

Automated Maternal–Fetal Health Analysis through Deep Neural Network Integration in Ultrasound Imaging

Shivanand S Gornale

Rani Channamma University, Belagavi

Priyanka kamat

`priyankakamat@rcub.ac.in`

Rani Channamma University, Belagavi

Rashmi Siddalingappa

York St John University

Kefang Li

Macao Polytechnic University

Khang Wen Goh

INTI International University

Research Article

Keywords: Ultrasound Fetal Medical Image, Parameter-based Image Classification, Trimester-based Image Classification, Deep Learning Techniques, Feature Extraction

Posted Date: April 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8461700/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Automated Maternal–Fetal Health Analysis through Deep Neural Network Integration in Ultrasound Imaging

Shivanand S. Gornale¹, Priyanka Kamat¹ (corresponding author), Rashmi Siddalingappa², Kefang Li³, and Khang Wen Goh⁴

¹Department of Computer Science, School of Mathematics and Computing Sciences, Rani Channamma University, Belagavi, Karnataka, India;

²York St John University, London campus, Clove crescent, E14 2BA, UK.

³Faculty of Applied Sciences, Macao Polytechnic University, R. de Luís Gonzaga Gomes, Macao, China.

⁴Faculty of Data Science and Information Technology, INTI International University, 71800 Nilai, Malaysia.

E-mails: ¹shivanand1971@rcub.ac.in, ¹priyankakamat@rcub.ac.in, ²r.siddalingappa@yorksj.ac.uk, ³kefengl@mpu.edu.mo, ⁴khangwen.goh@newinti.edu.my.

Abstract

Ultrasonography is a non-invasive and widely adopted imaging modality for monitoring fetal development and detecting potential complications during pregnancy. It provides essential information on fetal biometric measurements, anatomical structures, fetal movements, and growth abnormalities. This study addresses two tasks: (i) fetal biometric parameter-based classification (four categories) and (ii) trimester-based classification (three categories). Two datasets are employed: the publicly available HC-18 dataset and a custom dataset containing Head Circumference (HC), Femur Length (FL), Abdominal Circumference (AC), and Crown-Rump Length (CRL). Biometric parameters served as features for parameter-based classification, while HC, FL, and HC-18 are used for trimester-based classification into First Trimester (FT), Second Trimester (ST), and Third Trimester (TT). Feature extraction is performed using pre-trained deep learning models (ResNet50, VGG16, VGG19, and Xception), and a feature fusion strategy is applied to combine their representations. Gradient-weighted Class Activation Mapping (Grad-CAM) is incorporated to identify regions of interest (ROI) and improve model interpretability. The framework is evaluated using standard performance metrics, including accuracy, precision, recall, and F1-score. The parameter-based classification achieved an accuracy of 99.65%, while the trimester-based classification achieved 96.24% for HC, 97.70% for FL, and 93.33% for HC-18. The experimental findings demonstrate that the proposed feature-fusion framework achieves robust and reliable classification performance for fetal ultrasound analysis. The integration of Grad-CAM further enhances interpretability, making the method a comprehensive solution for the automated analysis of two-dimensional fetal ultrasound images.

Keywords: *Ultrasound Fetal Medical Image, Parameter-based Image Classification, Trimester-based Image Classification, Deep Learning Techniques, Feature Extraction.*

1. Introduction

Ultrasound imaging is a cornerstone of obstetric care, offering non-invasive, real-time visualization of the fetus and maternal anatomy[1, 2]. It is routinely used to monitor fetal growth, estimate gestational age, measure critical biometric parameters (Head Circumference (HC), Femur Length (FL), Abdominal Circumference (AC), and Crown-Rump Length (CRL)), evaluate fetal weight, and detect structural or developmental abnormalities. Pregnancy is divided into three trimesters: first (0–13 weeks), second (14–26 weeks), and third (27–40 weeks), during which fetal growth is closely tracked to support clinical decision-making and ensure maternal and neonatal safety[3]. Despite substantial reductions in global maternal mortality due to improved healthcare, advanced diagnostics, and increased prenatal care access, significant disparities persist in low-resource regions, where shortages of trained personnel, limited infrastructure, and restricted access to diagnostic tools

contribute to disproportionately high maternal death rates[4, 5].

Recent advances in artificial intelligence (AI) and deep learning have demonstrated significant potential in medical image analysis, encompassing tasks such as segmentation, classification, and anomaly detection. In fetal ultrasound, most prior studies have focused on single-parameter segmentation, particularly HC, or on fetal plane classification, providing limited insight into comprehensive fetal growth patterns across trimesters[6–9]. To address these limitations, this study proposes (i) a fully automated deep learning framework for (i) parameter-based classification using HC, FL, AC, and CRL, and (ii) trimester-based classification based on HC and FL[10]. The system integrates feature extraction from four widely used pretrained CNNs, such as ResNet50, VGG16, VGG19, and Xception, through a feature fusion strategy to capture both global and fine-grained features[11, 12]. Interpretability is enhanced with Grad-CAM visualization, while extensive data augmentation improves generalization and robustness, providing a reliable tool for automated fetal assessment[13,

14]. The main contributions of this study are summarized as follows:

1.1 Key Contributions

- Created a custom dataset of 3,587 fetal ultrasound images across four biometric parameters (HC, FL, AC, and CRL).
- Developed parameter-based and trimester-specific classification frameworks for accurate fetal growth assessment.
- Implemented deep learning models (ResNet50, VGG16, VGG19, Xception) for robust feature extraction.
- Proposed a feature fusion strategy combining multiple pretrained architectures to enhance classification accuracy.

2. Related Work

This section presents a comprehensive literature review of prior research conducted on fetal ultrasound image analysis, fetal biometry estimation, segmentation, and classification using deep learning techniques.

Several studies have explored deep learning for fetal ultrasound analysis. Baumgartner et al.[1] proposed SonoNet, an optimized VGG16-based framework for real-time detection of 13 fetal standard views, achieving an F1-score of 0.798 and localization accuracy of 77.8%. Al-Razgan et al.[2] developed an attention-guided CNN (AG-CNN) for maternal-fetal plane recognition, outperforming DenseNet169, ResNet50, and VGG16. Ishikawa et al.[13] used CNNs for automated fetal position estimation, while Rauf et al. [7] introduced a dual CNN fusion framework with Bayesian optimization for early abnormality detection. Sivasubramanian et al.[9] proposed a lightweight attention-based EfficientNet achieving 96.25% Top-1 accuracy, and Ghabri et al.[15] applied transfer learning with InceptionResNetV2, DenseNet, and ResNet50, reaching 99.78% accuracy. Other notable works include Gornale et al. [5] for fetal biometric parameter classification using U-Net and XceptionNet, Rathika et al. [16] with hybrid feature selection and Fast RBFNN, Lo et al.'s [17] FOAC-Net, Hasan et al.'s [18] U-Net ensemble, and Liang et al.'s [19] SPRNet.

Despite these advances, most studies focus on segmentation or measurement estimation rather than parameter-based classification (HC, AC, FL, CRL) or trimester-wise classification (FT, ST, TT). Manual interpretation still predominates for trimester assessment [10]. To address this gap, the proposed framework uses pretrained ResNet50, VGG16, VGG19, and Xception, combining their features through feature-map-level fusion[20–22]. Grad-CAM is applied to highlight important regions, improving interpretability. This method enhances classification accuracy, robustness, and clinical trust in fetal health assessment.

3. Materials and Methods

This study proposes a deep learning framework for classifying fetal ultrasound images into parameter-based

categories (HC, FL, AC, and CRL) and trimester-based classes (FT, ST, and TT), focusing on the key biometric indicators most relevant to fetal growth assessment. To enhance the model's ability to extract clinically meaningful information, we adopted a two-stage attention-guided classification approach. Grad-CAM maps are first generated from a baseline CNN to highlight the most discriminative fetal regions. Rather than using Grad-CAM solely for explainability, these activation maps are treated as attention-enhanced inputs, effectively suppressing noise and irrelevant background artifacts. This attention-refined representation enables the classifier to focus on essential anatomical structures, thereby improving robustness and overall prediction performance.

3.1 Dataset Description

Fetal biometric parameters are critical for monitoring growth and detecting potential complications. Existing datasets lack focused coverage of these parameters. To address this, a custom dataset was created, comprising images for HC, FL, AC, and CRL, with HC and FL specifically used for trimester-based classification. This dataset supports research in fetal biometric analysis and enhances the training of deep learning models for accurate parameter- and trimester-specific predictions.

3.1.1 Grand Challenge HC18

The HC18 dataset contains 1,334 fetal head ultrasound images (800×540 pixels) collected from 551 pregnant women across the first, second, and third trimesters at Radboud University Medical Center, Netherlands. All images were acquired using Voluson E8 and Voluson 730 ultrasound systems, with pixel sizes ranging from 0.052 to 0.326 mm. The dataset is divided into 999 training and 335 testing samples, and HC measurements are provided only for the training set. Expert sonographers manually annotated the skull boundaries using ellipses, and all procedures were approved by the CMO Arnhem-Nijmegen Ethics Committee in accordance with the Declaration of Helsinki[23]. Despite its usefulness, HC18 has limitations: it includes only the HC parameter, trimester labels are not explicitly provided, and the complexity of fetal images makes trimester categorization challenging. To address these gaps, a larger custom dataset was developed, incorporating four biometric parameters and enabling trimester-wise classification for both HC and FL images[3, 10].

3.1.2 Custom Dataset

The custom dataset makes a significant contribution to the field by adding more cases to improve the diversity and complexity of medical images. This provides a strong foundation for thorough evaluation. The dataset includes fetal age in weeks, and the images are labelled according to the corresponding trimester based on this information. It contains a total of 3,587 images divided into several categories. There are 1,426 HC images, which include 148 from the FT, 380 from the ST, and 898 from the TT. Then, there are 1,409 FL images, with 116 from the FT, 395 from the ST, and 898 from the TT.

Additionally, there are 553 CRL images and 199 AC images. The images were acquired using the VOLUSON P6 Ultrasound Machine at Metgud Hospital - Advanced Laparoscopy Centre and IVF in Belagavi, Karnataka, India. Each image, captured in JPG format, has an original resolution of 640×480 pixels, which is later cropped to 300×300 pixels for processing. Figure 1 presents representative samples of images categorized by their respective parameters.

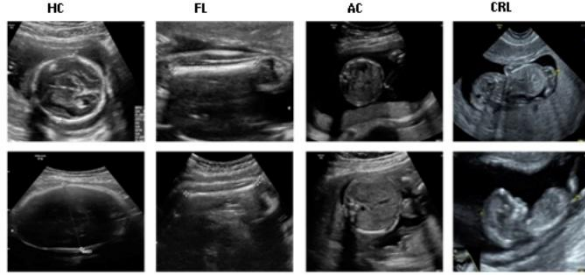


Figure 1. Samples of Parameter-wise Images (Custom Dataset)

3.2 Proposed Methodology

The proposed methodology classifies key fetal biometric parameters such as HC, FL, AC, and CRL. Also, it classifies the fetal trimester-based classification to enhance prenatal diagnostics. A pre-trained VGG16 CNN extracts features from fetal images and generates Grad-CAM visualizations to improve model transparency[13, 14]. Deep learning models, including VGG16, VGG19, ResNet50, and Xception, classify these images, with extracted features systematically stored for further analysis[24, 25]. A model fusion strategy integrates multiple feature representations to enhance classification accuracy and robustness. The proposed methodology is illustrated in Figure 2.

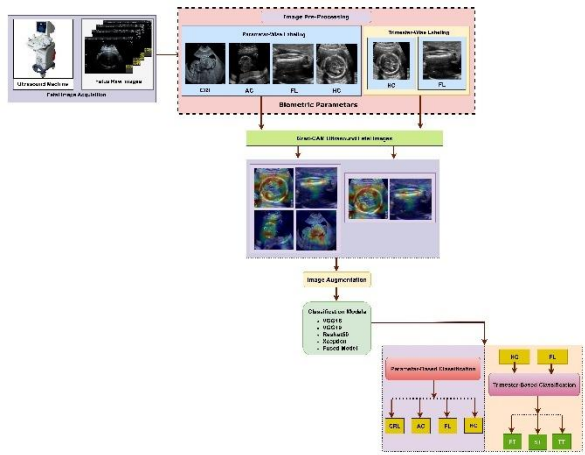


Figure 2. Block Diagram of Proposed Methodology

3.2.1 Fetal Image Acquisition

Ultrasound images are collected using the VOLUSON P6 Ultrasound Machine, a state-of-the-art imaging system known for its high resolution and precision in visualizing internal anatomical structures. The acquired images are exported in JPG format and systematically organized into well-structured folders, ensuring

efficient storage, easy access, and streamlined data management.

3.2.2 Pre-processing techniques

Preprocessing of fetal ultrasound images enhances quality, reduces artifacts, and highlights key anatomical features for deep learning. Images are organized by biometric parameters (HC, AC, FL, CRL) and trimester, then cropped to focus on the fetal head, removing irrelevant background. All images are resized to 300×300 pixels and converted to grayscale to simplify input and improve contrast [5, 26]. Intensity normalization and denoising standardize pixel values and suppress artifacts. These steps produce a consistent, high-quality dataset that improves model training, robustness, and reproducibility. Figure 3 illustrates the workflow, including intensity histograms showing enhanced contrast and edge clarity.

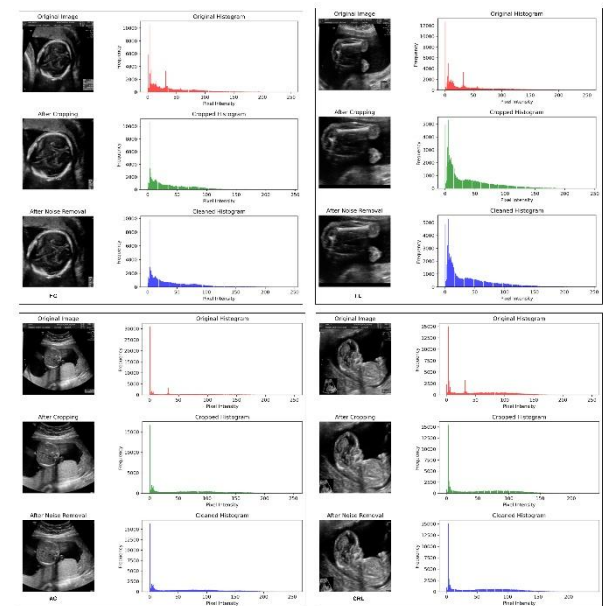


Figure 3. Preprocessing Representation of Fetal Biometric Parameters.

3.2.3 Grad-CAM

Interpretability is essential in deep learning models, particularly in medical image analysis, to establish trust and transparency in decision-making. Grad-CAM is a widely used technique for visualizing the regions of an image that most influence model predictions[13, 14]. The process involves three key stages. First, the input image is resized to 224×224 pixels, converted into a tensor, and preprocessed according to the requirements of VGG16. Second, gradients of the predicted class with respect to the final convolutional feature maps are computed, pooled, and combined to generate a class-specific heatmap. Finally, the heatmap is normalized and colorized with a jet colormap, then overlaid on the original image to highlight important regions. The resulting Grad-CAM outputs are saved for further analysis, as illustrated in Figure 4.

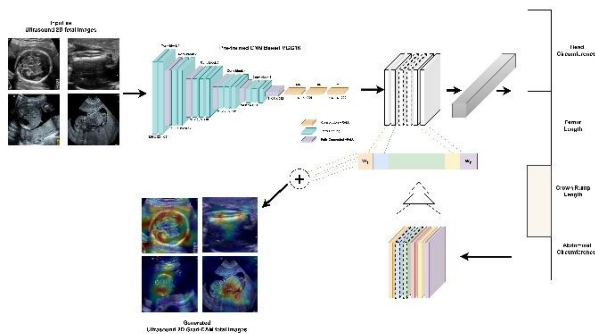


Figure 4. Gradient-Based Activation Map Diagram

Figure 5 shows Grad-CAM visualizations of fetal ultrasound images for HC, FL, AC, and CRL. Original grayscale images are on the left, with Grad-CAM overlays on the right highlighting regions of high attention; warmer colors indicate stronger focus. Pixel intensity histograms quantify attention distribution, showing that the model selectively concentrates on anatomically relevant regions[27]. These visualizations demonstrate how the model uses key features for accurate parameter- and trimester-based classification, providing interpretable and reliable insights into its predictions.

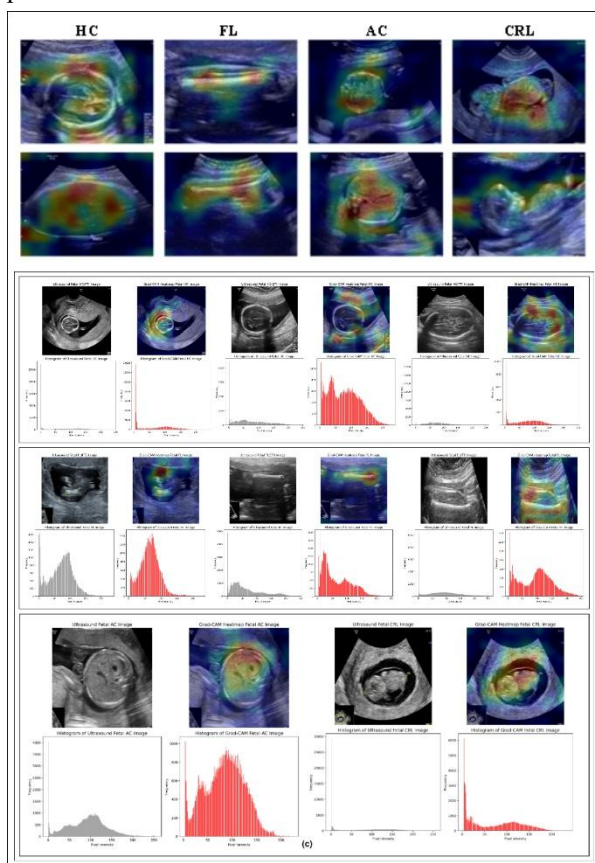


Figure 5. Samples of fetal ultrasound images with Grad-CAM activation maps and histogram analysis for all biometric parameters: (a) HC across three trimesters, (b) FL across three trimesters, and (c) AC and CRL with their corresponding visualizations.

3.2.4 Data Augmentation

The study implements a data augmentation pipeline to increase the diversity and size of fetal ultrasound image datasets. A total of 15 advanced augmentation

techniques is applied, encompassing geometric, photometric, and noise-based transformations to simulate real-world variations while preserving fetal anatomical structures. Geometric augmentations included rotation ($\pm 15^\circ$), scaling ($0.9\text{-}1.1\times$), translation ($\pm 10\%$ of image dimensions), shear ($\pm 10^\circ$), and horizontal and vertical flipping. Photometric adjustments involved modifications in contrast, hue, and saturation, while Gaussian noise and blurring are used to replicate common imaging artifacts. All images are normalized to the range 0-1 before being fed into deep learning models. These augmentation strategies are designed to enhance model generalization, prevent overfitting, and ensure methodological transparency and reproducibility in fetal medical image analysis[11]. The sample augmented images are depicted in Figure 6.

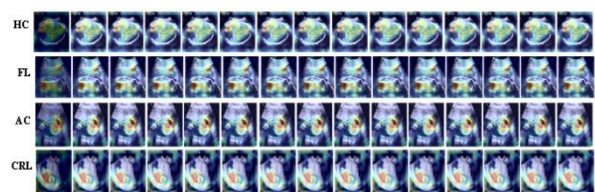


Figure 6. Samples of Augmented Grad-CAM images

3.3 Feature Extraction

Feature extraction is performed using ResNet50, Xception, VGG16, and VGG19, all pretrained on ImageNet, due to their ability to learn rich and discriminative features[20]. Extracted features are integrated into a proposed fused approach.

ResNet50: A 50-layer CNN with residual connections to mitigate vanishing gradients. Lower layers are frozen, and higher layers are fine-tuned for fetal-specific features with a Global Max Pooling layer, a 1024-neuron dense layer, and a softmax classifier[11, 28]. Misclassifications (e.g., AC as CRL/HC, FL as HC) arose from overlapping anatomy and attention to irrelevant regions. Figure 7 illustrates the model’s architecture and representative misclassifications.

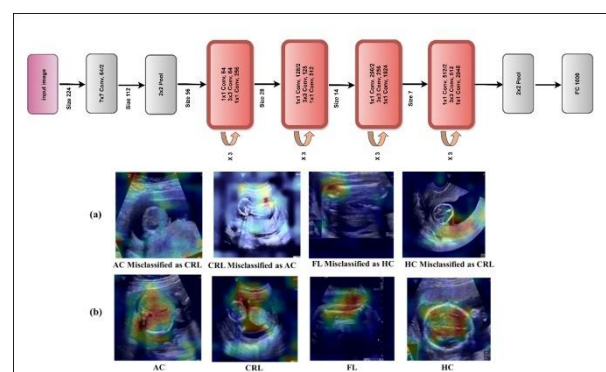


Figure 7. Architecture of the ResNet50 model with sample results: (a) misclassified images and (b) correctly classified images.

Xception: Uses depth wise separable convolutions for efficient feature extraction. Fine-tuned similarly, with global average pooling and dense layers[11]. Misclassifications occurred due to anatomical similarity

and texture overlap. Figure 8 shows the model's architecture and representative misclassifications.

VGG16 & VGG19: CNNs with 16- and 19-layers using 3×3 filters. Lower layers frozen, higher layers fine-tuned for fetal-specific patterns[20, 28]. Misclassifications (e.g., AC as HC, CRL as AC) arose from subtle intra-class similarity, class imbalance, and non-discriminative attention. VGG19 improved feature representation; ensemble approaches may further enhance robustness. Figures 9 and 10 demonstrate the model's architecture and representative misclassifications.

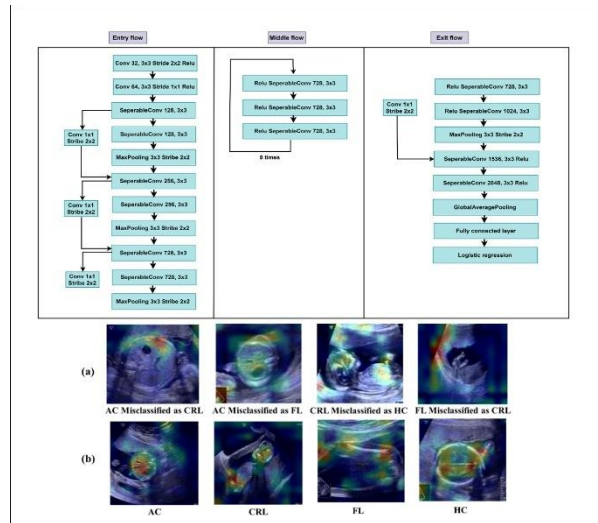


Figure 8. Architecture of the Xception model with sample results: (a) misclassified images and (b) correctly classified images.

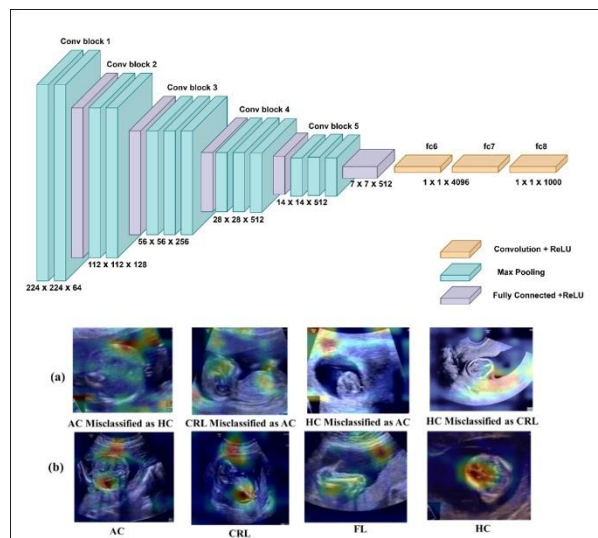


Figure 9. Architecture of the VGG16 model with sample results: (a) misclassified images and (b) correctly classified images.

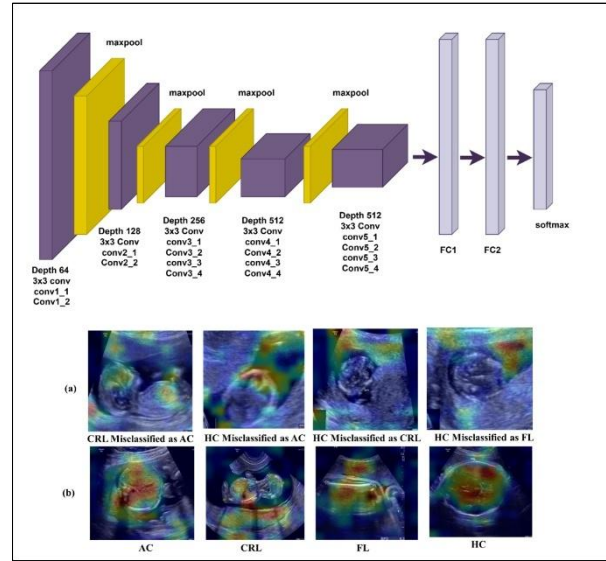


Figure 10. Architecture of the VGG19 model with sample results: (a) misclassified images and (b) correctly classified images.

Ensemble Framework: The ensemble combines VGG16, VGG19, Xception, and ResNet50 to leverage complementary strengths: VGG captures spatial hierarchies, Xception provides depthwise separable convolutions, and ResNet50 maintains stable gradient flow[29–31]. Features from the Dense1024 layers of all models are concatenated into a unified vector, passed through a dense classifier with dropout, batch normalization, and softmax. Figure 11 illustrates a block diagram for parameter-based classification, and Figure 12 outlines trimester-based classification.

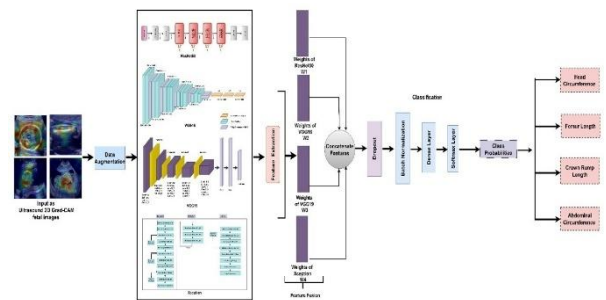


Figure 11. Fused Model Representation for Parameter-Based Classification

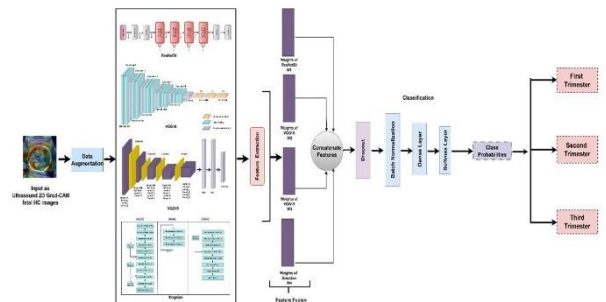


Figure 12. Fused Model Representation for Trimester-Based Classification

Let $X_i \in \mathbb{R}^{h \times w \times c}$ denote the i^{th} input ultrasound image, where h , w , and c represent the height, width, and number of channels, respectively. Each base CNN model $f_j(\cdot)$ extracts a feature map from X_i : and is given by equation (1).

$$F_i = f_i(X_i), \quad j = 1, 2, \dots, n \quad (1)$$

where n denotes the total number of base models in the ensemble ($n=4$).

The extracted feature maps F_j are concatenated to form a joint feature vector: and is defined by equation (2)

$$F_{concat} = Concat(F_1, F_2, \dots, F_n) \quad (2)$$

The concatenated feature vector is passed through a fully connected layer W followed by a non-linear activation function $\sigma(\cdot)$: and is given by equation (3).

$$Z = \sigma(W F_{concat} + b) \quad (3)$$

where b represents the bias term and σ is the ReLU activation function.

The final class probabilities are generated using a softmax activation function and are given by equation (4).

$$\hat{y} = Softmax(Z) \quad (4)$$

The network is trained using the categorical cross-entropy loss function to optimize classification performance and is defined by equation (5).

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^C y_{ik} \log(\hat{y}_{ik}) \quad (5)$$

where N denotes the total number of samples, C represents the number of classes, y_{ik} is the ground truth label for the i^{th} sample and k^{th} class ($y_{ik} = 1$ if sample i belongs to the class k , otherwise 0), and $\hat{y}_{ik}$ is the predicted probability for the class k obtained from the softmax output layer.

Optimization is performed using the Adam optimizer with a learning rate η : and is defined by equation (6)

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (6)$$

where $\hat{m}_t$ and $\hat{v}_t$ denote the bias-corrected first and second moment estimates, respectively.

This ensemble captures both fine-grained and high-level contextual patterns across fetal anatomical regions, improving robustness and classification accuracy while reducing overfitting[21, 30].

3.4.1 Grad-CAM Interpretation and Misclassification Analysis

The proposed methodology effectively classifies fetal parameters (AC, CRL, HC, FL) and trimesters (FT, ST, TT), but some misclassifications still occur, as shown in Figure 13. These errors arise mainly from overlapping anatomical appearances, low contrast, shadowing, occlusions, and variability in fetal positioning. For example, AC may appear similar to CRL, CRL may resemble FL, and HC may be misidentified when head boundaries are unclear. In trimester classification, subtle and gradual developmental changes can cause FT to be misclassified with ST or TT. Overall, variations in image quality, noise, and inherent anatomical similarities contribute to these classification challenges.

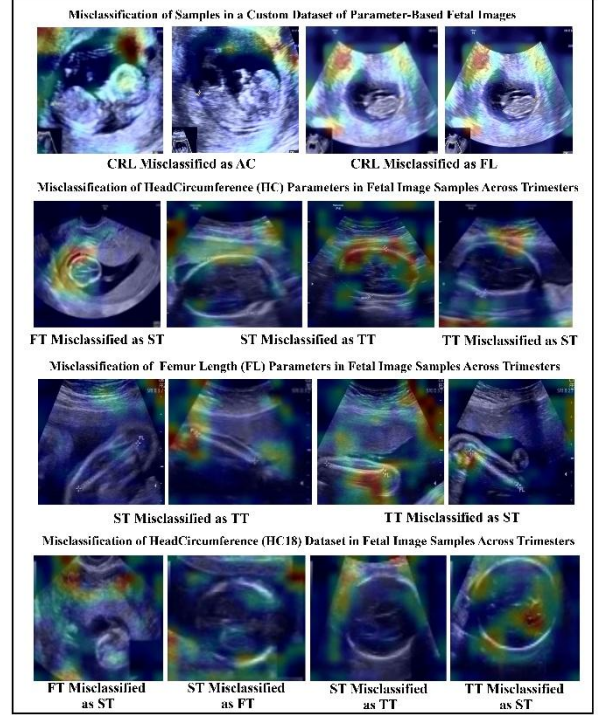


Figure 13. Misclassification of Sample Images Using the Proposed Method

Algorithm.

Parameter-based/trimester-based classification of fetus images using deep features.

Input: Fetal biometric parameter images (AC, CRL, HC, FL) / fetal trimester images focusing on HC (head circumference) and FL (femur length).

Output: Classification of fetal images into parameter-based (4-class) or trimester-based (3-class) categories.

Step 1: Design an image data pipeline that loads images from a specified directory and assigns labels based on the folder structure for systematic classification.

Step 2: Generate Grad-CAM images using the pre-trained VGG16 model to visualize the regions of the input images.

Step 3: Apply 15 different augmentation techniques to the generated Grad-CAM images to enhance their diversity and analyze the impact of augmentations on the visualized regions.

Step 4: The data is divided into training, validation, and test sets, with 70% of the images used for training, 15% for validation, and 15% for testing the network, ensuring this distribution is applied within each folder.

Step 5: Utilize feature extraction methods using VGG16, Xception, VGG19, and ResNet50 architectures.

Step 6: Feature fusion and classification.

Step 7: Analyze performance metrics to evaluate the test dataset.

Step 8: Tabularize the results.

END.

3.5 Implementation details

The complete model uses the Anaconda Navigator, a graphical user interface that accommodates various implementation platforms. Interactive computing notebooks on the web can be run through Jupyter Notebook (version 7.0.8). Python serves as the programming language for writing the code (<https://www.python.org/>). Libraries such as TensorFlow, Keras, sklearn, pandas, NumPy, and matplotlib are incorporated into the algorithm using the import statement.

3.6 Ablation Study

Table 1 presents the parameter-based ablation study conducted on VGG16, VGG19, ResNet50, Xception, and the fused model across four training configurations. Baseline training with original images yields moderate accuracy, whereas the inclusion of data augmentation provides consistent performance gains by increasing intra-class variability. The use of Grad-CAM-derived attention maps further enhances accuracy, indicating that spatially discriminative regions contribute to more effective feature learning. The highest accuracy (99.65%) is achieved when Grad-CAM images are combined with augmentation in the fused model. Overall, the results demonstrate that attention-guided representations and model fusion substantially strengthen fetal image classification performance.

Table 1. Ablation Results Showing Accuracy (%) Across Different Models and Training Settings for Parameter-Based Classification.

Method	VGG 16	VGG 19	ResNet 50	Xception	Fused model
Images	89.23	87.42	80.13	84.24	88.48
Images+ Augmentation	92.72	91.42	90.32	88.72	92.91
Grad Cam images	96.21	93.23	95.62	96.21	97.32
Grad Cam images+ Augmentation	99.28	99.04	99.14	98.89	99.65

The ablation results demonstrate consistent performance improvements across all tasks when using the proposed fused model compared to individual backbone networks. In parameter-based classification, the fused model achieves the highest accuracy (99.65%), outperforming VGG16, VGG19, ResNet50, and Xception. For trimester-based HC classification, the fused model again leads with 96.24%, indicating better generalization across gestational age-specific variations. A similar trend is observed in FL classification, where the fused model attains the top accuracy of 97.70%. In the HC18 dataset evaluation, although all models exhibit slightly lower performance due to dataset complexity, the fused model maintains superior results with 93.33%. Overall, the Table 2

confirms that feature fusion substantially enhances accuracy, precision, recall, and F1-score across diverse fetal biometric classification scenarios.

Table 2. Ablation results for baseline models and the proposed fused model across all classification tasks.

Model	Accuracy	Precision	Recall	F1-Score
Parameter-Based Classification				
VGG16	99.28%	99.18%	99.26%	99.12%
VGG19	99.04%	98.28%	98.66%	99.04%
ResNet50	99.14%	99.12%	99.03%	99.14%
Xception	98.89%	98.25%	98.63%	98.72%
Proposed Fused Model	99.65%	99.66%	99.66%	99.65%
Trimester-Based Classification (HC)				
VGG16	94.91%	94.38%	94.26%	94.12%
VGG19	95.28%	95.28%	95.66%	95.14%
ResNet50	95.37%	95.12%	95.03%	95.14%
Xception	95.34%	95.25%	95.22%	95.12%
Proposed Fused Model	96.24%	96.34%	96.25%	96.17%
Trimester-Based Classification (FL)				
VGG16	96.82%	96.68%	96.36%	96.42%
VGG19	96.28%	96.14%	96.06%	96.12%
ResNet50	97.21%	97.18%	97.13%	97.04%
Xception	97.33%	97.15%	97.25%	97.22%
Proposed Fused Model	97.70%	97.71%	97.70%	97.70%
Trimester-Based Classification (HC18)				
VGG16	91.47%	91.33%	91.26%	91.17%
VGG19	92.85%	92.34%	92.16%	92.22%
ResNet50	86.71%	86.62%	86.35%	86.34%
Xception	85.73%	85.65%	85.15%	85.52%
Proposed Fused Model	93.33%	93.35%	93.33%	93.25%

4. Experimental Results

The complete dataset statistics used in this study for both parameter-based and trimester-based classification tasks. For parameter-based classification, the dataset consisted of four biometric parameters: HC, FL, AC, and CRL, with 1,426 HC, 1,409 FL, 199 AC, and 553 CRL original images. To overcome severe class imbalance and strengthen model generalization, Grad-CAM-guided augmentation with 15 transformations is applied to each class. This expanded the dataset to 21,390 HC, 21,135 FL, 2,985 AC, and 8,295 CRL images, resulting in a total of 53,805 samples. The dataset is then partitioned using a 70:15:15 ratio, producing 37,650 training, 8,025 validation, and 8,130 testing images to ensure reliable model training and evaluation. For trimester-based classification, Table 3

also presents the distribution of HC, FL, and the public HC18 dataset across the three trimesters. The HC dataset originally contained 148 FT, 380 ST, and 898 TT images (1,426 total). After augmentation, it increased to 2,220 FT, 5,700 ST, and 13,470 TT images (21,390 total). Similarly, the FL dataset included 116 FT, 395 ST, and 898 TT images (1,409 total), which expanded to 1,740 FT, 5,925 ST, and 13,470 TT images (21,135 total). The HC18 dataset consisted of 165 FT, 693 ST, and 141 TT images (1,334 total), and augmentation increased this to 2,475 FT, 10,395 ST, and 2,115 TT images (20,010 total). All trimester-based datasets are split using the same 70:15:15 practice to maintain consistency across experiments.

Table 3 presents the comprehensive performance of the proposed fused model, which integrates VGG16, VGG19, ResNet50, and Xception feature representations. The fused model demonstrates consistently high accuracy across all evaluation tasks. In parameter-based classification, it achieves near-perfect performance with 99.65% accuracy. For trimester-based HC and FL classification, the model maintains strong generalization with accuracies of 96.24% and 97.70%, respectively. Although performance slightly decreases on the more challenging HC18 dataset, the fused model still attains the highest accuracy of 93.33%. Overall, the experimental results confirm that the fused architecture provides robust and reliable feature learning, substantially enhancing fetal biometric classification across diverse datasets.

Table 3. Performance Metrics of the Proposed Fused Model for All Classification Tasks.

Proposed Fused Model (VGG16+VGG19+ResNet50+Xception)	Accuracy	Precision	Recall	F1-Score
Parameter-Wised Classification	99.65%	99.66%	99.6%	99.65%
Trimester-Wised Classification (HC)	96.24%	96.34%	96.2%	96.17%
Trimester-Wised Classification (FL)	97.70%	97.71%	97.7%	97.7%
Trimester-Wised Classification (HC18)	93.33%	93.35%	93.3%	93.28%

Figure 14 illustrates the model evaluation results, including the learning curves, confusion matrices, and t-SNE feature visualization for both parameter-based and trimester-based tasks. These results demonstrate clear class separability, stable training behaviour, and improved discriminative performance, confirming that the proposed fused model effectively utilizes the expanded, balanced datasets to achieve robust classification outcomes.

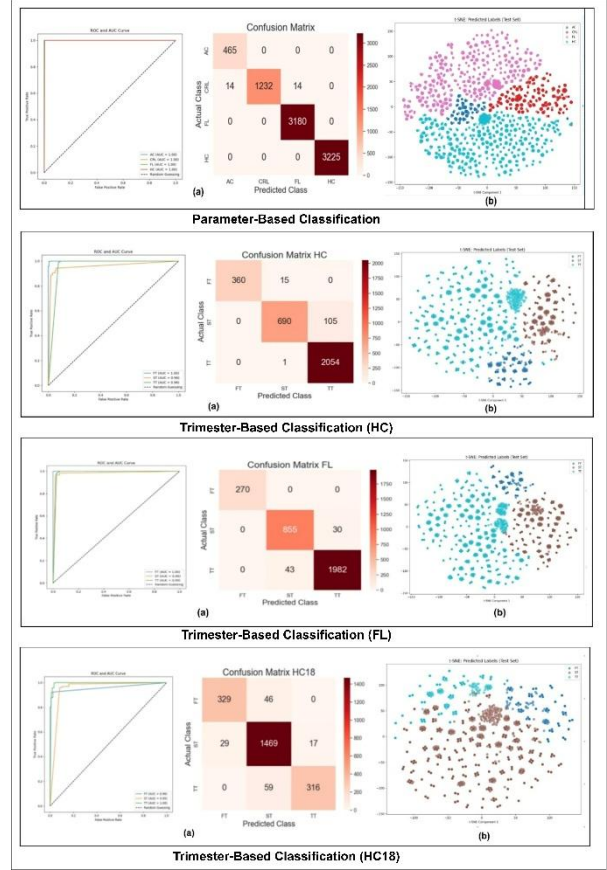


Figure 14. Fused model Evaluation results including learning curves, confusion matrices, and t-SNE plots for parameter-based and trimester-based classification.

4.2 Comparative Study

Table 4 presents a comparative analysis of recent fetal ultrasound classification studies. Gornale et al.[5] achieved 99.82% accuracy using XceptionNet on HC, FL, and AC, while Ghelich Oghli et al.[32] reported 95.56% with a CNN. Sivasubramanian et al.[9] used EfficientNetV2B0+MLP for multi-organ classification (96.25%), and Ghabri et al.[15] applied DenseNet169, reaching 99.78%. For fetal plane classification, Hasan et al.[18] attained 97.68%, and Burgos-Artizzu et al.[33] 93.6%. Gornale et al.[10] achieved 83.68–92.50% in trimester-based classification. The proposed fused model achieves 99.65% for parameter-wise and 93.33–97.70% for trimester-wise classification, outperforming existing methods and demonstrating robust, generalizable performance.

Table 4. Comparison of Proposed Work with Related Works

Authors	Dataset	Methodology	Parameters	Classification Results
Gornale, S., et al. (2024)[5]	Created Dataset	XceptionNet	HC, FL, AC	99.82%
Ghelich Oghli, M., et al. (2021)[32]	HC18 Grand Challenge Dataset and Created Dataset	CNN	HC, FL, AC	95.56%
Sivasubramanian, A., et al. (2024)[9]	Created Dataset	EfficientNetV2B0 + MLP	Multi-Organ	96.25%

Ghabri, H., et al. (2023)[15]	Created Dataset	DenseNet169	Multi-Organ	99.78%
Hasan, M. N., et al. (2024)[18]	FETAL_PL ANES_DB	Ensemble Model	Fetal Planes	97.68%
Burgos-Artizzu, X. P., et al. (2020)[33]	FETAL_PL ANES_DB	DenseNet169	Fetal Planes	93.6%
Gornale, S., et al. (2025)[10]	HC18 Grand Challenge dataset and Created Dataset (HC, FL)	Multi-input DenseNet121	Trimester-Based (FT, ST, TT)	HC18=83.68%
				HC=92.50%
				FL=90.60%
Proposed Work (Parameter-wise)	Custom Dataset	Fused Model	HC, FL, AC, CRL	99.65%
Proposed Work (Trimester-wise)	HC18 Grand Challenge dataset and Created Dataset (HC, FL)	Fused Model	Trimester-Based (FT, ST, TT)	HC18=93.33%
				HC=96.24%
				FL=97.70%

4.3 Five-Fold Cross-Validation Results of the Proposed Fused Model

Table 5 summarizes the proposed fused model's performance in parameter-wise and trimester-wise fetal biometric classification using five-fold cross-validation. For parameter-wise classification, the model achieved high consistency with a mean $\pm$ SD of $99.29 \pm 0.49\%$ accuracy, $99.24 \pm 0.54\%$ precision, $99.29 \pm 0.49\%$ recall, and $99.06 \pm 0.42\%$ F1-score. Trimester-wise evaluation showed robust results for HC ($95.92 \pm 0.48\%$ accuracy), FL ($96.87 \pm 0.63\%$), and the HC18 subset ($93.06 \pm 0.88\%$), reflecting its reliability across diverse fetal measures. These results demonstrate that the model is accurate and generalizable and can support automated clinical assessment of fetal growth.

Table 5: Five-fold cross-validation performance of the fused model for fetal biometric parameters, showing consistent accuracy and reliability across folds for both parameter-wise and trimester-wise ultrasound assessment.

Fused Model Evaluation	Accuracy	Precision	Recall	F1-Score
Parameter-Wise Performance				
Fold 1	99.26	98.57	99.26	99.29
Fold 2	99.27	99.52	99.27	98.94
Fold 3	99.70	99.79	99.70	99.09
Fold 4	98.52	98.75	98.52	98.43
Fold 5	99.72	99.56	99.72	99.54
Mean $\pm$ SD	99.29 ± 0.49	99.24 ± 0.54	99.29 ± 0.49	99.06 ± 0.42
Accuracy	99.65	99.66	99.66	99.65
Trimester-Wise Performance (HC)				

Fold 1	95.41	95.17	95.41	95.26
Fold 2	96.25	96.32	96.25	96.22
Fold 3	96.25	96.62	96.25	95.09
Fold 4	96.31	96.25	96.31	96.45
Fold 5	95.38	95.56	95.38	95.64
Mean $\pm$ SD	95.92 ± 0.48	95.98 ± 0.60	95.92 ± 0.48	95.73 ± 0.59
Accuracy	96.24	96.34	96.25	96.17
Trimester-Wise Performance (FL)				
Fold 1	97.18	97.43	97.18	97.19
Fold 2	96.37	96.22	96.37	96.24
Fold 3	96.08	96.29	96.08	96.29
Fold 4	97.62	97.45	97.62	97.48
Fold 5	97.12	97.26	97.12	97.34
Mean $\pm$ SD	96.87 ± 0.63	96.93 ± 0.57	96.87 ± 0.63	96.91 ± 0.56
Accuracy	97.70	97.71	97.70	97.70
Trimester-Wise Performance (HC18)				
Fold 1	92.84	92.54	92.84	92.63
Fold 2	93.12	93.32	93.12	93.75
Fold 3	93.25	93.62	93.25	93.58
Fold 4	94.12	94.72	94.12	94.66
Fold 5	91.95	91.79	91.95	91.88
Mean $\pm$ SD	93.06 ± 0.88	93.20 ± 1.10	93.06 ± 0.88	93.30 ± 1.06
Accuracy	93.33	93.35	93.33	93.25

4.4 Test for Statistical Significance

A significance test is used to determine whether the experimental results are statistically meaningful. Paired t-tests are performed on both datasets to validate the findings of this study[4, 34]. Table 6 presents the statistical significance of the fused model's classification results for the custom dataset and the HC18 dataset. The paired t-test compares the mean difference between two related groups, such as predicted and actual values, to evaluate whether the difference is statistically significant. In a paired t-test, the null hypothesis (H_0) assumes the mean difference μ_d between paired observations is zero, while the alternative (H_1) assumes $\mu_d \neq 0$. The t-statistic is calculated as $t = \bar{d}/(s_d/\sqrt{n})$, where $d_i = |\text{Actual}_i - \text{Predicted}_i|$, $\bar{d}$ is the mean difference, and s_d is the standard deviation of differences. This statistic is compared against a t-distribution with $df = n - 1$ to determine significance[10].

As shown in Table 6, statistical significance is evaluated using a two-tailed t-test at the 5% level ($\alpha = 0.05$). For the fused model (parameter-wise), the calculated t-value ($t = 1.03$) is lower than the critical value $t_{\text{table}} = 3.182$ for 3 degrees of freedom, indicating acceptance of the null hypothesis (H_0). Similarly, for the trimester-wise fused models of HC and FL, the calculated t-values ($t = 1.20$ and $t = 1.91$, respectively) are below the critical value $t_{\text{table}} = 4.303$ for 2 degrees of freedom, confirming that H_0 is accepted. Additionally, for the fused HC trimester-

wise evaluation, the calculated t-value ($t = 4.03$) also remains below the same critical threshold (4.303), reaffirming that H_0 is accepted. Overall, these results indicate that the differences between actual and predicted values are not statistically significant, and any observed variation is likely due to random fluctuations rather than a systematic deviation.

Table 6. Two-Tailed T-Test Results for Parameter-Wise and Trimester-Wise Fused Models.

Model	Class	Actual value	Predicted value	Difference (di)	Where,
Fused Model (Parameter-wise)	AC	465	465	0	$n = 4, df = 3,$ $\bar{d} = 7,$ $S_d = 13.64,$ $t_{calculated} = 1.03,$ $t_{table} = 3.182$ $t_{calculated} < t_{table}$ H_0 is accepted
	CRL	1260	1232	28	
	FL	3180	3180	0	
	HC	3225	3225	0	
Fused Model FL (Trimester-wise)	FT	375	360	15	$n = 3, df = 2,$ $\bar{d} = 40.33,$ $S_d = 58.03,$ $t_{calculated} = 1.20,$ $t_{table} = 4.303,$ $t_{calculated} < t_{table}$ H_0 is accepted
	ST	795	690	105	
	TT	2055	2054	1	
Fused Model HC (Trimester-wise)	FT	270	270	0	$n = 3, df = 2,$ $\bar{d} = 24.33, S_d = 2.05,$ $t_{calculated} = 1.91,$ $t_{table} = 4.303,$ $t_{calculated} < t_{table}$ H_0 is accepted
	ST	885	855	30	
	TT	2025	1982	43	
Fused Model HC18 (Trimester-wise)	FT	375	329	46	$n = 3, df = 2,$ $\bar{d} = 75.5,$ $S_d = 25.90,$ $t_{calculated} = 4.03,$ $t_{table} = 4.303, t_{calculated} < t_{table}$ H_0 is accepted
	ST	1515	1469	46	
	TT	375	316	59	

4.5 Discussion

Several deep learning approaches have been proposed for fetal ultrasound analysis. SonoNet [1] achieved an F1 score of 0.798, AG-CNN [2] reached 94% accuracy, and EfficientNet-based methods[9] reported Top-1 accuracy of 96.25%. Transfer learning with InceptionResNetV2, DenseNet, MobileNet, and XceptionNet[5, 15] achieved accuracies above 99% for parameter-based tasks. However, trimester-wise classification remains challenging due to subtle developmental changes; for example, Gornale et al.[10] reported 83.68% accuracy on HC18 and ~91% on custom datasets. In contrast, the proposed fused model attains 99.65% for parameter-wise classification and significantly improves trimester-wise performance,

achieving 93.33% on HC18 and 96.24–97.70% on custom datasets. These results demonstrate the framework's robustness and potential for reliable prenatal assessment.

5. Conclusion and Future Work

This study presents a robust deep learning framework for automated classification of fetal ultrasound images using key biometric parameters (HC, FL, AC, CRL) and trimester-based categories (FT, ST, TT). A novel contribution of this work is the use of Grad-CAM activation maps not only for explainability but also as attention-enhanced inputs, enabling the model to learn from clinically relevant regions while suppressing background noise. Additionally, the fusion of multiple pretrained models (VGG16, VGG19, ResNet50, Xception) provides a richer multi-level representation, significantly boosting performance.

The proposed method achieved high accuracy, precision, recall, F1-score, and ROC/AUC across both custom and benchmark datasets, demonstrating its effectiveness for reliable fetal biometric assessment. Remaining challenges include variations in image quality, fetal position, and limited diversity of single-parameter images. Future work will explore dataset expansion through generative models, integration of vision transformers for improved global feature extraction, and multi-center validation to advance the framework toward real-world clinical deployment for comprehensive fetal growth monitoring.

Declarations

Funding

This research received no external funding.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgment

The authors express their gratitude to KSTEPS, DST, Government of Karnataka, for their financial support and for granting a Ph.D. fellowship, which made this research possible. We also extend our sincere thanks to Dr. Vanita B. Metgud and Dr. Satwik B. Metgud from Metgud Hospital - Advanced Laparoscopy Centre and IVF, Belagavi, Karnataka, India, for providing the ultrasound fetal images and sharing valuable insights regarding the analysis and abnormalities in fetuses, which greatly contributed to this study.

References:

1. Baumgartner, C.F., Kamnitsas, K., Matthew, J., Fletcher, T.P., Smith, S., Koch, L.M., Kainz, B., Rueckert, D.: SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound. *IEEE Trans Med Imaging*. 36, 2204–2215 (2017). <https://doi.org/10.1109/TMI.2017.2712367>.
2. Al-Razgan, M., Ali, Y.A., Awwad, E.M.: Enhancing Fetal Medical Image Analysis through

- Attention-guided Convolution: A Comparative Study with Established Models. *Journal of Disability Research*. 3, (2024). <https://doi.org/10.57197/jdr-2024-0005>.
3. Kiserud, T., Piaggio, G., Carroli, G., Widmer, M., Carvalho, J., Neerup Jensen, L., Giordano, D., Cecatti, J.G., Abdel Aleem, H., Talegawkar, S.A., Benachi, A., Diemert, A., Tshefu Kitoto, A., Thinkhamrop, J., Lumbiganon, P., Tabor, A., Kriplani, A., Gonzalez Perez, R., Hecher, K., Hanson, M.A., Gülmezoglu, A.M., Platt, L.D.: The World Health Organization Fetal Growth Charts: A Multinational Longitudinal Study of Ultrasound Biometric Measurements and Estimated Fetal Weight. *PLoS Med*. 14, (2017). <https://doi.org/10.1371/journal.pmed.1002220>.
 4. Alzubaidi, M., Agus, M., Shah, U., Makhlof, M., Alyafei, K., Househ, M.: Ensemble Transfer Learning for Fetal Head Analysis: From Segmentation to Gestational Age and Weight Prediction. *Diagnostics*. 12, (2022). <https://doi.org/10.3390/diagnostics12092229>.
 5. Gornale, S., Kamat, P., Siddalingappa, R., Kumar, S.: Deep Learning Techniques for a Comprehensive Analysis of Fetal Biometric Parameters Across Trimesters. *Transactions on Machine Learning and Artificial Intelligence*. 12, 18–45 (2024). <https://doi.org/10.14738/tecs.123.16985>.
 6. Fiorentino, M.C., Villani, F.P., Di Cosmo, M., Frontoni, E., Moccia, S.: A Review on Deep-Learning Algorithms for Fetal Ultrasound-Image Analysis. (2022). <https://doi.org/10.1016/j.media.2022.102629>.
 7. Rauf, F., Attique Khan, M., Albarakati, H.M., Jabeen, K., Alsenan, S., Hamza, A., Teng, S., Nam, Y.: Artificial intelligence assisted common maternal fetal planes prediction from ultrasound images based on information fusion of customized convolutional neural networks. *Front Med (Lausanne)*. 11, (2024). <https://doi.org/10.3389/fmed.2024.1486995>.
 8. Harikumar, A., Surendran, S., Gargi, S.: Explainable AI in Deep Learning Based Classification of Fetal Ultrasound Image Planes. In: *Procedia Computer Science*. pp. 1023–1033. Elsevier B.V. (2024). <https://doi.org/10.1016/j.procs.2024.03.291>.
 9. Sivasubramanian, A., Sasidharan, D., V, S., Ravi, V.: Efficient feature extraction using light-weight CNN attention-based deep learning architectures for ultrasound fetal plane classification. *Phys Eng Sci Med* 48, 1079–1093 (2025). <https://doi.org/10.1007/s13246-025-01566-6>
 10. Gornale, S.S., Kamat, P.C., Hiremath, P.S., Siddalingappa, R.: A Hybrid Ensemble of Denoising Autoencoders and Deep Learning Models for Fetal Image Analysis. *Cureus Journal of Computer Science*. (2025). <https://doi.org/10.7759/s44389-025-09506-x>.
 11. Rahaman, M.M., Li, C., Yao, Y., Kulwa, F., Wu, X., Li, X., Wang, Q.: DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques. *Comput Biol Med*. 136, (2021). <https://doi.org/10.1016/j.compbio.2021.104649>.
 12. Jadhav, V.S., Yakkundimath, R., Saunshi, G., Konnurmath, G.: Deep Learning Based Classification of Cervical Cancer Stages Using Transfer Learning Models. *Indian Journal of Medical and Paediatric Oncology*. (2025). <https://doi.org/10.1055/s-0045-1809907>.
 13. Ishikawa, G., Xu, R., Ohya, J., Iwata, H.: Detecting a Fetus in Ultrasound Images using Grad CAM and Locating the Fetus in the Uterus. In: *International Conference on Pattern Recognition Applications and Methods*. pp. 181–189. Science and Technology Publications, Lda (2019). <https://doi.org/10.5220/0007385001810189>.
 14. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. (2016). <https://doi.org/10.1007/s11263-019-01228-7>.
 15. Ghabri, H., Alqahtani, M.S., Ben Othman, S., Al-Rasheed, A., Abbas, M., Almubarak, H.A., Sakli, H., Abdelkarim, M.N.: Transfer learning for accurate fetal organ classification from ultrasound images: a potential tool for maternal healthcare providers. *Sci Rep*. 13, (2023). <https://doi.org/10.1038/s41598-023-44689-0>.
 6. Rathika, S., Mahendran, K., Sudarsan, H., Ananth, S.V.: Novel neural network classification of maternal fetal ultrasound planes through optimized feature selection. *BMC Med Imaging*. 24, (2024). <https://doi.org/10.1186/s12880-024-01453-8>.
 17. Lo, J., Lim, A., Wagner, M.W., Ertl-Wagner, B., Sussman, D.: Fetal Organ Anomaly Classification Network for Identifying Organ Anomalies in Fetal MRI. *Front Artif Intell*. 5, (2022). <https://doi.org/10.3389/frai.2022.832485>.
 18. Hasan, M.N., Aowlad Hossain, A.B.M.: Fetal Brain Planes Classification Using Deep Ensemble Transfer Learning from U-Net Segmented Fetal Neurosonography Images. *International Journal of Image, Graphics and Signal Processing*. 16, 74–86 (2024). <https://doi.org/10.5815/ijigsp.2024.04.06>.
 19. Liang, J., Huang, R., Kong, P., Li, S., Wang, T., Lei, B.: SPRNet: Automatic fetal standard plane recognition network for ultrasound images. In: *Lecture Notes in Computer Science (including*

- subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 38–46. Springer (2019). https://doi.org/10.1007/978-3-030-32875-7_5.
20. Faghihi, A., Fathollahi, M., Rajabi, R.: Diagnosis of skin cancer using VGG16 and VGG19 based transfer learning models. *Multimed Tools Appl* 83, 57495–57510 (2024). <https://doi.org/10.1007/s11042-023-17735-2>
 21. Visuña, L., Yang, D., Garcia-Blas, J., Carretero, J.: Computer-aided diagnostic for classifying chest X-ray images using deep ensemble learning. *BMC Med Imaging*. 22, (2022). <https://doi.org/10.1186/s12880-022-00904-4>.
 22. Gupta, A., Dhanakshirur, R.R., Jain, K., Garg, S., Yadav, N., Seth, A., Das, C.J.: Deep Learning for Detecting and Subtyping Renal Cell Carcinoma on Contrast-Enhanced CT scans Using 2D Neural Network with Feature Consistency Techniques. *Indian Journal of Radiology and Imaging*. 35, 395–401 (2024). <https://doi.org/10.1055/s-0044-1800804>.
 23. van den Heuvel, T.L.A., de Bruijn, D., de Korte, C.L., van Ginneken, B.: Automated measurement of fetal head circumference using 2D ultrasound images. *PLoS One*. 13, (2018). <https://doi.org/10.1371/journal.pone.0200412>.
 24. Dubey, G., Srivastava, S., Jayswal, A.K., Saraswat, M., Singh, P., Memoria, M.: Fetal Ultrasound Segmentation and Measurements Using Appearance and Shape Prior Based Density Regression with Deep CNN and Robust Ellipse Fitting. *Journal of Imaging Informatics in Medicine*. 37, 247–267 (2024). <https://doi.org/10.1007/s10278-023-00908-8>.
 25. Walker, M.C., Willner, I., Miguel, O.X., Murphy, M.S.Q., El-Chaâr, D., Moretti, F., Dingwall Harvey, A.L.J., White, R.R., Muldoon, K.A., Carrington, A.M., Hawken, S., Aviv, R.I.: Using deep-learning in fetal ultrasound analysis for diagnosis of cystic hygroma in the first trimester. *PLoS One*. 17, (2022). <https://doi.org/10.1371/journal.pone.0269323>.
 26. Bano, S., Dromey, B., Vasconcelos, F., Napolitano, R., David, A.L., Peebles, D.M., Stoyanov, D.: AutoFB: Automating Fetal Biometry Estimation from Standard Ultrasound Planes. (2021). https://doi.org/10.1007/978-3-030-87234-2_22
 27. Yang, Y., Yang, P., Zhang, B.: Automatic segmentation in fetal ultrasound images based on improved U-net. In: *Journal of Physics: Conference Series*. IOP Publishing Ltd (2020). <https://doi.org/10.1088/1742-6596/1693/1/012183>.
 28. Kumaran S, Y., Jeya, J.J., Mahesh T, R., Khan, S.B., Alzahrani, S., Alojail, M.: Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam. *BMC Med Imaging*. 24, (2024). <https://doi.org/10.1186/s12880-024-01345-x>.
 29. Patil, G., Palaiahnakote, S., Gornale, S.S., Lopresti, D.P.: Altered Handwritten Text Detection in Document Images Using Deep Learning. *Intern J Pattern Recognit Artif Intell*. 38, (2024). <https://doi.org/10.1142/S0218001424520062>.
 30. Shen, X., Wei, L., Tang, S.: Dermoscopic Image Classification Method Using an Ensemble of Fine-Tuned Convolutional Neural Networks. *Sensors*. 22, (2022). <https://doi.org/10.3390/s22114147>.
 31. Kuzu, A., Santur, Y.: Early Diagnosis and Classification of Fetal Health Status from a Fetal Cardiotocography Dataset Using Ensemble Learning. *Diagnostics*. 13, (2023). <https://doi.org/10.3390/diagnostics13152471>.
 32. Ghelich Oghli, M., Shabanzadeh, A., Moradi, S., Sirjani, N., Gerami, R., Ghaderi, P., Sanei Taheri, M., Shiri, I., Arabi, H., Zaidi, H.: Automatic fetal biometry prediction using a novel deep convolutional network architecture. *Physica Medica*. 88, 127–137 (2021). <https://doi.org/10.1016/j.ejmp.2021.06.020>.
 33. Burgos-Artizzu, X.P., Coronado-Gutiérrez, D., Valenzuela-Alcaraz, B., Bonet-Carne, E., Eixarch, E., Crispí, F., Gratacós, E.: Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. *Sci Rep*. 10, (2020). <https://doi.org/10.1038/s41598-020-67076-5>.
 34. Kong, P., Ni, D., Chen, S., Li, S., Wang, T., Lei, B.: Automatic and efficient standard plane recognition in fetal ultrasound images via multi-scale dense networks. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp. 160–168. Springer Verlag (2018). https://doi.org/10.1007/978-3-030-00807-9_16.