

An Optimized Deep Learning Framework for Fetal Plane Classification using Hybrid Attention-Based Models

Harismithaa L R.^{1*}, Sudha Sadasivam G.²

¹Research Scholar, ²Professor, Department of Computer Science and Engineering, PSG College of Technology, Coimbatore, India.

E-mail: ^{1*}lrh.cse@psgtech.ac.in, ²sudhasadhasivampsg@gmail.com

Orcid ID: ^{1*}0009-0000-1537-8784, ²0000-0001-6845-1869

Abstract

Fetal plane classification is the foremost step in analyzing fetal health in ultrasound screening. Standard planes from fetal ultrasound scans are used to estimate fetal biometric parameters, such as femur length and overall fetal growth. This work develops an automated method for identifying standard and non-standard fetal anatomical planes in 2D ultrasound images to assist sonographers and reduce operator workload and variability. A pre-trained neural network is selected and trained for the fetal anatomy classification task. The trained network is integrated with the convolutional block attention module to improve the classification and localization of important features in ultrasound images. Further, to minimize computational demands, an efficiency-aware integration is proposed, in which a Ghost module offsets the parameter and FLOP overhead of CBAM within an EfficientNet-B0 backbone. The results have shown improved performance, as measured by detailed metrics, with a 15% reduction in trainable parameters and a 7% reduction in FLOPs relative to the attention-only model. A detailed class-wise analysis of sensitivity and specificity highlights the efficiency of the proposed model in complex fetal planes. The synergistic deep learning framework for automated fetal standard and non-standard ultrasound plane classification using a hybrid attention mechanism exhibits notably good computational performance and improved robustness compared to existing literature. The proposed model was validated on a four-class test split of the BCNatal cohort and an independent multicentre cohort acquired in five African countries. In addition, a four-way ablation study is presented to evaluate the performance and computational efficiency on GPU and resource-constrained hardware. The proposed learning framework strives to achieve a balance between computational overhead and classification efficiency within a lightweight framework for the fetal plane classification task.

Keywords: Attention, Convolutional Block Attention Module (CBAM), Depth-wise Separable Convolutions, Fetal Plane, Ghost Module, Ultrasound Imaging.

1. Introduction

Ultrasound images are the main indicator of fetal growth and development. Fetal anatomical planes reveal the anatomy of the fetus, from which various biometric parameters can be inferred. Fetal biometrics include femur length, fetal weight, and the presence of

* Corresponding Author

Journal of Innovative Image Processing, December 2026, Volume 8, Issue 4, Pages 1557-1581

DOI: <https://doi.org/10.36548/jiip.2026.4.011>

Received: 18.08.2026, received in revised form: 21.09.2026, accepted: 13.10.2026, published: 23.10.2026

© 2026 Inventive Research Organization. This is an open access article under the Creative Commons Attribution-NonCommercial International (CC BY-NC 4.0) License

fetal anomalies. The classification of proper fetal anatomy is a crucial phase. There is a chance that ultrasound technicians can be misled in identifying the standard anatomical plane of the fetus. With the advent of artificial intelligence in the medical field, the workload of certain strenuous clinical detection tasks performed by technicians is reduced. The ongoing demand for manpower in different branches of medicine has encouraged the use of artificial intelligence in clinical decision-support systems. This research work focuses on developing a deep learning-based model for classifying anatomical features in 2D fetal ultrasound images. Ultrasound imaging is a convenient, secure, and commonly used tool for monitoring fetal development and identifying the different fetal planes. Accurate identification of standard fetal planes, such as the abdomen, thorax, brain, and femur, is a prerequisite for reliable biometric measurement, which is subsequently interpreted by the clinician. AI-driven classification systems can assist sonography technicians by providing decision support and automating repetitive, tedious tasks.

The main objective of this research work is to analyze the performance of various deep-learning models for the fetal anatomical ultrasound dataset and design optimized deep-learning models for the classification of standard and non-standard fetal anatomical planes. The following sections of the paper discuss the existing works, methodologies, the architecture of the proposed models, implementation details, the results obtained, a discussion about the performance of the models, and concluding remarks.

2. Related Work

This section discusses existing works by researchers in fetal health analysis using artificial intelligence and deep learning tools.

Prominent ultrasound classes for fetal plane classification studies were based on five main anatomical structures: the abdomen, femur, maternal cervix, fetal brain, and thorax. FPUS23 is a synthetic open-source dataset that comprises common fetal plane categories for classification. These phantom datasets can be used to enhance performance on real-world datasets through transfer learning. A modified ResNet34 pre-trained model to classify fetal anatomical features and orientations was developed with techniques like model pruning and quantization to reduce model complexity and improve model efficiency on resource-constrained devices [1].

Deep convolutional neural networks (DCNNs) can be used for the classification of fetal ultrasound planes to achieve higher performance than traditional machine learning approaches [2]. However, the main challenge faced is the homogeneous nature of real-world data. DCNN models require diverse data, including abnormal cases, to differentiate them from normal classes. The study highlights the impact of data augmentation and the challenges in accurate fetal image classification. Further, explainable AI has been applied in prenatal healthcare [3] to address the black-box nature of CNNs and help improve the interpretability of deep learning models. Deeper networks with feature extraction blocks were proposed to capture multivariate features in low and high dimensions, followed by feature reduction modules with skip connections to reduce the spatial dimensions of the feature maps [4]. Additionally, an approach of integrating deep features from pre-trained networks with fixed and variable-sized kernels using a multilayer perceptron worked well on static structural patterns [5]. However, more enhancements were required for planes with variable structural anatomy. Pretrained networks like DenseNet show notable performance in capturing deeper patterns in comparison with other pre-trained networks

[6]. EfficientNet is a breakthrough for medical image analysis that uses a compound scaling strategy to consistently increase network depth and resolution. EfficientNet has variants ranging from B0 to B7 for different use cases and is shown to achieve superior accuracy and performance efficiency compared to previous Convolutional Neural Networks [7].

Attention mechanisms have shown great promise in medical imaging. The Convolutional Block Attention Module (CBAM) by Woo et al. [25] uses sequential attention in both the channel and spatial dimensions. By highlighting key areas in intricate medical images, this technique enables networks to significantly improve performance on classification tasks. Multi-scale attention networks (MSANs) are another noteworthy advancement. They use hierarchical attention methods to concurrently capture both fine-grained and high-level patterns, and results showed that these frameworks have proved beneficial for analysing multi-resolution data [10].

Another noteworthy study explored the impact of the Convolutional Block Attention Module (CBAM) and the squeeze excitation blocks integrated into a multi-head attention structure to classify breast tissues from ultrasound and histopathological data [12]. The study claims to achieve superior accuracy compared to existing works and confirms that the attention module greatly aids in learning meaningful spatial and contextual information from the data. A parallel work in this area integrates the convolutional block attention module into a residual inception-based UNet structure for breast tumour segmentation [13]. The Attention module is claimed to be highly effective in highlighting the region of interest, achieving notable Dice and IoU scores. A similar study coupled dilated convolutions with CBAM, in order to improve the model's capacity to capture contextual information [14]. Attention modules have been found efficient even when involving medical signals, such as in heart sound signal classification [15]. Observations reveal that strategically adding attention modules into early and mid-level convolutional blocks considerably boosts the classification efficiency and the model's capacity to extract highly discriminative features by clearly separating classes.

Recently, in the literature, various studies have been reported on the EfficientNet-B0 backbone, ghost module and standard fetal planes dataset using attention [9],[11],[18],[19]. These studies collectively show that attention mechanisms and lightweight backbones work well for classifying medical ultrasound and MRI images, but they all overlook how efficiency is measured. One study claims that CBAM with an EfficientNet backbone improves both performance and Grad-CAM interpretability for locating tumours [9]. EfficientNetV2B0 with a multi-layer perceptron has been shown to reduce the parameter count to roughly forty times fewer than the ensemble and transformer methods [11]. Ghost modules and CBAM have been shown to work well together in a YOLOv5 multi-task detector, with the Ghost module alone reducing parameters while CBAM notably improves accuracy, precision and recall over a ResNet-34 recognizer [18]. However, the recent studies only measure efficiency by the number of trainable parameters and do not report floating-point operations, inference latency, or throughput on target hardware, leaving the deployment cost of the attention and lightweight modules unverified. The studies involve an image-level split on the BCNatal dataset rather than a patient-level split, which is a stricter and more reliable protocol. The evaluation methods are also incomplete, as the studies fall back on reporting the Matthews correlation coefficient, which is important for class-imbalanced medical datasets.

While recent studies examine attention mechanism integration into lightweight

environments, including MobileNet, ShuffleNet, DenseNet, and ResNet, research on the sustainable integration of attention modules into lightweight backbones is an impactful field to explore, considering the computational cost factor is limited. The overhead incurred while integrating the attention module has not been quantified in previous works. There is an acute need to develop an optimized lightweight model with reduced computational cost and enhanced classification performance.

Table 1. Summary of recent works in comparison with the proposed work

Study	Backbone Model	Attention mechanism	Parameter optimization	Computation budget	Patient-level split	External dataset validation
Prabakaran et al., 2023 [1]	Modified ResNet-34	No	Yes (Model pruning and quantization)	Yes (Parameters, FLOPs)	No	Yes
Sengodan 2024 [24]	EfficientNet-V2	Yes (CBAM)	No	Not reported	No	No
Omega Boro et al., 2025 [13]	Residual inception U-Net	Yes (CBAM)	No	Not reported	No	No
Shih et al., 2025 [14]	DRA-UNet	Yes (CBAM)	No	Not reported	No	No
Lin et al., 2025 [18]	YOLOV5	No	No	Inference Time	No	No
Govardhan et al., 2026 [9]	EfficientNet-V2	Yes (CBAM)	No	Not reported	No	No
Sivasubramanian et al., 2025 [11]	EfficientNet-B0	Yes (MHA)	Yes (MLP)	Not reported	No	No
Proposed Method	EfficientNet-B0	Yes (CBAM)	Yes (Ghost module)	Yes (Parameters, FLOPs, Inference Latency, Memory usage, Throughput)	Yes	Yes

The summary of recent works with the proposed method is provided in Table 1. Upon thorough study of existing research in the field of fetal standard and non-standard plane identification, research gaps, limitations, and opportunities were identified as shown in Table 2. In order to address the key research gaps, this paper proposes an EfficientNet-B0 framework with a Convolutional Block Attention Module (CBAM) and a ghost module for fetal anatomical plane classification. The proposed approach focuses on computational efficiency by reducing redundant features with enhanced attention for discriminative feature learning. This combination enables robust learning with an improved accuracy- efficiency trade-off within a lightweight framework.

Table 2. Research gaps identified and contributions of the proposed work

Research Gaps	How the proposed work addresses it
G1: High Computational Cost of Deep Models: Complex medical data requires deeper models to identify patterns, but this increases model size and training parameters, making deployment on end-user systems difficult	Proposes an optimized EfficientNet-B0 framework combined with a Ghost module to effectively reduce redundant features and lower the computational parameter overhead
G2: Unquantified Attention Overhead: Existing studies integrate attention mechanisms into lightweight models but fail to quantify the added computational overhead in terms of parameter count, FLOPs throughput and Inference latency on target hardware.	Synergistically integrates the Ghost module specifically to reduce the parameter overhead incurred by CBAM, demonstrating that attention quality can be maintained within a reduced computing budget. A detailed computational budget is presented for the proposed work to evaluate the computational efficiency.
G3: High Misclassification Rates in Clinical Settings: There is a need to address the false positive (Type I) and false negative (Type II) error rates to build trust in AI-based automated annotation tools for medical data.	Deploys CBAM to intensify focus on pertinent fetal anatomical plane patterns and eliminate insignificant artifacts, enabling robust discriminative feature learning and an improved accuracy-efficiency trade-off.

3. Methodologies

To overcome the research gaps and limitations, the proposed work presents a synergistic integration of the components: EfficientNet-B0, CBAM and the Ghost module. The proposed work aims to solve a specific, under-studied problem: the computational overhead added by attention modules to lightweight backbones. The primary contributions of the work are:

1. A cost-effective attention integration: The Ghost module is efficiently integrated to reduce the parameter overhead incurred by CBAM on an EfficientNet-B0 backbone. This interaction is not quantified in prior fetal-plane work. This demonstrates that attention quality can be preserved with a reduced computing budget for fetal-plane classification.
2. A complete computational cost profile: Parameters, FLOPs, Inference latency, throughput, and memory size are reported on a GPU. Prior fetal-plane attention studies primarily reported parameter counts only.
3. Empirical evidence of complementarity: A four-way ablation is presented with the components CBAM and the Ghost module. The ablation shows that their combination improves balanced accuracy, macro-F1 and MCC with a cost-effective computational budget.
4. A stricter evaluation protocol: Evaluation uses a patient-level split on the BCNatal dataset and zero-shot external testing on the African fetal plane dataset, including an explicit domain shift analysis.

3.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are widely used for image-based applications. They are designed for visual data where data patterns are significant for tasks such as image classification, segmentation and object detection. Convolutional neural networks comprise convolution layers that perform convolution operations on the input images. To extract features from the input dataset, convolutional layers apply a set of learnable filters called kernels that

perform a dot product between the neuron weights and the input image pixels in varying strides. Activation functions are paired with convolution layers to introduce nonlinearity to the network. The rectified linear unit (ReLU) is the most predominantly used activation in the hidden layers of CNNs, as it applies the non-saturating activation function given by equation (1) and eliminates the negative values from an activation map by setting them to zero [7].

$$f(x) = \max(0, x); \text{output range} : [0, \infty) \quad (1)$$

The learned patterns are stacked as activation maps and passed on to the pooling operation. Pooling layers reduce the spatial dimensions of the learned features by down-sampling. The common pooling techniques used are max-pooling and average pooling. Following the convolution block, a flatten layer is used to flatten the learned patterns into a 1D feature vector. This feature vector is then forwarded to the fully connected layer neurons for final classification using logistic activation functions, such as sigmoid and softmax. Regularization techniques such as dropout layers, L1, L2 regularization, and early stopping can be deployed to prevent overfitting and generalize patterns in the input.

3.2 Pretrained Neural Networks

Pre-trained neural network models are a breakthrough in ML and DL. They widen the opportunity for applications with limited datasets for ML and DL tasks. Pre-trained models are designed to adapt to smaller datasets and provide precise decisions based on new learning. This concept is technically termed transfer learning. These models are first developed and trained on extensive datasets from a common domain. The model can be then fine-tuned to capture patterns in application-specific data. The transfer learning approach leverages the strength of pre-trained CNNs to capture discriminative features in various domains, such as image classification and time frequency-based classification [16]. However, in certain use cases, retraining the models from scratch can assist in enhancing performance where the underlying data distribution varies over time or demography. Recent works on ensemble architectures combining multiple pretrained CNN models to exploit complementary feature representations have shown improved results [17]. However, ensemble architectures significantly increase computational complexity and training cost. Also, ensemble architectures are not suitable for real-time or resource-constrained deployment.

3.3 Attention Mechanisms

The process of attention in human cognition works by directing human concentration on important components in the visual field while filtering out the unimportant aspects in the background. This inspired the innovative idea of attention mechanisms in deep learning models.

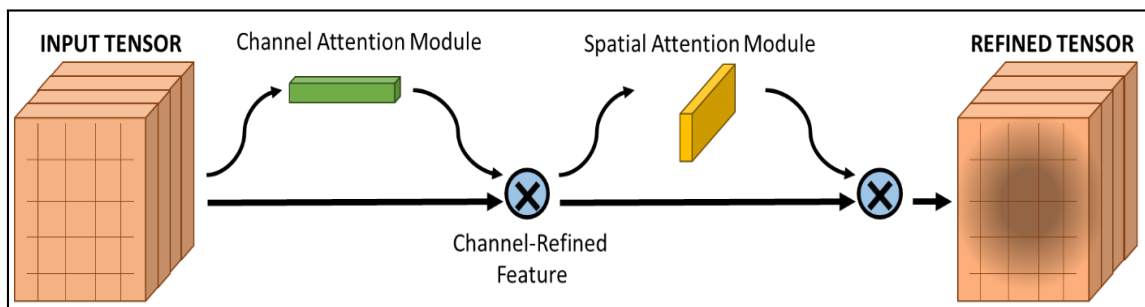


Figure 1. Convolution block attention module [25]

Attention mechanisms are notably effective techniques that allow models to focus more on important features and improve the model's decision-making process. Attention mechanisms emphasize relevant information by assigning calculated weights to significant features in the incoming sequence, which helps the model learn important patterns and dependencies in the data. Several research works have highlighted the benefits of hybrid mechanisms for critical data applications. The Convolutional Block Attention Module (CBAM) is a hybrid attention mechanism that combines two modules: channel attention and spatial attention. The channel attention module highlights the important feature channels critical for learning and the spatial attention module highlights the important spatial areas within the feature channels. CBAM falls under the soft attention category that calculates each input element's contribution to the final prediction using probabilities or attention weights. An overview of the Convolutional Block Attention Module (CBAM) by Woo et al., [25] is given in Figure 1.

4. Proposed Model Architecture

This section describes the comprehensive approach to classifying fetal ultrasound plane data. The proposed method employs a hybrid attention mechanism based on EfficientNet-B0 combined with a ghost module for cost-effective operations. The objective of the proposed architecture is to improve the model's prediction rate on real-world medical data with robust attention mechanisms and reduced computational cost. The EfficientNet-B0 pretrained model is chosen as the backbone to implement fetal ultrasound plane classification. A Convolutional Block Attention Module (CBAM) is proposed integration into the backbone model to improve the feature extraction process. CBAM proves vital in fetal plane identification, where identifying structural anomalies is critical. The proposed model integrates a ghost module to balance the computational cost of CBAM and enhance the prediction performance of the EfficientNet-B0 fetal plane classification model.

The EfficientNet-B0 pretrained model provides a compound scaling technique for computational efficiency. In the compound scaling method, network width, depth, and resolution are scaled uniformly using a compound coefficient. The hidden layers of the EfficientNet-B0 architecture are designed using MobileNet convolution (MB_conv) blocks of different configurations, which include depth-wise separable convolution operations and squeeze-excitation layers. This baseline model overcomes the drawbacks of other deep learning models by evenly scaling all three network requirements, providing efficient performance for critical applications such as medical image classification, critical object identification and image segmentation. This pretrained network is retrained on the fetal ultrasound planes dataset comprising six fetal plane classes. The following sections elaborate on the integration of the proposed methods with EfficientNet-B0.

4.1 CBAM-EfficientNet-B0 for Fetal Plane Classification

This section describes the use of the Convolutional Block Attention Module (CBAM) for enhancing the classification of fetal ultrasound planes. The channel attention and spatial attention modules of CBAM, shown in Figure 1, are elaborated below.

Channel Attention: The channel attention module highlights the more relevant channels in the incoming tensor while giving less weight to irrelevant features [23][24]. The first step in channel attention is performing two types of pooling operations: max pooling and average pooling, as given in Equation 2. Max pooling provides information about distinct

object features across channels, highlighting the strong activations across the feature maps. Average pooling provides the average value of each channel across all spatial locations in the tensor.

$$F_{max} = GMP(F); F_{avg} = GAP(F) \quad (2)$$

Where F is the input feature map. These two pooling operations are exploited together in the channel attention module to obtain better channel-wise attention descriptors and capture the global context in the fetal images. The two different channel attention maps obtained from the two pooling operations are forwarded to a multi-layered perceptron network (MLP). The MLP generates the channel attention scores $CA(F)$, which are then aggregated using element-wise addition and the sigmoid activation function(σ) given by equation 3.

$$CA(F) = \sigma(MLP(F_{max}) + MLP(F_{avg})) \quad (3)$$

Further, the resulting channel activation maps are merged with the input tensor using an element-wise multiplication operation to weight the original image feature map, F , based on the attention map activations, $CA(F)$, as in equation 4.

$$RF' = CA(F) \odot F \quad (4)$$

where RF' is the channel-refined feature representation and the element-wise multiplication operation is represented by $\odot$.

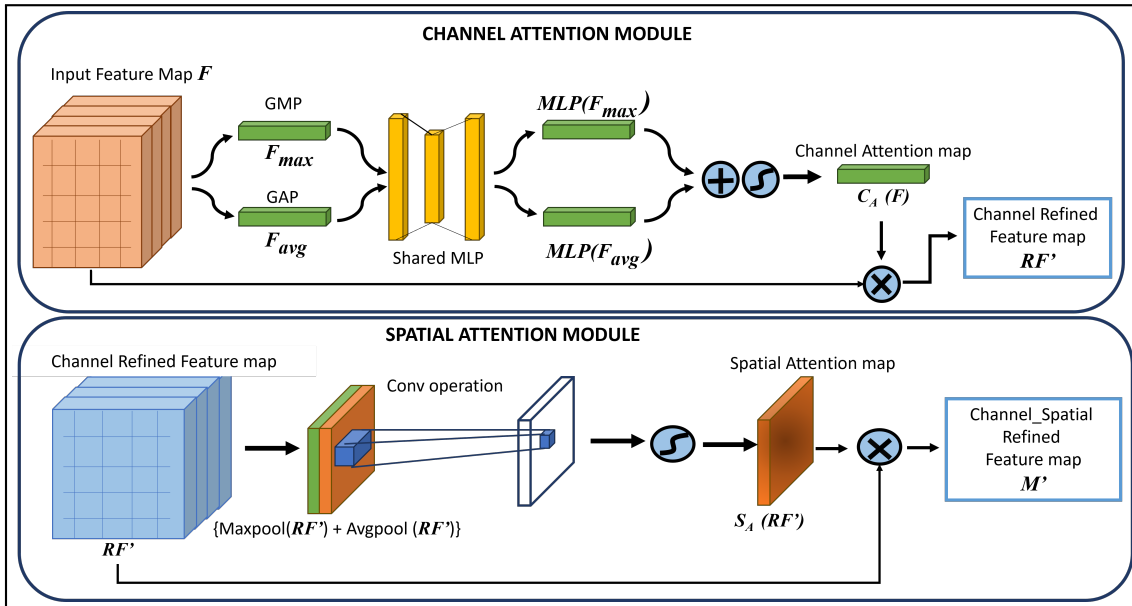


Figure 2. Channel and spatial attention modules

Spatial Attention: The spatial attention module highlights the information that is more relevant in each channel. Complementary to the channel attention module, spatial attention focuses on the spatial area where the relevant feature is present [21]. In this process, max-pooling and average pooling are first applied to the channel-refined features RF' . The feature descriptors obtained as a result of pooling operations are concatenated to effectively highlight the informative regions. The concatenated feature descriptor is subjected to a convolutional operation with a 7x7 kernel size and sigmoid activation σ , providing a 2D spatial refined attention map, $SA(RF')$ in equation 5. The spatially refined feature map is then merged with the channel-refined feature set by an element-wise multiplication operation.

$$SA(RF') = \sigma(f7 \times 7[MaxPool(RF'); AvgPool(RF')]) \quad (5)$$

$$M' = SA(RF') \odot RF' \quad (6)$$

Where M' represents the output refined feature map after applying channel and spatial attention as in equation 6. Figure 2 provides a detailed workflow of the channel and spatial attention modules.

4.2 Proposed Model: G-CBAM-EfficientNet-B0

The proposed hybrid model architecture integrates a ghost module with the Convolutional Block Attention Module (CBAM) to extract more refined features while maintaining a low computational cost of operations. With the ghost module integrated with CBAM, there's room to boost accuracy with channel and spatial attention while sustaining low FLOPs. The ghost module produces output feature maps, which are “ghosts” of intrinsic feature maps obtained from cost-efficient linear transformations [20]. The ghost module primarily uses a 2D convolution operation on the input tensor to generate intrinsic feature maps. Following the conv2D operation, the ghost layer performs a depth-wise 2D convolution on the generated feature maps to produce ghost features. The depth-wise 2D convolution is considered a cheap linear operation since it applies a convolution on each input channel independently, without combining data from different channels. Thus, its computational cost is much less than an ordinary convolution. This ghost layer augments significant features and amplifies the input channels. Finally, the extracted features from the conv operation and the ghost layer are concatenated. The Ghost module may be readily integrated with existing neural networks to reduce computing costs. In the proposed method, the Ghost module replaces the final MB_conv operation, which is block 7a in the EfficientNet-B0 backbone, followed by CBAM and global average pooling. Figure 3 illustrates the Ghost module architecture.

The input feature set from the EfficientNet backbone is passed to a convolution layer with a reduced number of filters. In a point-wise convolution operation, the input feature set, F is convolved with the filters, f , as in equation 7.

$$C = F * f + b \quad (7)$$

where b is the bias term and $*$ represents a point-wise convolution operation. C is the obtained feature map after the convolution of dimension $h \times w \times n$, where h and w are the height and width of the data and n is the number of channels in the feature map. In the ghost module, the point-wise convolution operation is performed on the feature set with a reduced number of filters, f' given by equation 8.

$$Fm = F * f' + b \quad (8)$$

The 2D Convolution layer in the ghost module produces a feature set, Fm , with m intrinsic feature maps where $m < n$. The other hyperparameters are similar to arbitrary convolution operations to maintain the spatial size of the feature map. The remaining $n-m$ feature maps are obtained by performing a cost-efficient linear operation using depth-wise 2D convolution on Fm as in equation 9.

$$FG = L(Fm) \quad (9)$$

where L represents the cost-effective linear operation (depth-wise 2D convolution) on each feature of F_m and FG is the ghost feature map with the remaining features. The features from the point-wise convolution and the linear operation are concatenated to obtain the ghost module output feature map given by equation 10.

$$FGM = FG \oplus F_m \quad (10)$$

where $\oplus$ represents the concatenation operation on the feature maps. The Ghost module output feature map is then passed on to the CBAM. The attention-weighted features are then passed on to the global average pooling layer of the EfficientNet-B0 architecture, followed by dense layers and a softmax activation function for final output fetal plane classification.

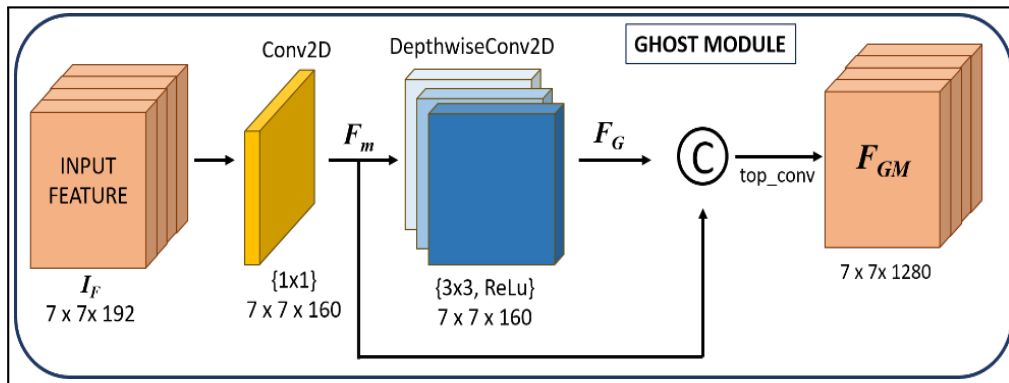


Figure 3. Ghost module architecture

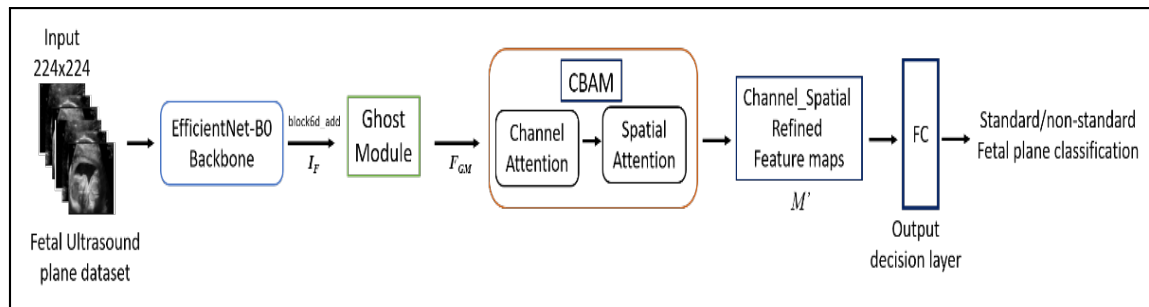


Figure 4. Workflow of the proposed hybrid G-CBAM EfficientNet-B0 architecture

CBAM aims to improve accuracy by highlighting the important features extracted by the EfficientNet-B0 backbone model in fetal ultrasound plane images, but some computational overhead is still incurred during processing. This might cause real-time delays in fetal ultrasound plane classifications, where efficient performance is expected with reduced computational cost. Increased computational overhead makes real-time deployment less efficient. This research gap is addressed in the proposed G-CBAM module for fetal plane classification. The proposed method is lightweight with reduced trainable parameters and FLOPs. It can be a promising candidate for testing in a real-time, resource-constrained environment. It is possible to scale down the upstream feature extraction with ghost Modules when integrated with CBAM. It makes CBAM's overhead cheaper by lowering parameters and FLOPs. By striking this balance, the model remains effective while retaining CBAM's attention quality. The overall architecture of the hybrid ghost module integrated with CBAM in the EfficientNet-B0 is given in Figure 4.

Placement Rationale: The last MB-Conv block (block 7a) is replaced with the ghost module for cost-effective attention integration. A standard convolution generating n feature maps from c input channels with $k \times k$ kernels over an $h \times w$ map costs on the order of $h \cdot w \cdot n \cdot c \cdot k^2$

operations. The Ghost module produces only m intrinsic maps by ordinary convolution ($m < n$) and generates the remaining $n-m$ “ghost” maps by cheap depthwise linear operations, whose per-channel cost is far smaller than a full convolution. The Ghost module first applies a 1×1 convolution that produces only C/ratio intrinsic feature maps (ratio = 2, i.e. half the channels), then generates the remaining maps with a cheap depth-wise 3×3 convolution and concatenates the two. Because the expensive point-wise convolution operates on half the channels and the rest are produced by a low-cost depth-wise operation, both the FLOPs and the parameters are reduced. The ghost module is followed by the CBAM. For an input feature tensor of shape $C \times H \times W$ ($C = 1280$, $H = W = 7$), channel attention applies global average and global max pooling to yield two $1 \times 1 \times C$ descriptors ($1 \times 1 \times 1280$ in the proposed model). Each pass through a shared MLP with a bottleneck reduction ratio of 16 ($r = C/16$) and re-weights the channels. Further, they are summed and passed through a sigmoid to give a $1 \times 1 \times C$ channel-attention map. Spatial attention then concatenates channel-pooled maps into a $2 \times H \times W$ tensor and applies a 7×7 convolution to produce a $1 \times H \times W$ spatial map.

5. Implementation of the Proposed Work

5.1 Features of the Dataset

The Fetal ultrasound plane dataset was originally collected from the fetal department of BCNatal Hospital in Spain [2]. The study involved maternal-fetal scans from expectant mothers in their second and third trimesters. This is the largest and most diverse available public dataset for fetal ultrasound plane images. The data were acquired according to proper ethical guidelines and with patient consent for experimental studies. A sample of the fetal plane data is given in Figure 5. The dataset consists of 12,400 ultrasound images collected from 1792 patients and annotated into six fetal plane classes. Each patient is typically scanned for several anatomical planes; consequently, the per-plane counts sum to 4,837 patient-plane instances across 1,792 unique patients. The dataset was split into a 70:30 ratio for training and testing, respectively. Since the dataset was collected in real time, it was imbalanced in nature. Hence, data augmentation was carried out to balance the dataset. The distribution of the dataset is given in Table 3.

5.2 Data Pre-processing

Ultrasound images carry critical medical data and might need intensive pre-processing to ease the work of the training model in extracting better patterns in the data. The implementation of a systematic data preparation pipeline was necessary to overcome the difficulties presented by fetal plane ultrasound image data. The fetal ultrasound dataset was subjected to the following pre-processing techniques:

- **Data Normalization:** All ultrasound images were resized to 224×224 pixels to match the EfficientNet-B0 input resolution. The pixel values in the images were normalized to the range $[0, 1]$ to guarantee consistency in the distribution of input data.
- **Train-Test split selection:** To prevent overlap between training and testing sets, patient-wise splitting was implemented in a 70:30 ratio on 1792 patients. Patient data were separated into non-overlapping training and test groups before model input. The training set comprised ultrasound planes from 1254 patients, and the test set comprised data from 538 patients. The dataset consists of annotated 2D fetal ultrasound images aligned with anatomical planes, excluding samples with significant

artifacts or poor anatomical views. Mild ultrasound noise patterns were retained in the training set to facilitate the learning of noise-invariant representations.

- **Data Augmentation:** To address the pronounced class imbalance in the fetal planes dataset, a data-level balancing strategy was applied after patient-level train-test data partitioning. Data Augmentation techniques such as rotation, scaling, and random flipping were carried out to reduce the risk of overfitting and to handle data imbalance. The exact range for the augmentations is as follows: rotation $\pm 15^\circ$, width/height shift $\pm 10\%$, zoom(scaling) $\pm 10\%$, horizontal flip with probability 0.5, and fill_mode='nearest'. Vertical flipping was not used as it is anatomically inappropriate for these planes. The aforementioned augmentation techniques were applied to the training data, resulting in a balanced training set of 3000 images per class. Thus, the data imbalance was mitigated in the training set.
- **Label Encoding:** One-hot encoding is used for label encoding. It is a vital preprocessing technique that converts the categorical label data into numerical data understandable by the learning models. By using one-hot encoding, the model can handle each category separately and any implicit assumptions about the order of classes are avoided.

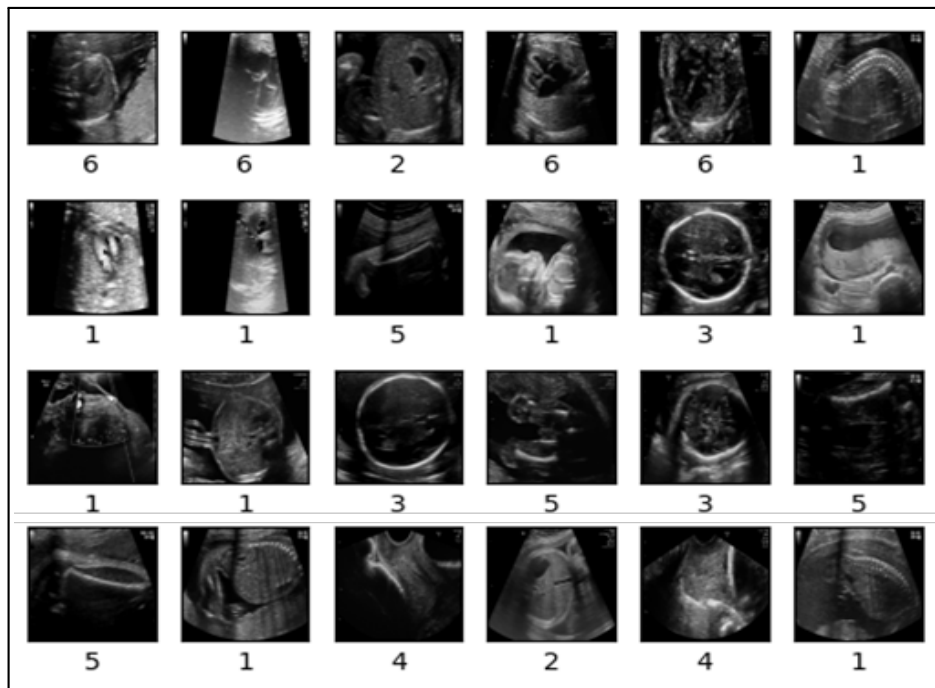


Figure 5. A sample view of fetal plane classes in the Fetal ultrasound plane dataset

5.3 Model Training and Optimization

The implementation uses TensorFlow with the Keras API. The proposed model was trained and tested on an NVIDIA RTX 4080 GPU workstation. The system was equipped with 32GB RAM and 16GB of GDDR6X VRAM with a 256-bit memory interface. This helps in processing large datasets and an intensive DL model pipeline in a faster and more efficient way. The EfficientNet-B0 model is originally designed for RGB images. Since the fetal ultrasound plane dataset comprises grayscale images, channel replication of input fetal plane images was carried out during data loading. The model was trained with selected hyperparameters based

on an experimental approach. The ReLU activation was used in the CBAM and dense layers of the model and the Swish function in the EfficientNet-B0 backbone model, while the Adam optimizer was used in the final training with a learning rate of 0.001. Categorical cross-entropy was used for classifying multiple classes of fetal planes.

Hyperparameters were tuned by grid search over batch size {16, 32, 64}, dropout {0.3, 0.4, 0.5}, and learning rate {1e-2, 1e-3, 1e-4}, selecting the configuration that minimized validation loss. Dropout was added to the fully connected layers to avoid overfitting and to enhance generalizability. Upon observing the impact on final classification, a range of dropout rates was tried and an optimal rate of 0.5 was fixed. In addition to dropout, early stopping was included to monitor validation loss with a patience of 5 epochs to prevent overfitting. After testing 16, 32, and 64 batch sizes, a batch size of 64 was selected with a learning rate of 1e-3 and the model was trained up to 50 epochs. The dataset was split in the optimal ratio of 70:30 for training and testing, respectively, with the validation split set to 20% of the training data.

Table 3. Features of the fetal ultrasound planes dataset

Fetal Plane	Clinical Use	No. of Images	Patients examined per plane*
Abdomen	Morphology, Fetal weight	711	595
Brain	Brain development	3092	1082
Maternal Cervix	Prematurity in growth	1626	917
Femur	Fetal physical development, Fetal weight	1040	754
Thorax	Heart and lung Development	1718	755
Other Artifacts	Other artifacts and incorrect anatomical structures	4213	734

*Each patient is scanned across multiple planes; per-plane totals therefore exceed the 1,792 unique patients.

6. Results and Discussion

6.1 Evaluation Metrics

The results obtained from the hybrid attention mechanism-based deep learning model for fetal ultrasound plane classification are given in this section. The performance metrics used for evaluating the model are accuracy, F1-score and Mathew's correlation coefficient (MCC). Accuracy is given by the ratio of the correct predictions by the model to the total predictions. The F1-score provides a balanced value between false positives and false negatives by integrating precision and recall into a single statistical measure. The Mathew's correlation coefficient (MCC) is a balanced metric that provides the statistical accuracy considering all the categories of the confusion matrix: TP, TN, FP, FN and provides a single value that evaluates the overall performance of the model. The formulas for the performance metrics are given by equations 11,12 and 13. Additionally, the macro-averaged F1-score and per-class AUC have been used for evaluating the robustness. These metrics treat all classes equally, regardless of support, preventing minority-class performance from being eclipsed by majority-class performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$Per - Class F1 score = 2 * \frac{Precision_i * Recall_i}{Precision_i + Recall_i} \quad (12)$$

For the six-class fetal plane classification task, the Matthews correlation coefficient

is computed in its multiclass generalisation (Gorodkin's R_K), which reduces to the familiar two-class expression when $K = 2$. Table 4 reports all aggregate metrics, including balanced accuracy, macro-averaged F1 and MCC were computed from prediction vectors of each model using scikit-learn. Multi-class MCC is used to evaluate the 6-class fetal plane classification given by the equation:

$$Multi-class MCC = \frac{c * s - \sum_k p_k t_k}{\sqrt{s^2 - \sum_k p_k^2} \sqrt{s^2 - \sum_k t_k^2}} \quad (13)$$

Where k is the total number of classes and c_{ij} represents the number of classes truly belonging to class i but predicted as class j . Further, the formula can be expanded for s, c, p_k, t_k as: s is the total number of samples given by $\sum_i \sum_j c_{ij}$; c is the total number of correctly predicted samples given by $\sum_k c_{kk}$; p_k is the number of true occurrences of class k given by $\sum_j c_{kj}$ and t_k is the total number of predictions for class k given by $\sum_i c_{ik}$.

6.2 Experimental Results

The results obtained from all the models experimented with for fetal ultrasound plane classification are shown in Table 4. Starting with a baseline CNN model with 16 layers, variants of ResNet and DenseNet were also executed and the results recorded. Among all the other pretrained models, the EfficientNet-B0 variant showed superior performance for the fetal planes data. Therefore, the EfficientNet-B0 model was selected for integration with the Convolutional Block Attention Module, which demonstrates performance enhancement in terms of accuracy. In addition to validation accuracy, test accuracy, balanced accuracy, Macro F1 and MCC values are reported for the existing and proposed models in Table 4. It can be noted from Table 4 that there is an increase in parameter size in the CBAM-integrated module. To reduce the overhead incurred by the CBAM module, a ghost module is integrated to balance the parameter increase. The tabulated results show a 15% reduction in trainable parameter size in the G-CBAM hybrid model. The results from Table 4 show that the proposed G-CBAM model achieves a balanced accuracy of 92.8% with 89.7% and 99.5% accuracy on the mutually exclusive test set and validation set.

Table 4. Comparison of results obtained from the existing and the proposed models

Model	Accuracy			Macro F1	Multi-class MCC	Params (M)
	Val	Test	Balanced			
Baseline Model 1 -CNN	0.789	0.691	0.717	0.66	0.62	11.99
ResNet-34	0.832	0.714	0.756	0.77	0.75	21.8
Resnet-50	0.851	0.79	0.81	0.68	0.74	24.1
Resnet-101	0.788	0.752	0.77	0.72	0.73	43.1
Densenet-169	0.937	0.80	0.89	0.82	0.79	14.5
Densenet-201	0.943	0.87	0.879	0.88	0.80	20.2
MobileNet_V2	0.948	0.866	0.898	0.856	0.835	2.23
ShuffleNet_V2	0.964	0.868	0.892	0.872	0.858	1.25
EfficientNet-B0	0.973	0.87	0.911	0.857	0.844	4.38
EfficientNet_B0 with CBAM	0.984	0.87	0.892	0.858	0.843	4.59
EfficientNet_B0 with G-CBAM (Proposed)	0.995	0.897	0.928	0.883	0.873	3.90

6.3 Performance Analysis

6.3.1 Class-wise Performance Analysis

The class-wise performance metrics are listed in Table 5. It can be inferred from the results that the proposed hybrid G-CBAM with EfficientNet-B0 for fetal plane classification shows notable performance and has a slight advantage over the classes denoting fetal brain, maternal cervix, and fetal femur planes in the dataset. The per-class sensitivity for the proposed model is 96.2% (Abdomen), 99.2% (Brain), 99.6% (Maternal Cervix), 91.7% (Femur), 94.6% (Thorax) and 75.3% (Other), yielding a balanced accuracy of 92.8%. Thus, the model achieves consistently high sensitivity across the five specific anatomical planes. Notably, the heterogeneous other class remains comparatively harder to recognise owing to its high intra-class variability. The CBAM-integrated EfficientNet-B0 variant shows neat performance in all classes, but the G-CBAM hybrid attention module has overall better performance. When compared to attention-integrated models, EfficientNet-B0 is observed to suffer from sensitivity despite having better precision measures. The model performance trend, confusion matrix results and per-class AUC of EfficientNet-B0 with the hybrid G-CBAM module are provided in Figure 6. The accuracy and loss graph trend shows progressive model learning during training. The loss is gradually minimized in the later epochs and is seen to follow an optimal trend.

Table 5. Class-wise performance analysis of the proposed models

Class	Model	Sensitivity	Specificity	Precision	F1
1 Other	Eff-b0	0.6809	0.9763	0.9368	0.7886
	CBAM	0.7276	0.9629	0.9099	0.8086
	G-CBAM	0.7536	0.9780	0.9463	0.8386
2 Abdomen	Eff-b0	0.9295	0.9651	0.6187	0.7429
	CBAM	0.8826	0.9734	0.6690	0.7611
	G-CBAM	0.9624	0.9735	0.6879	0.8023
3 Fetal brain	Eff-b0	0.9913	0.9903	0.9714	0.9813
	CBAM	0.9924	0.9820	0.9484	0.9699
	G-CBAM	0.9924	0.9932	0.9798	0.9861
4 Maternal Cervix	Eff-b0	0.9958	0.9894	0.9344	0.9642
	CBAM	0.9979	0.9876	0.9239	0.9595
	G-CBAM	0.9959	0.9975	0.9838	0.9898
5 Fetal Femur	Eff-b0	0.9262	0.9656	0.7118	0.8050
	CBAM	0.8269	0.9720	0.7308	0.7759
	G-CBAM	0.9167	0.9583	0.6682	0.7730
6 Thorax	Eff-b0	0.9436	0.9618	0.7993	0.8555
	CBAM	0.9262	0.9687	0.8266	0.8736
	G-CBAM	0.9456	0.9791	0.8791	0.9111

6.3.2 Comparison with the Strongest Baseline

Against the strongest baseline, the CBAM-integrated EfficientNet-B0 model, the proposed EfficientNet-B0 with G-CBAM model improves test accuracy from 0.870 to 0.897, balanced accuracy from 0.892 to 0.928, and MCC from 0.843 to 0.873. The gain is concentrated in the Femur (recall 0.827 $\rightarrow$ 0.917) and Other (0.728 $\rightarrow$ 0.754) classes, where CBAM-only regressed or improved the least. Computationally, the 15.0% parameter reduction relative to CBAM-only yields a 15.0% reduction in model memory and a 7.3% reduction in FLOPs. FLOPs fall less than parameters because the substituted layer operates at 7×7 spatial

resolutions. It holds a large share of parameters but a small share of the arithmetic, which is dominated by the unchanged high-resolution early blocks.

6.3.3 Class Separability and Feature-space Error Analysis

The confusion matrix in Figure 6 reports the classification behaviour of the proposed model on the test set. The training pipeline was repeated over five independent runs with five different random seeds and reports accuracy with a 95% confidence interval. The model correctly classified the large majority of samples, achieving a balanced accuracy of 92.8%. The 95% CI for the proposed model is 88.72% to 90.62%, with the test accuracy at 89.7%. Discrimination was near-perfect for the Brain and Maternal Cervix planes, with per-class recall of 99.2% and 99.6%, respectively, and strong for Thorax (94.6%) and Abdomen (96.2%). The other class attained a recall of only 75.3%. This observation is visible in the confusion matrix in Figure 6.

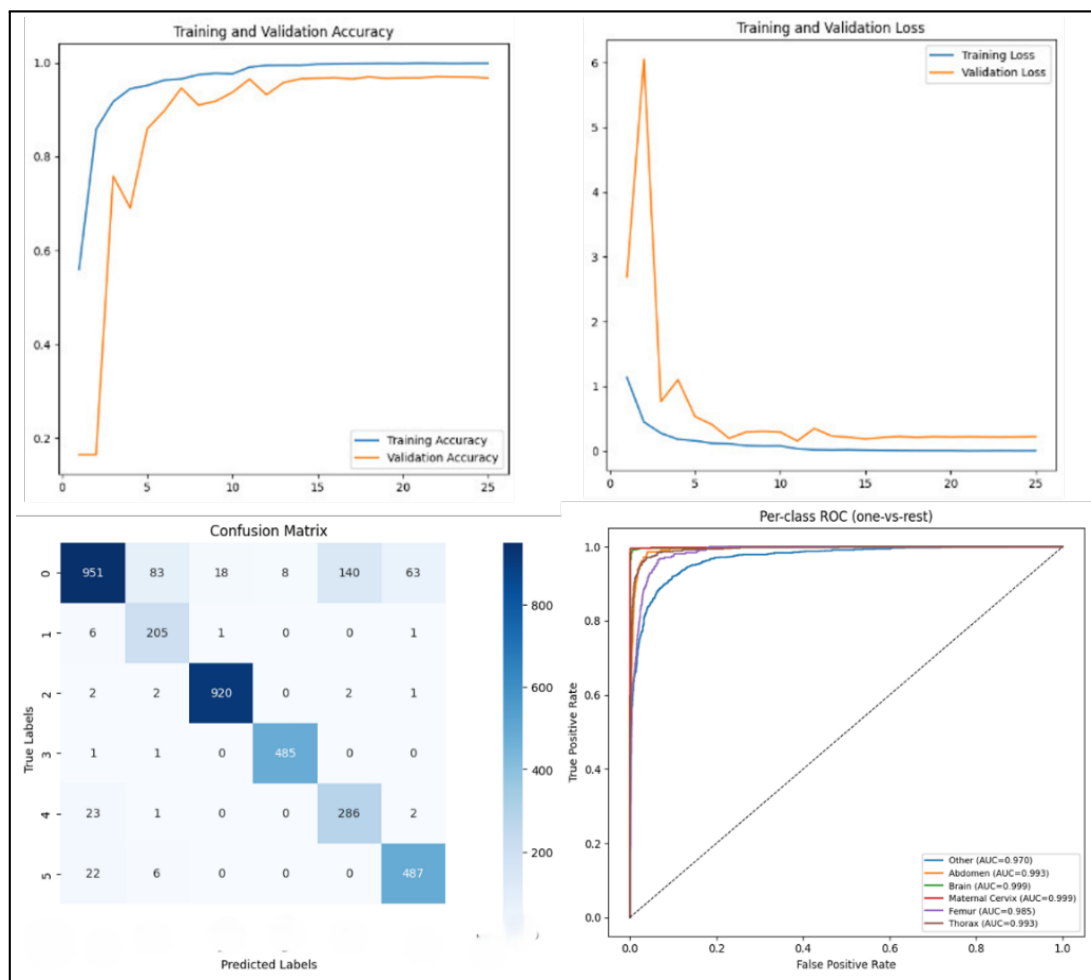


Figure 6. Hybrid G-CBAM - EfficientNetB0 model performance graph, confusion matrix and per-class AUC-ROC

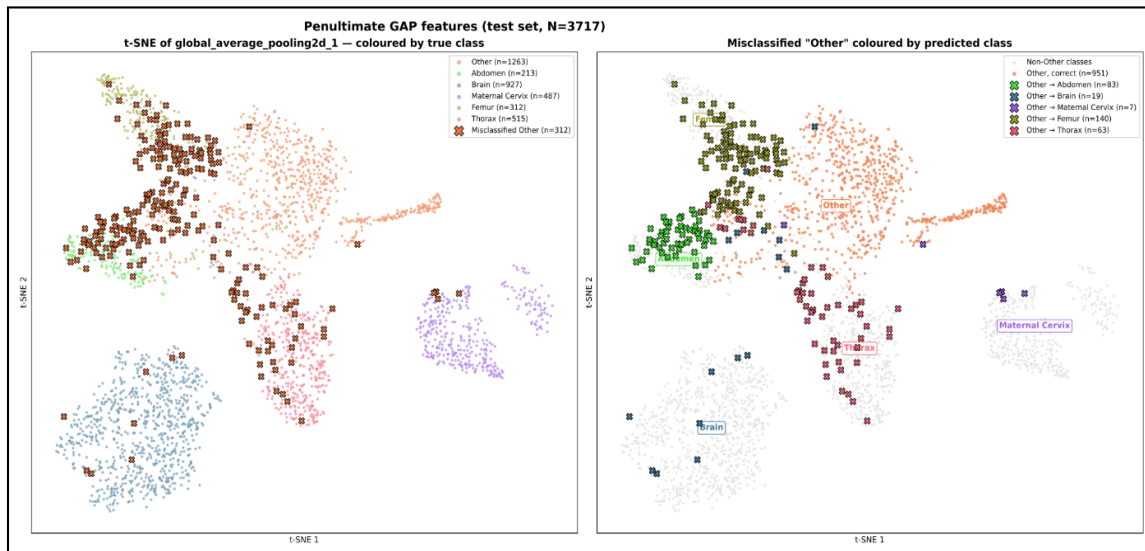


Figure 7. Feature-space error analysis using t-SNE plots of the test set images (N=3717). The misclassified other class samples (n=312) are shown in the left plot, and the colour-differentiated other class samples absorbed into specific classes (abdomen: n=83, brain: n=19, thorax: n=63, maternal cervix: n=7 and femur: n=140) are shown in the right plot

As observed from Figure 6, in the per-class ROC, all six classes achieved well-balanced discrimination, with AUCs of 0.999 for Brain and Maternal Cervix, 0.993 for Abdomen and Thorax, 0.985 for Femur, and 0.970 for Other, corresponding to a macro-averaged AUC of approximately 0.99. Each curve rises steeply toward the upper-left corner, indicating that high sensitivity is attained at very low false-positive rates. The marginally lower value for the other class is due to its heterogeneous, diverse artefacts data, which increases intra-class variability. These results confirm that the model maintains strong per-class separability under the test distribution.

The 312 instances were misassigned, principally to Femur (140), Abdomen (83) and Thorax (63). Relatively few images from the specific anatomical planes were misclassified as the other class, whereas a substantial fraction of true "Other" images was absorbed into a specific plane. This indicates that the residual error arises from the intrinsic visual variability of the artefact category rather than from confusion among the distinct anatomical planes, which the model separates reliably. The residual weakness on the femur and abdomen planes is attributed to the low femur and abdomen precision, which is caused by Other class images being absorbed into those classes, as can be observed in the t-SNE plots given in Figure 7. The reason points to the anatomical setting of these classes. The femur presents as a thin, orientation-dependent shaft that occupies a small image region, and the abdomen is frequently confused with the thorax owing to overlapping soft-tissue appearance. Most misclassified samples lie at cluster boundaries rather than inside the cores of other classes. This suggests that the errors come mainly from visual similarity between planes rather than a failure of the learned representation. This can be reduced by refining the "Other" class or adding a dedicated rejection protocol.

6.3.4 Explainability and Robustness Evaluation

To assess the interpretability of the proposed G-CBAM model, Grad-CAM visualizations were generated as in Figure 8. Grad-CAM heatmaps provide visual evidence that the attention mechanism focuses on the correct anatomical region. Grad-CAM maps were generated at the final conv layer, which yielded sharply focused activations that coincided with

the salient structures in each plane. This indicates that clinically meaningful evidence is best captured with CBAM attention.

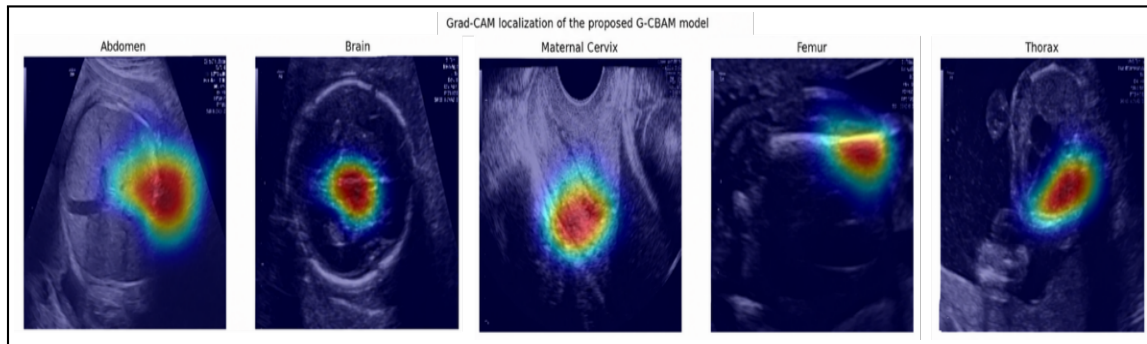


Figure 8. Grad-CAM visualization for the proposed EfficientNet-B0 G-CBAM model

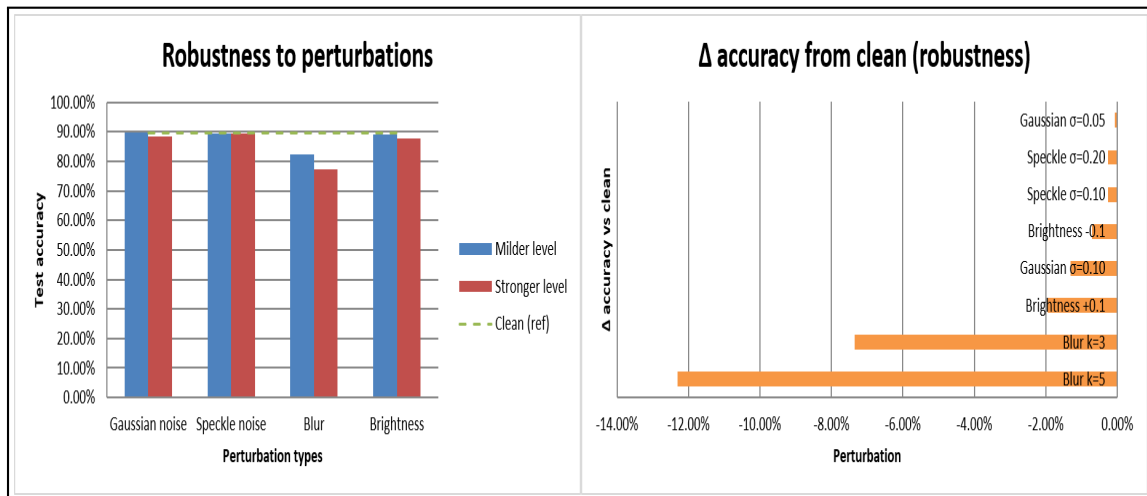


Figure 9. Noise Robustness analysis of the proposed G-CBAM model

The proposed model was evaluated under controlled perturbations on the test set without retraining to determine resilience to acquisition-related noise. The observation from Figure 9 shows that, in comparison to the clean test accuracy of 89.7%, accuracy dropped by a maximum of 1.3 percentage points under Gaussian noise (σ up to 0.10), by 0.2 points under speckle noise (σ up to 0.20), and by a maximum of 1.9 points under brightness shifts of ± 0.1 . However, performance remained stable under additive noise and illumination changes but was sensitive to blur. The model's susceptibility to spatial blurring can be noted, with accuracy dropping up to 12.3 points for Gaussian blur kernels of size 3 and 5. This trend is consistent with the ultrasound plane classification task, since class-discriminative information is mostly contained in small anatomical edges and borders that low-pass filtering attenuates. These findings show that the model can withstand the noise and intensity fluctuation common to ultrasound data and signify that picture sharpness is the main quality element influencing dependability under extreme defocus blur.

6.4 Multi-center Dataset Validation and Domain Shift Analysis

To assess generalizability beyond the trained BCNatal dataset, we performed an external, multi-centre evaluation on a publicly available maternal–fetal ultrasound dataset acquired in five African countries (Malawi, Egypt, Uganda, Ghana, and Algeria) using low-resource imaging equipment [8]. Since this dataset comprises only four of the six planes in our label space (abdomen, brain, femur, and thorax), evaluation was restricted to these shared classes.

The logits of the two absent categories (maternal cervix and other) were suppressed prior to prediction using masked softmax logits. The BCNatal dataset was also tested with masked logits, and an accuracy of 98.6% was observed. Without exposure to the target data during training, the model achieved 83.5% test accuracy, a macro-averaged F1 of 0.821, and a Matthews correlation coefficient (MCC) of 0.77 on the African dataset. The comparison of African cohort results with the four-class logit masked BCNatal cohort is given in Figure 10. These results indicate that the model retains considerable discriminative capability across significantly different demographic population, ultrasound machine variability and acquisition protocols. However, the 15% degradation in accuracy can be attributed to low-resolution data collection devices, different probes and presets, operator variability, image resolution and cropping and gestational-age distribution. Sendra-Balcells [8] highlighted the generalization gap between different demographic cohorts and the benefit of transfer learning in their study. To overcome the domain shift and improve generalization across cohorts, the proposed model can be fine-tuned with the African dataset during training for better performance efficiency over wider demographics.

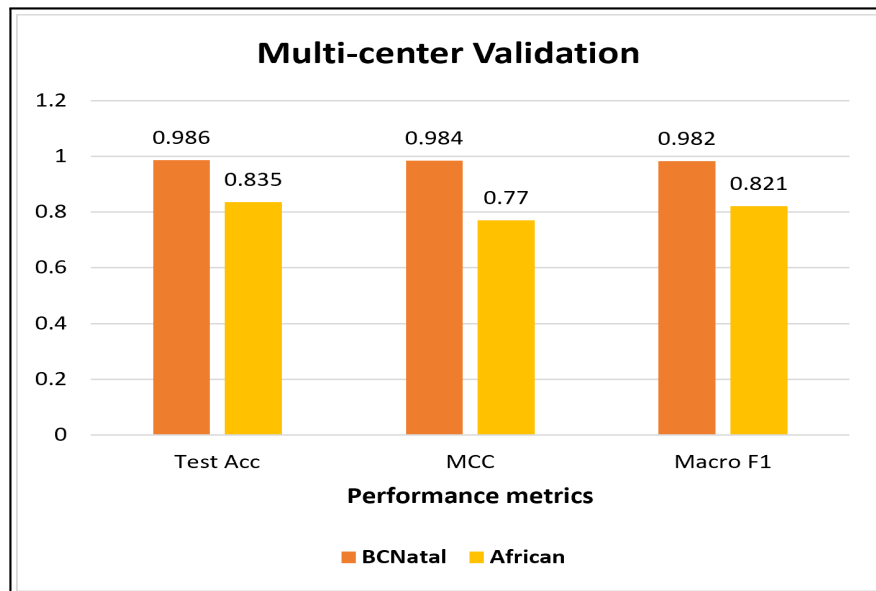


Figure 10. Multicenter validation on African demographic data

6.5 Ablation Study

To systematically analyze the contribution of each architectural component, an ablation study was conducted by progressively integrating the Ghost module and CBAM into the baseline EfficientNet-B0 model. The configurations evaluated were baseline EfficientNet-B0, EfficientNet-B0 with CBAM and the proposed EfficientNet-B0 with combined Ghost module and CBAM integration. The findings from Table 5 show that deploying the Ghost module decreases the computational workload while retaining equivalent classification accuracy, indicating its efficiency in removing duplicate representations of features.

To quantify the individual contribution of the Ghost module and the CBAM attention block, we conducted a controlled ablation evaluated on identical train and test datasets and hyperparameter settings. Four configurations were compared: (i) EfficientNet-B0 (baseline), (ii) EfficientNet-B0 with the Ghost module only, (iii) EfficientNet-B0 with CBAM only and (iv) the proposed model combining both (Ghost + CBAM). Table 6 reports the evaluated results of

performance and computational efficiency metrics on the four model configurations.

The Ghost module is responsible for enhancing the model's efficiency by consistently cutting down the computational cost by reducing the parameter count by 689K, which is a 15% reduction, when block7a is replaced. This reduction is evident regardless of CBAM inclusion. Conversely, CBAM acts mainly as a feature-refinement tool, introducing a slight computational overhead and a modest parameter increase of 206K.

It is evident from Table 4 that, among the pretrained backbones evaluated, EfficientNet-B0 offered the best accuracy for the fetal plane data and was therefore selected as the backbone for attention integration. However, integrating CBAM alone does not improve aggregate performance. The class-wise results in Table 5 show that CBAM improves recall on the heterogeneous "Other" category from 0.681 to 0.728, but this comes at the cost of recall on the Femur plane and an increase in the trainable parameters.

The Ghost module was introduced to address both issues. The combined G-CBAM configuration recovers Femur recall to 0.917 and a simultaneous gain in balanced accuracy (0.928), macro-F1 (0.883) and MCC (0.873), with the smallest parameter count of the three configurations (3.90 M). The two modules are therefore complementary, where CBAM refines the feature representation and the Ghost module provides compact, less redundant features upon which the attention refinement operates more effectively.

Table 6. Four-way ablation study on performance and computational efficiency of the proposed model vs individual baselines

Model	EfficientNet-B0	Ghost-only	CBAM only	G-CBAM (proposed)
Test Acc	0.87	0.87	0.87	0.897
Bal Acc	0.911	0.911	0.892	0.928
Macro F1	0.857	0.857	0.858	0.883
MCC	0.844	0.843	0.843	0.873
Param (M)	4.38	3.69	4.59	3.9
FLOPs	0.77	0.7138	0.7706	0.7147
Mem use (mb)	16.7	14.07	17.49	14.86
GPU latency (batch 1, ms)	1.52	1.68	1.62	1.78
GPU latency (batch 64, ms/batch)	18.3	18.6	18.9	19.1
GPU throughput (img/sec)	~3500	~3440	~3390	~3340
CPU Latency (batch 1, 4 threads, ms)	4.8	4.6	4.9	4.7
CPU throughput (img/sec)	~188	~197	~183	~191
Edge Latency (Android mobile, batch 1, ms)	7.7	7.4	8.0	7.6
Edge throughput (Android mobile, img/sec)	~110	~115	~107	~112

Further, the FLOPs, inference latency, throughput (FPS) and memory usage also show that the proposed model is superior to the other variants. The edge suitability was verified by deploying on a resource-constrained laptop with an Intel Core i5 and 16 GB RAM and on an Android mobile edge device. The model was converted into a TFLite FP32 file and was tested for latency and throughput. The inference latency and throughput results were observed on all

four variants, including the proposed G-CBAM model. G-CBAM achieved real-time inference on both platforms with 1.78 ms per image on the RTX 4080 GPU (about 3,340 images/s at batch 64) and 6.2 ms per image on the i7-1260P CPU (about 191 images/s). On the GPU, latency was slightly higher than EfficientNet-B0 (+0.26 ms) and throughput was about 4.6% lower. This is because the additional Ghost and attention operations increase kernel-launch overhead, which dominates small-batch GPU execution. On the CPU, where latency scales with arithmetic cost, the 7.2% FLOP reduction made G-CBAM 2.1% faster than EfficientNet-B0 and 4.1% faster than CBAM-only integration. On the Android mobile edge device, G-CBAM runs the TFLite FP32 model in about 7.6 ms per image (about 112 images/s), which shows that the Ghost network absorbs the cost of the attention module on mobile hardware too. Overall, G-CBAM delivers a 2.7 percentage-point accuracy gain with computational efficiency comparable to the baselines. Thus, in terms of performance and computational efficiency, it can be noted that the two modules work well together: CBAM enhances representational quality, whereas the Ghost module improves parameter and FLOP efficiency.

6.6 Discussion

Deep learning models typically require large datasets, but combining lightweight attention modules with deep networks can achieve ideal performance. This study addresses the lack of automation in fetal ultrasound classification by providing a deep learning solution that helps sonographers identify ultrasound data efficiently with reduced manpower and operator variability. The method combines ghost modules and attention mechanisms for effective feature learning, delivering competitive performance with lower computational costs for clinical use. Results showed that fetal brain, femur and maternal cervix anatomies were classified more accurately due to their distinctive patterns and clear organ outlines, which facilitated network learning. However, planes with subtle anatomical structures remain challenging due to reduced organ visibility and varying fetal positions. These findings align with previous research showing that planes with distinct anatomical outlines outperform those with varying features [4][5].

AI-based fetal ultrasound evaluations are intended to provide assistance to sonographers, reducing operator workload and variability. Compared to ensemble and transfer learning models, the proposed approach prioritizes computational efficiency, achieving competitive performance without relying on multiple heavyweight backbones. The results show that EfficientNet-B0 with a lightweight hybrid G-CBAM attention mechanism compares favourably with the feature-fusion CNN architectures by Srivastava et al., [4] and Rauf et al., [22] in terms of performance. From Table 6, it can be inferred that the parameter size of the proposed model is 3.90M, which is lower than the 4.59M parameters of EfficientNet-B0 with CBAM-only integration. This constitutes a 15% reduction in parameter size. Further, the ghost module constitutes a 7% reduction in FLOPs compared to EfficientNet-B0 with CBAM-only integration. The work also compares the proposed lightweight architecture in terms of parameters, FLOPs, latency and throughput efficiency relative to the CBAM-only and baseline configurations, deployed on GPU, resource-constrained CPU hardware and Android mobile edge device. Thus, the hybrid G-CBAM module with the EfficientNet-B0 backbone exhibits superiority in both performance and computational efficiency in comparison with the baseline individual models.

The primary scope of the proposed system is fetal plane classification and all downstream interpretation, such as fetal anomaly identification, is performed by the clinician.

The proposed model was evaluated on the African dataset [8] for multi-center validation. Generalization to additional demographics, equipment types and acquisition protocols remains to be established. Future research will aim for a broader validation across additional centres, ultrasound machines, and gestational-age ranges and more anatomical planes to ensure clinical generalizability.

7. Conclusion

The work aims to develop a synergistic deep learning framework for automated fetal ultrasound plane classification using a hybrid attention mechanism to assist sonographers and reduce operator workload and variability. The proposed model integrates a CBAM with an EfficientNet-B0 backbone to enhance model accuracy. To reduce computational overhead, a hybrid ghost module is integrated into CBAM. The proposed model was validated on a patient-level test partition of the BCNatal dataset and an independent multi-centre open-source dataset acquired in five African countries. Results show that the overhead incurred by the CBAM module is reduced by 15% in parameters and by 7% in FLOPs, while maintaining enhanced accuracy in fetal plane classification. Thus, the proposed model demonstrates enhanced performance and higher accuracy than individual baselines, indicating effective integration and edge device suitability of the Ghost and CBAM modules when deployed on GPU, resource-constrained CPU and Android mobile edge devices. External validation on a five-country African cohort yielded 83.5% accuracy, compared with 98.6% on the BCNatal internal four-class subset, revealing substantial domain-shift sensitivity. For future work, lighter framework enhancements can be designed with computationally less intensive operations for various real-time applications.

Declarations

Data Availability

The original dataset used for this study is the open-source Fetal Ultrasound Plane Dataset by Burgos-Artizzu et al., [2] and can be accessed at the link: <https://zenodo.org/records/3904280>. The African test dataset is open-sourced by Sendra-Balcells et al., [8] and is available at the link: <https://zenodo.org/records/7540448>

Author Contributions

Conceptualization, Methodology, Coding, Formal analysis, Investigation, Data curation, Visualization, Writing - Original Draft & Final draft editing: Harismithaa LR; Conceptualization supervision, Writing – Review & Overall Supervision: Dr G Sudha Sadasivam.

Funding

The authors did not receive any support or grants from any organization for the submitted work.

Conflict of Interest

The authors declare that they have no known competing interests to declare that are relevant to the content of this article.

References

- [1] Prabakaran, Bharath Srinivas, Paul Hamelmann, Erik Ostrowski, and Muhammad Shafique. "FPUS23: An Ultrasound Fetus Phantom Dataset with Deep Neural Network Evaluations for Fetus Orientations, Fetal Planes, and Anatomical Features." *IEEE Access* 11 (2023): 58308-58317.
- [2] Burgos-Artizzu, Xavier P., David Coronado-Gutiérrez, Brenda Valenzuela-Alcaraz, Elisenda Bonet-Carne, Elisenda Eixarch, Fatima Crispi, and Eduard Gratacós. "Evaluation of Deep Convolutional Neural Networks for Automatic Classification of Common Maternal Fetal Ultrasound Planes." *Scientific Reports* 10, no. 1 (2020): 10200.
- [3] Harikumar, Akshay, Simi Surendran, and S. Gargi. "Explainable AI in Deep Learning Based Classification of Fetal Ultrasound Image Planes." *Procedia Computer Science* 233 (2024): 1023-1033.
- [4] Srivastava, Somya, Ankit Vidyarthi, and Shikha Jain. "A Highly Densed Deep Neural Architecture for Classification of the Multi-Organs in Fetal Ultrasound Scans." *Neural Computing and Applications* 37, no. 28 (2025): 23619-23633.
- [5] Krishna, Thunakala Bala, and Priyanka Kokil. "Automated Classification of Common Maternal Fetal Ultrasound Planes Using Multi-Layer Perceptron with Deep Feature Integration." *Biomedical Signal Processing and Control* 86 (2023): 105283.
- [6] Ghabri, Haifa, Mohammed S. Alqahtani, Soufiene Ben Othman, Amal Al-Rasheed, Mohamed Abbas, Hassan Ali Almubarak, Hedi Sakli, and Mohamed Naceur Abdelkarim. "Transfer Learning for Accurate Fetal Organ Classification from Ultrasound Images: A Potential Tool for Maternal Healthcare Providers." *Scientific Reports* 13, no. 1 (2023): 17904.
- [7] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks." In *International conference on machine learning*, PmLR, 2019, 6105-6114.
- [8] Sendra-Balcells, Carla, Víctor M. Campello, Jordina Torrents-Barrena, Yahya Ali Ahmed, Mustafa Elattar, Benard Ohene-Botwe, Pempho Nyangulu et al. "Generalisability of Fetal Ultrasound Deep Learning Models to Low-Resource Imaging Settings in Five African Countries." *Scientific reports* 13, no. 1 (2023): 2728.
- [9] Govardhan, C., and V. Anantha Natarajan. "Interpretable and Efficient Brain Tumor Detection using Efficient Net Enhanced with CBAM and Grad-CAM Analysis." *International Journal of Artificial Intelligence and Machine Learning* 6, no. 1s (2026): 487-505.
- [10] Li, Junjie, Weiming Mi, Yucheng Guo, Xinyu Ren, Hao Fu, Tao Zhang, Hao Zou, and Zhiyong Liang. "Artificial Intelligence for Histological Subtype Classification of Breast Cancer: Combining Multi-Scale Feature Maps and the Recurrent Attention Model." *Histopathology* 80, no. 5 (2022): 836-846.
- [11] Sivasubramanian, Arrun, Divya Sasidharan, V. Sowmya, and Vinayakumar Ravi. "Efficient Feature Extraction Using Light-Weight CNN Attention-Based Deep Learning Architectures for Ultrasound Fetal Plane Classification." *Physical and Engineering Sciences in Medicine*

- 48, no. 3 (2025): 1079-1093.
- [12] Ogut, Zeki, Mucahit Karaduman, and Muhammed Yildirim. "Clinically Focused Computer-Aided Diagnosis for Breast Cancer Using Se and Cbam with Multi-Head Attention." *Tomography* 11, no. 12 (2025): 138.
 - [13] Omega Boro, Lal, and Gypsy Nandi. "CBAM-RIUnet: Breast Tumor Segmentation with Enhanced Breast Ultrasound and Test-Time Augmentation." *Ultrasonic Imaging* 47, no. 1 (2025): 24-36.
 - [14] Shih, Yi-Hsuan, Chih-Ying Wu, Ruei-Chi Lin, and Cheng-Ta Huang. "Segmentation of Tumor Regions in BUSI Breast Ultrasound Images based on DRA-UNet Model with CBAM." In *Proceedings of the 2025 7th international conference on image, video and signal processing*, pp. 105-113. 2025.
 - [15] Huai, Ximing, Lei Jiang, Chao Wang, Peng Chen, and Hanchi Li. "Heart Sound Classification Based on Convolutional Neural Network with Convolutional Block Attention Module." *Frontiers in Physiology* 16 (2025): 1596150.
 - [16] Jaba Deva Krupa, Abel, Samiappan Dhanalakshmi, Khin Wee Lai, Yongqi Tan, and Xiang Wu. "An IoMT Enabled Deep Learning Framework for Automatic Detection of Fetal QRS: A Solution to Remote Prenatal Care." *Journal of King Saud University Computer and Information Sciences* 34, no. 9 (2022): 7200-7211.
 - [17] Teoh, Jing Ru, Khairunnisa Hasikin, Khin Wee Lai, Xiang Wu, and Chong Li. "Enhancing Early Breast Cancer Diagnosis Through Automated Microcalcification Detection Using an Optimized Ensemble Deep Learning Framework." *PeerJ Computer Science* 10 (2024): e2082.
 - [18] Lin, Yirong, Xu Guo, and Peizhong Liu. "FBSPM-Net: Automatic Recognition of Fetal Brain Ultrasound Images in Early Pregnancy." In *Proceedings of the 2025 International Conference on Health Big Data*, 2025, 80-85.
 - [19] LR, Harismithaa, and G. Sudha Sadasivam. "An Intensive Review on Deep Learning Approaches in Non-Invasive Fetal Healthcare Monitoring." In *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, IEEE, 2024, 670-676.
 - [20] Han, Kai, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. "Ghostnet: More Features from Cheap Operations." In *2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, IEEE, 2020, 1577-1586.
 - [21] Chen, Long, Hanwang Zhang, Jun Xiao, Liqiang Nie, Jian Shao, Wei Liu, and Tat-Seng Chua. "Sca-cnn: Spatial and Channel-Wise Attention in Convolutional Networks for Image Captioning." In *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, IEEE, 2017, 6298-6306.
 - [22] Rauf, Fatima, Muhammad Attique Khan, Hussain M. Albarakati, Kiran Jabeen, Shrooq Alsenan, Ameer Hamza, Sokea Teng, and Yunyoung Nam. "Artificial Intelligence Assisted Common Maternal Fetal Planes Prediction from Ultrasound Images Based on Information Fusion of Customized Convolutional Neural Networks." *Frontiers in Medicine* 11 (2024): 1486995.
 - [23] Wu, Zizhang, Man Wang, Weiwei Sun, Yuchen Li, Tianhao Xu, Fan Wang, and Keke Huang.

- “CAT: Learning to Collaborate Channel and Spatial Attention from Multi-Information Fusion.” *IET Computer Vision* 17, no. 3 (2023): 309-318.
- [24] Sengodan, Naren. “Breast Cancer Histopathology Classification Using CBAM-EfficientNetV2 with Transfer Learning.” *arXiv preprint arXiv:2410.22392* (2024).
- [25] Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. “Cbam: Convolutional Block Attention Module.” In *European conference on computer vision*, Cham: Springer International Publishing, 2018, 3-19.