IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 31 January 2025, accepted 18 March 2025, date of publication 24 March 2025, date of current version 31 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3553548

RESEARCH ARTICLE

FetalMovNet: A Novel Deep Learning Model Based on Attention Mechanism for Fetal Movement Classification in US

MUSA TURKAN[1], EMRE DANDIL[2], FURKAN ERTURK URFALI[3], AND MEHMET KORKMAZ[4]
[1]Institute of Graduate, Department of Computer and Electronics Engineering, Bilecik Şeyh Edebali University, 11230 Bilecik, Türkiye
[2]Faculty of Engineering, Department of Computer Engineering, Bilecik Şeyh Edebali University, 11230 Bilecik, Türkiye
[3]Department of Radiology, Bursa City Hospital, 16200 Bursa, Türkiye
[4]Department of Interventional Radiology, Kütahya University of Health Sciences, 43020 Kütahya, Türkiye

Corresponding author: Emre Dandil (emre.dandil@bilecik.edu.tr)

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Non-Interventional Clinical Research of Kütahya University of Health Sciences, in July 2021 under Application 2021/12-07.

ABSTRACT Automated classification of fetal movements in ultrasound (US) videos is critical for assessing fetal well-being and detecting potential complications during pregnancy. This study introduces FetalMovNet, a novel deep learning model that incorporates an attention mechanism to improve the classification of fetal movement in US video sequences. The model integrates convolutional neural networks (CNN) for feature extraction and an attention mechanism to capture spatio-temporal patterns, significantly improving classification performance of fetal movements. To evaluate FetalMovNet, we construct a new dataset containing fetal movements in US across seven different anatomical structures-head, body, arm, hand, heart, leg, and foot. Experimental results show that FetalMovNet achieves an accuracy of 0.9887, precision of 0.9871, recall of 0.9910, and an F1-score of 0.9891, outperforming state-of-the-art CNN and CNN-LSTM architectures. Ablation studies confirm the effectiveness of the attention mechanism, with FetalMovNet achieving an area under curve (AUC) score of 0.9957, compared to 0.9471 for CNN and 0.9543 for CNN-LSTM. The proposed FetalMovNet model provides a robust and clinically applicable tool for real-time fetal movement monitoring, reducing the need for manual assessment and improving prenatal care.

INDEX TERMS Fetus, fetal movement detection, US video, deep learning, CNN, attention mechanism.

I. INTRODUCTION

The assessment of fetal movements is a crucial aspect of prenatal care, serving as an important indicator of fetal health and well-being. Fetal movements typically become noticeable by 15 weeks of gestation [1], [2]. These movements can include kicks, rolls and flutters and vary in strength and frequency depending on the fetus' age and activity level. Abnormal patterns or reduced frequency of fetal movements can signal potential complications, including fetal distress, growth restrictions, or even stillbirth [3]. Traditionally, the monitoring of fetal movements has relied on maternal perception and manual observation by healthcare professionals, methods that are inherently subjective and prone to variability.

The ability to detect, recognize and classify fetal tissue movements can be useful in monitoring fetal development and detecting abnormalities. Movement of fetus is a measure of fetal health [4]. Normal fetal movements reflect a properly functioning central nervous system and a healthy, risk-free fetus [5]. Therefore, fetal movement is an important indicator of healthy development. There is limited understanding of the relationship between fetal behaviour and fetal brain development. However, it is thought that fetal movements are fundamental to the development of most parts of the nervous system and muscle development [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Rajeswari Sundararajan.

Monitoring fetal movement is an important parameter in a healthy pregnancy. Fetal movements can be divided into general movements and movements that may be of concern. While general movements are a sign of a healthy process, worrying movements can be a sign of poor fetal health [7]. Reduced or absent fetal movements can be a sign of serious fetal risks [8]. It is therefore very important to understand how the fetus moves and to extract quantitative and qualitative patterns of movement during pregnancy. The fetus may respond to sound with different movements [9]. In this case, these movements can be used to check hearing health. On the other hand, some excessive fetal movements may be a sign of serious fetal disorders. As it is difficult for mothers to recognize this, intervention may be delayed [10]. Real-time ultrasound (US) has made it possible to assess large motor movements in healthy and intact fetuses in their natural environment. This has allowed the characterization of fetal movements in fetuses with growth restriction, fetuses destined for preterm delivery and fetuses with chromosomal abnormalities [11].

When US images are combined with appropriate image processing techniques, the images can be processed more efficiently and can assist professionals in diagnosis and treatment. US imaging is one of the most widely used diagnostic and therapeutic modalities in medicine [12]. US is preferred in most parts of the world, especially in prenatal imaging, because it is relatively safe, inexpensive and easy to use [13]. The advent of US technology has greatly improved the ability to visualize and monitor fetal movements in real time, but the interpretation of these movements remains largely manual, requiring expert analysis that can be both time consuming and prone to human error.

For large datasets, image-level labelling for classification and pixel-level labelling for segmentation are time-consuming operations. Working with US images involves more complex operations than working with other medical imaging images. Because the resolution of US images is very low, there is a lot of noise and no visual sign. For this reason, more appropriate specific algorithms need to be used to analyze US images [13]. Two-dimensional (2D) images of the fetus are obtained by placing an US transducer on the maternal abdomen along the axis of the fetal thorax and abdomen. The frequency, intensity and duration of fetal movements are related to the maternal perception of the movements. Movements of a certain shorter duration or intensity observed during the US scan is usually not perceived [11]. Although some work has been done to describe the pattern of fetal movements, a general framework has not yet been defined or is very limited in terms of the amount, ideal number and when these movements need to be assessed.

As deep learning algorithms have developed, their use in gesture recognition has become widespread. Many deep learning models have been proposed for movement classification. CNN and attention mechanisms are used to classify human movements and to detect specific movements from video [14], [15], [16]. On the other hand, it has also been shown that the combination of CNN and LSTM works effectively for video motion recognition tasks such as human action recognition [17]. The performance improvement from deep learning depends on the large datasets used for training. The generation of high-quality datasets for fetal movement detection and classification is an important process in this field. Although there are many publicly available datasets in the medical field, the available fetal datasets are very limited and this is one of the main bottlenecks in the development of deep learning applications in this field [18]. Therefore, there is a need for large datasets of standardized US images annotated with the fetus and fetal movements. However, there are a number of previous databases on the fetus. One of these, the HC18 dataset, has been created for the automatic measurement of fetal head circumference, annotated by experienced experts [19]. In another dataset, six different tissues in the fetus, such as the abdomen, brain, femur, thorax and chest, were annotated for classification from the fetal anatomical plane [20].

It has been observed that previous studies proposed for the evaluation and classification of movements on fetal videos are limited. Previous studies on the fetus have proposed clinical measurements of the fetus, segmentation of fetal tissues and evaluation of the quality of fetal images. Methods have also been developed to measure fetal head circumference, detect fetal lateral ventricles, and detect and measure fetal anatomy from 2D fetal images [19], [21], [22], [23], [24]. On the other hand, for tasks based on US images, it has been shown that direct analysis of video, rather than examining still images, is more successful [25]. Therefore, this study proposes a new model based on CNN and attention deep learning architectures for the classification of fetal movements from US videos. A study-specific dataset consisting of sequences of video frames for each of the fetal structures is generated. The contributions of this study are summarized below:

- A novel dataset of seven different anatomical structures of the fetus is generated from US movement videos.
- A new deep learning model FetalMovNet based on attention and CNN is proposed for fetal movement classification. This model offers higher performance compared to existing state-of-the-art methods such as CNN and CNN-LSTM by effectively processing both spatial and temporal information.
- Thanks to the attention mechanism, the proposed model achieves more accurate and reliable classification results by focusing only on important regions and movements in sequential US frames.
- By considering the temporal characteristics of fetal movements, the FetalMovNet model is able to capture critical patterns in movement sequences. This contributes to a more accurate understanding of sequential movements, especially in US videos.
- By providing a practical solution in a clinical setting, the model can help clinicians minimize inaccuracies

due to manual assessment. This enables faster and more objective analysis of fetal movement in prenatal care.
- The study provides a new state-of-the-art application example by demonstrating the potential of deep learning and attention mechanisms in US video classification.

This study consists of the following sections. Section II reviews the related work previously proposed for the classification of fetal movements. Section III is materials and methods and includes the architecture of the proposed FetalMovNet deep learning network for classification of fetal movements. In addition, the US video dataset specific to the study and the architecture of the proposed FetalMovNet model are presented in detail in this section. In Section IV, experimental studies are performed to evaluate the performance of FetalMovNet and the results are compared with the results of state-of-the-art CNN and CNN-LSTM architectures through detailed analyses and ablation studies. In Section V, the results are discussed, the obtained results are compared with literature results, and the spatio-temporal features extracted for FetalMovNet are represented by heatmaps using Grad-CAM. The last section is the conclusions, where the strengths and limitations are highlighted and an overall evaluation of the study is provided.

II. RELATED WORKS

Obtaining fetal biometric measurements from 2D US scans is essential for fetal monitoring. However, the acquisition of standardized head, abdominal and femoral planes is difficult due to variable image quality [26]. It has been shown that measurements obtained from first-trimester US scans can be used to predict high-risk pregnancies [27]. US images can be used to monitor the pregnancy process by detecting critical anatomical structures and tracking fetal movements. Fetal movement is an important process as it is used to detect an inactive heartbeat [28]. For these reasons, many studies have been conducted using fetal US scans to assist experts in fetal assessment [29].

Sobhaninia et al. [30] used a deep learning CNN architecture to measure fetal biometrics. In their study, they evaluated the measurement of head circumference for monitoring fetal growth and health. On the other hand, fetal heart movement is an important indicator for the detection and assessment of heart disease. Dozen et al. [31] performed image segmentation in fetal heart US videos using time series information and special region information. In the study, fetal heart beats and dynamically changing and small cardiac substructures were detected. They also presented a deep learning-based method for detecting the ventricular region. In addition, analysis and monitoring of fetal heart movements can also be performed using echocardiography data [32], [33], [34]. However, it has been shown that the evaluation of fetal cardiography is difficult due to the involuntary movements of the fetus, the very small size of the heart and the lack of expertise and experience in fetal cardiography [35].

Ishikawa et al. [36] proposed a CNN architecture based fetal part detection and classification method to estimate fetal position. In the study, fetal anatomical structures were classified as head, body, legs and other parts. In another study, van den Heuvel et al. [37] proposed a VGG-Net inspired deep learning CNN method to measure head circumference in each frame using US images. Ravishankar et al. [38] combined traditional anatomical structure detection and deep learning methods to detect and measure the abdominal region of the fetus from 2D US images. By using a CNN-based model, they were able to detect the abdominal region with higher performance than traditional methods. Yaqup et al. [39] proposed a deep learning method to detect possible brain abnormalities in the fetus. They developed a system that automatically detects whether the US images of the fetal brain are accurate and meet the required clinical standard. The study involves the correct localization of the fetal brain, the detection of the region where the structures of interest are located, and the recognition of the acoustic pattern in the regions that provide planar verification. Chen et al. [40] proposed a recurrent neural network-based hybrid method for accurate acquisition of fetal planes in US images, detection of the standard plane and accurate biometric measurements. In another study, Gustavo et al. [22] presented a method to detect and measure fetal anatomical structures in US images.

Chen et al. [21] proposed a deep learning-based framework that detects and quantifies the lateral ventricles from 2D US scans. In the study, after pixel-based segmentation of the lateral ventricle images, the number of pixels per centimeter was calculated by morphological operations in the light of previously determined information. It was highlighted that the experiments achieved more successful results than manual measurements. Arnaout et al. [41] proposed a deep learning method to detect congenital heart defects in the fetus from US images. In the study, the hearts with defects were detected in a test set obtained from the fetal examination. In addition, the cardiac measurements were correlated with measurements from normal and abnormal hearts. Plotka et al. [42] developed a model for estimating fetal weight by measuring fetal anatomical structures. The study compared the proposed deep learning predictions with manual measurements made by six human readers with different levels of expertise. It was confirmed that the difference between the expert measurements and the predictions of the deep learning model was 2.5%.

Previously proposed studies for fetal movement on US images can be broadly classified as fetal standard plane detection, anatomical structure analysis and biometric parameter estimation [23]. Considering these tasks, it is clear that there are few studies on movement detection of fetal anatomical structure. Although studies on the classification of fetal movement provide important benefits in prenatal care, the existing methods have significant limitations and shortcomings. Traditional approaches do not effectively combine the spatial and temporal information in US images. They often focus on a single image frame and do not

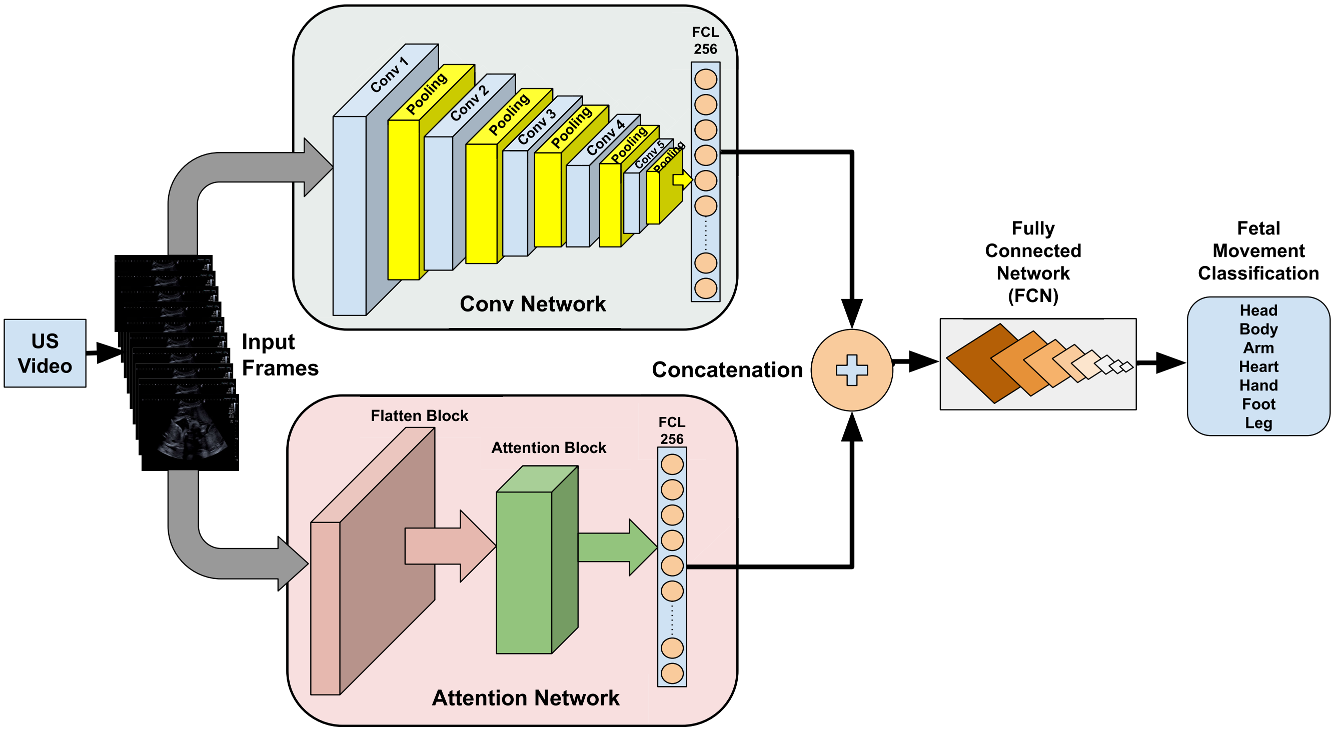

FIGURE 1. Architecture of the proposed FetalMovNet deep learning network for classification of fetal movements.

capture the full dynamics of movement, which can lead to misclassification. On the other hand, manually extracted features are often used to classify fetal movements. However, these features are usually insufficient to describe a limited number of movements and are subjective as they are based on human intervention. In addition, the data used to classify fetal movements are limited and often not accurately annotated. The lack of large and balanced datasets, especially for rare fetal movements, limits the performance of the models. The majority of existing studies do not take advantage of the attentional mechanism, which allows attention to focus only on specific regions to capture important information. This causes the model to focus equally on the whole image, which can lead to critical details being missed.

III. MATERIAL AND METHODS

In this study, a new deep learning model FetalMovNet was developed to classify the movements of fetal anatomical structures from US videos. In the process of fetal movement classification, an attention mechanism was applied to the frames used as input to the model, and movement classification was performed by combining with CNN. In addition, a new dataset was created from US videos containing the movements of 7 different fetal anatomical structures. The video frame sequences in this dataset were used as input to the proposed model and the model was used to classify fetal movements. The architecture of the proposed FetalMovNet deep learning network for fetal movement classification is shown in Fig. 1.

A. FETUS US DATASET

Creating an original dataset is an important step for deep learning applications. In previous studies, there are no or very limited datasets created from US videos to detect and classify the movements of many anatomical structures of the fetus. In this study, an original study-specific dataset was created using US videos of seven different fetal anatomical structures for fetal movement classification. Fetal images and videos were collected from Kütahya Evliya Çelebi Training and Research Hospital. Fetal US scans were obtained from 105 healthy pregnant women aged between 21 and 40 years. The fetuses ranged in age from 16 to 23 weeks, and the pregnant women's US scan dates ranged from March 2021 to April 2024. On the other hand, according to the decision of Ethics Committee of Non-Interventional Clinical Research of Kütahya University of Health Sciences dated 08.07.2021 and numbered 2021/12-07, the conduct of this study was ethically approved. Informed consent was obtained from all volunteer pregnant women.

In the proposed study, all pregnant women were selected from cases in the second trimester of pregnancy to facilitate assessment of fetal movements and consistency of patterns. The main extremities, diaphragm, swallowing and body, head and heart, movements were recorded as US video for all fetuses in the dataset. Fetal US evaluations in the dataset were performed using a Siemens Acuson S3000 with a 6C1 HD transducer. The US system used in the study includes a convex transducer with a frequency of 4-4.5 MHz and an obstetric US preset.

As part of the dataset creation process, fetal anatomical structures were first extracted from the US videos and sequential images of each anatomical structure were selected. The US scans of the fetal anatomical structures were then annotated by 2 experts to create true class labels for the anatomical structures. The images of each annotated anatomical structure of the fetus were divided into video sequences of 10 frames according to the sequential order of the video frames. As a result, a total of 587 video frame sequences were created with the help of 2 different experts for 7 different anatomical structure movements of the fetus, including head, body, arm, heart, hand, foot and leg, using videos collected from different patients during healthy pregnancy. The image sequences were created without skipping any frames while obtaining images from the video. Four random US frames representing fetal movements for the anatomical structures of head, body, hand, arm, heart, leg and foot in the dataset created for the study are shown in Fig. 2.

B. PRE-PROPOCESSING

To ensure high-quality input data for FetalMovNet, a structured pre-processing pipeline is essential. As part of the study, frames obtained from US videos of fetal anatomical structures such as head, body, arm, hand, heart, leg, foot in the dataset were annotated by experts prior to classification. Labelling was performed using frame-by-frame manual annotation based on movement patterns observed in US videos. In addition, 2 annotators cross-validated labels to reduce inter-observer variability. Afterwards, the US frames used for annotation in the dataset were resized to 300 × 300 to standardize input image dimensions for computational efficiency and consistency. Furthermore, the smaller sized frames were padded and the larger sized frames were resized accordingly. On the other hand, the frames from the US videos were used in greyscale. Finally, the video frames were normalized to floating-point values in the 0-1 range before being applied to the network to provide pixel intensity values to improve model generalization.

C. ARCHITECTURE OF PROPOSED FETALMOVNET MODEL

Discriminative feature extraction in US videos is known to be a powerful approach for object recognition and detection. CNN networks are highly complex and non-linear models with strong feature representation capacity [43]. The combination of attention and CNN layers can be a suitable model for motion classification. The CNN can be used to determine the distinctive features of objects, and the attention process can be used to distinguish important patterns of change that the motion reveals in each frame. Using the attention mechanism in a neural network allows different regions of the image to be weighted. It allows the network to focus on more significant features rather than unimportant features such as background. Attention also reduces the weight of unimportant parts and increases the weight of important parts. This makes it possible to focus on areas that are important for learning. Therefore, the attention mechanism can be applied to the frames to learn the important and distinctive features of the changes between fetal US video frames.

Traditional, simpler CNN-based architectures extract hierarchical features but lack temporal awareness for movement classification. Attention mechanisms improve feature selection by ensuring that the model focuses on relevant fetal structures rather than irrelevant regions. Integrating attention mechanisms improves feature representation and improves classification performance in medical imaging [44]. Similarly, a hybrid deep learning model can outperform conventional CNNs for complex biomedical applications [45]. On the other hand, CNN-LSTM models can capture temporal dependencies, but have higher computational costs and require longer training times. Consequently, attention mechanisms for US applications are computationally more efficient, focusing on critical frames without the overhead of sequential memory models such as LSTM.

In this study, a new deep learning model, named FetalMovNet, is developed using a combination of CNN and attention to classify the movements of fetal anatomical structures in US videos. In the architecture, firstly, regions of different fetal tissues are extracted from US videos in sequential frames and 10 consecutive frames are acquired to obtain a sequential of movements of each structure. Furthermore, in the proposed architecture, each frame is not only considered as an independent input. By evaluating 10 sequential frames together, the time-dependent changes of the movement are learned. An attention network is added to focus on the significant changes between frames and to highlight the distinctive features in the classification of fetal movements. Each of the 10 frames applied to the input of the attention network is flattened with the Flatten layer, multiplied by the Query and Value matrices, and then attention is applied. The attention process, applied to 10 sequential frames, generates an attention score for each frame. The Conv process, applied together with the sequential frames, allows the distinctive movement patterns in the image to be learned. The FetalMovNet model is then used to classify the movements of 7 different anatomical structures annotated from the fetal videos. The classification is completed by combining the output of the convolution (Conv) network with the output of the attention network. In the model, Conv layer and the attention layer, which extract the features of fetal anatomical structures, are applied to the input US frames obtained from US videos, and the outputs of these layers are combined and the classification process is performed with softmax using a fully connected network. The block diagram of the proposed FetalMovNet network for fetal motion classification from US video frames is shown in Fig. 3.

1) ATTENTION NETWORK

In this study, the attention mechanism was used to capture important variations between fetal US video frames. There-

FIGURE 2. Frames representing movements and changes in US videos for each anatomical structure in the dataset.

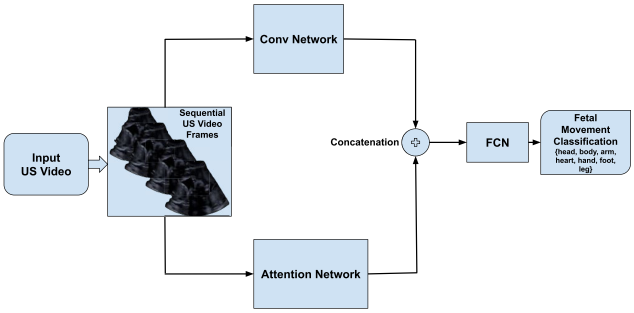

FIGURE 3. Block diagram of the proposed FetalMovNet network for classification of fetal movement from US video frames.

fore, in order to detect fetal movement, the attention network is used in the model to detect important movement-specific variations. The attention network of the proposed FetalMovNet architecture consists of a Flatten block, an Attention block and a Fully Connected Layer (FCL), as shown in Fig. 4. The Flatten block in the attention network is a critical step that facilitates the transition from high-dimensional feature representations to a format that can be processed by subsequent layers, thereby enhancing the model's ability to focus on relevant features while minimizing the impact of less important information. For tasks such as fetal movement classification in US, the output of an attention block needs to be flattened before being applied to the final layers of the network that perform the classification task.

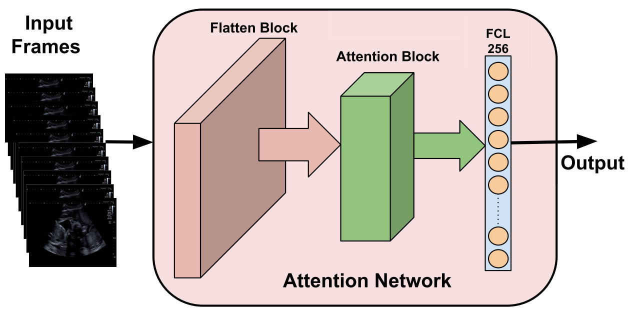

FIGURE 4. Components of the attention layer in the proposed FetalMovNet architecture.

When the human eye looks at an image, it perceives only the important parts and defines the image in this way. Similarly, the human brain pays attention to only the important parts of the incoming information. The attention mechanism works in a similar way to the human cognitive method [46]. Attention mechanism has been used in many machine learning tasks, such as classification, visual recognition, and text processing, due to its ability to process deep learning and large data inputs, and to capture long-term dependencies and relationships in data. In particular, the attention mechanism is used in image classification to highlight the important pixels in the image and suppress the unimportant ones. Although the attention mechanism is mainly developed for text translation, it is also widely used in image recognition tasks [47], [48], [49].

The Attention mechanism allows inputs to interact with each other to decide which need more attention. The outputs are the sum of these interactions and the Attention scores. It improves the performance of the network by determining which of the different input features require selective attention. Traditional models reason by giving equal weight to all inputs. Attention-based models, on the other hand, assign different weights to inputs according to their importance. This process allows unnecessary information to be discarded and only the important input features to be focused on. Attention-based systems contribute to higher accuracy by reducing the computational cost of video recognition [16], [49].

The query (Q), key (K) and value (V) matrices used in the attention mechanism are shown in Fig. 5. Here the normalized input value, denoted by X, represents each of the US video frames applied to the model. The 10 input frames representing the movement of the fetal anatomy are flattened using the Flatten layer to obtain the input values for Wq, Wk and Wv, which represent the weight values. Thus, each input multiplied by the weight matrices Wq, Wk and Wv is applied to the fully connected layer to obtain the matrices Q, K, V. As a result, for each frame in the US video, the attention mechanism captures the highlights of the entire fetal motion change as it is processed by looking at other sequential frames.

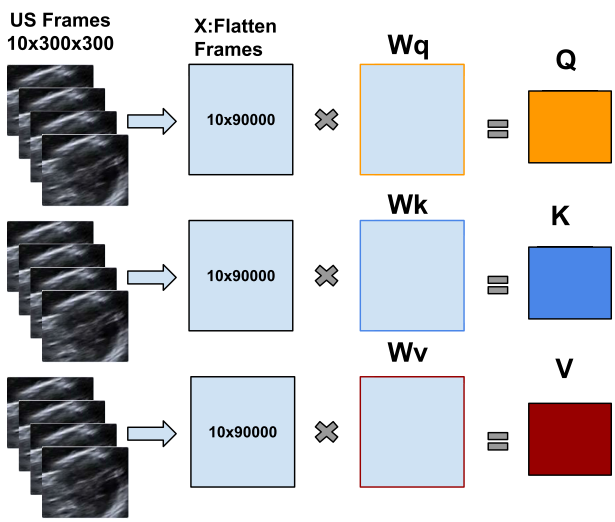

FIGURE 5. Creation of query (Q), key (K) and value (V) matrices in attention mechanism.

A typical Attention module receives n inputs and produces n outputs. In the Attention mechanism, as shown in Fig. 6, first 10 frames representing fetal movement are sequentially selected from the input US videos. Then, each frame is normalized and flattened by the Flatten layer to complete the preparation of the inputs. In the next step, the weights are multiplied by the prepared inputs to form the Q, K and V matrices. In the next step, the Q and K matrices are softmaxed and an attention score is calculated for each input US frame. This score is multiplied by the V matrix and the output Z in Fig. 7 is obtained.

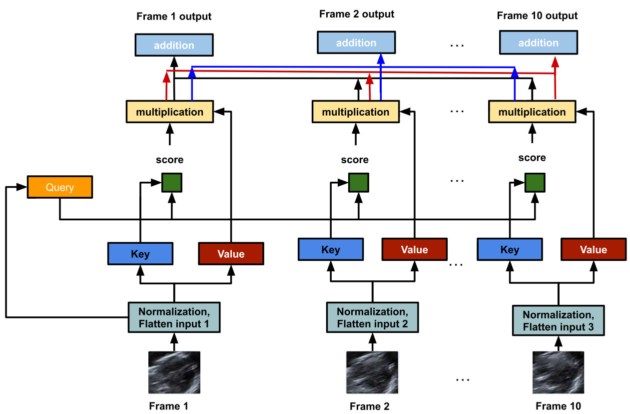

FIGURE 6. Attention block internal structure and functioning mechanism.

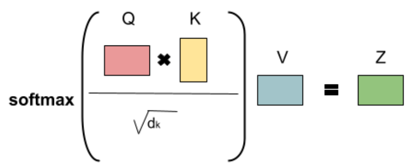

FIGURE 7. Obtaining the attention score by passing the Q, K and V matrices through the softmax unit in the attention mechanism.

The scores obtained in the attention mechanism produce results that allow focusing on important points in the image or on a feature map. In the attention structure, the dot product operation is applied. Thus, attention produces results that focus on the important points between US video frames. In the proposed study, the output produced by the attention network and the output produced by the Conv network are combined and the classification process is completed with softmax.

2) CONVOLUTION NETWORK

The convolutional network (Conv), which is a typical CNN model of the FetalMovNet architecture proposed in this study for the classification of fetal movements, consists of 5 different convolutional and pooling layers, as shown in Fig. 8. This network is designed to work with inputs consisting of 10 frames of 300 × 300 size. US video frame sequences of each anatomical structure of the fetus are provided as input to the Conv network. In the FetalMovNet architecture, an attention score is assigned to each frame applied to the input. These scores are passed through the fully connected layer and combined with the output of the convolution layer, and then the output is obtained with the softmax classifier. In parallel with the Conv layer, which learns the features of the input images, the attention process applied to each video frame allows the connections between the video frames to be learned. Thus, the intention is to focus on the motion-defining features for the video frames in the network. In the conv network, pooling layers are added between the conv layers to reduce the computational load. On the other hand, a 3 × 3 filter is used in the convolution process.

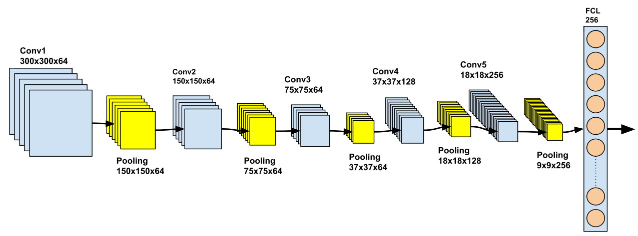

FIGURE 8. Conv layer used in the proposed FetalMovNet model.

3) FULLY CONNECTED NETWORK (FCN)

At the output of the attention and Conv networks shown in Fig. 9, the FCN used for classification produces a softmax layer that produces outputs for 7 different classes. In the layers of the FCN, the ReLU activation function is applied, except for the output. Prior to the classification stage, the outputs of the Conv and attention networks are combined and applied to the FCN as an input of size 512 and reduced to 256, 128, 64 and 7 respectively. Dropout layers of 0.2 are also added to prevent the network from memorizing.

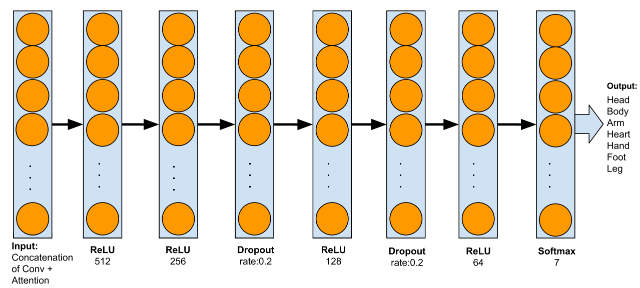

FIGURE 9. The structure of the FCN used for classification at the output of attention and Conv networks.

IV. EXPERIMENTAL AND RESULTS

In this study, we developed a novel deep learning model, FetalMovNet, improved by an attention mechanism, to classify fetal movements in US videos. The FetalMovNet model was trained on a carefully curated dataset comprising different fetal tissue movements captured in US video frame sequences.

Several experimental analyses were performed to evaluate the performance of the proposed FetalMovNet deep learning model with attention mechanism in classifying fetal movements. In addition, the effect of the model components was also evaluated in the experimental studies using an ablation study. The ablation study compares FetalMovNet, which consists of CNN and attention structures, with CNN-LSTM, a combination of CNN and long short-term memory (LSTM) architectures [50], and typical CNN architectures used for classification. The experimental studies were performed on the generated US video dataset and the performance of the model was analyzed according to key performance metrics. In the FetalMovNet architecture, ~70% (410) of the dataset was used as the training set and the remaining ~30% (177) as the test set for the training and test phases. Table 1 shows the number of video training and test sets for each anatomical structure of fetus. For all experimental studies in the study, a computer with Intel Core i5 4.10 GHz CPU, 16GB RAM, NVIDIA RTX A4000 16GB GPU and 1TB HDD + 500GB SSD disc hardware was used.

TABLE 1. The number of video training and test sets for each anatomical structure of fetus.

Anatomical structure of fetus	The number of videos in training set	The number of videos in test set	Total
Head	86	44	130
Body	68	31	99
Arm	57	25	82
Heart	54	20	74
Hand	58	16	74
Foot	43	21	64
Leg	44	20	64
Total	**410**	**177**	**587**

To ensure optimal performance and generalizability, FetalMovNet was trained using a well-defined strategy, incorporating hyperparameter tuning, optimizer selection, batch size considerations, activation functions, and training epochs. The proposed FetalMovNet deep learning network was trained for 50 epochs with the parameters given in Table 2 Model training was performed in the Keras environment using the Python programming language. During training, overfitting was prevented by using the early stopping method and training was stopped at the iterations where the model performed best. Batch size plays a crucial role in training efficiency, model performance and memory usage. When training the FetalMovNet network, the batch size was set to 4. This can lead to smoother convergence and improved generalization by introducing more noise into the gradient estimation, which prevents the model from overfitting. In addition, root mean square propagation (RMSprop) was used as an optimizer in the training of the network. RMSprop adjusts the learning rate for each weight based on the average of the recent magnitudes of the gradients for that weight. As CNNs often have many layers and parameters, RMSprop's ability to adjust learning rates helps to avoid vanishing or exploding gradient problems. RMSprop was chosen because it performs adaptive learning rate adjustments and works well with noisy datasets such as US videos. Categorical cross entropy is commonly used as a loss function for classification problems and measures the difference between the actual class and the predicted class for each class produced by the model. Categorical cross entropy is also suitable for multi-class classification. ReLU activation function avoids the vanishing gradient problem and speeds up convergence when used in the FCN block.

TABLE 2. Some hyperparameters and their values used in the training phase of the proposed fetalmovnet deep learning network.

Parameter	Value
Batch size	4
Optimizer	RMSprop
Loss	Categorical cross entropy
Epoch	50
Activation function	ReLU (in FCN block)

In this study, the layer structure in the architecture of the FetalMovNet deep learning network, which enables the classification of fetal movements from US video frames, is detailed in Table 3. In the model, each input video frame is subjected to attention processing with other frames in the sequence and a score is generated. These scores are passed through fully connected layers and combined with the result of the convolution layers. The softmax function is then used to complete the classification of fetal movement. By evaluating the loss and accuracy progress graphs in Fig. 10, it can be seen that the proposed FetalMovNet network can perform generalized learning on the sequential US image data generated when trained for 50 epochs. When the training time of the proposed FetalMovNet architecture in this study is evaluated, it is observed that the validation accuracy stabilizes after approximately 30 epochs. Similarly, it is observed that the training loss does not change much at the end of 30 epochs. Thus, it can be concluded that the proposed CNN and attention-based FetalMovNet architecture has a high performance in fetal movement detection from US videos on the generated dataset.

To evaluate the performance of the proposed attentional deep learning model FetalMovNet in classifying fetal movements, the outputs of different layers of the architecture can be considered. Fig. 11 shows some of the filters obtained at the output of the Conv4 layer of the FetalMovNet architecture. These filters can give a first idea of the training of the network and the results of the classification. The filters obtained from the Conv4 layer of the architecture extract more abstract, high-level features depending on the depth of the network. In particular, in deeper layers such as Conv4, the features presented by the filters highlight object parts, texture details, shapes or unique structural features in the input data due to the model's attention mechanism. Thus, the output of these filters may also contain the attentional focus that the model uses to identify salient and critical regions in the input data.

In the experimental studies, the key metrics Accuracy, Precision, Recall and F1-score were used to evaluate the performance of the proposed FetalMovNet architecture and other state-of-the-art models in the classification of fetal

TABLE 3. Layer structure of the fetalmovnet deep learning network for classification of fetal movements from us video frames.

Layer Type	Output Shape	Connected to
Input layer	(n,10,300,300)	-
Reshape	(n,300,300,10)	Input layer
Conv1	(n,300,300,64)	Reshape
Pooling1	(n,150,150,64)	Conv1
Conv2	(n,150,150,128)	Pooling1
Pooling2	(n,75,75,128)	Conv2
Conv3	(n,75,75,256)	Pooling2
Pooling3	(n, 37,37,256)	Conv4
Conv4	(n, 37,37,512)	Pooling3
Pooling4	(n, 18,18,512)	Conv4
Batch Norm1	(n, 9,9,256)	Pooling4
Flatten1	(n, 20736)	BatchNorm1
Dense1	(n, 256)	Flatten1
Reshape2	(n, 10,90000)	Input Layer
Dense(Q)	(n, 10,1024)	Reshape2
Dense(V)	(n, 10,1024)	Reshape2
Attention Layer	(n, 10,1024)	Dense(Q), Dense(V)
Flatten2	(n, 10240)	Attention Layer
Dense2	(n, 256)	Flatten2
BatchNorm2	(n, 256)	Dense2
Concatenate	**(n, 512)**	**Flatten1, BatchNorm2**
Dense3	(n,256)	Concatenate
DropOut1	(n,256)	Dense3
Dense4	(n,128)	Dropout1
Dropout2	(n,128)	Dense4
Dense5	(n,64)	Dropout2
SoftMax	(n,7)	Dense5

anatomical structures. These metrics are calculated using true positive (TP), true negative (TN), false positive (FP) and false negative (FN) obtained from the confusion matrix. Accuracy in Eq. (1) represents the correct prediction rate of the proposed model in all classifications, while Precision in Eq. (2) evaluates the correct positive predictions of the model. Recall in Eq. (3) gives the proportion of all positive examples correctly predicted by the model. Finally, the F1-score is the harmonic mean of the Precision and Recall. As this study is a multi-class classification problem, the average values of the metrics were taken into account by using macro-averaged in the calculation of the key metrics of Precision, Recall and F1-score. In the macro-averaged approach, when calculating a metric for all classes, the average of the values of this metric calculated on a class basis is taken into account [51].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

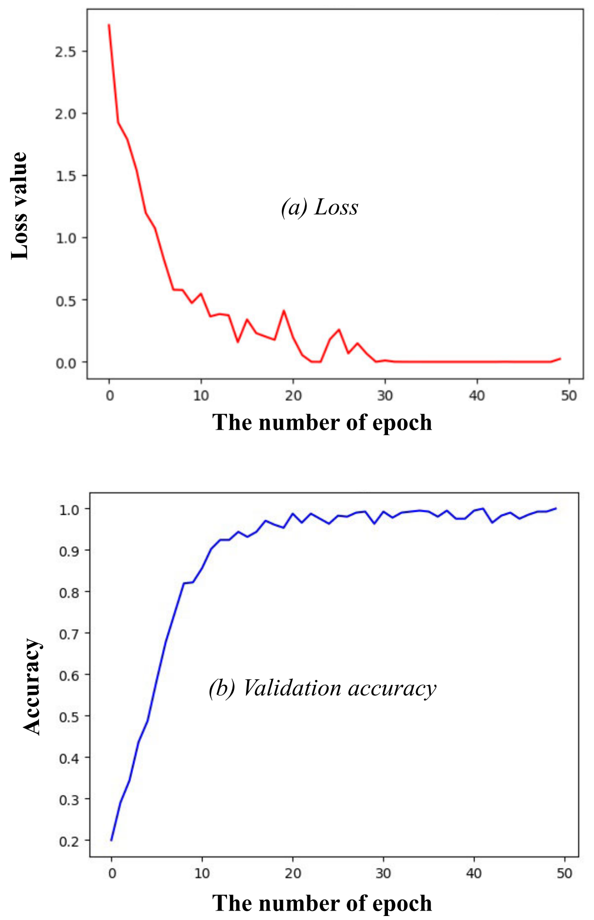

FIGURE 10. Variation of loss (a) and validation accuracy (b) values during the training of the proposed FetalMovNet architecture for 50 epochs.

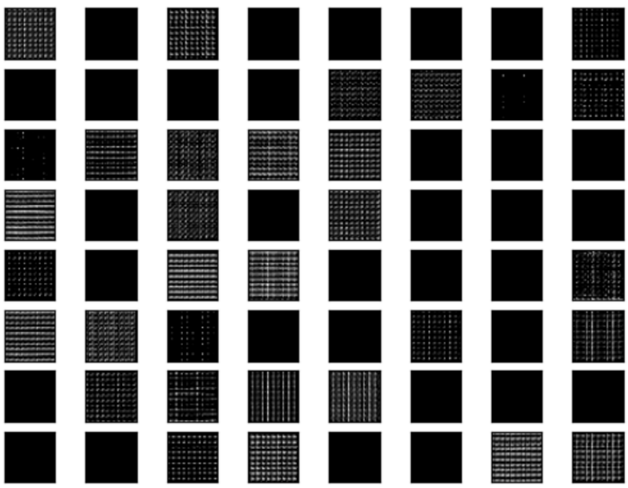

FIGURE 11. Filter patterns learnt at the output of the Conv4 layer of the FetalMovNet architecture.

For the FetalMovNet model proposed in this study, ablation was performed as part of experimental studies to evaluate the effect of different layers on performance results. The ablation process is used to evaluate the robustness of deep learning models and to understand which module shows how it works [52]. Firstly, in the FetalMovNet architecture proposed for ablation, a typical CNN model was created by removing the attention layer and using only the Conv and FCN layers. Secondly, another state-of-the-art architecture for ablation assessment was developed by combining a typical CNN architecture with the LSTM model. The performance of the results obtained with FetalMovNet, typical CNN and CNN-LSTM architectures were then compared. The confusion matrices showing the results obtained in classifying the anatomical structures of the fetus such as head, body, arm, heart, leg, foot and hand are shown in Fig. 12. While Fig. 12(a) shows the results obtained with only one CNN network, Fig. 12(b) shows the confusion matrix values of the classification results obtained with CNN-LSTM model. On the other hand, Fig. 12(c) denotes the confusion matrix for the results obtained using the proposed FetalMovNet architecture. With a typical CNN architecture, 1 US video for head, 4 for body, 3 for arm, 2 for leg, 2 for hand, 1 for heart and 2 for foot were misclassified. As a result, the CNN architecture misclassified 15 of the 177 US videos in the test set that contained movements of fetal anatomical structures. In addition, using CNN-LSTM, 13 out of a total of 177 US sequences in the test set were predicted to the wrong class. The CNN-LSTM architecture misclassified 3 for head, 2 for body, 3 for arm, 2 for leg and 3 for hand in US sequences in test set. On the other hand, for the classification of fetal movements, only 2 out of 177 US videos in the test set were misclassified using the proposed attention-based FetalMovNet architecture. 1 movement video in the head class was predicted in the body class and 1 movement video in the arm class was predicted in the hand class. When these three confusion matrices are evaluated together, it can be seen that the proposed FetalMovNet architecture is more successful than the typical CNN and CNN-LSTM architectures in classifying US videos consisting of 10-sequential frames containing fetal movements.

Table 4 shows the performance results of the typical CNN and the proposed FetalMovNet architectures for the test set using confusion matrices with key metrics. While the typical CNN and CNN-LSTM architectures achieved an average accuracy of 0.9153 and 0.9265, respectively, for anatomical structures in classifying fetal movements, the proposed FetalMovNet architecture achieved an accuracy of 0.9887. On the other hand, for the key metrics of precision, recall and F1-score, 0.9871, 0.9910 and 0.9891 scores were obtained using the proposed FetalMovNet architecture. These results show that the proposed FetalMovNet architecture is more successful in detecting and classifying the movements of fetal anatomical structures compared to CNN and CNN-LSTM architectures.

Receiver operating characteristic (ROC) curve can be used to assess the ability of the model to correctly classify fetal movements in US videos. It illustrates the balance between true positive and false positive rates. ROC curves provide a standardized way to compare their classification

TABLE 4. Comparison of the results obtained with typical CNN and proposed FetalMovNet architectures on the basis of key metrics after ablation procedure.

Model	Accuracy	Precision	Recall	F1-score
CNN	0.9153	0.9112	0.9083	0.9098
CNN-LSTM	0.9265	0.9191	0.9228	0.9209
FetalMovNet	0.9887	0.9871	0.9910	0.9891

performance. The area under the curve (AUC) can be used as a summary metric, with a higher AUC indicating a model that is better at discriminating between movement classes. Fig. 13 shows the AUC values obtained from the ROC curve for each anatomical structure in the test set. As can be seen in Fig. 13(a) and Fig. 13(b), an average AUC score of 0.9471 and 0.9543 for CNN and CNN-LSTM were obtained for the head, body, arm, hand, heart, leg and foot anatomical structures in the test set, respectively, while using the proposed FetalMovNet architecture, an average AUC score of 0.9957 is obtained for the same anatomical structures of the fetus, as seen in Fig. 13(c). On the other hand, using the proposed FetalMovNet architecture, AUC scores of 1.0 are achieved for the body, leg, heart, hand and foot movement classes, confirming the successful classification of fetal movements from US videos.

V. DISCUSSION

In this study, we introduced FetalMovNet, a novel deep learning model designed to classify fetal motion in US videos. The model uses a CNN-based architecture improved by an attention mechanism, which provides significant improvements in the accurate detection and classification of fetal anatomical movements compared to traditional CNN and CNN-LSTM architectures. The key advantage of FetalMovNet is the use of an attention mechanism that allows the model to focus on spatial and temporal patterns that are most relevant to fetal movement. This attention-based approach, combined with CNN layers, results in a model that can more accurately classify dynamic movements across frames, overcoming the limitations of conventional CNN models that can struggle to capture temporal dependencies in video data. Through a number of experiments, we have observed that FetalMovNet can distinguish between different fetal anatomical structures with high accuracy and reliability.

The ablation study showed that the inclusion of the attention mechanism is crucial for improving classification performance. When the attention layer was removed, leaving only the CNN and fully connected layers, the accuracy of the model decreased significantly, and the AUC for most classes also decreased, as shown in our confusion matrices. In addition, by using the CNN-LSTM architecture, which is a combination of CNN and LSTM, it was observed that the performance was slightly improved compared to the typical CNN. However, the results obtained with CNN-LSTM were still lower than the proposed FetalMovNet model. In contrast, FetalMovNet's attention layer allowed the

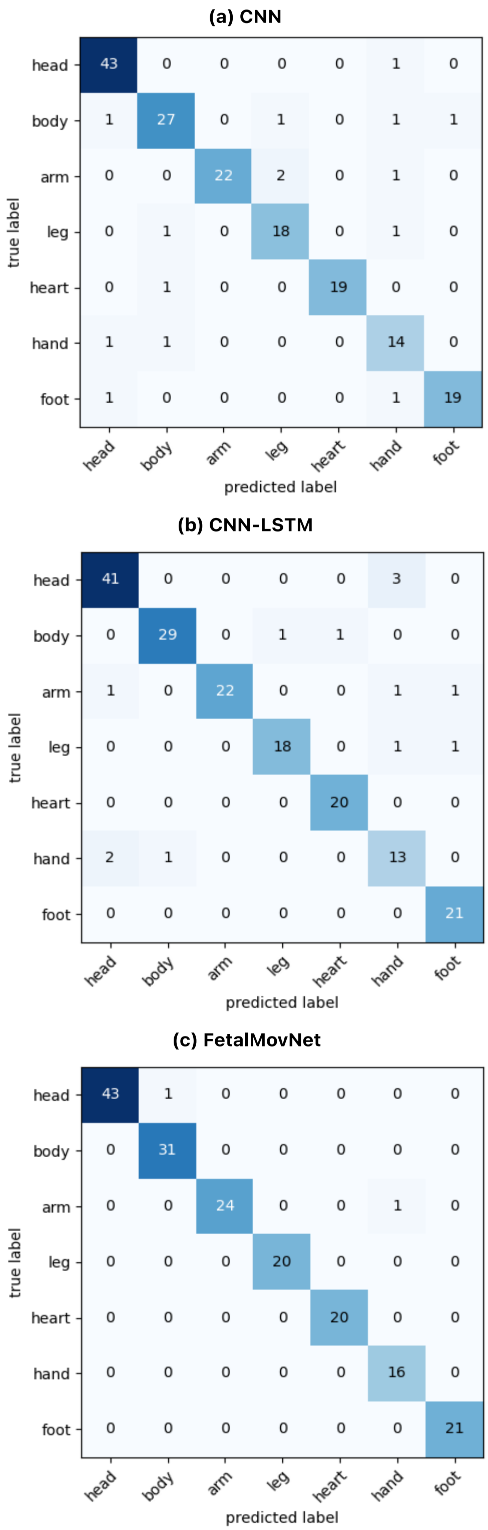

FIGURE 12. Confusion matrix results achieved using the models in test for the classification of the movements of the anatomical structures of the fetus. (a) a typical CNN architecture, (b) CNN-LSTM architecture, and (c) the proposed FetalMovNet architecture.

model to dynamically focus on relevant image sequences and regions, ensuring more accurate classification by minimizing false positives and false negatives. However, for these results to be meaningful, the statistical significance of the results obtained with CNN, CNN-LSTM and FetalMovNet needs to be evaluated and the results considered accordingly. There are many tests that can be used to assess the significance between data. The Wilcoxon signed-rank test is a non-parametric statistical test that evaluates the difference between the medians of two variables [53]. In this study, a two-sided Wilcoxon signed-rank test was used for the comparison and statistical significance of the results obtained using CNN and CNN-LSTM and the proposed FetalMovNet architectures. When comparing the results, the averages of the macro-averaged F1-scores obtained for each class using each model were taken into account. The comparison of typical CNN and FetalMovNet according to the results of the Wilcoxon signed-rank test yielded a p-value of 0.012. On the other hand, when comparing the Wilcoxon signed-rank test results of CNN-LSTM and FetalMovNet, the p-value was 0.028. Since the p-values obtained in both comparisons are less than 0.05, it can be said that the results obtained are statistically significant according to the Wilcoxon signed-rank test.

Using the FetalMovNet architecture proposed in this study, 175 of the 177 fetal US videos in the test set of fetal anatomical structures were correctly classified. However, as can be seen from some of the frames in Fig. 14, 1 US video containing fetal movement belonging to the Head anatomical structure was incorrectly classified as Body. On the other hand, another US video belonging to the Arm class is also misclassified as Hand. The reason for the incorrect recognition of these fetal movements by the proposed FetalMovNet architecture is that the sequential movement patterns in these videos are similar to the patterns of the incorrectly recognized class movements.

Previous studies have proposed various approaches to the detection and measurement of fetal anatomical structures from US scans. Most of these studies generated original study specific datasets. In one of these studies, Carneiro et al. [22] used a probabilistic boosting tree to detect and measure fetal anatomy such as head, body, abdomen, femur and humerus. Cuingnet et al. [64] used random forests and template deformation from 3D US images to determine the position of the fetal head with an average error of 5.8 mm. On the other hand, Ravishankar et al. [38] provided automatic segmentation of the fetal abdomen using a CNN-based hybrid approach. Sinclair et al. [29] proposed segmentation and detection of fetal head biometrics using CNN architecture on 2D US images. Alzubaidi et al [65] proposed FetSAM deep learning model for segmentation of fetal head biometrics from US scans. In addition, there are studies [24] proposed for quality assessment in fetal US scans and studies [66] proposed for data augmentation in US data for standard plane classification. Furthermore, in one of our previous studies, Dandıl et al. [67] provided anatomical plane recognition in fetal US scans for head, arm, heart and body using YOLOv5. In our other study, Turkan et al. [68] proposed fetal motion detection from trajectories for anatomical structures such as body, head, arm and heart using YOLOv5 and LSTM. In this

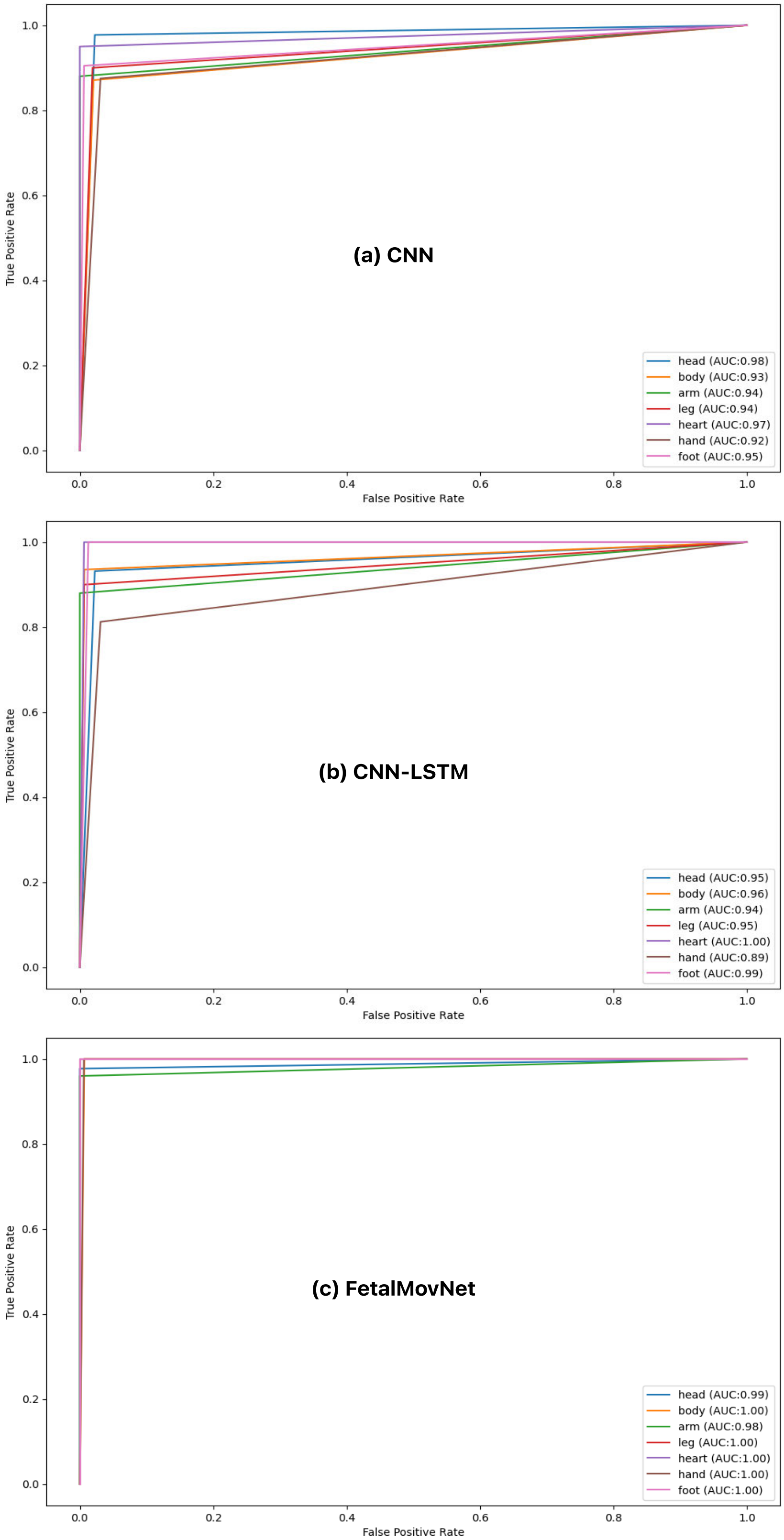

FIGURE 13. ROC curves and AUC scores of (a) typical CNN, (b) CNN-LSTM and (c) proposed FetalMovNet architectures for each anatomical structure in the testset for classification of fetal anatomical structures.

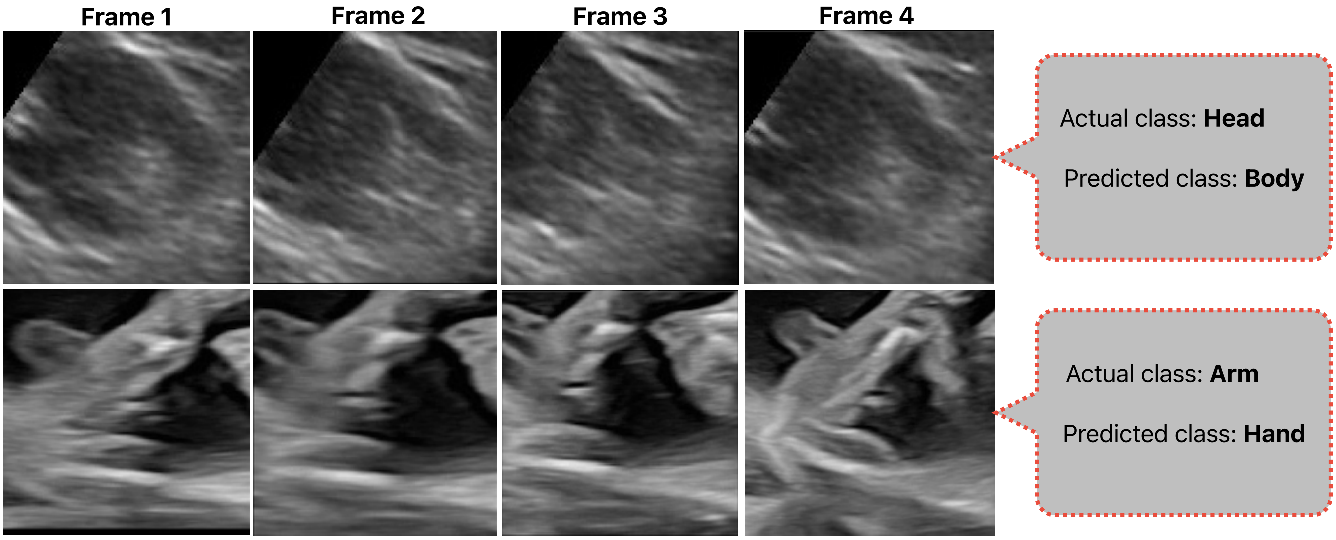

FIGURE 14. Evaluation of consecutive frames of incorrectly classified fetal movements in the test set using the proposed FetalMovNet architecture.

TABLE 5. Comparison of the results of the FetalMovNet model with existing studies previously proposed for the recognition and classification of fetal anatomical structures.

Study and Year	Dataset	The number of data	Method	Evaluation for anatomical structure of fetus	Performance accuracy
Linares et al. [54], 2004	Original dataset	101 US scans	*k*-NN based on feature selection via Relief-F	Texture classification using placenta US images	0.6071
Yaqub et al. [55], 2012	Original dataset	40 3D US volume	Discriminative model based on random forests	Automatic detection of fetal brain structures	0.9290
Yu et al. [56], 2016	Original dataset	1735 US images	deep convolutional networks (DCNN)	Recognition of fetal facial standard plane	0.9699
Gao et al. [43], 2016	Original dataset	323 fetal US video	CNN	Describing of fetus anatomical structures such as skull, abdomen, heart and others	0.9150
Yu et al. [57], 2018	Original dataset (FFSP dataset)	7267 US images	Deep CNN	Fetal facial standard plane recognition	0.9653
Lin et al. [58], 2019	Original dataset	1771 US images	multi-task learning framework using a faster R-CNN	Fetal head classification	0.9625
Burgos-Artizzu et al.[20], 2020	Original dataset	12400 US images	Several CNN architectures	classification of maternal fetal planes such as abdomen, brain, femur, cervix, thorax and others	0.9360
Qu et al. [59], 2020	Original dataset	19142 US images	Differential CNN	Standard plane identification in fetal brain	0.9311
Zhang et al. [60], 2021	Original dataset	4101 US images	CNN-based approaches	Assessment of fetal standard planes such as head, abdomen and heart	0.9671
Li et al. [61], 2024	Original dataset (video)	3360 2D fetal US images	FHUSP-NET (a deep learning architecture based on YOLO)	Fetal standard plane recognition for heart anatomical structure	0.9640
Yasrab et al. [62], 2024	PULSE dataset [63]	250 US video	CNN	Fetal tissue recognition and fetal biometry measurement	0.8700
Proposed	Original US dataset (video)	587 (each US video contains 10 frames)	FetalMovNet	Automatic classification of fetal anatomical structures such as head, body, heart, arm, foot, leg and hand	0.9887

study, unlike our previous studies, the classification of fetal motion by detection of anatomical structures from fetal US videos is achieved using FetalMovNet, a novel proposed architecture.

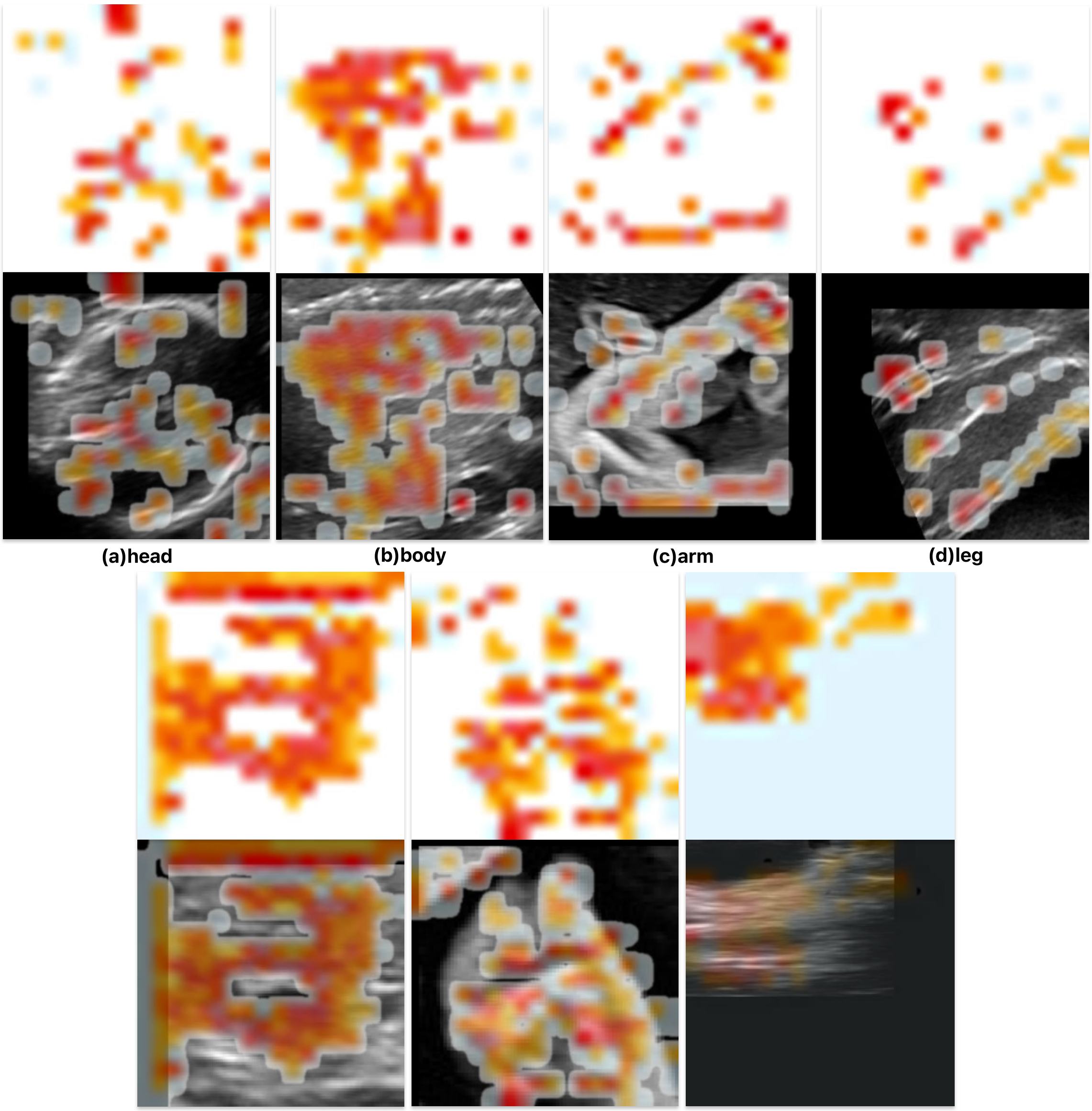

FIGURE 15. The heatmap of feature outputs of FetalMovNet for fetal anatomical structures such as (a) heart, (b) body, (c) arm, (d) leg, (e) heart, (f) hand, and (g) foot in a frame of US sequences with Grad-CAM.

The results of the FetalMovNet model are compared with existing studies previously proposed for the detection and classification of fetal anatomical structures, as shown in Table 5. Although CNN and similar architectures have been used in previous studies to successfully detect and classify fetal anatomical structures, it is observed that the classification performance remains within a certain accuracy and these methods usually focus on basic classification rather than understanding the anatomical context of the movements. In contrast, the FetalMovNet model outperforms the other approaches with an average accuracy of 0.9887, effectively learning both spatial and temporal features through the attention mechanism. These results show that the proposed model makes a significant contribution not only to the detection of fetal movement, but also to the accurate classification of anatomical structures.

For clinical adoption, deep learning models need to be explainable to radiologists and obstetricians. While attention mechanisms improve accuracy, it is sometimes difficult for clinicians to trust automated decisions. Therefore, visualizing how the automated system makes decisions contributes to the understandability of the process. In this study, gradient-weighted class activation mapping (Grad-CAM), a discriminative localization technique, is used to show which spatio-temporal features the proposed FetalMovNet architecture takes into account for the classification of fetal movements using US sequences, and to generate heatmaps of these important features. Grad-CAM is widely used

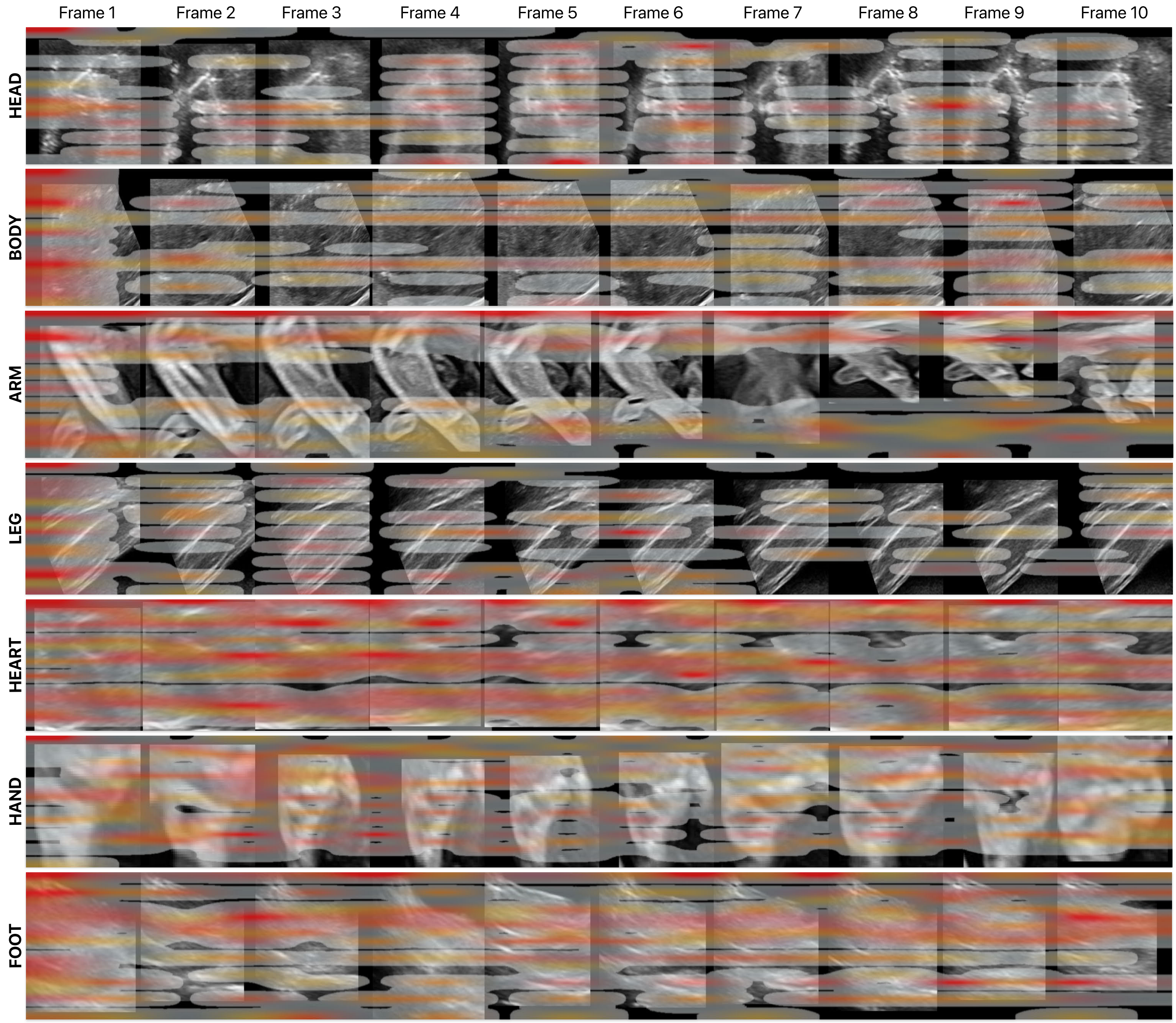

FIGURE 16. The heatmap of feature outputs of FetalMovNet for fetal anatomical structures such as heart, body, arm, leg, heart, hand, and foot in US sequences with Grad-CAM.

in classification problems to highlight key regions of an image or video frame to enable prediction of a particular class [69], [70]. In addition, Grad-CAM can be adapted to any CNN model on images and can provide a more transparent heatmap through the generation of visual descriptions of the images [71]. Grad-CAM provides a mechanism to highlight the key regions of the image/frame that contribute to the prediction of a particular class. Rather than being distracted by surrounding fetal anatomical structures, Grad-CAM prompts the model to increase its attention to regions that contribute most to the correct classification of the fetal movements. Grad-CAM highlights salient regions in US images that contribute to fetal movement classification and helps clinicians understand which anatomical features influence model predictions. Because Grad-CAM provides spatial heatmaps overlaid on US images, it shows which temporal frames contribute most to fetal movement classification. Fig. 15 shows the heatmap of the feature outputs of FetalMovNet for fetal anatomical structures in one frame of US sequences with Grad-CAM. It can be seen that the heatmaps of the fetal anatomical structures of heart, body, arm, leg, heart, hand and foot from a single frame of US sequences with Grad-CAM are emphasized and highlighted, especially the regions of movement. On the other hand, Fig. 16 shows the heatmap of FetalMovNet's feature outputs for fetal anatomical structures such as heart, body, arm, leg, heart, hand and foot in US sequences with Grad-CAM. It can be seen that the features of the regions of interest with Grad-CAM are similar to the heatmaps of the movements in sequential 10-frame sequences. In conclusion, the classification result of FetalMovNet and the heatmap features output by Grad-CAM are compatible. By integrating Grad-CAM into the FetalMovNet architecture, the model gained the ability to better understand and highlight the characteristic details of fetal anatomical structures.

FetalMovNet integrates CNN-based feature extraction and an attention mechanism, which improves accuracy but can increase computational requirements. Therefore, as the number of video in the test set increases, computational time may increase in real-time clinical applications. The speed at which the model processes US images is critical. To provide meaningful clinical insights, FetalMovNet must operate at real-time or near real-time speeds. The total test/computation time of the proposed FetalMovNet model on a total of 177 sequential 10-frame US sequences in the test set is 11.6 seconds. This results in an average of 65 ms for a single US sequence. The latency for a single US frame is therefore 6.5 ms. Latency is critical for real time applications and the average time taken per frame is acceptable as 10 ms - 50 ms for each frame. Therefore, FetalMovNet meets real-time requirements ($\leq$ 100ms latency per frame) and can be integrated into US machines for real-time monitoring. On the other hand, real-time inference can be achieved by using high-end GPUs.

VI. CONCLUSION

In this study, a proposed deep learning model, FetalMovNet, based on the attention mechanism, is used to detect fetal anatomical structures such as head, body, arm, hand, foot, leg and heart from US videos and to classify fetal movements of these structures. A CNN network was used to extract features at each level in the sequence of US videos. The attention layer is used to capture the important connections between the video frames, focusing on the differences caused by the movements. The results of this study demonstrate that FetalMovNet is an effective tool for fetal movement classification, offering higher accuracy and robustness than traditional CNN approaches. Through the automation of fetal movement assessment with reliable and interpretable models, FetalMovNet has the potential to improve antenatal care and make fetal monitoring more accurate, accessible and objective.

Fetal movements are an important indicator of fetal well-being. Recognition fetal movements in this study can help monitor fetal well-being during pregnancy. In addition to providing a baseline for automated fetal health assessment, this study highlights the value of attention-based deep learning models in medical image analysis. With further validation and optimization, FetalMovNet could become an indispensable tool in clinical practice, helping clinicians to improve pregnancy outcomes through timely and accurate analysis of fetal movement. The clinical implications of FetalMovNet are promising. In traditional fetal health monitoring, the assessment of fetal movement relies heavily on subjective and manual observation, which is prone to error and time constraints. By automating this process, FetalMovNet could help clinicians make faster and more accurate assessments of fetal well-being, reducing the need for continuous manual monitoring.

Abnormal fetal movement patterns can indicate conditions such as fetal distress, neurological and cardiac abnormalities. In addition, with its high accuracy and precision, FetalMovNet has the potential to alert clinicians to abnormal movement patterns, enabling earlier intervention in cases where fetal distress or developmental problems may be present. In addition, expectant mothers often worry about fetal activity and rely on subjective perceptions of movement. FetalMovNet could provide objective feedback, reducing unnecessary anxiety and helping to detect abnormalities early.

The FetalMovNet model proposed in this study for the classification of fetal movements from US sequences has several strengths. Firstly, the model achieves higher performance for accuracy, precision, recall, and F1-score, suggesting excellent classification capability when compared with state-of-the-art CNN and CNN-LSTM models. On the other hand, the integration of an attention mechanism into FetalMovNet improves the spatio-temporal feature extraction with a higher AUC score. Furthermore, the study introduces a new dataset with seven different anatomical movements, ensuring a diverse representation of fetal movements. Using this dataset, FetalMovNet was shown to generalize well to different anatomical movements, highlighting its potential as a tool for real-time clinical monitoring.

Despite its strengths, FetalMovNet has several limitations that could be acknowledged. Firstly, the model was trained on a specific dataset, which may limit its generalizability across different populations and imaging setups. Therefore, performance may decrease when applied to US devices with different imaging parameters or when used in real clinical settings. The model may struggle with unseen real-world cases and requires further validation in larger, multicenter datasets. To overcome this, FetalMovNet can be trained on multi-institutional datasets with different US machines, imaging protocols and patient demographics, using domain adaptation techniques such as adversarial training or feature alignment. In addition, model quantization and pruning can be used to reduce memory and computational requirements. However, data from different hospitals and machines can have significant variations in resolution, contrast and noise, requiring robust pre-processing techniques. Secondly, in experimental studies, the average computation time for a single US sequence in the test set is 65 ms. The use of attention mechanisms and deep CNNs increases the computational burden, potentially making real-time deployment on standard US devices difficult. Therefore, as the number of video in the test set increases, the computational time of the FetalMovNet model may increase in real-time clinical applications. Another limitation is that the study mainly compared FetalMovNet with CNN and CNN-LSTM, but not with other state-of-the-art deep learning models used in medical imaging. This may limit a broader assessment of its competitiveness against more advanced architectures. Finally, the video frame sequences in the dataset are limited to 10 frames. In this case, it may be difficult to detect long-period movements that span multiple frames. A solution to this problem could be to increase the number of frames used in the dataset or to sample the video frame sequence at a certain rate. This may lead to overfitting of the model if the training time is increased.

Future research could extend this study in a number of ways. First, additional anatomical classes and movement patterns could be added to the dataset to make FetalMovNet more comprehensive and adaptable to a wider range of fetal assessments. In addition, future implementations could benefit from optimizing the model to reduce inference time without compromising accuracy, thereby enabling its use in resource-constrained environments. Furthermore, future work could investigate the use of alternative attentional mechanisms, such as self-attention or multi-head attention, which may provide additional improvements in handling complex motion sequences. On the other hand, detailed analyses of fetal US videos may help to identify movements that are characteristic of different types of disease. The enrichment of the generated dataset may allow the use of motion analysis for different tasks. Finally, the nature of motion of anatomical structures in US videos differs from normal movement. The movement in the video may vary depending on how the user uses the US device and the quality of the device's detection of anatomical structures. Movement can be caused not only by the anatomical structure itself, but also by the movements of the device user.

ACKNOWLEDGMENT

The authors would like to thank Kütahya Evliya Çelebi Training and Research Hospital for sharing US dataset for fetal movement classification and all the patients who volunteered to participate.

REFERENCES

[1] A. B. Lüchinger, M. Hadders-Algra, C. M. van Kan, and J. I. P. de Vries, "Fetal onset of general movements," *Pediatric Res.*, vol. 63, no. 2, pp. 191–195, Feb. 2008.

[2] I. Tsakiridis, C. Zerva, A. Mamopoulos, I. Kalogiannidis, A. Athanasiadis, and T. Dagklis, "Maternal perception of fetal movements: Onset and associated factors," *J. Perinatal Med.*, vol. 50, no. 9, pp. 1174–1179, Nov. 2022.

[3] J. M. Turner, V. Flenady, D. Ellwood, M. Coory, and S. Kumar, "Evaluation of pregnancy outcomes among women with decreased fetal movements," *JAMA Netw. Open*, vol. 4, no. 4, Apr. 2021, Art. no. e215071.

[4] A. Bekiou and K. Gourounti, "Reduced fetal movements and perinatal mortality," *Materia Socio Medica*, vol. 32, no. 3, p. 227, 2020.

[5] J. M. Turner, R. Cincotta, J. Chua, G. Gardener, S. Petersen, J. Thomas, A. Lee-Tannock, and S. Kumar, "Decreased fetal movements—The utility of ultrasound to identify infants at risk and prevent stillbirth is poor," *Amer. J. Obstetrics Gynecol. MFM*, vol. 5, no. 2, Feb. 2023, Art. no. 100782.

[6] A. Kadic and A. Kurjak, "Cognitive functions of the fetus," *Ultraschall Der Medizin-Eur. J. Ultrasound*, vol. 39, no. 2, pp. 181–189, Apr. 2018.

[7] Y. Bai, "Object tracking & fetal signal monitoring," Doctor of Philosophy Dissertion, Dept. Elect. Comput. Eng., Graduate School, Southern Illinois Univ., Carbondale, IL, USA, 2016. [Online]. Available: https://www.proquest.com/docview/1830772289?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

[8] J. You, Q. Li, Z. Guo, and R. Zhao, "Smart fetal monitoring," in *Proc. Int. Conf. Inf. Sci. Appl.*, Cham, Swizerland. Springer, Jan. 2017, pp. 494–503.

[9] F. Ovalı, "Fetus ve yenidoğanda işitme: Temel kavramlar ve perspektifler," *Turkiye Klinikleri J. Pediatr*, vol. 14, no. 3, pp. 138–149, 2005.

[10] C. L. Whitehead, N. Cohen, G. H. A. Visser, and D. Farine, "Are increased fetal movements always reassuring?" *J. Maternal-Fetal Neonatal Med.*, vol. 33, no. 21, pp. 3713–3718, Nov. 2020.

[11] M. D. Velazquez and W. F. Rayburn, "Antenatal evaluation of the fetus using fetal movement monitoring," *Clin. Obstetrics Gynecology*, vol. 45, no. 4, pp. 993–1004, Dec. 2002.

[12] Q. Huang, F. Zhang, and X. Li, "Machine learning in ultrasound computer-aided diagnostic systems: A survey," *BioMed Res. Int.*, vol. 2018, no. 1, pp. 1–10, 2018.

[13] S. Saha, "Ultrasound image processing for CAD and other applications," M.Tech. dissertation, Dept. Elect. Eng., Indian Inst. Technol., Mumbai, India, 2014.

[14] J. Zhang and Z. Liu, "Detecting abnormal motion of pedestrian in video," in *Proc. Int. Conf. Inf. Autom.*, Jun. 2008, pp. 81–85.

[15] J. Zhang and Z. J. Liu, "Abnormal behavior of pedestrian detection based on fuzzy theory," Deakin Univ., 2023.

[16] S. Vosta and K.-C. Yow, "KianNet: A violence detection model using an attention-based CNN-LSTM structure," *IEEE Access*, vol. 12, pp. 2198–2209, 2024.

[17] M. Majd and R. Safabakhsh, "Correlational convolutional LSTM for human action recognition," *Neurocomputing*, vol. 396, pp. 224–229, Jul. 2020.

[18] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261–275, Feb. 2019, doi: 10.1016/j.eng.2018.11.020.

[19] T. L. A. van den Heuvel, D. de Bruijn, C. L. de Korte, and B. V. Ginneken, "Automated measurement of fetal head circumference using 2D ultrasound images," *PLoS ONE*, vol. 13, no. 8, Aug. 2018, Art. no. e0200412.

[20] X. P. Burgos-Artizzu, D. Coronado-Gutiérrez, B. Valenzuela-Alcaraz, E. Bonet-Carne, E. Eixarch, F. Crispi, and E. Gratacós, "Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes," *Sci. Rep.*, vol. 10, no. 1, p. 10200, Jun. 2020.

[21] X. Chen, M. He, T. Dan, N. Wang, M. Lin, L. Zhang, J. Xian, H. Cai, and H. Xie, "Automatic measurements of fetal lateral ventricles in 2D ultrasound images using deep learning," *Frontiers Neurol.*, vol. 11, p. 526, Jul. 2020.

[22] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, pp. 1342–1355, Sep. 2008, doi: 10.1109/TMI.2008.928917.

[23] M. C. Fiorentino, F. P. Villani, M. Di Cosmo, E. Frontoni, and S. Moccia, "A review on deep-learning algorithms for fetal ultrasound-image analysis," *Med. Image Anal.*, vol. 83, Jan. 2023, Art. no. 102629.

[24] L. Wu, J.-Z. Cheng, S. Li, B. Lei, T. Wang, and D. Ni, "FUIQA: Fetal ultrasound image quality assessment with deep convolutional networks," *IEEE Trans. Cybern.*, vol. 47, no. 5, pp. 1336–1349, May 2017.

[25] D. Ouyang, B. He, A. Ghorbani, N. Yuan, J. Ebinger, C. P. Langlotz, P. A. Heidenreich, R. A. Harrington, D. H. Liang, E. A. Ashley, and J. Y. Zou, "Video-based AI for beat-to-beat assessment of cardiac function," *Nature*, vol. 580, no. 7802, pp. 252–256, Apr. 2020.

[26] M. Ahmed and J. A. Noble, "Fetal ultrasound image classification using a bag-of-words model trained on sonographers' eye movements," *Proc. Comput. Sci.*, vol. 90, pp. 157–162, Jan. 2016.

[27] P. Looney, G. N. Stevenson, K. H. Nicolaides, W. Plasencia, M. Molloholli, S. Natsis, and S. L. Collins, "Automatic 3D ultrasound segmentation of the first trimester placenta using deep learning," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 279–282.

[28] J. Wróbel, T. Kupka, K. Horoba, A. Matonia, D. Roj, and J. Jeżewski, "Automated detection of fetal movements in Doppler ultrasound signals versus maternal perception," *J. Med. Informat. Technol.*, vol. 23, pp. 1–8, Jan. 2014.

[29] M. Sinclair, C. F. Baumgartner, J. Matthew, W. Bai, J. C. Martinez, Y. Li, S. Smith, C. L. Knight, B. Kainz, J. Hajnal, A. P. King, and D. Rueckert, "Human-level performance on automatic head biometrics in fetal ultrasound using fully convolutional neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 714–717.

[30] Z. Sobhaninia, S. Rafiei, A. Emami, N. Karimi, K. Najarian, S. Samavi, and S. M. Reza Soroushmehr, "Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 6545–6548.

[31] A. Dozen, M. Komatsu, A. Sakai, R. Komatsu, K. Shozu, H. Machino, S. Yasutomi, T. Arakaki, K. Asada, S. Kaneko, R. Matsuoka, D. Aoki, A. Sekizawa, and R. Hamamoto, "Image segmentation of the ventricular septum in fetal cardiac ultrasound videos based on deep learning using time-series information," *Biomolecules*, vol. 10, no. 11, p. 1526, Nov. 2020.

[32] A. Patra and J. A. Noble, "Multi-anatomy localization in fetal echocardiography videos," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1761–1764.

[33] J. Kaliappan, A. R. Bagepalli, S. Almal, R. Mishra, Y.-C. Hu, and K. Srinivasan, "Impact of cross-validation on machine learning models for early detection of intrauterine fetal demise," *Diagnostics*, vol. 13, no. 10, p. 1692, May 2023.

[34] A. Kuzu and Y. Santur, "Early diagnosis and classification of fetal health status from a fetal cardiotocography dataset using ensemble learning," *Diagnostics*, vol. 13, no. 15, p. 2471, Jul. 2023.

[35] P. Garcia-Canadilla, S. Sanchez-Martinez, F. Crispi, and B. Bijnens, "Machine learning in fetal cardiology: What to expect," *Fetal Diagnosis Therapy*, vol. 47, no. 5, pp. 363–372, 2020.

[36] G. Ishikawa, R. Xu, J. Ohya, and H. Iwata, "Detecting a fetus in ultrasound images using Grad CAM and locating the fetus in the uterus," in *Proc. 8th Int. Conf. Pattern Recognit. Appl. Methods*, 2019, pp. 181–189.

[37] T. L. A. van den Heuvel, H. Petros, S. Santini, C. L. de Korte, and B. van Ginneken, "Automated fetal head detection and circumference estimation from free-hand ultrasound sweeps using deep learning in resource-limited countries," *Ultrasound Med. Biol.*, vol. 45, no. 3, pp. 773–785, Mar. 2019.

[38] H. Ravishankar, S. M. Prabhu, V. Vaidya, and N. Singhal, "Hybrid approach for automatic segmentation of fetal abdomen from ultrasound images using deep learning," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 779–782.

[39] M. Yaqub, B. Kelly, A. T. Papageorghiou, and J. A. Noble, "A deep learning solution for automatic fetal neurosonographic diagnostic plane verification using clinical standard constraints," *Ultrasound Med. Biol.*, vol. 43, no. 12, pp. 2925–2933, Dec. 2017.

[40] H. Chen, Q. Dou, D. Ni, J. Cheng, J. Qin, S. Li, and P. Heng, "Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.*, Cham, Swizerland. Springer, Jan. 2015, pp. 507–514.

[41] R. Arnaout, L. Curran, Y. Zhao, J. Levine, E. Chinn, and A. Moon-Grady, "Expert-level prenatal detection of complex congenital heart disease from screening ultrasound using deep learning," *MedRxiv*, vol. 2020, pp. 1–45, Jun. 2020.

[42] S. S. Płotka, M. K. Grzeszczyk, P. I. Szenejko, K. Żebrowska, N. A. Szymecka-Samaha, T. Łęgowik, M. A. Lipa, K. Kosińska-Kaczyńska, R. Brawura-Biskupski-Samaha, I. Išgum, C. I. Sánchez, and A. Sitek, "Deep learning for estimation of fetal weight throughout the pregnancy from fetal abdominal ultrasound," *Amer. J. Obstetrics Gynecol. MFM*, vol. 5, no. 12, Dec. 2023, Art. no. 101182.

[43] Y. Gao, M. A. Maraci, and J. A. Noble, "Describing ultrasound video content using deep convolutional neural networks," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 787–790.

[44] M. Kaur, D. Singh, A. A. Alzubi, A. Shankar, and U. Rawat, "DARNet: Deep attention module and residual block-based lung and colon cancer diagnosis network," *IEEE J. Biomed. Health Informat.*, early access, Nov. 20, 2024, doi: 10.1109/JBHI.2024.3502636.

[45] D. Singh, R. R. Regatte, and M. V. Zibetti, "HDNLS: Hybrid deep-learning and non-linear least squares-based method for fast multi-component T1ρ mapping in the knee joint," *Bioengineering*, vol. 12, no. 1, p. 8, 2024.

[46] M. Corbetta and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature Rev. Neurosci.*, vol. 3, no. 3, pp. 201–215, Mar. 2002.

[47] R. Azad, L. Niggemeier, M. Hüttemann, A. Kazerouni, E. K. Aghdam, Y. Velichko, U. Bagci, and D. Merhof, "Beyond self-attention: Deformable large kernel attention for medical image segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2024, pp. 1276–1286.

[48] F. Chen, Z. Luo, L. Zhou, X. Pan, and Y. Jiang, "Comprehensive survey of model compression and speed up for vision transformers," 2024, *arXiv:2404.10407*.

[49] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Jun. 2017, pp. 5998–6008.

[50] Y.-T. Su, Y. Lu, M. Chen, and A.-A. Liu, "Spatiotemporal joint mitosis detection using CNN-LSTM network in time-lapse phase contrast microscopy images," *IEEE Access*, vol. 5, pp. 18033–18041, 2017.

[51] S. Farhadpour, T. A. Warner, and A. E. Maxwell, "Selecting and interpreting multiclass loss and accuracy assessment metrics for classifications with class imbalance: Guidance and best practices," *Remote Sens.*, vol. 16, no. 3, p. 533, Jan. 2024.

[52] R. Meyes, M. Lu, C. W. de Puiseau, and T. Meisen, "Ablation studies in artificial neural networks," 2019, *arXiv:1901.08644*.

[53] J. D. Cunha, S. Perdikis, S. Halder, and R. Scherer, "Post-adaptation effects in a motor imagery brain–computer interface online coadaptive paradigm," *IEEE Access*, vol. 9, pp. 41688–41703, 2021.

[54] P. A. Linares, P. J. McCullagh, N. D. Black, and J. Dornan, "Feature selection for the characterization of ultrasonic images of the placenta using texture classification," in *Proc. 2nd IEEE Int. Symp. Biomed. Imag., Macro Nano*, vol. 2, Apr. 2004, pp. 1147–1150.

[55] M. Yaqub, R. Napolitano, C. Ioannou, A. T. Papageorghiou, and J. A. Noble, "Automatic detection of local fetal brain structures in ultrasound images," in *Proc. 9th IEEE Int. Symp. Biomed. Imag. (ISBI)*, May 2012, pp. 1555–1558.

[56] Z. Yu, D. Ni, S. Chen, S. Li, T. Wang, and B. Lei, "Fetal facial standard plane recognition via very deep convolutional networks," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 627–630.

[57] Z. Yu, E.-L. Tan, D. Ni, J. Qin, S. Chen, S. Li, B. Lei, and T. Wang, "A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 3, pp. 874–885, May 2018.

[58] Z. Lin, S. Li, D. Ni, Y. Liao, H. Wen, J. Du, S. Chen, T. Wang, and B. Lei, "Multi-task learning for quality assessment of fetal head ultrasound images," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101548.

[59] R. Qu, G. Xu, C. Ding, W. Jia, and M. Sun, "Standard plane identification in fetal brain ultrasound scans using a differential convolutional neural network," *IEEE Access*, vol. 8, pp. 83821–83830, 2020.

[60] B. Zhang, H. Liu, H. Luo, and K. Li, "Automatic quality assessment for 2D fetal sonographic standard plane based on multitask learning," *Medicine*, vol. 100, no. 4, Jan. 2021, Art. no. e24427.

[61] F. Li, P. Li, X. Wu, P. Zeng, G. Lyu, Y. Fan, P. Liu, H. Song, and Z. Liu, "FHUSP-NET: A multi-task model for fetal heart ultrasound standard plane recognition and key anatomical structures detection," *Comput. Biol. Med.*, vol. 168, Jan. 2024, Art. no. 107741.

[62] R. Yasrab, H. Zhao, Z. Fu, L. Drukker, A. T. Papageorghiou, and J. A. Noble, "Automating the human action of first-trimester biometry measurement from real-world freehand ultrasound," *Ultrasound Med. Biol.*, vol. 50, no. 6, pp. 805–816, Jun. 2024.

[63] L. Drukker, H. Sharma, R. Droste, M. Alsharid, P. Chatelain, J. A. Noble, and A. T. Papageorghiou, "Transforming obstetric ultrasound into data science using eye tracking, voice recording, transducer motion and ultrasound video," *Sci. Rep.*, vol. 11, no. 1, p. 14109, Jul. 2021.

[64] R. Cuingnet, O. Somphone, B. Mory, R. Prevost, M. Yaqub, R. Napolitano, A. Papageorghiou, D. Roundhill, J. A. Noble, and R. Ardon, "Where is my baby? A fast fetal head auto-alignment in 3D-ultrasound," in *Proc. IEEE 10th Int. Symp. Biomed. Imag.*, Apr. 2013, pp. 768–771.

[65] M. Alzubaidi, U. Shah, M. Agus, and M. Househ, "FetSAM: Advanced segmentation techniques for fetal head biometrics in ultrasound imagery," *IEEE Open J. Eng. Med. Biol.*, vol. 5, pp. 281–295, 2024.

[66] L. H. Lee, Y. Gao, and J. A. Noble, "Principled ultrasound data augmentation for classification of standard planes," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, Cham, Swizerland. Springer, Jan. 2021, pp. 729–741.

[67] E. Dandıl, M. Turkan, F. E. Urfalı, İ. Biyik, and M. Korkmaz, ''Fetal movement detection and anatomical plane recognition using YOLOv5 network in ultrasound scans,'' *Avrupa Bilim Ve Teknoloji Dergisi*, vol. 2021, no. 26, pp. 208–216, 2021.

[68] M. Turkan, F. E. Urfalı, and E. Dandıl, ''Automatic fetal motion detection from trajectory of U.S. videos based on YOLOv5 and LSTM,'' in *Proc. Explainable Mach. Learn. Multimedia Based Healthcare Appl.*, Cham, Swizerland. Springer, Jan. 2023, pp. 1–20.

[69] D. Tang, J. Chen, L. Ren, X. Wang, D. Li, and H. Zhang, ''Reviewing CAM-based deep explainable methods in healthcare,'' *Appl. Sci.*, vol. 14, no. 10, p. 4124, May 2024.

[70] E. Dandıl, B. T. Baştuğ, M. S. Yıldırım, K. Çorbacı, and G. Güneri, ''MaskAppendix: Backbone-enriched mask R-CNN based on grad-CAM for automatic appendix segmentation,'' *Diagnostics*, vol. 14, no. 21, p. 2346, Oct. 2024.

[71] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, ''Grad-CAM: Visual explanations from deep networks via gradient-based localization,'' *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Feb. 2020.

MUSA TURKAN received the bachelor's degree in electronics education from Sakarya University, in 2006, and the master's degree in computer engineering from the Institute of Science, Bilecik Şeyh Edebali University, Bilecik, Türkiye, in 2014. Currently, he teaches with the Department of Management Information Systems and Computer Programming, Bilecik Şeyh Edebali University. His research interests include deep learning, software development, and image processing.

EMRE DANDIL received the bachelor's degree in electronic and computer sciences, in 2007, the M.Sc. degree from Suleyman Demirel University, Türkiye, and the Ph.D. degree in computer engineering from Sakarya University, Türkiye, in 2015. He is currently an Associate Professor with the Department of Computer Engineering, Bilecik Şeyh Edebali University. He joined as a Lecturer with Bilecik Şeyh Edebali University, in 2010, having previously taught at the Ministry of Education for three years. He also visited the Center for Secure Information Technologies (CSIT), Institute of Electronics, Communication and Information Technology (ECIT), Queen's University Belfast, U.K., to research Postdoctoral study as a Visiting Researcher, from 2016 to 2017. He has studied there in a project relating to fetal movement detection on US scans. His research and teaching interests include artificial intelligence, image processing, pattern recognition, biomedical, machine learning, computer-aided detection, medical informatics, data mining, and computer vision.

FURKAN ERTURK URFALI received the degree from the Faculty of Medicine, Eskişehir Osmangazi University, in 2007. He completed his Radiology Specialty Training with the Faculty of Medicine, Department of Radiology, Eskişehir Osmangazi University, in 2013. In 2022, he worked with the Department of Radiology, Kütahya University of Health Sciences, as an Associate Professor. He is currently with the Interventional Radiology Department, Bursa City Hospital.

MEHMET KORKMAZ received the European Diploma degree in radiology, and the Radiology degree from Istanbul OkmeydanıTraining and Research Hospital, in 2010. He completed further training in vascular, non-vascular, and neuro-interventional radiology with the Interventional Radiology Department, School of Medicine, Ege University, from 2013 to 2015. He was an Associate Professor, in 2018 and a Full Professor, in 2023. He is currently the Head of the Department of Radiology and Interventional Radiology, Kütahya University of Health Sciences. Additionally, he has a diploma and experience in advanced cardiac imaging.