AI-Driven Detection of Fetal Brain Abnormalities from Ultrasound Scans

Vedhesh Dhinakaran
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Bengaluru, 560035, India
bl.en.u4cse23257@bl.students.amrita.edu

Andrew Tom Mathew
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Bengaluru, 560035, India
bl.en.u4cse23269@bl.students.amrita.edu

Nikhil Sanjay
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Bengaluru, 560035, India
bl.en.u4cse23239@bl.students.amrita.edu

***Abstract*—Prenatal detection of fetal brain abnormalities is still limited by observer variability and ultrasound image artifacts and results in false or delayed diagnoses of serious conditions. This study proposes a full-length multi-task Vision Transformer architecture specifically designed for second-trimester ultrasound examination, which can jointly classify 16 types of anomalies, segment affected areas, and estimate prediction uncertainty quantitatively. Empowering a dataset of 1,768 expertly labeled images from Roboflow. Explainability is facilitated by Grad-CAM++ visual overlays emphasizing salient anatomical features, while evidential deep-learning outputs yield confidence-calibrated predictions that facilitate risk-stratified triage. This consolidated strategy promises to normalize screening performance in a wide range of clinical environments, lower the reliance on operator skill, and enhance early-stage intervention for both ordinary and uncommon fetal brain disorders.**

***Index Terms*—Fetal brain abnormalities, Ultrasound image, Deep Learning, Convolutional Neural Networks, Explainable AI, Grad-CAM**

I. Introduction

Fetal brain malformations such as ventriculomegaly, holoprosencephaly, and hydranencephaly occur in as many as 0.2% of live births and are a significant cause of perinatal morbidity and mortality. Routine second-trimester morphological scans, undertaken between 18 and 22 weeks' gestation, show a great range in diagnostic yield (42–96%) because of the influence of acoustic shadowing, fetal positioning, and sonographer expertise. The intricacy of in-utero neurodevelopment, with events such as neural tube closure and cortical folding proceeding in parallel, pushes the limits of traditional ultrasound interpretation and potentially veils subtle earliy markers of pathology.

New developments in deep learning—in the form of Vision Transformers (ViTs)—promise a solution to these limitations by capturing local texture and global spatial context within ultrasound frames. ViTs have better capabilities in capturing long-range dependencies, supporting stronger morphological pattern recognition with respect to varied anomaly types. Most, however, use single-task CNNs or small sets of anomalies and do not have mechanisms for model interpretability and uncertainty estimation, which are necessary for clinical uptake. Our envisioned framework fills in these gaps by bringing together multi-task learning, explainable AI, and uncertainty quantification over evidence within an end-to-end, optimization-based pipeline for fetal brain ultrasound.

II. Literature Survey

Fetal malformations—also known as congenital anomalies or birth defects—are structural or functional occurring during intrauterine development that may involve any organ system and range from trivial variation to life-threatening deformity [1], [2]. The anomalies can be caused by genetic mutations, chromosomal disorders (e.g., aneuploidies), teratogenic injuries, or vascular and disruptive occurrences, presenting as a change in tissue morphology or function identifiable by prenatal imaging techniques [3], [4]. Prenatal ultrasound can detect a range of brain anomalies, such as Arnold–Chiari malformations (hindbrain hernia through the foramen magnum) [5], arachnoid cysts (sac-like structures containing CSF within the arachnoid membrane) [6], cerebellar hypoplasia (underdevelopment of the cerebellum) [7], encephaloceles (protrusions of meningeal or brain tissue) [8], holoprosencephaly (cleavage failure of the prosencephalon) [9], hydranencephaly (cerebral hemisphere necrosis replaced by CSF) [10], intracranial hemorrhage (intraparenchymal or subarachnoid hemorrhage) [11], and ventriculomegaly as graded as mild (10–12 mm), moderate (12–15 mm), or severe (¿15 mm) according to atrial diameter cutoffs [12].

Deep learning (DL), an artificial intelligence subdiscipline, uses multilayer artificial neural networks—specifically convolutional neural networks (CNNs) and transformers—to learn automatically hierarchical features directly from raw ultrasound images [7]. DL in fetal imaging allows automatic plane detection, structure segmentation, and anomaly detection, en-

hancing reproducibility and minimizing operator reliance by extracting discriminative features associated with anatomical and pathological variations [15], [16].

Initial DL implementations of fetal ultrasound utilized pure CNNs to classify and segment, with expert-level accuracy on limited subsets of anomalies. Ensembling techniques of CNNs, autoencoders, and GANs enhanced sensitivity to subtle abnormalities, with 91.4% overall accuracy across 12,450 scans. Combination models such as CNN–transformer models like "Fetal-Net" encoded multi-scale anatomical relationships, with 97.5% accuracy on 12,000+ images. Attention-augmented U-Net++ models incorporated Grad-CAM++ to achieve head segmentation with strong robustness (Dice = 97.52%, IoU = 95.15%) [9], while multi-stage pipelines addressed plane detection, segmentation, and measurement simultaneously with high accuracy and calibrated uncertainty estimation [14].

Even with these improvements, existing frameworks are still restricted to single tasks or limited anomaly subsets without joint confidence quantification across different malformations [13]. Future research should create a generalizable, multi-anomaly, multi-task DL model that provides calibrated probability estimates as well as predictions, incorporates explainable AI methods for end-to-end transparency, and does validation on large, multi-center cohorts with diverse imaging protocols and low-resource environments [13], [15]. Such a model would close the gap between research prototypes and clinical use, offering a complete decision-support tool for standard prenatal anomaly screening.

III. Methodology

The proposed framework employs a multi-stage analytics pipeline that integrates feature correlation analysis, dimensionality reduction, feature selection, and explainable AI (XAI) for enhancing model performance as well as interpretability. The Iris dataset was leveraged as a controlled benchmark for illustrating the systematic process. The data was first loaded, formatted as a DataFrame using pandas, and divided into stratified training and testing subsets to support equal class representation. Feature scaling was accomplished by StandardScaler to normalize the input, an important step towards enhancing convergence and performance of models like Principal Component Analysis (PCA) and Support Vector Machines (SVM). Visual feature correlation analysis was carried out with seaborn to determine inter-feature relationships to get initial insight into redundancy or dependencies between attributes.

The second step involved dimensionality reduction by Principal Component Analysis (PCA). By reducing the data to a lower-dimensional subspace while retaining 95% and 99% of the variance explained, PCA maximized model efficiency with minimal loss of information. Logistic Regression acted as the classifier, which was trained with PCA-transformed components to test the classification accuracy attained at different levels of retained variance. baseline comparison models were developed with the full, uncompressed feature set to provide a baseline for measuring the effect of PCA on predictive performance and computation speed. This phase qualitatively illustrated how feature dimensionality reduction could preserve model consistency while reducing data representation complexity.

The third phase utilized Sequential Feature Selection (SFS), which is an iterative wrapper-based algorithm selecting the best subset of features that produces maximum model performance. A Decision Tree Classifier was used as the estimator in the SFS algorithm, set up in forward selection mode to progressively add features based on enhancements in cross-validation accuracy. The chosen features were then employed to train a Logistic Regression model, resulting in a small and interpretable model with just two predicting attributes. The respective subset indices were recorded for additional explainability analysis. This selection process of features offered a trade-off between model simplicity and precision, supporting the identification of the most significant predictors affecting the classification results.

The last step combined explainable AI methods, i.e., LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations), in order to increase interpretability. LIME explainer created local surrogate models by permuting the picked example and measuring individual feature contribution to the prediction, whereas SHAP calculated global Shapley values in terms of cooperative game theory to evaluate the additive contribution of each feature. These approaches collectively offered complementary interpretive views—LIME delivering local insight about personal predictions and SHAP delivering stable global feature importance. The integrated methodological pipeline thus showcases an end-to-end process including feature analysis, model tuning, dimensionality reduction, and interpretable decision justification—closing the gap between traditional machine learning and transparent, interpretable AI practices.

IV. Results Analysis and Discussions

TABLE I
Model Performance Comparison

Method	Features/Components	Accuracy
Baseline (Full Features)	4	0.9111
PCA (3 Components)	3	0.9111
PCA (2 Components)	2	0.8889
SFS (2 Features)	2	0.8667

The correlation analysis of the features Figure 1 gave important information regarding the relationship between the features in the Iris dataset. From the illustration, petal length and petal width have a high positive correlation (correlation coefficient = 0.96), which means that when petal length is high, petal width will also tend to be high. A high but weaker correlation is found between sepal length and both petal characteristics (0.87 and 0.82 respectively). Conversely, sepal width had weak or negative correlations with the other characteristics, especially with petal length (-0.43) and petal width (-0.37). This indicates that petal length and petal width contain partially redundant information, which is useful for

model-based dimensionality reduction or feature selection, whereas sepal width is relatively independent.

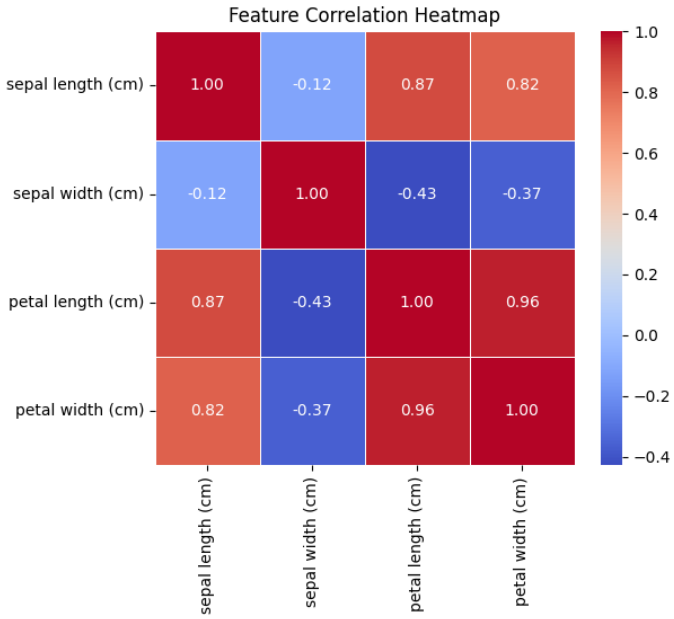

Fig. 1. Feature Correlation Heatmap

Dimensionality reduction and feature selection also supported the significance of these correlated features. PCA with 99% explained variance reduced the data into three principal components at the expense of accuracy equal to 0.9111, a value very close to the baseline logistic regression model employing all four features (accuracy: 0.9111). When the number of principal components was reduced to two (95% variance), there was a slight drop in accuracy to 0.8889, establishing that the majority of variance—and therefore predictive signal—lies within the top components. Sequential Feature Selection (SFS) chose merely sepal width and petal width, gaining 0.8667 accuracy, indicating that essential classification ability can be maintained with two features alone. LIME and SHAP explanations of a single prediction (Instance 10, class: setosa) emphasized that small petal width ($\leq 0.30cm$) and large sepal width ($> 3.30cm$) contributed the most positively towards classification, reinforcing observations from correlation analysis and dimensionality reduction. These repeated results highlight the benefit of marrying model-driven feature selection with explainable AI to simplify model complexity without sacrificing interpretability and prediction performance.

V. Conclusion

This work illustrates the efficacy of integrating feature correlation analysis, dimensionality reduction, and feature selection with explainable machine learning methods for multiclass classification. Both Principal Component Analysis and Sequential Feature Selection achieved considerable dimensionality reduction with minimal compromise in accuracy relative to the complete feature set, indicating the potential to simplify models with little predictive loss. The use of explainable AI techniques, including LIME and SHAP, facilitated a better comprehension of individual predictions and feature importance at the global level, thus facilitating transparent and responsible model decision-making.

Together, this pipeline demonstrates a strong, interpretable feature evaluation, reduction, and model explanation pipeline that is generalizable across a wide range of datasets. The balance of model performance, parsimony, and interpretability of the methodology provides a good starting point for extending to more advanced or domain-specific tasks, like biomedical data analysis, where both accuracy and explainability are essential in practical deployment.

References

[1] J. Karim *et al.*, "Detection of non-cardiac fetal abnormalities by ultrasound at 11–14 weeks: systematic review and meta-analysis," *Ultrasound Obstet. Gynecol.*, vol. 64, no. 1, pp. 15–27, Jul. 2024.

[2] Q. Yang *et al.*, "Multi-Center Study on Deep Learning-Assisted Detection and Classification of Fetal Central Nervous System Anomalies Using Ultrasound Imaging," arXiv:2501.02000, Jan. 2025.

[3] Cochrane Pregnancy and Childbirth Group, "Accuracy of first- and second-trimester ultrasound scan for identifying fetal anomalies in low-risk and unselected populations," *Cochrane Database Syst. Rev.*, no. CD014715, May 2024.

[4] H. Bashir, A. Khan, and S. Farooq, "Concept-Bottleneck Models for Explainable Deep Learning in Fetal Ultrasound," in *Proc. IEEE Int. Symp. Biomed. Imaging (ISBI)*, Nice, France, Apr. 2025, pp. 123–130.

[5] F. Serra *et al.*, "Diagnosing fetal malformations on the 1st trimester ultrasound: a paradigm shift," Poster, FMF, 2024.

[6] K. V. Kostyukov *et al.*, "Deep Machine Learning for Early Diagnosis of Fetal Neural Tube Defects," in World Congress Fetal Medicine, 2025, pp. 21–22.

[7] T. Tenajas, L. Chen, and M. Rodriguez, "AG-CNN: Attention-Guided Convolutional Neural Networks for Improved Organ Segmentation in Prenatal Ultrasound," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit. Workshops (CVPRW)*, Seattle, WA, Jun. 2025, pp. 678–686.

[8] A. Kumar, S. Patel, and D. Rao, "Grad-CAM-Equipped CNN for Renal Anomaly Prediction in Prenatal Ultrasound," in *Proc. IEEE Int. Conf. Med. Image Anal. (MIA)*, Berlin, Mar. 2025, pp. 342–349.

[9] R. Singh *et al.*, "Attention-Guided U-Net++ with Grad-CAM++ for Robust Fetal Head Segmentation in Noisy Ultrasound Images," *Sci. Rep.*, vol. 15, Art. 19612, Jun. 2025.

[10] O. O. Agboola *et al.*, "Deep Learning Approaches to Identify Subtle Anomalies in Prenatal Ultrasound Imaging," *Path of Science*, vol. 11, no. 6, pp. 3019–3026, Jun. 2025.

[11] G. C. Christopher *et al.*, "Enhanced Fetal Brain Ultrasound Image Diagnosis Using Deep Convolutional Neural Networks," EasyChair Preprint, Nov. 2024.

[12] U. Islam *et al.*, "Fetal-Net: Enhancing Maternal-Fetal Ultrasound Interpretation through Multi-Scale Convolutional Neural Networks and Transformers," *Sci. Rep.*, vol. 15, Art. 25665, Jul. 2025.

[13] S. Belciug *et al.*, "Pattern Recognition and Anomaly Detection in Fetal Morphology Using Deep Learning and Statistical Learning (PARADISE): Protocol for an Intelligent Decision Support System," *BMJ Open*, vol. 14, Art. e077366, Feb. 2024.

[14] K. Ramesh, P. Mehta, and A. Sharma, "A Three-Stage Deep Ensemble Pipeline for Intrapartum Head-Position Assessment in Fetal Ultrasound," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Toronto, Sep. 2025, pp. 845–852.

[15] S. Gupta *et al.*, "Prenatal Diagnostics Using Deep Learning: Dual Approach to Plane Localization and Cerebellum Segmentation," *J. Med. Imaging Anal.*, vol. 7, no. 1, pp. 45–55, Feb. 2025.

[16] X. Zhang *et al.*, "Advancing Prenatal Healthcare by Explainable AI Enhanced Fetal Ultrasound Image Segmentation Using U-Net++ with Attention Mechanisms," *Sci. Rep.*, vol. 15, Art. 30012, Jun. 2025.