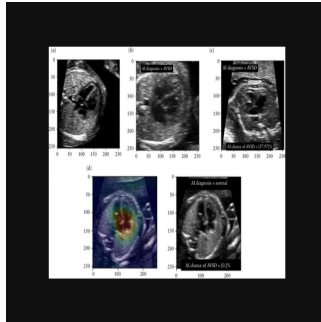


---

## Collaborative Role of AI and US Clinicians in Detecting Congenital Heart Defects



---

The antenatal detection of fetal congenital heart disease (CHD) is crucial for reducing postnatal morbidity and mortality. Despite efforts like mid-trimester ultrasound screenings, detection rates vary widely across regions. In the UK, the Fetal Anomaly Screening Programme identifies around 69.2% of atrioventricular septal defects (AVSDs) through ultrasound, and up to 79.4% with combined screening approaches.

Artificial intelligence (AI), particularly convolutional neural networks, shows promise in automating the detection of fetal CHD from ultrasound images. However, ultrasound interpretation remains reliant on human operators due to the nuanced nature of clinical decision-making. Collaborative use of AI alongside clinicians may enhance diagnostic accuracy, though recent studies indicate mixed results in other medical imaging fields like chest X-rays.

To optimise the clinician-AI partnership in ultrasound interpretation, it's crucial to explore how AI assistance affects overall performance in diagnosing conditions like fetal AVSD. Additionally, providing clinicians with more detailed insights into AI model outputs—such as confidence levels and influential image areas—could mitigate risks like algorithm aversion (underutilising AI) or automation bias (overreliance on AI). There is limited research on how humans interact with AI specifically in ultrasound interpretation, highlighting the need for further investigation in this area.

[A recent study](#) was conducted at a specialised foetal cardiology centre with access to a comprehensive image archive of cardiac abnormalities, enabling training and validation of an AI model for classifying foetal ultrasound images into normal and atrioventricular septal defect (AVSD) categories.

### Development of AI Models

An AI model was developed using a manually labelled dataset of 173 fetal ultrasound scans (98 normal hearts, 75 AVSD cases). The dataset included 121,130 images extracted from multiple ultrasound videos, focusing on high-quality four-chamber views. Technical details of model development are provided in Appendix S1.

For the experiment, a subset of 500 four-chamber images from 36 fetuses (16 AVSD, 20 normal hearts) were selected from the test and validation sets of the AI model. These images were presented to clinicians in four different conditions:

- Plain image with no AI information.
- Image with AI's binary classification (normal or AVSD).
- Image with AI's classification and temperature-scaled confidence score indicating the likelihood of AVSD.
- Image with AI's classification, confidence score, and gradient-weighted class activation map (grad-CAM) highlighting influential areas.

Clinicians from the foetal cardiology unit participated, including consultants, trainee doctors, and sonographers, with varying years of experience in foetal ultrasound. Each clinician evaluated 2000 images (500 per condition), randomly ordered, and classified them as normal or AVSD. They received training on interpreting AI outputs and the significance of confidence scores and grad-CAM images.

### Outcome Measurement

Accuracy of diagnosis was assessed by comparing clinician classifications across the four conditions. Statistical analysis using the McNemar test evaluated differences in accuracy between conditions, with significance set at  $P \leq 0.05$ . Accuracy was defined as the proportion of correct

© For personal and private use only. Reproduction must be permitted by the copyright holder. Email to [copyright@mindbyte.eu](mailto:copyright@mindbyte.eu).

classifications relative to the total number of images.

This study aimed to determine whether AI assistance, along with additional information like confidence scores and visual aids (grad-CAM), improves diagnostic performance in distinguishing fetal AVSD from normal heart conditions during ultrasound examinations.

### **Impact of AI Assistance on Clinician Performance in Foetal Ultrasound Diagnosis of AVSD**

Ten clinicians participated in the study, including two consultant foetal cardiologists, five non-medical sonographers, and three paediatric cardiology trainee doctors. Consultant foetal cardiologists had extensive experience, ranging from 20 to 29 years post-qualification, and specific expertise in foetal cardiology. Sonographers had an average of 27.4 years post-qualification and 14.2 years in foetal ultrasound, all completing Foetal Medicine Foundation training. Trainee doctors had an average of eight years post-qualification, with 2.7 years in paediatric cardiology and less than one year in fetal cardiology.

The AI model achieved an accuracy of 0.798 (95% CI, 0.760–0.832) in diagnosing AVSD from foetal ultrasound images, with sensitivity at 0.868 and specificity at 0.728. Clinicians, when interpreting images without AI assistance, performed significantly better than the AI model with an accuracy of 0.844 (95% CI, 0.834–0.854), particularly among more experienced participants (accuracy of 0.873,  $P < 0.001$ ).

Introducing AI's binary classification alongside images (Condition 2) improved overall clinician performance to an accuracy of 0.865 ( $P < 0.001$ ). However, providing additional information such as model confidence (Condition 3) led to a decrease in accuracy compared to binary classification (0.850 vs 0.865,  $P = 0.002$ ). Similar trends were observed with grad-CAM (Condition 4), especially among more experienced clinicians.

Stratifying by AI correctness, clinicians performed significantly better when AI advice was correct (accuracy 0.908 vs 0.761 when incorrect,  $P < 0.001$ ). Incorrect AI advice resulted in a notable decline in performance (accuracy 0.693), exacerbated by additional information (model confidence or grad-CAM).

The study highlights that while AI assistance can enhance diagnostic accuracy, especially with correct advice, additional information beyond binary classification may not consistently improve clinician performance and could even diminish it, particularly among more experienced practitioners.

### **Integrating AI for AVSD Detection and Clinician Collaboration**

Researchers demonstrated a significant improvement in diagnostic performance when clinicians received AI assistance in interpreting foetal ultrasound images, compared to relying on either AI or clinicians alone. This enhancement was observed across both experienced foetal cardiology experts and less-experienced trainee doctors, even when the AI's standalone performance was inferior to that of clinicians. This suggests that integrating AI into foetal ultrasound screening for structural malformations, such as atrioventricular septal defects (AVSDs), could be beneficial, despite neither humans nor AI achieving perfect accuracy.

The study found that experienced foetal cardiology clinicians outperformed AI alone, whereas less-experienced operators performed similarly to AI alone. The impact of AI assistance may vary in different clinical settings, which warrants further investigation. Similar variability has been observed in other medical imaging studies concerning the effectiveness of AI in improving diagnostic accuracy.

When AI provided incorrect diagnoses, clinician performance significantly declined due to increased false positives and negatives. This underscores the importance of "trust calibration," where clinicians learn to trust AI outputs appropriately based on their accuracy. However, providing additional AI model information, such as confidence scores or grad-CAM images highlighting influential areas, did not improve clinician trust calibration in this study. In fact, it sometimes led to clinicians incorrectly trusting AI outputs when they were incorrect—a phenomenon known as "automation bias."

The study acknowledges several limitations, including its retrospective nature using still images rather than real-time ultrasound videos, which may affect the generalizability of findings to clinical practice. Additionally, the dataset focused on antenatally diagnosed cases and may not represent cases typically missed in screening programmes. Future research aims to develop AI techniques for real-time clinical use and to expand AI models to cover a wider spectrum of congenital heart defects.

The findings suggest that a collaborative approach between clinicians and AI could improve diagnostic accuracy in fetal ultrasound, supporting potential clinical utility. However, addressing challenges such as AI trust calibration and the impact of incorrect AI advice is crucial for safe integration into clinical workflows. Further research is needed to refine AI models and optimise their implementation in diverse clinical settings.

**Source & Image Credit:** [Ultrasound in Obstetrics and Gynecology](#)

Published on : Thu, 4 Jul 2024