AI-Assisted Detection of Lateral Ventricular Abnormalities in 2D Fetal Brain Ultrasound Project

Introduction

Accurate prenatal evaluation of the fetal brain is of critical importance for the detection, characterization, and prognostic assessment of congenital intracranial anomalies, directly influencing both prenatal counseling and perinatal management (1). Two-dimensional (2D) ultrasonography (USG) remains the cornerstone imaging modality for fetal brain assessment worldwide due to its wide availability, real-time imaging capability, and absence of ionizing radiation (1,2). However, the diagnostic performance of fetal brain USG is inherently limited by several factors, including operator experience, fetal position, acoustic window quality, and variability in image acquisition techniques (2). Consequently, subtle or atypical abnormalities may be overlooked, and substantial inter-observer variability may arise both in the acquisition of appropriate imaging planes and in image interpretation (1,2).

Over the past decade, rapid advances in artificial intelligence (AI), particularly in deep learning, have fundamentally transformed medical image analysis and have been increasingly integrated into obstetric ultrasonography (3–5). AI-based systems have demonstrated expert-level or near–expert-level performance in a variety of fetal ultrasound applications, including automatic recognition of standard imaging planes, organ segmentation, biometric measurements, and anomaly classification when evaluated in retrospective datasets (3,4). These technologies offer the potential to reduce operator dependence, improve reproducibility of measurements and interpretations, shorten examination time, and provide real-time decision support during routine screening or targeted assessments (4–6). In the context of fetal imaging, AI models developed for fetal brain ultrasound have reported promising results, particularly for the detection of conditions such as ventriculomegaly and related ventricular abnormalities (7,8). Nevertheless, a substantial proportion of existing studies are limited by their focus on a single imaging plane or specific lesion types, relatively small sample sizes, or a lack of systematic evaluation using heterogeneous datasets that reflect real-world clinical practice (1,9).

Given the ongoing need for reliable, standardized, and scalable tools to support the interpretation of fetal brain ultrasonography in routine clinical settings, there is a strong rationale for the development and validation of AI-assisted diagnostic models utilizing large archives of 2D fetal brain ultrasound images (1,4,9). Accordingly, the objective of this study was to develop an AI-assisted diagnostic model for the analysis of archived 2D fetal brain ultrasound images, with the aim of improving the detection of **lateral ventricular abnormalities** and contributing to greater standardization of image interpretation across operators and varying examination conditions. This study was deliberately designed as an anatomy-focused, proof-of-concept initial model, with the first phase concentrating on the **lateral ventricular plane**, which is highly reproducible, clinically relevant, and routinely acquired in fetal neurosonography. This approach is intended to provide a scalable foundation for subsequent development of a more comprehensive, multiplanar AI-assisted neurosonography framework incorporating the posterior fossa, midline structures, and additional intracranial planes.

Study Objectives and Hypothesis (Lateral ventricle–focused)

The primary objective of this study was to evaluate the diagnostic performance of an **artificial intelligence–assisted model trained on expert-labeled archived 2D fetal lateral ventricular (Vp) ultrasound images** for the detection of **lateral ventricular abnormalities**. Specifically, we aimed to assess the model's diagnostic accuracy, sensitivity, specificity, and error patterns in distinguishing normal from abnormal findings using real-world clinical datasets, as an anatomy-focused proof-of-concept initial model for subsequent expansion to a comprehensive multiplanar fetal neurosonography framework (1–4).

The secondary objective was to explore whether **AI-assisted analysis of lateral ventricular imaging** could reduce inter-observer variability and contribute to improved standardization of fetal brain ultrasound interpretation, addressing a key limitation of operator-dependent neurosonographic assessment and providing a scalable foundation for broader clinical implementation (1,2,4).

Additionally, we sought to identify **specific lateral ventricular abnormality patterns or image characteristics** associated with misclassification—insights that may guide refinement of AI-based diagnostic tools and support their translation into clinical workflows (3–5,7).

We hypothesized that an AI-assisted diagnostic model trained on a large set of expert-labeled archived fetal brain ultrasound images would (1) achieve high diagnostic accuracy comparable to expert interpretation for **lateral ventricular abnormality detection**, (2) improve the detection of lateral ventricular abnormalities relative to traditional operator-dependent assessment, and (3) contribute to greater standardization and reproducibility in fetal neurosonography by reducing interpretive variability (1–5,7).

Methods

This observational diagnostic study was conducted between **October 1, 2025, and November 30, 2025**, using previously acquired and archived **2D fetal brain ultrasound images** obtained during routine prenatal examinations. All ultrasound scans were performed using a **Voluson E8 system (2021, GE Healthcare, Zipf, Austria)**. No patients were re-examined, and no new ultrasound acquisitions or measurements were performed specifically for the purposes of this study.

The study evaluated a **deep learning–based artificial intelligence model** designed to classify **fetal lateral ventricular findings** as **normal**

or abnormal on 2D ultrasound images. Approximately **3,000 fetal brain ultrasound images** containing intracranial structures were initially screened and deemed suitable for expert review. From this archive, images demonstrating the lateral ventricular plane were selected for subsequent analysis.

Abnormality was defined as a **lateral ventricular width greater than 10 mm** measured in the **standard axial transventricular plane**, in accordance with guidelines established by the **International Society of Ultrasound in Obstetrics and Gynecology (ISUOG)** (10).

Ethical approval for the study was obtained from the **MEF University Ethics Committee** (Approval No. **E-47749665-050.04-4465**). The study was registered in a clinical trials registry, and a **ClinicalTrials.gov identification number (NCT07261618)** was obtained.

The study followed an **observational diagnostic accuracy design** using **fully anonymized ultrasound images** captured between **18 and 24 weeks of gestation** as part of routine prenatal care. All images were retrieved from the institutional digital archive and contained no patient identifiers. No new imaging procedures or patient interactions occurred during the study. All data handling complied with ethical standards governing the secondary use of clinical imaging data.

Eligibility Criteria

Eligibility criteria were defined *a priori* according to the clinical trial protocol and methodological framework. Included images met all of the following criteria:

- Archived **2D fetal brain ultrasound images** obtained between **18 and 24 weeks of gestation**
- Acquired from individuals aged **18 to 45 years** at the time of imaging
- Adequate visualization of the **lateral ventricles**, primarily in the **transventricular and transthalamic planes**

- Diagnostic image quality sufficient for structural assessment
- Fully anonymized, with all identifying information removed
- Suitable for **ground-truth classification** by expert reviewers

A **normal case** was defined as one demonstrating stored freeze-frame ultrasound images in standard axial planes—**transventricular, transthalamic, and transcerebellar**—acquired in accordance with **ISUOG guidelines**, with **lateral ventricular width <10 mm** and normal ventricular morphology (10).

Images were excluded if they:

- exhibited insufficient diagnostic quality or motion artifacts,
- were incomplete, duplicated, or corrupted,
- lacked standard axial-plane images or contained color Doppler or measurement caliper overlays,
- had uncertain gestational age,
- contained any form of identifiable metadata, or
- were misfiled records unrelated to the fetal brain.

Expert Review and Ground Truth Definition

Each eligible image was independently reviewed by a clinician (**Bilge Çetinkaya Demir, MD, Assoc. Prof.**) with more than **15 years of experience in fetal neurosonography**. Images were classified as **normal or abnormal**, and abnormal cases were further subcategorized based on structural characteristics of the lateral ventricles. Any discrepancies were resolved through consensus review.

These **expert-labeled classifications served as the ground truth** for all subsequent model training, validation, and evaluation procedures.

AI Model Development

A deep learning–based classification framework was developed using a standardized and reproducible preprocessing pipeline. All **2D fetal brain ultrasound images focusing on the lateral ventricular plane** were resized to a fixed spatial resolution and intensity-normalized prior to model training to ensure consistency across heterogeneous acquisition conditions.

A major methodological challenge in this study was the pronounced **class imbalance inherent to lateral ventricular abnormality detection**, with substantially fewer abnormal cases compared to normal images. To mitigate this limitation, **synthetic abnormal lateral ventricular images (n = 116)** were generated following evidence-based strategies for handling imbalanced medical imaging datasets, as described in recent comprehensive reviews of imbalance mitigation techniques in healthcare AI (11).

Importantly, all synthetic images represented **abnormal lateral ventricular cases only** and were incorporated **exclusively into the training dataset**. No synthetic images were used during validation or real-only testing to preserve the **clinical integrity and external validity** of model evaluation.

The final dataset used for model development consisted of **529 real ultrasound images** (413 normal and 116 abnormal) and **116 synthetically generated abnormal images**, yielding a total of **645 images**. The dataset was split using an **85% / 15% train–test ratio**, with strict separation between training and evaluation sets. All validation and testing subsets consisted exclusively of **real clinical images**, ensuring that reported performance metrics reflect real-world diagnostic behavior.

Prior to final model selection, multiple model families were explored and comparatively evaluated using cross-validation, including large-scale vision transformer–based architectures (e.g., *google/*

vit_huge_patch14_224_in21k) and multimodal generative vision–language models (e.g., *unsloth/LLaMA-3.2-11B-Vision-Instruct*) fine-tuned for binary image classification. However, these approaches did not achieve stable or clinically acceptable performance for **lateral ventricular assessment**. A key limitation observed with large-scale transformer and generative models was their reliance on aggressive image upscaling, which introduced pixel-level artifacts and degraded fine anatomical detail critical for accurate ventricular evaluation in ultrasound imaging.

Similarly, conventional convolutional neural network (CNN) architectures **without imbalance-aware optimization strategies** were insufficient to achieve robust performance, particularly for the minority abnormal class. To address these limitations, a **custom hybrid CNN-based architecture optimized for grayscale medical imaging** was developed and adopted (12). This approach combined convolutional feature extraction with modern regularization strategies to improve discrimination under limited and imbalanced data conditions.

Data augmentation techniques applied during training included **small rotational perturbations, horizontal flipping, random resized cropping, controlled brightness and contrast adjustments, and localized random erasing**. All augmentations were carefully constrained to preserve **anatomical plausibility of the lateral ventricular structures**. All synthetically generated abnormal images were independently reviewed and validated by fetal sonography specialists, including **Nefise Nazlı Yenigül, MD, Assoc. Prof and Bilge Çetinkaya Demir, MD, Assoc. Prof.**, prior to inclusion in the training process.

Model optimization employed **class-weighted loss functions, adaptive learning-rate scheduling, early stopping, and mixed-precision training** to improve generalizability and training stability. In addition, a **dynamic probability decision-threshold optimization strategy** was implemented at inference time to reduce classification bias and improve fairness across classes, particularly for the underrepresented abnormal category (12). This dynamic thresholding approach aligns with established principles of **responsible and bias-aware AI design in medical diagnostics** (13).

All data preprocessing, data visualization, exploratory analysis, synthetic anomaly generation pipelines, cross-validation experiments, model architecture design, training procedures, and evaluation workflows—particularly the development and application of the dynamic probability-threshold optimization strategy for responsible and bias-aware clinical inference—were **designed and implemented by AI Engineer Erdeniz Ünvan**.

Full technical specifications, hyperparameters, and implementation details are provided in the study appendix to ensure transparency and reproducibility.

Model Evaluation

Model performance was evaluated by comparing AI-generated classifications with **expert-defined ground-truth labels** established by an experienced fetal neurosonography specialist. Diagnostic performance metrics included **overall accuracy, sensitivity (recall), specificity, precision (positive predictive value), negative predictive value, F1-score, receiver operating characteristic area under the curve (ROC-AUC), and confusion matrix analysis**.

The final dataset consisted of **645 images**, including **529 real clinical ultrasound images** (413 normal, 116 abnormal) and **116 synthetically generated abnormal images**. The dataset was partitioned using an **85% / 15% train–test split**. Accordingly, the training set comprised **548 images** (347 normal and 201 abnormal, including synthetic cases), while the held-out test set included **97 images** (66 normal and 31 abnormal).

Within the held-out test set, **all normal images were real clinical ultrasound images**, whereas the abnormal class consisted of both **real abnormal cases (n = 18)** and **synthetic abnormal cases (n = 13)**, reflecting the class-imbalance mitigation strategy applied during model development. Synthetic images were included in this mixed evaluation solely as part of the predefined data split.

To ensure **fair and clinically responsible evaluation under class imbalance**, a **dynamic probability decision-threshold optimization strategy** was employed at inference time instead of a fixed 0.5 classification threshold. The optimal decision threshold was selected via a **threshold scan** to maximize balanced diagnostic performance while reducing bias against the underrepresented abnormal class.

All misclassified cases, including **false-positive and false-negative predictions**, were subsequently reviewed qualitatively by domain experts to identify contributing factors such as **borderline lateral ventricular measurements, suboptimal acoustic windows, reduced image resolution, or intrinsic anatomical variability**.

The **primary outcome** of the study was the **diagnostic accuracy of the AI-assisted model in classifying lateral ventricular findings in fetal brain ultrasound images as normal or abnormal**, using expert-established ground-truth labels. This evaluation served as an **anatomy-focused, proof-of-concept initial model** for potential future expansion to a comprehensive multiplanar fetal neurosonography framework.

Secondary outcomes included assessment of the model's **sensitivity and specificity for detecting lateral ventricular abnormalities**, its potential to **reduce inter-observer variability** by providing standardized AI-assisted interpretation, identification of **error patterns and image characteristics associated with misclassification**, and evaluation of the model's **generalizability across heterogeneous, real-world variations in ultrasound image quality**.

Results

Approximately **10,000 prenatal ultrasound images** were initially screened for model development. Of these, approximately **3,000 images** corresponded to fetal brain intracranial structures. Following predefined inclusion criteria, **773 lateral ventricular images** were identified. Among these, **529 images** (413 normal and 116 abnormal) were selected through expert review as high-confidence ground-truth cases for model development.

To address the pronounced class imbalance inherent in **lateral ventricular abnormality detection**, **116 synthetic abnormal images** were generated and incorporated **exclusively into the training dataset**. No synthetic images were used during validation or real-only testing. This resulted in a final dataset of **645 images**, comprising **529 real clinical images** and **116 synthetic abnormal images**. All images represented routine mid-gestation fetal neurosonographic examinations performed between **18 and 24 weeks of gestation**.

The dataset was partitioned using an **85% / 15% train–test split**. The training set contained **548 images** (347 normal and 201 abnormal, including synthetic anomalies), while the held-out test set comprised **97 images** (66 normal and 31 abnormal). Within the held-out test set, **all normal images were real clinical cases**, whereas the abnormal class included both **real abnormal cases (n = 18)** and **synthetic abnormal cases (n = 13)**, reflecting the imbalance-mitigation strategy applied during model development.

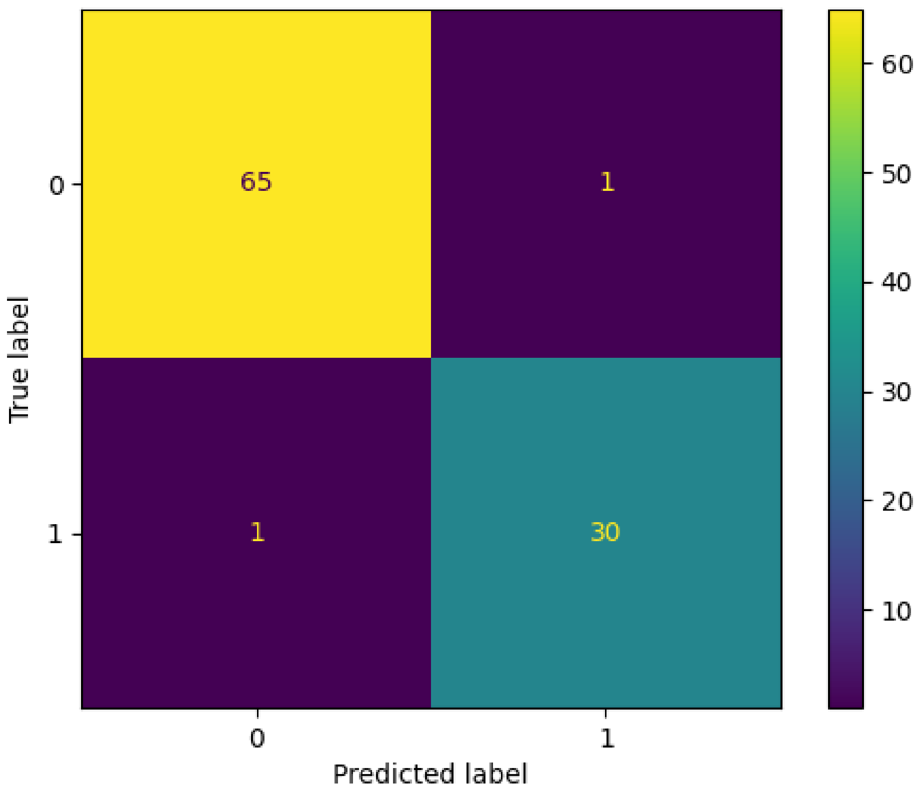
Evaluation Frameworks and Confusion Matrices

Two independent evaluation scenarios were conducted to assess model performance:

Mixed Test Evaluation (Real + Synthetic; n = 97)

This evaluation was derived from the primary 85/15 split of the total augmented dataset (N = 645).

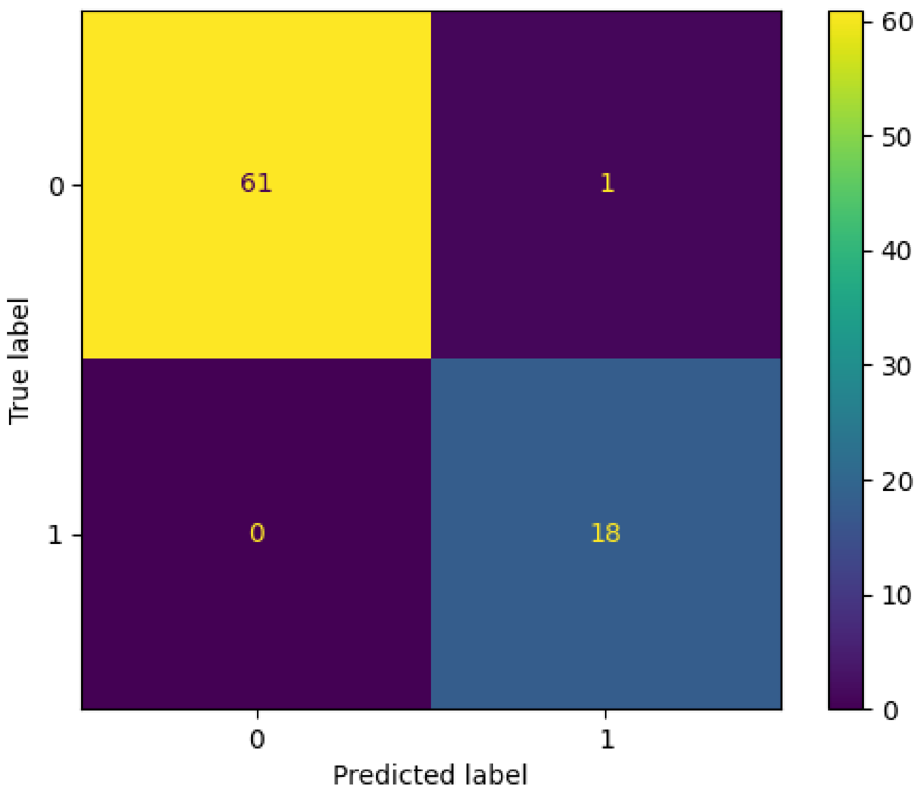
- **Distribution:** 66 normal and 31 abnormal cases (18 real + 13 synthetic).
- **Performance:** Using dynamic probability-threshold optimization, the model achieved an overall accuracy of **97.94%**, correctly classifying **65 of 66 normal cases** and **30 of 31 abnormal cases**. The resulting confusion matrix was **[[65, 1], [1, 30]]**, corresponding to one false-positive and one false-negative prediction.
- **ROC-AUC: 0.9976**, indicating excellent discriminatory performance.



Real-Only Test Evaluation (Unseen Real Images; n = 80)

To assess real-world generalizability, a secondary evaluation was performed using **exclusively real clinical images** that were never included in training. This real-only test set consisted of **62 normal and 18 abnormal images**, representing approximately **15% of the original real dataset (80/529)**.

- **Performance:** Under dynamic threshold optimization, the model achieved an accuracy of **98.75%**, sensitivity of **100%**, specificity of **98.4%**, F1-score of **0.973**, and ROC-AUC of **0.9955**.
- **Confusion matrix: [[61, 1], [0, 18]]**, with only one false-positive case and **no false-negative predictions**, indicating complete detection of abnormal lateral ventricular cases in the real-only evaluation.



Statistical Rationale

The variation in the number of normal cases between the mixed test set (66 cases) and the real-only test set (62 cases) reflects independent random shuffling and differing denominators. In the mixed evaluation, the 15% split was applied to the total augmented dataset ($645 \times 0.15 \approx 97$), whereas in the real-only evaluation it was applied solely to the real clinical subset ($529 \times 0.15 \approx 80$). This dual-validation strategy confirms that the use of synthetic data during training enhanced feature learning without introducing bias into the clinical evaluation phase.

Data Leakage Verification and Error Analysis

To ensure strict separation between training and evaluation datasets, a file-level overlap analysis was performed following dataset partitioning. All filenames associated with real training images and real test images were programmatically compared, confirming **zero overlap** and excluding any possibility of data leakage. Synthetic images were restricted exclusively to the training set and were never included in validation or real-only testing.

Qualitative expert review of the single misclassified real-image case revealed **borderline lateral ventricular measurements** and mild image degradation related to **suboptimal acoustic windows**, highlighting known challenges in fetal neurosonographic interpretation.

Discussion

This study demonstrates that an **AI-assisted diagnostic model developed exclusively using archived 2D fetal brain ultrasound images from the lateral ventricular plane** can achieve **very high diagnostic performance** in detecting **lateral ventricular abnormalities**. In real-only clinical evaluation, the model correctly classified **98.4% of normal cases and 100% of abnormal cases**, yielding an overall diagnostic accuracy of **98.75%**. Despite relying on a single imaging plane, these results indicate that **targeted, anatomy-focused deep-learning approaches** can provide substantial diagnostic value even at early stages of fetal neurosonographic assessment.

Recent literature has shown that artificial intelligence applications in obstetric imaging can reduce observer variability, improve diagnostic consistency, and limit operator dependence (14). AI-based tools applied to complex anatomical regions, including the fetal brain, have demonstrated potential to support more standardized assessment and interpretation (15,16). The high diagnostic accuracy achieved in this study using a carefully curated lateral ventricular dataset is consistent with these findings and further supports the clinical relevance of focused AI-assisted screening tools.

Several prior studies investigating AI applications in fetal brain imaging provide important context. For example, Xie et al. reported diagnostic accuracies exceeding 96% for normal–abnormal classification using axial fetal brain ultrasound images (7), while Drukker et al. highlighted the substantial operator dependency inherent in routine ultrasound workflows and the potential role of AI-based standardization tools (2). Together, these studies support the targeted lateral ventricular classification strategy employed in the present work and suggest that even a single-plane, anatomy-focused approach may offer meaningful clinical decision support.

From a methodological standpoint, AI models developed for fetal imaging often involve segmentation, identification of key anatomical structures, and anomaly classification (14,16). The architecture adopted in this study aligns with this framework and provides a **scalable foundation for future expansion** to include additional planes such as the posterior fossa, midline structures, and cortical development. The systematic expert review of misclassified cases further reflects best practices for iterative model refinement and clinical translation.

The strengths of this study include the use of a **real-world, archived clinical dataset**, rigorous **expert-based annotation**, and evaluation strategies aligned with routine clinical practice. The exclusive use of synthetic images during training effectively addressed class imbalance without inflating test performance or introducing evaluation bias, consistent with emerging best practices in medical AI development.

Several limitations should be acknowledged. The model was trained exclusively on **lateral ventricular planes** and therefore does not represent a comprehensive fetal neurosonographic assessment. The relatively limited number of abnormal cases may influence the precision of certain performance metrics, and the single-center design warrants external validation across different ultrasound systems, operators, and patient populations.

In conclusion, this study provides strong evidence for the feasibility of **AI-assisted detection of lateral ventricular abnormalities in 2D fetal brain ultrasound** as a standardized, anatomy-focused screening approach. Future work incorporating multiplanar imaging, classification of specific ventricular anomaly subtypes, and prospective clinical validation will be essential to further strengthen the clinical applicability of this approach.

References

1. Yousefpour Shahrivar R, Karami F, Karami E. Enhancing fetal anomaly detection in ultrasonography images: a review of machine learning-based approaches. Biomimetics (Basel). 2023;8(7):519. doi:10.3390/biomimetics8070519
2. Drukker L, Sharma H, Karim JN, et al. Clinical workflow of sonographers performing fetal anomaly ultrasound scans: deep-learning-based analysis. Ultrasound Obstet Gynecol. 2022;60(6):759–765. doi:10.1002/uog.24975
3. Enache IA, Iovoaica-Rămescu C, Ciobanu ȘG, et al. Artificial intelligence in obstetric anomaly scan: heart and brain. Life (Basel). 2024;14(2):166. doi:10.3390/life14020166
4. Horgan R, Nehme J, Abuhamad A. Artificial intelligence in obstetric ultrasound: a scoping review. Prenatal Diagnosis. 2023;43(10):1176–1219.
5. Chen Z, Liu Z, Du M, Wang Z. Artificial intelligence in obstetric ultrasound: an update and future applications. Frontiers in Medicine. 2021;8:733468. doi:10.3389/fmed.2021.733468
6. Miskeen E, Alfaifi J, Alhuian DM, et al. Prospective applications of artificial intelligence in fetal medicine: a scoping review of recent updates. International Journal of General Medicine. 2025;18:237–245. doi:10.2147/IJGM.S490261
7. Xie HN, Wang N, He M, et al. Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal. Ultrasound Obstet Gynecol. 2020;56(4):579–587. doi:10.1002/uog.21967
8. Vahedifard F, Awate SP, Roy AG, et al. Review of deep learning and artificial intelligence models in fetal brain MRI. Frontiers in Radiology. 2023;3:1370533. doi:10.3389/fradi.2023.1370533

9. Belciug S, Ivanescu RC, Serbanescu MS, et al. Pattern recognition and anomaly detection in fetal morphology using deep learning and statistical learning (PARADISE): protocol for the development of an intelligent decision support system. BMJ Open. 2024;14(2):e077366. doi:10.1136/bmjopen-2023-077366

10. Paladini D, Malinger G, Monteagudo A, Pilu G, Timor-Tritsch I, Toi A. Sonographic examination of the fetal central nervous system: guidelines for performing the "basic examination" and the "fetal neurosonogram". Ultrasound Obstet Gynecol. 2007;29(1):109–116.

11. Salmi M, Atif D, Oliva D, Abraham A, Ventura S. Handling imbalanced medical datasets: review of a decade of research. Artificial Intelligence Review. 2024;57(10):273.

12. Rashid AB, Kausik MAK. AI revolutionizing industries worldwide: a comprehensive overview of its diverse applications. Hybrid Advances. 2024;7:100277.

13. Hanna MG, Pantanowitz L, Jackson B, Palmer O, Visweswaran S, Rashidi HH. Ethical and bias considerations in artificial intelligence and machine learning. Modern Pathology. 2025;38(3):100686.

14. Kim HY, Cho GJ, Kwon HS. Applications of artificial intelligence in obstetrics. Ultrasonography. 2023;42(1):2–9.

15. Yannaeva NE, Bokeria EL, Sencha AN, Petrovna E. Artificial intelligence in the evaluation of the fetal heart. Ultrasound in Medicine & Biology. 2025;51(Suppl):S111.

16. Yasrab R, Fu Z, Zhao H, Lee LH, Sharma H, Drukker L, Noble JA. A machine learning method for automated description and workflow analysis of first trimester ultrasound scans. IEEE Transactions on Medical Imaging. 2022;42(5):1301–1313.