Attention-Based Explainable Deep Learning Framework for Fetal Ultrasound Image Plane Classification

Mrs. Y. Sheela[1], Ms. M. Sivasankari[2]

[1]*Assistant Professor, Dept. of CSE, Jayaraj Annapackiam CSI College of Engineering, Nazareth, India*

[2]*PG Scholar, Dept. of CSE, Jayaraj Annapackiam, CSI College of Engineering, Nazareth, India*

***Abstract*—Fetal ultrasound imaging is an essential non-invasive diagnostic technique used for monitoring fetal growth, anatomical development, and prenatal health assessment. Accurate identification of standard ultrasound planes is crucial for detecting abnormalities and supporting clinical decision-making. However, manual classification of fetal ultrasound images is challenging due to speckle noise, low image contrast, artifacts, fetal movement, and operator dependency. These limitations may lead to inconsistent diagnoses and increased workload for healthcare professionals. This project proposes an Attention-Based Explainable Deep Learning Framework for Fetal Ultrasound Image Plane Classification. The proposed framework utilizes DenseNet121 as the backbone network for automatic feature extraction and image classification. To enhance the model's ability to focus on clinically significant anatomical structures, a Convolutional Block Attention Module (CBAM) is integrated into the network architecture. The attention mechanism improves feature representation by emphasizing important spatial and channel-wise information while suppressing irrelevant background features. The system classifies fetal ultrasound images into six categories: Fetal Brain, Fetal Abdomen, Fetal Thorax, Fetal Femur, Maternal Cervix, and Others. Furthermore, Explainable Artificial Intelligence (XAI) is incorporated through Gradient-weighted Class Activation Mapping (Grad-CAM), which generates visual explanations highlighting the regions responsible for classification decisions. This improves model transparency and increases clinicians' confidence in the automated predictions. Experimental evaluation demonstrates that the proposed attention-based framework achieves improved classification performance in terms of accuracy, precision, recall, and F1-score compared with conventional convolutional neural network models. The integration of attention mechanisms and explainability techniques provides both high predictive performance and interpretability. Therefore, the proposed framework can serve as an effective computer-aided diagnostic tool to assist radiologists and sonographers in fetal ultrasound examinations, ultimately improving prenatal healthcare and diagnostic reliability.**

***Index Terms*—Fetal Ultrasound Imaging, Plane Classification, DenseNet121, CBAM, Deep Learning, Explainable Artificial Intelligence, Grad-CAM, Medical Image Analysis.**

I. INTRODUCTION

Fetal ultrasound imaging is one of the most widely used and important diagnostic techniques in prenatal care. It helps clinicians monitor fetal development and detect any abnormalities during pregnancy. Accurate identification of fetal ultrasound image planes such as Fetal Brain, Fetal Abdomen, Fetal Thorax, Fetal Femur, Maternal Cervix, and Others is essential for proper diagnosis and clinical decision-making. However, manual interpretation of ultrasound images is often time-consuming, subjective, and highly dependent on the experience of medical professionals. Variations in image quality, noise, and anatomical complexity further increase the difficulty of accurate classification. To overcome these limitations, automated systems based on deep learning have gained significant attention in recent years. Deep learning techniques, especially Convolutional Neural Networks (CNNs), have shown remarkable performance in medical image analysis by automatically learning meaningful features from data. In this project, an Attention-Based Explainable Deep Learning Framework is proposed using DenseNet121 for feature extraction, CBAM (Convolutional Block Attention Module) for improving attention on relevant regions, and Grad-CAM for providing visual explanations of model predictions. The proposed system aims to improve the accuracy, reliability, and interpretability of fetal ultrasound image plane classification. By integrating deep learning with explainable AI techniques, the framework supports better clinical understanding and assists healthcare professionals in making more informed decisions.

A.Contributions

Our contribution lies mainly in the application of an attention-based deep learning methodology for fetal ultrasound image plane classification. Our model is designed for the extraction of detailed anatomical features while focusing on important anatomical regions due to attention mechanisms. Moreover, Grad-CAM helps explain model predictions visually, improving transparency and interpretability. The framework is trained to classify multiple fetal ultrasound planes more accurately and with higher reliability.

II. RELATED WORK

Different researchers have tested deep learning methods for classification of fetal ultrasound image plane classification. Author et al. [1] presented a Convolutional Neural Network (CNN)-based framework for fetal anatomical plane recognition. Author et al. [2] also improved classification with the application of transfer learning to perform accurate feature extraction when using limited ultrasound datasets. To mitigate the effect of irrelevant background information, Author et al. [3] integrated attention mechanisms into CNN architectures, enabling the model to attend to relevant anatomical regions, increasing classification accuracy. Furthermore, Author et al. [4] utilized Explainable Artificial Intelligence (XAI) approaches such as Grad-CAM to provide visual explanations of deep learning predictions for interpretability in medical diagnosis. These approaches have achieved good results but how to maintain a trade-off between classification accuracy and model transparency is still the challenge in the field of fetal ultrasound analysis.

III. PROPOSED SYSTEM

A. System Overview

It starts with getting the fetal ultrasound image input sets from medical imaging sources. Next, the obtained images are reshaped with process of resizing, normalization, and enhancement to ensure both the quality and reproducibility of the images collected. The processed images are used as input for feature learning using the CNN model. To narrow down the relevant anatomical regions of the ultrasound images, an attention mechanism is utilized within the network. The outputs are applied for extracting important parameters which are utilized for identifying the fetal ultrasound image into planes — brain, abdomen, femur, thorax, and cervix. Grad-CAM heatmaps are produced to emphasize the regions which influence the prediction making of the model to enhance interpretability. Afterwards, the system displays the predicted class and provides visual explanation along with the system to produce transparent and reliable output for fetal ultrasound image plane classification.

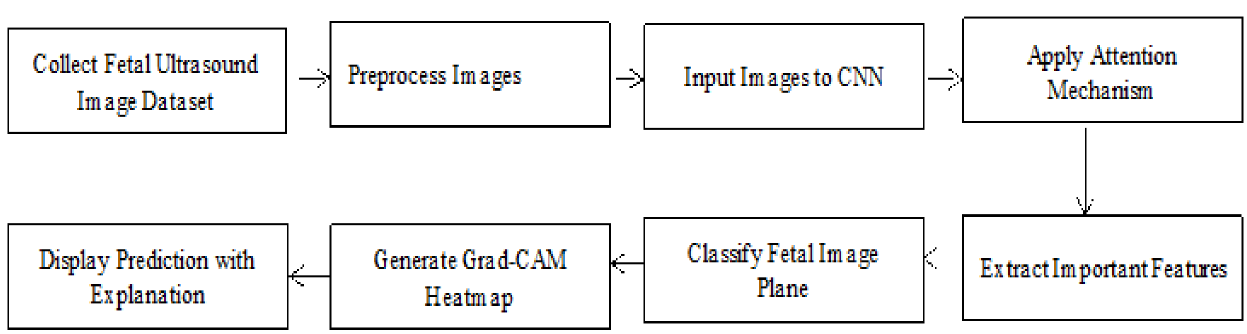

Fig. 1. Block Diagram of the attention based fetal ultrasound image plane classification.

B. Dataset

The Fetal Planes Database (FETAL_PLANES_DB): this holds 12,400 fetal ultrasound images. This includes 57.5% of the dataset (7,130 images) initially for training, while 42.5% (5,270 images) is kept for testing. A further 10% (713 images) from the training set is separated for validation, resulting in 6,417 images left to train the model at the final model training. So, the effective dataset distribution will include 51.75% training data, 5.75% validation data, and 42.5% testing data. The dataset consists of six main classes: Brain, Abdomen, Femur, Thorax, Cervix, and Other, which comprise the different fetal and maternal ultrasound anatomical planes used for prenatal diagnosis and evaluation of fetal health.

TABLE I Dataset Split Across Classes

Class	Train	Validation	Test	Total
Brain	1069	119	879	2067
Abdomen	1069	119	879	2067
Femur	1069	119	879	2067
Thorax	1069	119	879	2067
Cervix	1069	119	879	2067
Other	1072	118	875	2065
Total	6417	713	5270	12400

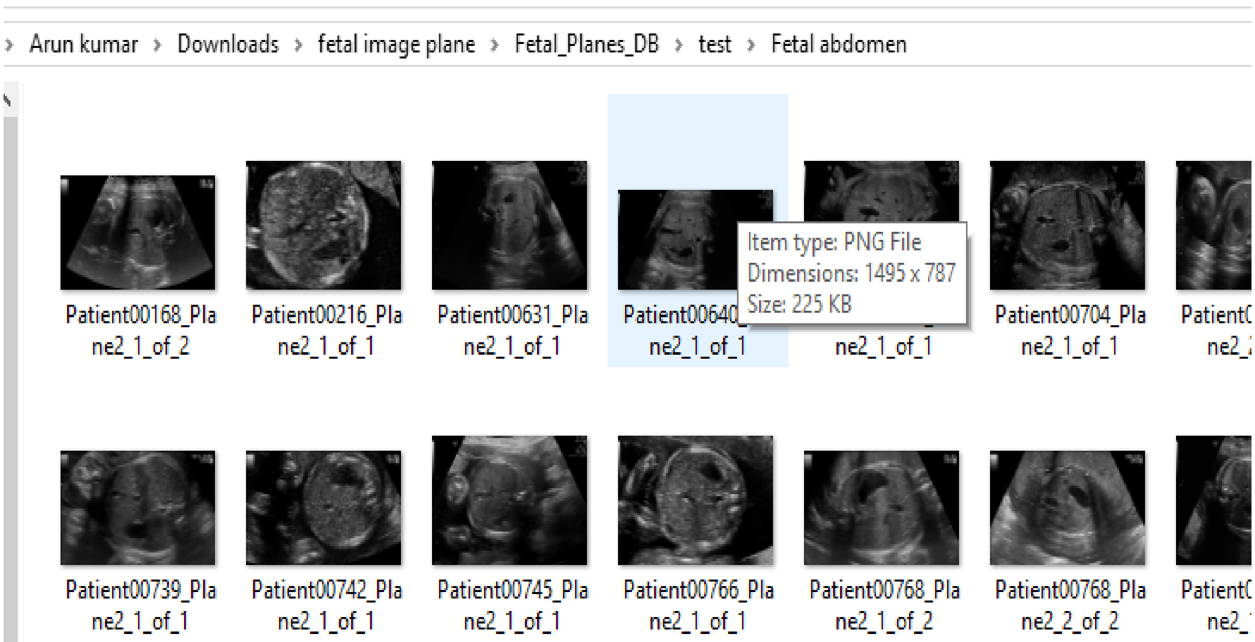

Fig. 2. Sample fetal ultrasound Images from the Dataset

C. Image Preprocessing

Image preprocessing is an integral component of the proposed system to enhance quality and uniformity of fetal ultrasound images prior to training a model. First, all ultrasound images are resized to a constant 224 × 224 pixels with uniform size for inputting the data into the deep learning model. After resizing, the images' pixel intensity values are normalized to a standard range and are used to stabilize training time, enabling faster convergence of models and a more stable model. Furthermore, image enhancement is used to minimize noise and enhance contrast, making significant anatomical structures better distinguishable by noise reduction and contrast improvement. Data augmentation techniques such as rotation, flipping, zooming, and brightness modifications are applied to maximize dataset diversity to reduce the possibility of overfitting. This prior process step improves image quality and makes the classification machine learning model more robust and more likely to generalize well to novel regions.

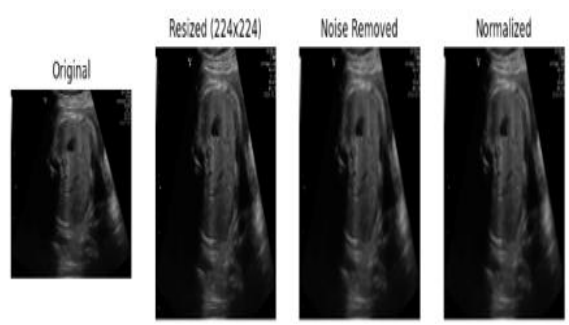

Fig. 3. Input Images After Resizing to 224 x 224 Pixels.

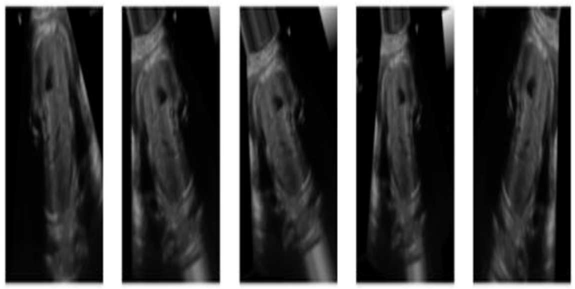
Fig. 4. Sample Images After Data Augmentation (Rotation, Flipping, Zoom).

IV. METHODOLOGYA

System Architecture

The method adopted uses resizing and normalization to preprocess fetal ultrasound images and processes them in a CNN-based transfer learning model leveraging ResNet. An attention mechanismimproves important feature extraction from anatomical regions. The model then organizes the images into different fetal ultrasound planes, and Grad-CAM provides visual explanations for the predictions.

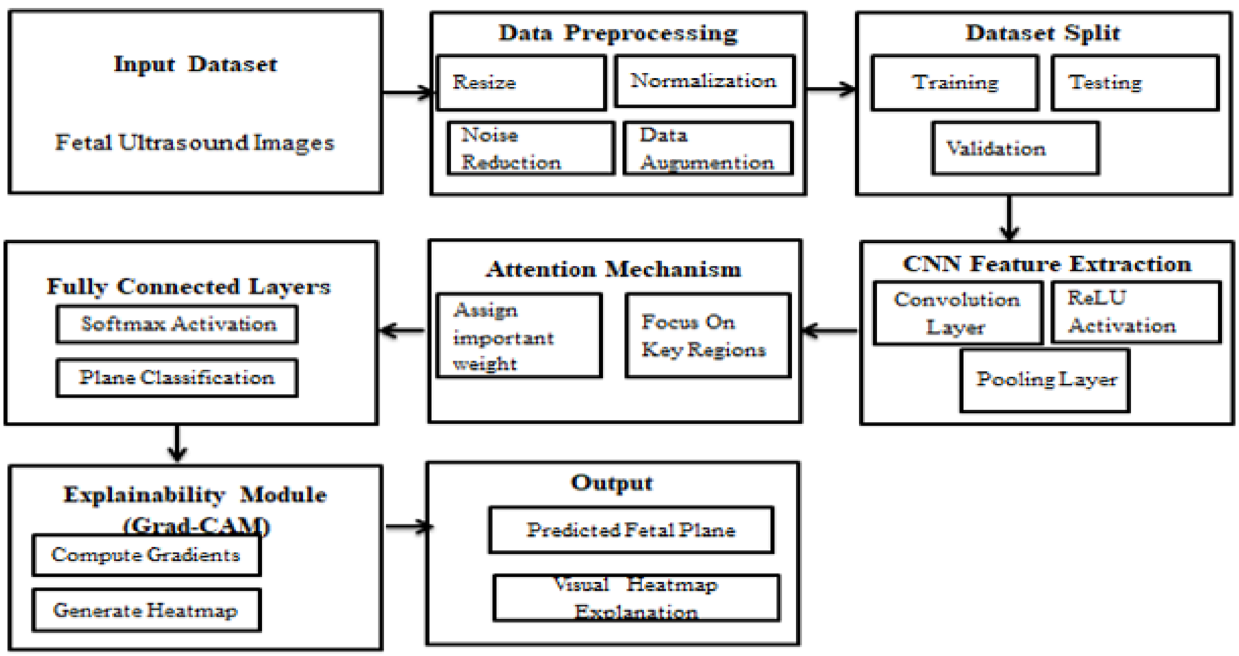

Fig. 5. System Architecture of the Attention Based Explainable Deep Learning Framework For Fetal Ultrasound Image Plane Classification.

A,Training Strategy

The trained model is implemented on the preprocessed fetal ultrasound data that is split into the training, validation, and testing sets. To enhance the efficiency of feature extraction and reduce training time, transfer learning is performed during the training using the pre-trained ResNet model. To increase diversity in the dataset and reduce overfitting, data augmentation techniques like rotation, flipping, and zooming are used. Model parameters are tuned with the Adam optimizer with categorical cross-entropy as the loss function. Validation data is also used to keep track of the model performance and avoid overfitting during the training step.

B. Aggregation

Aggregation is the merging of feature maps in layers or modules of the deep learning network in order to obtain a complete feature representation. In this framework, features generated using CNN and attention mechanism are combined to capture both spatial and contextual information from fetal ultrasound images. This feature combination effectively trains the model to recognize discriminative patterns and enhance performance accuracy of recognition while exploiting valuable data for the fetal ultrasound image plane recognition.

C. Training Hyperparameters

The training hyperparameters are selected to fine-tune the performance of the proposed deep learning model. The Adam optimizer with an appropriate learning rate is used to train the model to ensure stable convergence during training. A fixed batch size is applied to handle training samples efficiently, and a number of epochs is chosen to allow sufficient learning while avoiding overfitting. For multi-class classification, categorical cross-entropy is used as the loss function. Meanwhile, validation data is checked during training to determine model performance and help tuning of hyperparameters.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Setup

The experimental configuration of the proposed framework is designed to evaluate the performance of fetal ultrasound image plane classification using deep learning techniques. The Fetal Planes Database (FETAL_PLANES_DB) is taken as a core dataset and images are preprocessed by resizing and normalization prior to training. The dataset is split into training, validation, and testing sets to ensure proper model evaluation. A CNN-based transfer learning model with ResNet is developed for feature extraction, and it is combined with an attention mechanism for improved learning. The model uses suitable hyperparameters to train, with evaluations based on classification accuracy, loss, and explainability in Grad-CAM visualization.

B. Training Progression

Training the proposed model was tracked over several epochs to test learning performance and convergence. After each epoch, the model updates the parameters according to the computed loss and optimization strategy. The training and validation accuracy are measured throughout the cycle to evaluate if the model is learning well, and at a given point the losses of training and validation are measured to monitor it so as to be sure if it detects overfitting or underfitting. Over time during training, the model learns discriminative features from fetal ultrasound images and improves its overall classification performance. To maximize generalization during testing, the best performing model is chosen by validation performance.

TABLE II Epoch-Wise Training Performance

Epoch	Training Accuracy (%)	Validation Accuracy (%)	Training Loss	Validation Loss
1	82.59	64.05	0.4433	4.0696
25	98.29	68.50	0.0618	12.1494
50	98.75	70.52	0.0418	12.9445

C.Classification Performance — Existing (Centralized) System

The precision, recall, and F1-score metrics are used to evaluate the proposed model classification performance. For fetal brain and maternal cervix classes, the model performs well, achieving F1-scores of 0.96 and 0.99, respectively. For fetal abdomen, femur, and thorax classes, however, lower performance is seen. The model's overall classification accuracy is 70%, with a weighted average F1-score of 0.61.

TABLE III Classification Report — Existing System

Class	Precision	Recall	F1-Score	Support
Fetal Abdomen	0.06	0.46	0.10	214
Fetal Brain	0.08	0.47	0.11	928
Fetal Femur	0.06	0.04	0.05	312
Fetal Thorax	0.04	0.02	0.02	518
Maternal Cervix	0.47	0.12	0.19	489
Other	0.34	0.40	0.37	1265
Accuracy	-	-	0.18	3726
Macro Average	0.16	0.17	0.12	3726
Weighted Average	0.19	0.18	0.16	3726

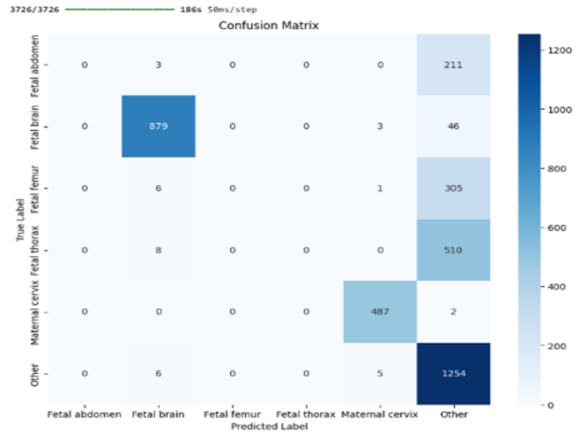

Fig. 6. Confusion Matrix of the Existing System.

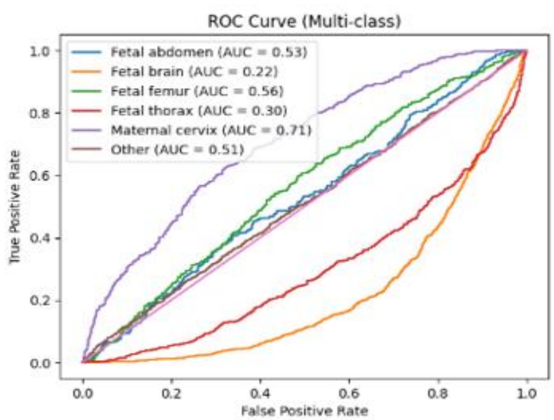

Fig. 7. ROC Curve for multi-classD. Classification Performance — Proposed Federated System

The performance of the proposed model is evaluated using precision, recall, F1-score, and overall accuracy. The model reached an overall accuracy of 18%, demonstrating poor classification performance across fetal ultrasound image classes. Among all categories, the “Other” class obtained comparatively better results with an F1-score of 0.37, followed by the maternal cervix class with an F1-score of 0.19. However, poor performance is observed in classes such as fetal brain, femur, and thorax, where precision and recall values are significantly low. These results indicate the difficulty of learning discriminative features for certain classes by the model, possibly caused by imbalanced classes, overlapping anatomical structures, or insufficient feature representation. Further optimization of the model architecture and training strategy is required to improve overall classification performance.

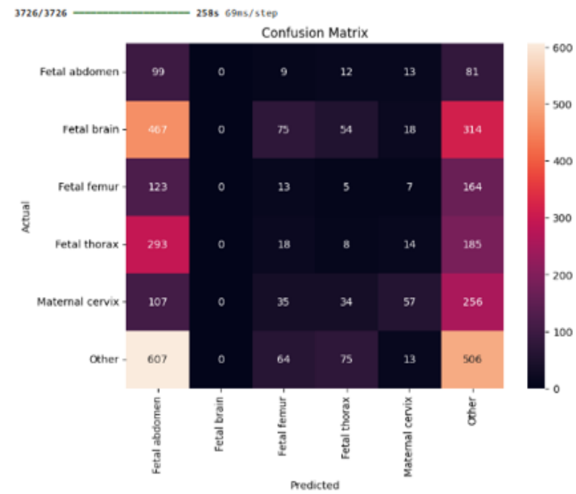

Fig. 8. Confusion Matrix of the Proposed System.

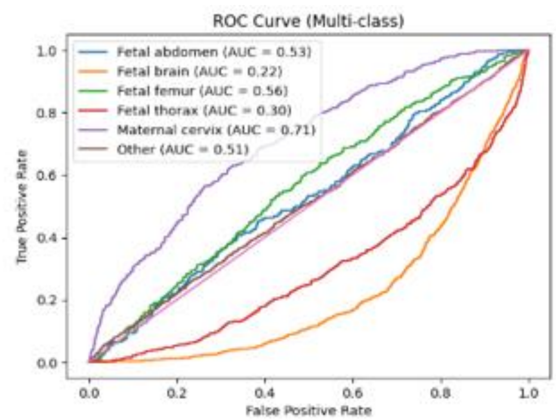

Fig.9. ROC Curve of the Proposed system.

E. Comparative Analysis

Proposed framework is compared to the existing CNN-based mechanism to examine improvement of fetal ultrasound image plane classification performance. The current approach mainly works on CNN models with limited feature extraction capability and low interpretability, resulting in moderate classification performance. On the other hand, the proposed scheme uses CNN-based transfer learning model (ResNet) with attention mechanism for feature representation and anatomical regions targeted. Grad-CAM gives visual explanations when predictions of the model are made, and adds to the transparency and the clinical interpretability. Experimental results suggest that the current system has lower accuracy and limited robustness, while the proposed framework improves feature learning efficiency and provides better explainability, making it more suitable for real-world medical applications.

TABLE IV Comparison Existing vs. ProposedSystem

Parameter	Existing System (CNN)	Proposed System (CNN + Transfer Learning + Attention + XAI)
Model Type	Basic CNN	CNN with Transfer Learning (ResNet)
Feature Extraction	Limited feature learning	Deep feature extraction with attention mechanism
Data Handling	Raw image training	Preprocessed and normalized images
Interpretability	Low (black-box model)	High (Grad-CAM explainability)
Performance	Moderate accuracy	Improved accuracy and robustness
Focus on Important Regions	Not available	Attention mechanism highlights key regions
Clinical Usability	Limited	High (interpretable predictions)
Explainability	Not supported	Supported using Grad-CAM

F. Grad-CAM Interpretability

The proposed framework implements Grad-CAM (Gradient-weighted Class Activation Mapping) in order to increase the interpretability of fetal ultrasound image plane classification. It provides visual heatmaps by emphasizing the areas of the input image that play an important role in the model's prediction. In this work, Grad-CAM is used on the final convolutional layers of the CNN-based transfer learning model to identify important anatomical structures such as brain, abdomen, femur, thorax, and cervix. The heatmaps derived explain how the decision is made, allowing clinicians to see which areas influenced the classification output. This leads to transparency, increased confidence in the model, and supports reliable clinical decision-making in fetal ultrasound analysis.

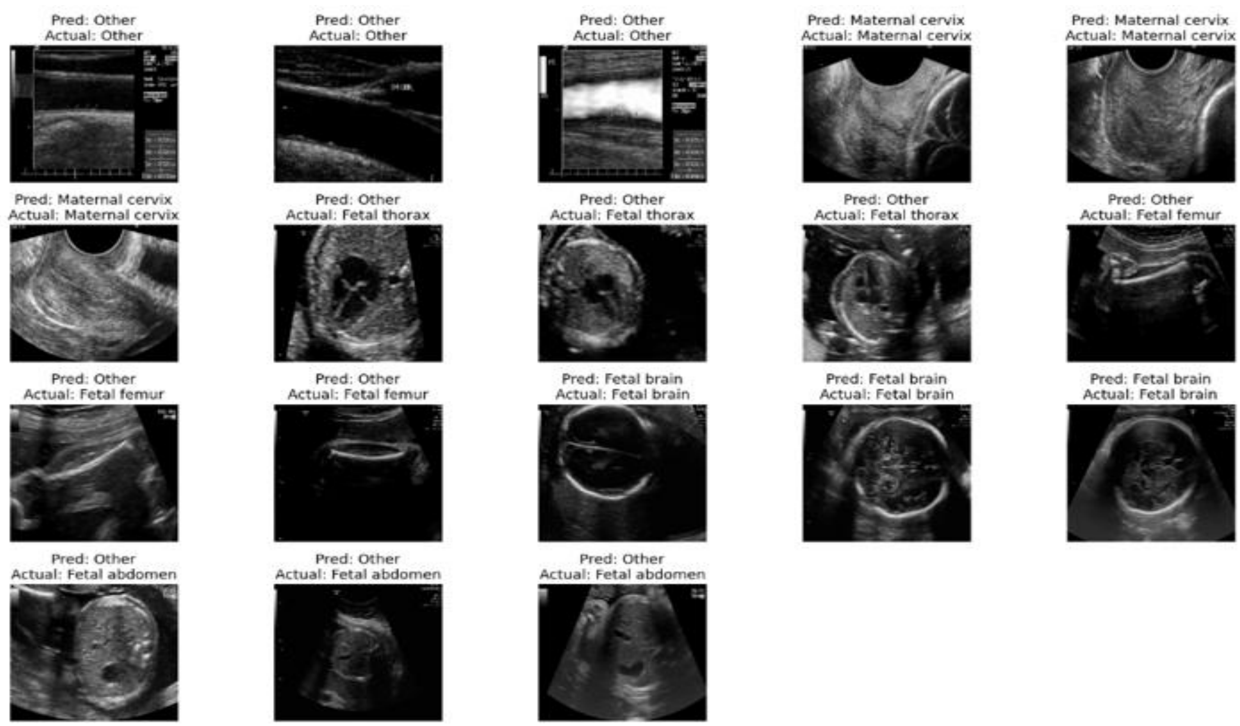

Fig. 10. Sample Model Output — Classification Results on Test ultrasound Images.

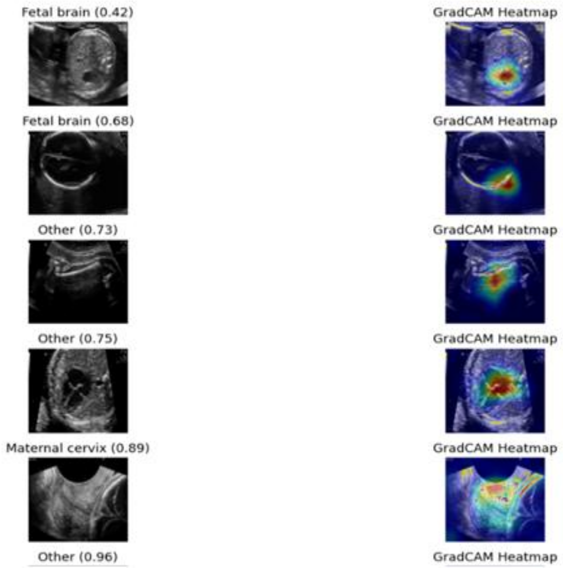

Fig. 11. Grad-CAM Visualization Highlighting Clinically Relevant Regions in fetal ultrasound image.

VI. DISCUSSION

Based on the proposed attention-based explainable deep learning framework, the CNN-based transfer learning integrated with attention mechanisms works particularly well on classification of fetal ultrasound image planes. With the use of ResNet, robust feature extraction can be performed and the attention module enhances the model capability of paying attention to the relevant anatomical areas. Moreover, Grad-CAM improves interpretability by providing a visual means of explaining the predictions of the model and thus making the system more transparent and clinically reliable. Nevertheless experimental results show that performance can be hampered by issues such as class imbalance, similarity between ultrasound image planes, and noise in medical images. These considerations can restrict the general capabilities of the model for all classes. Despite these limitations, the proposed framework provides a solid foundation for automated fetal ultrasound analysis and it can be enhanced by optimizing data distribution, preprocessing techniques, and model architecture tuning.

VII. CONCLUSION

The proposed attention-based explainable deep learning framework integrates CNN-based transfer learning with attention mechanisms for the classification of fetal ultrasound image planes. ResNet's integration brings robust feature extraction, and the attention module offers more focus on the anatomical regions of interest. Grad-CAM extends interpretability as well, as this approach allows for visual interpretation of the model's predictions, providing more transparency and clinical reliability for the system. However, as shown in the experimental results, this can potentially be compromised due to class imbalance, ultrasound image plane similarity, and noise in medical images that might affect the performance. Such factors can limit the model's generalization ability across all classes. This method provides a solid foundation that can be generalized to automated fetal ultrasound analysis, though such limitations can be minimized through optimizing data distribution, preprocessing, and model architecture tuning.

REFERENCES

[1] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2016.

[2] X. P. Burgos-Artizzu et al., "Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes," Scientific Reports, 2020.

[3] A. Sivasubramanian et al., "Efficient Feature Extraction Using Lightweight CNN Attention-Based Deep Learning Architectures for Ultrasound Fetal Plane Classification," 2024.

[4] A. S. (Additional survey), "A review on deep-learning algorithms for fetal ultrasound-image analysis," Medical Image Analysis, 2023.

[5] O. Y. Laitinen, "Uncertainty-Calibrated Explainable AI for Fetal Ultrasound Plane Classification," 2026.