LBCAM: A Channel Attention Embedded Sensor Fusion Architecture & Its Applications in Fetal Movement Monitoring

Praditha M. Alwis, *Member, IEEE*, Isuru P. Thilakasiri, *Member, IEEE*, Rahal T. Nanayakkara, *Member, IEEE*, Roshan I. Godaliyadda, *Senior Member, IEEE*, Mervyn Parakrama B. Ekanayake, *Senior Member, IEEE*, Chathura J. Rathnayake, Janaka V. Wijayakulasooriya *Member, IEEE*

***Abstract*— Fetal movement monitoring is critical to ensure the health of a fetus. A well-known approach to measuring fetal health is to keep track of fetal kicks on a regular basis. As a result, various devices and algorithms that can count the number of fetal kicks are being developed. In this article, we introduce a novel channel attention architecture that can learn pertinent information by observing the evolution of each channel with time. This novel channel attention-aided architecture is embedded in the introduced novel model to forecast the fetal kick count based on a set of accelerometric data collected from forty-four pregnant mothers. The dataset utilized for training was collected using our own multi-sensory device. With the more powerful channel attention and sensor fusion in the proposed architecture, we were able to outperform previous state-of-the-art models in terms of accuracy. Notably, the proposed channel attention architecture offers broad applicability in scenarios where channel prioritization based on temporal significance is required. Hence, this work represents a substantial improvement in the field of fetal monitoring and has the potential to be applied in contexts beyond fetal kick counting.**

***Index Terms*— channel attention, LSTM, fetal movement, multi-sensory data, sensor fusion**

I. Introduction

Monitoring fetal movements (FMs) is an important routine task during pregnancy, which contributes to monitor the fetal condition and thereby helps reduce the risk of adverse perinatal outcomes [1], [2]. Studies have been conducted to evaluate the correlation between reduced fetal movements (RFMs), stillbirths [1], preterm births [2], neonatal intensive care admissions, neonatal Seizures, neonatal Syndromes [3] as well as low birth weight [4], [5]. The presence of healthy FMs indicates a healthy fetus, while irregular FMs may indicate an unhealthy fetus or fetal distress [6]–[10].

Praditha M. Alwis, Isuru P. Thilakasiri, Rahal T. Nanayakkara, Roshan I. Godaliyadda, Mervyn Parakrama B. Ekanayake, Janaka V. Wijayakulasooriya with the Department of Electrical and Electronic Engineering, University of Peradeniya, Sri Lanka (e-mails: e16019@eng.pdn.ac.lk, e16367@eng.pdn.ac.lk, e16246@eng.pdn.ac.lk, roshang@eng.pdn.ac.lk, mpbe@eng.pdn.ac.lk, jan@eng.pdn.ac.lk respectively).

Chathura J. Rathnayake is with the Department of Obstetrics and Gynaecology, University of Peradeniya, Sri Lanka (email: chathura67@hotmail.com).

Monitoring of FMs on a regular basis helps to identify abnormalities or distresses in the fetus and take remedial action promptly. As such, pregnant mothers are advised to keep a log of the number of kicks by the fetus as a robust indicator of FMs, especially after the 28th week of pregnancy [11], [12]. This involves manually and consciously counting the number of kicks by the mother. Keeping track of the fetal kicks and counting them while engaging in day-to-day activities is a tedious task [13]. Moreover, the mother may not be able to feel or recognize fetal kicks during the first pregnancy. It is not abnormal not to feel the kicks until the 25th week of the pregnancy [14] or even later or not feel any such at all [15].

Therefore, automatic fetal movement counting systems are employed to better evaluate the condition of the fetus. These counting systems are employed [16]–[18] with the expectation of monitoring and raising alarm about possible absence or irregularities in the number of fetal movements so that they may be referred to a physician for a clinical evaluation.

Fetal movement monitoring can be conducted using many clinical techniques such as the use of Cardiotocography (CTG) [19], Magnetic Resonance Imaging (MRI) [20], [21], and Ultrasound Scanners (USS) [21]. CTG scanning provides a waveform of the contractions and can only capture the mother's contractions, while MRI and USS provide a visual image. These devices also produce electromagnetic waves and when used repeatedly, ultrasound scanners may have a variety of negative side effects [22]. Furthermore, all these being clinical diagnostic equipment, their usage is restricted to a clinical setting under the expertise of a professional practitioner. Therefore, while being highly reliable and accurate in the conclusions, access to them is often limited as the patient has to visit the location of the clinical practice in person; which may be highly restricted in some cases due to the irregular distribution of facilities and resources. Access to those facilities may also be limited due to financial reasons, socio-economic constraints, or even stigma and beliefs.

Under these considerations, systems that enable routine, or continuous, monitoring of FMs and/or kick counts under non-clinical settings are invaluable. Additionally, such a device being non-invasive (i.e., not emitting radiation, light, sound, vibration, or application of any other electromagnetic or mechanical energy form, or any chemical or biological

involvement), being wearable, and having a simple interface, makes it possible to make prolonged measurements of FMs safely.

Studies have been done to identify fetal movements using Inertial Measurement Units (IMUs) with the detection of fetal movement using both a single accelerometric sensor and with the use of multiple sensors mounted on the pregnant mother's abdomen [17], [23], [24]. The positioning, and number of sensors have also been evaluated in some studies [25] and variables such as different arrangements, force, and duration have also been investigated [26].

Several techniques have been evaluated to be used for the preprocessing, detection, and analysis of FMs using accelerometric data collected via IMUs. Kurtosis-based analysis, and nonnegative matrix factorization [17], independent component analysis (ICA), discrete wavelet transform (DWT), support vector machines (SVM), decision tree, random forest [24], bayesian optimization with hyper-parameter tuning [27], recurrent neural networks [28] have also been investigated in the recent years.

Furthermore, it is critical to distinguish between fetal movements and other maternal movements such as cough, positional movements, breathing, hiccups, abdominal gas, and bloating which could result in incorrect recognition of prenatal movements. Such movements must be identified and labeled algorithmically to improve the accuracy of the process of recognizing fetal movements [17], [29].

Although many works (e.g. [30]–[34]) have utilized some recent developments in the field of machine learning (ML), in particular the recent advances in deep learning (DL), it does not compare to the extent in which advancements in ML has been utilized in other fields such as computer vision. As an example, taking inspiration from the human visual system, attention mechanisms in DL have been used to obtain better performance in various applications such as computer vision. Soft attention, spatial attention, channel attention, and mixed attention mechanisms have been introduced in the past few years and they have been able to overcome some of the inherent limitations of ordinary DL network architectures [35]–[37]. Although most of these have been used in computer vision-related applications, only some attempts have been successful in utilizing these methods on applications stemming from biomedical signals [38]–[42].

Therefore, this shows a clear lack of penetration of DL-based AI in biomedical problems as a whole, including in the fetal monitoring problem [43]–[48]. Even when considering the works in biomedical signal processing that have utilized some DL architectures, it is evident that many works use generic ML algorithms without utilizing the contextual biomedical basis to modify the architectures to suit the precise nature of the problem [49], [50].

In contrast, the research significance of this paper is that we take inspiration from state-of-the-art DL models from computer vision and make fundamental contextual modifications to the AI architecture and the associated algorithm for our specific biomedical problem of FM detection via kick counts. Both the raw signal data as well as their spectrograms are utilized to capture more descriptive temporal and frequency characteristics.

As such, the main contribution of this paper is to propose a novel channel attention architecture that can learn pertinent information by utilizing the time-evolution of each channel in a manner specific to the biomedical problem of fetal movement detection and thereby predict fetal kick count with a better accuracy compared to the state-of-the-art. As exemplified and elaborated in the results and discussion sections, the following outcomes of the paper can be highlighted:

- Construct spectrograms as a descriptive feature set, which incorporates time and frequency attributes, allowing them to be used effectively by considering the temporal significance through deep learning
- Enabling the sensor fusion to be learnable by considering the entire multi-sensory network as a bundle of sensory inputs as opposed to the common practice of fusion using rudimentary operations such as concatenations, averaging, etc.
- Making the DL pipeline contextual for the application by incorporating a temporal Long Short-Term Memory (LSTM), by noting the fact that spectrograms are images and thereby drawing inspiration from an architecture, CBAM [37], originally designed for computer vision applications
- Utilizing a global type of learning and not a mere individual-specific learning which adds to the reusability of the algorithms introduced by the study.

The above-mentioned modifications to the CBAM architecture are proposed to reliably assess the use of channel attention in the biomedical signal processing context demonstrating promising prospects for these types of future applications. These modifications have been introduced in an explainable manner to improve the interpretability of the algorithm allowing for further improvement as a future aspect of the study.

II. Materials & Methods

A. Recording Device

The signal acquisition device used in this project was developed to be lightweight, and non-invasive. Considering these factors, a wearable device consisting of four MPU-6050 inertial measurement modules was designed that cost around $24. This method not only ensures the cost to be a minimum, but also allows pregnant mothers to monitor fetal count with the comfort of their own homes, minimizing the need for frequent hospital trips. Each MPU-6050 module contains a 3-axis accelerometer and 3-axis gyroscope, as well as an inbuilt analog-to-digital converter and a serial peripheral interface. The data stream from the four sensor modules is transferred to a central microcontroller via the serial peripheral interface for processing and storage. The sampling rate used by the microcontroller was 32 samples per second.

The first prototype of the recording device was in the form of a wearable belt, where the sensor modules were woven into the fabric of the belt. However, this gave rise to various practical complications, such as making the device difficult to sterilize. Thus, with the advice of medical health professionals, a second prototype device was constructed in which, the sensor

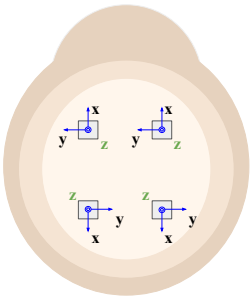

Fig. 1. Arrangement of accelerometers on the mother's abdomen. Arrows mark the orientation of the axes of the sensor

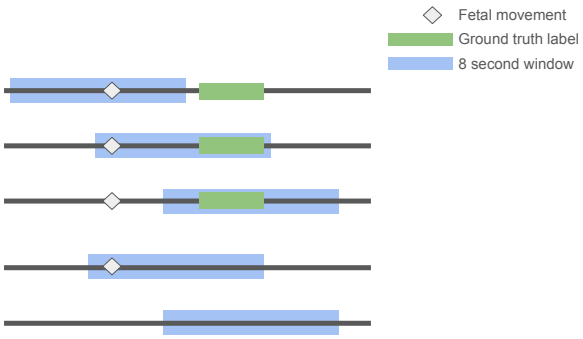

Fig. 2. Possible correct labeling/ mislabeling scenarios

modules were enclosed in plastic, and could be placed on the mother's abdomen and secured using medical tape.

In both devices, the sensor modules were placed on the mother's abdomen according to the configuration shown in Fig. 1. This configuration was selected as it covers all four quadrants of the mother's abdomen and enables the device to detect all fetal movement, even as the baby moves around inside the womb.

The device was constructed to have the capability of recording up to four types of FMs of choice (trunk movement, general body movements, isolated limb movements, movements in the 1st quadrant, etc.) and recording other movements that occur during the recording process such as cough, laugh, other positional movements. This wide flexibility of recording movement types was added to aid the data pre-processing steps.

B. Acquired Dataset

The dataset from the above-mentioned process, "A Multi-Sensory Inertial Measurement Unit Dataset for Fetal Condition Monitoring" [51] consists of seventy-one files including the data description. Beyond analysis of the dataset, we believe that the respiratory movements can also be captured from the IMU readings of the data and we consider that to be a future exploratory aspect of the dataset.

C. Signal Acquisition Process

During the data collection phase, recordings were collected from forty-four pregnant mothers, who volunteered for the study providing written and verbal consent. They were inpatients at the Professorial Unit of the Gynecology Ward, Teaching Hospital, Peradeniya, Sri Lanka. Two types of recording sessions were conducted to develop the dataset. The first type of readings were collected with the mother operating the device and pressing a button when a fetal kick occurs. Hence, the ground truth is the mother's perception of a fetal kick occurring, and its accuracy depends on the mother's sensitivity to such events and how such an event is perceived by the individual. The second type of reading was taken while a doctor was performing an ultrasound scan. With ultrasound measurements, the fetal movements can be distinguished precisely as they are clearly visible on the screen and the occurrence of such events can be accurately marked as ground truth. Hence, the findings made using the ultrasound observations serve as the ground truth of the second type of recording. The accuracy of the second type of recording is significantly higher than the first, as it does not depend on the mother's sensitivity to fetal kicks. In that sense, the first type of recording is a subjective measurement and the second type of recording is an objective measurement. In both cases, a typical recording session ran for approximately 20 minutes. More details about the recording sessions can be found by accessing the dataset.

Ethical Clearance: This study was approved by the Ethics Review Committee, Faculty of Medicine, University of Peradeniya, Sri Lanka. Approval was granted to conduct research project No. 2018/EC/43 entitled "Fetal movement analysis for condition monitoring" at the Teaching Hospital, Peradeniya, Sri Lanka.

D. Signal Preprocessing

A certain delay is inherently present between the actual occurrence of the fetal movement and the ground truth label due to human errors in the recording process. This delay is smaller for USS recordings than for mothers' perception ground truth recordings. As a preprocessing step, a separate algorithm was used to deal with this delay and the corner cases that come with it. In this step, each accelerometer recording of the dataset was broken down into overlapping windows with a length of eight seconds and a stride of one second. This delay was observed to be approximately one second between the actual occurrence of a fetal kick and the subsequent pressing of the ground truth button. These scenarios are shown in Fig. 2.

Windows that had no fetal kick ground truth within its entire 8-second duration, as well as the 2 seconds following it, were marked as windows with no kicks. The additional 2 seconds were considered to account for the 1 second delay with a tolerance. Windows which had a marked ground truth within its first 2 seconds, and within the 2 seconds after the window were discarded due to their ambiguity. Finally, windows that had a positive ground truth label in their latter 6 seconds were marked as windows that contained fetal movements. The distribution of kicks in the selected windows was non-uniform as many of the windows did not contain any fetal movements. To mitigate this, a subset of windows was randomly sampled from the set of all windows that had no fetal movements to

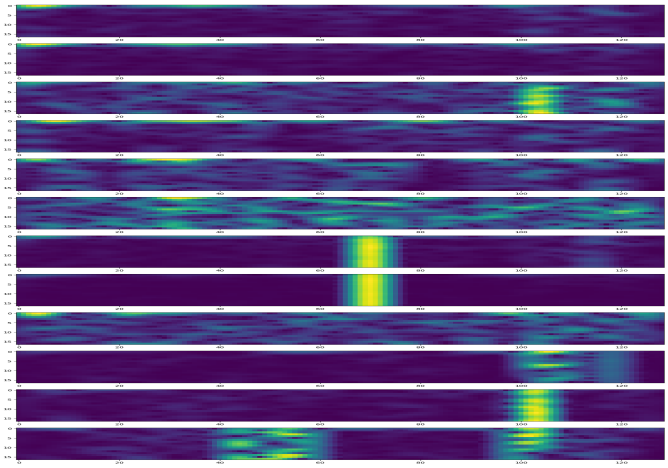

Fig. 3. Twelve spectrograms of a sample kick window

make the processed dataset more uniform, as it was required to create robustly trained networks to better identify fetal movements.

1) Feature Analysis: A sliding Hanning window of 16 samples with a stride of 1 was used to calculate the STFT and thus construct the magnitude spectrogram of each accelerometer channel of the selected windows as shown in Fig. 3. The magnitude spectrogram was utilized in this case since it is one of the most expressive signal representation methods representing the intensity plot of frequencies of a signal that fluctuates with time [52]. Since narrow windows have a short time duration but a wide bandwidth, a narrow window length of 8 samples will have a good time resolution but a coarse frequency resolution. On the other hand, a broad window length of 32 samples will result in a fine frequency resolution but a coarse time resolution. In order to account for this window effect, a window length of 16 samples was used. These spectrograms were used as the inputs of deep neural networks when attempting to identify the occurrence of a fetal kick.

E. Deep Learning Architecture

It is a widely accepted fact that human perception depends significantly on attention. Human vision does not process an entire scene at once and instead, to better understand visual organization, a series of sporadic glimpses is used to choose key segments. To capture this crucial aspect of human vision, the Convolutional Block Attention Map (CBAM) [37] has been developed. CBAM sequentially infers attention maps along the channel and spatial dimensions from an intermediate feature map. The attention maps are then multiplied by the input feature map for adaptive feature refinement. The main flaw in this approach is that the time series data contained in a signal fails to be handled by this module. To account for this, we have produced a modified version with RNNs that is more adept at capturing time series data.

1) LSTM Based Channel Attention Map (LBCAM): It is evident that while assigning priority to each channel, the previous architecture has failed to give essential consideration to time series data. In CBAM, each channel is assigned a weight mostly based on the channel's maximum pool and average pool values. Here, two spatial information descriptors are used. These descriptors: F^c_{avg} and F^c_{max} which denote average-pooled features and max-pooled features respectively, are forwarded into a shared network of multi-layer perceptron (MLP) with a single hidden layer.

In a nutshell, channel attention is calculated as,

$$M_c = \sigma(W_1 \times ReLU(W_0 \times F^c_{avg}) + W_1 \times ReLU(W_0 \times F^c_{max}))$$

where,
$ReLU$ denotes the ReLU function
σ denotes the sigmoid function
$W_0 \in \mathbb{R}^{C/r \times C}$ and $W_1 \in \mathbb{R}^{C \times C/r}$ denote the MLP weights.

In our novel LSTM Based Channel Attention Map (LBCAM), the F^c_{max} and F^c_{avg} have been replaced by quantities that can capture time series data. The architecture of the LSTM-based channel attention map (LBCAM) is shown in Fig. 4. Here, the input data shape is $C \times N \times M$, where C is the number of channels, N is the number of time samples, and M is the number of frequency samples. To capture the time series significance of each channel, C channels are split and routed through parallel LSTM layers. The output of the k^{th} LSTM channel is calculated as,

$$f_{t,k} = \sigma(W_{f,k} \times (h_{t-1,k} \oplus x_{t,k}) + b_{f,k})$$

$$i_{t,k} = \sigma(W_{i,k} \times (h_{t-1,k} \oplus x_{t,k}) + b_{i,k})$$

$$\tilde{C}_{t,k} = \tanh(W_{C,k} \times (h_{t-1,k} \oplus x_{t,k}) + b_{C,k})$$

$$C_{t,k} = f_{t,k} \times C_{t-1,k} + i_{t,k} \times \tilde{C}_{t,k}$$

$$h_{t,k} = \sigma(W_{o,k} \times (h_{t-1,k} \oplus x_{t,k}) + b_{o,k}) \times \tanh(C_{t,k})$$

where,
x is the input spectrogram
i is the output of the input gate layer of an LSTM cell
h is the output of the hidden cell of an LSTM layer
W and b denote the weights and the bias
t denotes the t^{th} time sample
k denotes the k^{th} channel
$\tanh$ denotes the tanh function
σ denotes the sigmoid function
$\oplus$ denotes the concatenation operation

Then, the last hidden cell output (h_N) of each LSTM channel is sent through separate dense layers to obtain a single value for each channel. This architecture tries to carry out learned long-term dependencies through this value. In CBAM, this value is obtained using max pooling or average pooling which doesn't carry much detail about the time series progression.

The output of the linear layers is sent through a shared MLP layer to learn attention weights considering cross-channel long-term dependencies. Finally, the output of the shared MLP

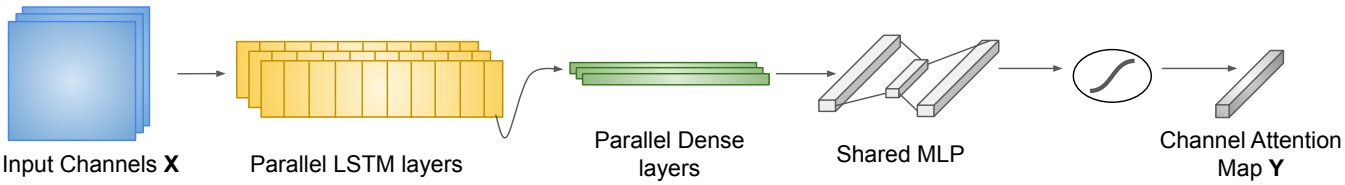

Fig. 4. Architecture of the LSTM Based Channel Attention Map (LBCAM)

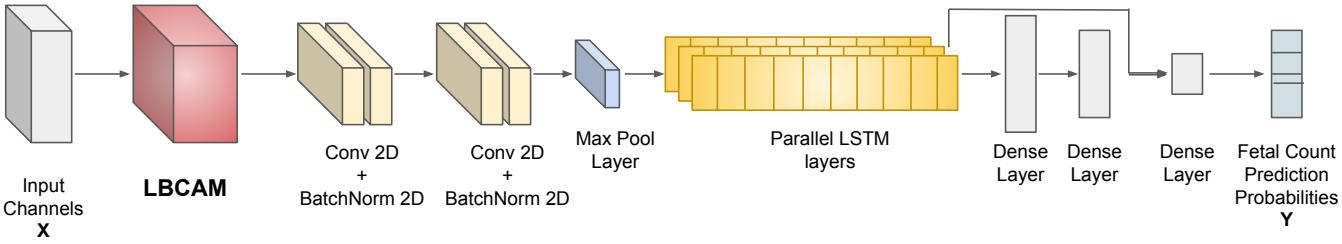

Fig. 5. Architecture of the classification model

is sent through a sigmoid layer to obtain the final channel attention map.

$$Y_k = \sigma(M_k) = \frac{1}{1 + e^{-M_k}}$$

where,
Y_k denotes the channel attention map
σ denotes the sigmoid function
M_k denotes the output of the shared MLP corresponding to the k^{th} channel

The goal of this architecture is to facilitate learning meaningful data from each spectrogram by looking at the progressions along the time axis.

2) Classification Model: A novel classification model was used to predict the number of fetal kicks within a given window. The classification model was trained to predict four separate classes, including windows with no kicks, single kicks, two kicks, and more than two kicks.

The spectrograms that were weighted by the LBCAM module were fed into this novel classification model. The initial breakthrough was weighting each channel based on the signal's evolution along the time axis. The following objective was to predict the number of fetal kicks in each window. The classification model employs a sensor fusion technique to boost the productivity of the learning process. This model primarily attempts to acquire information from each spectrogram collectively.

First, the generated spectrograms are sent through the LBCAM to weigh the channels according to their spectral significance as shown in Fig. 5. The set of weighted spectrogram is sent through two sets of channel-wise 2-dimensional convolutional layers followed by a 2-dimensional batch normalization layer. Our LBCAM is tuned such that MLP tries to squeeze the number of channels from 12 to 6 and excite to 12 again. Hence, it can be interpreted as the LBCAM block trying to weigh the 6 most significant channels. Hence, the two 2-dimensional convolutional layers in the classification model are implemented such that the number of channels in the output is reduced to 6. In other words, this approach attempts to exclude six spectrograms that have the least spectral significance. This output is then passed via a max pool layer in the same order. It is envisioned that the model will try to decrease the spatial complexity of the spectrograms by selecting only high-magnitude spatial data. Here, the max pooling is tuned in such a way that each vector will be reduced to a vector of $T \times 1$. Hence, the output shape will be $N \times 6 \times T \times 1$ where N and T represent the batch size and the number of samples along the time axis. In other words, the 2-dimensional spectrograms are reduced to 1-dimensional time signals considering their long-term dependencies and spectral significance. After that, the channels are split and routed through parallel LSTM layers. The outputs of each LSTM layer are concatenated together and sent through 2 dense layers activated by the ReLU function. After that, the current output and the outputs of the last hidden cell of previous LSTM layers are concatenated and sent through another dense layer activated by a softmax function.

$$Z = W^T(D \oplus h_N) + b$$

$$Y_i = S(Z_i) = \frac{e^{Z_i}}{\sum_{j=1}^{n} e^{Z_j}}$$

where,
W and b denote the weights and bias of the second dense layer
D denotes the output of the second dense layer
h_N denotes the last hidden layer output of the parallel LSTM layers
$\oplus$ denotes the concatenation operation
Z denotes the output of the final dense layer
S denotes the softmax function
Y denotes the fetal kick count prediction probability vector

The fundamental idea behind the three dense layers in this case is to perform sensor fusion to get the most out of the input channels. The last hidden cell outputs of the parallel LSTM

TABLE I
VALIDATION ACCURACY, SENSITIVITY AND SPECIFICITY OF THE MODELS TRAINED ON THE DATASET WITH MOTHER'S PERCEPTION AS THE GROUND TRUTH

Model	Validation Accuracy	Sensitivity	Specificity
Model	50%	0.66	0.54
Squeeze and Excitation + model	48%	0.54	0.61
CBAM + model	53%	0.58	0.64
LBCAM + Novel model	**69%**	**0.83**	0.81
xresnet1D	62%	0.56	**0.87**
inception1D	59%	0.6	0.76
STS	54%	0.76	0.57

TABLE II
VALIDATION ACCURACY, SENSITIVITY AND SPECIFICITY OF THE MODELS TRAINED ON THE DATASET WITH ULTRASOUND SCAN AS THE GROUND TRUTH

Model	Validation Accuracy	Sensitivity	Specificity
Model	55%	0.83	0.55
Squeeze and Excitation + model	51%	0.65	0.51
CBAM + model	58%	0.7	0.59
LBCAM + Novel model	**80%**	**0.91**	**0.8**
xresnet1D	69%	0.48	0.71
inception1D	61%	0.43	0.63
STS	58%	0.78	0.58

layers are passed to the final dense layer in order to give more weight to long-term dependencies. In other words, this sensor fusion algorithm learns the fusion while attempting to learn more from the accumulative time series information grasped at the last hidden cell of the parallel LSTM layers.

III. EXPERIMENTAL TESTS AND RESULTS

As discussed in section II-C, our dataset contains two sets of readings. The first dataset was created using the mother's perception as the ground truth, and the second was created using an ultrasound scan as the ground truth. Our model was trained and validated on both the dataset separately. Moreover, the model was cross-validated between the two datasets.

In order to verify the performance of the proposed novel architecture, results were compared with several state-of-the-art architectures in the field of fetal monitoring and in other biomedical applications such as xresnet1D [53], inception1D [53], SleepPoseNet (STS) [54], Squeeze and Excitation [36], Convolutional Block Attention Module (CBAM) [37].

Each of these state-of-the-art architectures was modified in such a way that they can be used to predict the fetal kick count. Furthermore, architectures such as xresnet1D and inception1D, which were previously utilized for different ECG signal predictions, were modified to predict fetal kick count using the accelerometric data recorded by our signal acquisition device. In this case, the same signal windowing functions were utilized. However, since xresnet1D and inception1D only forecast on 1-dimensional data, windowed signals were not transformed into spectrograms.

For the performance analysis, prediction accuracy, sensitivity and specificity were chosen as the performance metrics.

TABLE III
CROSS-VALIDATION ACCURACIES

Model	Validation Accuracy	
	Trained on mother's perception dataset and validated on ultrasound dataset	Trained on ultrasound dataset and validated on mother's perception dataset
Novel model	43%	49%
Squeeze and Excitation + model	39%	37%
CBAM + model	45%	49%
LBCAM + model	**58%**	**64%**
xresnet1D	52%	58%
inception1D	48%	52%
STS	47%	51%

Since neural network-based predictions are slightly randomized, each model was trained and tested four times, and average values were obtained. Prediction accuracy was obtained considering the four classes as mentioned in II-E.2. When calculating sensitivity and specificity, predicting the presence of a fetal kick was considered a positive result, whereas predicting no fetal kick was considered a negative result. Sensitivity and specificity were calculated as follows.

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives}$$

Each of these models was trained and validated on the two datasets. Table I and II show the validation accuracy, sensitivity and specificity of our model and the other architectures.

While high sensitivity and specificity values are crucial, particular emphasis is placed on specificity. In the context of fetal movement prediction, lower sensitivity implies that the model may predict a reduced fetal movement, prompting the mother to consult with a doctor to address any concerns about the fetus's health. On the other hand, lower specificity, indicating the model predicts heightened fetal movement when the actual movement is low, could potentially lead to a mother choosing to stay at home, posing a risk of overlooking critical situations.

From Table II, it is evident that the LBCAM + model surpasses other models across all performance metrics. As shown in Table I, the xresnet1D exhibits the highest specificity at the expense of lower sensitivity, indicating a bias towards predicting 0. While the increased specificity suggests a reduced likelihood of false positives, the trade-off is an increased rate of false negatives, necessitating more frequent consultations with a doctor for the mother. Notably, Table I illustrates that the specificity of the LBCAM + model is at an acceptable level. Considering this, along with specificity and other metrics, it becomes apparent that the LBCAM + model outperforms other models.

For a comprehensive evaluation of the models, cross-validation was performed on the datasets. Table III shows the validation accuracies of cross-validation results.

It is evident that our novel LBCAM combined with our novel prediction model has outperformed other models in accuracy metrics. However, xresnet1D has come closer to the accuracy level of the LBCAM + model.

Since xresnet1D is a deep neural network, it has a larger parameter space relative to our data space. As a result, it has a high tendency to overfit. Table I and II, show that while xresnet1D has good validation accuracy on the same dataset, it has performed badly in cross-validation. In simple terms, the xresnet1D model has overfitted the train set to a significant extent.

Moreover, a further study was carried out to observe this tendency to overfit. Here, a minor change was introduced to the data set. During the pre-processing, windows were split from the raw signal with a stride of 1 second. Hence, there are partially overlapping windows. Precautions were taken to limit these sets of overlapping windows to the train set or the test set. Otherwise, the model may overfit to certain data windows in the train set and accurately predict the fetal count in an overlapping window that is present in the test set. Here, this proactive approach was removed, and all the windows were mixed together and randomly selected for the train and test datasets. As a result, the train and test datasets now have a significant quantity of data with partially overlapping windows.

TABLE IV
VALIDATION ACCURACIES WITH THE NEW DATASETS

Model	**Validation Accuracy**	
	Trained and validated on modified mother's perception dataset	Trained and validated on modified ultrasound dataset
LBCAM + model	77%	85%
xresnet1D	**91%**	**94%**

Our novel LBCAM + model and the xresnet1D model were trained again using this modified dataset. The validation accuracies are shown in table IV. It can be observed that the xresnet1D model has a higher tendency to overfit when the results from tables I, II and IV are compared. The xresnet1D tends to learn and overfit to an overlapping part of two time-domain windows. On the other hand, LBCAM tries to learn temporal progressions in a spectrogram. Although there can be overlapping windows in the time domain, the two corresponding spectrograms may have different temporal progressions. Hence, it is very unlikely for our model to overfit to a certain overlapping part of two windows. Considering these aspects, our novel LBCAM and model will outperform xresnet1D in this biomedical application.

IV. DISCUSSION

A. Improvement of channel attention with LSTM addition

The performance metrics in Table I and II show that the LBCAM + novel model has surpassed all architectures. The LBCAM module has played an important part in this. Table II shows that adding a CBAM block to the novel model has only increased the validation accuracy from 55% to 58%. Nonetheless, the addition of an LBCAM block has boosted the validation accuracy from 55% to 80%. However, this improvement is accompanied by a reduction in sensitivity from 0.83 to 0.7. In essence, the CBAM model tends to predict a lack of fetal movement more frequently. In CBAM, mainly max-pool and avg-pool layers are used to grasp significant information from a specific channel. While it is a common procedure in deep learning to use max pooling and average pooling, however, when those are used on a specific data stream like spectrograms, it may not have any physical meaning. Therefore, the usage of an LSTM layer that encapsulates the underline structure will become more feasible. Hence, our novel LBCAM architecture could improve the channel attention significantly.

In CBAM architecture, there is a spatial attention block to learn where to focus. However, if we consider this spectrogram to be a picture, we anticipate x and y spatial coordinates to be equivalent to t and f; time-frequency coordinates that are not necessarily equivalent in a physical sense. Therefore, while we can be inspired by an image-like measure, direct plug-and-play may not be possible here. Our main target was to learn from this spectrogram in an orderly fashion because this has a certain temporal structure. These temporal structures can be learned by the LSTM layers in LBCAM. As a result, rather than introducing a max pooling or average pooling layer or a spatial attention block, the addition of an LSTM layer has improved the strength of the channel attention for this application.

To examine the influence of LBCAM on the training process, cross-entropy curves were plotted. Fig. 6 depicts the training cross-entropy loss curves.

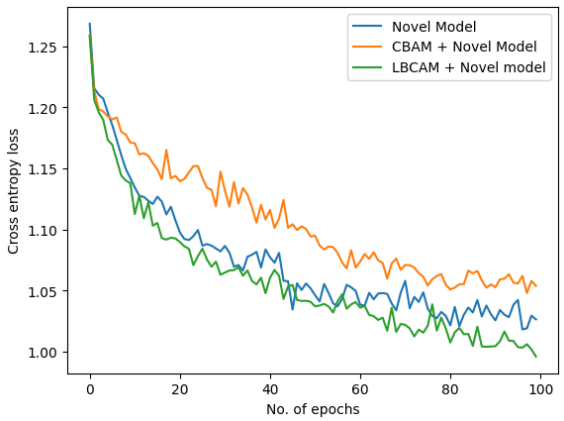

Fig. 6. Cross-entropy loss curves of novel model, novel model with CBAM, novel model with LBCAM

As shown in Fig. 6, the addition of a CBAM layer has greatly diminished the training rate of our model. This phenomenon could be attributed to the CBAM's attempt to learn more data by treating the spectrogram as an image. Given that the fetal kick count is predominantly represented in the temporal structure of the spectrogram, providing channel attention becomes challenging when treating the spectrogram solely as an image. On the other hand, the addition of LBCAM has enhanced the training rate. This could be linked to LBCAM's capacity to prioritize each channel based on its temporal

significance. As it turns out, it is clear that the addition of an LSTM layer has substantially enhanced the ease of training as well as the performance metrics.

B. Recordings with inherent noise

Noisy recordings in our dataset may pose significant barriers and may have a significant effect on accuracy. They can be electrical interference, equipment limitations, and sensor inaccuracies in the device. Furthermore, noise may be created due to the imprecise nature of human perception in situations when mothers operate the device and press a button when they detect a fetal kick.

The primary issue with the dataset obtained from the mother's perspective as the ground truth is that it may contain a significant amount of human error. As a result, training on this dataset will be very challenging. The model may attempt to learn to forecast errors. On the other hand, the dataset recorded using an ultrasound scan as the ground truth may contain only a handful of human errors. Errors such as electrical interference, equipment limitations, sensor inaccuracies and so on may account for miniature errors in the dataset.

The effect of noise on the dataset is readily apparent in the results of Tables I and II. Most of the performance metrics shown in Table I are lower than those shown in Table II.

To summarise, coping with noisy recordings in fetal movement monitoring is an important part of this field's research. The quality and dependability of data can be improved by being aware of the sources of noise, understanding its potential effects, and utilizing a combination of technological and methodological solutions. This opens up new avenues for further exploration and research in the field.

V. Conclusion

In this paper, we have proposed a novel channel attention architecture as well as a model to predict fetal kick count. Our new channel attention architecture learns by observing the evolution of each channel with time. The two datasets used for the training process were gathered using our own multi-sensory device. Our architecture could predict fetal kick count with high accuracy because of the more powerful channel attention in our innovative LBCAM architecture and sensor fusion in our novel model.

The performance of the proposed method was verified by comparing prediction results with several state-of-the-art architectures in the field of fetal monitoring and in other biomedical applications. The results demonstrated that the proposed model outperforms all state-of-the-art models. Further studies suggested that the likelihood of xresnet1D for overfitting will reduce the significance of xresnet1D for this application.

In future work, we are planning on recording more fetal movement data with high accuracy and less noise. There is tremendous potential for the usage of AI and deep learning architectures specifically in biomedicine. Our work and other related works highlight the importance of contextual knowledge of biomedicine and applying appropriate architectures, algorithms, functions, etc. for the given application. The whole rationale is that when AI is used in biomedicine, there is less potential for success when other algorithms are simply reused. The AI model should have explainable steps. Moreover, those explainable steps should be aligned with physical reality.

While this solves certain problems and introduces a new way of thinking, what we have really done is pioneer in a certain direction. So, as the future work, there is a lot more work to be done for the future researchers if they want to go down this lane. Compared to much of the work that is being used in biomedicine, researchers may tend to perform transfer learning. Transfer learning is not AI design. There is a significant difference between transfer learning vs a dedicated or a specific AI that connects us to a specific problem. We are not going for a one-fits-all architecture. We want to promote this in another way where contextual bio-medical knowledge, contextual medical realities, and practical realities coupled with the inherent meaning of each of the components in the neural network, all matter.

References

[1] J. E. Lawn, H. Blencowe, P. Waiswa, A. Amouzou, C. Mathers, D. Hogan, V. Flenady, J. F. Frøen, Z. U. Qureshi, C. Calderwood *et al.*, "Stillbirths: rates, risk factors, and acceleration towards 2030," *The Lancet*, vol. 387, no. 10018, pp. 587–603, 2016.

[2] J. E. Norman, A. E. Heazell, A. Rodriguez, C. J. Weir, S. J. Stock, C. J. Calderwood, S. C. Burley, J. F. Frøen, M. Geary, F. Breathnach *et al.*, "Awareness of fetal movements and care package to reduce fetal mortality (affirm): a stepped wedge, cluster-randomised trial," *The Lancet*, vol. 392, no. 10158, pp. 1629–1638, 2018.

[3] E. M. Mizrahi, P. Plouin, and R. Clancy, "Neonatal seizures," *Pediatric epilepsy: Diagnosis and treatment (third edition) pp*, pp. 229–40, 2008.

[4] V. Smith, K. Muldoon, V. Brady, and H. Delaney, "Assessing fetal movements in pregnancy: A qualitative evidence synthesis of women's views, perspectives and experiences," *BMC pregnancy and childbirth*, vol. 21, no. 1, pp. 1–14, 2021.

[5] S. Efkarpidis, E. Alexopoulos, L. Kean, D. Liu, and T. Fay, "Case-control study of factors associated with intrauterine fetal deaths," *Medscape General Medicine*, vol. 6, no. 2, 2004.

[6] V. Flenady, P. Middleton, G. C. Smith, W. Duke, J. J. Erwich, T. Y. Khong, J. Neilson, M. Ezzati, L. Koopmans, D. Ellwood *et al.*, "Stillbirths: the way forward in high-income countries," *The Lancet*, vol. 377, no. 9778, pp. 1703–1717, 2011.

[7] B. F. Bradford, R. S. Cronin, L. M. McCowan, C. J. McKinlay, E. A. Mitchell, and J. M. Thompson, "Association between maternally perceived quality and pattern of fetal movements and late stillbirth," *Scientific reports*, vol. 9, no. 1, p. 9815, 2019.

[8] M. Levy, M. Kovo, G. Barda, O. Gluck, L. Koren, J. Bar, and E. Weiner, "Reduced fetal movements at term, low-risk pregnancies: is it associated with adverse pregnancy outcomes? ten years of experience from a single tertiary center," *Archives of gynecology and obstetrics*, vol. 301, pp. 987–993, 2020.

[9] A. Aiob, R. Toma, M. Wolf, Y. Haddad, and M. Odeh, "Cerebroplacental ratio and neonatal outcome in low-risk pregnancies with reduced fetal movement: A prospective study," *European Journal of Obstetrics & Gynecology and Reproductive Biology: X*, vol. 14, p. 100146, 2022.

[10] M.-C. Malm, H. Lindgren, and I. Rådestad, "Losing contact with one's unborn baby—mothers' experiences prior to receiving news that their baby has died in utero," *OMEGA-Journal of Death and Dying*, vol. 62, no. 4, pp. 353–367, 2011.

[11] J. Warland, L. M. O'Brien, A. E. Heazell, E. A. Mitchell, and S. consortium, "An international internet survey of the experiences of 1,714 mothers with a late stillbirth: the stars cohort study," *BMC pregnancy and childbirth*, vol. 15, pp. 1–11, 2015.

[12] B. A. Winje, J. Røislien, and J. F. Frøen, "Temporal patterns in count-to-tenfetal movement charts and their associations with pregnancy characteristics: a prospective cohort study," *BMC Pregnancy and Childbirth*, vol. 12, no. 1, pp. 1–12, 2012.

[13] Z. R. Hijazi, S. E. Callan, and C. E. East, "Maternal perception of foetal movement compared with movement detected by real-time ultrasound: An exploratory study," *Australian and New Zealand Journal of Obstetrics and Gynaecology*, vol. 50, no. 2, pp. 144–147, 2010.

[14] M. Sheikh, S. Hantoushzadeh, and M. Shariat, "Maternal perception of decreased fetal movements from maternal and fetal perspectives, a cohort study," *BMC pregnancy and childbirth*, vol. 14, no. 1, pp. 1–7, 2014.

[15] B. Bradford and R. Maude, "Maternal perception of fetal movements in the third trimester: A qualitative description," *Women and Birth*, vol. 31, no. 5, pp. e287–e293, 2018.

[16] A. Alim and M. H. Imtiaz, "Wearable sensors for the monitoring of maternal health—a systematic review," *Sensors*, vol. 23, no. 5, p. 2411, 2023.

[17] U. Delay, T. Nawarathne, S. Dissanayake, S. Gunarathne, T. Withanage, R. Godaliyadda, C. Rathnayake, P. Ekanayake, and J. Wijayakulasooriya, "Novel non-invasive in-house fabricated wearable system with a hybrid algorithm for fetal movement recognition," *Plos one*, vol. 16, no. 7, p. e0254560, 2021.

[18] E. Somathilake, U. H. Delay, J. B. Senanayaka, S. L. Gunarathne, R. I. Godaliyadda, M. P. Ekanayake, J. Wijayakulasooriya, and C. Rathnayake, "Assessment of fetal and maternal well-being during pregnancy using passive wearable inertial sensor," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[19] N. Daly, D. Brennan, M. Foley, and C. O'Herlihy, "Cardiotocography as a predictor of fetal outcome in women presenting with reduced fetal movement," *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 159, no. 1, pp. 57–61, 2011.

[20] A. Singh, S. S. M. Salehi, and A. Gholipour, "Deep predictive motion tracking in magnetic resonance imaging: application to fetal imaging," *IEEE transactions on medical imaging*, vol. 39, no. 11, pp. 3523–3534, 2020.

[21] J. Lai, N. C. Nowlan, R. Vaidyanathan, C. J. Shaw, and C. C. Lees, "Fetal movements as a predictor of health," *Acta obstetricia et gynecologica Scandinavica*, vol. 95, no. 9, pp. 968–975, 2016.

[22] J. S. Abramowicz, "Benefits and risks of ultrasound in pregnancy," in *Seminars in perinatology*, vol. 37, no. 5. Elsevier, 2013, pp. 295–300.

[23] X. Zhao, X. Zeng, L. Koehl, G. Tartare, and J. De Jonckheere, "A wearable system for in-home and long-term assessment of fetal movement," *IRBM*, vol. 41, no. 4, pp. 205–211, 2020.

[24] M. Mesbah, M. S. Khlif, S. Layeghy, C. E. East, S. Dong, A. Brodtmann, P. B. Colditz, and B. Boashash, "Automatic fetal movement recognition from multi-channel accelerometry data," *Computer methods and programs in biomedicine*, vol. 210, p. 106377, 2021.

[25] M. Altini, P. Mullan, M. Rooijakkers, S. Gradl, J. Penders, N. Geusens, L. Grieten, and B. Eskofier, "Detection of fetal kicks using body-worn accelerometers during pregnancy: Trade-offs between sensors number and positioning," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 5319–5322.

[26] Y.-C. Du, L. B. Yen, P.-L. Kuo, and P.-Y. Tsai, "A wearable device for evaluation of relative position, force, and duration of fetal movement for pregnant woman care," *IEEE Sensors Journal*, vol. 21, no. 17, pp. 19 341–19 350, 2021.

[27] S. Liang, J. Peng, Y. Xu, H. Ye *et al.*, "Passive fetal movement recognition approaches using hyperparameter tuned lightgbm model and bayesian optimization," *Computational intelligence and neuroscience*, vol. 2021, 2021.

[28] L. I. Thilakasiri, D. Alwis, R. Nanayakkara, G. Godaliyadda, M. Ekanayake, J. Wijayakulasooriya, and R. Rathnayake, "Fetal movement identification using spectrograms with attention aided models and identifying a set of correlating parameters with gestational age," in *2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS)*. IEEE, 2023, pp. 227–232.

[29] M. S. Khlif, B. Boashash, S. Layeghy, T. Ben-Jabeur, M. Mesbah, C. East, and P. Colditz, "Time-frequency characterization of tri-axial accelerometer data for fetal movement detection," in *2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE, 2011, pp. 466–471.

[30] C. Cao, F. Liu, H. Tan, D. Song, W. Shu, W. Li, Y. Zhou, X. Bo, and Z. Xie, "Deep learning and its applications in biomedicine," *Genomics, proteomics & bioinformatics*, vol. 16, no. 1, pp. 17–32, 2018.

[31] P. Baldi, "Deep learning in biomedical data science," *Annual review of biomedical data science*, vol. 1, pp. 181–205, 2018.

[32] P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov, "Applications of deep learning in biomedicine," *Molecular pharmaceutics*, vol. 13, no. 5, pp. 1445–1454, 2016.

[33] I. R. I. Haque and J. Neubert, "Deep learning approaches to biomedical image segmentation," *Informatics in Medicine Unlocked*, vol. 18, p. 100297, 2020.

[34] M. Wainberg, D. Merico, A. Delong, and B. J. Frey, "Deep learning in biomedicine," *Nature biotechnology*, vol. 36, no. 9, pp. 829–838, 2018.

[35] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," *Advances in neural information processing systems*, vol. 28, 2015.

[36] H. Jie, S. Li, S. Gang, and S. Albanie, "Squeeze-and-excitation networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 8, pp. 2011–2023, 2017.

[37] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[38] M. Liu, Y. Lu, S. Long, J. Bai, and W. Lian, "An attention-based cnn-bilstm hybrid neural network enhanced with features of discrete wavelet transformation for fetal acidosis classification," *Expert Systems with Applications*, vol. 186, p. 115714, 2021.

[39] M. R. Mohebbian, S. S. Vedaei, K. A. Wahid, A. Dinh, H. R. Marateb, and K. Tavakolian, "Fetal ecg extraction from maternal ecg using attention-based cyclegan," *IEEE journal of biomedical and health informatics*, vol. 26, no. 2, pp. 515–526, 2021.

[40] L. Li, J. Wan, J. Zheng, and J. Wang, "Biomedical event extraction based on gru integrating attention mechanism," *Bmc Bioinformatics*, vol. 19, no. 9, pp. 93–100, 2018.

[41] X. He, P. Tai, H. Lu, X. Huang, and Y. Ren, "A biomedical event extraction method based on fine-grained and attention mechanism," *BMC bioinformatics*, vol. 23, no. 1, pp. 1–17, 2022.

[42] X. He, L. Li, J. Wan, D. Song, J. Meng, and Z. Wang, "Biomedical event trigger detection based on bilstm integrating attention mechanism and sentence vector," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 651–654.

[43] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis *et al.*, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, 2018.

[44] J. Chai, H. Zeng, A. Li, and E. W. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Machine Learning with Applications*, vol. 6, p. 100134, 2021.

[45] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, and R. Socher, "Deep learning-enabled medical computer vision," *NPJ digital medicine*, vol. 4, no. 1, p. 5, 2021.

[46] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[47] C. H. Chen, *Handbook of pattern recognition and computer vision*. World scientific, 2015.

[48] T. Saba, "Computer vision for microscopic skin cancer diagnosis using handcrafted and non-handcrafted features," *Microscopy Research and Technique*, vol. 84, no. 6, pp. 1272–1283, 2021.

[49] M. Piriyajitakonkij, P. Warin, P. Lakhan, P. Leelaarporn, N. Kumchaiseemak, S. Suwajanakorn, T. Pianpanit, N. Niparnan, S. C. Mukhopadhyay, and T. Wilaiprasitporn, "Sleepposenet: Multi-view learning for sleep postural transition recognition using uwb," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 1305–1314, 2020.

[50] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ecg analysis: Benchmarks and insights from ptb-xl," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1519–1528, 2020.

[51] R. Nanayakkara, I. Thilakasiri, P. Alwis, R. Godaliyadda, P. Ekanayake, J. Wijayakulasooriya, and C. Ratnayake, "A multi-sensory inertial measurement unit dataset for fetal condition monitoring," 2023. [Online]. Available: https://doi.org/10.7910/DVN/QHFHYC

[52] C. Baumgartner, K. J. Blinowska, A. Cichocki, H. Dickhaus, P. J. Durka, P. V. E. McClintock, G. Pfurtscheller, A. Stefanovska, and S. Tong, "Discussion of "time-frequency techniques in biomedical signal analysis: A tutorial review of similarities and differences"," *Methods Inf. Med.*, vol. 52, no. 04, pp. 297–307, 2013.

[53] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ECG analysis: Benchmarks and insights from PTB-XL," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 5, pp. 1519–1528, May 2021.

[54] M. Piriyajitakonkij, P. Warin, P. Lakhan, P. Leelaarporn, N. Kumchaiseemak, S. Suwajanakorn, T. Pianpanit, N. Niparnan, S. C. Mukhopadhyay, and T. Wilaiprasitporn, "SleepPoseNet: Multi-View learning for sleep postural transition recognition using UWB," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 4, pp. 1305–1314, Apr. 2021.