International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(13s) 2026 pp. 640-654
https://doi.org/10.70917/ijcisim-2026-4102
Article

A Novel Approach for Multi-Class Fetal ECG Arrhythmia Classification by ResNet-50 Combined with Transformer

Yogesh Sharma[1,*], Harsh Dev[2], Anurag Tiwari[3]

[1,2]BBD University, Lucknow, India, 227105
[3]BBDITM, Lucknow, India, 227105
Email: yogeshbbdit@gmail.com, drharshdev@bbdu.ac.in, anuragrktiwari@bbdnitm.ac.in
***Corresponding Email:** yogeshbbdit@gmail.com

Abstract: Accurate classification of fetal arrhythmias from non-invasive fetal electrocardiogram (NI-fECG) signals is essential for the early diagnosis and prenatal management of congenital cardiac abnormalities. However, reliable automated fetal arrhythmia classification remains challenging due to the low signal-to-noise ratio of fetal ECGs, maternal ECG interference, class imbalance, and the limited availability of annotated fetal ECG datasets. To address these challenges, this study proposes a patient-wise deep transfer learning framework for multiclass fetal ECG arrhythmia classification using adult ECG pretraining and a hybrid ResNet50–Transformer network. The proposed framework first pretrains the hybrid model on the MIT-BIH Arrhythmia Database to learn generalized cardiac electrophysiological representations from a large-scale adult ECG dataset. The pretrained model is then fine-tuned on the Non-Invasive Fetal ECG Arrhythmia Database (NIFEADB) to adapt the learned representations for fetal cardiac rhythm analysis. Prior to classification, NI-fECG signals are preprocessed and transformed into time-frequency spectrograms using the Short-Time Fourier Transform (STFT). ResNet50 is employed to extract discriminative spatial features from the spectrograms, while the Transformer encoder models global contextual relationships through self-attention to improve the discrimination of subtle arrhythmic patterns. To eliminate data leakage and provide a clinically realistic evaluation, the proposed framework adopts a strict patient-wise validation protocol, ensuring that recordings from the same subject are never simultaneously included in both the training and testing sets. The framework performs multiclass classification of four clinically significant fetal cardiac conditions, namely Normal Rhythm, Premature Atrial Contractions (PACs), Tachyarrhythmia, and Bradyarrhythmia/Heart Block (HB). Experimental results demonstrate that the proposed patient-wise transfer learning framework effectively leverages knowledge learned from adult ECGs to improve fetal arrhythmia classification despite the limited availability of fetal training data, providing a robust and clinically reliable solution for automated prenatal cardiac screening.
Keywords: Arrhythmia; Electrocardiogram; Self Attention; Transformer

1. Introduction

Congenital cardiac disorders are among the most common fetal abnormalities and affect approximately 0.8% of live births worldwide. Despite significant advances in prenatal diagnosis and fetal cardiac care, cardiac rhythm abnormalities continue to pose substantial risks to fetal development and neonatal survival. Early and reliable identification of fetal arrhythmias enables timely clinical intervention, appropriate pregnancy management, and improved neonatal outcomes, thereby reducing the risk of severe complications such as fetal hydrops, heart failure, and intrauterine fetal demise [1]. Fetal arrhythmia refers to any abnormality in the rhythm or rate of the fetal heartbeat. Under normal physiological conditions, the fetal heart rate ranges from 110 to 160 beats per minute. Persistent deviations from this range often indicate underlying electrophysiological abnormalities that require clinical attention. Fetal arrhythmias encompass a broad spectrum of rhythm disorders, including Premature Atrial Contractions (PACs), Tachyarrhythmia, Bradyarrhythmia, Atrioventricular Re-entrant Tachycardia (AVRT), and Congenital Complete Heart Block (CCHB). Although some arrhythmias are transient and benign, others may significantly increase fetal morbidity and mortality if left undetected [2], [3]. Fetal arrhythmias may be detected throughout pregnancy, with a higher prevalence between the 28th and 32nd weeks of gestation, making continuous prenatal cardiac monitoring particularly important [4].

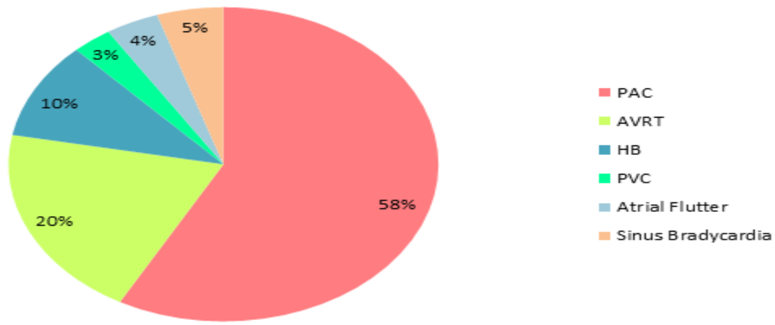

Figure1. Investigated World-wide Distribution of Multi-Class Fetal Arrhythmia

reported in previous epidemiological studies [5]. Among these, Premature Atrial Contractions (PACs) represent the most frequently observed fetal rhythm abnormality, followed by Tachyarrhythmia and Bradyarrhythmia/Heart Block. This unequal distribution highlights the inherent class imbalance that characterizes fetal ECG datasets and increases the difficulty of automated multiclass classification.

The electrocardiogram (ECG) is a widely adopted non-invasive modality for monitoring cardiac electrical activity. In fetal applications, the non-invasive fetal ECG (NI-fECG) provides valuable temporal and morphological information regarding fetal cardiac function. As illustrated in

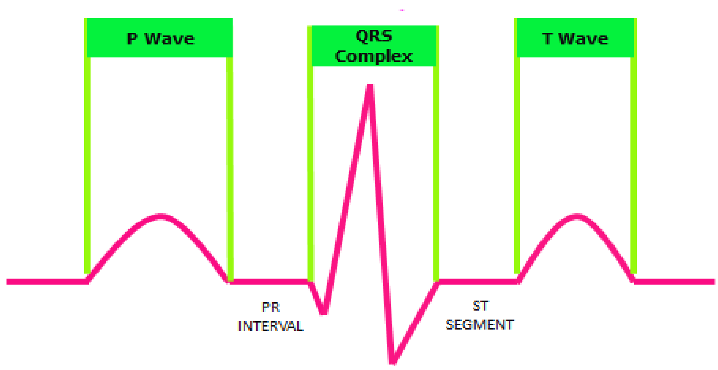

Figure 2. Representation of ECG waveform P-QRS-T

Fig. 2, a normal cardiac cycle consists of the P wave, QRS complex, and T wave, representing atrial depolarization, ventricular depolarization, and ventricular repolarization, respectively. Morphological variations in these waveform components often indicate specific cardiac rhythm abnormalities and therefore provide clinically relevant information for automated arrhythmia detection [6], [7]. Despite its clinical significance, automated fetal ECG analysis remains considerably more challenging than adult ECG analysis. Fetal ECG signals exhibit extremely low amplitudes and are heavily contaminated by maternal ECG, fetal movement artifacts, baseline wander, muscle noise, and power-line interference. Furthermore, publicly available annotated fetal ECG datasets are extremely limited, with only a small number of patients available for each arrhythmia category. Such limited data make it difficult for deep learning models to learn generalized discriminative representations while simultaneously increasing the risk of over fitting.

2. Literature Review

The Multi-class classification of fetal arrhythmias presents a significantly greater challenge compared to binary classification, as it requires distinguishing between multiple abnormal cardiac conditions with subtle differences in ECG patterns. The complexity further increases due to variations in signal quality and class imbalance within available datasets. Recent research has explored various deep learning approaches for ECG-based arrhythmia classification. [8]

Proposed work on multi-arrhythmia classification of human heart using ECG, they have used a fusion model of CNN and transformer for training and testing purposes, mainly considered five types of arrhythmia as a threat to human beings, and for training purposes, took the MIT-BIH dataset and achieved 99.2% accuracy. [9] They were created an automated deep learning model capable of accurately classifying ECG signals into three categories: cardiac arrhythmia (ARR), congestive heart failure (CHF), and normal sinus rhythm (NSR). [10], [11] to achieve this, ECG data were taken from the MIT-BIH and BIDMC databases available on PhysioNet. Pre-trained models, ResNet50 and AlexNet, were fine-tuned and configured to achieve optimal classification results. In this study, [12] they used the adult ECG MIT-BIH dataset to train a deep learning model, and the fetal ECG NIFEADB dataset was used as input for the detection of arrhythmia. [13] proposed a model of detection and identification of arrhythmia & they developed a small CNN model of 12 layers, and its performance was compared against pre-trained models like GoogLeNet. In the context of fetal ECG analysis, some studies have utilized a combination of adult and fetal datasets to enhance model generalization [16]. [17] Proposed a deep convolutional neural network-based approach (DCNN) that uses the Fourier transform to design a two-dimensional convolutional deep CNN architecture after transforming the ECG signal to the time-frequency domain. [18] The architecture trains a specific DCNN for ECG samples of a specific length to learn the features of ECG wave signals from the time domain and perform classification in an end-to-end manner. [19] Conducted classification studies using a convolutional recurrent neural network (CRNN) and SVM, and proposed an ECG classification algorithm that combined a bidirectional recurrent neural network (BiRNN) with a CNN. [20], [21] Achieved superior classification performance using a bidirectional long short-term memory (bi-LSTM) model.

TABLE 1. PREVIOUS STUDIES OF CLASSIFICATION OF FETAL ECG ARRHYTHMIA

S.No	**Authors**	**Techniques**	**Dataset**	**Accuracy**	**Limitations**	**Training Data**
1.	Shu et al.[23] (2018)	CNN & LSTM	NIFEADB	98.10%	Limited generalization due to use small dataset.	Segment Wise Classification
2.	Mohebian et al.[12] (2019)	CNN	MIT-BIH & NIFEADB	92%	Work on the spatial feature only.	Segment Wise Classification
3.	Y. Li et al [15] (2022)	CNN & Transfer Learning	MIT-BIH	86%	Human multi class Arrhythmia classification.	Patient Wise Classification
4.	Sara N et al.[33] (2022)	CNN	NIFEADB	83.3%	Over-fitting and lack of temporal feature extraction.	Patient Wise Classification
5.	Jie Liu et al. [22] (2023)	Multi-domain Feature Extraction	NIFEADB	96.0%	Models tend to Over-fit domain-specific patterns, reducing cross-dataset robustness.	Segment Wise Classification
6.	Rai et al. [14] (2023)	CNN & Transfer Learning	NIFEADB	98.56%	Lack of time-frequency feature extraction.	Segment Wise Classification
7.	Rami S. et al.[12] (2023)	CNN & BiLSTM	MIT-BIH	98.6%	Patient wise generalization to other datasets was not investigated.	Patient Wise Classification
8.	N.Tulyakova et al.[16] (2024)	CNN Attention & BiLSTM	MIT-BIH	98.24%	Same data may be in training and testing	Segment Wise Classification

9.	Negin et al. [26] (2022)	CNN & LSTM	MIT-BIH	98.5%	This model does not capture long-range global dependencies.	Train–Test Split
10	Xu et al. [29] (2024)	CNN & LSTM	MIT-BIH	98.52%	Complexity of the model was not validated using patient-wise cross-validation.	10-fold Cross Validation

The approaches and associated limitations of recent research on the classification of fetal ECG arrhythmias are collected in the above table. Morphological characteristics and heart rate variability (HRV) analysis are the mainstays of the majority of current methods for the prediction and categorization of fetal cardiac arrhythmias [24], [25]. Nevertheless, previous research has not extensively integrated hybrid deep learning frameworks that combine transfer learning and attention-based architectures [27]. The current work uses an advanced hybrid architecture that blends a Transformer model with ResNet50 in order to close this gap. While the Transformer uses self-attention processes to capture long-range dependencies and contextual interactions among ECG data, ResNet50 is used for deep spatial feature extraction. The suggested hybrid approach improves the precision and resilience of fetal arrhythmia prediction and multi-class classification by utilizing many feature representations concurrently.

3. MATERIALS & Methods

The In heart defect different diagnosis tools are used to monitoring health of heart that is based on different AI (Artificial Intelligent) Techniques [30], [31]. Fetal arrhythmia detection and classification by analysis of non-invassive ECG signal through advance deep learning techniques like CNN, RNN & Transformer provided the significant information

a. Overall Proposed Framework

The proposed study presents a patient-wise deep transfer learning framework for multiclass fetal ECG arrhythmia classification using adult ECG pretraining and a hybrid ResNet50–Transformer network.

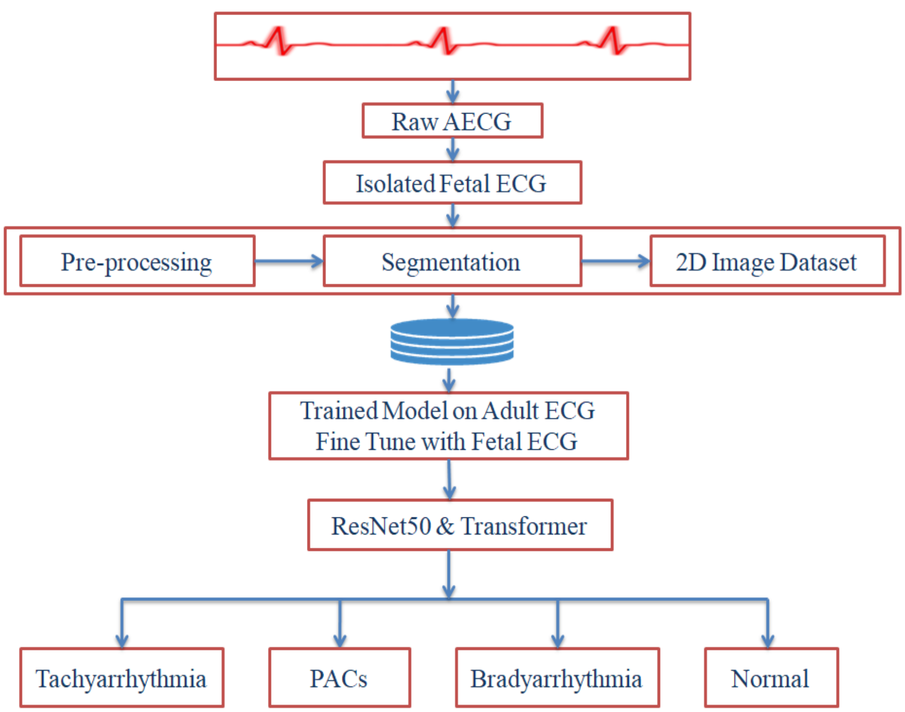

Figure 3. Proposed architecture for multi-class fetal ECG arrhythmias classification

The overall workflow of the proposed framework is illustrated in Fig. 3. Unlike conventional deep learning approaches that directly train models using limited fetal ECG recordings, the proposed framework adopts a two-stage transfer learning strategy to improve feature learning and generalization under small-data conditions.

In the first stage, a hybrid ResNet50–Transformer model is pretrained using the MIT-BIH Arrhythmia Database, a large-scale adult ECG dataset containing diverse cardiac rhythm patterns. This pretraining stage enables the network to learn generalized cardiac electrophysiological representations, including morphological characteristics of the P wave, QRS complex, and T wave, which are fundamental to both adult and fetal cardiac rhythm analysis.

In the second stage, the pretrained model is transferred to the Non-Invasive Fetal ECG Arrhythmia Database (NIFEADB) and fine-tuned using fetal ECG recordings. Prior to training, abdominal ECG recordings are preprocessed using band-pass filtering and Independent Component Analysis (ICA) to suppress maternal ECG interference and other noise components while preserving diagnostically relevant fetal cardiac activity. The extracted fetal ECG signals are subsequently transformed into two-dimensional time–frequency spectrograms using the Short-Time Fourier Transform (STFT), enabling the utilization of deep convolutional feature learning.

The generated spectrograms are processed by the ResNet50 backbone to extract hierarchical spatial representations, whereas the Transformer encoder models long-range contextual relationships through multi-head self-attention. The complementary integration of convolutional feature extraction and Transformer-based contextual modeling enables the proposed framework to capture both local morphological characteristics and global temporal dependencies of fetal cardiac rhythms. Finally, the learned feature representations are forwarded to a Softmax classification layer for multiclass prediction of four clinically significant fetal rhythm categories: Normal Rhythm, Premature Atrial Contractions (PACs), Tachyarrhythmia, and Bradyarrhythmia/Heart Block (HB).

b. Adult ECG Dataset and Pretraining Strategy

The proposed framework employs a two-stage deep transfer learning strategy to address the limited availability of annotated fetal ECG recordings. In the first stage, the hybrid ResNet50–Transformer model is pretrained using a large-scale adult ECG dataset to learn generalized cardiac electrophysiological representations. These learned representations are subsequently transferred and fine-tuned for multiclass fetal ECG arrhythmia classification. [11] The MIT-BIH Arrhythmia Database, obtained from PhysioNet, was selected as the source dataset for model pretraining. This database is one of the most widely used benchmark datasets for automated ECG analysis and contains 48 half-hour ambulatory ECG recordings collected from 47 subjects, sampled at 360 Hz. The recordings include a broad range of normal and abnormal cardiac rhythms, providing sufficient morphological diversity for learning robust ECG feature representations. The dataset contains annotations for various heartbeat categories, enabling the model to capture discriminative characteristics associated with different cardiac rhythm abnormalities. [32], [35] to ensure realistic model evaluation and prevent subject-level information leakage, the adult ECG recordings were partitioned using a patient-wise data splitting strategy, where recordings from the same subject were assigned exclusively to either the training or validation set. This protocol allows the pretrained model to learn generalized ECG characteristics from previously unseen individuals rather than memorizing patient-specific waveform patterns. Prior to training, the one-dimensional ECG signals were converted into two-dimensional time–frequency spectrograms using the Short-Time Fourier Transform (STFT). The spectrogram representation preserves both temporal and frequency-domain information, enabling the use of convolutional neural networks originally developed for image analysis. These spectrograms served as the input to the hybrid ResNet50–Transformer model during the pretraining stage.

c. Fetal ECG Dataset

The target dataset used for multiclass fetal arrhythmia classification is the Non-Invasive Fetal ECG Arrhythmia Database (NIFEADB)**,** publicly available through PhysioNet. This database contains synchronized non-invasive abdominal electrocardiogram (NI-fECG) recordings collected from pregnant women with both normal and abnormal fetal cardiac rhythms. Unlike adult ECG databases, NIFEADB specifically focuses on fetal cardiac electrophysiological activity, making it an appropriate benchmark for developing automated fetal arrhythmia classification systems. The recordings were acquired using multiple abdominal electrodes placed on the maternal abdomen together with reference maternal chest electrodes to facilitate fetal ECG extraction. Depending on the recording protocol, the signals were sampled at 500 Hz or 1000 Hz, providing sufficient temporal resolution for detailed cardiac rhythm analysis. Each recording contains maternal ECG interference, fetal ECG components, and various physiological and environmental noise sources, making fetal arrhythmia detection considerably more challenging than conventional adult ECG analysis.

TABLE 2. SUMMARY OF ARRHYTHMIA AND NORMAL PATIENT IN DATASET

Arrhythmia Class	Number of Patients
Normal Rhythm	14
Premature Atrial Contractions (PACs)	3
Tachyarrhythmia	6
Bradyarrhythmia / Heart Block (HB)	3
Total	**26**

d. Patient-Wise Deep Transfer Learning Strategy

Deep learning models generally require a large volume of annotated training data to learn robust and generalized feature representations. However, publicly available fetal ECG databases contain only a limited number of patients, making the direct training of high-capacity deep neural networks prone to overfitting and poor generalization. To overcome this limitation, the proposed framework employs a patient-wise deep transfer learning strategy**,** in which knowledge acquired from a large-scale adult ECG dataset is transferred to the fetal ECG classification task.

The proposed transfer learning framework consists of two sequential stages, namely source-domain pretraining and target-domain fine-tuning**,** as illustrated in Fig. 5**.** During the first stage, the hybrid ResNet50–Transformer network is pretrained using the MIT-BIH Arrhythmia Database to learn generalized representations of cardiac electrical activity. Rather than learning fetal-specific characteristics, the network captures fundamental ECG morphology, including the P wave, QRS complex, T wave, rhythm variations, and temporal dependencies that are common across human cardiac signals.

Let (A) denote the Adult ECG dataset and (F) represent the Fetal ECG dataset. The model was first pre-trained on the Adult ECG dataset.

M_A = Train(A)

where (M_A) represents the trained model containing the learned feature representations from Adult ECG signals.

Thereafter, the pre-trained model was fine-tuned using the fetal ECG dataset:

$$M_F = Fine\_tune\ (M_A, F)$$

e. Data Acquisition and Pre-processing

ECG (fECG) signals are typically acquired from abdominal ECG (AECG) recordings using electrodes placed on the maternal abdomen. These recordings are often contaminated by various types of noise, including baseline wander, maternal muscle activity, and power-line interference, which can significantly degrade signal quality. Therefore, before feature extraction and further analysis, a band-pass filter was initially applied to suppress baseline wander, motion artifacts, and high-frequency noise. The abdominal ECG signal was filtered within the range of 5–15 Hz, which preserves the essential components of both maternal and fetal cardiac activity while attenuating powerline interference and electromyographic noise. However, since maternal ECG exhibits higher amplitude and overlaps spectrally with fetal ECG, Therefore, advanced blind source separation techniques, Independent Component Analysis (ICA) were employed to extract the fetal ECG component from the composite abdominal signal. These methods exploit the statistical independence and morphological differences between maternal and fetal cardiac signals, enabling effective suppression of maternal interference while retaining diagnostically relevant fetal QRS complexes [22].

f. Segmentation

The NIFEADB database contains long-duration non-invasive fetal ECG recordings acquired at sampling frequencies of 500 Hz or 1000 Hz. Following signal preprocessing and fetal ECG extraction, each recording was segmented into fixed-length windows to preserve the morphological characteristics of individual cardiac cycles while providing sufficient temporal context for arrhythmia analysis. To generate robust two-dimensional representations suitable for deep learning, the Short-Time Fourier Transform (STFT) was employed. STFT was selected because it simultaneously preserves the temporal and spectral characteristics of ECG signals, making it particularly suitable for analyzing the non-stationary nature of fetal cardiac activity. Unlike one-dimensional waveform analysis, STFT transforms each ECG segment into a time–frequency spectrogram, enabling convolutional neural networks to learn both morphological and frequency-domain characteristics associated with fetal arrhythmias. Prior to STFT computation, fetal heartbeats was localized using QRS complex detection, allowing the extraction of individual cardiac segments. Accurate localization of the QRS complex is essential because it contains the highest signal energy and provides reliable estimation of R–R intervals, which represent the temporal interval between two consecutive ventricular depolarizations and constitute an important indicator of cardiac rhythm abnormalities. A Hamming window of 300 samples with a hop length of 150 samples (50% overlap) was employed to minimize spectral leakage while maintaining an appropriate balance between temporal and frequency resolution. The overlapping window strategy preserves the continuity of ECG morphology and improves the representation of transient rhythm variations. Each segmented fetal ECG signal was transformed into a two-dimensional time–frequency spectrogram and resized to 224 × 224 pixels, matching the input dimensions required by the ResNet50 backbone for feature extraction.

g. Feature Extraction using ResNet-50

The adoption of transfer learning facilitates faster convergence and improves generalization by leveraging pre-learned visual representations. Furthermore, the residual connections in ResNet-50 effectively mitigate the vanishing-gradient problem, enabling stable training of deep architectures while adapting to domain-specific fetal ECG spectrogram representations. [28] The pre-processed 2-D time–frequency spectrograms of fetal ECG signals are provided as input to ResNet-50 to extract hierarchical and discriminative local feature representations associated with fetal cardiac dynamics. During the fine-tuning process, only the final convolutional block is updated, while the earlier layers are frozen to retain generic low- and mid-level visual features learned from large-scale natural images.ResNet-50 is a 50-layer deep convolutional neural network originally proposed by [29], [34]. Its primary innovation lies in the introduction of residual (skip) connections, which facilitate efficient training of deep networks by alleviating gradient degradation. Instead of directly learning a mapping H(x), ResNet learns a residual function F(x), expressed as:

$$H(X) = F(X) + X \tag{1}$$

Where x denotes the input to a residual block, and F(x) represents the learned nonlinear transformation composed of convolutional layers, batch normalization, and ReLU activation. This residual formulation allows the network to learn incremental feature refinements rather than complete transformations, thereby enhancing convergence stability and representational capacity.

h. Feature Extraction Though Transformer

To enhance the contextual representation learned by the convolutional backbone, the proposed framework incorporates a Transformer encoder after the ResNet50 feature extraction stage. While ResNet50 effectively captures local spatial patterns from the STFT spectrograms, its convolutional operations have limited capability to model long-range dependencies among the extracted features. Therefore, a Transformer encoder is introduced to capture global contextual relationships using the self-attention mechanism.

The STFT spectrograms are first processed by the ResNet50 backbone to generate high-level feature representations. These feature embeddings are subsequently projected into a suitable embedding space and enriched with positional encoding to preserve the sequential order of the extracted feature vectors. The resulting sequence is then passed to a single Transformer encoder layer. The Transformer encoder consists of a multi-head self-attention module followed by a position-wise feed-forward network (FFN). Multi-head attention enables the model to simultaneously focus on different regions of the feature sequence, thereby learning both local morphological characteristics and long-range contextual dependencies. This complementary integration allows ResNet50 to extract discriminative spatial features, while the Transformer encoder models global feature interactions, leading to improved fetal arrhythmia classification performance.

Table 3: Transformer encoder architecture and hyper parameter configuration.

Parameter	**Value**

No. of Transformer encoder layers	1
No. of attention heads	4
Key dimension	64
Feed-forward network size	512
Dropout	0.2
Dense layer	256 neurons

The proposed framework employs a single Transformer encoder layer following the ResNet50 feature extraction stage to capture global contextual dependencies among the extracted feature embeddings. The encoder utilizes multi-head self-attention with four attention heads, where each head has a key dimension of 64. The extracted feature embeddings are processed through a feed-forward network (FFN) comprising 512 hidden units with the Gaussian Error Linear Unit (GELU) activation function, followed by a projection back to the original embedding dimension. The input feature representation is first combined with positional encoding to preserve the sequential order of the ECG feature vectors. The resulting input to the Transformer encoder is expressed as

$Z = E + PE$ (2)

three parameters which are mathematical represented as:

$Query: \ Q = Zw$ (3)

$keys: \ K = Z\,w$ (4)

$Values: \ V = Zw$ (5)

In Equations (3), (4), and (5), the Query (Q) represents the feature vector that seeks relevant information from other sequence elements, the Key (K) encodes the content of each heartbeat to determine its relevance, and the Value (V) contains the actual feature information that is weighted and propagated forward through the attention mechanism.

WQh, WKh, Wvh are learnable weight matrices for each head. The scaled dot-product attention is mathematical represented as:

$$A = Softmax\left(\frac{Q}{}\right)$$

Where K, V & d_k are Key matrix, Value matrix and Dimension of the key vectors respectively. Softmax normalizes the attention scores into a probability distribution. Here, were concatenated, which mathematical is represented as:

$$Z = conc(A, A, \dots \dots \dots \dots \dots \dots A)$$

Where W_o is a learnable weight matrix for output projection.

$$Output = Z + X$$

i. ECG Signal Processing Through Hybrid Model ResNet-50 and Transformer

After fine-tuning on the NIFEADB fetal ECG dataset, the ResNet50 backbone serves as a deep feature extractor for fetal ECG spectrograms. Let (XF) denote the input fetal ECG spectrogram generated through the STFT process. The deep feature representation extracted by ResNet50 is expressed as feature embedding is represented as:

$$Z = M_F\ (X_F)$$

The feature embedding Z is reshaped into sequential tokens and fed into the Transformer encoder for global contextual learning.

$$T = Trasnformer\ (Z)$$

The Transformer output is combined with the learned feature representation and forwarded to the classification layer.

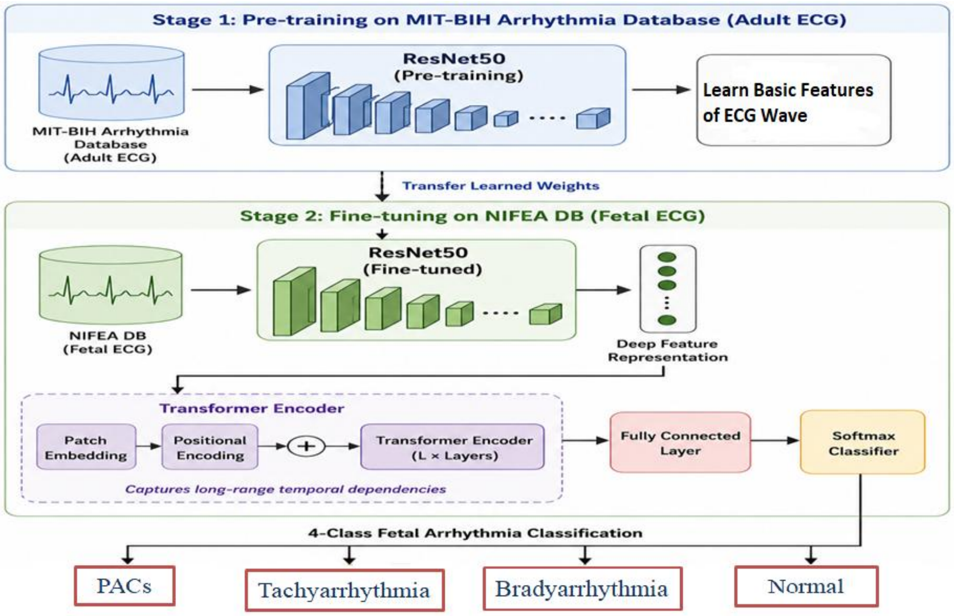

Figure 4: Hybrid Framework of ResNet-50 and Transformer

ResNet50–Transformer framework for multiclass fetal ECG arrhythmia classification, the framework consists of three sequential stages. First, the ResNet50 backbone is initialized with ImageNet-pretrained weights and subsequently pre-trained on the MIT-BIH Arrhythmia Database to learn generalized ECG feature representations from adult cardiac signals. Second, the pretrained network is fine-tuned using the NIFEADB fetal ECG database, enabling adaptation to fetal-specific cardiac morphology and rhythm characteristics. Finally, the extracted deep feature embeddings are processed by a Transformer encoder, which employs positional encoding and multi-head self-attention to model global contextual dependencies. The contextualized feature representations are then passed through a fully connected dense layer followed by a Softmax classifier to categorize each fetal ECG recording into one of four rhythm classes: Normal Rhythm, Premature Atrial Contractions (PACs), Tachyarrhythmia, and Bradyarrhythmia/Heart Block.

j. Classification Layer

The contextual feature representations generated by the Transformer encoder integrate both the discriminative local morphological characteristics extracted by the ResNet50 backbone and the global contextual dependencies learned through the multi-head self-attention mechanism. These enriched feature embeddings are subsequently forwarded to a fully connected (dense) layer, which projects the learned representations into class-specific logits. The final classification is performed using the Softmax activation function, which converts the output logits into a normalized probability distribution over the four fetal rhythm classes. The predicted class is determined by selecting the category with the highest posterior probability.

$$Y = Softmax\,(WT + b)$$

Where T denotes the feature vector obtained from the Transformer encoder, W represents the learnable weight matrix, b is the bias term, and Y corresponds to the predicted probability distribution in four classes Normal, PACs, Tachyarrhythmia and Bradyarrhythmia.

4. Results & Discussion

a. Computational Setup & Deployment Analysis

The proposed ResNet50–Transformer hybrid model was implemented using the TensorFlow and Keras deep learning frameworks in Python. All experiments were conducted on a workstation equipped with an Intel Core i7-

9750H processor, 32 GB RAM, and an NVIDIA GeForce GTX 1650 Max-Q GPU (4 GB GDDR5 VRAM). The software environment consisted of Anaconda 2023.07.1, Python 3.11, TensorFlow, and Jupyter Notebook 7.3.2. The proposed model contains approximately 24.7 million trainable parameters, including the ResNet50 backbone and the Transformer encoder. During transfer learning, the ResNet50 backbone was initialized with ImageNet pretrained weights and subsequently fine-tuned using the fetal ECG image dataset. The average training time per epoch was approximately 45–60 seconds on the above hardware configuration, depending on the number of trainable layers during fine-tuning. The complete model converged within 50 epochs, requiring approximately 22–30 minutes for training. For practical deployment, the average inference time of the proposed model was approximately 18–25 ms per ECG image, enabling real-time classification of fetal ECG arrhythmias. These computational results indicate that the proposed framework provides an effective trade-off between classification performance and computational efficiency, making it suitable for real-time clinical decision support system.

b. The Patient-Wise Evaluation Protocol

To ensure subject-independent evaluation and eliminate potential data leakage, a Leave-One-Patient-Out Cross-Validation (LOPO-CV) strategy was adopted. During each iteration, all heartbeat segments belonging to one subject were reserved exclusively for testing, while the remaining subjects were used for model training and validation. This process was repeated until every subject had been used once as the test subject. Unlike segment-wise random splitting, the proposed evaluation protocol prevents ECG segments originating from the same patient from appearing simultaneously in both training and testing sets. Consequently, the reported performance reflects the true generalization capability of the proposed framework in unseen fetal subjects. The final classification performance was computed by averaging the results obtained across all LOPO iterations. Accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC) were used as performance evaluation metrics.

Table-4 the total number of generated samples for each class is summarized (Behar et al., 2019)

Class	**Subjects**	**Generated Segments**	**Heart Rate**	**Features of Image** **mm:ss** **RR Interval**
Normal Rhythm	14	14*300=4200	110–160 bpm	10:05 to 10:27
PACs	3	3*300=900	Mostly normal but irregular	10:04 to 12:37
Tachyarrhythmia	6	6*300=1800	>180 bpm	08:01 to 32:05
Bradyarrhythmia	3	3*300=900	<110 bpm	10:05 to 11:05

A major challenge in fetal ECG classification is the possibility of subject-level data leakage. If ECG segments originating from the same patient appear simultaneously in both training and testing datasets, the reported classification performance may be artificially inflated. To eliminate this issue, a patient-wise validation protocol was adopted in this study. All ECG segments belonging to a particular subject were assigned exclusively to either the training or testing set. Therefore, no patient contributed samples to both datasets simultaneously. This subject-independent evaluation strategy provides a realistic assessment of model generalization performance and prevents information leakage between training and testing phases. Considering the limited number of fetal arrhythmia subjects available in NIFEADB, Leave-One-Patient-Out Cross Validation (LOPO-CV) was employed for performance evaluation. During each iteration, one subject was selected for testing. The remaining subjects were used for training and validation. The process was repeated until every subject had been used once as the testing subject. The final classification performance was obtained by averaging the results across all LOPO iterations. This evaluation strategy is particularly suitable for small medical datasets and ensures unbiased estimation of subject-independent classification performance.

C. Evaluation of Experimental Metrics

For the assessment of multi-class classification performance, standard evaluation metrics are derived based on the comparison between true and predicted class labels. Correct classifications are represented by true positives (TP) and true negatives (TN), whereas incorrect classifications are represented by false positives (FP) and false negatives (FN). A correct classification occurs when the predicted label matches the true label. An incorrect classification occurs when an arrhythmic signal is predicted as normal, which is categorized as a false negative (FN), or when a normal signal is predicted as arrhythmic, which is categorized as a false positive (FP). These misclassifications correspond to type-II and type-I errors, respectively.

Table 5. Representation of Evaluation metrics

S. No.	**Name of Metrics**	**Explanation of Metrics**	**Formulation of Metrics**
1.	Precision	It measures the correctly classified value of true prediction.	$\frac{TP}{TP + FP}$
2.	Recall	It measures the positive fraction in total sample of (Positive+ Negative).	$\frac{TP}{TP + FN}$
3.	F1-Score	It plays the important role to count false cases in total sample.	$\frac{2 \times Recall \times Precision}{Recall + Precision}$
4.	Accuracy	It is the main evaluation of correct classification in total sample.	$\frac{TP + TN}{TP + TN + FN + FP}$

The training and validation accuracy of ResNet50 Transformer concerning the number of epochs is shown in Figure 8. From the first five epochs, the training and validation accuracies of ResNet50 Transformer are reported as 80% and 67%, respectively. This demonstrates that the use of transfer learning allows the model to begin training from a higher-level feature representation, thereby reducing computational cost.

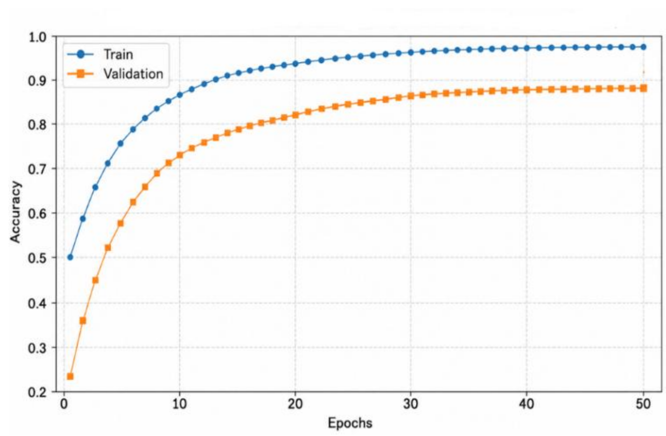

Figure 5. Training and validation accuracies by ResNet50 Transformer

It is understood from the mentioned figure 5 that the proposed model achieved maximum accuracy 93.4% at epoch 50. As we applied ResNet50 with transformer has worked more efficiently for multilevel fetal ECG arrhythmia classification. During training loss function indicate how well model perform the classification of image, in our model to reduce the loss in training we have used cross entropy loss function which is widely used in multi-class classification.

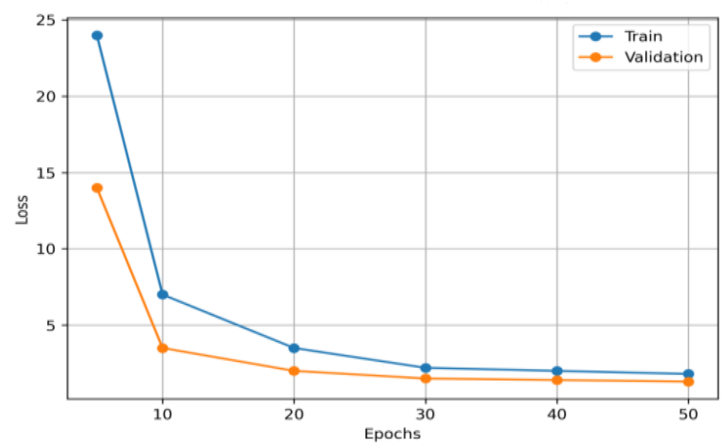

Figure 6. Training loss by ResNet50 Transformer for multi class of fetal ECG

Figure 6 shows the training and validation loss curves of the proposed model. Both losses decrease steadily with increasing epochs and reach their minimum values around epoch 50, indicating stable convergence and effective optimization without significant over fitting. Another way to show model's performance is confusion matrix, for multi classification heat-map visualize different class and its classification accuracy by darkness of color. Each class show inside the cell if color of cell is more darkness than model predicted efficiently.

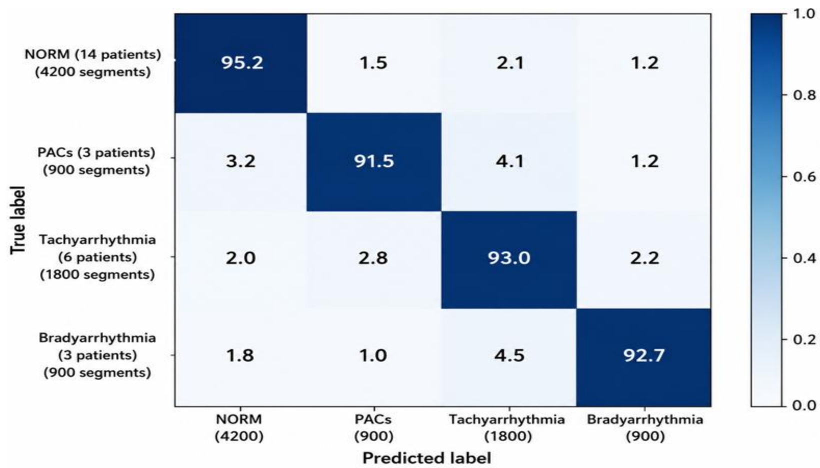

Figure 7. Confusion matrix for multiclass fetal ECG arrhythmia

The confusion matrix provides a detailed visualization of class-wise prediction performance. Darker diagonal cells indicate a higher number of correctly classified samples, whereas the off-diagonal cells represent misclassifications. The confusion matrix demonstrates that most samples were correctly assigned to their respective classes, with only a small number of misclassifications observed among PACs and Bradyarrhythmia due to class imbalance.

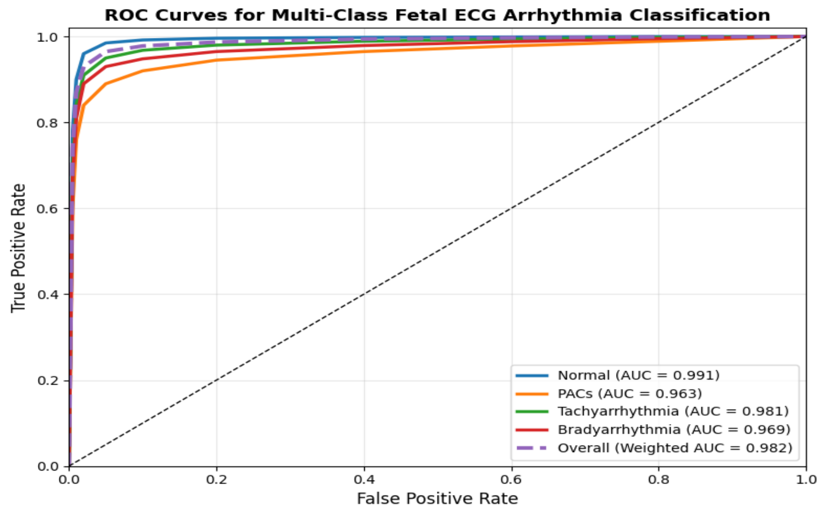

Figure 8. Representation of ROC curve for multiclass classification

Figure 8 illustrates the Receiver Operating Characteristic (ROC) curves of the proposed ResNet50-Transformer model for four-class fetal ECG arrhythmia classification. The ROC curves are concentrated toward the upper-left corner, indicating a high true positive rate (sensitivity) and a low false positive rate, which demonstrates the excellent discriminative capability of the proposed model. The model achieved a macro-average AUC of 0.976 and a weighted-average AUC of 0.982**,** confirming robust classification performance despite the class imbalance in the dataset. Among the four classes**,** Normal achieved the highest AUC, while Tachyarrhythmia, Bradyarrhythmia, and PACs also exhibited high AUC values, indicating reliable discrimination between normal and arrhythmic fetal ECG patterns. These results demonstrate that the proposed ResNet50-Transformer framework provides accurate and robust multi-class classification of fetal ECG arrhythmias.

Table 6. Performance of the Proposed ResNet50-Transformer Model on the Validation Dataset

S.No	**Category**	**Precision**	**Recall**	**F1-score**
1.	Norm	94.54	93.64	94.09
2.	PACs	88.30	85.20	86.72
3.	Tachyarrhythmia	92.45	91.23	91.84
4.	Bradyarrhythmia	88.34	82.27	85.20

The above table shown class wise accuracy of fetal ECG, the performance of multi classification is shown variation because the dataset of each class have not balanced in NIFEADB, The image dataset of normal fetal ECG exist large amount of images in comparison of fetal arrhythmia ECG image dataset so it has shown variation in performance but overall accuracy represent 93.4%. The statistical analysis of the five experimental runs demonstrates the consistency and robustness of the proposed ResNet50-Transformer model for multi-class fetal ECG arrhythmia classification.

Table 7: Statistical analysis of the proposed ResNet50–Transformer model based on five experiments

S.No.	**Experiment Cycle**	**Average Accuracy of 4 Classes (%)**
i	Exp-1 in 50 epoch	93.40
ii	Exp-2 in 40 epoch	92.40

iii	Exp-3 in 35 epoch	91.80
iv	Exp-4 in 60 epoch	93.10
v	Exp-5 in 45 epoch	92.70

The average classification accuracy across all experiments was 92.68%, with a standard deviation of 0.62%, indicating low performance variation and stable model behaviour under different training epochs. Accordingly, the performance of the proposed model can be expressed as 92.68 ± 0.62% Furthermore, the 95% confidence interval (CI) was calculated as 91.91% – 93.45% suggesting that the true mean classification accuracy is expected to lie within this range with 95% confidence. These statistical results confirm that the proposed framework provides reliable and reproducible performance for multi-class fetal ECG arrhythmia classification. The performance of the proposed ResNet50–Transformer framework was compared with several widely used deep learning architectures, including a conventional CNN, standalone ResNet50 and Transformer

Table 8: Comparative evaluation of different deep learning models for multi-class fetal ECG arrhythmia classification.

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
CNN	88.60	88.20	87.80	87.95
ResNet50	91.30	91.00	90.70	90.85
Transformer	90.20	89.90	89.50	89.70
Proposed ResNet50 + Transformer	93.40	91.90	90.60	91.00

The experimental results demonstrate that the proposed hybrid framework achieved the highest classification accuracy of 93.4%, outperforming the standalone ResNet50, Transformer, conventional CNN. he improved performance can be attributed to the complementary strengths of the two architectures. ResNet50 effectively extracts hierarchical spatial features from fetal ECG spectrograms, whereas the Transformer encoder captures long-range contextual dependencies through its multi-head self-attention mechanism. The integration of these complementary feature representations enables the proposed model to distinguish subtle morphological differences among Normal, PACs, Tachyarrhythmia, and bradarrhythmia more effectively than either architecture alone.

5. Conclusions

The preprocessed fetal ECG spectrograms were first processed by the ResNet50 backbone to extract discriminative spatial features. These feature embeddings were subsequently refined by the Transformer encoder, which captured long-range contextual dependencies through the self-attention mechanism. The data underwent pre-processing to remove noise and artifacts that could adversely affect model performance. The clean fetal ECG signals were then converted into a 2D dataset and fed into the transformer network for feature extraction using its attention mechanism, which effectively models the temporal characteristics of the signals. Meanwhile, the ResNet-50 architecture was employed to extract global features, and its integration with the transformer further improved the multi-class classification performance. The combination of ResNet-50 and transformer architectures provides a robust and efficient framework for addressing the complexity of fetal ECG arrhythmia classification. The proposed model achieved a maximum classification accuracy of 93.4% after 50 epochs, demonstrating its strong potential for accurate and reliable fetal ECG.

References:

1. T. Miyoshi, "Fetal arrhythmias: Current evidence of prenatal diagnosis and management of a novel active separation control approach." Journal of Obstetrics and Gynaecology Research, vol. 51, no. 3, (2025).
2. T. K. Braley, G. Dadlani, J. Geiger, K. Douglas, and M. Mehta,"Introduction to fetal echocardiography." Progress in Pediatric Cardiology, vol. 58, p. 101278, (2020).
3. N. Dia et al.,"Fetal heart rate estimation by non-invasive single abdominal electrocardiography in real clinical conditions."Biomedical Signal Processing and Control, vol. 71, no. 3,p. 103187–103187(2021).

4. “The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals.”Annals of Internal Medicine, vol. 113, no. 6, p. 487,(1990).
5. C. Sen, O. Uzun, GokhanGoynumer, and B. Beattie, “Systematic Appraisal of Diagnosis and Management of Arrhythmias in the Fetus.” Donald School Journal of Ultrasound in Obstetrics and Gynecology, vol. 9, no. 3,p. 314–326,(2015).
6. M. S. Sklansky, “Fetal Cardiology Simplified: A Practical Manual.” Future Cardiology, vol. 9, no. 5,p. 611–613,(2013).
7. A. Veduta et al.,“Treatment of Fetal Arrhythmias.” Journal of Clinical Medicine, vol. 10, no. 11,p. 2511(2021).
8. F. Zhou et al.,“Heartbeat classification method combining multi-branch convolutional neural networks and transformer.”iScience, vol. 27, no. 3, p. 109307,(2024).
9. Ö. Yıldırım, P. Pławiak, R.-S. Tan, and U. R. Acharya “Arrhythmia detection using deep convolutional neural network with long duration ECG signals.”Journal of Obstetrics and Gynaecology Research, vol. 51, no. 3, p. 411-420, (2018).
10. G. B. Moody and R. G. Mark, “The impact of the MIT-BIH Arrhythmia Database.”IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, (2001).
11. A. L. Goldberger et al., “PhysioBank, PhysioToolkit, and PhysioNet.”Circulation, vol. 101, no. 23, (2001).
12. M. R. Mohebbian, H. R. Marateb, and K. A. Wahid, “Semi-supervised active transfer learning for fetal ECG arrhythmia detection.” Computer Methods and Programs in Biomedicine Update, vol. 3, p.100096, (2023).
13. N. Katal et al., “Deep-Learning-Based Arrhythmia Detection Using ECG Signals: A Comparative Study and Performance Evaluation.”Diagnostics, vol. 13, no. 24, p.3605, (2023).
14. R. K. Rai, A. Singh, R. Srivastva, and G. Kumar, “Fetal ECG Arrhythmia Detection Based on DensNet Transfer Learning.”Frontiers in Biomedical Technologies, vol.10, (2023).
15. Y. Li, R. Qian, and K. Li,“Inter-patient arrhythmia classification with improved deep residual convolutional neural network.”Computer methods and Programs in Biomedicine, vol. 214, p.106582, (2022).
16. N. Tulyakova and O. Trofymchuk, “Real-time filtering adaptive algorithms for non-stationary noise in electrocardiograms.”Biomedical Signal Processing and Control, vol. 72, p.103308, (2022).
17. J. Zhang, et al.,“Fine-grained ECG Classification Based on Deep CNN and Online Decision Fusion.” arXiv:1901.06469,vol.3,(2019).
18. J. Zhang, et al.,“Fine-grained ECG Classification Based on Deep CNN and Online Decision Fusion.” arXiv:1901.06469,vol.3,(2019).
19. M. Limam et al., "Atrial fibrillation detection and ECG classification based on convolutional recurrent neural network." Computing in Cardiology (CinC), Rennes, France, pp. 1-4, (2017).
20. P. Xie et al., "Bidirectional Recurrent Neural Network and Convolutional Neural Network (BiRCNN) For ECG Beat Classification." 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, pp. 2555-2558, (2018).
21. B. Ganguly, A. Ghosal, A. Das, D. Das, D. Chatterjee, and D. Rakshit, “Automated Detection and Classification of Arrhythmia from ECG Signals using Feature Induced Long Short-Term Memory Network.” IEEE Sensors Letters,pp. 1-1, (2020).
22. J. Guan, W. P. Feng, X. Wang, and W. Wang, “Low-Dimensional Denoising Embedding Transformer for ECG Classification.”ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1285-1289, (2021).
23. J. Liu, H. Xu, J. Wang, X. Peng, and C. He, “Non-invasive diagnosis of fetal arrhythmia based on multi-domain feature and hierarchical extreme learning machine.”Biomedical Signal Processing and Control, vol. 79, P.104191 (2022).
24. Shu Lih Oha, Eddie Y.K. Ng, Ru San Tan, U. RajendraAcharya, “Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats.” Computers in Biology and Medicine P. 278–287, (2018).
25. M. T. Anis and V. Sharma, “Classification of ECG Signal Using CNN Algorithm.”pp. 185–189, (2025).
26. Chenyang Qiu et al.,“Enhancing ECG classification with continuous wavelet transform and multi-branch transformer.” Heliyon, vol. 10, (2024).
27. S.-H. Wang, J. Zhou, and Y.-D. Zhang, “Community-Acquired Pneumonia Recognition by Wavelet Entropy and Cat Swarm Optimization.” Mobile Networks and Applications, vol. 29, no. 4, pp. 1122-1139, (2022).
28. M. Wasimuddin, K. Elleithy, A. Abuzneid, M. Faezipour, and O. Abuzaghleh, “Stages-Based ECG Signal Analysis from Traditional Signal Processing to Machine Learning Approaches: A Survey.”IEEE Access, pp.1-1, (2020).
29. K. Ma, C. A. Zhan, and F. Yang, “Multi-classification of arrhythmias using ResNet with CBAM on CWGAN-GP augmented ECG Gramian Angular Summation Field.”Biomedical Signal Processing and Control, vol. 77, p.103684, (2022).
30. Xu, C., Wu, J., Zhang, F. et al.,“A deep image classification model based on prior feature knowledge embedding and application in medical diagnosis”,Sci Rep 14, 13244 (2024).
31. M. A. Mishra, M. H. Khan, W. Khan, M. Z. Khan, and N. K. Srivastava, “A Comparative Study on Data Mining Approach Using Machine Learning Techniques: Prediction Perspective,” in Pervasive Healthcare, EAI/Springer Innovations in Communication and Computing. Springer, Cham, pp. 153-165, (2022).
32. S. R. Khan and M. W. Khan, “A Comprehensive Review of Privacy-Preserving Techniques in Health Data Analysis,” 2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI), Greater Noida, India, pp. 104–109, (2025).

33. J. A. Behar, L. Bonnemains, V. Shulgin, J. Oster, O. Ostras, and I. Lakhno, "Noninvasive fetal electrocardiography for the detection of fetal arrhythmias."Prenatal Diagnosis, vol. 39, no. 3, pp.178-187, (2019).
34. Sara, N.; Kohei, Y.; Tomoaki, O. "Fetal Arrhythmia Detection based Deep Learning using Fetal ECG Signals". In Proceedings of the IEEE GLOBECOM, Rio de Janeiro, pp. 2266-2271, (2022).