Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Interpretable Attention-Guided Graph Convolutional Networks for Fetal Anatomical Plane Classification using Ultrasound Images

**Faris Kateb**
King Abdulaziz University
**Rania M. Alhazmi**
King Abdulaziz University
**Sami Binyamin**
The Applied College, King Abdulaziz University
**Mona Alqassim**
Taif University
**Khalid Allehaibi**
King Abdulaziz University
**Mahmoud Ragab**
mragab@kau.edu.sa

King Abdulaziz University

**Article**

**Keywords:** Fetus, Ultrasound Images, Feature Fusion, Attention-Classification, Hyperparameter Tuning, Explainable Artificial Intelligence

**Posted Date:** May 28th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-9525341/v1

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

**Additional Declarations:** No competing interests reported.

# Interpretable Attention-Guided Graph Convolutional Networks for Fetal Anatomical Plane Classification using Ultrasound Images

**Faris Kateb[1], Rania M. Alhazmi[1], Sami Binyamin[2], Mona Alqassim[3], Khalid Allehaibi[4], and Mahmoud Ragab[1,*]**

[1]Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia; fakateb@kau.edu.sa , ralhazmi@kau.edu.sa, mragab@kau.edu.sa

[2]Computer and information technology department, The Applied College, King Abdulaziz University, Jeddah, Saudi Arabia; ssbinyamin@kau.edu.sa

[2]Department of Technology, Applied College, Taif University, Taif, Saudi Arabia; myqassim@tu.edu.sa

[3]Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia; kallehaibi@kau.edu.sa

*Corresponding author: mragab@kau.edu.sa

## Abstract

The precise classification of fetal ultrasound (US) anatomical planes is a crucial for dependable prenatal evaluation and decision-making in medical. Many metrics attained from maternal-fetal images are normally utilized for monitoring fetal development. The most widely employed biomarkers in medical training to screen fetal anomalies are fetal biometrics, prediction of fetal mass, etc. Moreover, US study procedures are time-intensive, inefficient, and prone to errors. Thus, an automated method executing this operation would improve cost efficiency and may decrease mistakes. Artificial Intelligence (AI) has experienced significant growth over the past decade, especially with the advent of deep learning (DL) and its extraordinary achievement in image classification tasks through convolutional neural networks (CNNs). In recent times, CNNs have exposed their advantage in a broad range of clinical usages, like radiology, dermatology, and/or to recognize lesions and organs from clinical images. This study proposes the Attention-Guided Graph Convolutional Network Framework for Multi-Modal Fetal Organ Classification (AGGCN-MMFOC). The proposed AGGCN-MMFOC model aims to effectively classify fetal organs captured from high-resolution ultrasound images, while providing interpretable predictions for clinically relevant anatomical planes. For accomplishing that, initially image preprocessing is performed, including image resizing and normalization, noise removal with modified anisotropic diffusion, and contrast enhancement using CLAHE. Subsequently, a learnable weighted fusion strategy is designed to integrate features from three models: EfficientNet-B7, Twins-SVT, and Inception-V3, enabling the capture of discriminative and significant features for the classification task. The fused features are then passed to the

attention-graph convolutional network for the accurate classification of six anatomical plane classes: Maternal Cervix, Fetal Abdomen, Fetal Brain, Fetal Femur, Fetal Thorax, and Other. To improve the performance of the AGGCN-MMFOC model, the parameter optimization process takes place using the Nadam optimizer for the classification model. Finally, Score-CAM is used to interpret the decision-making process of the DL approach leveraged for organ classification from ultrasound images. The experimental outcome examination of the AGGCN-MMFOC method was evaluated using various benchmark datasets. The simulation results described the improved results of the AGGCN-MMFOC system in terms of multiple metrics.

**Keywords:** Fetus; Ultrasound Images; Feature Fusion; Attention-Classification; Hyperparameter Tuning; Explainable Artificial Intelligence

## 1. Introduction

Development of the fetus is an essential phase in human evolution, in which any anomaly could lead to crucial health difficulties. The inaccuracies and subjectivity of clinical technicians and sonographers in interpreting the ultrasonography images frequently result in misdiagnoses [1]. Fetal abnormalities could be specified as fundamental anomalies in prenatal development that manifest in many key anatomical sites, like the fetal heart, lungs, central nervous system (CNS), and kidneys [2]. These abnormalities can arise during different stages of pregnancy and could be affected by various genetic and environmental factors, or an integration of both, which are known complex diseases [3]. Ultrasound (US) has developed as the primary tool for prenatal imaging diagnosis with its better performance, along with its non-radiative and non-invasive structure, real-world use, low cost, and convenience [4]. During pregnancy, prenatal US is the main significant imaging analysis for it appraises the fetal status, identifies congenital defects, and supports doctors in diagnosis through imaging the fetus and its appendages. These methods assist clinicians very quickly in intervening in the disease's progression [5]. Common US analyses for women between pregnancies could efficiently decrease the congenital disability prevalence of congenital disabilities [6]. Moreover, fetal US meets some difficulties in the medical pipeline. Some features are affecting the precision of analyses, like high fetal motion, the exorbitant abdominal layer's thickness of pregnant women, and divergences among observers, among other things [7]. Obtaining high-quality ordinary planes, precise biomedical measurements, and standard disease diagnosis are time-intensive and labour-intensive for ultrasonographers [8]. Therefore, optimization of the prenatal US analysis procedure reduces sonographers' workload and enhances medical effectiveness and consistency.

Recently, artificial intelligence (AI) has been progressively used in the fetal US domain. The appearance of AI-driven US imaging is estimated to accelerate the prenatal check-up stage [9]. AI refers to solving difficulties or obtaining knowledge, similar to the intelligence of humans. Machine learning (ML) is a subset of AI that concentrates on learning and inferring rules from collective data and decision-making, and making estimates about novel data [10]. In contrast to traditional ML, which is extremely reliant on data-expertise and data-quality, deep learning (DL) represents the major branch of ML, separates significant features from large volumes of raw data, and is exceedingly effective at image categorization, image segmentation, and target detection [11]. In recent times, it obtained popularity in clinical imaging. Convolutional neural networks (CNNs) are the main component of DL models in computer vision, which delivered extraordinary outcomes in a variety of pattern identification difficulties, object localization, classification, and segmentation [12]. CNN automatically learns a significant feature representation and combines feature representation and classification processes into a unified pipeline proficient human intervention. However, utilizing CNN with various hidden layers to separate major features from restricted training models and attain higher-performance forecasting [13].

To enhance the reliability of fetal organ classification, this study proposes a novel Attention-Guided Graph Convolutional Network Framework for Multi-Modal Fetal Organ Classification (AGGCN-MMFOC) using high-resolution US images. The goal of the model is to proficiently extract fine-grained features for precise anatomical plane classification, enhancing model explainability to support physicians in the diagnostic process. The proposed AGGCN-MMFOC model adopts a multi-faceted approach that consists of image preprocessing, feature extraction and fusion, advanced DL-based classification, model optimization, and explainability analysis. The performance assessment of the AGGCN-MMFOC system is carried out on a benchmark dataset, and the results are examined under various evaluation metrics. The core contributions of this research is as follows:

- Design a learnable weighted fusion strategy by integrating three models, namely EfficientNet-B7, twin spatially separable vision transformer **(**Twins-SVT), and Inception-V3, which extract fine-grained and discriminative features from high-resolution US images.
- Combine attention mechanism and graph convolutional network (Attn-GCN) models for robust anatomical classification, by leveraging both local image features and global spatial context.

- Utilize the Nadam optimizer for hyperparameter tuning, which helps accomplish better convergence and enhanced performance.
- Integrate Score-CAM for interpreting the DL system's decision-making process in anatomical plane classification detection from US images.
- Evaluate the AGGCN-MMFOC technique on the FETAL_PLANES_DB dataset and validate its excellence over other-art methods across diverse metrics.

## 2. Related Studies

This section presents a brief synthesis of the literature on fetal US classification techniques. Kaur and Kaur [14] proposed an automated and intelligent fetal health classification architecture using deep learning (DL) methods to forecast fetal quality by CTG-derived characteristics. With the presented methodology, examiners use different data pre-processing procedures beginning with feature standardization and continuing with class balancing via SMOTE, and then feature selection (FS) was performed by applying ANOVA F-score. The analysis pushes AI-supported prenatal diagnostic improvement and delivers basic approaches for using real-time fetal screening systems in both remote areas and clinical facilities. Ciobanu et al. [15] presented the potential of an AI framework in automatically detecting abnormalities in fetal standard scanning planes and structures, mainly concentrating on the abdominal conditions. US imageries obtained from a prospective cohort analysis have been pre-processed utilizing CV2 and Keras-Optical Character Recognition (OCR) to remove textual features and artifacts. OCR has been used to detect text and remove textual components, and then inpainting utilizing contiguous pixels. Turkan et al. [16] proposed FetalMovNet, an innovative DL approach that integrate an attention-based models to enhance the fetal activities classification in US video sequences. The method incorporates CNN architecture for feature representation and an attention-based model to learn spatial and temporal relations, substantially refining the recognition effectiveness of fetal movements. To examine the FetalMovNet, the author develops a novel dataset comprising fetal movements in US.

Krishna et al. [17] introduce an innovative method utilizing a hybrid vision transformer (H-ViT) for automated fetal anatomical plane detection to meet these concerns. The presented models use hierarchical attributes separated from DenseNet121 that are provided as an input to the ViT to examine the intricate spatial structures in fetal US imagery. Through integrating local as well as global attributes, the presented approach improves feature discrimination, as a

consequence, alleviating lower inter-class differences. Alasmawi et al. [18] presented the problems posed by the manual classification of fetal US images by presenting an unsupervised technique, the fetal US semantic clustering (FUSC) model. The main motive is to gather extensive US imageries to diverse fetal sites, decreasing or removing the necessity for time-consuming manual labeling. Rathika et al. [19] proposed a hybrid optimization method for FS and presented a new Radial Basis Function NN (RBFNN) for dependable maternal-fetal US plane detection. A huge data of maternal fetal screening US imageries has been gathered from open-access sources and classified into six categories. Feature extraction is executed utilizing the GLCM, and optimization techniques are used to identify the major relative attributes. The optimized characteristics from all models are then passed into presented and traditional DL techniques.

Ghabri et al. [20] introduce a DL algorithm for automating fetal organ detection from US images. The author validated the technique on fetal US image datasets. Author utilized a training process on labeled-datasets with signs for fetal organs like abdomen, femur, thorax, and brain, along with the maternal part of the cervical. Sarker et al. [21] examined a DL-driven image detection framework named COMFormer to detect maternal-fetal and anatomical structures of the brain in 2D fetal US images. The author presented a framework that depends on a transformer-driven model that exploits global and spatial features through employing a recently developed residual cross correlation attention (R-XCA) block. This block presents a modern XCA-based model for capturing a high-range representation by input utilizing global features.Table 1 highlights existing fetal classification models, detailing the technique, datasets, results and research gap

**Table 1** Literature review of fetal classification model

| Authors | Years | Techniques | Datasets | Results | Research gap |
|---|---|---|---|---|---|
| Kaur and Kaur [14] | 2026 | CNN and LSTM models. | CTG image dataset. | Accuracy of 95.47%. | Lack of real-time clinical validation and deployable, explainable CTG-based fetal monitoring systems. |
| Ciobanu et al. [15] | 2025 | DL model | | Accuracy of 79.89%. | Limited comparison of AI-based fetal abdominal plane detection with clinically established automatic abdominal circumference measurement methods. |

| | | | | | |
|---|---|---|---|---|---|
| Turkan et al. [16] | 2025 | LSTM and CNN methods. | Fetus US Dataset. | Accuracy of 0.9887%. | Existing studies insufficiently address model robustness and generalization of fetal movement classification across diverse US acquisition conditions. |
| Krishna et al. [17] | 2025 | ViT model | Maternal-fetal US image dataset. | Accuracy of 96.60% | Current fetal US identification approaches lack robust modeling of complex spatial relationships, resulting in reduced accuracy when anatomical classes appear visually similar. |
| Alasmawi et al. [18] | 2024 | Self-supervised learning model | Fetal US image dataset. | Achieves higher outcomes. | There is a lack of automated, unsupervised techniques capable of organizing large-scale fetal US images into meaningful anatomical views without manual annotation. |
| Rathika et al. [19] | 2024 | ANN, CNN, and RBF techniques. | Zendo dataset | Accuracy of 98.07%. | Prior studies lack an optimized and reliable framework that jointly combines effective FS and DL to accurately classify diverse maternal–fetal US planes. |
| Ghabri et al. [20] | 2023 | TL and CNN methods. | Maternal-fetal US image dataset. | Accuracy of 99.84%. | Accurately identifying fetal organs from US images is still difficult and relies on experts. |
| Sarker et al. [21] | 2023 | CNn and transformer techniques. | Fetal US dataset. | Accuracy of 96.33%. | Existing approaches do not fully leverage long-range and local features for US classification. |

## 3. Proposed AGGCN-MMFOC Architecture

The proposed AGGCN-MMFOC approach follows a systematic approach for classifying fetal organs from US images. Fig. 1 showcases the end-to-end architecture of the AGGCN-MMFOC approach. As illustrated in Fig. 1, the process includes preprocessing of input US images,

feature extraction and fusion, classification, parameter tuning, and explainability analysis as described below:

- **Image preprocessing:** The US images were pre-processed to improve quality and reduce noise before analysis. First, image resizing and normalization are performed. Second, a modified anisotropic diffusion (MAD) filter is utilized to remove noise by preserving edge details. Third, CLAHE is applied to enhance contrast and structural details.
- **Feature Extraction and Fusion:** The resultant pre-processed images are further passed to the EfficientNet-B7, Twins-SVT, and Inception-V3 models to capture relevant diagnostic features. Once the features are extracted, a learnable weighted feature fusion strategy is utilized to aggregate the features from the three models for obtaining unified representations.
- **Classification:** The extracted features are provided to the Attn-GCN model. This hybrid design allows the system to learn image features and global spatial contextual representations, ensuring precise classification.
- **Parameter Tuning:** The Nadam optimizer is adopted for model optimization, enhancing the overall performance of the Attn-GCN model, resulting in improved convergence, stability, and generalization.
- **XAI Analysis:** To ensure improved model interpretability, the Score-CAM model is incorporated to offer precise understanding into the model's predictions and enhance transparency for physicians.

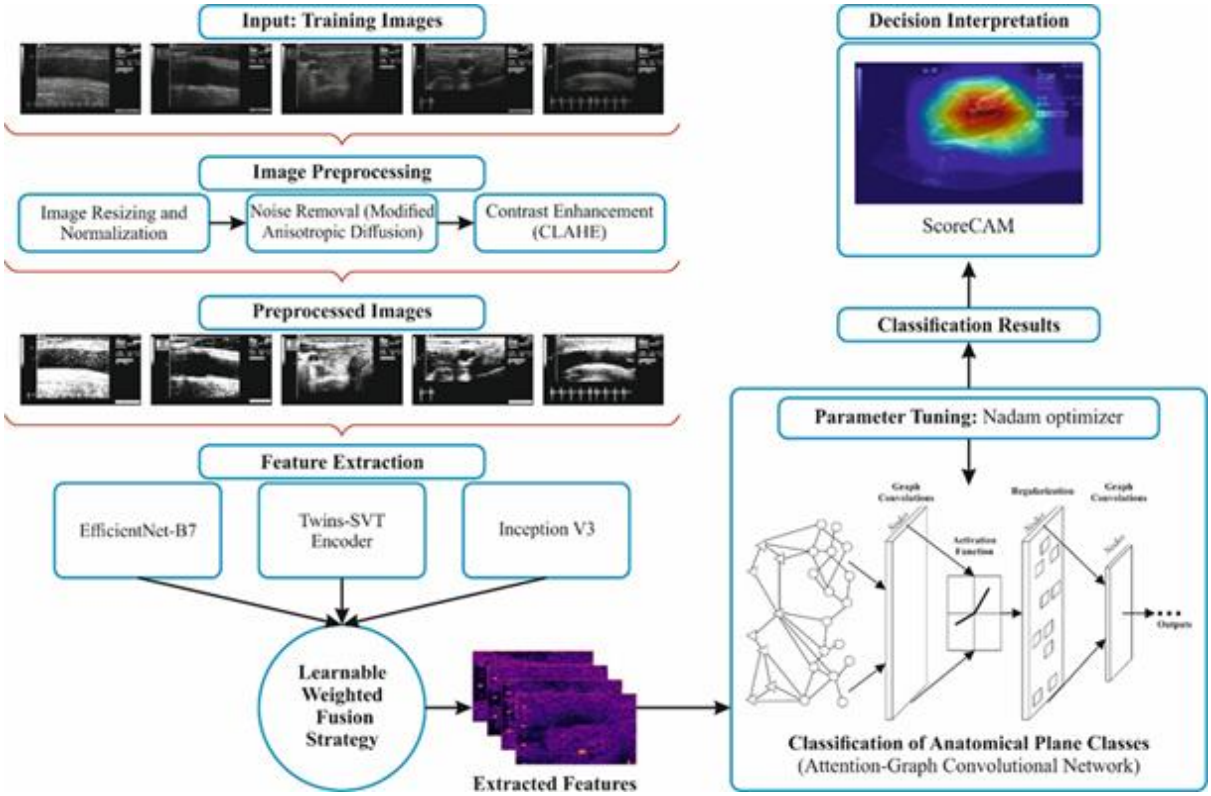



**Fig. 1.** End-to-end architecture of the AGGCN-MMFOC model

### 3.1. Ultrasound Image Preprocessing

Image preprocessing is vital for any DL model for producing high-quality results. In this study, the process involved image resizing and normalization, followed by noise removal using modified anisotropic diffusion, and contrast enhancement via CLAHE, all of which are detailed below.

**Resizing and Normalization:** Each input image can be normalized effectively and resized to a set size [22]. With the help of ensuring that the pixel intensity values are decreased in the normal range, the standardization improves the training effectiveness of model.

$$I_{\text{norm}} = \frac{I_{resized} - mean(I)}{std(I)} \tag{1}$$

This stage confirms that the images are normalized, which can be more crucial when handling the images.

**Noise removal via MAD:** In US images, the edges and local data are the primary outcomes to classify the fetal organs [23]. Then, a practical requirement of eliminating the noise, maintaining, and improving the edges for the purpose of accurate analysis of fetal diseases. In order to remove the noise, different methods are presented like nonlinear and linear filtering. Anisotropic diffusion filtering is an extensively utilized method. This method depends on heat diffusion equations and includes an edge detection and multi-scale smoothing method. The higher-level gradients of the images have been decreased because of a diffusion operation that relies on the image gradient. The diffusion operation can smooth the fatal abnormalities of the pictures that lead to similar data in all directions, and the mathematical form is represented by,

$$\frac{\partial I}{\partial t} = div(D \cdot \nabla I) \tag{2}$$

Here, the term $(\frac{\partial I}{\partial t})$ signifies the intensity modification corresponding to time, and $D$ refers to the diffusion tensor. When $D$ remains homogeneous, the diffusion is called isotropic. However, the complexity of the standard method is that the noise removal and blurred edges also occur in the images, which leads to inferior diagnosis of fatal diseases. An innovative model substitutes the conventional one, where the intensity is continuous in the occurrence of smooth images $= d(|\nabla I|)$ . The anisotropic diffusion expression can be specified by,

$$\begin{cases} d \to 0 \; for \; |\nabla I| \to \infty \\ d \to 1 \; for \; |\nabla I| \to 0 \end{cases} \tag{3}$$

Afterward, Perona and Malik defined the mathematical form of $d(|\nabla I|)$ as given below,

$$d(|\nabla I|) = \frac{1}{1 + (\frac{|\nabla I|}{K})^2} \tag{4}$$

$$d(|\nabla I|) = e^{-(\frac{|\nabla I|}{k})^2} \tag{5}$$

In this mathematical Eq. (4), it can determine the higher-level contrast edges and low-level contrast edges, and Eq. (5) explains the extensive field over the small region that signifies the diffusion scale.

**Contrast Enhancement (CE):** Contrast Limited Adaptive Histogram Equalization (CLAHE) is a local CE method that is based on AHE [24]. Whereas AHE improves the micro-level image data, it generally results in extreme noise intensification within the dark regions. CLAHE encounters these difficulties by minimizing the histogram and also a clip limit parameter,

thereby ensuring that the CE remains measured. In CLAHE, the image is separated into numerous local grids, and the AHE of every grid can be clipped through a predetermined threshold after intensity reallocation. The intensity alteration in CLAHE has been denoted by Eq. (6):

$$l_{CLAHE}(x,y) = CDF_{\gamma}\big(l(x,y)\big) \cdot (L-1) \tag{6}$$

Here, the term $CDF_{\gamma}$ means the cumulative-distribution function with clip limits $\gamma$, and $L$ signifies the number of intensity ranges.

### 3.2. Multi-Encoder Feature Extraction and Fusion

Feature extraction is performed using three advanced models, such as EfficientNet-B7, Twins-SVT, and Inception-V3, to capture high-level discriminative features crucial for classification. Then, the feature fusion is performed using a learnable weighted fusion strategy to integrate from each model to obtain unified representations.

**EfficientNet-B7:** It is a highly accurate model gained through scaling up EfficientNet-B0. In this analysis, EfficientNet-B7 has been designated as the benchmark model to attain the feature extraction process precision [25]. The framework of the MBConv model is unlike conventional residual mechanisms because its input as well as output feature maps are wider on the both sides in the middle. The MBConv component implements the batch normalization (BN) and swish activation process following convolution layers. BN could standardize the data and increase the rapidity of the system merging while training. In contrast swish activation function could present data's non-linearity and prevent over-fitting. Additionally, the SE attention from the MBConv mechanism integrates the SE attention mechanism (AM), improving the system's feature representational capacity through learning channel attentions. The SE-AM uses the SE attention to estimate the significance of every channel in maps for the present task, which is attained through below stated process.

Squeeze: The $C$x$H$x$W$ input feature map experiences global average pooling, lessening each 2D channel to a single scalar, thereby summarizing the spatial distribution of each channel into a 1x1x$C$ feature vector.

Excitation: Using a multilayer perceptron, the model maps the squeezed representation through non-linearities, using ReLU and sigmoid operations to compute activated weights.

Scale: The weights are employed for weighing all channels within feature maps through executing dot multiplication.

With these three processes, the model allocates weighted to every channel depending on its influence on classification precision, improving the efficiency of feature channels and down-weighting weak ones.

**Twins-SVT:** The proposed model Twins-SVT as another backbone for feature extraction. [26]. The Twins-SVT implements an adjustable pyramid hierarchical model that can be mainly separated into 4 phases for produce feature maps of various channels and scales. Every phase includes position encoding generators (PEGs), patch embedding, and numerous transformer encoder units. In every phase, the outcomes from the feature maps with different sizes. The patch embedding unit is primarily processed the input images. In this network architecture, original input image is represented as $I \in R^{H \times W \times C}$ and defined as a sequence of $H$x$W$ 1D vectors regarding $C$ dimension. To decrease patch tokens on convolution layers, the size of ever patch token must be improved. From first to fourth phases of the Twins-SVT architecture, there are patch token sequences of 128x128,64x64,32x32, and 16x16, with respect to sizes 64, 128, 256, and 512. The patch embedding unit is a major role of transformer encoders. The major phases of the unit that comprise the features. The input patch tokens can be primarily transformed into 3 3D feature maps. After the count of patch tokens are decreased and every token size has been elevated through convolutional and LayerNorm. In conclusion, the $3 - D$ feature maps are gained through convolution that can be transformed into 1D patch token vectors for input and also successive contextual representation generators for this operation of the model.

**Inception-V3:** Inception-V3 is selected due to its capability for extracting sophisticated features with many variations of filters with an efficient integration of various kinds of convolutional process [27]. Also, the design of InceptionV3 provides feature elimination without affecting the system potential by utilizing two (3x3) layers rather than one (5x5) convolution layers. When employing an equal number of filters, a (5x5) convolution requires approximately $25/9 \approx 2.78$ times more computation than a (3x3) convolution. The implication of unbalanced convolution outcomes in more time reduction. Additionally, rather than (3x3) convolutions, $a$ (1x3) convolution layer follows (3x1) convolution layers, which decreases the processing expenses. For (3x3) convolutions, the whole number of parameters found is 9. Conversely, a (1x3) convolution layer that uses $a$ (1x3) convolutions lead to

complete parameter count of $(3x1) + (1x3) = 6$, also minimizes the entire parameter counts number. Fig. 2 represents the model diagram of Inception-V3 architecture.

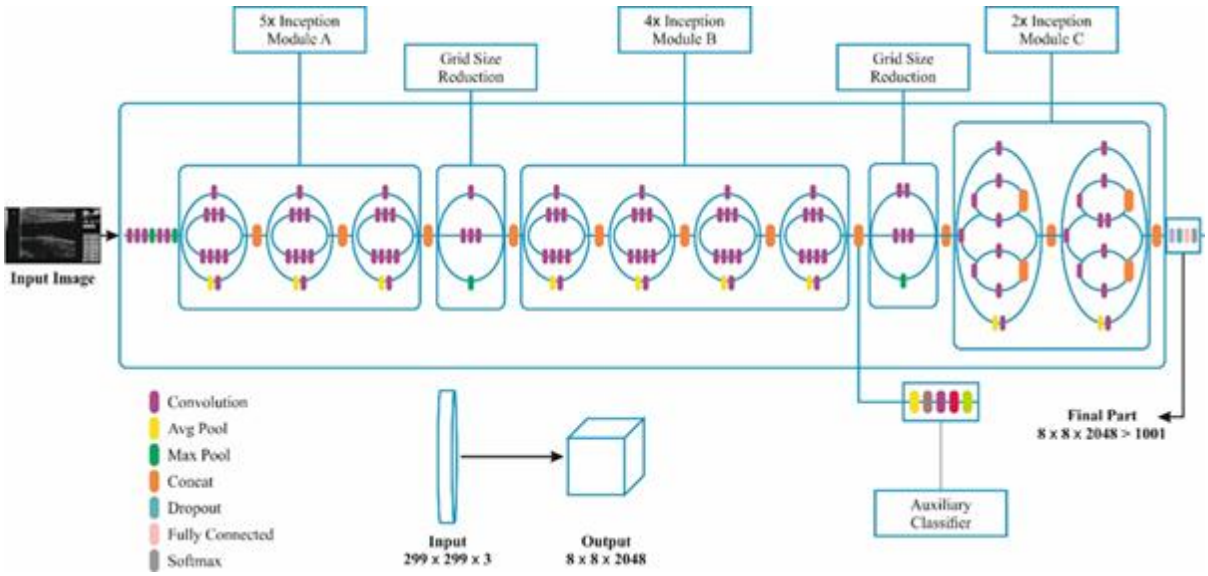



**Fig. 2.** Model diagram of Inception-V3 architecture

Regarding the processing expenses and ability, the Inception-V3 framework further used an effectual grid minimization model to assist problems of traditional pre-trained systems. Because general pretrained method employs max pooling and convolution layers, which is too materialistic, conversely, the usage of max pooling next to convolution layers are unaffordable. To overcome these concerns, Inception-V3 carried out the effective grid minimization employing discrete convolution and pooling functions. For feature representation operation, firstly, the images are input to the Inception-V3 without the classifier module, which outcomes in textural attributes, $Inception_{feature_texture}$ from final concatenate layers (mixedlO) and the transition of texture features into a $1 - D$ vector has been achieved through flattened layers that produced a result of $Y_{Inception}$ as showed in Eqs. (7) and (8):

$$Inception_{feature_texture} = Inception_{mixed10_layer(I)} \tag{7}$$

$$Y_{Inception} = \text{flatten}\left(Inception_{feature_texture}\right) \tag{8}$$

**Learnable Weighted Fusion:** In order to integrate the balancing features from the 3 encoders, namely EfficientNet-B7, Twins-SVT, and Inception-V3, a learnable weighted fusion approach has been used. Every feature map has been primarily predicted into a normal embedding space and then combined by utilizing learnable scalar weights:

$$\mathrm{F}_{fusion} = w_E \mathrm{F}_E + w_T \mathrm{F}_T + w_I \mathrm{F}_I \tag{9}$$

Where, the parameters $w_E, w_I$ , and $w_T$, describe learning parameters that can be enhanced in training process. This method permits the model to adaptively equal the role of every encoder, while maintaining the fusion lightweight method and varying in the classification phases.

### 3.3. Attention-Guided Graph-Based Classification

The fused features are fed into the Attn-GCN model for the effectual classification of anatomical planes. The Attention-GCN classification layer is formed from a basic self-AM and a GCN. This can convert the gridded feature maps into nodes, calculate the self-attention coefficient, and produce a dynamic adjacent matrix for gaining the graph-structural data [28].

Input preprocessing: The feature maps have been examined through a sliding window for generating a tensor that the network can be processed. The dimension with respect to the count of channels, spatial size, and batch size.

Data transformation: The GCN works on nodes instead of grid-like features; the tensor has been resized as a 2D node-feature format, where every individual node denotes a spatial position within the feature maps.

Linear transformation: A dense layer is mapped every individual feature vector of node into a novel space for making it in the GCN and attention mechanism.

Dynamic adjacency matrix construction: The adjacent matrix can be produced by employing attention scores that can be calculated in the node features. This matrix encodes the correlations or similarity among various nodes in the graph.

$$\text{Attention}(H) = \text{Sigmoid}(H \times H^T) \tag{10}$$

Later, the graph-structured data is produced, and the GCN layer is extracted to produce relative features.

Graph convolution: The adjacent matrix is normalized for preventing the gradient explosion or vanishing. The Multiplication of the standardized adjacent matrix alongside feature matrices permits every node to combine the attributes by using adjacent nodes. A ReLU and fully connected (FC) layer activation changes the node features and represents the following layer.

$$H^{(l+1)} = \delta\left(\widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)} + b^{(l)}\right) \tag{11}$$

Where, $\tilde{A}$ indicates the adjacent matrix with self-connections, the symbol δ refers to the activation function, $\widetilde{D}$ represents the degree matrix, $H^{(l)}$ remains node feature matrix at layer l, and $W^{(l)}$ implies weighted matrix.

Output adjustment: The output is reformed and transferred to compare the original size of features.

The Attention-GCN is incorporated with convolutional layers in the DL method. This architecture comprises 1 FC layer, 4 convolutional layers, 1 input layer, 1 max-pooling layer, and2 Attention-GCN layers. Convolutional layers are proficiently extracted local features, whereas the Attention-GCN combines node connection and gains the connectivity patterns among features. The Max-pooling decreases computational complexity and spatial dimensions,

$$M(i,j) = \max_{u=0}^{f-1} \max_{v=0}^{f-1} A(i \cdot s + u, j \cdot s + v) \tag{12}$$

In conclusion, a dropout layer (the probability ranges 0.1) has been previously included the FC layer to avoid overfitting. The Attention-GCN layers and modified convolutional layers permit the adaptive computation of node connections, thereby integrating the global and local features for increasing classification accuracy of fetal organ structures.

### 3.4. Hyperparameter Tuning Strategy

The parameter optimization process is carried out using Nadam optimizer for the classification model, resulting in enhanced generalization and convergence. The Nadam is an optimization method that integrates the adaptive learning rate method of Adam with the lookahead gradient method of Nesterov accelerated gradient (NAG) [29]. Whereas conventional momentum approaches modify the updates of parameters depending on collected gradients from prior stages. Nadam combines this expectation into Adam's adaptive update architecture, thereby enabling the parameters to converge more efficiently near the optimum outcomes. This incorporation produces rapid convergence and increased stability in the training process, particularly in non-convex optimizer issues or once handling noisy gradients. Accordingly, Nadam is widely utilized as an improved optimizer in NN training. It is paramount in enhancing training and overall model effectiveness.

$$\theta_{t+1} = \theta_t - \eta \left( \frac{\beta_1 \hat{m}_t + (1 - \beta_1) g_t}{\sqrt{\hat{v}_t} + \epsilon} \right) \tag{13}$$

where $\theta_t$implies model parameters at iteration $t$, $\hat{m}_t$and $\hat{v}_t$are the bias-corrected 1st and 2nd moment estimates, respectively, $\eta$is the learning rate, $g_t$is the gradient, $\beta_1$is the momentum coefficient, and $\epsilon$is a small constant for numerical stability.

### 3.5. Explainable AI Using Score-CAM

In order to consider that the model's predictions can be comprehensible and suitable for medical analysis, Score-CAM has been utilized effectively [30]. The Score-CAM generates the heatmaps that show medical experts precisely how AI achieved this outcome. The processes are involved as mentioned in below:

Feature Map Representation: The captured feature maps are gained after final convolution layers.

Computation of Class Scores: The involvement of the feature map is presented in the last class score $S_c$ that can be calculated by applying this mathematical expression,

$$w_k = \mathrm{R}eLU\left(\frac{\partial S_c}{\partial A_k}\right) \tag{14}$$

Heatmap Generation: The CAM is produced by the weighted summation of feature maps that can be overlapped with the source images.

$$CAM = \sum_k w_k A_k \tag{15}$$

For improving reliability and allowing real-world medical decision-making in inaccessible or undesirable conditions, the medical specialists have analyzed the major regions that influence the AI detection by overlapping the CAM under input images. The overlapping CAMs on input imageries with Score-CAM increase the comprehensibility of AI diagnosis through detecting important regions and also predictions. This method upsurges medical management by resolving the black box complexity and offering strong image identification.

## 4. Results and Evaluation

In this part, the experimental evaluation of the proposed model is provided in detail with an experimental dataset and performance metrics. The assessment shows the feasibility and applicability of the proposed model in the anatomical plane classification task.

### 4.1. Dataset Used

The performance evaluation of the proposed technique has been examined under FETAL_PLANES_DB dataset [31]. This dataset of regularly obtained maternal-fetal screening US imagery gathered from 2 various hospitals by many operators and US machines. Every image can manually categorized by a professional maternal fetal doctor. Images are separated into six classes such as 4 of the most broadly employed fetal anatomical planes (Abdomen, Thorax, Brain and Femur), the mother's cervix (broadly utilized for prematurity screening) and a common classification to comprise any other less common image plane. Fetal brain imagery are more classified into the three most general fetal brain planes (Trans-ventricular, Trans-thalamic, and Trans-cerebellum) to analyse well grain classification solution. The dataset consists 12400 records from six classes as illustrate in Table 2. Fig. 3 shows the sample US fetal images. Fig. 4 demonstrates the pre-processed images and feature map representations.

**Table 2** Details of datasets

| **Classes** | **Sample Images** |
|---|---|
| "Maternal Cervix" | 1626 |
| "Fetal Abdomen" | 711 |
| "Fetal Brain" | 3092 |
| "Fetal Femur" | 1040 |
| "Fetal Thorax" | 1718 |
| "Other" | 4213 |
| **Total** | **12400** |

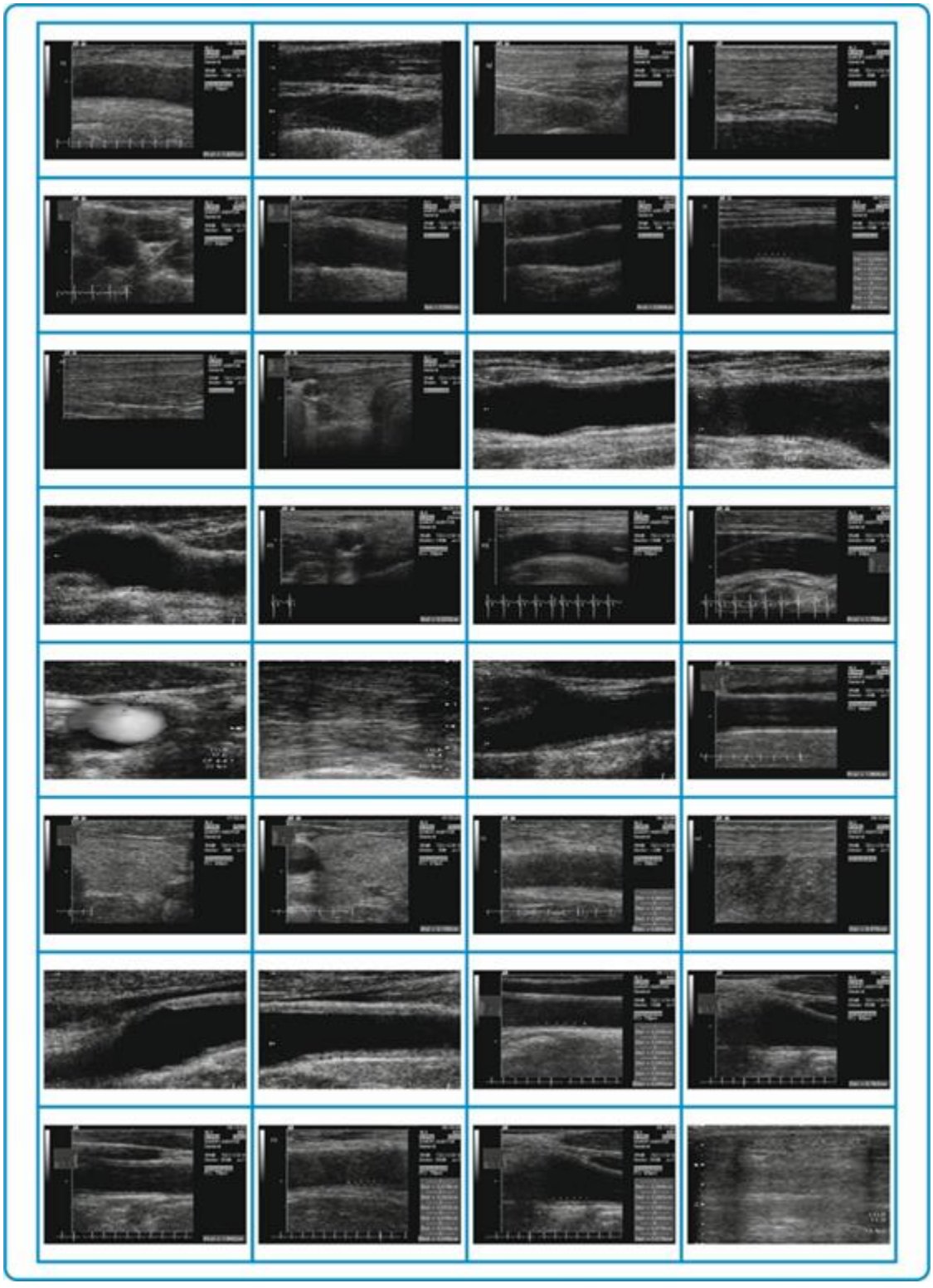


**Fig. 3.** Sample ultrasound fetal images

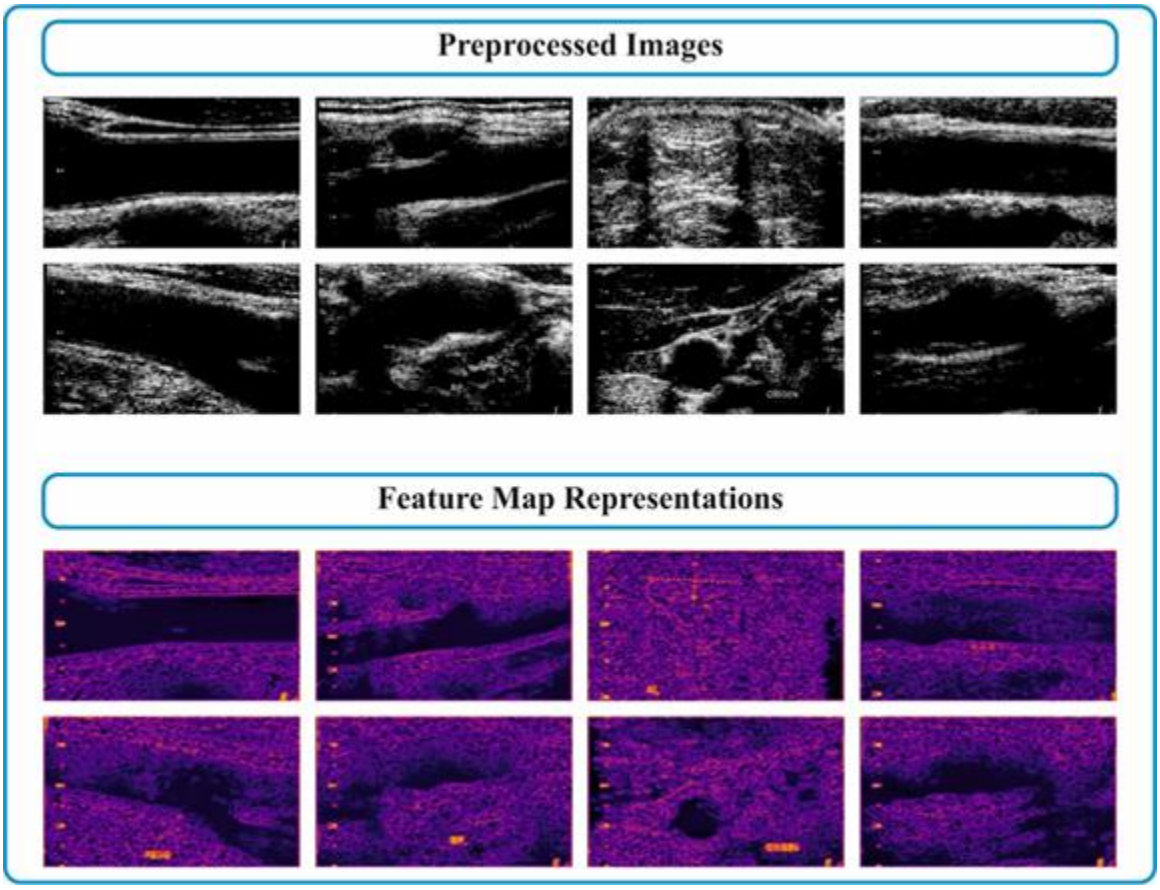



**Fig. 4.** Pre-processed images and feature map representations

### 4.2. Evaluation Metrics

In order to assess the performance of entire models, the precision, sensitivity, accuracy, MCC, F-Measure, and specificity have been effectively utilized. The mathematical forms of every metrics can be explained in below.

Now, $TP$ refers to the true positive, $TN$ denotes the true negative, $FP$ represents the false positive, and $FN$ signifies the false negative.

Accuracy is the proportion of appropriately categorized instances to the total counts of instances:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

Precision is the ratio of TP that can be divided by total predictable positives, also called as positive predictive values

$$\text{Precision} = \frac{TP}{TP + FP} \tag{17}$$

Sensitivity is the proportion of positive instances termed as true. Recall is specified as a true positive rate (TPR), the mathematical form is given by,

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{18}$$

Specificity measures the number of negative instances classified as true. It is known as a true negative rate (TNR),

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{19}$$

F-Measure is the harmonic mean of precision and sensitivity and derived in mathematically.

$$\text{F- Measure } = 2 \cdot \frac{Precision \cdot \text{sensitivity}}{Precision + \text{sensitivity}} \tag{20}$$

MCC is the relationship among the actual and predictable classifications:

$$\text{MCC} = \frac{[(TP \cdot TN) - (FP \cdot FN)]}{\sqrt{[(TP + FP)(TP + FN)(TN + FP)(TN + FN)]}} \tag{21}$$

### 4.3. Classification Performance Results

The performance evaluation of the AGGCN-MMFOC technique is examined under FETAL_PLANES_DB dataset. Fig. 5 demonstrate the confusion matrices generated by the AGGCN-MMFOC system below 80:20 TRPA/TEPA. The outcomes require that the AGGCN-MMFOC algorithm has effective identification and detection of each class accurately.

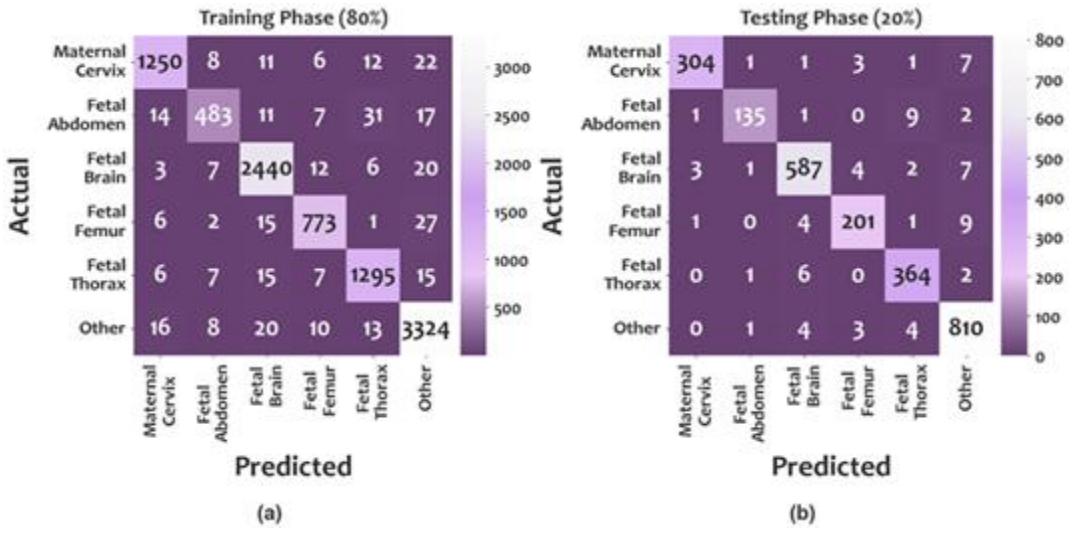



**Fig. 5.** Confusion matrices of AGGCN-MMFOC technique (a-b) 80% and 20% of TRPA/TEPA

Table 3 and Fig. 6 illustrate the fetal classifier of AGGCN-MMFOC system on 80:20. The outcomes suggest that the AGGCN-MMFOC model precisely categorized the instances. On 80% TRPA, the AGGCN-MMFOC approach provides average results with $accur_y$ of 98.81%, $preci_n$ of 95.78%, $sensi_y$ of 94.58%, $speci_y$ of 99.24%, $F_{measure}$of 95.15%, and MCC of 94.42%, respectively. Concurrently, on 20% TEPA, the AGGCN-MMFOC technique presents average results with $accur_y$ of 98.94%, $preci_n$ of 96.74%, $sensi_y$ of 95.58%, $speci_y$ of 99.31%, $F_{measure}$ of 96.14%, and MCC of 95.48%, respectively.

**Table 3** Fetal classification of AGGCN-MMFOC model with 80:20 split

| Class Labels | $Accur_y$ | $Preci_n$ | $Sensi_y$ | $Speci_y$ | $F_{measure}$ | $MCC$ |
|---|---|---|---|---|---|---|
| **Training Phase (80%)** | | | | | | |
| Maternal Cervix | 98.95 | 96.53 | 95.49 | 99.48 | 96.01 | 95.40 |
| Fetal Abdomen | 98.87 | 93.79 | 85.79 | 99.66 | 89.61 | 89.11 |
| Fetal Brain | 98.79 | 97.13 | 98.07 | 99.03 | 97.60 | 96.79 |
| Fetal Femur | 99.06 | 94.85 | 93.81 | 99.54 | 94.33 | 93.82 |
| Fetal Thorax | 98.86 | 95.36 | 96.28 | 99.27 | 95.82 | 95.16 |
| Other | 98.31 | 97.05 | 98.02 | 98.45 | 97.54 | 96.25 |
| **Average** | **98.81** | **95.78** | **94.58** | **99.24** | **95.15** | **94.42** |
| **Testing Phase (20%)** | | | | | | |
| Maternal Cervix | 99.27 | 98.38 | 95.90 | 99.77 | 97.12 | 96.72 |
| Fetal Abdomen | 99.31 | 97.12 | 91.22 | 99.83 | 94.08 | 93.77 |
| Fetal Brain | 98.67 | 97.35 | 97.19 | 99.15 | 97.27 | 96.39 |
| Fetal Femur | 98.99 | 95.26 | 93.06 | 99.56 | 94.15 | 93.60 |
| Fetal Thorax | 98.95 | 95.54 | 97.59 | 99.19 | 96.55 | 95.94 |
| Other | 98.43 | 96.77 | 98.54 | 98.37 | 97.65 | 96.48 |
| **Average** | **98.94** | **96.74** | **95.58** | **99.31** | **96.14** | **95.48** |

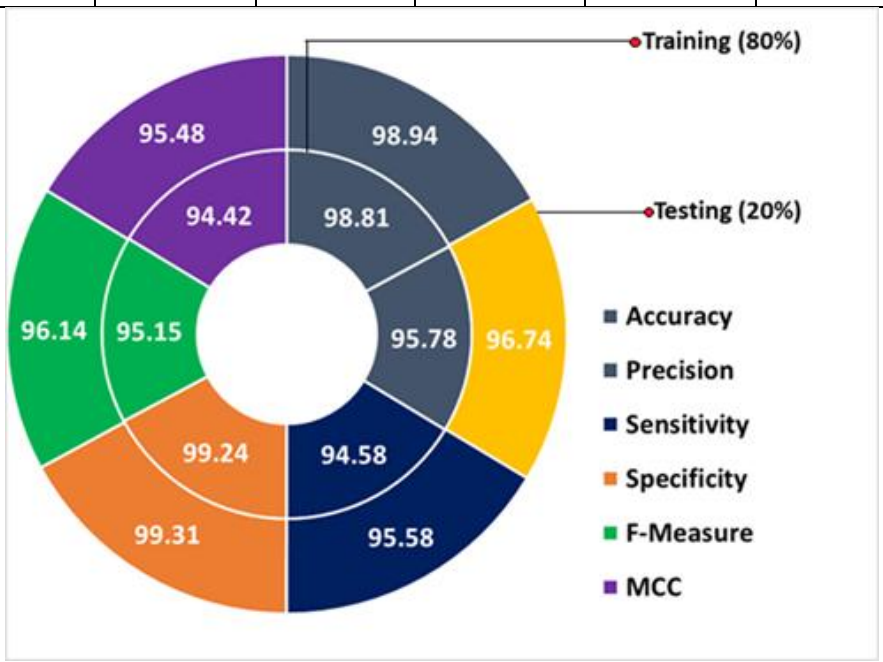



**Fig. 6.** Average of AGGCN-MMFOC method

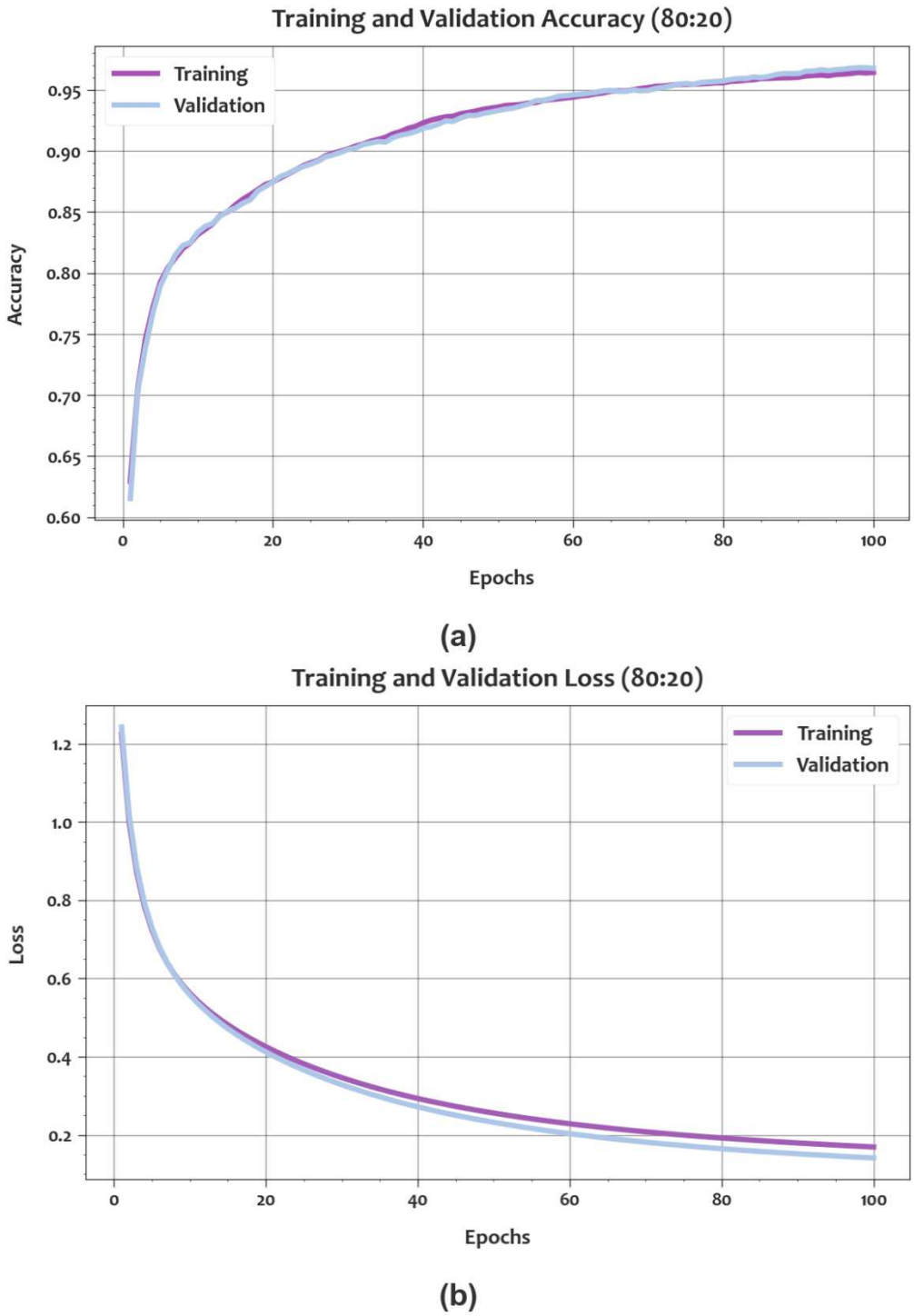



**Fig. 7.** Curve of (a) $Accur_y$ and (b) Loss of the AGGCN-MMFOC method with 80: 20

Fig. 7 shows the loss and accuracy curves of the AGGCN-MMFOC system with 80:20 across 100 epochs. Fig. 7a illustrate the training (TRAIN) and validation (VALID) $accur_y$ curves of the AGGCN-MMFOC system. TRAIN accuracy improves gradually and attains a better level, representing efficient learning and convergence. The VALID accuracy nearly follows the training trend with a stable and small gap, signifying minimal overfitting and good

generalization. Fig. 7b determines the TRAIN and VALID loss curves of AGGCN-MMFOC technique. Both losses decrease sharply while the starting epochs, signifying quick learning, and then progressively decline in a stable and smooth manner, reflecting efficient convergence. The near alignment among training and validation loss, with only a small consistent gap, recommends strong generalization capability and minimal overfitting.

Fig. 8 illustrate the confusion matrices made by the AGGCN-MMFOC system below 70%:30% TRPA/TEPA. The results specify that the AGGCN-MMFOC method has effective identification and detection of every classes accurately.

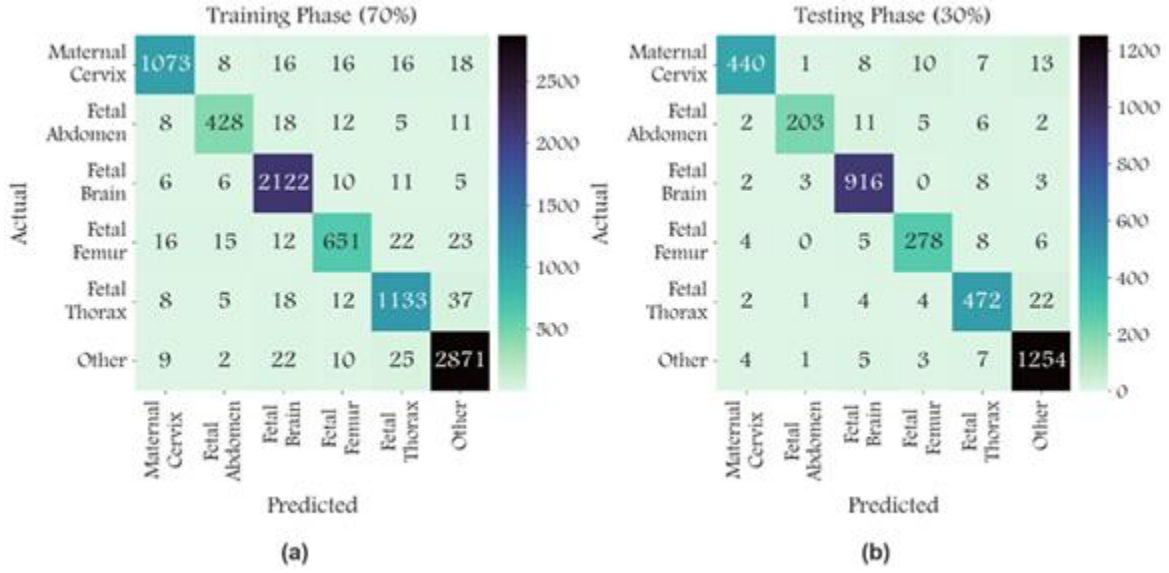



**Fig. 8.** Confusion matrices of AGGCN-MMFOC system (a, b) 70%:30% of TRPA/TEPA

Table 4 and Fig. 9 demonstrate the fetal classifier of AGGCN-MMFOC technique with 70%:30%. The outcomes inferred that the AGGCN-MMFOC system precisely classified the instances. On 70% TRPA, the AGGCN-MMFOC system provides average results with $accur_y$ of 98.46%, $preci_n$ of 94.34%, $sensi_y$ of 93.29%, $speci_y$ of 99.03%, $F_{measure}$ of 93.80%, and MCC of 92.86%, correspondingly. Concurrently, on 30% TEPA, the AGGCN-MMFOC approach presents average results with $accur_y$ of 98.59%, $preci_n$ of 95.43%, $sensi_y$ of 93.84%, $speci_y$ of 99.09%, $F_{measure}$ of 94.59%, and MCC of 93.75%, individually.

**Table 4** Fetal classification of AGGCN-MMFOC approach with 70:30 split

| Class Labels | $Accur_y$ | $Preci_n$ | $Sensi_y$ | $Speci_y$ | $F_{measure}$ | $MCC$ |
|---|---|---|---|---|---|---|
| **Training Phase (70%)** | | | | | | |
| Maternal Cervix | 98.61 | 95.80 | 93.55 | 99.38 | 94.66 | 93.87 |
| Fetal Abdomen | 98.96 | 92.24 | 88.80 | 99.56 | 90.49 | 89.96 |
| Fetal Brain | 98.57 | 96.11 | 98.24 | 98.68 | 97.16 | 96.22 |
| Fetal Femur | 98.29 | 91.56 | 88.09 | 99.24 | 89.79 | 88.88 |
| Fetal Thorax | 98.17 | 93.48 | 93.40 | 98.94 | 93.44 | 92.38 |
| Other | 98.13 | 96.83 | 97.69 | 98.36 | 97.26 | 95.84 |
| **Average** | **98.46** | **94.34** | **93.29** | **99.03** | **93.80** | **92.86** |
| **Testing Phase (30%)** | | | | | | |
| Maternal Cervix | 98.58 | 96.92 | 91.86 | 99.57 | 94.32 | 93.55 |
| Fetal Abdomen | 99.14 | 97.13 | 88.65 | 99.83 | 92.69 | 92.35 |
| Fetal Brain | 98.68 | 96.52 | 98.28 | 98.82 | 97.40 | 96.52 |
| Fetal Femur | 98.79 | 92.67 | 92.36 | 99.36 | 92.51 | 91.85 |
| Fetal Thorax | 98.15 | 92.91 | 93.47 | 98.88 | 93.19 | 92.12 |
| Other | 98.23 | 96.46 | 98.43 | 98.12 | 97.44 | 96.09 |
| **Average** | **98.59** | **95.43** | **93.84** | **99.09** | **94.59** | **93.75** |

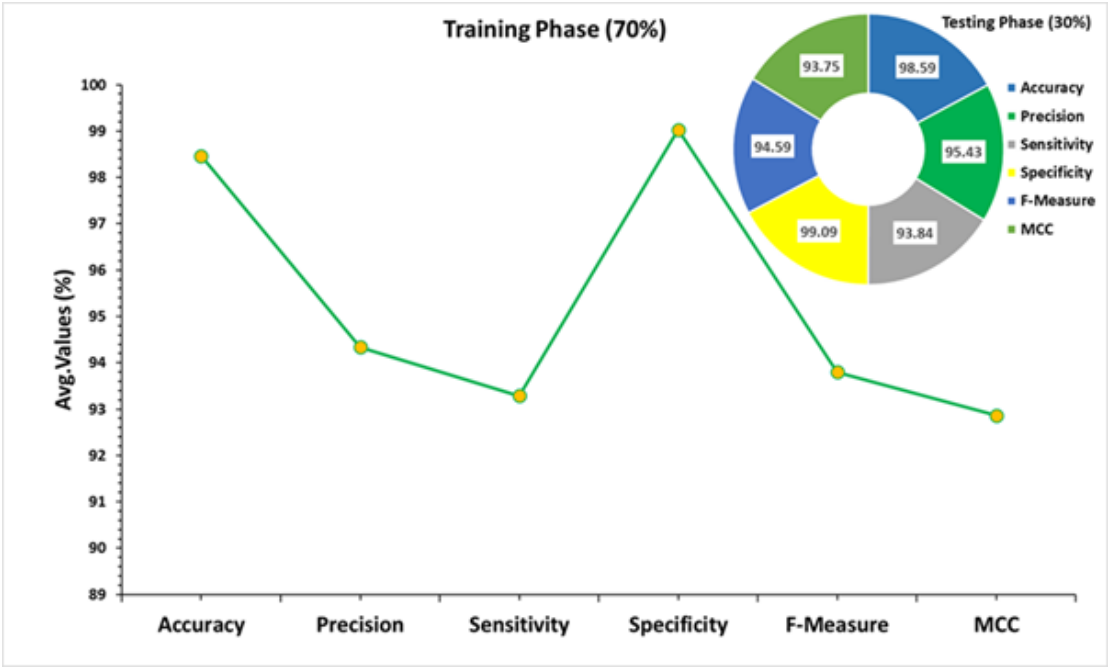



**Fig. 9.** Average of AGGCN-MMFOC method

Fig. 10 exhibits the accuracy and loss curves of the AGGCN-MMFOC technique with 70:30 across 100 epochs. Fig. 10a illustrate the TRAIN and VALID accuracy curves of the AGGCN-MMFOC method. TRAIN accuracy improves gradually and attains a better level, representing efficient learning and convergence. The VALID accuracy nearly follows the training trend with a stable and small gap, representing minimal overfitting and good generalization. Fig. 10b

displays the TRAIN and VALID loss curves of the AGGCN-MMFOC system. Both losses decrease sharply while the first epochs, signifying quick learning, and then progressively decline in a stable and smooth manner, reflecting efficient convergence. The near alignment among training and validation loss, with only a small consistent gap, recommend strong generalization and minimal overfitting capability.

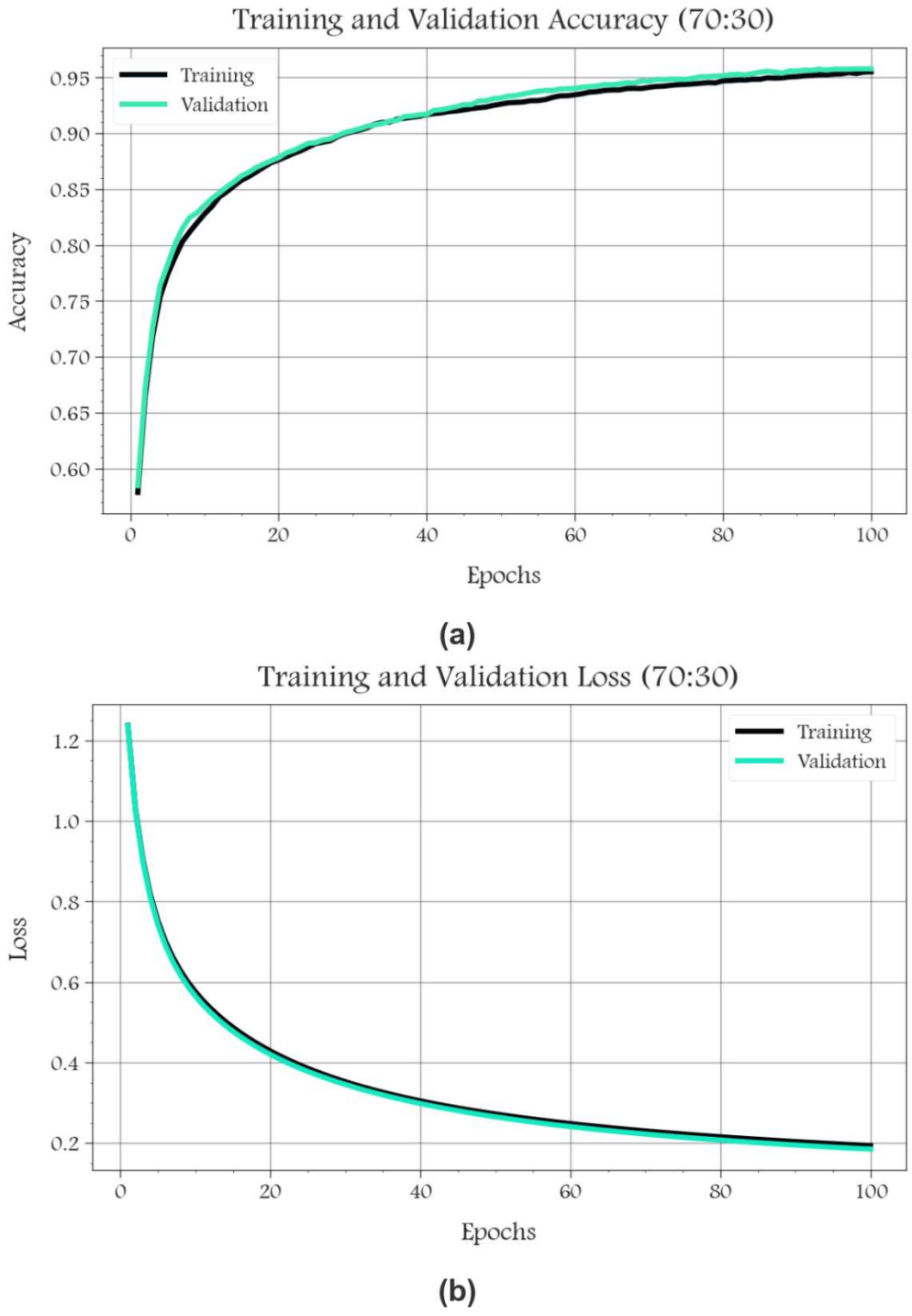



**Fig. 10.** Curve of (a) $Accur_y$ and (b) Loss of the AGGCN-MMFOC model with 70: 30

The comparative examination of AGGCN-MMFOC approach with present methods are illustrated in Table 5 [32, 33]. The simulation results stated that the AGGCN-MMFOC system outperformed higher performances.

**Table 5** Comparative analysis of AGGCN-MMFOC approach with existing technique

| **Model** | $Accur_y$ | **Loss Rate** | $Preci_n$ | $Sensi_y$ | $Speci_y$ | $F_{measure}$ | **Time Consuming (sec)** |
|---|---|---|---|---|---|---|---|
| ANN | 94.06 | 5.94 | 94.74 | 93.44 | 94.70 | 94.08 | 3.10 |
| RBF | 96.84 | 3.16 | 97.65 | 96.00 | 97.68 | 96.82 | 9.74 |
| Fetal plane detection | 88.90 | 11.10 | 89.83 | 90.40 | 88.56 | 91.47 | 11.96 |
| COMFormer | 95.64 | 4.36 | 91.06 | 93.63 | 90.73 | 93.51 | 10.46 |
| BP-Net | 92.35 | 7.65 | 88.87 | 91.20 | 95.68 | 90.45 | 11.50 |
| DAViT | 95.08 | 4.92 | 94.64 | 90.68 | 88.78 | 88.79 | 11.97 |
| POCFormer | 93.90 | 6.10 | 88.27 | 91.80 | 93.55 | 89.39 | 3.55 |
| AGGCN-MMFOC | 98.94 | 1.06 | 96.74 | 95.58 | 99.31 | 96.14 | 1.03 |

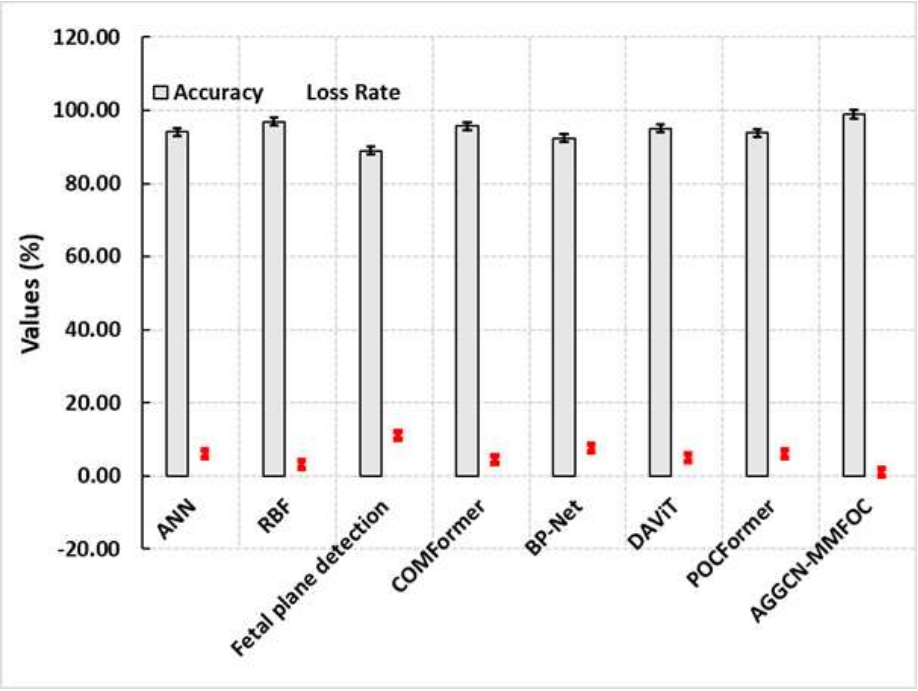



**Fig. 11.** Comparative results of AGGCN-MMFOC model with $Accur_y$ and Loss rate

Fig. 11 shows the comparative analysis of AGGCN-MMFOC technique with $accur_y$ and loss rate. Depend on $accur_y$, the AGGCN-MMFOC system have attained better performance with $accur_y$ of 98.94% whereas the ANN, RBF, Fetal plane detection, COMFormer, BP-Net, DAViT, and POCFormer methods have the slightly lower performance. Concurrently, relay on

loss rate, the AGGCN-MMFOC approach have attained lesser performance with loss rate of 1.06% while the ANN, RBF, Fetal plane detection, COMFormer, BP-Net, DAViT, and POCFormer systems have gains better results.

Fig. 12 exhibits the comparative research of AGGCN-MMFOC technique with different measures. The BP-Net and POCFormer methods has gain lower performance. Furthermore, the Fetal plane detection model has revealed slightly better results with with $preci_n$ of 89.83%, $sensi_y$ of 90.40%, $speci_y$ of 88.56%, and $F_{measure}$ of 91.47%. Also, ANN, RBF, COMFormer, and DAViT approaches have outperformed reasonable and nearly results. Finally, The AGGCN-MMFOC technique have attained better performance with $preci_n$ of 96.74%, $sensi_y$ of 95.58%, $speci_y$ of 99.31%, and $F_{measure}$ of 96.14%.

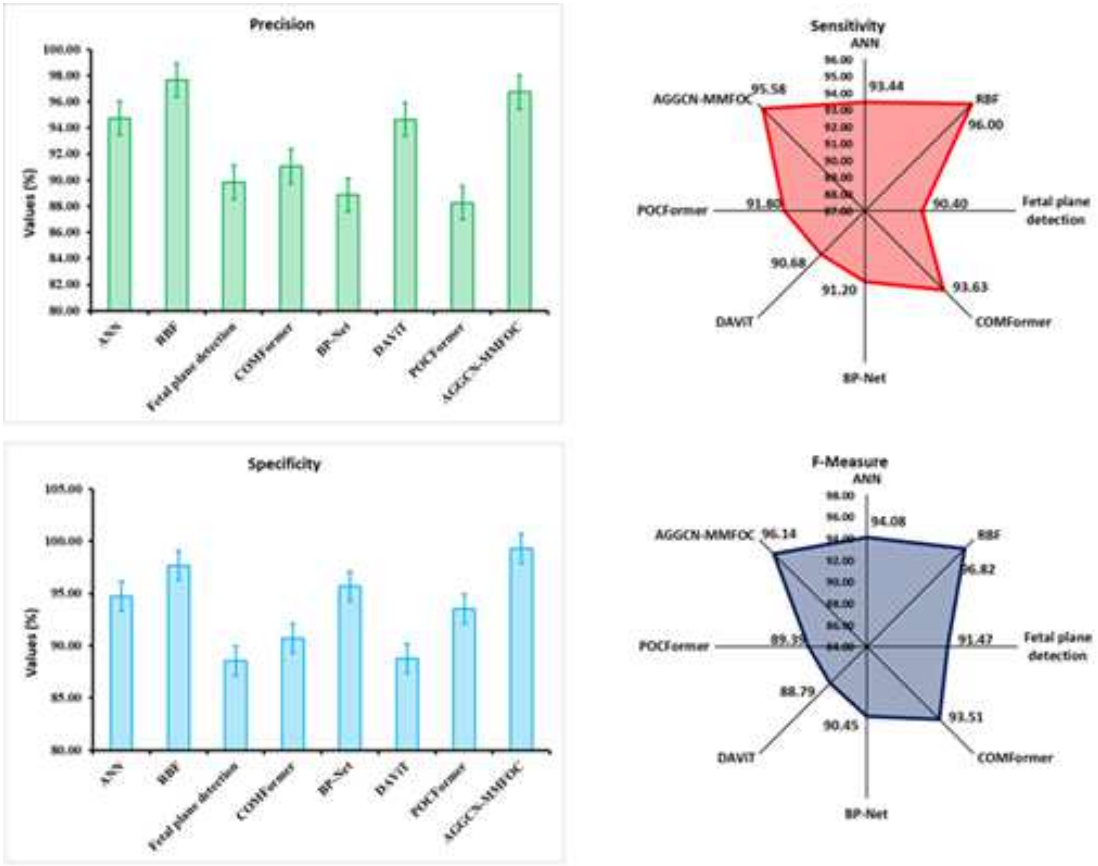



**Fig. 12.** Comparison result of AGGCN-MMFOC model with various metrics

Fig. 13 shows the time-consuming of AGGCN-MMFOC approach with current system. Depend on time, the AGGCN-MMFOC model gives lesser value of 1.03sec whereas the ANN, RBF, Fetal plane detection, COMFormer, BP-Net, DAViT, and POCFormer method reaches

higher time of 3.10sec, 9.74sec, 11.96sec, 10.46sec,11.50sec, 11.97sec, and 3.55sec, individually.

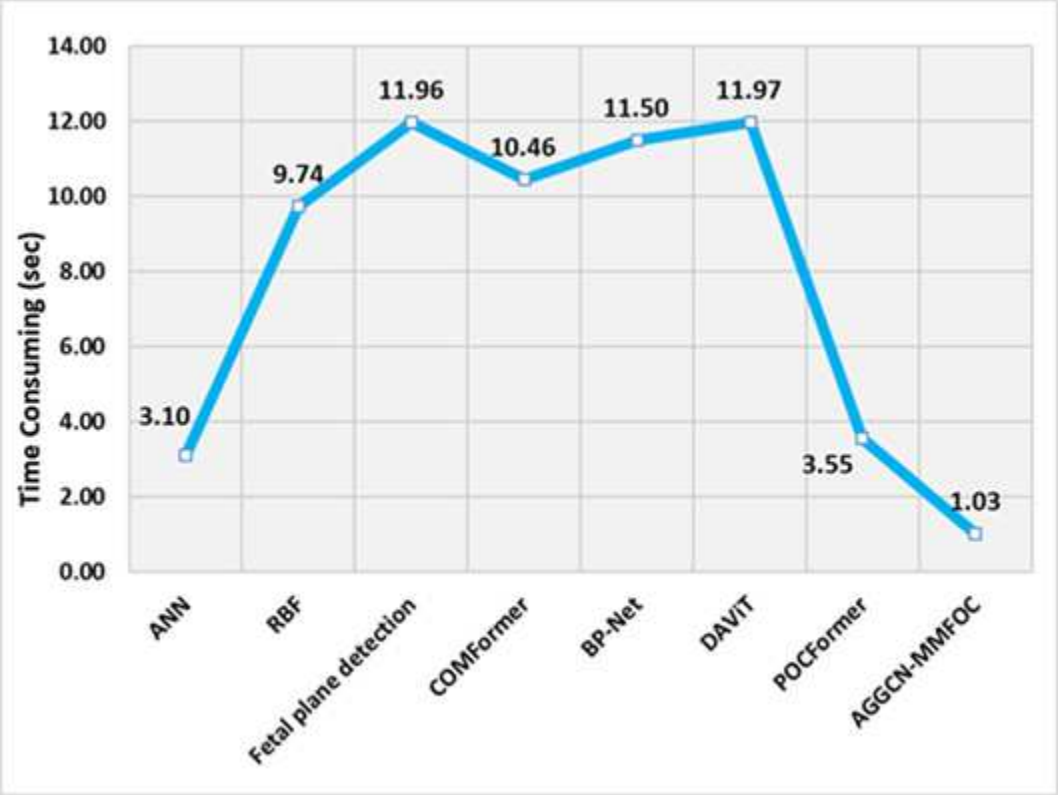



**Fig. 13.** Time consumption analysis of AGGCN-MMFOC model

Fig. 14 illustrate a qualitative explanation of how Score-CAM is utilized for understanding the DL model's decision on obstetric US images. The left column demonstrates the original US images belonging to various anatomical classes, containing other, fetal thorax, fetal femur, fetal brain, fetal abdomen, and maternal cervix. The right column illustrates the corresponding Score-CAM visualizations, where a color heatmap is overlaid on every image for representing the regions that most strongly influenced the model's prediction. In such heatmaps, red and yellow regions signify areas of high importance, meaning the method depends heavily on such anatomical structures for classification, while blue regions designate less contribution. The visualizations show that the technique concentrate on medical related regions. For instance, the skull area in fetal brain images or the elongated bright structure in fetal femur images rather than unrelated background.

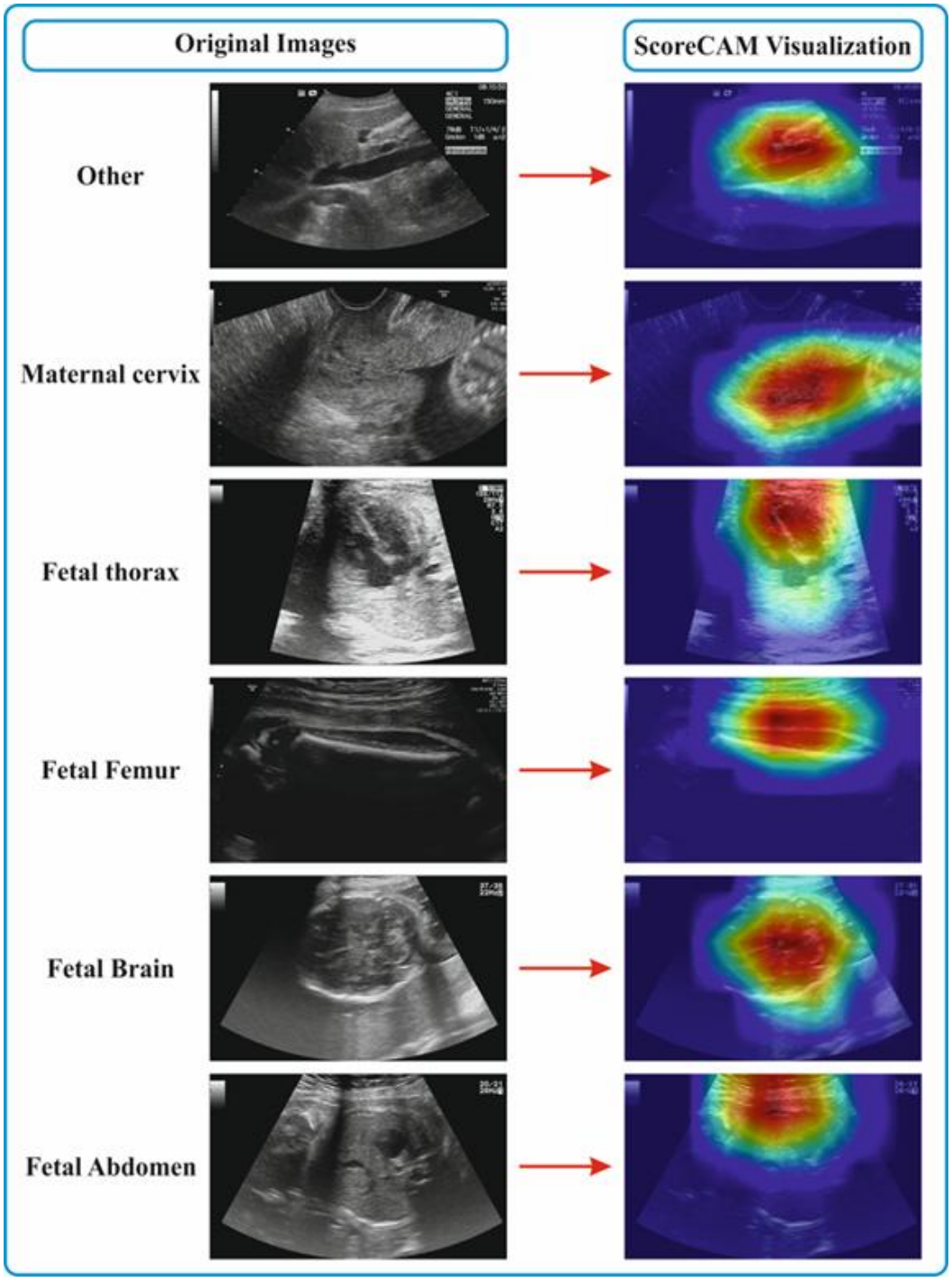



**Fig. 14.** ScoreCAM visual interpretation results of XAI models under 6 classes

## 5. Conclusion

In this paper, the AGGCN-MMFOC model has been proposed for the efficient classification of fetal organs from high-resolution US images, while offering interpretable predictions for clinically relevant anatomical planes. To achieve that, input US are first preprocessed using image resizing and normalization, noise removal, and contrast enhancement techniques. Then,

a learnable weighted fusion strategy is adopted to combine features from three architectures: EfficientNet-B7, Twins-SVT, and Inception-V3, which enables the extraction of discriminative features for the classification. The fused features are further fed into the Attn-GCN for the precise classification of anatomical planes into corresponding classes. To enhance the efficiency of the proposed AGGCN-MMFOC framework, the parameter optimization process is carried out using Nadam optimizer for the classification model. By integrating Score-CAM into the proposed model, the model's predictions become more explainable, assisting radiologists in assessing DL-based diagnoses and increasing confidence in automated anatomical plane classification systems. The experimental result analysis of the AGGCN-MMFOC model is conducted on the benchmark dataset. The simulation results pointed out the greater performance of the AGGCN-MMFOC technique with an accuracy of 98.94% and loss rate of 1.06 across multiple measures.

**Data Availability Statement:** The data that support the findings of this study are openly available at https://zenodo.org/records/3904280, reference number [31].

**Funding:** This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia under grant no. (IPP-000-611-2025). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

**Conflict of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript. All authors contributed to the manuscript and approved the final version.

**Ethics Approval:** This article does not involve any studies with human participants or animals performed by any of the authors.

**Consent to Participate:** Not applicable.

**Informed Consent:** Not applicable.

**Clinical Trial Registration:** Not applicable. The research data used in this study were obtained from previously published sources, and no human subjects were involved or harmed.

**Authorship contributions**

Faris Kateb: Conceptualization, methodology development, experiment, formal analysis, investigation, writing.

Rania M. Alhazmi: Formal analysis, investigation, validation, visualization, writing.

Sami Binyamin: Formal analysis, review and editing.

Mona Alqassim: Methodology, investigation.

Khalid Allehaibi: Review and editing.

Mahmoud Ragab: Conceptualization, methodology development, investigation, supervision, review and editing.

All authors have read and agreed to the published version of the manuscript.

**References**


[1] Vatanparast, A., Fateh, M. and Mashayekhi, H., 2026. AC_Branch_MobileNet: an effective hybrid branch network for plane classification in fetus brain ultrasound image. Biomedical Signal Processing and Control, 112, p.108586.

[2] Zheng, Q., Li, L., Huang, Y., Lyu, G. and Zeng, P., 2026. FCNSU-NET: Comprehensive recognition of the first trimester fetal CNS ultrasound standard planes and anatomical structures with linear transformer attention. Biomedical Signal Processing and Control, 114, p.109260.

[3] Judi, A.A., Suresh, P. and Raj, T.A.B., 2026. FB-UNet++: federated biometric UNet++ model for segmentation and classification network of fetal anomaly detection in prenatal care. Biomedical Signal Processing and Control, 112, p.108706.

[4] Testi, M., Fiorentino, M.C., Ballabio, M., Visani, G., Ciccozzi, M., Frontoni, E., Moccia, S. and Vessio, G., 2025. FetalMLOps: operationalizing machine learning models for standard fetal ultrasound plane classification. Medical & Biological Engineering & Computing, pp.1-16.

[5] Mushtaq, G. and Veningston, K., 2025. Fetal MoAt Net: a light-weight deep learning model for fetal diagnostic plane classification. International Journal of Computers and Applications, 47(11), pp.927-942.

[6] Zhu, S., Cai, J., Xiong, R., Zheng, L. and Ma, D., 2025. Singular pooling: a spectral pooling paradigm for second-trimester prenatal level II ultrasound standard fetal plane identification. IEEE Transactions on Circuits and Systems for Video Technology.

[7] Abdillahi, M.A., Baker, M.R., Doğru, A. and Buyrukoğlu, S., 2025. Enhanced Fetal Plane Classification in Ultrasound Imaging via Prototypical Networks and Few-Shot Learning. Journal of Imaging Informatics in Medicine, pp.1-18.

[8] Krishna, T.B. and Kokil, P., 2024. Standard fetal ultrasound plane classification based on stacked ensemble of deep learning models. Expert Systems with Applications, 238, p.122153.

[9] Sindhu, K.G., 2024. Ensemble-based advancements in maternal fetal plane and brain plane classification for enhanced prenatal diagnosis. International Journal of Information Technology, pp.1-17.

[10] Li, F., Li, P., Wu, X., Zeng, P., Lyu, G., Fan, Y., Liu, P., Song, H. and Liu, Z., 2024. FHUSP-NET: a multi-task model for fetal heart ultrasound standard plane recognition and key anatomical structures detection. Computers in Biology and Medicine, 168, p.107741.

[11] Sindhu, K.G. and Annamalai, R., 2024, March. Enhanced Multi-Class Fetal Plane Detection with Limb Localization in Ultrasound Images. In 2024 IEEE International Conference on Contemporary Computing and Communications (InC4) (Vol. 1, pp. 1-6). IEEE.

[12] Thomas, S. and Harikumar, S., 2024. An ensemble deep learning framework for foetal plane identification. International Journal of Information Technology, 16(3), pp.1377-1386.

[13] Castellanos-Díaz, N.O., Pérez-González, J., Arámbula-Cosío, F., Camargo-Marín, L., Guzmán-Huerta, M., Escalante-Ramírez, B., Olveres-Montiel, J., García-Ramírez, J., Medina-Bañuelos, V. and Valdés-Cristerna, R., 2024, October. Automatic Standard Plane Detection in Fetal Ultrasound Improved by Fine-Tuning. In Latin American Conference on Biomedical Engineering (pp. 368-377). Cham: Springer Nature Switzerland.

[14] Kaur, M. and Kaur, S., 2026. Hybrid deep learning models for automated fetal health prediction using cardiotocography signals. In Recent Advances in Computational Methods in Science and Technology (pp. 181-187). CRC Press.

[15] Ciobanu, Ș.G., Enache, I.A., Iovoaica-Rămescu, C., Berbecaru, E.I.A., Vochin, A., Băluță, I.D., Istrate-Ofițeru, A.M., Comănescu, C.M., Nagy, R.D., Șerbănescu, M.S. and Iliescu, D.G., 2025. Automatic identification of fetal abdominal planes from ultrasound images based on deep learning. Journal of Imaging Informatics in Medicine, pp.1-8.

[16] Turkan, M., Dandil, E., Urfali, F.E. and Korkmaz, M., 2025. FetalMovNet: A Novel Deep Learning Model Based on Attention Mechanism for Fetal Movement Classification in US. IEEE Access.

[17] Krishna, T.B., Poreddy, A.K.R., Sindhu, K.G. and Kokil, P., 2025. Automated Maternal Fetal Ultrasound Image Identification Using a Hybrid Vision Transformer Model. In International Conference on Pattern Recognition (pp. 253-269). Springer, Cham.

[18] Alasmawi, H., Bricker, L. and Yaqub, M., 2024. Fusc: Fetal ultrasound semantic clustering of second-trimester scans using deep self-supervised learning. Ultrasound in Medicine & Biology, 50(5), pp.703-711.

[19] Rathika, S., Mahendran, K., Sudarsan, H. and Ananth, S.V., 2024. Novel neural network classification of maternal fetal ultrasound planes through optimized feature selection. BMC Medical Imaging, 24(1), p.337.

[20] Ghabri, H., Alqahtani, M.S., Ben Othman, S., Al-Rasheed, A., Abbas, M., Almubarak, H.A., Sakli, H. and Abdelkarim, M.N., 2023. Transfer learning for accurate fetal organ classification from ultrasound images: a potential tool for maternal healthcare providers. Scientific Reports, 13(1), p.17904.

[21] Sarker, M.M.K., Singh, V.K., Alsharid, M., Hernandez-Cruz, N., Papageorghiou, A.T. and Noble, J.A., 2023. COMFormer: Classification of Maternal–Fetal and Brain Anatomy Using a Residual Cross-Covariance Attention Guided Transformer in Ultrasound. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 70(11), pp.1417-1427.

[22] Jadon, R., Gollapalli, V.S.T., Srinivasan, K., Chauhan, G.S. and Budda, R., 2025. Interpretable AI for skin lesion detection: Enhancing diagnostic accuracy with CNN and Score-CAM in IoMT systems. Journal of Ubiquitous Computing and Communication Technologies, 7(1), pp.1-18.

[23] Nagpal, D., Pattanaik, P.A., Madaan, V. and Agrawal, P., 2024. Enhancement of retinal images through modified anisotropic diffusion. Journal of Medical Artificial Intelligence, 7.

[24] Ramadani, R., Harkespan, I., Hartomo, K.D. and Arthur, C., 2026. LUMEN: Low-light Unified Multi-stage Enhancement Network to Improve RetinaFace-Based Face Detection. Journal of Future Artificial Intelligence and Technologies, 2(4), pp.565-586.

[25] Huang, Z., Su, L., Wu, J. and Chen, Y., 2023. Rock image classification based on EfficientNet and triplet attention mechanism. Applied Sciences, 13(5), p.3180.

[26] Zhang, B., Wu, Q., Wu, F., Huang, J. and Wang, C., 2024. A lightweight pyramid transformer for high-resolution SAR image-based building classification in port regions. Remote Sensing, 16(17), p.3218.

[27] Dutta, P., Sathi, K.A., Hossain, M.A. and Dewan, M.A.A., 2023. Conv-ViT: a convolution and vision transformer-based hybrid feature extraction method for retinal disease detection. Journal of Imaging, 9(7), p.140.

[28] Rui, W., Linfu, X., Yongsheng, L., Jianbang, W., Qun, Y. and Xiangjin, R., 2025. Mineral prospectivity prediction based on the dynamic relation model Atten-GCN: A case study of gold prospecting in the Yingfengjie area, Shaanxi province (northern China). Ore Geology Reviews, 176, p.106399.

[29] Tu, J., Yasin, A. and Mansor, N.S., 2025. NadamClip: A Novel Optimization Algorithm for Improving Prediction Accuracy and Training Stability. Processes, 13(7), p.2145.

[30] Jadon, R., Gollapalli, V.S.T., Srinivasan, K., Chauhan, G.S. and Budda, R., 2025. Interpretable AI for skin lesion detection: Enhancing diagnostic accuracy with CNN and Score-CAM in IoMT systems. Journal of Ubiquitous Computing and Communication Technologies, 7(1), pp.1-18.

[31] https://zenodo.org/records/3904280

[32] Rathika, S., Mahendran, K., Sudarsan, H. and Ananth, S.V., 2024. Novel neural network classification of maternal fetal ultrasound planes through optimized feature selection. *BMC Medical Imaging*, *24*(1), p.337.

[33] Vafaeezadeh, M., Behnam, H. and Gifani, P., 2024. Ultrasound image analysis with vision transformers. *Diagnostics*, *14*(5), p.542.