Sensors and Materials, Vol. 38, No. 8 (2026) 4673–4696
MYU Tokyo

S & M 4588

Multi-source Medical Sensor Data Fusion Using Transformer for Preoperative Assessment of Placenta Accreta Spectrum

Hua Liu,[1] Weihan Ge,[1*] and Mingyu Zhao[2]

[1]Second Affiliated Hospital of Mudanjiang Medical University, Mudanjiang 157009, China
[2]Heilongjiang Academy of Sciences, Harbin 150001, China

(Received May 11, 2026; accepted August 12, 2026)

Keywords: placenta accreta spectrum, medical sensors, sensing system, deep learning, cross-center generalization, multi-source sensor fusion

The prenatal diagnosis of placenta accreta spectrum (PAS) is essential to prevent life-threatening maternal hemorrhage. Recent deep learning models using single-modality ultrasound or magnetic resonance imaging (MRI) have achieved high accuracy, but each modality has inherent limitations: Ultrasound suffers from acoustic shadowing, MRI is prone to motion artifacts, and current methods lack mechanisms to handle missing modalities in routine workflows. To overcome these limitations, we developed a biomimetic medical sensor data fusion model that addresses these limitations through three innovations: (1) a structural prior step using morphological dilation to isolate the placental–myometrial interface, (2) a bidirectional cross-attention module that dynamically aligns acoustic and electromagnetic features, and (3) a missing-modality robust training strategy. On an external validation cohort (N = 60), the model achieved an area under the curve (*AUC*) of 0.889 [95% confidence interval (*CI*): 0.824–0.941], outperforming single-modality models (ultrasound's *AUC* = 0.812; MRI's *AUC* = 0.843) and previous late-fusion networks (*AUC* = 0.857).

1. Introduction

Placenta accreta spectrum (PAS) is one of the most catastrophic conditions in high-risk obstetrics, defined by a clinicopathological continuum of abnormalities at the placental–decidual interface.[1,2] Pathologically, PAS involves the abnormal invasion of placental villi into the uterine myometrium, sometimes extending beyond the serosal layer to adjacent organs. PAS is associated with life-threatening hemorrhage,[3] massive transfusion requirements,[4] and high maternal morbidity.[5] These risks underscore the importance of early prenatal identification and planned management at specialized centers.

PAS diagnosis is not merely an imaging challenge but a multi-sensor problem. Ultrasound remains the primary modality for screening,[6] but its diagnostic accuracy is highly operator-dependent and variable.[7,8] Magnetic resonance imaging (MRI) provides complementary electromagnetic sensing data, particularly valuable for posterior placentation

*Corresponding author: e-mail: 15245329000@163.com
https://doi.org/10.18494/SAM6423

ISSN 0914-4935

and deep pelvic invasion.[9] Ultrasound and MRI need to be understood not as isolated imaging tools but as components of a broader medical sensor system.[10] However, a significant limitation in current practice is the signal gap between acoustic data captured by ultrasound transducers and electromagnetic signals acquired by MRI coils. To address this limitation, advanced sensor fusion algorithms capable of integrating heterogeneous data streams are required.[11]

Existing approaches often oversimplify PAS diagnosis by treating it as a conventional image classification task, overlooking the complexity of multi-source sensor data. MRI is inappropriate as a universal screening tool because multimodal datasets are frequently incomplete, and missing modalities where one type of scan is unavailable are common in clinical practice, introducing substantial selection bias. Moreover, many models lack external validation, which leads to performance degradation when deployed across different hardware systems or sensing environments.[12] Deep neural networks widely applied to medical imaging also tend to exhibit overconfidence, where a high area under the curve (*AUC*) does not necessarily translate into trustworthy risk probabilities.[13] For clinical decision-making, however, the reliability of sensor outputs is paramount, as predicted probabilities must justify initiating high-risk management pathways.[14,15]

The prenatal assessment of PAS relies mainly on ultrasound and MRI. However, existing deep learning models for these modalities do not fully address their limitations. Ultrasound is highly operator-dependent, has limited tissue penetration in posterior placentas, and is often compromised by acoustic shadowing from maternal tissue or pelvic bone. MRI, in contrast, is susceptible to fetal motion artifacts, offers lower temporal resolution, varies in performance across field strengths [1.5 and 3.0 Tesla (T)], and is costly to implement. Previous multimodal approaches combine ultrasound and MRI through static feature concatenation or decision-level voting ensembles. However, these strategies do not account for spatial misalignment between two-dimensional ultrasound planes and 3D MRI volumes, and they cannot function effectively when one modality is missing during emergency presentations.

To overcome these limitations, we employed a biomimetic cross-attention Transformer. This method enables dynamic weight adjustment based on high-resolution spatial details from MRI against acoustic flow indicators from ultrasound within a morphologically constrained interface region of interest (ROI). This design enables the reliable fusion of complementary signals and ensures robust performance even when one modality is unavailable.

In this article, PAS diagnosis is defined as a risk probability problem in multi-source medical sensing. Accordingly, we developed a medical sensor data fusion model incorporating five optimization strategies: (1) structural prior optimization, (2) unimodal representation optimization, (3) cross-modal alignment optimization, (4) missing-modality robustness optimization, and (5) trustworthy calibration optimization.[16] These strategies enable this model to support sensor technology applications in obstetrics, providing a more robust and reliable tool for the intelligent analysis of multi-source medical data. We evaluated whether sensor fusion models outperform unimodal approaches, whether calibrated sensing systems deliver greater clinical utility,[17] and whether structural priors and missing-modality modeling mitigate degradation in cross-center generalization.

2. Theoretical Basis

The most important diagnostic information on PAS is specific anatomical interfaces, including the placenta, uterine wall boundary, scar tissue regions, and uterine–bladder junction. Ultrasound markers (a loss of the clear zone, placental lacunae, bridging vessels, and myometrial thinning) and MRI markers (T2-weighted imaging in hypointense bands, uterine bulging, and bladder wall involvement) are used to detect abnormalities at these interfaces.[6,9] For data fusion, this means that ROI is not the whole image but the interface zone where signals converge. By focusing on the ROI, we can reduce background noise and device-specific variability, improving robustness across different hospitals and imaging systems.[8,9]

PAS diagnosis must be conducted on the basis of multiple sensor inputs rather than a single image. In practice, Doppler ultrasound serves as the primary medical sensor, capturing both acoustic and flow signals, whereas MRI functions as a complementary sensor that provides electromagnetic structural data. Physiological sensors such as photoplethysmograms or electrocardiograms supply vital signs, and contextual metadata, including gestational age, placental location, and prior cesarean history, provide essential information. PAS risk assessment depends more on the combined sensor data than on local imaging abnormalities. Clinical guidelines emphasize early risk stratification to prepare for hemorrhage and surgical complexity.[1,3] From a technological perspective, the multi-source data fusion method used in this study mirrors the functionality of biomimetic sensing arrays, wherein heterogeneous transducers, specifically acoustic and electromagnetic, are integrated to capture and synchronize the complex physical and biological signals characteristic of the placental–decidual interface.[18,19]

For the effective integration of multi-source sensor data, the choice of fusion method is critical. Early fusion, which simply concatenates raw inputs, is highly sensitive to device variability and acquisition differences. Late fusion, which combines decisions after independent processing, fails to capture the structural relationships between modalities. Intermediate fusion, implemented through cross-modal attention at ROI, better presents the mutually reinforcing signals between ultrasound and MRI, yielding more robust representations.[11]

A persistent challenge in clinical datasets is the presence of missing modalities. MRI is not universally performed, meaning that data fusion using MRI images is often incomplete. These missing inputs reflect the constraints of clinical pathways, a structured plan or roadmap for how patients with a specific condition should be managed in healthcare. Consequently, models must be explicitly designed to handle incomplete sensor data, ensuring resilience and reliability in practical deployment.[16] Another challenge is the trustworthiness of predictions. Deep neural networks often produce overconfident outputs, where high *AUC* scores do not necessarily correspond to reliable risk probabilities.[13] In this case, calibration methods can improve probability quality.[14] Decision curve analysis is also conducted to assess predicted risk to clinical action, ensuring that medical sensor data are meaningful for initiating high-risk management pathways.[17]

To conduct PAS diagnosis as a multi-sensor fusion task, structured ROI analysis, robust fusion strategies, and trustworthy calibration are required. The model developed in this study

integrates structural priors, unimodal optimization, cross-modal alignment, missing-modality robustness, and calibrated prediction.[16] The model advances the application of sensor technology in obstetrics, enabling a reliable tool for intelligent risk assessment in PAS.

3. Methods

The developed model in this study enables the preoperative risk assessment of PAS. According to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis + AI (TRIPOD + AI)[20] and Checklist for AI in Medical Imaging (CLAIM) 2024 guidelines,[21] the model was designed to process and fuse heterogeneous data from multi-source medical sensors. The model integrates four types of sensor data: acoustic signals from ultrasound, complementary electromagnetic signals from MRI, low-frequency physiological signals, and metadata.

3.1 Data optimization

The sensing data corpus is constructed on a per-patient basis, where $i = 1,\ldots, N$ represents the patient index. The multivariate sensor input vector X_i is defined as

$$x_i = \{x_i^{US}, x_i^{MRI}, x_i^{PHY}, x_i^{META}\}. \tag{1}$$

Here, x_i^{US} denotes the ultrasound data, x_i^{MRI} the MRI data, x_i^{PHY} the physiological data, and x_i^{META} the metadata. To account for sensor unavailability, a modal availability mask m_i is defined as

$$m_i = \left(m_i^{US}, m_i^{MRI}, m_i^{PHY}, m_i^{META}\right) \in \{0,1\}. \tag{2}$$

Here, 1 indicates available data and 0 indicates missing data. The joint optimization objective for the sensing system is defined as

$$L_{total} = \lambda_{prior} L_{prior} + \lambda_{uni} L_{uni} + \lambda_{fusion} L_{fusion} + \lambda_{robust} L_{robust}. \tag{3}$$

Here, L_{prior} isolates high-value sensing regions through structural spatial filtering, L_{uni} supervises independent unimodal feature extraction, L_{fusion} guides cross-modal signal alignment and multi-task prediction, and L_{robust} maintains stability during missing modality failures. We set the weighting coefficients $\lambda_{fusion} = 1.0$, $\lambda_{prior} = 0.5$, $\lambda_{uni} = 0.5$, and $\lambda_{robust} = 0.5$ to balance multi-task feature extraction and prevent primary task gradient starvation.[22]

3.2 Sensor data filtering and interface ROI

To enhance the signal-to-noise ratio (*SNR*) in high-value anatomical regions, we used a structural filter by constructing an ROI. The model employs two segmentation processors, f_{seg}^{US} and f_{seg}^{MRI}, to identify the placenta (M_p) and myometrium (M_m) markers.

$$\left\{M_p^{US}, M_m^{US}\right\} = f_{seg}^{US}\left(x_i^{US}\right); \left\{M_p^{MRI}, M_m^{MRI}\right\} = f_{seg}^{MRI}\left(x_i^{MRI}\right) \tag{4}$$

Here, x_i^{US} denotes the input ultrasound image (raw acoustic sensor data), f_{seg}^{US} and f_{seg}^{MRI} denote the segmentation function applied to ultrasound and MRI, respectively, which detects and separates relevant regions, $\{M_p^{US}, M_m^{US}\}$ denotes the output markers from ultrasound segmentation, $\{M_p^{MRI}, M_m^{MRI}\}$ denotes the output markers from MRI segmentation, again distinguishing placental and myometrial features, and x_i^{MRI} denotes the input MRI image (electromagnetic sensor data).

The interface ROI, which represents the critical sensing zone for invasion detection, is defined using morphological dilation.

$$ROI = \left(M_p \oplus \delta\right) \cap \left(M_m \oplus \delta\right) \tag{5}$$

Here, $\oplus\delta$ represents the morphological dilation operation with the structuring element δ that expands the boundaries of each marker region slightly to account for uncertainty or variability in segmentation and $\cap$ is the intersection operator to find the overlapping area between the dilated placental region and the dilated myometrial region. ROI is defined as the overlap zone between the dilated placental and myometrial marker regions.

We evaluated segmentation accuracy for f_{seg}^{US} and f_{seg}^{MRI} against ground-truth manual annotations outlined by senior radiologists using the Dice similarity coefficient (*DSC*) and intersection over union (IoU). *DSC* is a spatial overlap index that is used to measure the similarity between the model's predicted segmentation and the ground-truth region.[23] Values range from 0 (no overlap) to 1 (perfect alignment). IoU is calculated to measure the overlap of two boundaries by dividing their intersection by their union.[24] It penalizes false positives (*FP*s) and false negatives (*FN*s) more severely than *DSC*.

$$DSC = \frac{2\left|M_{pred} \cap M_{gt}\right|}{\left|M_{pred}\right| + \left|M_{gt}\right|} \tag{6}$$

$$IoU = \frac{\left|M_{pred} \cap M_{gt}\right|}{\left|M_{pred} \cup M_{gt}\right|} \tag{7}$$

Here, M_{pred} denotes the predicted binary mask segmented automatically by the model (f_{seg}^{US} or f_{seg}^{MRI}). M_{gt} is the ground-truth binary mask manually annotated and verified by senior radiologists. $| M_{pred} \cap M_{gt} |$ represents the intersection, that is, the spatial overlap where both the predicted and ground-truth masks agree [true positive (*TP*) region]. $| M_{pred} | + | M_{gt} |$ is the sum of cardinalities, the total number of pixels or voxels contained in both masks. $| M_{pred} \cup M_{gt} |$ denotes the union, which is the combined spatial area covered by either mask or both. Table 1 shows the segmentation performance across modalities. The segmentation processor achieved robust boundary delineation, which, when combined with morphological dilation (δ= 5 mm), ensured target ROI coverages of 93.1% in ultrasound and 95.2% in MRI.

3.3 Unimodal representation and learning

Before data fusion, the independent risk features are extracted to reduce sensitivity to unstable sensor data. The ROI-constrained encoders for ultrasound (E_{US}) and MRI (E_{MRI}) process the filtered signals. The following equation describes the feature extraction step in the model:

$$z_i^{US} = E_{US}\left(x_i^{US} \odot ROI^{US}\right); z_i^{MRI} = E_{MRI}\left(x_i^{MRI} \odot ROI^{MRI}\right). \tag{8}$$

Here, z_i^{US} is the ultrasound feature vector. ROI^{US} and ROI^{MRI} are the ROIs derived from ultrasound and MRI, respectively. $\odot$ is the elementwise masking operation, meaning the raw image is restricted to the ROI. $E_{US}(\cdot)$ and $E_{MRI}(\cdot)$ are respectively the ultrasound encoder, which extracts feature representations from the ROI-masked input, and the MRI encoder, which extracts feature representations from the ROI-masked input.

The physiological sensor data x_i^{phy}, such as heart rate and blood pressure, are processed by a two-layer fully connected network E_{phy}.

$$z_i^{phy} = E_{phy}\left(x_i^{phy}\right) \tag{9}$$

Table 1
ROI segmentation accuracy across modalities.

Modality	Target	*DSC*	IoU	Interface ROI coverage
Ultrasound	Placenta	0.871	0.772	0.931
	Myometrium	0.853	0.744	
	Interface ROI	0.862	0.758	
MRI	Placenta	0.895	0.81	0.952
	Myometrium	0.873	0.775	
	Interface ROI	0.884	0.793	

Here, z_i^{phy} is the resulting feature vector that captures the essential physiological information for fusion with imaging and contextual data, x_i^{phy} represents the raw physiological input data from the photoplethysmogram and electrocardiogram or other vital signal sensors, and $E_{phy}(\cdot)$ is the physiological encoder that transforms raw physiological signals into a structured representation (features). Metadata, such as 1.5 or 3.0 Tesla MRI field strength, is transformed on the embedding layer to generate z_i^{META}. Unimodal auxiliary heads provide supervision (L_{uni}) using the task label to ensure each sensor path develops independent discriminative power.

3.4 Cross-modal data fusion

To fill the signal gap between acoustic (ultrasound) and electromagnetic (MRI) data,[18,19] we employed a bidirectional cross-attention mechanism. $\hat{T}_{US}$ and T_{MRI} are projected into a shared latent space using the final equation

$$\hat{T}_{US} = T_{US} W_{US};\ \hat{T}_{MRI} = T_{MRI} W_{MRI}. \tag{10}$$

Here, $\hat{T}_{US}$ is the transformed ultrasound feature representation, ready for fusion, T_{US} is the feature matrix extracted from ultrasound (acoustic sensor data), W_{US} is the learned weight matrix specific to ultrasound, acting as a transformation filter to adjust the feature space to emphasize the most relevant signals, $\hat{T}_{MRI}$ is the transformed MRI feature representation, T_{MRI} is the feature matrix extracted from MRI (electromagnetic sensor data), and W_{MRI} is the learned weight matrix specific to MRI. The fusion engine calculates the interaction between modalities using query (Q), key (K), and value (V) matrices.

$$Attention(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{11}$$

Here, d_k is the scaling factor used to stabilize training. The cross-modal fusion backbone is constructed using four Transformer layers, where each layer contains eight parallel attention heads, a model embedding dimension (d_{model}) of 256, and a feed-forward layer dimension (d_{ff}) of 1024 with a dropout rate of 0.1. Bidirectional fusion enables MRI data to refine ultrasound features and vice versa, yielding a synchronized cross-modal representation. The final patient-level fusion vector Z_{fuse} combines all sensor data streams as follows.[25]

$$Z_{fuse} = \left[z_i^{img}, z_i^{phy}, z_i^{meta} \right] \tag{12}$$

Here, z_i^{img} represents the imaging fusion vector (ultrasound and MRI), z_i^{phy} represents the physiological sensor features (photoplethysmogram and electrocardiogram), and z_i^{meta} represents the metadata features (gestational age, placental location, and prior cesarean history). These form the complete patient-level fusion vector, integrating all sensor streams into one representation.

3.5 Missing modality learning

To ensure the sensing system remains operational if a sensor fails, we introduce the robustness loss L_{robust}. In training, Bernoulli random variables $b \in \{0, 1\}$ are used to simulate sensor dropout.

$$\tilde{z}_i^{MRI} = b_i^{MRI} \cdot z_i^{MRI} \tag{13}$$

The model is trained to minimize the Kullback–Leibler Divergence between the full-modal output and the missing-modal output. This allows the model to maintain consistent risk assessment even with reduced sensing inputs.[16]

3.6 Post-processing probability calibration

For a sensing system to be trustworthy in clinical decision-making, its output must be calibrated.[18] We apply temperature scaling as a post-processing step on a held-out validation set. The calibrated probability is derived from the raw logit using a scalar parameter and the Sigmoid function. This optimization does not change the model's ranking (*AUC*) but aligns the predicted probabilities with actual observed frequencies, essential for threshold-based surgical planning.[17] This workflow of the model transitions from anatomical understanding to risk sensing at the following stages.

- Stage 1: Pre-training spatial filters
- Stage 2: Independent sensor feature extraction
- Stage 3: Multi-source fusion and robustness training
- Stage 4: System reliability calibration

This phased approach ensures the model evolves from raw signal processing to a reliable, intelligent sensing system for PAS assessment.[16,26]

3.7 Data and experiment

3.7.1 Data collection

We constructed the experimental dataset using a multicenter high-risk obstetric cohort. A total of 642 patients were included, comprising 236 PAS-positive patients (36.8%), 406 non-PAS patients (63.2%), and 165 patients with severe hemorrhage or high-risk surgical outcomes (25.7%). The multisource sensor data include the following.

- Ultrasound data: available for all 642 patients
- MRI data: available for 351 patients (54.7%)
- Physiological signals: available for 565 patients (88.0%)
- Complete multimodal samples: 298 patients (46.4%) with ultrasound, MRI, and physiological data, and metadata

We recruited high-risk pregnant patients with suspected PAS from three tertiary academic medical centers in China, specializing in high-risk obstetrics and maternal–fetal medicine.

- Primary institution (Hospital A: internal validation set): 240 patients recruited between January 2018 and December 2023, randomly partitioned into training and internal test/validation ($N = 60$) subsets.
- Partner Center 1 (Hospital B: external validation set): 35 patients recruited between March 2020 and November 2023.
- Partner Center 2 (Hospital C: external validation set): 25 patients recruited between June 2020 and October 2023.

The data from the validation patient group were kept entirely separate from model derivation, parameter tuning, and probability calibration. All patients met the following inclusion criteria: (1) with the documented diagnosis or suspicion of PAS based on prior cesarean delivery and coexisting placenta previa, (2) with available preoperative paired ultrasound B-mode images and T2-weighted MRI sequences, and (3) with definitive intraoperative or histopathological staging confirmed at delivery according to the International Federation of Gynecology and Obstetrics guidelines. Data collection and anonymization protocols were approved by the Institutional Review Boards (IRBs) of all participating institutions (Primary IRB Approval No. 2023OB084). Deidentified imaging files and clinical records were transferred under formal interinstitutional data transfer agreements.

3.7.2 Experiment

Although MRI is not a universal screening tool in PAS clinical pathway determination, its data are reserved for suspected or complex cases. Consequently, the dataset exhibits missing modalities and selection bias, reflecting real-world conditions. This characteristic provided the rationale for designing the missing-modality robust learning strategy in this study. For model development, the data were partitioned at the patient level to prevent slice-level or frame-level information leakage. The training and validation datasets contained the data of 420 and 102 patients, respectively. The internal and external datasets each included the data of 60 patients. To evaluate model performance, we compared the developed medical sensor data fusion model with non-Transformer models. For performance comparison, we used the following non-Transformer models. All non-Transformer baselines were trained on identical data splits and spatial inputs.

- Convolutional neural network (CNN)– long short-term memory (LSTM) model: CNN feature extractors paired with an LSTM network to process concatenated multi-sensor feature sequences[27]
- Graph neural network (GNN) model: a graph attention network (GAT) constructing relational graph nodes across ultrasound, MRI, and physiological signal embeddings[28]
- Multimodal variational autoencoder (MVAE): a generative multimodal fusion model using a Product-of-Experts joint latent space inference network[29]

The experiment was conducted using PyTorch and executed on an NVIDIA RTX 3090 graphics processing unit. For model training, the Adam optimizer was used with an initial learning rate of 1×10^{-4}, a batch size of 8, and 120 epochs. All models were executed on a

workstation equipped with an NVIDIA RTX 3090 graphics processing unit (GPU) [24 gigabytes (GB) of virtual random access memory (VRAM)] and an Intel Core i9-12900K CPU (64 GB RAM). Training the complete multi-source Transformer model across 120 epochs with a batch size of 8 took approximately 3.4 h. Parameter complexity and floating-point operations (FLOPs) were benchmarked using the PyTorch FlopCountAnalysis tool. Direct monetary training and deployment costs were omitted because of high variability in regional electricity tariffs and cloud platform pricing structures; hardware-agnostic metrics (FLOPs, latency, and parameter counts) are reported instead to allow standardized evaluation across deployment environments.

Data preprocessing and augmentation were applied to improve model robustness and generalization. Image intensity normalization was performed to standardize pixel values across different devices. Random rotations within ±10°, random flipping, and Gaussian noise perturbation (applied to MRI data) were introduced to simulate natural variability and enhance resilience against cross-device and cross-center differences. These steps ensured that the model could adapt to heterogeneous imaging conditions encountered in real clinical practice.

To simulate scenarios where not all modalities are consistently available, a random modality dropout strategy was incorporated during training. In each batch, the MRI modality was randomly omitted with a probability of 0.3, whereas the physiological signal modality was randomly omitted with a probability of 0.2. This approach enforced consistency constraints and strengthened the model's ability to handle incomplete sensor data, thereby improving robustness in practical deployment.

To evaluate the effectiveness of the developed model, a comprehensive set of comparison models was constructed. Two baseline models were included: the treat-all model, which classifies every patient as PAS-positive and represents a universal intervention approach, and the treat-none model, which classifies every patient as PAS-negative and represents a no-intervention approach. These baselines serve as reference points in decision curve analysis (DCA), illustrating the extremes of overtreatment and undertreatment. Other than these baseline models, single-modal models were implemented, including an ultrasound-only model and an MRI-only model, representing unimodal clinical approaches. Traditional fusion strategies were tested through early fusion, which concatenates inputs at the raw data level, and late fusion, which combines decisions at the output level. To isolate the contributions of specific architectural components, ablation models were designed: a no-ROI model without ROI constraints, a no-cross-modal attention model without semantic alignment, and a no-robustness mechanism model without missing-modality training. Finally, the model developed has all key design elements integrated—ROI constraints, cross-attention mechanisms, robustness strategies, and calibration procedures—representing the full architecture proposed in this study. All models were trained and evaluated using identical data partitions and training parameters to ensure fair comparison. The details of model settings are summarized in Table 2.

Model performance was evaluated for discriminative ability, generalization capacity, calibration quality, and clinical utility. Discriminative ability was evaluated using *AUC*, sensitivity, specificity, and F1-score. The model's capacity was assessed to distinguish PAS-positive from non-PAS patients. Generalization capacity was quantified by the generalization gap, defined as the performance difference between internal and external test sets, to assess stability across different clinical centers. Calibration quality was evaluated using

Table 2
Comparison of model settings.

Model	Input data	Method
Treat-all	No	Classifying every patient as PAS-positive (universal intervention baseline)
Treat-none	No	Classifying every patient as PAS-negative (no intervention baseline)
Ultrasound-only	Ultrasound	Ultrasound unimodal modeling
MRI-only	MRI	MRI unimodal modeling
Early fusion	All	Input-level concatenation fusion
Late fusion	All	Decision-level weighted fusion
No-ROI	All	No ROI constraints (whole image)
No-cross-modal attention	All	No cross-modal attention
No-robustness mechanism	All	No missing modality training
Model developed in this study	All	ROI + cross-modal attention + robust + calibration

the Brier score (*BS*) and the expected calibration error (*ECE*), which reflect the alignment between predicted probabilities and actual event incidence rates. Clinical utility was analyzed through net benefit (NB) and DCA, providing insight into whether the model offers genuine decision-making value under varying risk thresholds. Through multi-dimensional evaluation under identical conditions to guarantee consistency, the models were compared in terms of discriminativeness, generalizability, calibrating ability, and clinical applicability. Evaluation indicators are presented in Table 3.

3.8 Evaluation metrics

To evaluate diagnostic performance and probability reliability, we assessed predictions using both classification and calibration metrics. Classification metrics included accuracy, *FP* rate, and *FN* rate. Accuracy represents the proportion of all patients that were correctly classified, calculated as the sum of *TP*s and true negatives (*TN*) divided by the total number of cases ($TP + TN + FP + FN$).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

The *FP* rate is used to measure the proportion of non-PAS patients incorrectly classified as PAS. It is calculated as the number of *FP*s divided by the total number of actual negatives ($TN + FP$). This is equivalent to one minus specificity.

$$FP\,rate = \frac{FP}{TN + FP} = 1 - Specificity \tag{15}$$

The *FN* rate is used to quantify the proportion of PAS patients incorrectly classified as non-PAS. It is calculated as the number of *FN*s divided by the total number of actual positives ($TP + FN$). This is equivalent to one minus sensitivity.

Table 3
Evaluation indicators used in model evaluation and comparison.

Dimension	Metric	Symbol	Description
Discrimination	*AUC*	*AUC*	Discriminative ability
	Sensitivity	Sen	Positive identification ability
	Specificity	Spe	Negative identification ability
	F1-score	F1	Comprehensive performance
Generalization	Gap	ΔAUC	Cross-center performance drop
Calibration	*BS*	*BS*	Probability error
	Calibration error	ECE	Probability consistency
Clinical applicability	*NB*	*NB*	Clinical benefit
	DCA	—	Decision curve

$$FN\ rate = \frac{FN}{TP + FN} = 1 - Sensitivity \tag{16}$$

All decision metrics were calculated at a default risk decision threshold of t = 0.50. Probability calibration metrics were also used in this study. *BS* is used to measure the mean squared difference between predicted risk probabilities (p_i) and actual binary outcomes (y_i) across all N patients. Lower values indicate more accurate probability estimates.[30]

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2 \tag{17}$$

ECE is used to evaluate how well predicted probabilities align with observed outcomes.[31] Predictions are grouped into ten equal-width probability bins (B_m), and for each bin, the difference between empirical accuracy [acc(B_m)] and average predicted confidence [conf(B_m)] is weighted by the bin size. The sum across all bins provides the overall calibration error.

$$ECE = \sum_{m=1}^{M}\frac{|B_m|}{N}\left|\mathrm{acc}\left(B_m\right) - \mathrm{conf}\left(B_m\right)\right| \tag{18}$$

These metrics were used to assess the discriminative ability of the model (classification accuracy, sensitivity, and specificity) and the reliability of its probability estimates (*BS* and *ECE*), ensuring the comprehensive evaluation of diagnostic performance in clinical research.

4. Results

4.1 Performance comparison

The developed model was evaluated for the reliability and robustness of multi-source sensor data fusion. The model results were analyzed for classification outputs in heterogeneous signal

environments. To evaluate statistical reliability, we computed 95% confidence interval (*CI*) using 1000 bootstrap resamples. The developed model showed superior discriminative ability with an internal *AUC* of 0.928 (95% *CI*: 0.884–0.963) and an external *AUC* of 0.889 (95% *CI*: 0.824–0.941) (Table 4). The developed model demonstrated superior discriminative ability with an internal *AUC* of 0.928 and an external *AUC* of 0.889. This indicates that the model distinguished between patients with a 92.8% accuracy on its data and an 89% accuracy on completely new, unfamiliar data. The generalization gap (Δ*AUC*) is critical for medical sensor deployment, since it shows the difference between a model's performance on the internal development dataset and its performance on an independent, external test set. In the developed model, the gap was minimized to 0.039, the lowest among the compared models. This stability suggests that the model effectively mitigated the hardware variability found in external data, mitigating differences in ultrasound transducer sensitivity and MRI radiofrequency coil. Whereas unimodal models suffer from significant performance degradation across centers,[12] the multi-source approach of the developed model maintains a consistent *SNR* regardless of the sensing environment.

The model developed in this study outperformed all non-Transformer models (Table 5). The CNN–LSTM model showed an external *AUC* of 0.848 (Δ*AUC* = 0.050). This model's sequential processing limits the ability to capture non-temporal, spatially heterogeneous interactions

Table 4
Performance comparison of different models for preoperative PAS risk assessment with 95% *CI*s.

Model	*AUC* (internal, *CI*)	*AUC* (external, *CI*)	Δ*AUC*	Decrease in *BS*	Decrease in *ECE*
Ultrasound-only	0.871 [0.815–0.918]	0.812 [0.732–0.879]	0.059	0.168	0.129
MRI-only	0.889 [0.834–0.932]	0.843 [0.768–0.905]	0.046	0.154	0.121
Early fusion	0.903 [0.852–0.946]	0.851 [0.781–0.912]	0.052	0.149	0.118
Late fusion	0.909 [0.860–0.951]	0.857 [0.789–0.916]	0.052	0.147	0.112
No-ROI	0.898 [0.845–0.942]	0.848 [0.775–0.908]	0.05	0.144	0.098
No-cross-modal attention	0.912 [0.863–0.954]	0.865 [0.796–0.922]	0.047	0.142	0.089
No-robustness mechanism	0.895 [0.841–0.939]	0.841 [0.766–0.902]	0.054	0.15	0.102
Developed model in this study	0.928 [0.884–0.963]	0.889 [0.824–0.941]	0.039	0.141	0.114

Table 5
Performance comparison of different models for preoperative PAS assessment.

Model	*AUC* (internal)	*AUC* (external)	Δ*AUC*	Decrease in *BS*	Decrease in *ECE*
Ultrasound-only	0.871	0.812	0.059	0.168	0.129
MRI-only	0.889	0.843	0.046	0.154	0.121
Early fusion	0.903	0.851	0.052	0.149	0.118
Late fusion	0.909	0.857	0.052	0.147	0.112
CNN–LSTM fusion	0.898	0.848	0.05	0.144	0.098
GNN fusion	0.912	0.865	0.047	0.142	0.089
MVAE	0.895	0.841	0.054	0.15	0.102
Model developed in this study	0.928	0.889	0.039	0.141	0.114

between ultrasound and MRI signals. MVAE reached an external *AUC* of 0.841 ($\Delta AUC = 0.054$) because its evidence lower bound emphasizes input reconstruction over fine-grained diagnostic boundary discrimination. The GNN model performed relatively well (external *AUC* = 0.865), showing the benefit of relational modeling across modalities, but its fixed node-edge structures cannot adaptively reweight cross-modal attention when local ROI signal quality varies across hardware providers. In contrast, the Transformer cross-attention computes pairwise matrix products (QK^T) between ultrasound and MRI feature tokens, enabling dynamic bi-directional modulation. High-resolution MRI features refined ambiguous acoustic boundaries at the placental–myometrial interface, improving diagnostic precision.

Table 6 shows the parameter complexity, computational footprint (FLOPs), training execution time, and per-patient inference latency across baseline and proposed models. The model developed in this study maintained a compact size (14.2 million parameters) and computational load (18.6 giga FLOPs). The per-patient inference latency averaged 42 ms, encompassing ROI spatial extraction, cross-attention fusion, and post-processing calibration. This minimal computational overhead enables seamless integration into standard workstation GPUs or hospital PACS workstations without specialized computing infrastructure.

4.2 Ablation study results

To verify the contribution of model configurations, we conducted a structural ablation study. Each configuration, such as ROI constraint, cross-modal attention, robustness mechanism, or calibration, was systematically removed to assess its impact on performance (Table 7). When the ROI filter was excluded, the external *AUC* decreased to 0.861, with a ΔAUC of 0.040. This confirms that ROI constraints are essential for isolating high-value sensing signals at the placental–myometrial interface. Without this structural configuration, background acoustic and electromagnetic noise diluted the discriminative power of the model, leading to weaker calibration (an *ECE* decrease of 0.067 compared with 0.114 for the developed model). Eliminating cross-modal attention reduced the external *AUC* to 0.854 ($\Delta AUC = 0.039$). This demonstrates that bidirectional refinement between ultrasound and MRI data is critical for capturing mutually reinforcing structural semantics. Without attention, the modalities contributed independently,

Table 6
Computational footprint, parameter complexity, and inference latency evaluation.

Model	Number of parameters (million)	FLOPs (giga, billion)	Training time (h)	Inference latency (ms/patient)
Ultrasound-only	11.2	11.4	1.8	12
MRI-only	11.8	14.2	2.1	18
Early fusion	13.1	16.5	2.6	24
Late fusion	13.5	16.8	2.7	26
CNN–LSTM fusion	15.8	22.1	4.1	38
GNN fusion	16.4	24.5	4.8	52
MVAE	17.2	21	3.9	35
Model developed in this study	14.2	18.6	3.4	42

Table 7
Ablation study results.

Model configuration	*AUC* (internal)	*AUC* (external)	Δ*AUC*	Decrease in *BS*	Decrease in *ECE*
ROI + cross-modal attention + robust + calibration	0.928	0.889	0.039	0.141	0.114
ROI	0.901	0.861	0.04	0.136	0.067
Cross-modal attention	0.893	0.854	0.039	0.139	0.071
Robust	0.887	0.832	0.055	0.145	0.084
Calibration	0.896	0.857	0.039	0.138	0.073

resulting in less effective fusion and diminished calibration consistency (an *ECE* decrease of 0.071). Without the robustness mechanism, the most significant degradation is observed with an external *AUC* decrease to 0.832 and a Δ*AUC* increase to 0.055. In sensor data, this highlights the vulnerability of multimodal systems when one sensor stream is missing or incomplete. The robustness mechanism acts as safeguard against signal drift, ensuring diagnostic stability even when MRI or physiological data are unavailable. The absence of this mechanism led to inefficient calibration (an *ECE* decrease of 0.084) and higher probability error (a *BS* decrease of 0.145).

The complete architecture of the developed model (ROI + cross-modal attention + robustness mechanism + calibration) showed the best performance, with an internal *AUC* of 0.928, an external *AUC* of 0.889, and the smallest Δ*AUC* (0.039). Calibration metrics were also improved substantially, with decreases in *BS* (0.141) and *ECE* (0.114). This confirms that the integration of structural priors, semantic alignment, robustness mechanisms, and calibration strategies collectively enables superior discriminative ability, generalization, and trustworthy probability estimates.

We evaluated clinical utility across probability decision thresholds ($t \in [0.1,0.9]$) using *NB*.

$$NB = \frac{TP}{N} - \frac{FP}{N}\left(\frac{t}{1-t}\right) \tag{19}$$

Here, *TP* is the *TP* count, *FP* is the *FP* count, and *N* is the total sample count. Table 8 shows *NB* values across clinical decision thresholds. The calibrated model showed a higher *NB* across all decision thresholds (0.279 at $t = 0.3$, 0.191 at $t = 0.5$, and 0.112 at $t = 0.7$), outperforming both default intervention baselines and uncalibrated multimodal models.

At the low-threshold tier ($t = 0.20-0.30$), a risk level of 20–30% represents the point where clinicians prepare for possible complications. This includes transferring the patient to a tertiary care hospital, arranging blood supplies, and securing a specialized operating suite. At $t = 0.30$, the calibrated model achieves an *NB* of 0.279, compared with 0.182 for the treat-all approach. This improvement means the model correctly identifies about 10 more high-risk patients out of every 100, without triggering unnecessary alarms for surgical preparation. At the high-threshold tier ($t = 0.50-0.70$), a risk level of 50–70% presents the decision boundary for an aggressive preventive procedure, such as balloon occlusion catheter placement in the pelvic arteries or planned hysterectomy. At $t = 0.50$, the calibrated model maintains $NB = 0.191$, compared with

Table 8
Ablation study results.

Model	*NB*				
	$t = 0.1$	$t = 0.3$	$t = 0.5$	$t = 0.7$	$t = 0.9$
Treat-all	0.321	0.182	0.095	−0.021	−0.214
Treat-none	0	0	0	0	0
Ultrasound-only	0.312	0.214	0.126	0.041	0.008
MRI-only	0.328	0.228	0.139	0.058	0.012
Early fusion	0.339	0.241	0.152	0.069	0.015
Late fusion	0.344	0.247	0.158	0.073	0.018
Developed model (uncalibrated)	0.358	0.262	0.173	0.089	0.022
Developed model (calibrated)	0.372	0.279	0.191	0.112	0.031

0.095 for the treat-all approach. This difference reduces unnecessary invasive procedures in patients incorrectly flagged as high risk, ensuring sensitivity for severe accreta or percreta patients.

4.3 Reliability

For clinical or industrial sensor deployment, the raw data must be trustworthy and discriminative. Calibration was conducted to meet such a requirement by aligning predicted confidence scores with the true likelihood of events. In this study, temperature scaling was applied as a reliability assessment of the sensing system. The results in Table 5 show that calibration did not alter *AUC* (0.889), meaning the model's ranking capability, the ability to distinguish PAS-positive from PAS-negative patients, remained unchanged. However, calibration substantially improves the quality of probability estimates. Specifically, *ECE* decreased from 0.114 to 0.048, and *BS* from 0.122 to 0.103 (Table 9). These changes indicate that the model's predicted probabilities became much closer to the actual incidence rates observed in the data. Without calibration, the model behaves as a binary classifier, outputting probabilities that might be over- or under-confident. With calibration, the model can be used for high-precision medical risk evaluation, where confidence values correspond to the true physical probability of PAS. This ensures that clinicians can interpret the model's outputs as reliable risk estimates, supporting decision-making under varying thresholds of intervention.

The reduction in *BS* from 0.122 to 0.103 (a 15.6% improvement) has direct implications for preoperative surgical management. In a cohort with an empirical PAS event rate of about 30%, a non-informative null model that predicts only overall prevalence yields a reference *BS* as follows.

$$BS_{ref} = p(1-p)^2 + (1-p)p^2 \approx 0.210 \tag{20}$$

The calibrated *BS* of 0.103 indicates that predicted probabilities represent true individual risk distributions rather than overconfident classification outputs. For surgical teams, this probability accuracy guides whether to mobilize specialized multidisciplinary resources, such as cell saver autologous transfusion systems and gynecologic oncology personnel, before skin incision. *NB* values across decision threshold ranges also correspond to the tiers of clinical intervention.

Table 9
Comparison before and after calibration of sensor data.

Calibration	*AUC*	Decrease in *BS*	Decrease in *ECE*
With	0.889	0.103	0.114
Without	0.889	0.122	0.048

4.4 Clinical applicability and *NB*

The clinical applicability of the model was evaluated using DCA, which quantifies the *NB* of different models across clinically relevant risk thresholds. *NB* integrates *TP*s and *FP*s into a single measure, reflecting whether a model provides a meaningful decision-making value compared with extreme strategies. The treat-all model assumes that every patient is treated as PAS-positive and undergoes intervention. Although the models maximize sensitivity, they present substantial overtreatment, unnecessary surgical interventions, and wasted hospital resources. At higher thresholds (t = 0.7 s), *NB* becomes negative (–0.021), showing that indiscriminate treatment harms more than it helps. The treat-none model assumes that no patient is treated as PAS-positive. It avoids overtreatment but misses true PAS patients entirely. Its *NB* is consistently zero, meaning it provides no clinical benefit.

Single-modal models (ultrasound-only and MRI-only) outperformed the treat-all and treat-none strategies, but their *NB* values remain modest (ultrasound-only *NB* = 0.041 at t = 0.7 s). This reflects the limited discriminative power of relying on a single sensor modality. Traditional fusion models (early fusion and late fusion) improved *NB* further, with late fusion slightly outperforming early fusion (*NB* = 0.073 vs 0.069 at t = 0.7 s). However, they lacked the robustness and calibration of the developed model. The uncalibrated model developed in this study showed a higher *NB* across all thresholds (0.089 at t = 0.7 s), demonstrating the value of ROI constraints, cross-modal attention, and robustness mechanisms. The calibrated model showed the highest *NB* at every threshold, reaching 0.279 at t = 0.3 s, 0.191 at t = 0.5s, and 0.112 at t = 0.7 s. The results indicate that calibration improved probability reliability and clinical applicability, enabling the efficient allocation of resources and the avoidance of unnecessary interventions (Table 10).[17,32]

4.5 Failure mode and subgroup analysis

Despite its overall robustness, the model developed showed performance variability across clinical scenarios (Table 11). The error rate for posterior placentas (0.182) was higher than that for anterior placentas (0.109). This difference occurs because ultrasound signals must traverse more maternal tissue and fluid when the placenta is posterior, leading to the larger attenuation and scattering of acoustic waves. As a result, posterior placentas are more likely to be missed, with an *FN* rate of 0.121 compared with 0.071 for anterior placentas.

Clinical history also affects the model's performance. Patients with ≥2 prior cesarean sections (CS) for deliveries showed a lower error rate (0.098) than those with 0–1 prior CS (0.137). This reflects structural changes in the uterine wall after multiple surgeries, which provide stronger

Table 10
NB values of different models at threshold intervals (t).

Model	*NB* (t = 0.3 s)	*NB* (t = 0.5 s)	*NB* (t = 0.7 s)
Treat-all	0.182	0.095	−0.021
Treat-none	0.000	0.000	0.000
US-only	0.214	0.126	0.041
MRI-only	0.228	0.139	0.058
Early fusion	0.241	0.152	0.069
Late fusion	0.247	0.158	0.073
Developed model (uncalibrated)	0.262	0.173	0.089
Developed model (calibrated)	0.279	0.191	0.112

Table 11
Error distribution across different subgroups.

Subgroup	*Error rate*	*FN* rate	*FP* rate
Posterior placenta	0.182	0.121	0.061
Anterior placenta	0.109	0.071	0.038
≥2 prior CS	0.098	0.062	0.036
0–1 prior CS	0.137	0.082	0.055
Low image quality	0.214	0.143	0.071
High image quality	0.087	0.052	0.035

imaging cues for PAS detection. However, *FN* and *FP* rates (0.062 and 0.036, respectively) indicate that whereas detection improves, misclassifications still occur, underscoring the need for robust multimodal fusion. Image quality is an important determinant of the accuracy of the model. Low-quality images yielded the highest error rate (0.214), with *FN* and *FP* rates of 0.143 and 0.071, respectively. These limitations stem from physical constraints such as impedance mismatch between the ultrasound probe and maternal skin or insufficient contrast in MRI sequences.[26] In contrast, high-quality images reduced the error rate to 0.087, with *FN* and *FP* rates dropping to 0.052 and 0.035, respectively, highlighting the direct correlation between sensor fidelity and diagnostic reliability.

By examining *FN* and *FP* rates with overall error rates, we found that the model's performance was better understood. *FN* errors represent missed PAS patients, which have a significant clinical risk due to undetected hemorrhage. *FP* errors represent overdiagnosis, which might lead to unnecessary surgical interventions and resource strain. By analyzing the errors, we evaluated the model in terms of its accuracy, safety, and clinical applicability.

4.6 Visual analysis of ROI and boundaries

To demonstrate the structural advantage of the developed model, several correctly identified PAS-positive MRI cases were analyzed. According to the design principle of the model, accurate imaging information in PAS is obtained from the placenta–uterine myometrial interface and adjacent hazardous regions. The model prioritizes the interface-related target areas. In Fig. 1, the highlighted ROIs are located at the placental–myometrial interface and the lower uterine segment. This alignment with the structure-prior ROI confirms that the model can stably capture clinically significant cues, reinforcing its interpretability and consistency with clinical logic.

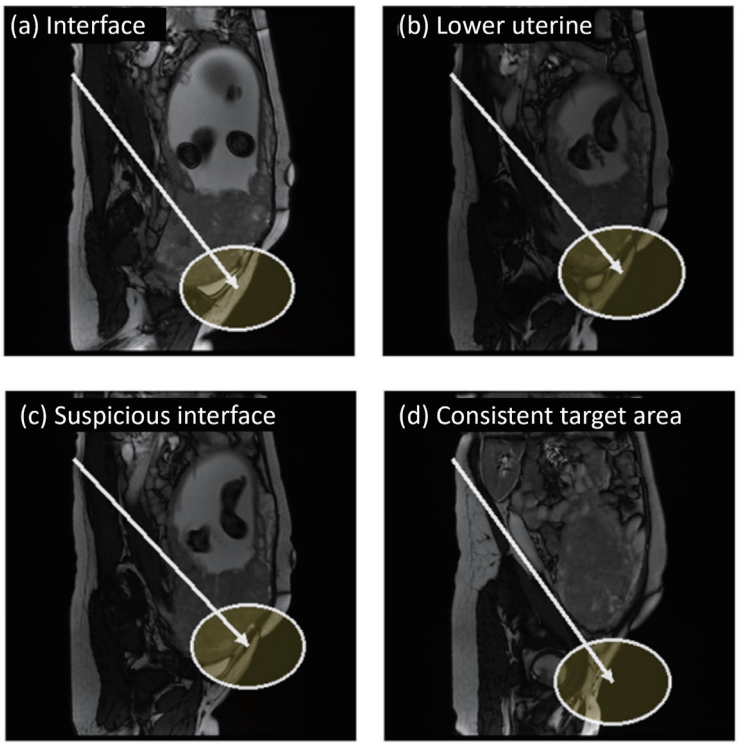

Fig. 1. (Color online) MR images of PAS-positive patients correctly identified. Representative sagittal MR slices highlighting diagnostic features of PAS: (a) Clear placental–myometrial interface, (b) lower uterine involvement, (c) suspicious interface changes, and (d) consistent target area. Arrows indicate ROIs used for model recognition.

Whereas these visual presentations show the main performance results, the developed model improves *AUC* values on the external test dataset and increases *NB* at the thresholds. The advantages of the model include a stable focusing ability on interface-related structural regions. The model does not rely solely on whole-image texture differences for black-box discrimination. Instead, under the guidance of the structure prior, it shows risk areas around the high-value placental–myometrial interface, enhancing the consistency between discriminative outcomes and clinical interpretation. Several cases remain highly challenging under structure-prior constraints. When the interface boundary is blurred, local signs are weak, or when the abnormal region is not sufficiently concentrated, the expression of structural cues diminishes, increasing uncertainty in model identification. Figure 2 shows such difficult patients, where the interface-related regions are concealed and the abnormal signals are less prominent. These conditions weaken the model's ability to stably capture structures and are consistent with the stratified error statistics, which show complex boundaries, weak signs, and low-salience abnormalities.

Figures 1 and 2 show contrasting scenarios that highlight the strengths and limitations of the developed model. In the correctly identified patients, the model demonstrates a stable ability to form structural foci around the placenta–uterine interface, capturing clinically significant cues in line with the structure-prior ROI design. In the challenging patients, weak abnormalities, blurred boundaries, or atypical local signs diminish the clarity of structural cues, reducing the

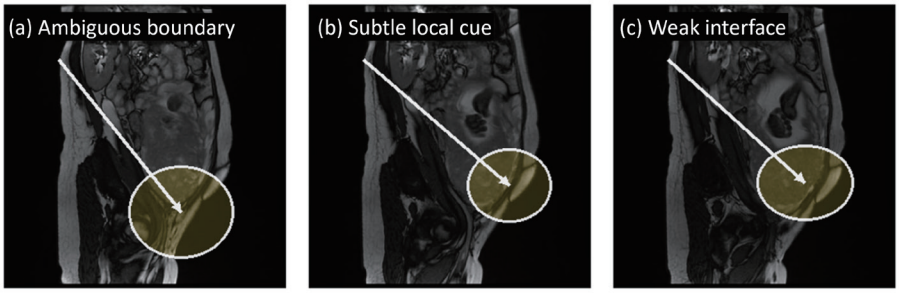

Fig. 2. (Color online) MR images of patients with challenging PAS diagnosis. Examples of difficult diagnostic cases where boundaries are less distinct: (a) ambiguous tissue interface, (b) subtle local cue, and (c) weak placental–myometrial boundary. Arrows mark ROIs illustrating diagnostic uncertainty.

model's discriminative stability. These observations confirm that the performance improvement of the model is not the result of unconstrained whole-image learning, but rather stems from its reliance on the interface ROI. Through the continuous refinement of the model's capacity to represent weak interface regions and adapt to complex scenarios, model robustness and clinical interpretability can be further enhanced.

5. Discussion

The performance of the model developed depends on the interaction between sensing hardware and biological materials. Ultrasound transducers, built on high-sensitivity piezoelectric elements, capture microstructural changes in the uterine wall, whereas MRI coils detect electromagnetic signatures associated with tissue vascularity.[26] By treating these distinct sensing modalities as components of a unified biomimetic array, the model overcomes the inherent limitations of individual sensors and demonstrates how heterogeneous hardware can be orchestrated into a coherent diagnostic system.[19]

The results of this study provide a methodology for heterogeneous sensor data fusion. Traditionally, ultrasound (acoustic) and MRI (electromagnetic) data are processed as independent observations. In this study, the data were fused as complementary inputs within a unified sensing array by using a bidirectional cross-attention mechanism. The results show that the data fusion enhances signal resolution in high-attenuation environments, such as the posterior placental–decidual interface, where unimodal sensing often fails owing to acoustic shadowing or field inhomogeneities. Through data fusion, the reliability of multi-source sensing systems can be enhanced through robustness learning and calibration. The implementation of missing-modality robust learning provides a technical basis for resilient sensing. By training the model to maintain consistent risk-ranking outputs even when secondary sensors are unavailable, we can enhance operational stability and diagnostic accuracy across diverse hardware environments, regardless of sensor-specific failures or incomplete data streams.

The minimal generalization gap (0.039) can be a new benchmark for medical sensor development. By leveraging structural components such as spatial filters, the model isolates biological signals from noise signatures introduced by different device manufacturers and

transducer materials. This is crucial for the standardization of intelligent sensing systems, aligning with TRIPOD + AI and CLAIM 2024 guidelines to ensure that diagnostic performance reflects biological reality rather than hardware artifacts.

To enhance clinical trust and algorithmic transparency, we visualized spatial attention distributions across cross-modal fusion layers using gradient-weighted class activation mapping and cross-attention token weight matrices.[33] Feature attribution heatmaps present representative *FP* and *TP* PAS patients across ultrasound and MRI modalities. In *TP* patients, the model's highest attention density (a weight distribution > 85%) concentrated along the loss of the retroplacental clear zone and hypervascular intraplacental lacunae, matching radiologist annotations. In ambiguous or artifact-heavy acoustic regions, cross-attention maps demonstrated that the network dynamically shifted feature reliance toward high-resolution MRI spatial tokens, verifying that fusion decisions are driven by biologically plausible features rather than background noise.

It is still required to refine signal processing methods for posterior placentas and to integrate real-time calibration to further minimize the generalization gap between sensing materials and devices.[16] This involves exploring advanced piezoelectric materials with broader bandwidths and developing adaptive electromagnetic sensing protocols, enhancing the fidelity and resilience of the biomimetic sensors. In deploying multimodal deep learning models, challenges exist when deployed across different clinical sites, related to hardware heterogeneity, algorithmic bias, and patient data privacy. Imaging inputs in this study were acquired using diverse hardware systems, including ultrasound transducers with center frequencies ranging from 2.0 to 6.0 MHz and MRI scanners operating at 1.5 or 3.0 T. Vendor-specific signal processing, such as the use of speckle reduction filters or gain compression curves, introduces domain shifts that cause models to overfit to site-specific noise profiles. To reduce the shifts, z-score intensity normalization and spatial domain alignment are introduced before feature extraction in this study. Clinical deployment, however, requires ongoing monitoring across demographic groups and vendor platforms to prevent bias against underrepresented patients or lower-resolution imaging systems.

Multi-center data centralization poses strict regulatory and privacy concerns under the Health Insurance Portability and Accountability Act and the General Data Protection Regulation. To enable multi-center collaborative training without raw data centralization, federated learning with local differential privacy must be integrated with the developed model.[34] We can scale multi-institution validation by transmitting encrypted gradient updates rather than raw patient imaging, while safeguarding patient confidentiality and maintaining compliance with local privacy directives.

The multi-sensor fusion model developed in this study serves as a transferable framework for diverse biological sensing applications where acoustic, electromagnetic, and physiological signals intersect. In obstetrics, this paradigm translates directly to the early prediction of fetal growth restriction and preeclampsia. These conditions involve dynamic placental vascular dysfunction, requiring the real-time synchronization of acoustic Doppler velocity profiles (umbilical and uterine arteries) with maternal physiological signals. The ROI-constrained cross-attention mechanism enables the spatial isolation of the retroplacental clearance zone while dynamically updating risk estimates using continuous maternal vitals. In cardiovascular diagnostics, the developed model can be applied to carotid plaque vulnerability and aortic

aneurysm risk stratification. Evaluating plaque stability requires fusing acoustic intravascular ultrasound data detailing tissue echo density with magnetic resonance angiography measuring wall shear stress and local blood pressure waveforms. The missing-modality robust learning strategy ensures that diagnostic models maintain stable risk stratification even when high-resolution imaging modalities are unavailable during point-of-care screening.[35]

6. Conclusion

There are three challenges in the preoperative risk assessment of PAS: the insufficient utilization of multimodal information, limited cross-center generalization, and the inadequate credibility of risk probabilities. To overcome these issues, we developed a multi-source medical sensor data fusion model. Focusing on the placental–myometrial interface ROI, the model offers a pipeline from structure-prior constraint to trustworthy risk output, incorporating unimodal representation learning, bidirectional cross-modal attention, missing-modality robust optimization, and post-processing calibration through temperature scaling. The validation results showed strong discriminative performance, with *AUC*s of 0.928 and 0.889 on internal and external test datasets, respectively. The model consistently outperformed traditional unimodal fusion strategies and ablation variants, exhibiting only a small cross-center performance decrease. Ablation study results showed that the interface ROI constraint, cross-modal interaction, and robust learning mechanisms contributed substantially to performance gains, underscoring that the advantages of the developed model stem from structure-prior-guided fusion rather than simple multimodal stacking. Furthermore, probability calibration significantly improved reliability, reducing *BS* and *ECE*, and enhancing clinical *NB* at key thresholds.

Beyond clinical applicability, the model developed provides a methodology for fusing heterogeneous medical sensor data, addressing the semantic problem between acoustic impedance signals from ultrasound and electromagnetic relaxation signals from MRI. The results of this study, conducted by implementing missing-modality robust learning, provide a basis for resilient sensor development for stable performance even under sensor failure or incomplete data conditions. A narrow generalization gap (0.039) of the model shows its ability to isolate biological signatures from hardware-specific artifacts, such as variations in piezoelectric transducer sensitivity or MRI coil performance. The calibrated sensing output ($ECE = 0.048$) presents the reliability of risk assessment, outperforming traditional fusion methods in both diagnostic accuracy and net clinical benefit.

The intelligent sensing model developed for PAS diagnosis can process ultrasound and MRI data as components of a biomimetic sensing array, not as isolated imaging modalities. By fusing acoustic, electromagnetic, and physiological data, we can overcome the inherent limitations of unimodal sensing and realize a foundation for the deployment of intelligent, multi-source sensing systems in high-stakes medical and industrial diagnostics. In further studies, resolution at the placental–myometrial interface can be considerably enhanced by integrating real-time sensor calibration and introducing advanced piezoelectric materials with higher bandwidths and adaptive electromagnetic protocols.

References

1 A. G. Cahill, R. Beigi, R. P. Heine, R. M. Silver, and J. R. Wax: Obstet. Gynecol. **132** (2018) e259. https://doi.org/10.1097/aog.0000000000002983
2 E. Jauniaux and G. Burton: Clin. Obstet. Gynecol. **61** (2018) 743. https://doi.org/10.1097/grf.0000000000000392
3 E. Jauniaux, F. Chantraine, R. M. Silver, and J. Langhoff-Roos: Int. J. Gynaecol. Obstet. **140** (2018) 265. https://doi.org/10.1002/ijgo.12407
4 I. M. Usta, E. M. Hobeika, A. A. Abu Musa, G. E. Gabriel, and A. H. Nassar: Am. J. Obstet. Gynecol. **193** (2005) 1045. https://doi.org/10.1016/j.ajog.2005.06.037
5 A. G. Eller, M. A. Bennett, M. Sharshiner, C. Masheter, A. P. Soisson, M. Dodson, and R. M. Silver: Obstet. Gynecol. **117** (2011) 331. https://doi.org/10.1097/aog.0b013e3182051db2
6 S. A. Shainker, B. Coleman, I. E. Timor-Tritsch, A. Bhide, B. Bromley, A. G. Cahill, M. Gandhi, J. L. Hecht, K. M. Johnson, D. Levine, J. Mastrobattista, J. Philips, L. D. Platt, A. A. Shamshirsaz, T. D. Shipp, R. M. Silver, L. L. Simpson, J. A. Copel, and A. Abuhamad: Am. J. Obstet. Gynecol. **224** (2021) B2. https://doi.org/10.1016/j.ajog.2020.09.001
7 F. D'Antonio, C. Iacovella, and A. Bhide: Ultrasound Obstet. Gynecol. **42** (2013) 509. https://doi.org/10.1002/uog.13194
8 A. M. Maged, A. El-Mazny, N. Kamal, S. I. Mahmoud, M. Fouad, N. El-Nassery, A. Kotb, W. S. Ragab, A. I. Ogila, A. A. Metwally, Y. Lasheen, R. M. Fahmy, M. Katta, E. K. Shaeer, and N. Salah: BMC Pregnancy Childbirth. **23** (2023) 354. https://doi.org/10.1186/s12884-023-05675-6
9 H. Lin, L. Li, Y. Lin, and W. Wang: Comput. Math Methods Med. **2022** (2022) 2751559. https://doi.org/10.1155/2022/2751559
10 J. E. Fitzgerald and H. Fenniri: Comput. Math. Methods Med. **2022** (2022) 27515559. https://doi.org/10.1039/C6RA16403J
11 L. Gong, A. R. Wright, K. Hynynen, and D. E. Goertz: Sensors **25** (2025) 5417. https://doi.org/10.3390/s25175417
12 A. C. Yu, B. Mohajer, and J. Eng: Radiol. Artif. Intell. **4** (2022) e210064. https://doi.org/10.1148/ryai.210064
13 C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger: PMLR **70** (2017) 1321. https://proceedings.mlr.press/v70/guo17a.html.
14 B. Van Calster, D. J. McLernon, M. van Smeden, L. Wynants, and E. W. Steyerberg: BMC Med. **17** (2019) 230. https://doi.org/10.1186/s12916-019-1466-7
15 B. Van Calster, D. Nieboer, Y. Vergouwe, B. De Cock, M. J. Pencina, and E. W. Steyerberg: J. Clin. Epidemiol. **74** (2016) 167. https://doi.org/10.1016/j.jclinepi.2015.12.005
16 H. Haick and N. Tnag: ACS Nano **15** (2021) 3557. https://doi.org/10.1021/acsnano.1c00085
17 A. J. Vickers and E. B. Elkin: Med. Decis. Making **26** (2006) 565. https://doi.org/10.1177/0272989x06295361
18 H. Yu, H. Li, X. Sun, and L. Pan: Biomimetics **8** (2023) 293. https://doi.org/10.3390/biomimetics8030293
19 K. A. Shah, Z. Halim, S. Anwar, C.-C. Hsu, and I. Rida: Inform. Fusion **125** (2025) 103503. https://doi.org/10.1016/j.inffus.2025.103503
20 G. S. Collins, K. G. M. Moons, P. Dhiman, R. D. Riley, A. L. Beam, B. Van Calster, M. Ghassemi, X. Liu, J. B. Reitsma, M. van Smeden, A.-L. Boulesteix, J. C. Camaradou, L. A. Celi, S. Denaxas, A. K. Denniston, B. Glocker, R. M. Golub, H. Harvey, G. Heinze, M. M. Hoffman, A. P. Kengne, E. Lam, N. Lee, E. W. Loder, L. Maier-Hein, B. A. Mateen, M. D. McCradden, L. Oakden-Rayner, J. Ordish, R. Parnell, S. Rose, K. Singh, L. Wynants, P. Logullo: BMJ **385** (2024) e078378. https://doi.org/10.1136/bmj-2023-078378
21 A. S. Tejani, M. E. Klontzas, A. A. Gatti, J. T. Mongan, L. Moy, S. H. Park, C. E. Kahn Jr: Radiol. Artif. Intell. **6** (2024) e240300. https://doi.org/10.1148/ryai.240300
22 R. Cipolla, Y. Gal, and A. Kendall: Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition (IEEE, 2018) 7482. https://doi.org/10.1109/CVPR.2018.00781
23 L. R. Dice: Ecology **26** (1945) 297. https://doi.org/10.2307/1932409
24 Y. Iwasa, T. Iwashita, Y. Takeuchi, H. Ichikawa, N. Mita, S. Uemura, M. Shimizu, Y.-T. Kuo, H.-P. Wang, and T. Hara: J. Clin. Med. **10** (2021) 3589. https://doi.org/10.3390/jcm10163589
25 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin: Proc. 31st International Conf. Neural Information Processing Systems (NIPS, 2017) 6000. https://doi.org/10.5555/3295222.3295349
26 M. Wang, C. Zhang, W. Zhang, C. Cheng, Y. Hu, X. Li, Q. Liang, Q. Zhang, and Y. Hou: Rev. Mater. Res. **1** (2025) 100062. https://doi.org/10.1016/j.revmat.2025.100062
27 (M. T. Hayat, Y. M. Allawi, W. Alamro, S. M. Sultan, A. Abadleh, H. Kang, and A. I. Zreikat: Diagnostics **15** (2025) 2673. https://doi.org/10.3390/diagnostics15212673

28 M. Vaida and Z. Huang: Front. Artif. Intell. **8** (2025) 1716706.
29 M. Da Silva–Filarder, A. Ancora, M. Filippone, and P. Michiardi: Proc. 20th IEEE Int. Conf. Machine Learning and Applications (IEEE, 2021) 1069. https://doi.org/10.1109/ICMLA52953.2021.00175
30 G. W. Brier: Mon. Wea. Rev. **78** (1950) 1. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2
31 M. Pakdaman Naeini, G. Cooper, and M. Hauskrecht: Proc. AAAI Conf. Artif. Intell. **29** (2015) 9602. https://doi.org/10.1609/aaai.v29i1.9602
32 B. Vasey, M. Nagendran, B. Campbell, D. A. Clifton, G. S. Collins, S. Denaxas, A. K. Denniston, L. Faes, B. Geerts, M. Ibrahim, X. Liu, B. A. Mateen, P. Mathur, M. D. McCradden, L. Morgan, J. Ordish, C. Rogers, S. Saria, D. S. W. Ting, P. Watkinson, W. Weber, P. Wheatstone, P. McCulloch, and DECIDE-AI expert group: BMJ **377** (2022) e070904. https://doi.org/10.1136/bmj-2022-070904
33 R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra: Int. J. Comput. Vis. **128** (2020) 336. https://doi.org/10.1007/s11263-019-01228-7
34 R. Kumar, C.-S. Shieh, P. Chakrabarti, A. Kumar, J. Moolchandani, and R. Sinha: EPJ Web Conf. **328** (2025) 01066. https://doi.org/10.1051/epjconf/202532801066
35 A. Gunjal and T. Judgi: MethodsX **15** (2025) 103678. https://doi.org/10.1016/j.mex.2025.103678