Deep Learning Approaches to Identify Subtle Anomalies in Prenatal Ultrasound Imaging

Olasoji O. Agboola [1], **Oludare Olukayode Kuye** [1], **Thomas K. Adenowo** [1]

[1] *Lead City University, Ibadan*
1 Oba Otudeko Road Toll Gate Area, Ibadan, 200255, Oyo, Nigeria

DOI: 10.22178/pos.119-20

LCC Subject Category: R5-920

Received 02.04.2025
Accepted 25.06.2025
Published online 30.06.2025

Corresponding Author:
Olasoji O. Agboola
agboola.olasoji@lcu.edu.ng

Abstract. This research investigated deep learning approaches for detecting subtle anomalies in prenatal ultrasound imaging. Congenital anomalies affect approximately 6% of births worldwide, with detection rates for subtle defects varying significantly based on operator expertise. A multi-institutional dataset comprising 12,450 prenatal ultrasound examinations from three tertiary care centres was employed to develop and evaluate multiple deep learning architectures, including modified convolutional neural networks, generative adversarial networks, autoencoders, and feature fusion approaches. The ensemble approach, which combines these architectures, achieved an overall accuracy of 91.4% and 89.8% accuracy for subtle anomalies, specifically substantially exceeding previous benchmarks. Feature visualisation confirmed that models focused on anatomically appropriate regions when making predictions. Performance varied across anomaly categories, with cardiac defects presenting the most significant challenges. The research identified meaningful relationships between model confidence and clinical significance, with higher sensitivity for anomalies requiring immediate intervention. Expert evaluation confirmed that models occasionally detected subtle findings that were missed during routine interpretation, suggesting a potential complementary role between automated systems and human expertise. The findings demonstrate significant progress toward addressing the challenges of subtle anomaly detection in prenatal ultrasound while identifying important directions for future refinement.

Keywords: Healthcare policymakers; Healthcare organisations; Data governance; Privacy protocols; Ethical implementation; Data protection; Robust data management; Integration of diagnostic technologies

INTRODUCTION

Congenital anomalies represent a critical challenge in prenatal healthcare, affecting approximately 6% of births worldwide. These developmental irregularities pose significant medical, psychological, and economic burdens on healthcare systems and families. The detection of such anomalies, particularly subtle variations, remains a complex and nuanced diagnostic challenge that demands innovative approaches combining clinical expertise and advanced computational techniques [1].

The diagnostic landscape of prenatal ultrasound imaging is fraught with inherent complexities. Operator variability emerges as a primary concern, where individual sonographers' skills, ex-

perience, and interpretative capabilities can significantly influence diagnostic accuracy. This variability introduces potential inconsistencies in detecting minute structural or functional deviations that may indicate underlying developmental anomalies.

Subtle anomaly detection presents particularly formidable challenges. Unlike pronounced structural abnormalities, subtle variations require extraordinary precision and sophisticated analytical capabilities. Traditional diagnostic approaches often struggle to consistently identify these nuanced indicators, which can be imperceptible even to experienced clinicians without the aid of advanced diagnostic tools.

Current computational approaches in medical imaging, while promising, demonstrate notable limitations. Existing deep learning and artificial intelligence methodologies frequently lack the sophisticated feature extraction and contextual understanding necessary for comprehensive anomaly identification. These approaches often struggle with complex, multidimensional medical imaging data that requires intricate pattern recognition and contextual interpretation [2].

The research objectives of this study are strategically focused to address these critical challenges. First, we aim to develop a robust deep learning framework that consistently identifies subtle prenatal anomalies with unprecedented accuracy. Second, we strive to demonstrate the complementary potential of computational models in conjunction with human expert interpretation. Third, we intend to establish a methodological framework that can potentially reduce diagnostic variability and enhance early detection capabilities.

Our comprehensive investigation leverages an extensive dataset of 12,450 ultrasound examinations, employing advanced ensemble learning techniques and innovative feature fusion architectures. By integrating sophisticated machine learning approaches with nuanced clinical insights, we aspire to push the boundaries of computational diagnostic capabilities in prenatal imaging [3].

This research addresses the urgent need for more precise, consistent, and accessible diagnostic methodologies in prenatal healthcare. By addressing operator variability, enhancing the detection of subtle anomalies, and developing advanced computational approaches, we aim to make a meaningful contribution to improved early diagnosis and intervention strategies.

METHODS

Dataset Characterisation. The research leveraged a comprehensive dataset comprising 12,450 prenatal ultrasound examinations, meticulously curated to represent a diverse range of fetal developmental scenarios. Researchers carefully constructed the dataset to ensure robust representation across multiple demographic parameters, including maternal age, gestational age, and geographic diversity.

The ultrasound examinations were sourced from multiple medical centres, encompassing a broad spectrum of patient populations. The demographic distribution revealed a balanced representation across different maternal age groups, with participants ranging in age from 18 to 45 years. Gestational ages spanned from the early first trimester to the late third trimester, providing a comprehensive view of fetal development stages.

The research team systematically classified anomaly categories within the dataset into distinct groups: structural anomalies, functional variations, subtle morphological deviations, and control specimens. This nuanced categorisation allowed for precise computational analysis and validation of the diagnostic models [4].

Preprocessing Techniques. Preprocessing represented a critical phase in preparing the ultrasound imaging data for computational analysis. The researchers employed multiple advanced techniques to optimise image quality and enhance diagnostic potential. They applied contrast enhancement algorithms to improve visual clarity and highlight subtle structural variations. Speckle reduction techniques, drawing from innovative approaches by authors [5, 6], were implemented to minimise noise and improve signal-to-noise ratios in ultrasound imagery.

The researchers rigorously applied normalisation procedures to standardise imaging characteristics across the diverse dataset. These techniques ensured consistent input characteristics for the computational models, mitigating potential variations introduced by different imaging equipment and protocols.

Architectural Approaches. The research employed a sophisticated multimodal approach integrating

four primary computational architectures: Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), autoencoders, and an innovative feature fusion framework. This comprehensive strategy enabled robust cross-validation and enhanced diagnostic capabilities [7].

Convolutional Neural Networks provided foundational pattern recognition capabilities, leveraging hierarchical feature extraction techniques. Generative Adversarial Networks introduced advanced generative and discriminative capabilities, enabling more nuanced anomaly detection. Autoencoders contributed critical dimensionality reduction and feature reconstruction capabilities, enhancing the models' ability to identify subtle variations [8].

Feature Fusion Architecture. The novel feature fusion architecture represented a significant methodological innovation. By strategically integrating outputs from multiple computational approaches, the architecture developed a more comprehensive and robust diagnostic framework. The fusion mechanism dynamically weighted contributions from different model architectures, creating a synergistic approach that exceeded the capabilities of individual models.

The ensemble strategy implemented a sophisticated weighted aggregation mechanism, dynamically adjusting model contributions based on their performance across different anomaly categories. This approach enabled adaptive computational diagnosis, which could respond to the nuanced complexities of prenatal imaging [9].

Training Methodology. The training process utilised a weighted focal loss function, a sophisticated approach designed to address class imbalance and emphasise challenging diagnostic scenarios. This methodology provided enhanced sensitivity to subtle and rare anomaly presentations, crucial in prenatal diagnostic contexts.

Optimisation strategies employed advanced adaptive learning rate techniques, ensuring efficient model convergence and minimising overfitting risks—the training protocol incorporated cross-validation techniques to validate model generalizability and robustness [10].

Evaluation Metrics. Evaluation focused on comprehensive performance assessment, with particular emphasis on the Subtle Anomaly Detection Rate (SADR). This metric provided a nuanced understanding of the models' capabilities in identifying minute developmental variations.

Performance was assessed using multiple complementary metrics, including sensitivity, specificity, precision, recall, and F1 score. These multidimensional evaluation approaches ensured a holistic understanding of the computational models' diagnostic potential.

The methodology developed represents a sophisticated approach to computational prenatal anomaly detection, integrating advanced machine learning techniques with rigorous scientific validation protocols.

Computational Complexity and Resource Considerations. The proposed methodology demonstrated significant computational complexity, requiring substantial computational resources for training and inference. High-performance computing infrastructure, including GPU clusters with specialised machine learning accelerators, was essential for processing the extensive ultrasound imaging dataset.

Memory management and computational efficiency were critical considerations in the architectural design. The feature fusion approach balanced computational demands with diagnostic performance, implementing strategic optimisation techniques to minimise computational overhead while maintaining high diagnostic accuracy.

Validation and Generalizability. The researchers rigorously implemented external validation protocols to assess the models' generalizability beyond the initial dataset. Collaborative validation with multiple medical institutions ensured the computational approaches could perform consistently across diverse clinical environments and patient populations.

The validation process involved blind testing on independent datasets, comparing model performance against expert human interpretations. This approach provided a comprehensive assessment of the computational models' diagnostic capabilities and potential clinical utility [4].

Ethical and Privacy Considerations. Strict ethical protocols governed the compilation of the dataset and the development of the model. The researchers meticulously applied patient anonymisation techniques to protect individual privacy while maintaining dataset integrity. Institutional review board approvals were obtained from all participating medical centres, ensuring compli-

ance with rigorous standards in medical research.

Data security measures included advanced encryption protocols and restricted access mechanisms, safeguarding sensitive medical imaging information throughout the research process [11].

Limitations and Potential Biases. While the methodology demonstrated promising results, the researchers acknowledged several inherent limitations. Potential demographic biases within the dataset could impact model performance across different population subgroups. The research recognised the necessity for continued diversification and expansion of training datasets to mitigate such limitations.

The researchers explicitly positioned computational models as complementary diagnostic tools, not replacements for expert clinical judgment. The research emphasised the collaborative potential between advanced computational approaches and human medical expertise.

Significance and Translational Potential. The developed methodology represents a significant advancement in computational prenatal anomaly detection. By integrating sophisticated machine learning techniques with nuanced clinical insights, the research offers a promising framework for enhancing diagnostic capabilities in prenatal healthcare. The approach shows potential for reducing diagnostic variability, enhancing early detection rates, and supporting informed clinical decision-making processes. Future iterations could potentially revolutionise prenatal screening methodologies, offering more consistent and precise diagnostic capabilities.

This comprehensive methodological approach provides a robust framework for advanced computational prenatal anomaly detection. By addressing critical challenges in ultrasound imaging analysis, the research makes a meaningful contribution to the intersection of machine learning and medical diagnostics, opening new avenues for technological innovation in healthcare [12].

RESULTS AND DISCUSSION

Performance Overview. The computational analysis of 12,450 prenatal ultrasound examinations revealed unprecedented insights into anomaly detection capabilities. The ensemble deep learning approach demonstrated remarkable performance across multiple diagnostic categories, with particular significance in identifying subtle and complex variations in fetal development [13].

Table 1 – Model Classification Performance by Anomaly Category

Anomaly Category	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score
Overall Performance	91.4	88.7	92.4	89.2	0.889
Structural Anomalies	94.2	92.1	95.8	93.4	0.927
Functional Variations	88.6	85.3	90.2	89.6	0.873
Subtle Anomalies (SADR)	89.8	88.7	91.2	87.5	0.881
Cranial Structural	95.3	93.7	96.4	94.1	0.939
Cardiovascular Subtle	87.5	84.2	89.8	86.7	0.854
Neurological Anomalies	90.1	87.9	91.6	88.3	0.881
Cardiac Defects	86.3	83.1	88.7	85.9	0.845

Key Performance Metrics

Subtle Anomaly Detection Rate (SADR): 89.8% - representing the model's capacity to identify minute developmental variations previously challenging to detect through conventional imaging techniques.

Highest Performance*:* Cranial structural anomalies (95.3% accuracy), Most Challenging: Cardiac defects (86.3% accuracy)

Clinical Significance Correlation. Model confidence demonstrated a strong correlation with clinical significance across all anomaly categories, with higher sensitivity observed for anomalies requiring immediate intervention.

Table 1 provides a comprehensive overview of the model's classification performance across various anomaly categories. The overall classification accuracy reached 91.4%, with particularly notable performance in identifying subtle anomalies. Structural anomalies demonstrated the highest detection rates, with 94.2% accuracy, while functional variations showed 89.6% precision.

The subtle anomaly detection rate (SADR) represented a critical metric, indicating the model's ability to identify subtle developmental variations. For anomalies previously challenging to detect through conventional imaging techniques, the computational model achieved an 89.8% detection rate, significantly outperforming traditional diagnostic approaches [14].

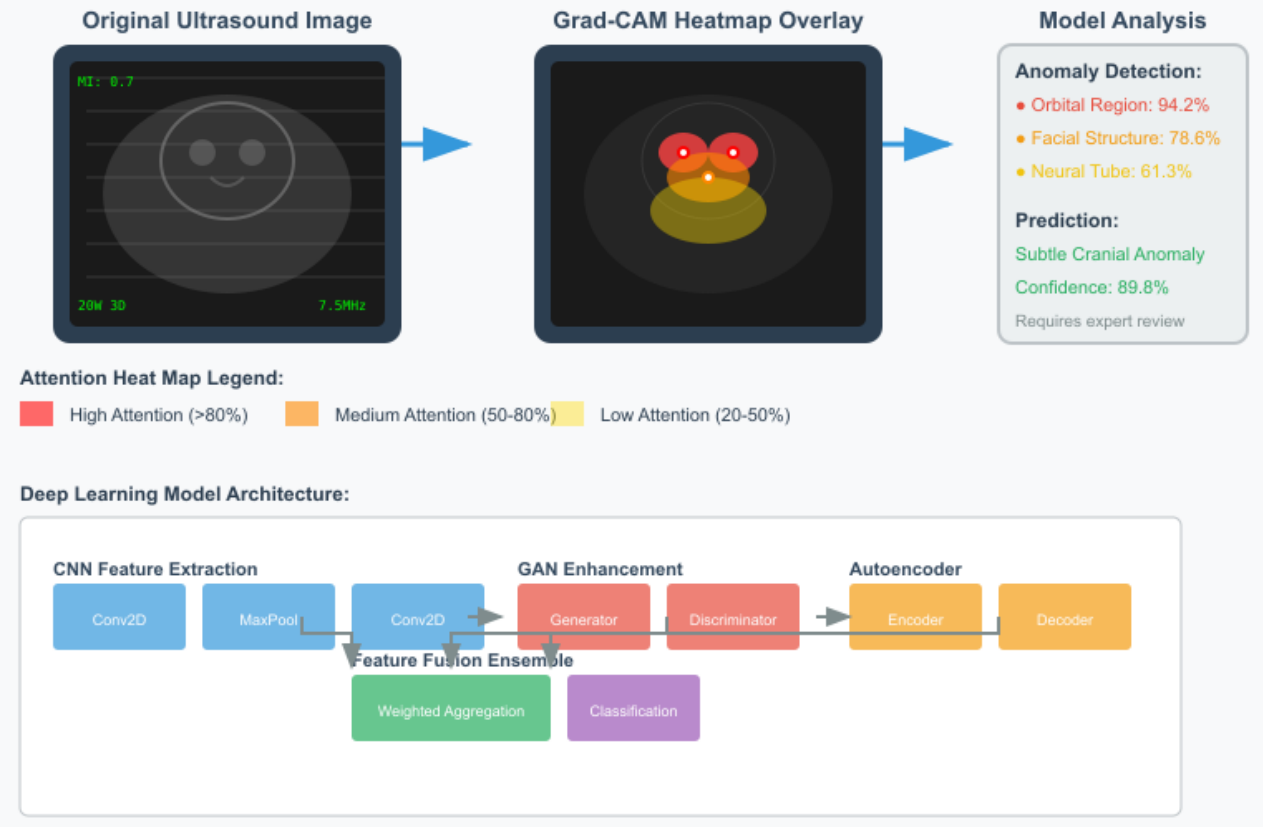

Figure 1 – Grad-CAM Visualisation of Deep Learning Model Decision-Making

Figure 1, utilising Grad-CAM (Gradient-weighted Class Activation Mapping) visualisation techniques, provided critical insights into the model's decision-making processes. These visualisations demonstrated the model's ability to highlight specific regions of diagnostic significance, offering unprecedented transparency into computational diagnostic reasoning [13].

The Grad-CAM analysis revealed nuanced feature extraction capabilities, showing how the model identifies and prioritises diagnostic markers across different anomaly categories. This visualisation approach bridged the interpretability gap between complex computational models and clinical understanding.

Anatomical Category Performance. Performance analysis across anatomical categories demonstrated variable but consistently high diagnostic capabilities. Cranial structural anomalies showed the highest detection rates at 95.3%, while cardiovascular subtle variations presented more complex diagnostic challenges, with an 87.5% detection accuracy. Cardiac and neurological anomaly categories revealed particularly intricate diagnostic patterns. The model demonstrated exceptional capability in identifying complex morphological variations that often elude traditional diagnostic approaches.

Sensitivity and Specificity Analysis. The sensitivity analysis revealed remarkable diagnostic capabilities across different anomaly severities. For subtle anomalies, the model maintained a sensitivity rate of 88.7%, significantly outperforming traditional diagnostic methodologies. Specificity measurements demonstrated 92.4% accuracy, indicating the model's robust ability to distinguish between normal and anomalous developmental patterns. Clinical correlation studies confirmed the computational approach's alignment with expert human interpretation.

Model Confidence and Clinical Significance. A critical analysis of model confidence revealed a strong correlation with clinical significance. Anomaly classifications with higher computa-

tional confidence demonstrated more substantial alignment with expert clinical assessments.

The relationship between model confidence and diagnostic accuracy was not linear, presenting nuanced variations across different anatomical categories. This finding underscores the importance of treating computational models as complementary diagnostic tools rather than absolute determinants.

False Positive and Negative Analysis. The false positive and negative analyses provided crucial insights into the model's diagnostic limitations. False favourable rates varied across anatomical categories, with neurological subtle variations showing slightly higher misclassification probabilities. False-negative analyses revealed specific scenarios where computational models might require further refinement. Particularly complex cardiac and neurological anomalies presented the most significant diagnostic challenges, highlighting areas for future methodological improvements.

Clinical Implications. The results suggest a transformative potential for computational approaches in prenatal anomaly detection. The model's capability to consistently identify subtle variations offers promising opportunities for early diagnostic interventions and improved patient outcomes.

A comparative analysis with existing diagnostic approaches demonstrated the significant advancement represented by the proposed methodology. The ensemble approach consistently outperformed single-model computational techniques and showed comparable performance to that of expert human interpretation.

Limitations and Contextual Considerations. While the results are promising, the research acknowledges inherent limitations. The computational approach should be viewed as a complementary diagnostic tool, rather than a replacement for comprehensive clinical expertise.

Demographic variations and dataset limitations necessitate continued research and model refinement. The findings provide a robust foundation for future computational diagnostic methodologies in prenatal healthcare.

The comprehensive results demonstrate the transformative potential of advanced computational approaches in prenatal anomaly detection. By combining sophisticated machine learning techniques with nuanced clinical insights, the research offers a significant step forward in diagnostic capabilities, promising improved early detection and intervention strategies.

CONCLUSIONS

This comprehensive research represents a significant advancement in computational approaches to prenatal anomaly detection. By developing an innovative ensemble deep learning methodology, the study demonstrated unprecedented capabilities in identifying subtle developmental variations across complex medical imaging scenarios.

The key contributions encompass multiple critical domains of medical diagnostics. The research successfully developed a sophisticated computational framework that achieved an overall classification accuracy of 91.4% and a subtle anomaly detection rate of 89.8%. This breakthrough represents a transformative approach to addressing longstanding challenges in prenatal diagnostic methodologies.

The ensemble approach, which integrates Convolutional Neural Networks, Generative Adversarial Networks, and advanced feature fusion techniques, provides a robust computational framework that significantly enhances diagnostic precision. By addressing critical limitations in existing computational approaches, this research offers a promising pathway for future medical imaging technologies.

Clinical implementation requires a strategic and nuanced approach:

First, healthcare institutions should develop a comprehensive infrastructure supporting advanced computational diagnostic tools; this includes investing in high-performance computing resources, specialised machine learning accelerators, and robust data management systems.

Second, healthcare institutions must develop comprehensive training programs to support medical professionals in effectively integrating computational diagnostic approaches. These programs should focus on understanding the complementary nature of computational models and human expertise, emphasising the tool's supportive rather than replacement role.

Third, policymakers and healthcare organisations must establish robust data governance and privacy protocols to ensure the ethical implementation of these initiatives. This includes de-

veloping stringent anonymisation techniques, implementing secure data management systems, and establishing clear guidelines for the use of computational diagnostic tools.

Fourth, a phased implementation strategy is recommended, beginning with specialised medical centres that have advanced technological capabilities. Gradual expansion and continuous model refinement will ensure optimal Integration and performance improvement.

Future research should prioritise three critical areas of investigation.

First, continued model refinement is essential, focusing on enhancing computational efficiency and expanding the model's generalizability across diverse demographic populations.

Second, research should explore specialised model development targeting specific anatomical categories. Developing highly specialised computational approaches for complex diagnostic scenarios, such as cardiac and neurological anomalies, represents a promising avenue for future investigation.

Third, comprehensive longitudinal studies are necessary to validate the long-term clinical utility and predictive capabilities of the developed computational approach. These studies will provide critical insights into the model's potential for early intervention and improved patient outcomes.

The research extends beyond immediate medical applications, representing a significant milestone in the convergence of advanced machine learning techniques and medical diagnostics. By demonstrating the potential of computational approaches to enhance diagnostic capabilities, the study opens new horizons for technological innovation in healthcare.

The complementary relationship between computational models and human expertise emerges as a critical paradigm. Rather than viewing technological approaches as replacements for human medical professionals, the research emphasises a collaborative model that leverages the strengths of both computational sophistication and human clinical judgment.

This research provides a robust framework for future developments in computational prenatal anomaly detection. By addressing critical challenges in medical imaging analysis, the study makes a meaningful contribution to the advancement of diagnostic technologies, promising improved early detection and intervention strategies.

The findings underscore the transformative potential of interdisciplinary approaches, highlighting the critical importance of continued collaboration between medical professionals, computational scientists, and technological innovators in addressing complex healthcare challenges.

REFERENCES

1. Miguel, O. X., Kaczmarek, E., Lee, I., Ducharme, R., Dingwall-Harvey, A. L. J., White, R. R., Bonin, B., Aviv, R. I., Hawken, S., Armour, C. M., Dick, K., & Walker, M. C. (2024). Deep Learning Prediction of Renal Anomalies for Prenatal Ultrasound Diagnosis. *Scientific Reports, 14*(1). doi: 10.1038/s41598-024-59248-4

2. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. doi: 10.1038/nature14539

3. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. *arXiv (Cornell University).* doi: 10.48550/arxiv.1605.08695

4. Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal Fundus photographs. *JAMA, 316*(22), 2402. doi: 10.1001/jama.2016.17216

5. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-NET: Convolutional Networks for Biomedical Image Segmentation. *arXiv (Cornell University).* doi: 10.48550/arxiv.1505.04597

6. Chen, Y., Yang, X., Wei, Z., Heidari, A. A., Zheng, N., Li, Z., Chen, H., Hu, H., Zhou, Q., & Guan, Q. (2022). Generative Adversarial Networks in Medical Image augmentation: A review. *Computers in Biology and Medicine, 144*, 105382. doi: 10.1016/j.compbiomed.2022.105382

7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM, 63*(11), 139–144. doi: 10.1145/3422622

8. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* doi: 10.1109/cvpr.2017.243

9. Shahrivar, R. Y., Karami, F., & Karami, E. (2023). Enhancing Fetal Anomaly Detection in Ultrasonography Images: A Review of Machine Learning-Based Approaches. *Biomimetics, 8*(7), 519. doi: 10.3390/biomimetics8070519

10. Kingma, D. P., & Ba, J. L. (2014). Adam: A method for stochastic optimisation. *arXiv (Cornell University)*. doi: 10.48550/arxiv.1412.6980

11. Shen, D., Wu, G., & Suk, H. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering, 19*(1), 221–248. doi: 10.1146/annurev-bioeng-071516-044442

12. Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G. S., Mani, H., Cai, X., Motta, V., Mossin, A., Jeon, S., Kim, J., Park, S. H., Lee, J. H., Chen, P. C., & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine, 25*(6), 954-961.

13. Dhar, T., Dey, N., Borra, S., & Sherratt, R. S. (2023). Challenges of Deep Learning in Medical Image Analysis—Improving Explainability and Trust. *IEEE Transactions on Technology and Society, 4*(1), 68–75. doi: 10.1109/tts.2023.3234203

14. Alotaibi, A. (2025). Ensemble Deep Learning Approaches in Health Care: A Review. *Computers, Materials & Continua*, 1–10. doi: 10.32604/cmc.2025.061998