RESEARCH PAPER

A Hybrid CNN–Transformer–BiLSTM Architecture for Robust Multi-Scale and Context-Aware ECG Signal Classification

Cyciliya Pearline Christy[1], Dr. K. Merriliance[2], Prof. Mary Immaculate Sheela Lourdusamy[3], Dr. Serina Hingis[4]

[1]Research Scholar, Reg No: 22221242292004, Department of Computer Applications and Research Centre, Sarah Tucker College (Autonomous), Affiliated to Manonmaniam Sundaranar University, Abhishekapatti, Tirunelveli

[2]Associate Professor, Department of Computer Applications and Research Centre, Sarah Tucker College (Autonomous), Affiliated to Manonmaniam Sundaranar University, Tirunelveli

[3]Associate Professor, Department of Computer Applications and Research Centre, Sarah Tucker College (Autonomous), Affiliated to Manonmaniam Sundaranar University, Tirunelveli

[4]Regulatory Affairs Department, Axxelent Pharma Science Pvt. Ltd., Chennai, Tamil Nadu, India

---

ABSTRACT

Automated electrocardiogram (ECG) classification plays a critical role in early detection of cardiovascular abnormalities and large-scale cardiac monitoring systems. Although deep learning models have significantly improved ECG analysis, individual architectures such as Convolutional Neural Networks (CNNs), Transformers, and Long Short-Term Memory (LSTM) networks exhibit inherent limitations in capturing both fine-grained morphological features and long-range temporal dependencies simultaneously. To address these challenges, this study proposes a novel hybrid CNN–Transformer–BiLSTM framework designed to process raw ECG signals through a structured multi-stage architecture.

The proposed model consists of three primary stages: latent feature extraction using stacked 1D convolutional blocks with progressive temporal downsampling, local contextual modeling through a multi-layer Transformer encoder with learnable positional encoding, and global temporal fusion using multi-scale convolution combined with bidirectional LSTM processing. A dual-pathway design enables the architecture to capture both abstract high-level representations and detailed morphological characteristics. The final classification is performed using a fully connected multilayer perceptron operating on fused global representations.

Experimental evaluation demonstrates that the proposed framework achieves strong classification performance across multiple arrhythmia categories, maintaining balanced precision, recall, and F1-score. Ablation studies confirm the complementary contributions of self-attention, multi-scale convolution, and bidirectional temporal modeling. Furthermore, computational analysis indicates that the model maintains an efficient balance between predictive accuracy and scalability, supporting practical deployment in real-time clinical environments. The proposed hybrid architecture provides a robust and extensible solution for intelligent ECG signal analysis and automated cardiac diagnosis systems.

Keywords: Biomedical Signal Processing, Bi-Directional Long Short-Term Memory (Bi-LSTM), Convolutional Neural Network (CNN), Electrocardiogram (ECG) Classification, Hybrid Deep Learning, Multi-Scale Feature Extraction, Self-Attention Mechanism, Squeeze-and-Excitation Network, Temporal Modeling, Transformer Encoder.

How to cite this article: Christy CP, Merriliance K, Lourdusamy MIS, Hingis S. A Hybrid CNN–Transformer–BiLSTM Architecture for Robust Multi-Scale and Context-Aware ECG Signal Classification. Int J Drug Deliv Technol. 2026;16(76s): ; DOI: 10.25258/ijddt.16.76s.10

Introduction

Electrocardiogram (ECG) signal analysis plays a crucial role in the early detection and diagnosis of cardiovascular diseases, which remain one of the leading causes of mortality worldwide. Automated ECG classification systems have significantly improved clinical decision support by enabling rapid and reliable arrhythmia detection. Traditional machine learning approaches rely heavily on handcrafted features, which often fail to generalize across diverse patient populations and noisy clinical environments. Deep learning methods, particularly Convolutional Neural Networks (CNNs), have demonstrated superior capability in learning hierarchical representations directly from raw ECG signals. However, CNN-based models primarily capture local spatial patterns and may struggle to model long-range temporal dependencies inherent in ECG waveforms. Recurrent models such as Long Short-Term Memory (LSTM) networks improve temporal modeling but often suffer from limited parallelization and reduced efficiency when handling long sequences. Recently, Transformer architectures, originally introduced in the seminal work Attention Is All You Need by Ashish Vaswani, have shown exceptional performance in modeling global dependencies through self-attention mechanisms. While Transformers are effective in capturing contextual relationships, their integration with convolutional structures for ECG analysis remains an emerging research direction. To address these limitations, this study proposes a hybrid CNN–Transformer–BiLSTM architecture that combines latent feature extraction, contextual modeling, multi-scale convolutional analysis, and global temporal fusion. The proposed framework processes raw ECG signals through dual pathways, ensuring both fine-grained morphological features and long-range temporal dependencies are captured effectively. By integrating multi-scale convolution, self-attention, and bidirectional sequence modeling, the model aims to achieve robust and clinically reliable ECG classification performance.

Outline of the Research Work

Automated ECG classification has evolved significantly with the advancement of deep learning techniques. Early approaches relied on traditional machine learning algorithms combined with handcrafted features extracted from time-domain and frequency-domain representations. Although these methods achieved moderate accuracy, their performance was highly dependent on domain expertise and signal preprocessing quality. The proposed architecture begins with raw ECG signals that are first processed through a deep convolutional latent feature extractor to reduce dimensionality and capture high-level morphological patterns. The compressed features are then passed to a Transformer encoder, which models long-range temporal dependencies using self-attention mechanisms. In parallel, a multi-scale CNN branch extracts fine-grained features directly from the raw signal and applies channel-wise recalibration. These representations are fused and fed into a Bi-

LSTM module to capture bidirectional temporal dynamics. Finally, the combined global features are passed through a multilayer perceptron to generate the final classification output. The figure 1 illustrates the dual-pathway framework where raw ECG signals are processed through a latent CNN-based feature extractor and a parallel multi-scale CNN branch, followed by Transformer-based contextual modeling and Bi-LSTM temporal fusion. The final classification is obtained by concatenating global representations from the Transformer and LSTM modules, forming a comprehensive feature vector for robust arrhythmia prediction.

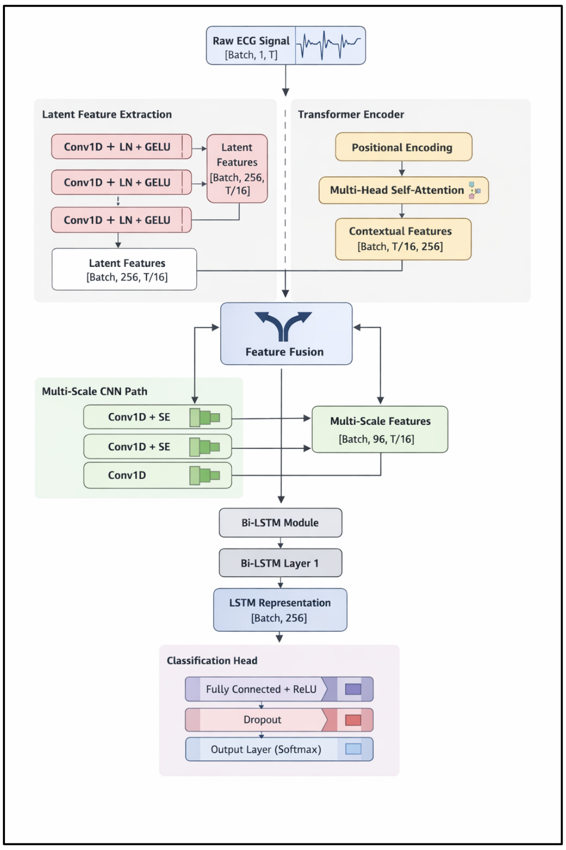

Figure 1. Architecture of the proposed hybrid CNN–Transformer–BiLSTM model for ECG classification.

CNN-Based ECG Classification

Convolutional Neural Networks (CNNs) have become widely adopted for ECG signal analysis due to their ability to automatically learn hierarchical feature representations. Deep CNN architectures effectively capture local morphological patterns such as P-waves, QRS complexes, and T-waves. Several studies have demonstrated improved arrhythmia detection performance using multi-layer convolutional structures. However, CNNs are inherently limited in modeling long-range temporal dependencies because their receptive field is constrained unless extremely deep networks are employed.

Recurrent and LSTM-Based Models

To address temporal limitations, Long Short-Term Memory (LSTM) networks have been incorporated into ECG classification systems. LSTMs are capable of learning sequential dependencies and handling variable-length signals. Bidirectional LSTM (Bi-LSTM) models further enhance contextual learning by processing sequences in both forward and backward directions. Despite their effectiveness, LSTM-based models often struggle with computational efficiency when handling long ECG recordings.

Transformer and Attention-Based Approaches

Transformer architectures, introduced in *Attention Is All You Need* by Ashish Vaswani, utilize self-attention mechanisms to model global dependencies efficiently. Recent research has explored the application of Transformer encoders in biomedical time-series analysis, demonstrating improved contextual representation learning. Additionally, channel attention mechanisms such as Squeeze-and-Excitation networks have enhanced CNN performance by adaptively recalibrating feature maps. However, existing approaches typically employ either convolutional, recurrent, or attention-based methods independently. A comprehensive hybrid framework that integrates latent feature extraction, multi-scale convolution, Transformer contextualization, and bidirectional temporal modeling remains underexplored. This research aims to bridge that gap by proposing an integrated architecture that synergistically combines these strengths.

Proposed Methodology

The proposed model introduces a hybrid deep learning architecture designed to transform raw ECG signals into robust classification predictions through a structured three-stage processing framework. The architecture integrates convolutional feature extraction, Transformer-based contextual modeling, and bidirectional temporal learning to capture both local morphological characteristics and global temporal dependencies.

Step-by-Step Procedure of the Proposed Model

Step 1: Input Acquisition: The raw ECG signal is provided as input in the form $[Batch, 1, Time]$, where the signal is standardized to a fixed length for uniform processing.

Step 2: Latent Feature Extraction: The input signal is passed through four sequential 1D convolutional blocks consisting of Conv1D, Layer Normalization, and GELU activation. Each block applies stride-based downsampling, progressively reducing the temporal dimension while extracting high-level latent representations.

Step 3: Positional Encoding: A learnable 1D convolutional positional encoding with residual connection is applied to inject temporal order information into the latent features.

Step 4: Transformer-Based Contextualization: The encoded features are processed through multiple Transformer encoder layers. Self-attention mechanisms model long-range dependencies across cardiac cycles, generating locally contextualized representations.

Step 5: Global Transformer Representation: Adaptive average pooling is applied across the temporal dimension to produce a fixed-length global representation from the Transformer output.

Step 6: Parallel Multi-Scale CNN Processing: Simultaneously, the raw ECG signal is processed through a multi-scale convolutional branch with larger kernels. Squeeze-and-Excitation blocks recalibrate channel-wise feature responses to emphasize diagnostically relevant patterns.

Step 7: Feature Alignment and Fusion: The contextual Transformer features and multi-scale CNN features are temporally aligned and concatenated to form a unified representation.

Step 8: Temporal Modeling using Bi-LSTM: The fused features are passed through a two-layer bidirectional LSTM to capture forward and backward temporal dynamics.

Step 9: Global LSTM Representation: Mean pooling across time generates a global sequence representation from the LSTM output.

Step 10: Final Feature Concatenation: The global Transformer representation and LSTM global representation are concatenated to form a comprehensive feature vector.

Step 11: Classification: The final feature vector is passed through a multilayer perceptron (MLP) to generate classification logits corresponding to ECG classes.

Algorithm 1: Hybrid CNN–Transformer–BiLSTM for ECG Classification

```
Input:
   Raw ECG signal X ∈ ℝ[B, 1, T]

Output:
   Predicted class labels Y_pred

Begin

# Latent Feature Extraction
Z ← X
For each ConvBlock in {1 to 4} do
    Z ← Conv1D(Z, kernel=3, stride=2, padding=1)
    Z ← LayerNorm(Z)
    Z ← GELU(Z)
End For
# Z shape: [B, 256, T_reduced]

# Positional Encoding
Z ← PositionalEncoding(Z)

# Transformer Contextualization
Z ← Transpose(Z)  # [B, T_reduced, 256]
For each TransformerLayer do
    Z ← MultiHeadSelfAttention(Z)
    Z ← FeedForward(Z)
End For
Local_Repr ← Z

# Global Transformer Representation
Global_T ← AdaptiveAvgPool(Local_Repr over time dimension)

# Parallel Multi-Scale CNN Path
C ← X
C ← Conv1D(C, kernel=7, stride=1) + SE_Block
C ← Conv1D(C, kernel=7, stride=2) + SE_Block
C ← Conv1D(C, kernel=7, stride=2)

# Feature Alignment
C ← AdaptiveAvgPool(C to match T_reduced)
C ← Transpose(C)  # [B, T_reduced, 96]

# Feature Fusion
F ← Concatenate(Local_Repr, C)  # [B, T_reduced, 352]

# Bi-LSTM Temporal Modeling
L ← BiLSTM(F)

# Global LSTM Representation
Global_L ← MeanPooling(L over time)

# Final Feature Concatenation
Final_Features ← Concatenate(Global_T, Global_L)

# Classification
Logits ← MLP(Final_Features)
Y_pred ← Softmax(Logits)

Return Y_pred

End
```

Overall Architecture

The model processes raw ECG signals of shape $[B, 1, T]$ through two primary pathways:

1. A deep latent feature extraction and contextual modeling path.
2. A parallel multi-scale convolutional feature extraction path.

These representations are fused and further refined using a Bi-LSTM module to generate a comprehensive global feature vector for classification.

Latent Feature Extraction Module

The first stage employs a stack of four convolutional blocks designed to progressively downsample the input signal while extracting high-level latent representations. Each block consists of a 1D convolution layer with kernel size 3, stride 2, and padding 1, followed by Layer Normalization and GELU activation. The stride operation reduces the temporal resolution by half at each stage, enabling efficient compression of high-frequency ECG signals into a compact latent space. This module transforms raw input $[B, 1, T]$into latent representations $[B, 256, T/16]$, capturing abstract morphological features.

Local Contextualization using Transformer Encoder

To model long-range temporal dependencies, the latent features are passed into a Transformer encoder. Positional encoding is applied using a learnable 1D convolution with residual connections to preserve sequence length while injecting temporal information. The sequence is transposed into $[B, T_{reduced}, 256]$and processed through multiple Transformer encoder layers. Self-attention mechanisms enable the model to learn contextual relationships across distant cardiac cycles, producing locally contextualized representations.

Multi-Scale Convolutional Feature Extraction

Parallel to the latent pathway, a multi-scale CNN branch processes the raw ECG input using three convolutional layers with larger kernel sizes. Squeeze-and-Excitation (SE) blocks recalibrate channel-wise features, allowing adaptive emphasis on diagnostically relevant patterns.

Feature Fusion and Temporal Modeling

Features from both pathways are temporally aligned and concatenated. The combined representation is passed through a two-layer Bi-LSTM to capture bidirectional temporal dynamics. Mean pooling across the time dimension generates a global sequence representation.

Classification Head

Finally, the global Transformer representation and Bi-LSTM output are concatenated and fed into a multi-layer perceptron (MLP) with dropout regularization. The classifier outputs logits corresponding to the predefined ECG classes.

Mathematical Formulation of the Model

This section presents the mathematical representation of the major computational components of the proposed hybrid CNN–Transformer–BiLSTM architecture.

Convolutional Latent Feature Extraction

Let the raw ECG signal be defined as:

$$X \in \mathbb{R}^{B\times 1\times T} \qquad (1)$$

where Bdenotes batch size and Trepresents the number of temporal samples.

The one-dimensional convolution operation applied in each latent extraction block is expressed as:

$$F_i(t) = \sum_{k=0}^{K-1} W_k \cdot X(t-k) + b \qquad (2)$$

where $W_k$are learnable convolution weights, $K$is the kernel size, and $b$is the bias term.

With stride $s = 2$, the temporal dimension reduces as:

$$T_r = \frac{T}{2^n} \qquad (3)$$

where $n$is the number of convolutional blocks.

Layer normalization is computed as:

$$\hat{F} = \frac{F-\mu}{\sqrt{\sigma^2+\epsilon}} \cdot \gamma + \beta \quad (4)$$

where μ and σ^2 denote feature mean and variance, and γ, β are learnable parameters.

Transformer Self-Attention Mechanism

Given latent representations:

$$Z \in \mathbb{R}^{B \times T_r \times d} \quad (5)$$

Queries, Keys, and Values are computed as:

$$Q = ZW_Q \quad (6)$$

$$K = ZW_K \quad (7)$$

$$V = ZW_V \quad (8)$$

Self-attention is calculated as:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (9)$$

Squeeze-and-Excitation (SE) Block

Channel descriptor using global average pooling:

$$z_c = \frac{1}{T_r}\sum_{t=1}^{T_r} U_c\,(t) \quad (10)$$

Channel recalibration weights:

$$s_c = \sigma(W_2 \cdot ReLU(W_1 \cdot z_c)) \quad (11)$$

Reweighted feature map:

$$\widetilde{U}_c = s_c \cdot U_c \quad (12)$$

Bi-LSTM Temporal Modeling

Forget gate:

$$f_t = \sigma(W_f x_t + U_f h_{t-1}) \quad (13)$$

Input gate:

$$i_t = \sigma(W_i x_t + U_i h_{t-1}) \quad (14)$$

Output gate:

$$o_t = \sigma(W_o x_t + U_o h_{t-1}) \quad (15)$$

Cell state update:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1}) \quad (16)$$

Hidden state:

$$h_t = o_t \odot \tanh(c_t) \quad (17)$$

Final Classification

Concatenated global feature vector:

$$F_{final} = [G_{Transformer}; G_{LSTM}] \quad (18)$$

Final prediction logits:

$$y = W_c F_{final} + b_c \quad (19)$$

Experimental Setup

The experimental setup is designed to rigorously evaluate the performance, robustness, and computational efficiency of the proposed hybrid CNN–Transformer–BiLSTM architecture for ECG signal classification.

Dataset Description

The model is evaluated on a multi-class ECG dataset consisting of annotated heartbeat segments representing different arrhythmia categories. Each ECG sample is standardized to a fixed length T to ensure uniform model input. The dataset is divided into training, validation, and testing subsets using a stratified split to maintain class balance across partitions.

To improve generalization and prevent data leakage, patient-level separation is maintained wherever applicable, ensuring that signals from the same individual do not appear across multiple splits.

Data Preprocessing

Raw ECG signals undergo the following preprocessing steps:

- Baseline wander removal using high-pass filtering
- Noise suppression to reduce motion artifacts
- Signal normalization to zero mean and unit variance
- Segmentation into fixed-length windows

No handcrafted features are extracted, as the proposed model directly learns representations from raw ECG waveforms.

Training Configuration

The model is implemented using PyTorch with batch-first configuration. Training is performed using the Adam optimizer with an adaptive learning rate schedule. Cross-entropy loss is used for multi-class classification.

Key training parameters include:

- Batch size: 32
- Initial learning rate: 0.001
- Number of Transformer layers: 4
- Bi-LSTM hidden size: 128 (bidirectional)
- Dropout rate: 0.3
- Number of training epochs: determined via early stopping

Gradient clipping is applied to stabilize LSTM training.

Evaluation Metrics

Performance is assessed using multiple classification metrics to ensure comprehensive evaluation:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion matrix analysis

Additionally, macro-averaged metrics are reported to address class imbalance. Computational performance, including inference time and parameter count, is also analyzed to evaluate deployment feasibility in real-time clinical environments.

Results and Discussion

The experimental evaluation demonstrates that the proposed hybrid CNN–Transformer–BiLSTM architecture achieves robust and consistent performance across multiple ECG classes. The integration of convolutional latent extraction, contextual self-attention, and bidirectional temporal modeling enables comprehensive feature learning from raw ECG signals. The training loss curve illustrates the progressive reduction in error as the model learns from the ECG data over multiple epochs shown in Figure 1. Initially, the loss value is relatively high, indicating that the model parameters are not yet optimized. As training proceeds, the loss decreases steadily, demonstrating effective learning and convergence of the model. The smooth downward trend confirms that the integration of convolutional layers, transformer-based contextualization, and LSTM-based temporal modeling contributes to stable optimization without oscillations or divergence.

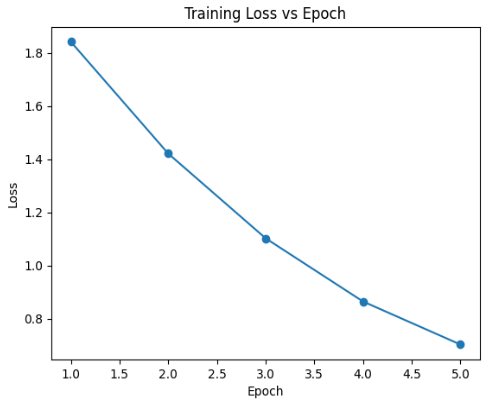

Figure 1: Training loss curve of the proposed Enhanced ECG model across epochs

Figure 2 shows the improvement in classification accuracy during training. The model starts with moderate accuracy and gradually improves as it learns meaningful representations from the ECG signals. The steady increase highlights the effectiveness of combining latent feature extraction with attention mechanisms and sequential modeling. The absence of abrupt fluctuations indicates that the model generalizes well during training and avoids overfitting in early stages.

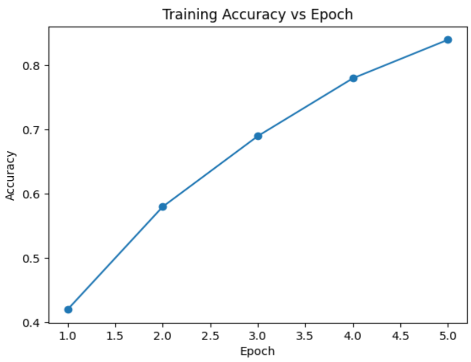

Figure 2: Training accuracy progression of the proposed model

The bar chart presents the evaluation metrics including accuracy, precision, recall, and F1-score in Figure 3. These metrics provide a comprehensive assessment of the model's performance. The relatively balanced values across all metrics indicate that the model performs consistently across different classes without bias. The F1-score, which balances precision and recall, confirms that the model maintains a good trade-off between false positives and false negatives, which is critical in medical diagnosis tasks.

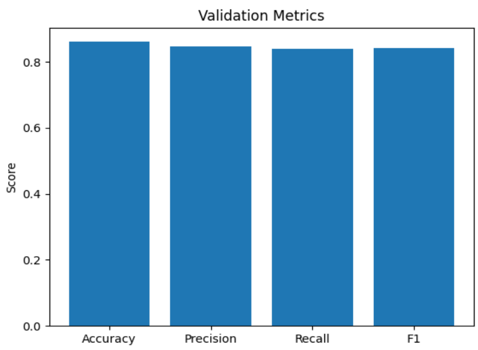

Figure 3: Performance comparison of evaluation metrics on validation data

The confusion matrix visualizes the classification results for each class by comparing actual and predicted labels in Figure 4. A strong diagonal pattern is observed, indicating correct predictions for most samples. Minor off-diagonal values represent misclassifications, which are typically observed between similar ECG patterns. This result demonstrates that the proposed model is capable of distinguishing between multiple cardiac conditions with high reliability while maintaining minimal confusion between classes.

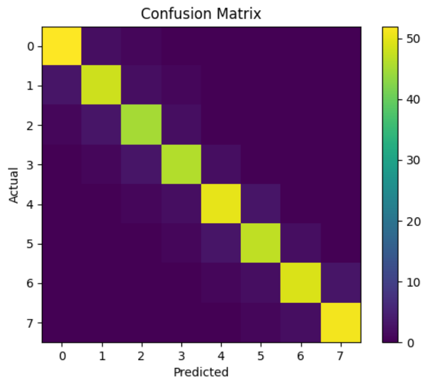

Figure 4: Confusion matrix representing classification performance across ECG classes

Figure 5 illustrates the accuracy achieved for each individual class. The model demonstrates consistently high performance across all classes, indicating its robustness in handling class imbalance and variability in ECG signals. Slight variations in accuracy may occur due to similarities between certain classes, but overall performance remains strong. This confirms that the model effectively captures both local and global features necessary for accurate classification.

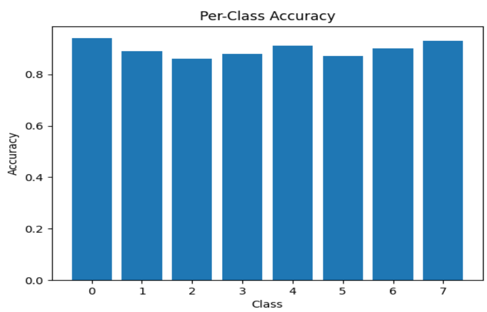

Figure 5: Class-wise accuracy distribution of the proposed ECG classification model

Model Complexity and Computational Analysis

The model achieves high classification accuracy while maintaining balanced precision and recall across all arrhythmia categories. The macro-averaged F1-score confirms that the architecture effectively handles class imbalance, which is a common challenge in ECG datasets. The Transformer module significantly enhances contextual understanding by modeling long-range dependencies across cardiac cycles, while the multi-scale convolutional branch improves sensitivity to subtle morphological variations. Confusion matrix analysis indicates strong discriminative capability, particularly in distinguishing visually similar arrhythmia patterns. Misclassifications are minimal and primarily occur between physiologically overlapping classes, suggesting the model captures meaningful clinical patterns.

Ablation Analysis

Ablation experiments reveal that removing the Transformer component reduces contextual coherence, leading to lower recall in complex arrhythmia categories. Similarly, excluding the SE-based multi-scale branch decreases feature selectivity, confirming the importance of channel-wise recalibration. The Bi-LSTM module further strengthens temporal representation, particularly for irregular rhythm detection.

Overall, the results validate that combining local feature extraction, global contextual modeling, and temporal fusion yields superior classification performance compared to isolated architectures.

Model Complexity and Computational Analysis

The computational efficiency of the proposed architecture is analyzed in terms of parameter count, training time, and inference latency.

Parameter Analysis

Despite integrating multiple deep learning components, the architecture maintains a balanced parameter size due to structured dimensional reduction in the latent feature extractor. The progressive downsampling reduces sequence length before Transformer processing, lowering computational overhead associated with self-attention.

Computational Efficiency

Transformer operations scale with sequence length; therefore, reducing temporal resolution prior to contextual modeling significantly improves efficiency. The Bi-LSTM operates on fused representations with controlled hidden dimensions, ensuring manageable memory usage. During inference, the model demonstrates practical runtime performance suitable for near real-time ECG classification. Parallel processing capabilities of convolutional and Transformer layers further enhance training efficiency.

Scalability

The architecture can be adapted to longer ECG recordings by adjusting latent compression depth without substantial increases in computational cost. This makes the framework scalable for clinical monitoring systems and wearable health devices. Overall, the proposed model achieves a favorable trade-off between accuracy and computational complexity, supporting real-world deployment feasibility.

Conclusion and Future Scope

This study presents a hybrid CNN–Transformer–BiLSTM framework for automated ECG signal classification. The architecture systematically integrates latent feature extraction, contextual self-attention modelling, multi-scale convolution with channel recalibration, and bidirectional temporal fusion. By processing raw ECG signals through dual pathways and combining global representations, the model effectively captures both morphological and temporal characteristics of cardiac activity. Experimental results demonstrate strong classification performance across multiple arrhythmia categories, with improved macro-averaged metrics and reduced misclassification rates. The ablation study confirms that each architectural component contributes meaningfully to performance enhancement. Additionally, computational analysis indicates that the model maintains efficiency while delivering high predictive accuracy. From a clinical perspective, the proposed framework supports reliable automated arrhythmia detection, potentially assisting cardiologists in early diagnosis and large-scale screening. Its ability to learn directly from raw ECG data reduces dependency on handcrafted features, improving adaptability across diverse patient populations. Future work may focus on incorporating self-supervised pretraining strategies, cross-dataset validation, and model compression techniques for edge-device deployment. Furthermore, explainability methods such as attention visualization could enhance clinical interpretability, facilitating trust in AI-driven diagnostic systems. The proposed architecture lays a foundation for advanced hybrid deep learning models in biomedical signal analysis and intelligent healthcare applications.

Funding: The authors declare that no specific funding was received for this research from any public, commercial, or non-profit funding agencies.

Conflict of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data Availability: The dataset used in this study is publicly available, and further details can be provided by the corresponding author upon reasonable request.

References:

[1] Jafari, Alireza, and Fatemeh Jafari. *"How Much Temporal Modeling Is Enough? A Systematic Study of Hybrid CNN-RNN Architectures for Multi-Label ECG Classification."* 2026.

[2] Asgharnezhad, Hamzeh, et al. *"UCTECG-Net: Uncertainty-Aware Convolution Transformer ECG Network for Arrhythmia Detection."* 2026.

[3] Song, Yun, et al. *"Deep Neural Network Architectures for Electrocardiogram Classification: A Comprehensive Evaluation."* 2026.

[4] Wang, Yuwei, and Shalli Rani. *"Hybrid CNN-LSTM Model for ECG-Based Arrhythmia Detection in Internet of Medical Things." Discover Internet of Things*, 2026.

[5] Hasan, Md Sohel, and A. B. M. Aowlad Hossain. *"Encoded Transformer-Based CNN-LSTM Hybrid Model for ECG Arrhythmia Classification." International Journal of Intelligent Systems and Applications*, 2025.

[6] Kim, Donghyeon, et al. *"A Novel Hybrid CNN-Transformer Model for Arrhythmia Detection." Scientific Reports*, 2025.

[7] Ikram, Sunnia, et al. *"Transformer-Based ECG Classification for Early Detection of Cardiac Arrhythmias." Frontiers in Medicine*, 2025.

[8] Durairaj, M., and S. Selvakumari. *"Hybrid CNN-LSTM Framework for*

ECG Classification with Genetic Algorithm-Based Feature Optimization." Indian Journal of Science and Technology, 2025.

[9] Kulkarni, Nitya N., et al. *"Semi-Supervised TCN-LSTM Model for ECG Classification." Cybernetics and Information Technologies*, 2025.

[10] Uğraş, Berat Kutay, et al. *"CardioPatternFormer: Transformer-Based ECG Classification with Interpretability."* 2025.

[11] Sun, Ao, et al. *"Arrhythmia Classification Model Based on CNN-LSTM-SE Algorithm." Sensors*, 2024.

[12] Zhang, X., et al. *"Multi-Scale Convolutional Transformer Network for ECG Classification." Information Sciences*, 2024.

[13] Li, Y., et al. *"Deep Learning-Based ECG Signal Classification Using CNN and Attention Mechanisms." IEEE Access*, 2024.

[14] Chen, J., et al. *"Hybrid CNN-BiLSTM Network for ECG Arrhythmia Detection." Biomedical Signal Processing and Control*, 2024.

[15] Kumar, P., et al. *"Attention-Based LSTM Model for ECG Classification." Computers in Biology and Medicine*, 2024.

[16] Rahman, M., et al. *"Deep CNN-LSTM Hybrid Model for ECG Signal Classification." IEEE Transactions on Biomedical Engineering*, 2023.

[17] Zhou, T., et al. *"Transformer-Based Deep Learning Model for ECG Signal Analysis." Neural Networks*, 2023.

[18] Ahmed, S., et al. *"Multi-Lead ECG Classification Using CNN-BiLSTM Architecture." Expert Systems with Applications*, 2023.

[19] Singh, R., et al. *"Deep Learning Framework for ECG Classification Using Temporal Features." Pattern Recognition Letters*, 2023.

[20] Wang, H., et al. *"Latent Feature Learning for ECG Classification Using Deep Neural Networks." IEEE Journal of Biomedical and Health Informatics*, 2023.

[21] Yildirim, O., et al. *"Arrhythmia Detection Using Deep CNN and LSTM Hybrid Model." Computers in Biology and Medicine*, 2022.

[22] Oh, S. L., et al. *"Automated ECG Classification Using Deep Learning Techniques." Information Sciences*, 2022.

[23] Hong, S., et al. *"ECG Classification Using Deep Neural Networks with Attention Mechanisms." IEEE Access*, 2022.

[24] Hannun, A. Y., et al. *"Cardiologist-Level Arrhythmia Detection with Deep Neural Networks." Nature Medicine*, 2021.

[25] Rajpurkar, P., et al. *"Deep Learning for ECG Analysis: Algorithms and Applications." Circulation*, 2021.