outputs back to input dimensionality.

3.3.3 LocalWindow_YOLOv11: Based on these improvements, we propose the LocalWindow_YOLOv11 model (Figure 3), which integrates the LocalWindow module into the tail of the Backbone. This design improves the model's ability to focus on local lesion areas—especially boundary regions, thereby addressing challenges posed by multi-scale lesions and improving detection accuracy.

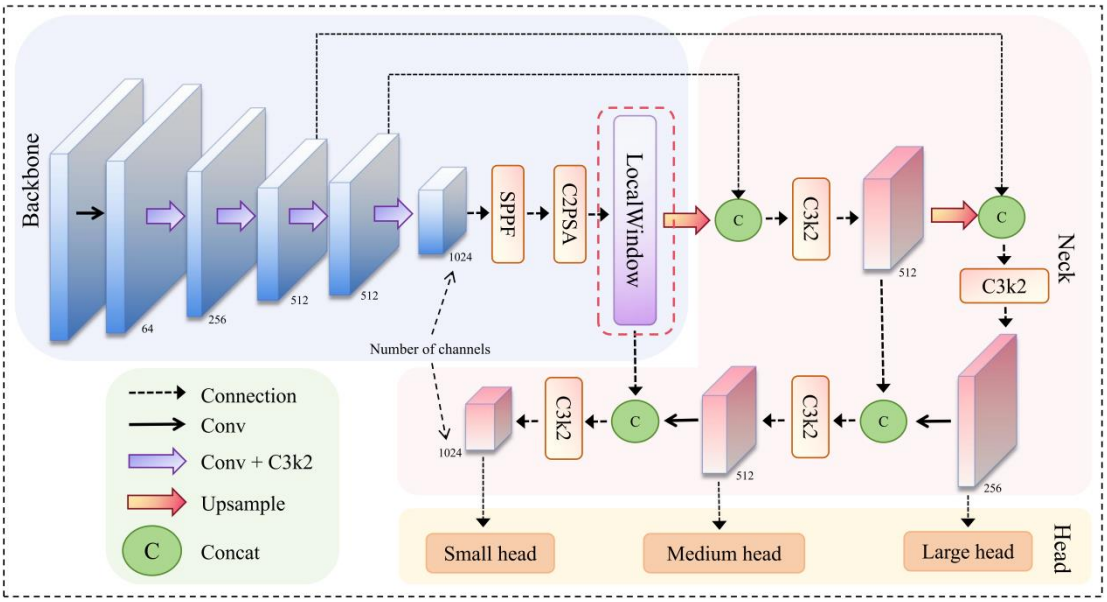

Figure 3 Network architecture of LocalWindow_YOLOv11

3.4 Experimental Setup and Metrics

The experiments were conducted in the PyCharm environment using Python 3.10.16 and PyTorch 2.2.2. The hardware setup included an AMD EPYC 7K62 processor, an NVIDIA RTX 3090 GPU with 24 GB VRAM, and 64 GB DDR4-3200 memory.

To comprehensively evaluate performance, we used Recall, Precision, F1-score, mean Average Precision (mAP), giga floating-point operations per second (GFLOPs), number of model parameters (Params), and frames per second (FPS) as metrics. GFLOPs quantify the model's computational complexity, Params represent the total number of parameters, and FPS measures the number of image frames processed per second. Together, these metrics provide a comprehensive assessment of the model's detection accuracy, efficiency, and computational demands, enabling a more precise evaluation of its practical applicability.

To more rigorously assess the performance gains achieved through the MONAI augmentation strategy, this study employs the TIDE (Task-specific Integrated Detection Evaluation) framework[27]. TIDE systematically isolates six categories of detection errors—classification (Cls), localization (Loc), combined classification and localization (Cls+Loc), duplicate detections (Duplicate), background misclassification (Bkgd), and missed detections (Missed)—and quantifies the impact of each on overall model performance. As illustrated in Figure 4, smaller metric values indicate a lower

negative contribution from the corresponding error type. For example, Cls denotes correct localization with an incorrect class label, while Missed refers to undetected ground-truth targets.

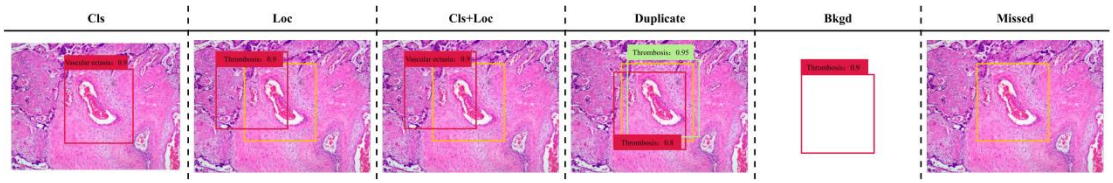

Figure 4 Error type definitions(example shown for a Thrombosis lesion). Bounding box color scheme: Red: False positive detection, Yellow: Ground truth, Green: True positive detection

3.5 Model performance evaluation and statistical analysis

3.5.1 Entropy-weighted TOPSIS algorithm

In this study, the entropy-weighted TOPSIS method is adopted to objectively evaluate the integrated performance of various attention modules and model architectures. Unlike traditional evaluation approaches that rely heavily on subjective weighting, this method leverages information entropy to automatically determine indicator weights, thereby minimizing human bias and ensuring greater scientific rigor[28]. Performance metrics—including Precision, Recall, F1-score, mAP50, and mAP50-95—are jointly considered. By calculating each model's relative proximity to an ideal best solution and distance from an ideal worst solution, the method produces a comprehensive performance score. A higher score indicates closer alignment with optimal performance, enabling a robust, data-driven selection of the best model configuration[29].

3.5.2 Wilcoxon signed-rank test

Because the model performance in this study was evaluated across seven lesion types, the sample size for each type was relatively small, and the normality of the performance metrics could not be guaranteed. Therefore, we employed a nonparametric method that does not require a normal distribution—the Wilcoxon signed-rank test—to perform paired comparisons between each improved model in the ablation experiments and the baseline model[30]. This test analyzes paired values of Precision, Recall, F1-score, mAP50, and mAP50-95 for the improved and baseline models to determine whether the observed performance gains are statistically significant ($P < 0.05$), thereby providing more robust evidence for the effectiveness of the model improvements.

The overall research workflow is outlined in Figure 5, integrating all aforementioned methodological components into a cohesive evaluation pipeline.

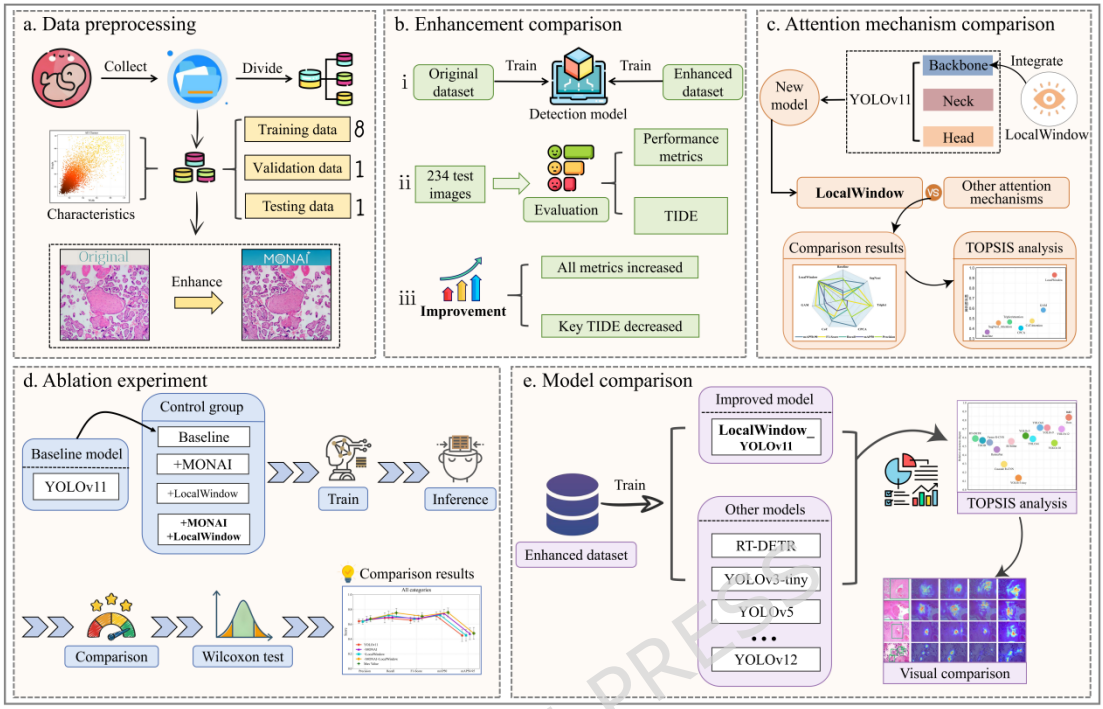

Figure 5 Study workflow diagram

4. Experiments and Results

4.1 Lesion Analysis of the Samples

We performed a statistical analysis of lesion quantities and size distributions across various lesion types in this study. As illustrated in Figure 6, lesion type distribution is markedly imbalanced: villous stromal-vascular karyorrhexis and avascular villi are most prevalent (2,211 and 2,210 instances), whereas thrombosis and intramural fibrin deposition are rare, with only 295 instances each. Lesion size distribution is similarly skewed—most lesions are small, but some are substantially larger and more scattered in distribution.

In summary, the dataset shows a significant imbalance in both class frequency and lesion size distribution, which not only increases the difficulty of model training but also imposes higher demands on the model's generalization ability and real-world diagnostic robustness.

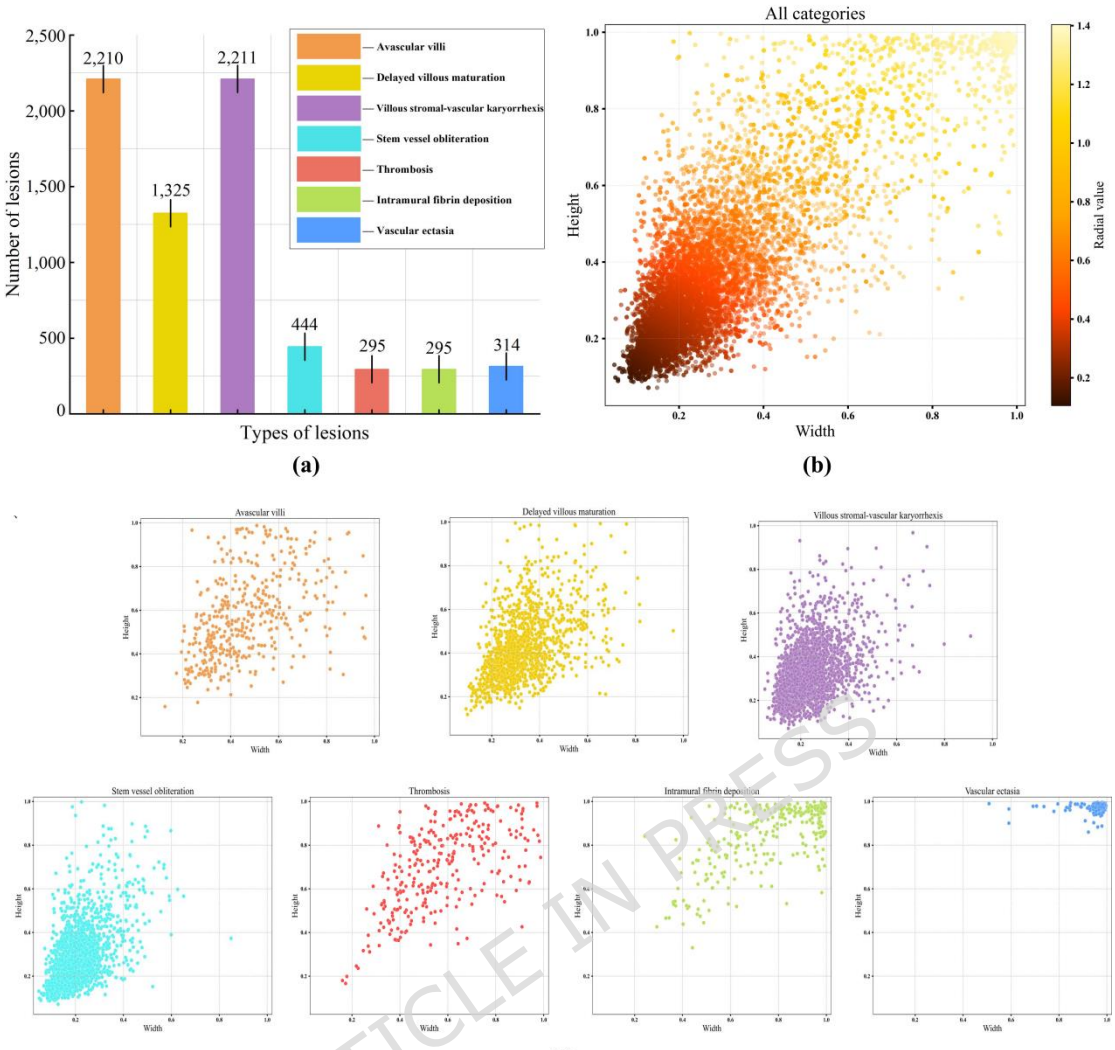

Figure 6 Lesion count and size distribution in the dataset

4.2 MONAI augmentation analysis

4.2.1 Image augmentation effects

In this experiment, data augmentation was performed on the 1,892 images in the original training set using the MONAI framework. Each image underwent a series of augmentation operations, including elastic deformation, random flipping, Gaussian smoothing, Gaussian sharpening, and contrast adjustment, with the parameters of each transformation randomly sampled within predefined ranges to generate one augmented counterpart for each original image. The expanded training set, therefore, contained 1,892 original images and 1,892 augmented images, for a total of 3,784 images. Meanwhile, the sizes of the validation and test sets were kept unchanged to ensure objective and unbiased model evaluation. As illustrated in Figure 7, the augmented images strictly preserved the integrity of pathological features while effectively simulating real-world conditions encountered during medical image acquisition, such as tissue deformation, viewpoint variations caused by differences in slide orientation,

and fluctuations in scanner image quality. According to evaluations by pathologists, the MONAI-generated images showed a high degree of consistency with the original slides in key diagnostic regions. Through these augmentation procedures, the size of the training set was effectively increased, providing richer and higher-quality medical image data for subsequent model training.

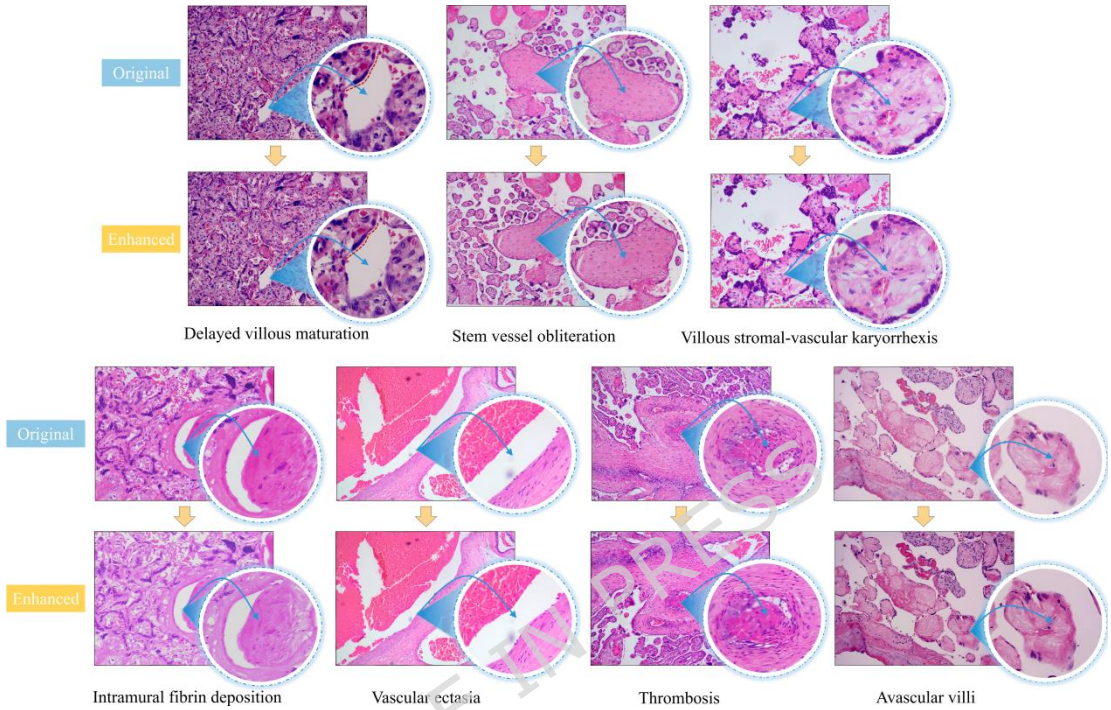

Figure 7 Training image samples before vs. after data augmentation

4.2.2 Impact of MONAI augmentation on performance

To evaluate the impact of MONAI-based image augmentation on detection performance, we designed a set of comparative experiments comprising three settings: baseline, data duplication, and data augmentation. The results are summarized in Table 2, where the percentages indicate the relative improvement of each metric compared with the YOLOv11 baseline model.

The experimental results showed that, after training with the "data duplication" strategy, the model's performance declined markedly compared with the baseline. This indicates that merely increasing the dataset size by duplicating samples not only fails to improve performance but may also exacerbate overfitting on a limited dataset and thereby impair the model's generalization ability. By contrast, the model trained with the MONAI augmentation strategy exhibited clear performance gains. Its mAP50 and mAP50-95 increased to 0.7440 and 0.4698, respectively, and both Precision and Recall improved steadily. In terms of error types, classification errors E_{Cls}, duplicate detection errors E_{Dupe}, and background errors E_{Bkg} all decreased. Notably, missed detection errors E_{Miss} dropped from 0.24 to 0.03, corresponding to a reduction of

87.50%, indicating a substantial enhancement in the model’s ability to detect lesions.

Taken together, these findings indicate that the observed performance improvement is primarily attributable to the increased data diversity introduced by the MONAI augmentation strategy, rather than to a mere increase in sample size. This confirms that, for FVM lesion detection, data diversity is a more critical driver of model performance than data quantity alone.

Table 2 Model performance before vs. after MONAI augmentation

Category	Metric	YOLOv11	YOLOv11 + Duplication	YOLOv11 +MONAI
Core Performance	Precision	0.6329	0.6071 (-4.08%)	0.6622 (+4.63%)
	Recall	0.6847	0.6633 (-3.13%)	0.6968 (+1.77%)
	F1-Score	0.6578	0.6340 (-3.62%)	0.6791 (+3.24%)
	mAP50	0.7117	0.6905 (-2.98%)	0.7440 (+4.54%)
	mAP50-95	0.4463	0.4638 (+3.92%)	0.4698 (+5.27%)
CVPR-standard TIDE framework	E_{Cls} ↓	6.30	4.37 (-30.63%)	2.57 (-59.21%)
	E_{Loc} ↓	5.55	6.22 (+12.07%)	6.38 (+14.95%)
	E_{Both} ↓	0.93	1.04 (+11.83%)	0.73 (-21.51%)
	E_{Dupe} ↓	0.90	1.08 (+20.00%)	0.63 (-30.00%)
	E_{Bkg} ↓	6.30	5.40 (-14.29%)	5.92 (-6.03%)
	E_{Miss} ↓	0.24	0.28 (+16.67%)	0.03 (-87.50%)

4.3 Comparative Analysis of Attention Modules

To assess the effectiveness of the LocalWindow attention module for lesion

detection, we conducted head-to-head comparisons with several mainstream modules: multi-scale separable convolution (SegNext), multi-dimensional interaction (Triplet), channel–position convolution (CPCA), convolution–attention fusion (Coordinate), and global attention (GAM). As shown in Table 3, LocalWindow achieved the highest recall (0.7516), substantially exceeding that of all other modules, and improved mAP50 to 0.7582, confirming its ability to significantly enhance detection performance. While its precision was marginally lower than CPCA and Coordinate, it still surpassed the baseline model, reflecting a clear overall advantage.

Table 3 Performance of different attention modules

Model	Precision	Recall	F1-Score	mAP 50	mAP 50-95
Baseline	0.6622	0.6968	0.6791	0.7440	0.4698
SegNext[31]	0.6826	0.7275	0.7043	0.7378	0.4635
Triplet[32]	0.6020	0.7483	0.6672	0.7408	0.4737
CPCA[33]	**0.6953**	0.6880	0.6916	0.7439	0.4682
Coordinate[34]	0.6755	0.7271	0.7004	0.7462	0.4548
GAM[35]	0.6581	0.7445	0.6986	0.7486	0.4596
LocalWindow	0.6716	**0.7516**	**0.7094**	**0.7582**	**0.4759**

To more objectively evaluate the comprehensive performance of each attention module, this experiment further adopted the entropy-weighted TOPSIS algorithm for multi-metric analysis. Figure 8 shows that LocalWindow attained both the shortest distance to the positive ideal solution and the greatest distance from the negative ideal solution, yielding a relative closeness score of 0.93, the highest among all evaluated modules. This demonstrates that the LocalWindow module significantly outperforms

other mainstream attention modules and baseline models across multiple performance metrics, further validating its robust effectiveness in lesion detection tasks.

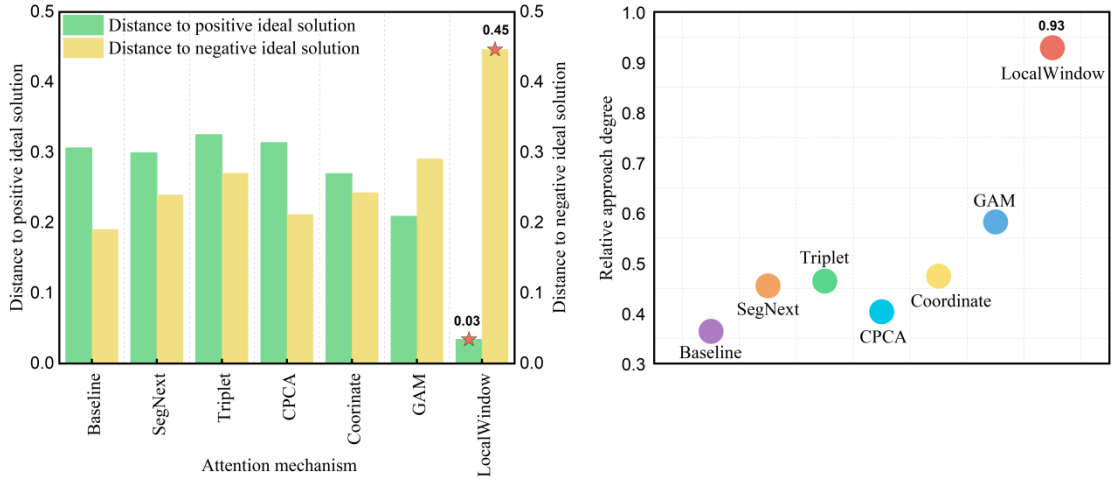

Figure 8 Comparative performance of different attention modules

4.4 Ablation Study

We systematically evaluated the synergistic effect of the MONAI augmentation strategy and the LocalWindow module using ablation experiments (Table 4), where percentages indicate improvement over the YOLOv11 baseline.

Individually, both MONAI augmentation and the LocalWindow module boosted all key performance metrics, demonstrating their respective benefits for detection accuracy and generalization. When integrated, the combined system achieved optimal results, with recall rising to 0.7516 and mAP50 to 0.4759—improvements of 9.77% and 6.63%, respectively.

Table 4 Ablation study results

Model	Precision	Recall	F1-Score	mAP50	mAP50-95
YOLOv11	0.6329	0.6847	0.6578	0.7117	0.4463
+ MONAI	0.6622 (+4.63%)	0.6968 (+1.77%)	0.6791 (+3.24%)	0.7440 (+4.54%)	0.4698 (+5.27%)
+LocalWindow	0.6346 (+0.27%)	0.7194 (+5.07%)	0.6743 (+2.51%)	0.7212 (+1.33%	0.4488 (+0.56%)

				)	
+LocalWindow + MONAI	**0.6716** (+6.11%)	**0.7516** (+9.77%)	**0.7094** (+7.84%)	**0.7582** (+6.53%)	**0.4759** (+6.63%)

To verify that the performance gains described above were statistically significant, we applied the Wilcoxon signed-rank test to assess the effects of the model modifications, with the results summarized in Table 5. The statistical analysis showed that when MONAI-based data augmentation alone (model 2) or the attention module alone (model 3) was introduced, most improvements in performance metrics did not reach statistical significance. However, when both components acted in combination (model 4), the improvements in Recall, F1-score, mAP50, and mAP50-95 all reached statistical significance. Although the improvement in Precision did not achieve statistical significance, it still represented a 6.11% relative increase compared with the baseline model (Table 4), and its trend was consistent with the overall upward pattern observed for the other metrics.

Table 5 Significance testing of ablation experiments

Metric	Model 2 vs Model 1	Model 3 vs Model 1	Model 4 vs Model 1
Precision	0.4688	0.5938	0.5938
Recall	0.2891	0.0547	0.0112*
F1-Score	0.2891	0.1484	0.0391*
mAP50	0.0547	0.1484	0.0391*
mAP50-95	0.0234*	0.3438	0.0234*

To qualitatively verify the above performance improvements, Figure 9 presents a comparison of the detection outputs of each experimental group for the seven typical lesion types. From the visualizations, it can be seen that the synergistic integration group (model 4) not only shows higher prediction confidence but also achieves more complete lesion detection than the other models.

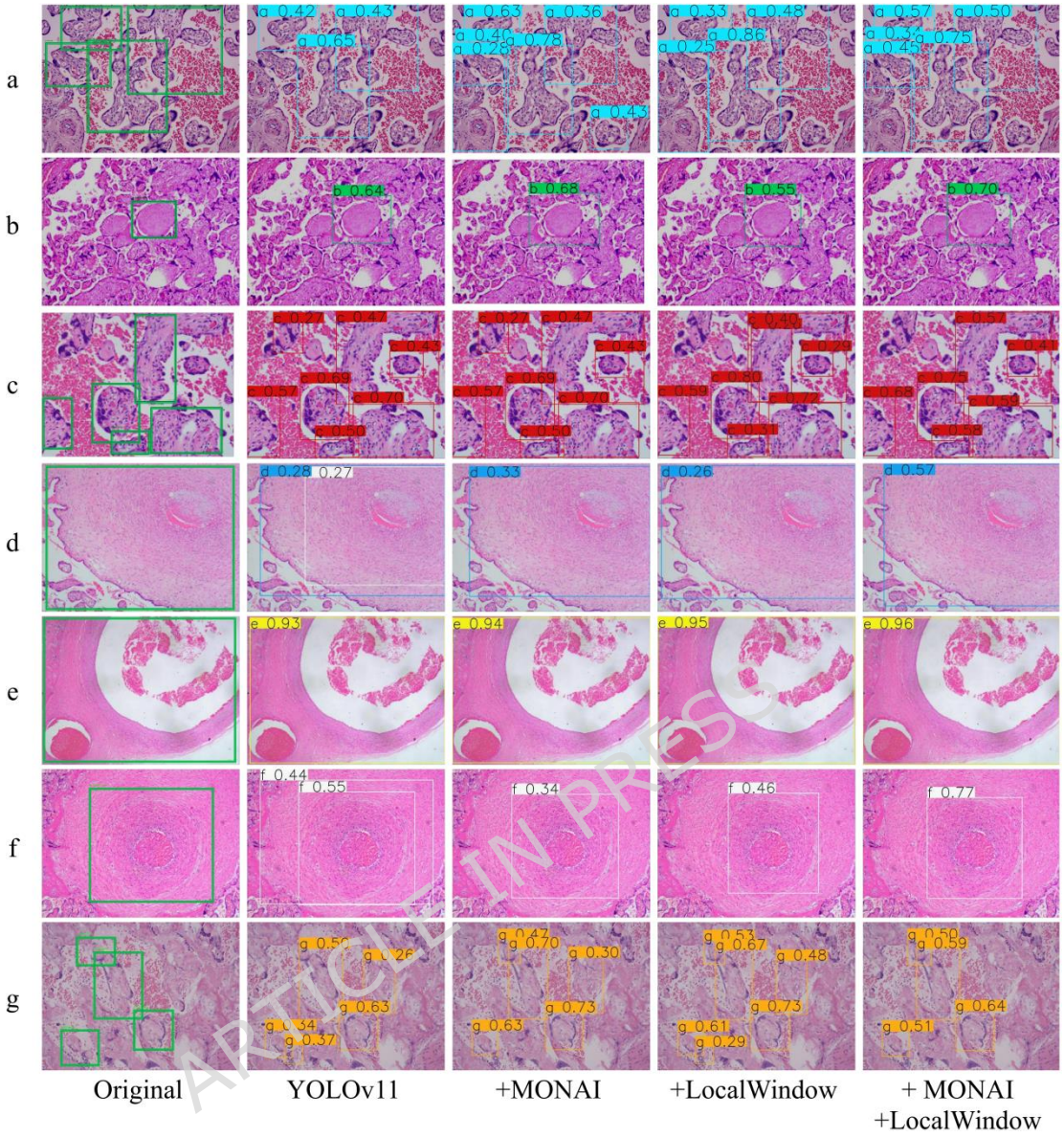

Figure 9 Comparison of detection results in the ablation experiments. Panels a-g correspond, respectively, to Delayed villous maturation, Stem vessel obliteration, Villous stromal-vascular karyorrhexis, Intramural fibrin deposition, Vascular ectasia, Thrombosis, and Avascular villi.

Figure 10 illustrates that the combined system excelled in detecting thrombosis and intramural fibrin deposition while maintaining high accuracy in other lesion types, such as delayed villous maturation and avascular villi. These findings underscore its robustness and adaptability across diverse pathological conditions.

Overall, the synergy between MONAI augmentation and the LocalWindow module not only enhanced comprehensive detection capability but also demonstrated good adaptability and consistency across various lesion detection tasks.

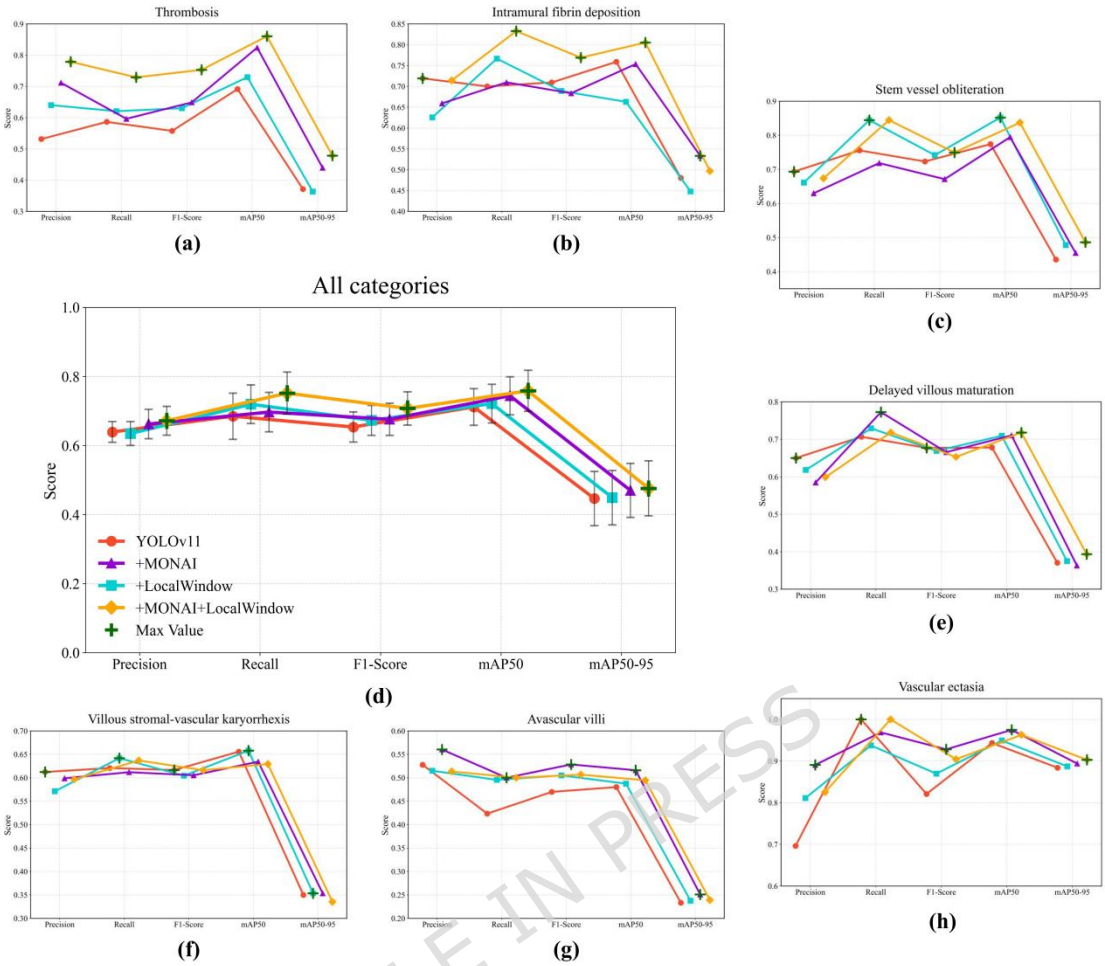

Figure 10 Performance comparison of each experimental group in detecting different lesion types. The error bars in the figure represent the mean±standard error.

4.5 Comparative Evaluation with Other Models

To comprehensively evaluate the performance of LocalWindow_YOLOv11, we compared it with models from the YOLO family as well as other mainstream object detection frameworks. The experimental results are summarized in Table 6. In terms of detection performance, LocalWindow_YOLOv11 achieved 0.7582 and 0.4759 on the two key metrics mAP50 and mAP50-95, respectively, outperforming all other models and demonstrating its superior capability in handling FVM lesions with diverse scales and morphologies. Meanwhile, the model attained an F1-Score of 0.7094, indicating a favorable balance between Precision and Recall and suggesting that it can effectively reduce both missed and false detections of lesions. With respect to computational efficiency, LocalWindow_YOLOv11 maintains relatively low computational cost and parameter count, which is advantageous for deployment and application in clinical

settings. Although its inference speed (153.47 FPS) is slightly lower than that of some lightweight models, it still offers strong real-time processing capabilities and can meet the detection speed requirements of practical clinical applications.

Table 6 Performance comparison of different detection models

Model	Precision	Recall	F1-Score	mAP50	mAP50-95	GFLOPs	Params(M)	FPS
RT-DETR	0.6565	0.7054	0.6801	0.6828	0.4375	103.46	32.00	96.71
Faster R-CNN	0.6280	0.7070	0.6652	0.6980	0.4270	178.00	41.38	50.00
Cascade R-CNN	0.5810	0.7300	0.6470	0.6800	0.4340	205.00	69.24	32.60
TOOD	0.6330	0.7090	0.6688	0.6960	0.4460	168.00	32.03	26.80
RetinaNet	0.5370	0.7240	0.6166	0.6270	0.3750	176.00	36.45	51.30
RTMDet	0.6280	0.7040	0.6638	0.7050	0.4440	8.03	4.88	51.80
YOLOv3-tiny	0.5486	0.6669	0.6020	0.6432	0.3265	18.93	12.13	260.84
YOLOv5	0.6531	0.6906	0.6713	0.7236	0.4646	7.07	2.50	296.33
YOLOv6	**0.7333**	0.6307	0.6781	0.7115	0.4448	11.78	4.23	138.72
YOLOv8	0.6573	0.7267	0.6903	0.7259	0.4496	8.09	3.01	119.49
YOLOv9	0.6610	0.7208	0.6896	0.7252	0.4581	7.61	1.97	127.83
YOLOv10	0.6772	0.6505	0.6636	0.7078	0.4373	6.53	2.27	276.97
YOLOv12	0.6235	**0.7626**	0.6861	0.7354	0.4621	6.33	2.56	114.90
Ours	0.6716	0.7516	**0.7094**	**0.7582**	**0.4759**	7.77	3.03	153.47

To further quantify the overall performance of each detection model, we conducted a comprehensive evaluation of multiple performance metrics using an entropy-weighted TOPSIS method, as shown in Figure 11. It can be seen that LocalWindow_YOLOv11 ranks first in terms of the composite score across all metrics and is closest to the ideal model in the TOPSIS space (a schematic of the ideal distance

is provided in Supplementary Figure 2). This further confirms that its overall performance surpasses that of other mainstream detection models and indicates greater potential for practical application.

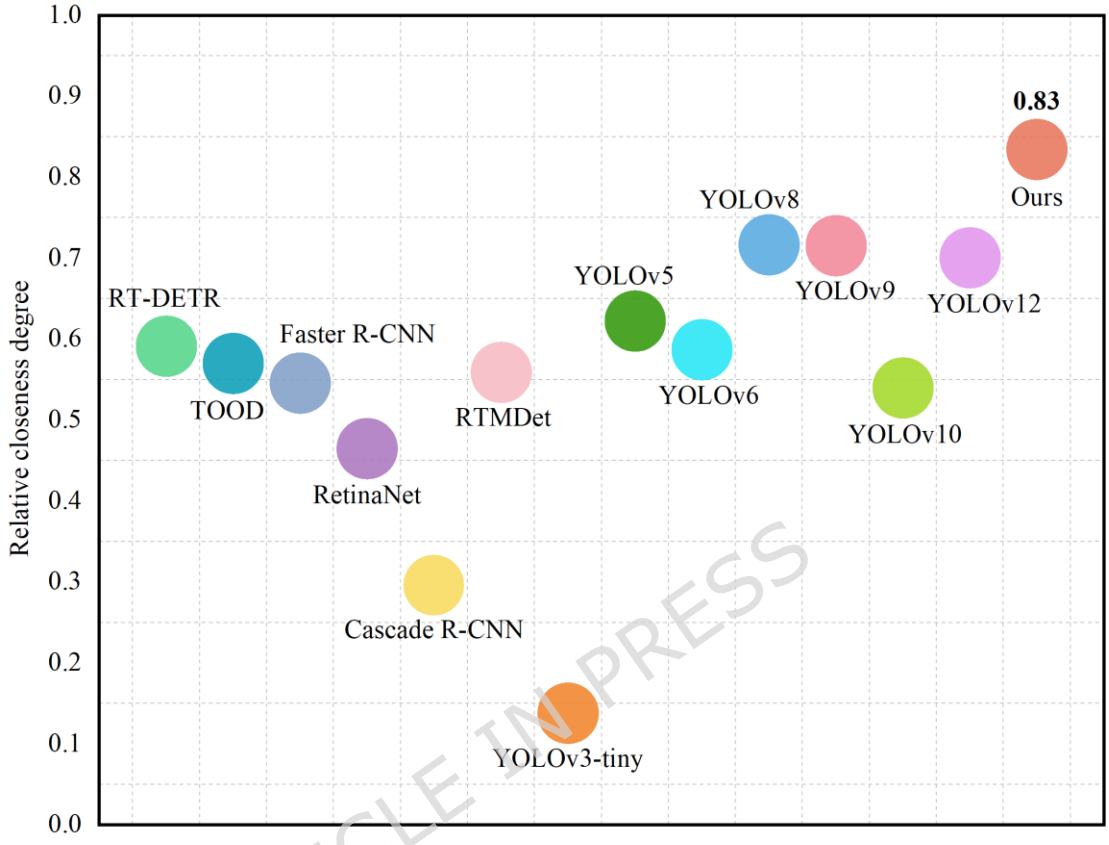

Figure 11 Comparison of the overall performance of different detection models

Finally, a comparative analysis of detection results between LocalWindow_YOLOv11 and models such as YOLOv8, YOLOv9, and RT-DETR is shown in Figure 12. Heatmaps show that YOLOv8, YOLOv9, and RT-DETR often produced dispersed or weak responses in certain lesion detections, increasing the risk of misdetection and missed detection. In contrast, LocalWindow_YOLOv11 exhibited stronger response intensity in lesion areas, accurately locating and identifying corresponding lesions, demonstrating its stability in complex lesion detection tasks.

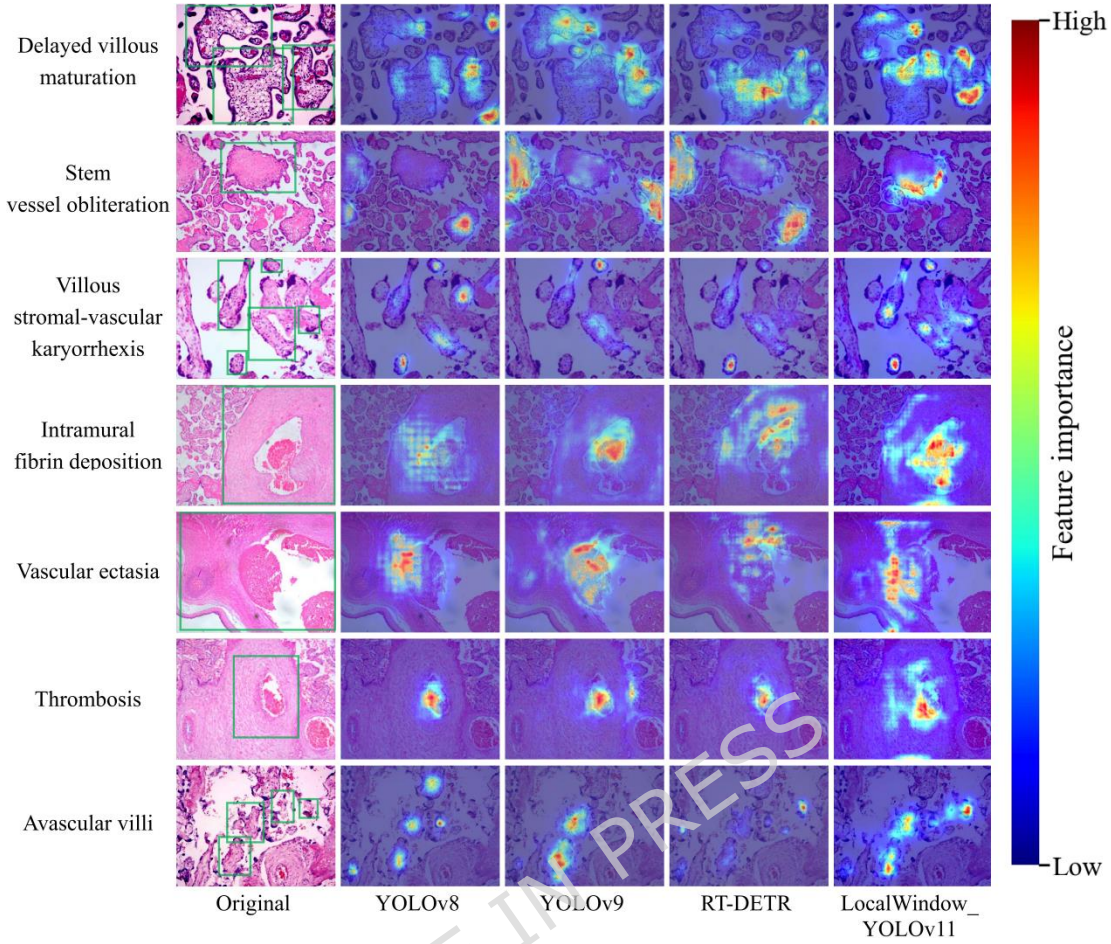

Figure 12 Heatmap comparison of lesion detection: YOLOv8 vs. YOLOv9 vs. RT-DETR vs. LocalWindow_YOLOv11

In conclusion, LocalWindow_YOLOv11 achieved a good balance between detection accuracy, operational efficiency, and model complexity. Compared to mainstream object detection models, it not only showed higher accuracy in lesion detection tasks but also balanced inference speed and resource consumption, offering strong application value and broad prospects for promotion in lesion screening and diagnosis.

5. Discussion

5.1 Effectiveness of the MONAI Augmentation Strategy

Despite notable advances in automated pathological image detection, progress remains constrained by challenges in data acquisition and limited sample sizes. To mitigate data scarcity, data augmentation is widely applied in medical imaging[36,37] .

Traditional augmentation, largely based on geometric transformations and pixel-level manipulations, can modestly increase variability but fails to capture the complex variations encountered in real-world clinical imaging. Recent work has shifted toward deep learning-based augmentation—such as Generative Adversarial Networks (GANs) and autoencoders[38-40]—which generate more realistic variants and enhance model robustness in complex scenarios.

Against this background, we adopted MONAI, a deep learning framework specifically designed for medical imaging. It integrates a variety of targeted augmentation operations that more faithfully reflect the diversity of tissue structures and the variability of imaging conditions encountered in clinical practice[23]. Compared with conventional augmentation methods, MONAI can generate plausible image variants while preserving the authenticity of pathological features, thereby reducing, to some extent, the risk that generic augmentation techniques may distort or obscure key diagnostic characteristics. Our experimental results further demonstrate that this MONAI-based augmentation strategy not only improves the model's recall and detection accuracy but also reduces critical error types such as missed and false detections. In addition, its dictionary-based pipeline architecture enables synchronized processing of images and annotations, ensuring that coordinate information is accurately updated during spatial transformations. Moreover, MONAI supports independent random-state management. Unlike traditional approaches that rely on a single global random seed, MONAI allows separate random states to be specified for the augmentation pipeline, ensuring that augmentation operations are not affected by other parts of the program and thereby enhancing the rigor and reproducibility of the experiments.

5.2 Effectiveness of Model Structure Improvement

Placental pathological images often present large size variability, indistinct lesion boundaries, and complex morphological patterns. These characteristics make it difficult for single-scale feature extraction methods to balance the fine-grained features of small lesions and the global information of large lesions, leading to missed detections and false positives. To address these detection challenges, existing research has attempted to introduce multi-scale feature fusion, densely connected networks, and Transformer architectures to improve detection performance[41-43]. However, these methods still have certain limitations in handling blurred boundaries and capturing fine-grained features, particularly in feature extraction and information integration.

To address the above issues, this study introduces a LocalWindow attention module into the YOLOv11 model to enhance its ability to detect multi-scale and complex lesions in placental histopathology images. This module draws on the local-window partitioning strategy in the Swin Transformer and, by dividing the feature map into multiple local windows, enables the model to focus more precisely on fine-grained features in local lesion regions[26]. Within each local window, the CGA module combines grouped and multi-head attention mechanisms, effectively improving the model's ability to extract features from lesions of different scales and structures and achieving effective fusion of local details with global information. Meanwhile, the token interaction layer in each local window further strengthens information exchange among features within the window, enabling the model to more sensitively distinguish subtle differences between lesions and normal tissue and to reduce the occurrence of false positives and missed detections. It is noteworthy that the introduction of the LocalWindow module not only optimizes the model's detection performance, but also shows important clinical value in terms of model interpretability. By constraining attention computation within local windows, the module effectively reduces the model's dependence on background noise and compels it to focus on fine-grained features in the lesion regions. This pathology-based decision mechanism is highly consistent with the reasoning logic used by clinicians during manual diagnosis, thereby enhancing the interpretability of the algorithm's decisions.

Empirical results confirm that MONAI augmentation and LocalWindow attention independently improve detection accuracy in the FVM task. When combined, their synergy yields greater gains—enhancing adaptability to diverse lesion morphologies and imaging conditions, and refining key feature recognition. This dual-strategy approach parallels successes in other domains: Guo et al. integrated low-light enhancement and model pruning in IMMP-YOLO, boosting vehicle detection under poor lighting[44]; Chen et al. paired multiple augmentation strategies with an improved YOLOv5 for efficient facade defect detection[45]. These parallels reinforce the broader applicability of coupling advanced augmentation with targeted architectural enhancements for robust detection in complex imaging scenarios.

5.3 Clinical Value of Automated Detection

Our LocalWindow_YOLOv11 model, as an automated detection tool, provides a feasible technical solution for assisting in the diagnosis of FVM. Its main strength lies in its ability to perform automated, high-throughput analysis of placental

histopathology images and to generate objective, reproducible quantitative results. When large numbers of samples need to be processed, such automated detection is expected to improve the efficiency of pathological image analysis and reduce inconsistencies arising from observer fatigue or inter-observer variability, thereby helping to alleviate, to some extent, the shortage of pathology specialists and the strain on medical resources in primary care settings. In addition, by identifying and counting different types of FVM lesions within the images, the model may help to further clarify the potential association between these lesions and adverse pregnancy outcomes, providing evidentiary support for neonatal follow-up and for the management of subsequent pregnancies in affected women.

To evaluate the practical utility of the model, we compared its performance with the diagnostic level of human experts. According to quality-control data from the pathology department of our hospital, the diagnostic accuracies of junior, intermediate, and senior pathologists in placental pathology are approximately 60–79%, 80–90%, and 90% or higher, respectively. Our experimental results showed that the YOLOv11 baseline model achieved Precision, Recall, and mAP50 values of 0.6329, 0.6847, and 0.7117, respectively, which are broadly comparable to the level of junior pathologists. After the synergistic introduction of MONAI-based data augmentation and the LocalWindow module, these metrics increased to 0.6716, 0.7516, and 0.7582, respectively, approaching the diagnostic performance of intermediate pathologists. This suggests that our combined strategy of data augmentation and architectural refinement effectively enhances detection performance. In addition, we designed an independent expert validation procedure to improve the objectivity and reliability of model evaluation. Specifically, we randomly selected representative cases from the test set and submitted the model's detection results to two independent pathologists, who had not been involved in the initial annotations, for double-blind review. Results on which the two reviewers agreed were accepted directly, whereas discrepant cases were adjudicated by a senior pathologist, who provided the final decision.

With regard to deployment feasibility, the LocalWindow_YOLOv11 model shows good practical usability. It requires only 7.77 GFLOPs of computation and 3.03 M parameters, which is far lower than traditional detectors such as Faster R-CNN (Table 6). In our test environment, the model achieved an inference speed of 153 FPS, which can meet the requirement for real-time batch image processing in pathology departments. More importantly, in view of the subjective variability that may occur in

manual diagnosis, the model analyzes images according to a unified algorithmic standard, ensuring that under the same conditions it always produces exactly the same detection results for the same input pathological image, thereby achieving reproducibility of the model in the diagnostic process. This combination of lightweight design and reproducibility gives the model the potential for wide deployment from tertiary hospitals to primary healthcare institutions.

In summary, through the synergistic optimization of data augmentation and model architecture, this study improves detection performance to a level approaching that of intermediate pathologists and demonstrates promising feasibility for deployment and clinical application, providing a valuable technical reference for the intelligent development of FVM diagnosis.

5.4 Conclusion and Future Outlook

This study addresses the challenge of automated FVM detection by proposing an integrated approach combining MONAI-based data augmentation strategies with architectural model improvements. Our findings demonstrate that MONAI augmentation effectively expands both sample quantity and heterogeneity, enhancing model adaptability to complex clinical samples. The LocalWindow attention module further elevates lesion detection precision by strengthening local feature extraction capabilities. Through synergistic integration, our framework achieves state-of-the-art performance across all evaluation metrics. Compared to mainstream detection models, our approach achieves an optimal balance between detection accuracy and computational efficiency while maintaining robust performance.

Although our method significantly improves performance, overall detection effectiveness has yet to reach an ideal level. This limitation may stem from the modest dataset scale, inconsistent image quality, and inter-observer variability in lesion annotation. Additionally, training and evaluation utilized only single-center data, precluding multicenter external validation — thereby limiting assessment of generalizability and clinical applicability.

Future work will focus on 1) optimizing data acquisition and annotation protocols to increase sample diversity; 2) validating generalizability through multicenter datasets; and 3) advancing clinical translation. We anticipate that this data-algorithm co-optimization paradigm will deliver robust technical support for automated FVM diagnosis, accelerating intelligent maternal-fetal health management.

Acknowledgments: We thank the doctors from the Department of Obstetrics and

Gynecology and the Department of Pathology at the Huizhou First Maternal and Child Health Care Hospital for their assistance in clinical information collection and blood testing.

Author contributions: Xuxuan Li and Zhifa Jiang contributed equally and were jointly responsible for the original draft, Resources, Methodology, and Conceptualization. Fengchao Chen: Writing – review & editing, Visualization. Jingwen Liu: Resources, Methodology, Investigation. Jianfeng Peng: Writing – original draft, Formal analysis. Ruoping Lin and Xiangyun Ye: Resources, Investigation, Data curation. Zhen Zhang: Supervision, Methodology, Conceptualization. All authors read and approved the final manuscript.

Funding: This work was supported by [the Guangdong Basic and Applied Research Foundation] grant numbers [2023A1515140146; 2022A1515110138; 2023A1515140184; 2024A1515140193].

Data availability: We have made the complete source code associated with this study open source. It has been uploaded to GitHub and is available at: https://github.com/zzsjbme/FVM_YOLOv11_LocalWindows.

Declaration

Ethics declarations: This study was approved by the Medical Ethics Committee of Huizhou First Maternal and Child Health Care Hospital (Ethics Approval No. 20240328A14), and all participants provided written informed consent. All methods were performed in accordance with the relevant guidelines and regulations.

Consent for publication: Not applicable.

Competing interests: The authors declare no competing interests.

Author details:

School of computer science and engineering, Huizhou University, Huizhou, 516000, Guangdong, China. Xuxuan Li, Fengchao Li, Zhen Zhang.

Obstetrics and Gynaecology, Huizhou First Maternal and Child Health Care Hospital, Huizhou, 516000, Guangdong, China. Zhifa Jiang, Jingwen Liu, Jianfeng Peng, Ruopign Lin, and Xiangyun Ye.

References

1. Heider, A. Fetal vascular malperfusion. *Arch. Pathol. Lab. Med.* **141,** 1484–1489 (2017).
2. Slack, J. C. & Boyd, T. K. Fetal vascular malperfusion due to long and hypercoiled umbilical cords resulting in recurrent second-trimester pregnancy loss: a case series and literature review. *Pediatr. Dev. Pathol.* **24**, 12–18 (2021).
3. Redline, R. W. & Ravishankar, S. Fetal vascular malperfusion, an update. *APMIS* **126,** 561–569 (2018).
4. Turowski, G. et al. The structure and utility of the placental pathology report. *APMIS* **126,** 638–646 (2018).
5. Khong, T. Y. et al. Sampling and definitions of placental lesions: Amsterdam placental workshop group consensus statement. *Arch. Pathol. Lab. Med.* **140,** 698–713 (2016).
6. Peck, M. et al. Review of diagnostic error in anatomical pathology and the role and value of second opinions in error prevention. *J. Clin. Pathol.* **71,** 995–1000 (2018).
7. Chen, W., Luo, J., Zhang, F. & Tian, Z. A review of object detection: datasets, performance evaluation, architecture, applications and current trends. *Multimed. Tools Appl.* **83**, 65603–65661 (2024).
8. Feng, Y. et al. Review of object detection algorithm improvement in deep learning. *J. Comput. Eng. Appl.* **59** (2023).
9. Chen, X. HD-YOLO: real-time detection of benign and malignant endometrial lesions using YOLO11. *Signal Image Video Process.* **19,** 535 (2025).
10. Li, H. et al. An algorithm for cardiac disease detection based on magnetic resonance imaging. *Sci. Rep.* **15,** 4053 (2025).
11. Wu, W. et al. Enhancing placental pathology detection with GAMatrix-YOLOv8 model. *Heliyon* **11** (2025).

12. Vanea, C. et al. Mapping cell-to-tissue graphs across human placenta histology whole slide images using deep learning with HAPPY. *Nat. Commun.* **15,** 2710 (2024).

13. Lee, Y. J., Lee, M. Y., Chung, J. H., Won, H. S., & Park, B. Birth weight prediction using artificial intelligence-based placental assessment from macroscopic photo: a retrospective study. *Placenta* (2025).

14. Andreasen, L. A., Feragen, A., Christensen, A. N., Sperling, L., & Tolsgaard, M. G. Multi-centre deep learning for placenta segmentation in obstetric ultrasound with multi-observer and cross-country generalization. *Sci. Rep.* **13,** 2221 (2023).

15. Fujii, Y. et al. Effectiveness of data-augmentation on deep learning in evaluating rapid on-site cytopathology at endoscopic ultrasound-guided fine needle aspiration. *Sci. Rep.* **14,** 22441 (2024).

16. Hong, J. E., Kim, Y. E., Kang, Y. S., Park, J. Y., & Yu, H. W. SMOTE-augmented machine learning model predicts recurrent and metastatic breast cancer from microbiome analysis. *Sci. Rep.* **15,** 33096 (2025).

17. Marouf, M. et al. Realistic in silico generation and augmentation of single-cell RNA-seq data using generative adversarial networks. *Nat. Commun.* **11,** 166 (2020).

18. Zhang, L., Jindal, B., Alaa, A., Clifton, D. A. & Yang, Y. Generative AI enables medical image segmentation in ultra low-data regimes. *Nat. Commun.* **16,** 6486 (2025).

19. Mobadersany, P., Cooper, L. A. D. & Goldstein, J. A. GestAltNet: aggregation and attention to improve deep learning of gestational age from placental whole-slide images. *Lab. Invest.* **101,** 942–951 (2021).

20. Parathanath, A. J. & Anu, M. CBMNet: a dual-attention enhanced ConvNeXt model for accurate GV Black type I–III classification in intraoral periapical radiographs. *Sci. Rep.* **15,** 39287 (2025).

21. Abrar, M., Salam, A., Ullah, F., Alraddady, F. & Al-Otaibi, A. N. Enhancing brain tumor segmentation using attention based convolutional UNet on MRI images. *Sci. Rep.* **15,** 36603 (2025).

22. Pinaya, W. H. L. et al. Generative AI for medical imaging: extending the MONAI framework. Preprint at https://arxiv.org/abs/2307.15208 (2023).

23. Cardoso, M. J. et al. Monai: an open-source framework for deep learning in healthcare. Preprint at https://arxiv.org/abs/2211.02701 (2022).

24. Khanam, R. & Hussain, M. YOLOv11: an overview of the key architectural enhancements. Preprint at https://arxiv.org/abs/2410.17725 (2024).

25. Liu, X. et al. EfficientViT: memory efficient vision transformer with cascaded group attention. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* 14420–14430 (2023).

26. Liu, Z. et al. Swin transformer: hierarchical vision transformer using shifted windows. In *Proc. IEEE/CVF Int. Conf. Comput. Vis.* 10012–10022 (2021).

27. Bolya, D., Foley, S., Hays, J. & Hoffman, J. Tide: a general toolbox for identifying object detection errors. *In Eur. Conf. Comput. Vis.* 558–573 (Springer, Cham, 2020).

28. Chen, P. Effects of the entropy weight on TOPSIS. *Expert Syst. Appl.* **168,** 114186 (2021).

29. Zhou, X. & Lu, M. Risk evaluation of dynamic alliance based on fuzzy analytic network process and fuzzy TOPSIS. (2012).

30. Woolson, R. F. Wilcoxon signed-rank test. In *Wiley Encyclopedia of Clinical Trials* 1–3 (Wiley, 2007).

31. Guo, M. H. et al. SegNeXt: Rethinking convolutional attention design for semantic segmentation. *Adv. Neural Inf. Process. Syst.* **35,** 1140–1156 (2022).

32. Misra, D., Nalamada, T., Arasanipalai, A. U. & Hou, Q. Rotate to attend: convolutional triplet attention module. *In Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.* 3139–3148 (2021).

33. Huang, H. et al. Channel prior convolutional attention for medical image segmentation. *Comput. Biol. Med.* **178,** 108784 (2024).

34. Hou, Q., Zhou, D. & Feng, J. Coordinate attention for efficient mobile network design. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* 13713–13722 (2021).

35. Liu, Y., Shao, Z. & Hoffmann, N. Global attention mechanism: retain information to enhance channel-spatial interactions. Preprint at https://arxiv.org/abs/2112.05561 (2021).

36. Abdollahi, B., Tomita, N. & Hassanpour, S. Data augmentation in training deep learning models for medical image analysis. In *Deep learners and deep learner descriptors for medical applications* 167–180 (Springer, Cham, 2020).

37. Nalepa, J., Marcinkiewicz, M. & Kawulok, M. Data augmentation for brain-tumor segmentation: a review. *Front. Comput. Neurosci.* **13,** 83 (2019).

38. Shorten, C. & Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **6,** 1–48 (2019).

39. Frid-Adar, M. et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **321,** 321–331 (2018).

40. Saldanha, J. et al. Data augmentation using variational autoencoders for improvement of respiratory disease classification. *PLoS One* **17,** e0266467 (2022).

41. Hussain, I., Tan, S. & Huang, J. Few-shot based learning recaptured image detection with multi-scale feature fusion and attention. *Pattern Recognit.* **161,** 111248 (2025).

42. Zhou, T. et al. Dense convolutional network and its application in medical image analysis. *Biomed*. *Res. Int.* **2022,** 2384830 (2022).

43. Shou, Y. et al. Object detection in medical images based on hierarchical transformer and mask mechanism. *Comput. Intell. Neurosci.* **2022,** 5863782 (2022).

44. Guo, L. & Zhou, X. IMMP-YOLO: a real-time object detection algorithm incorporating image enhancement and pruning strategy. In *2024 8th CAA Int. Conf. Vehicular Control and Intelligence (CVCI)* 1–6 (IEEE, 2024).

45. Chen, Y. & Li, D. Disease detection on exterior surfaces of buildings using deep learning in China. *Sci. Rep.* **15,** 8564 (2025).