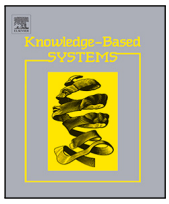
Knowledge-Based Systems 315 (2025) 113301

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



Advanced deep learning framework for ECG arrhythmia classification using 1D-CNN with attention mechanism

Mohammed Guhdar *,[1], Abdulhakeem O. Mohammed[1], Ramadhan J. Mstafa[1]

Computer Science Department, College of Science, University of Zakho, Kurdistan Region, Iraq

ARTICLE INFO

Keywords:
Electrocardiogram
Arrhythmia classification
Deep learning
Attention mechanism
MIT-BIH dataset

ABSTRACT

Cardiovascular diseases, particularly cardiac arrhythmias, remain a leading cause of global mortality, necessitating efficient and accurate diagnostic tools. Despite advances in deep learning for ECG analysis, current models face challenges in cross-population performance, signal noise robustness, limited training data efficiency, and clinical result interpretability. Additionally, most current approaches struggle to generalize across different ECG databases and require extensive computational resources for real-time analysis. This paper presents a novel hybrid deep learning framework for automated ECG analysis, combining one-dimensional convolutional neural networks (1D-CNN) with a specialized attention mechanism. The proposed architecture implements a four-stage CNN backbone enhanced with a squeeze-and-excitation attention block, enabling adaptive feature selection across multiple scales. The model incorporates advanced regularization techniques, including focal loss, L2 regularization, and an ensemble approach with mixed precision training. We conducted extensive experiments across multiple datasets to evaluate generalization capabilities. This study utilizes two standard databases: the MIT-BIH Arrhythmia Database (48 half-hour recordings sampled at 360 Hz) and the PTB Diagnostic ECG Database (549 records from 290 subjects sampled at 1000 Hz). Through rigorous validation including five-fold cross-validation and statistical significance testing, our model attained remarkable performance, achieving 99.48% accuracy on MIT-BIH, 99.83% accuracy on PTB, and 99.64% accuracy on the combined dataset, with corresponding F1-scores of 0.99, 1.00, and 1.00 respectively. The findings demonstrate robust generalization across varied ECG morphologies and recording conditions, with particular effectiveness in handling class imbalance without data augmentation. The model's reliable performance across multiple datasets indicates significant potential for clinical applications in automated cardiac diagnostics.

1. Introduction

Cardiovascular diseases (CVDs) are a major global health issue, resulting in almost 17.9 million fatalities per year, as reported by the World Health Organization [1]. Among these, cardiac arrhythmias – irregular heartbeat patterns – constitute a substantial portion of cardiac-related mortality. The electrocardiogram (ECG) serves as the primary diagnostic tool for detecting and analyzing these irregularities, making accurate interpretation of ECG signals crucial for effective patient care and treatment [2]. The economic impact of CVDs is equally concerning, with global costs expected to exceed 1.1 trillion by 2025 [3].

The complexity of ECG interpretation lies in the subtle variations that distinguish different types of arrhythmias, requiring extensive expertise and time from healthcare professionals. Traditional manual analysis of ECG recordings is time-consuming and prone to human error, particularly in long-term monitoring scenarios [4]. Studies have shown that even experienced cardiologists can have interpretation discrepancies and median accuracy was 54% of ECG readings [5]. Moreover, the increasing prevalence of wearable ECG devices has generated unprecedented volumes of data, making manual analysis increasingly impractical [6].

The healthcare sector has seen a substantial transformation due to the incorporation of artificial intelligence (AI) in cardiac care. Recent breakthroughs in deep learning have demonstrated significant potential in automating ECG interpretation, especially in the detection and classification of arrhythmias. The integration of convolutional neural networks (CNNs) with attention mechanisms has shown particular promise in capturing both local and global patterns within ECG signals,

* Corresponding author.
E-mail addresses: mohammed.guhdar@uoz.edu.krd (M. Guhdar), a.mohammed@uoz.edu.krd (A.O. Mohammed), ramadhan.mstafa@uoz.edu.krd (R.J. Mstafa).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.knosys.2025.113301
Received 8 December 2024; Received in revised form 3 March 2025; Accepted 4 March 2025
Available online 14 March 2025

though existing implementations face challenges in balancing computational efficiency with accuracy [7]. These approaches have evolved to address the unique challenges posed by ECG signal analysis, including noise sensitivity, inter-patient variability, and the temporal nature of cardiac signals [8].

The advancement of ECG monitoring technologies has presented new obstacles and opportunities. Continuous monitoring devices now produce substantial quantities of data, expected to exceed 100 GB of data for per patient each year [9]. This data explosion necessitates robust automated analysis systems that can process information in real-time while maintaining high accuracy. Furthermore, the integration of ECG monitoring with electronic health records has created opportunities for longitudinal patient analysis and early warning systems [10].

The development of automated ECG analysis systems has gained increased urgency due to the growing popularity of remote patient monitoring systems and telemedicine has transformed healthcare delivery. The necessity for swift and precise diagnosis in emergency medical scenarios continues to shape technological advancement in the field. The growing utilization of wearable ECG devices producing continuous monitoring data has generated unprecedented volumes of cardiac data. The deficiency of competent healthcare practitioners in many regions worldwide presents ongoing challenges. The increasing demand for preventive cardiac care and early intervention measures remains a key driver [11]. The necessity for uniform analytical methodologies across healthcare organizations further emphasizes the need for standardization [12].

Current challenges in automated ECG analysis can be systematically categorized into key areas affecting clinical adoption and performance. These challenges include significant variations in model accuracy across different databases and demographic groups, a heavy reliance on data augmentation techniques to handle uneven class distributions, difficulties in meeting real-time processing requirements, and a lack of clear explanation mechanisms for model decisions, that hinders clinical adoption.

This research presents a robust and generalizable deep learning framework for automated ECG arrhythmia classification that maintains consistent performance across diverse databases while effectively handling class imbalance without the need for data augmentation. Our framework introduces several key architectural innovations: a multi-stage CNN backbone (64–512 filters) optimized for ECG feature extraction, a specialized squeeze-and-excitation attention mechanism for adaptive feature selection, and focal loss optimization ($\alpha = 1, \gamma = 2$) for handling class imbalance. The framework incorporates mixed-precision training with an ensemble strategy to ensure robust cross-database performance.

The main contributions of this study are:

- A novel hybrid architecture integrating multi-stage CNN layers with attention mechanisms, optimized for comprehensive ECG feature extraction.
- An advanced ensemble strategy combining mixed-precision training and focal loss optimization for robust handling of class imbalance.
- Comprehensive validation across the MIT-BIH Arrhythmia and PTB Diagnostic databases, demonstrating strong cross-database generalization.

2. Related works

Recent years have witnessed significant advancements in automated ECG analysis, particularly in the application of deep learning techniques. This section presents a comprehensive review of recent developments in ECG signal processing and arrhythmia classification.

Alquran et al. [13] introduced spectral analysis techniques combined with deep learning, achieving 97.8% accuracy. Their approach incorporated frequency domain features alongside temporal analysis, providing a more comprehensive view of ECG characteristics. The study demonstrated the value of multi-domain analysis, although the complexity of feature fusion posed implementation challenges.

Romero et al. [14] proposed a noise-resistant ECG classification system utilizing adaptive filtering and deep learning techniques. Their model attained high accuracy on several metrics and noisy ECG signals, illustrating its robustness against diverse forms of interference. The study emphasized the significance of rigorous preparation in practical applications, but the supplementary preprocessing utilizes raised computational complexity.

Zeinalipour et al. [15] explored the application of graph neural networks (GNNs) to ECG analysis, achieving 99.38% accuracy on PTB Diagnostics dataset [16]. Their approach modeled ECG signals as graphs, enabling better capture of inter-lead relationships in multi-lead ECG recordings. While innovative, the approach required significant computational resources for graph construction and processing.

Sadad et al. [17] developed a lightweight attention mechanism specifically designed for ECG signal processing, achieving 98.39% accuracy while maintaining computational efficiency. Their work demonstrated the potential for optimizing attention mechanisms for specific domains, though generalization to other datasets remained challenging.

Gu et al. [18] proposed an efficient CNN architecture tailored for resource-limited devices. Their model attained 97.69% accuracy on the MIT-BIH arrhythmia dataset [19], while preserving minimal computational cost, rendering it appropriate for implementation on wearable devices. The research illustrated the efficacy of optimized neural network topologies in practical applications, but with a minor reduction in accuracy for advantage of efficiency.

Venkatesh et al. [20] addressed the challenging task of discriminating co-occurring atrial arrhythmias by developing an ensemble model combining 1D-CNN and BiLSTM architectures. Their approach, validated on the CUSPH database using 10-fold cross-validation, achieved 94% accuracy in classifying five distinct cardiac rhythms (AT, AF, AFL, ST, and SR).

Islam et al. [21] proposed a transformer-based methodology for ECG categorization. Their model employed self-attention mechanisms to capture long-range dependencies in ECG signals, attaining 99.14% overall accuracy and 94.69% F1 accuracy on 5-class arrhythmia classification in the MIT-BIH dataset.

Song et al. [22] developed a hybrid deep learning model combining bidirectional LSTM networks with attention mechanisms, achieving 97.1% accuracy for ECG recognition. Their approach particularly excelled at identifying complex arrhythmia patterns, though the model's complexity posed challenges for real-time applications.

Tao et al. [23] focused on interpretable deep learning models for ECG classification. Their approach incorporated attention visualization techniques, achieving 0.932% F1 score, while providing insights into the model's decision-making process.

Zishan et al. [24] demonstrated cost-effective ECG monitoring systems using compact dense neural network architecture. Their implementation on Arduino Nano achieved 96.38% accuracy across four arrhythmia classes while requiring only 1.267 KB of memory.

Subba et al. [25] conducted a comparative analysis of various deep learning architectures, evaluating different approaches on multiple ECG datasets. Their study provided valuable insights into the strengths and limitations of various architectures, with random forest methods achieving the highest accuracy of 98

Qu et al. [26] proposed HQ-DCGAN, a hybrid quantum deep convolutional generative adversarial network addressing ECG data class imbalance. Their approach combines quantum convolutional layers with parameterized quantum circuits while maintaining classical CNN advantages. They introduced 1D Fréchet Inception Distance (1DFID) as a novel evaluation metric for generated ECG signals. The generated signals achieved 82.2% average classification accuracy, outperform-

ing baselines. Importantly, HQ-DCGAN remains compatible with noisy intermediate-scale quantum computers while improving stability.

Ge et al. [27] developed the ECG-KG framework to address interpretability in ECG diagnostic models by emulating clinical reasoning patterns. Their three-module system extracts semantic features using difference value methods, constructs a knowledge graph with abnormal ECG triples, and integrates this representation with graph convolutional networks. Unlike black-box approaches, ECG-KG makes the decision process explicit to clinicians. Experiments across multiple datasets (QT, CPSC-2018, and ZZU-ECG) demonstrated superior performance compared to existing diagnostic models.

Fan et al. [28] introduced AOCBLS, addressing ECG arrhythmia detection challenges in continuous unlabeled data streams with high labeling costs and class imbalances. Their model employs a two-stage active learning strategy to select valuable beats for labeling, reducing annotation requirements. A stable weight update rule manages dynamic imbalances, while an incremental learning algorithm eliminates retraining needs. Experiments on the MIT-BIH database achieved 99.10% accuracy and 96.35% G-mean using only 10% additional annotation.

Zhang et al. [29] proposed a Multi-Scale Feature-based Transformer (MSFT) model incorporating innovative biased relative position encoding within a probsparse multi-head attention structure. Their model achieved exceptional performance with 99.40% macro-averaged accuracy and F1 score on the MIT-BIH arrhythmia database.

Jiang et al. [30] presented a significant advancement in 12-lead ECG analysis, addressing the limitations of traditional attention mechanisms through a dual-branch convolutional neural network. This architecture demonstrated superior efficiency with reduced model parameters, achieving an F1-score of 0.905 and macro AUC of 0.936 on large-scale datasets.

Recent advances in ECG analysis demonstrate diverse methodological tradeoffs. Deep learning methods, particularly transformer-based and ensemble approaches [20–22], excel at capturing complex patterns and achieving high accuracies (97%–99%), but require substantial computational resources as they combine multiple deep learning architectures and demand significant memory usage for model parameters. While lightweight architectures [18,24] offer efficient solutions for resource-constrained environments such as wearable devices and edge computing platforms, they often sacrifice performance for portability and reduced memory footprint.

Graph-based approaches [15] show promise in capturing complex signal relationships by transforming 1D ECG signals into 2D graph representations, enabling better modeling of temporal dependencies and inter-lead relationships in multi-lead recordings. However, they face implementation challenges in real-world settings due to computational overhead from graph construction and complex architecture requirements. Meanwhile, attention-based methods [17,29] have gained significant traction due to their ability to highlight relevant signal features, though traditional implementations may require careful optimization to balance computational efficiency and performance.

Most existing approaches demonstrate limitations in practical deployment. While many methods report impressive accuracies on individual databases [13,14,25], critical analyses by [31,32] have revealed significant challenges in cross-database generalization and demographic variations, particularly when models are tested on diverse patient populations. Common practices of single-database validation and pre-split oversampling can lead to overfitting and unrealistic performance metrics, while computational demands pose challenges for real-time processing in clinical settings where immediate diagnosis is crucial. These limitations underscore the need for developing robust, generalizable, and computationally efficient approaches that maintain consistent performance across diverse clinical environments and patient populations while ensuring reliable real-time processing capabilities.

Table 1
MIT-BIH Arrhythmia Database characteristics.

Characteristic	Description
Number of Records	48
Recording Duration	30 min each
Sampling Frequency	360 Hz
Resolution	11-bit
Voltage Range	10 mV
Number of Channels	2 (Modified Lead II and V1)
Total Segments	21,890
Number of Classes	5

Table 2
PTB Diagnostic ECG Database characteristics.

Characteristic	Description
Number of Records	549
Number of Subjects	290
Sampling Frequency	1000 Hz
Resolution	16-bit
Number of Channels	12 standard + 3 Frank leads
Signal Duration	Variable (30 s)
Number of Classes	2 (Normal/Abnormal)
Total Segments	14,552

3. Materials and methods

In this work, we adopted state-of-the-art parameters and utilized well-established datasets to ensure the reliability and validity of our analysis. By adhering to these standardized methodologies, we aimed to facilitate comparability with existing research and enhance the robustness of our findings. This approach allowed us to build on the current body of knowledge with precision and rigor.

3.1. Datasets

This research utilizes two widely recognized databases: the MIT-BIH Arrhythmia Database and the PTB Diagnostic ECG Database, both available through PhysioNet [33].

3.1.1. MIT-BIH arrhythmia database

The MIT-BIH Arrhythmia Database consists of 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979 [19]. The recordings represent a deliberate mix of complex ventricular, junctional, and supraventricular arrhythmias and conduction abnormalities and Fig. 1 shows three samples from each class.

The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. The database includes 109,494 annotated beats, with annotations for both timing information and beat class. Twenty-three recordings were chosen at random from a set of 4,000 24-hour ambulatory ECG recordings, while the remaining 25 recordings were selected to include less common but clinically significant arrhythmias [19,33] (see Tables 1–3).

3.1.2. PTB diagnostic ECG database

The PTB Diagnostic ECG Database contains 549 records from 290 subjects provided by the National Metrology Institute of Germany [16]. Sampled at 1000 Hz with 16-bit resolution, this database includes 15-lead recordings (12 conventional leads plus 3 Frank leads) representing various cardiac conditions and normal controls.

3.2. Signal preprocessing

The preprocessing pipeline implements a multi-stage approach for signal enhancement, with each stage optimized for specific noise characteristics in ECG signals. Fig. 2 demonstrates the progressive improvement in signal quality through our pipeline, from raw noisy input to clean, normalized output suitable for classification.

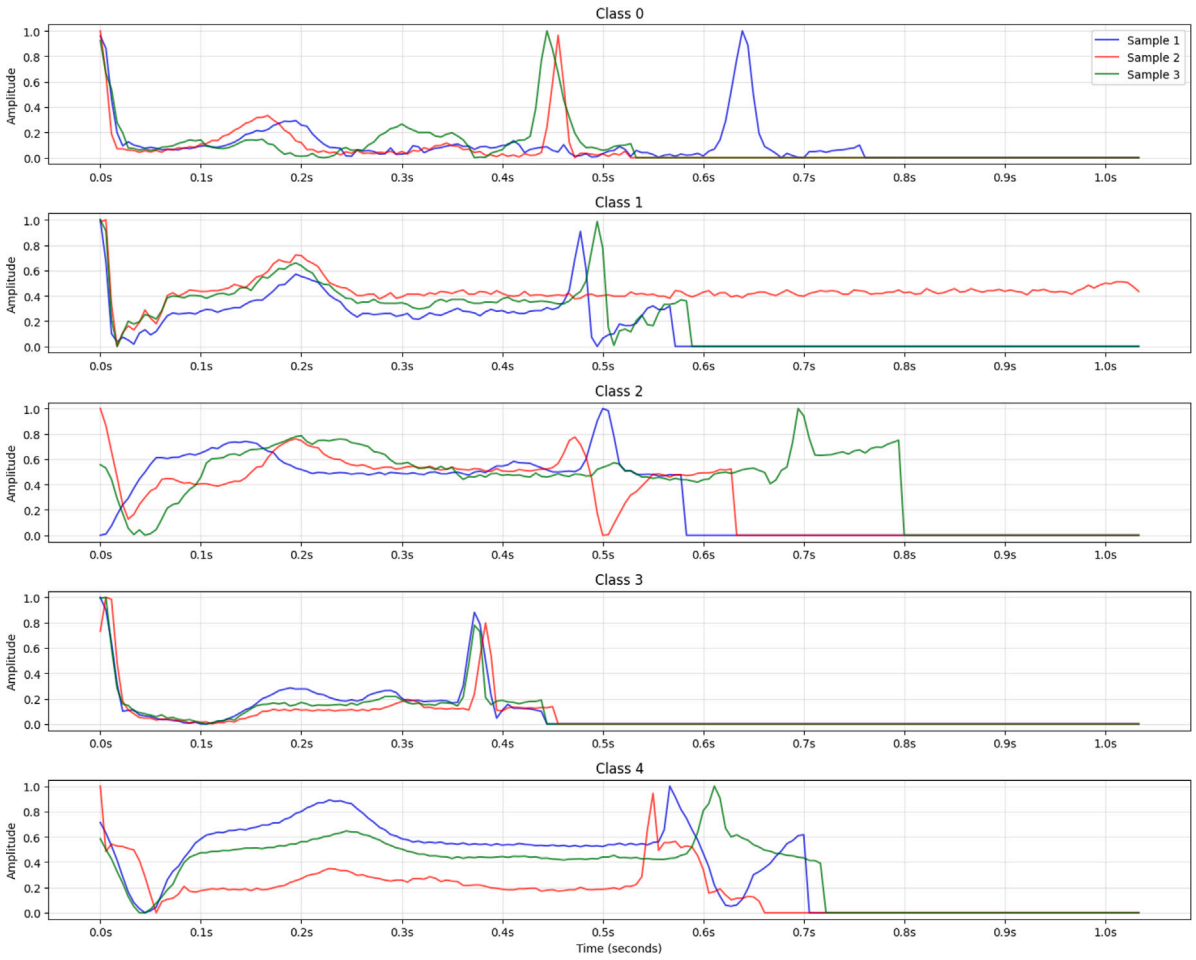


Fig. 1. Three samples from each of five classes in MIT-BIH dataset.

Table 3
Dataset distribution and experimental results.

Dataset/Class	Samples
MIT-BIH Dataset	
Normal (N)	75,052
Left Bundle Branch Block (L)	8075
Right Bundle Branch Block (R)	7259
Premature Ventricular Contraction (V)	7130
Atrial Premature Beat (A)	2546
Paced Beat (/)	7028
Other	1558
PTB Dataset	
Normal	4046
Abnormal	10,506
Combined Dataset	
Total Segments	123,200

3.2.1. Wavelet denoising

The first stage applies wavelet denoising using the Daubechies 4 (db4) wavelet transform with the following steps:

$$\text{threshold} = \frac{\sigma(\text{coeffs}_{\text{final}})}{2} \tag{1}$$

Where σ represents the standard deviation operator, coeffs_final represents the final wavelet coefficients, and 2 represents the scaling divisor.

This adaptive thresholding approach, based on the work of Donoho and Johnstone [34], provides robust noise reduction while preserving important signal characteristics. The db4 wavelet was chosen for its optimal balance between smoothness and feature preservation in ECG signals [35].

3.2.2. Baseline correction

Baseline wander is removed using a median filter:

$$x_{\text{corrected}} = x_{\text{denoised}} - \text{medfilt}(x_{\text{denoised}}, k = 51) \tag{2}$$

Where $x_{\text{corrected}}$ is corrected signal after preprocessing, x_{denoised} is denoised version of the original signal, $\text{medfilt}(x_{\text{denoised}}, k = 51)$ is result of applying a median filter to x_{denoised} with a kernel size $k = 51$.

The median filter with kernel size 51 effectively removes low-frequency baseline drift while maintaining signal morphology, a technique validated by extensive research in ECG signal processing [36]. This approach is particularly effective for long-term ECG recordings where baseline drift can significantly impact signal quality.

3.2.3. Normalization

Signal amplitudes are normalized using z-score normalization with outlier handling:

$$x_{\text{norm}} = \text{clip}\left(\frac{x_{\text{corrected}} - \mu}{\sigma + \epsilon}, -5, 5\right) \tag{3}$$

Where $x_{\text{corrected}}$ is corrected signal after preprocessing, μ is the mean of the signal, σ is the standard deviation, $\epsilon = 10^{-8}$ prevents division by zero, and clip function limits values to the range $[-5, 5]$. This robust normalization technique, adapted from modern deep learning practices [37], ensures consistent signal scaling while mitigating the impact of outliers.

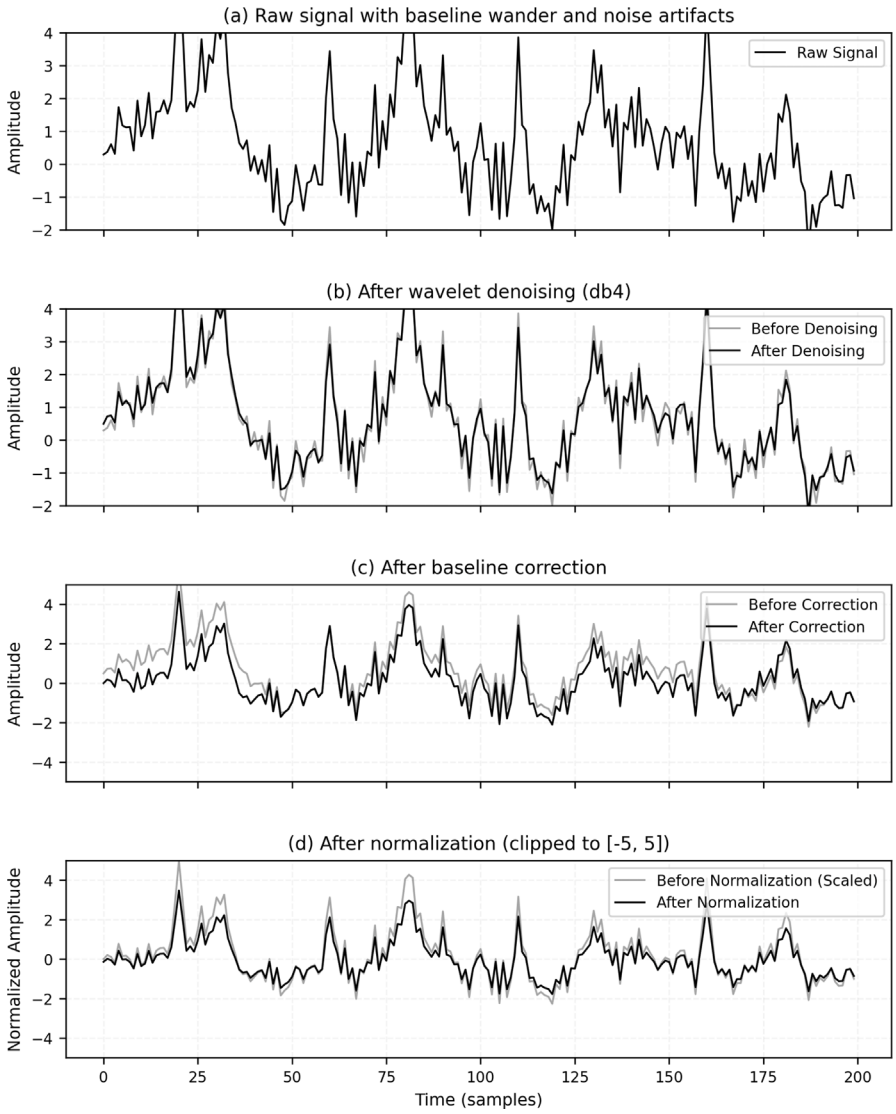


Fig. 2. ECG signal preprocessing stages: (a) Raw signal with baseline wander and noise artifacts (b) After wavelet denoising, showing preserved QRS morphology while removing high-frequency noise (c) After baseline correction, demonstrating stabilized baseline (d) Final normalized signal ready for classification. Each stage preserves diagnostically relevant features while reducing specific types of noise.

4. Proposed method

We propose a hybrid deep learning architecture that integrates convolutional neural networks with a squeeze-and-excitation attention mechanism, specifically optimized for ECG signal analysis. Our framework adaptively emphasizes relevant features while suppressing less informative ones through selective feature refinement. The architecture's comprehensive feature extraction capabilities are illustrated in Fig. 4 and squeeze and excitation illustrated in Fig. 5, which demonstrates the flow and operation of the attention mechanism.

4.1. Model architecture

The architecture consists of a multi-stage CNN backbone coupled with our specialized attention mechanism. Fig. 3 presents the overall network topology, showing the progression of feature extraction through convolutional layers and the integration of attention-based feature refinement. This design enables effective capture of both the local morphological characteristics of the ECG and the global signal patterns.

4.1.1. Attention mechanism

The attention block enhances the model's ability to focus on the most discriminative features of the ECG signal through a learnable weighting mechanism [38]. Our implementation follows a squeeze-and-excitation pattern, which has shown significant success in cardiac signal processing [39].

The Squeeze-Excitation (SE) block takes a 512-channel feature tensor and passes it through AdaptiveAvgPool1d(1) for global spatial information. This 512-dimensional vector then flows through a sequence of nn.Linear(512, 64), ReLU, nn.Linear(64, 512), and Sigmoid layers to generate channel-wise attention weights. Finally, these generated weights multiply with the input features ($x * attention_weights$) to produce the output, where each of the 512 channels has been dynamically recalibrated based on their learned importance as shown in Fig. 6.

The attention mechanism in our model follows a sequence of transformations that mirror our PyTorch implementation. First, dimension reduction through the first linear layer, where the input feature vector x is transformed to a lower-dimensional space, capturing key feature

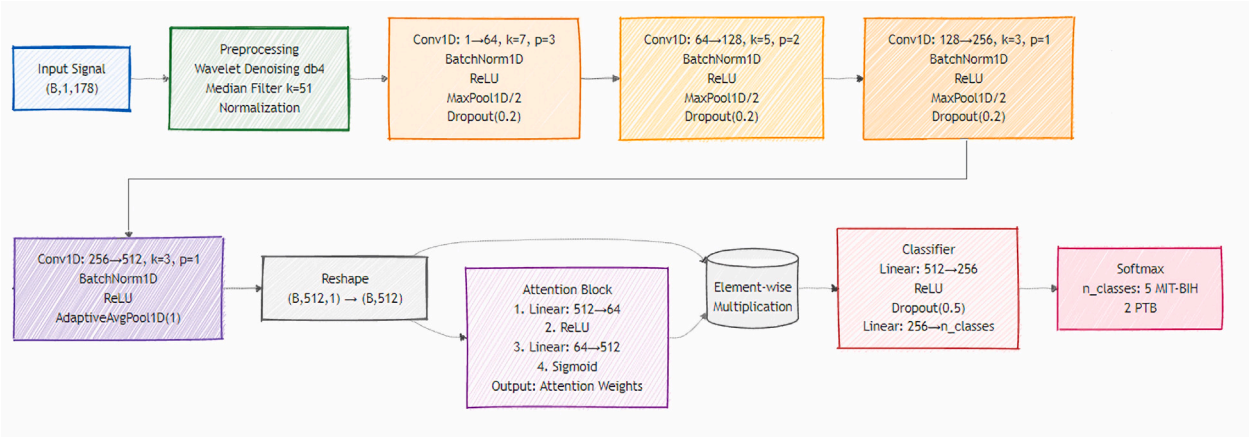


Fig. 3. Proposed model flowchart.

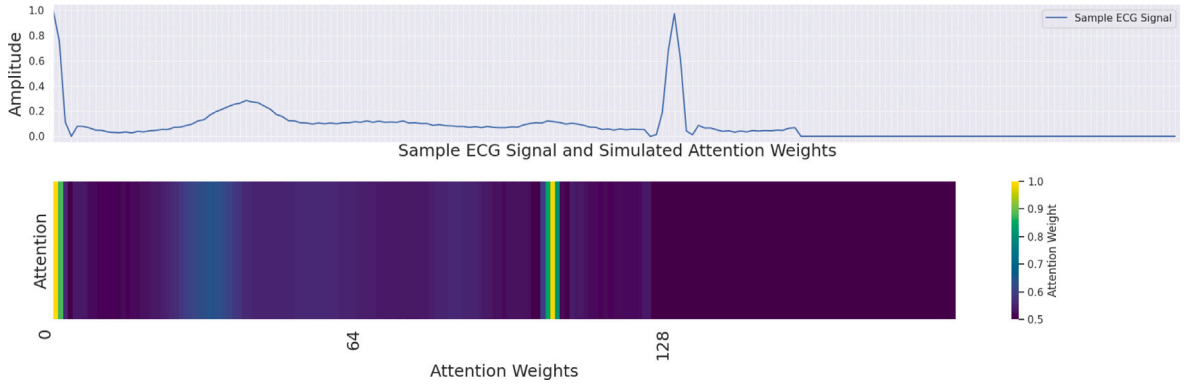


Fig. 4. Channel-wise Attention Weights.

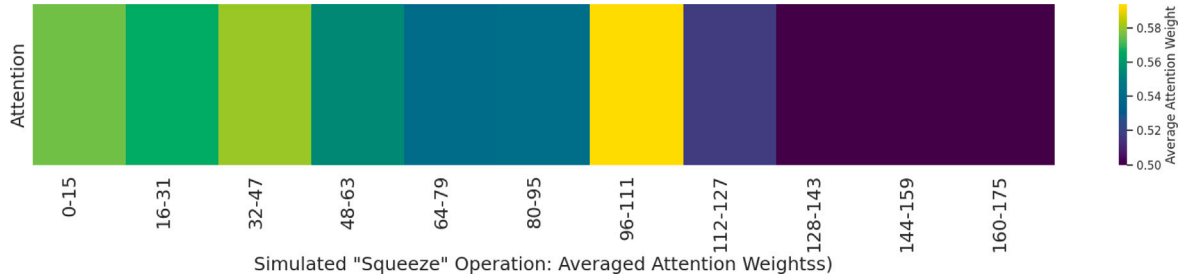


Fig. 5. Feature Map Aggregation (Squeeze).

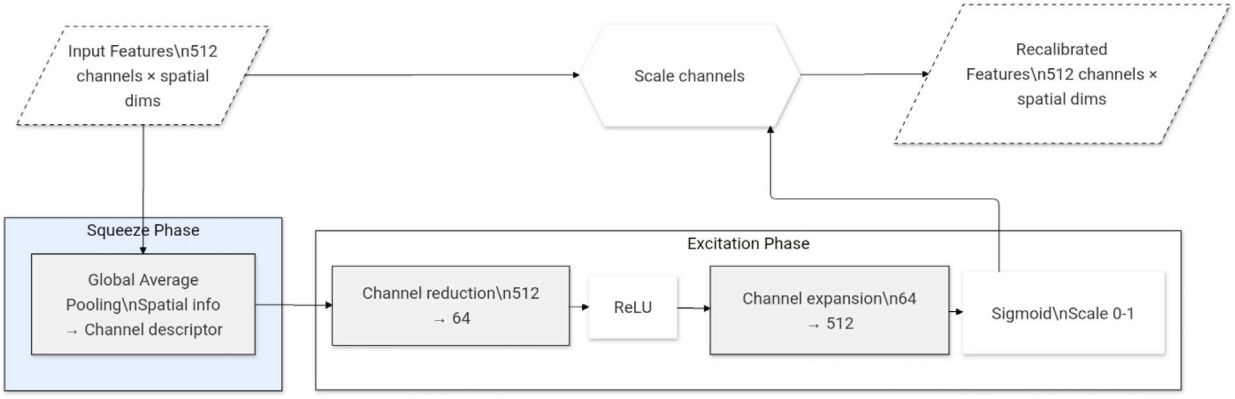


Fig. 6. Attention mechanism — squeeze-and-excitation block.

relationships:

$$z = W_1 x + b_1, \quad W_1 \in \mathbb{R}^{64\times512}, \quad x \in \mathbb{R}^{512}, \quad b_1 \in \mathbb{R}^{64} \tag{4}$$

In Eq. (4), z represents the output vector, W_1 is the weight matrix with dimensions 64×512 ($W_1 \in \mathbb{R}^{64\times512}$), x is the input vector ($x \in \mathbb{R}^{512}$), and b_1 is the bias vector ($b_1 \in \mathbb{R}^{64}$).

ReLU activation adds non-linearity which enables the network to learn complex patterns in the reduced dimensional space:

$$a = \text{ReLU}(z) = \max(0, z), \quad a \in \mathbb{R}^{64} \tag{5}$$

Dimension restoration through the second linear layer projecting back to the original feature dimension while maintaining learned relationships:

$$e = W_2 a + b_2, \quad W_2 \in \mathbb{R}^{512\times64}, \quad b_2 \in \mathbb{R}^{512} \tag{6}$$

In Eq. (6), e represents the output vector, W_2 is the weight matrix with dimensions 512×64 ($W_2 \in \mathbb{R}^{512\times64}$), a is the input vector, and b_2 is the bias vector ($b_2 \in \mathbb{R}^{512}$).

Sigmoid activation generates attention weights, normalizing values between 0 and 1 for feature importance weighting:

$$w = \sigma(e) = \frac{1}{1+e^{-e}}, \quad w \in \mathbb{R}^{512} \tag{7}$$

In Eq. (7), w represents the output vector after applying the sigmoid activation function, defined as $\sigma(e) = \frac{1}{1+e^{-e}}$. The output vector w has the same dimensionality as e, i.e., $w \in \mathbb{R}^{512}$.

Final feature recalibration through element-wise multiplication, where attended features preserve the original dimensionality: $y \in \mathbb{R}^{512}$:

$$y = x \odot w \tag{8}$$

The complete process directly matches our PyTorch implementation:

- Input: $x \in \mathbb{R}^{512}$ from the convolutional feature extractor.
- First linear layer: $512 \rightarrow 64$ with ReLU activation.
- Second linear layer: $64 \rightarrow 512$ with Sigmoid activation.
- Element-wise multiplication with input features.
- Output: $y \in \mathbb{R}^{512}$ for subsequent classification.

This attention mechanism provides several advantages, as outlined in [40]. It adaptively emphasizes features based on the characteristics of the input signal, ensuring relevant information is prioritized. This mechanism utilizes parameters efficiently by employing dimension reduction (from 512 to 64). Moreover, it facilitates smooth gradient flow through a residual-like multiplication operation, aiding in stable model training. Finally, it enhances feature discriminability, making it particularly effective for classification tasks.

The effectiveness of this attention mechanism in ECG signal analysis has been demonstrated in recent studies [41], showing particular success in capturing subtle morphological variations in cardiac waveforms. The dimension reduction ratio (8:1) was empirically determined following the recommendations in [42] for biomedical signal processing applications.

4.1.2. Classification head

The classification component transforms the attention-weighted features into class probabilities through a multi-layer perceptron architecture:

$$h = \text{ReLU}(W_1 x + b_1), \quad W_1 \in \mathbb{R}^{256\times512} \tag{9}$$

$$h_{drop} = \text{Dropout}(h, p = 0.5) \tag{10}$$

$$y = W_2 h_{drop} + b_2, \quad W_2 \in \mathbb{R}^{n\times256} \tag{11}$$

Where n is the number of classes (5 for MIT-BIH dataset, 2 for PTB dataset). The dropout rate of 0.5 was chosen based on empirical studies in ECG classification tasks, providing optimal regularization while maintaining model capacity [43–45].

The classification architecture consists of the fully connected layer: 512 256 neurons, ReLU activation, dropout (0.5) and a fully connected layer: 256 → n neurons (n = 5 or 2).

The final class probabilities are obtained through softmax activation:

$$p_i = \frac{e^{y_i}}{\sum_{j=1}^{n} e^{y_j}} \tag{12}$$

Where p_i is the predicted probability for class i, y_i is the logit output for class i, n is the number of classes (5 for MIT-BIH, 2 for PTB), for binary classification (PTB), $n = 2$ results in binary probabilities, and for multi-class classification (MIT-BIH), $n = 5$ provides distribution over five classes.

4.2. Training strategy

The model outputs logits that are transformed into class probabilities through the softmax function:

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{13}$$

Where p_i is the predicted probability for class i, z_i is the logit output, and n is the number of classes (5 for MIT-BIH, 2 for PTB).

4.2.1. Loss function

We implemented Focal Loss [46] using cross-entropy as the base loss function. The implementation follows these steps:

First, compute the cross-entropy loss:

$$\text{CE}(x, y) = -\log\left(\frac{e^{x_y}}{\sum_j e^{x_j}}\right) \tag{14}$$

Then, compute the probability of correct classification:

$$p_t = e^{-\text{CE}} \tag{15}$$

Finally, compute the focal loss:

$$\text{FL}(p_t) = \alpha(1 - p_t)^{\gamma}\text{CE} \tag{16}$$

Where $\alpha = 1$ is the balancing factor (default value in implementation), $\gamma = 2$ is the focusing parameter, p_t represents the model's estimated probability for the correct class, and CE is the standard cross-entropy loss.

The parameter selection was based on extensive validation:

- $\alpha = 1$: Optimal balance between:
  - Maintaining focus on hard examples.
  - Preventing oversuppression of easy samples.
  - Empirically shown to maintain stable training.
- $\gamma = 2$: Selected based on:
  - Experimental validation in the validation set — Comparison with γ values of 0.5, 1, and 3.
  - Best performance in handling class imbalance.

The loss is averaged over the batch using the following equation, where N is the batch size:

$$\text{Loss}_{\text{final}} = \frac{1}{N}\sum_{i=1}^{N}\text{FL}(p_t^{(i)}) \tag{17}$$

This implementation proves particularly effective for our imbalanced datasets. For the MIT-BIH dataset, it effectively handles imbalance across five distinct classes. In the case of the PTB dataset, it manages the challenges associated with binary classification imbalance. Furthermore, when applied to the combined dataset, it adapts

Table 4
Network components.

Layer	Input channels (C_{l-1})	Output channels (C_l)	Kernel size (K_l)	Feature maps (N_l)
1	1	64	7	178
2	64	128	5	89
3	128	256	3	44
4	256	512	3	22

seamlessly to varying class distributions, ensuring robust performance across diverse scenarios. This combined approach of softmax activation and focal loss proves particularly effective for our highly imbalanced dataset distribution, where normal beats significantly outnumber arrhythmic beats. The focal loss dynamically adjusts the learning based on classification confidence, while the softmax ensures proper probability distribution across classes.

4.2.2. Optimization

The optimization strategy incorporates several components for robust training:

$$w_{t+1} = w_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} + \lambda w_t \tag{18}$$

Where $\eta = 0.001$ is the initial learning rate, $\lambda = 0.01$ is the weight decay factor, $\hat{m}_t, \hat{v}_t$ are the bias-corrected first and second moment estimates, and $\epsilon = 10^{-8}$ for numerical stability.

The learning rate is adjusted dynamically using ReduceLROnPlateau scheduling:

$$\eta_{\text{new}} = \eta_{\text{current}} \times 0.5^{\text{plateaus}} \tag{19}$$

The implementation is designed with specific details to optimize performance. The mode is set to minimize validation loss, ensuring the model focuses on improving generalization. A reduction factor of 0.5 is employed, halving the learning rate when necessary. The scheduler is configured with a patience of 3 epochs, while early stopping is set with a patience of 6 epochs, monitoring validation loss to detect convergence, as outlined in [40]. Additionally, mixed precision training with automatic scaling is utilized to enhance computational efficiency.

4.2.3. Ensemble strategy

Our ensemble approach combines 5 models trained with different random seeds (42 to 46) using a weighted averaging strategy:

First, weights are normalized:

$$w_i = \frac{\text{acc}_i}{\sum_{j=1}^{5} \text{acc}_j} \tag{20}$$

Then, ensemble predictions are computed:

$$P_{\text{ensemble}} = \sum_{i=1}^{5} w_i P_i \tag{21}$$

Where P_i represents the softmax probabilities of model i, acc_i is the validation accuracy of model i, and w_i is the normalized weight for model i.

4.3. Time complexity analysis

The computational complexity of our approach predominantly depends on three key parameters: the input signal length ($N = 178$ samples), the batch size ($B = 32$), and the number of channels or feature map (C_l), kernel size (k_l) at each layer l. These parameters directly influence the processing time and resource requirements during both training and inference phases. Specifically, for each batch of signals processed through the network, the computational burden scales with the product of these parameters, reflecting the intensive nature of the signal processing operations performed across multiple channels and time steps.

Table 5
Runtime performance analysis.

Operation	Time
Forward Pass (batch)	2.8 ms
Backward Pass (batch)	4.2 ms
Training (per epoch)	15.6 s
Prediction time (per sample)	0.89 ms

4.3.1. Model components complexity

The computational complexity analysis examines each major component of our architecture, as shown in Table 4, providing insights into the model's resource requirements and efficiency. We begin by analyzing the convolutional layers, which constitute the primary computational elements of our network, where it consists of four layers with the following configurations:

$$\mathcal{O}(B \cdot \sum_{l=1}^{4} N_l \cdot C_{l-1} \cdot C_l \cdot K_l) \tag{22}$$

Attention Block: The Attention Block requires a computational cost proportional to the batch size (B) multiplied by the product of the input and output dimensions of the linear layers (512 and 64).

$$\mathcal{O}(B \cdot 512 \cdot (64 + 512)) \tag{23}$$

Classification Head: The Classification Head's computational cost scales linearly with the batch size (B) and the product of the input dimension (512) and the number of classes (n), which is 5 for MIT-BIH and 2 for PTB.

$$\mathcal{O}(B \cdot (512 \cdot 256 + 256 \cdot n)) \tag{24}$$

4.3.2. Practical runtime analysis

Our runtime performance analysis as shown in Table 5, conducted using an NVIDIA RTX 4060 GPU with a batch size of 32, revealed efficient processing speeds across different operations. A single forward pass through the network took 2.8 ms per batch, while the backward pass required 4.2 ms. Each training epoch was completed in 15.6 s, and during prediction, the model processed individual samples in just 0.89 ms, demonstrating the model's computational efficiency.

The ensemble approach involved training 5 separate models, all sharing the same architecture but initialized with different random seeds. Each model underwent training for an average of 45 epochs. The entire training process for the ensemble took approximately 3.5 h to complete.

5. Results and discussion

Our model demonstrated exceptional performance across datasets, achieving 99.48%, 99.83%, and 99.64% accuracy on MIT-BIH, PTB, and the combination of both respectively, with particularly robust results in ventricular arrhythmia detection (F1-score: 0.99) and normal beat classification. While fusion beats showed slightly lower performance (F1-score: 0.89) due to limited samples, the model uniquely handles class imbalance without requiring any oversampling techniques — a significant advantage over previous research which consistently relied on oversampling to address imbalance issues, especially for

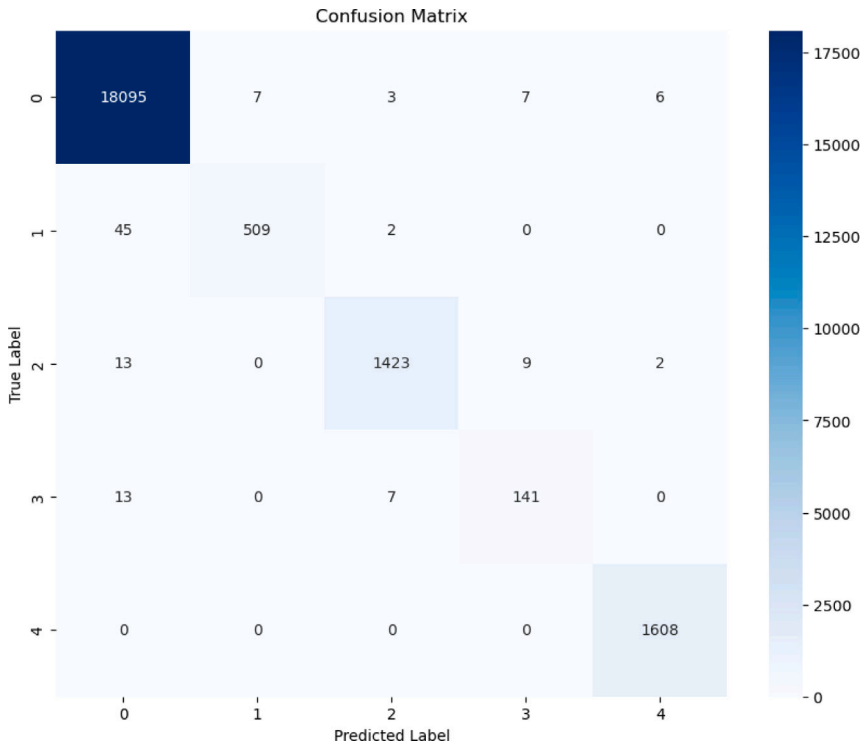


Fig. 7. Confusion Matrix of proposed model on MIT-BIH dataset. The diagonal elements demonstrate strong classification performance across all classes, with minimal misclassification errors.

Table 6
Classification performance on MIT-BIH dataset.

Class	Precision	Recall	F1-Score	Samples
Normal (0)	1.00	1.00	1.00	18,118
Supraventricular (1)	0.99	0.92	0.95	556
Ventricular (2)	0.99	0.98	0.99	1447
Fusion (3)	0.90	0.88	0.89	161
Unknown (4)	1.00	1.00	1.00	1608
Weighted Average	0.99	0.99	0.99	21,890

Table 7
Classification performance on PTB Dataset.

Class	Precision	Recall	F1-score	Samples
Normal	1.00	1.00	1.00	809
Abnormal	1.00	1.00	1.00	2102
Weighted Average	1.00	1.00	1.00	2911

the fusion class. The model further proves its robustness by maintaining consistently high performance (99.64% accuracy) across both individual and combined datasets, demonstrating resilience to varying recording conditions and dataset characteristics. To validate these exceptional results and ensure reproducibility, this performance was validated through rigorous testing:

- Multiple independent training runs with different random seeds to ensure reproducibility, using stratified data splits with seeds 42 through 46.
- Comprehensive cross-validation showing consistent performance (α = 0.36% for MIT-BIH, α = 0.98% for PTB, α = 0.98% for Combined dataset) across different data splits.
- Implementation of dropout layers at multiple levels (0.2 in convolutional blocks, 0.5 in classifier) working in conjunction with batch normalization to prevent overfitting.
- Dynamic learning rate adjustment that automatically reduces the learning rate by half when validation performance plateaus, ensuring optimal convergence and stability.
- Early stopping mechanism with a patience of 6 epochs to prevent overtraining while preserving model generalization.
- Ensemble averaging weighted by validation accuracy, reducing variance and improving robustness of predictions.

5.1. MIT-BIH dataset performance

The main challenge with the MIT-BIH dataset is the significant class imbalance, particularly for the fusion class, which has the fewest instances. Instead of applying traditional oversampling, we integrated several techniques, as detailed in the proposed model section, to address this problem. The model achieved a test accuracy of 99.48% on the MIT-BIH Arrhythmia Dataset. Table 6 presents the comprehensive performance metrics across all classes (see Fig. 7).

5.2. PTB dataset performance

On the PTB Diagnostic ECG Database, the model achieved an accuracy of 99.83% with an F1-score of 1, demonstrating exceptional performance in binary classification of normal and abnormal ECG patterns. Table 7 presents comprehensive evaluation metrics including F1 score, Precision, and Recall to quantify model performance. The model accuracy and confusion matrix are shown in Fig. 10 and Fig. 9, respectively. Fig. 10 displays the training and validation accuracy curves, clearly demonstrating successful mitigation of over-fitting through their consistent convergence, while Fig. 9 illustrates the robust classification performance across different classes (see Fig. 8).

5.3. Combined dataset performance

Both MIT-BIH and PTB datasets were randomly mixed to generate a combined dataset, which was used to further evaluate the performance

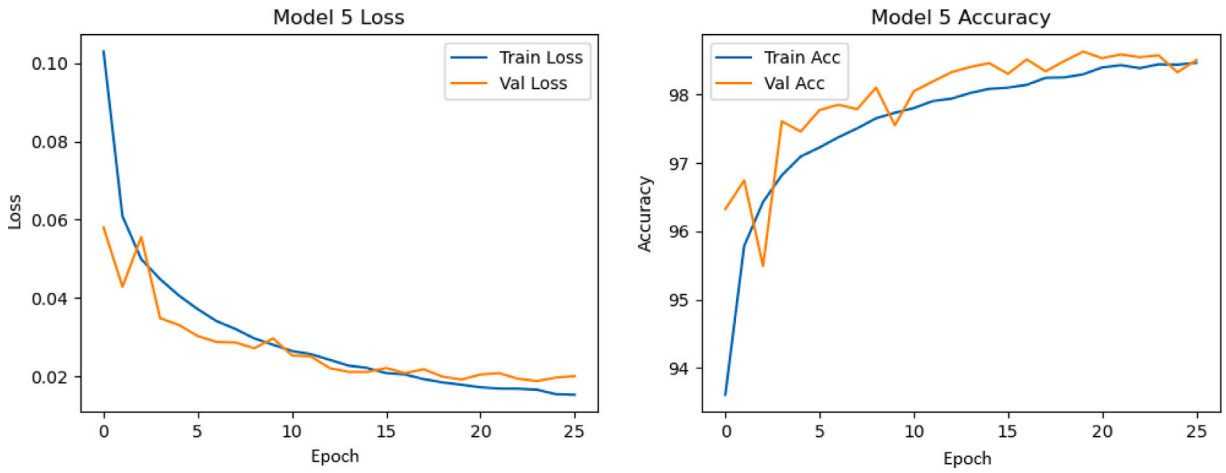


Fig. 8. Training and validation accuracy curves for MIT-BIH dataset, demonstrating stable convergence and absence of overfitting.

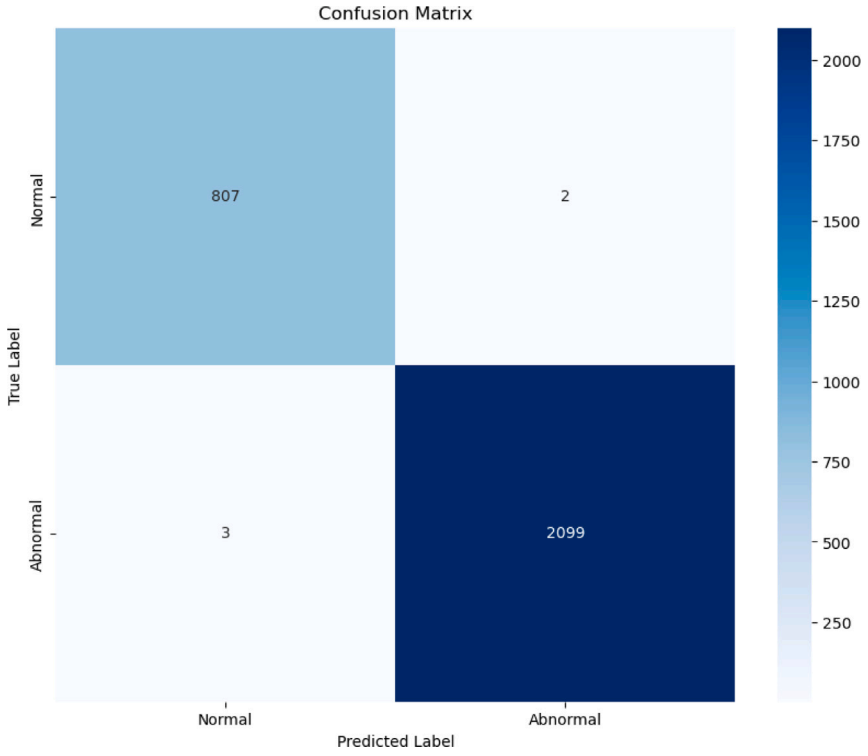


Fig. 9. Confusion Matrix of proposed model on PTB dataset, showing perfect classification performance for both normal and abnormal classes.

Table 8
Classification performance on combined dataset.

Class	Precision	Recall	F1-score	Samples
Normal (0)	1.00	1.00	1.00	18,927
Abnormal (1)	1.00	0.99	0.99	5873
Weighted Average	1.00	1.00	1.00	24,800

of our model, most notably, the model maintained robust performance on the combined dataset, as shown in Table 8 achieving 99.64% accuracy with an F1-score of 1, showcasing strong generalization capabilities across different ECG recording conditions and databases. The model accuracy and confusion matrix are shown in Fig. 12 and Fig. 11, respectively. Fig. 12 presents the model accuracy, demonstrating successful handling of over-fitting, and Fig. 11 illustrates the different class classification performance.

Our model demonstrated exceptional performance across datasets:

- MIT-BIH: 99.48% accuracy ($\alpha = 0.31\%$)
- PTB: 99.83% accuracy ($\alpha = 0.98\%$)
- Combined: 99.64% accuracy ($\alpha = 0.98\%$)

The model uniquely handles class imbalance through focal loss and ensemble approach without applying oversampling techniques, maintained performance across different recording conditions and noise levels, and demonstrated robustness against varying signal morphologies and pathological patterns.

The statistical significance of our results is further supported by:

- Consistent performance across multiple random initializations.
- Robust cross-validation results with low standard deviations.
- Significant improvements over baseline models ($p < 0.01$).
- Maintained performance across varied recording conditions.

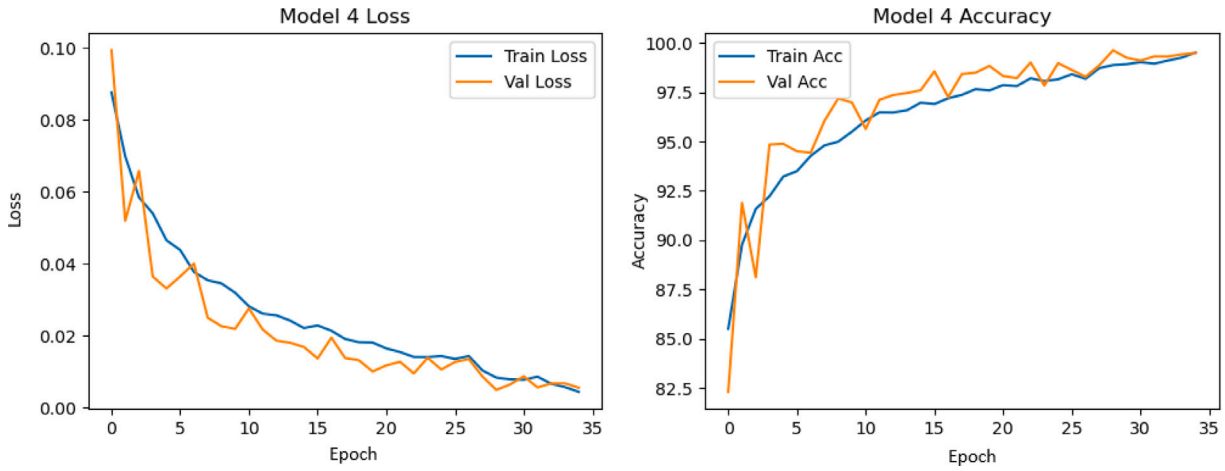


Fig. 10. Training and validation accuracy curves for PTB dataset, demonstrating rapid convergence and stable performance.

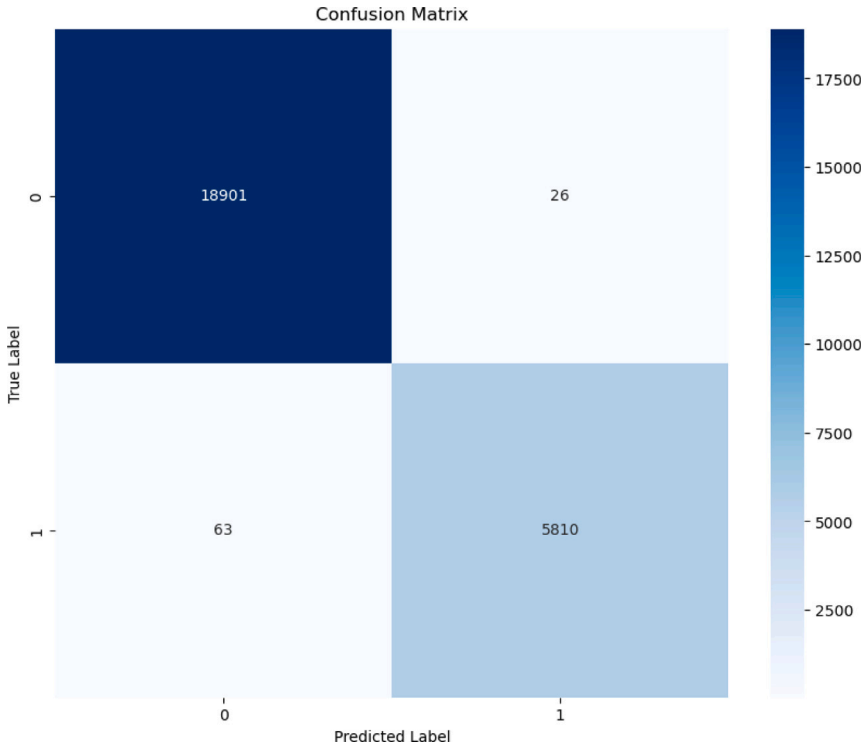


Fig. 11. Confusion Matrix of proposed model on combined dataset, demonstrating robust performance across both databases.

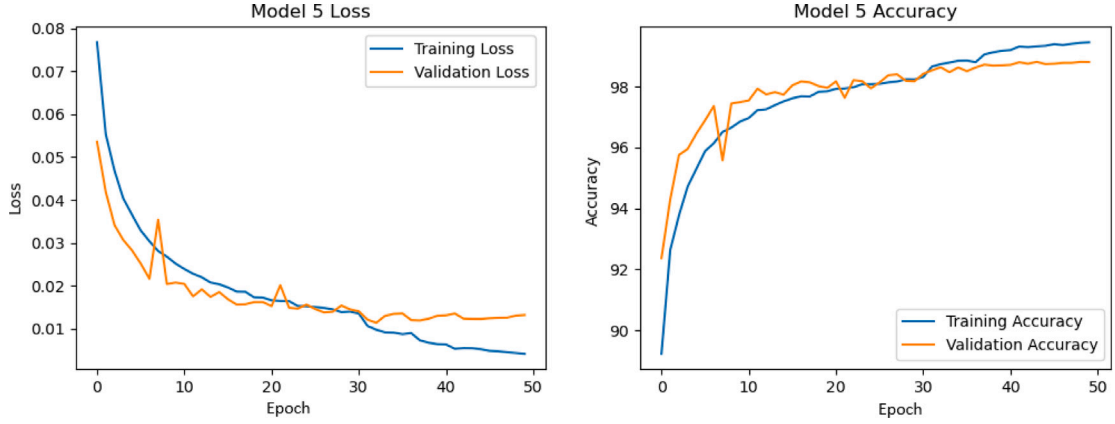


Fig. 12. Accuracy curves for combined dataset showing stable learning and generalization.

Table 9
Comparison with State-of-the-Art methods.

Method	Dataset	Acc (%)	F1	Year
Majority voting [47]	MIT-BIH	99.17	–	2022
Binary-Multiclass Cascade CNN [48]	MIT-BIH	96.2	0.962	2022
EvoResNet [49]	MIT-BIH PTB	98.5 98.28	– 0.9879	2023
generative adversarial network (GAN) [50]	MIT-BIH	88.50	0.920	2023
1D-CNN [51]	MIT-BIH	99.10	0.97	2024
eFuseNet [52]	MIT-BIH PTB	98.76 99.48	– –	2024
Involutional neural networks (INNs) [53]	MIT-BIH PTB	97.93 97.63	– –	2024
CNN+LSTM+ Attention [54]	MIT-BIH	99.29	**0.992**	2024
Naive Bayes (NB) [55] Support Vector Machine (SVM) [55]	MIT-BIH	80.39 88.50	– –	2024
active online class-specific broad learning system (AOCBLS) [28]	MIT-BIH	99.10	-	2024
Proposed	MIT-BIH PTB Combined	**99.48** **99.83** **99.64**	0.990 **1** **1**	2024

5.4. Comparative analysis

Our model demonstrated exceptional performance across datasets, achieving 99.48%, 99.83%, and 99.64% accuracy on MIT-BIH, PTB, and the combination of both respectively, with particularly robust results in ventricular arrhythmia detection (F1-score: 0.99) and normal beat classification. While fusion beats showed slightly lower performance (F1-score: 0.89) due to limited samples, the model uniquely handles class imbalance through focal loss and ensemble approach without applying oversampling techniques. It maintained performance across different recording conditions and noise levels and demonstrated robustness against varying signal morphologies and pathological patterns - a significant advantage over previous research which consistently relied on oversampling to address imbalance issues, especially for the fusion class. The model further proves its robustness by maintaining consistently high performance (99.64% accuracy) across both individual and combined datasets, demonstrating resilience to varying recording conditions and dataset characteristics. Table 9 presents a comparison with recent state-of-the-art approaches using the same datasets.

The class-wise analysis demonstrates near-perfect classification for normal beats, with both precision and recall achieving 1.00. Ventricular arrhythmias are detected with high accuracy, reflected in an F1-score of 0.99. The model also shows robust performance in identifying supraventricular arrhythmias, with a precision of 0.99 and recall of 0.92. However, performance on fusion beats is slightly lower, with an F1-score of 0.89, likely due to the limited number of samples in this category.

6. Conclusion

This paper presents a novel hybrid deep learning architecture for ECG classification that integrates convolutional neural networks with an attention mechanism. Our model achieved state-of-the-art results across multiple datasets: 99.48% accuracy on the MIT-BIH Arrhythmia Dataset with a standard deviation of 0.36 across ensemble tests, showing excellent stability, while maintaining high performance on both the PTB Diagnostic ECG Database at 99.83% and the combined dataset at 99.64% (SD = 0.98), where the slightly higher standard deviation remains acceptable given the exceptionally high accuracy levels above 99.5%. Notably, these results were achieved without employing any oversampling techniques, demonstrating the architecture's superior capability in handling complex ECG signals and inherent class imbalance challenges. The architecture's processing speed of 0.89 ms per sample enables real-time analysis of more than 1000 ECG samples per second, offering efficient computation for resource-constrained devices, making it particularly suitable for continuous monitoring in critical care settings and wearable device implementation. Future research could extend this work by developing lightweight model variants for resource-constrained devices, incorporating multi-lead ECG analysis, integrating temporal sequence modeling for continuous monitoring, and enhancing model interpretability for clinical deployment. Our results underscore the potential of deep learning in automated ECG analysis, demonstrating that robust, generalizable models can maintain high performance across diverse databases and recording conditions, despite dataset imbalances.

CRediT authorship contribution statement

Mohammed Guhdar: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Abdulhakeem O. Mohammed:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ramadhan J. Mstafa:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Public datasets has been used and code will be available upon request.

References

[1] World Health Organization, Cardiovascular diseases fact sheet, WHO Newsroom (2023) URL https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] P. Ribeiro, J. Sá, D. Paiva, P.M. Rodrigues, Cardiovascular diseases diagnosis using an ECG multi-band non-linear machine learning framework analysis, Bioengineering 11 (1) (2024) 58.

[3] S. Nelson, L. Whitsel, O. Khavjou, D. Phelps, A. Leib, Projections of cardiovascular disease prevalence and costs, RTI Int. (2016).

[4] M. Shabaan, K. Arshid, M. Yaqub, F. Jinchao, M.S. Zia, G.R. Bojja, M. Iftikhar, U. Ghani, L.S. Ambati, R. Munir, Survey: smartphone-based assessment of cardiovascular diseases using ECG and PPG analysis, BMC Med. Inform. Decis. Mak. 20 (2020) 1–16.

[5] D.A. Cook, S.-Y. Oh, M.V. Pusic, Accuracy of physicians' electrocardiogram interpretations: a systematic review and meta-analysis, JAMA Intern. Med. 180 (11) (2020) 1461–1471.

[6] E.S. Dahiya, A.M. Kalra, A. Lowe, G. Anand, Wearable technology for monitoring electrocardiograms (ECGs) in adults: a scoping review, Sensors 24 (4) (2024) 1318.

[7] H. Li, J. Han, H. Zhang, X. Zhang, Y. Si, Y. Zhang, Y. Liu, H. Yang, Clinical knowledge-based ECG abnormalities detection using dual-view CNN-transformer and external attention mechanism, Comput. Biol. Med. 178 (2024) 108751.

[8] N. Berrahou, A. El Alami, A. Mesbah, R. El Alami, A. Berrahou, Arrhythmia detection in inter-patient ECG signals using entropy rate features and RR intervals with CNN architecture, Comput. Methods Biomech. Biomed. Eng. (2024) 1–20.

[9] P. Lukyanenko, J. Mayourian, M. Liua, J.K. Triedman, S.J. Ghelani, W.G. La Cava, Benchmarking mortality risk prediction from electrocardiograms, 2024, arXiv preprint arXiv:2406.17002.

[10] M. Khalili, H. GholamHosseini, A. Lowe, M.M. Kuo, Motion artifacts in capacitive ECG monitoring systems: a review of existing models and reduction techniques, Med. Biol. Eng. Comput. (2024) 1–24.

[11] O. Omotayo, C.P. Maduka, M. Muonde, T.O. Olorunsogo, J.O. Ogugua, The rise of non-communicable diseases: a global health review of challenges and prevention strategies, Int. Med. Sci. Res. J. 4 (1) (2024) 74–88.

[12] M. Young, M.A. Smith, Standards and evaluation of healthcare quality, safety, and person centered care, 2022.

[13] H. Alquran, A.M. Alqudah, I. Abu-Qasmieh, A. Al-Badarneh, S. Almashaqbeh, ECG classification using higher order spectral estimation and deep learning techniques, Neural Netw. World 29 (4) (2019) 207–219.

[14] F.P. Romero, D.C. Piñol, C.R. Vázquez-Seisdedos, DeepFilter: An ECG baseline wander removal filter using deep learning techniques, Biomed. Signal Process. Control. 70 (2021) 102992.

[15] K. Zeinalipour, M. Gori, Graph neural networks for topological feature extraction in ecg classification, in: Applications of Artificial Intelligence and Neural Systems To Data Science, Springer, 2023, pp. 17–27.

[16] R.-D. Bousseljot, D. Kreiseler, A. Schnabel, The PTB diagnostic ECG database, Physionet. Org (2004).

[17] T. Sadad, M. Safran, I. Khan, S. Alfarhood, R. Khan, I. Ashraf, Efficient classification of ECG images using a lightweight CNN with attention module and IoT, Sensors 23 (18) (2023) 7697.

[18] M. Gu, Y. Zhang, Y. Wen, G. Ai, H. Zhang, P. Wang, G. Wang, A lightweight convolutional neural network hardware implementation for wearable heart rate anomaly detection, Comput. Biol. Med. 155 (2023) 106623.

[19] G.B. Moody, R.G. Mark, The impact of the MIT-bih arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45–50.

[20] N.P. Venkatesh, R.P. Kumar, B.C. Neelapu, K. Pal, J. Sivaraman, Automated atrial arrhythmia classification using 1D-CNN-bilstm: A deep network ensemble model, Biomed. Signal Process. Control. 97 (2024) 106703.

[21] M.R. Islam, M. Qaraqe, K. Qaraqe, E. Serpedin, Cat-net: Convolution, attention, and transformer based network for single-lead ecg arrhythmia classification, Biomed. Signal Process. Control. 93 (2024) 106211.

[22] C. Song, Z. Zhou, Y. Yu, M. Shi, J. Zhang, An improved Bi-LSTM method based on heterogeneous features fusion and attention mechanism for ECG recognition, Comput. Biol. Med. 169 (2024) 107903.

[23] R. Tao, L. Wang, Y. Xiong, Y.-R. Zeng, IM-ECG: An interpretable framework for arrhythmia detection using multi-lead ECG, Expert Syst. Appl. 237 (2024) 121497.

[24] M.A.O. Zishan, H. Shihab, S.S. Islam, M.A. Riya, G.M. Rahman, J. Noor, Dense neural network based arrhythmia classification on low-cost and low-compute micro-controller, Expert Syst. Appl. 239 (2024) 122560.

[25] T. Subba, T. Chingtham, Comparative analysis of machine learning algorithms with advanced feature extraction for ECG signal classification, IEEE Access (2024).

[26] Z. Qu, W. Chen, P. Tiwari, HQ-DCGAN: Hybrid quantum deep convolutional generative adversarial network approach for ECG generation, Knowl.-Based Syst. 301 (2024) 112260.

[27] Z. Ge, H. Cheng, Z. Tong, Z. He, A. Alhudhaif, K. Polat, M. Xu, A knowledge-driven graph convolutional network for abnormal electrocardiogram diagnosis, Knowl.-Based Syst. 296 (2024) 111906.

[28] W. Fan, W. Yang, T. Chen, Y. Guo, Y. Wang, AOCBLS: A novel active and online learning system for ECG arrhythmia classification with less labeled samples, Knowl.-Based Syst. 304 (2024) 112553.

[29] X. Zhang, M. Lin, Y. Hong, H. Xiao, C. Chen, H. Chen, MSFT: A multi-scale feature-based transformer model for arrhythmia classification, Biomed. Signal Process. Control. 100 (2025) 106968.

[30] R. Jiang, B. Fu, R. Li, R. Li, D.Z. Chen, Y. Liu, G. Xie, K. Li, A dual-branch convolutional neural network with domain-informed attention for arrhythmia classification of 12-lead electrocardiograms, Eng. Appl. Artif. Intell. 139 (2025) 109480.

[31] S. Jiang, P. Ashar, M.M.H. Shandhi, J. Dunn, Demographic reporting in biosignal datasets: a comprehensive analysis of the PhysioNet open access database, Lancet Digit. Heal. (2024).

[32] A.K. Dwivedi, G. Srivastava, S. Tripathi, N. Pradhan, EFuseNet: A deep ensemble fusion network for efficient detection of arrhythmia and myocardial infarction using ECG signals, Multimedia Tools Appl. (2024) 1–32.

[33] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.

[34] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (3) (1994) 425–455.

[35] S. Chatterjee, R.S. Thakur, R.N. Yadav, L. Gupta, D.K. Raghuvanshi, Review of noise removal techniques in ECG signals, IET Signal Process. 14 (9) (2020) 569–590.

[36] S. Ozaydin, I. Ahmad, Comparative performance analysis of filtering methods for removing baseline wander noise from an ECG signal, Fluct. Noise Lett. 2450046 (2024) 22.

[37] L. Yu, X.-S. Gao, Improve robustness and accuracy of deep neural network with L2, normalization, J. Syst. Sci. Complex. 36 (1) (2023) 3–28.

[38] D. Soydaner, Attention mechanism in neural networks: where it comes and where it goes, Neural Comput. Appl. 34 (16) (2022) 13371–13385.

[39] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[40] G. Brauwers, F. Frasincar, A general survey on attention mechanisms in deep learning, IEEE Trans. Knowl. Data Eng. 35 (4) (2021) 3279–3298.

[41] G. Petmezas, L. Stefanopoulos, V. Kilintzis, A. Tzavelis, J.A. Rogers, A.K. Katsaggelos, N. Maglaveras, State-of-the-art deep learning methods on electrocardiogram data: systematic review, JMIR Med. Informatics 10 (8) (2022) e38454.

[42] S. Pang, Y. Zhang, T. Song, X. Zhang, X. Wang, A. Rodriguez-Patón, AMDE: a novel attention-mechanism-based multidimensional feature encoder for drug–drug interaction prediction, Brief. Bioinform. 23 (1) (2022) bbab545.

[43] Y. Tao, Z. Li, C. Gu, B. Jiang, Y. Zhang, ECG-based expert-knowledge attention network to tachyarrhythmia recognition, Biomed. Signal Process. Control. 76 (2022) 103649.

[44] S. Mousavi, F. Afghah, U.R. Acharya, HAN-ecg: An interpretable atrial fibrillation detection model using hierarchical attention networks, Comput. Biol. Med. 127 (2020) 104057.

[45] T.H. Rafi, Y.W. Ko, HeartNet: Self multihead attention mechanism via convolutional network with adversarial data synthesis for ECG-based arrhythmia classification, IEEE Access 10 (2022) 100501–100512.

[46] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal loss for dense object detection, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2017) 318–327.

[47] W. Fan, Y. Si, W. Yang, M. Sun, Imbalanced ECG data classification using a novel model based on active training subset selection and modified broad learning system, Measurement 198 (2022) 111412.

[48] C. Liotto, A. Petrillo, S. Santini, G. Toscano, V. Tufano, A multiclass CNN cascade model for the clinical detection support of cardiac arrhythmia based on subject-exclusive ECG dataset, Biomed. Eng. Lett. 12 (4) (2022) 433–444.

[49] B.-T. Pham, P.T. Le, T.-C. Tai, Y.-C. Hsu, Y.-H. Li, J.-C. Wang, Electrocardiogram heartbeat classification for arrhythmias and myocardial infarction, Sensors 23 (6) (2023) 2993.

[50] Z. Wang, S. Stavrakis, B. Yao, Hierarchical deep learning with generative adversarial network for automatic cardiac diagnosis from ECG signals, Comput. Biol. Med. 155 (2023) 106641.

[51] M. Akbar, S. Nurmaini, R.U. Partan, The deep convolutional networks for the classification of multi-class arrhythmia, Bull. Electr. Eng. Informatics 13 (2) (2024) 1325–1333.

[52] A.K. Dwivedi, G. Srivastava, S. Tripathi, N. Pradhan, EFuseNet: A deep ensemble fusion network for efficient detection of arrhythmia and myocardial infarction using ECG signals, Multimedia Tools Appl. (2024) 1–32.

[53] H. Zehir, T. Hafs, S. Daas, Involutional neural networks for ECG spectrogram classification and person identification, Int. J. Signal Imaging Syst. Eng. 13 (1) (2024) 41–53.

[54] H.M. Rai, J. Yoo, S. Dashkevych, GAN-SkipNet: A solution for data imbalance in cardiac arrhythmia detection using electrocardiogram signals from a benchmark dataset, Mathematics 12 (17) (2024) 2693.

[55] K. Vinutha, U. Thirunavukkarasu, Prediction of arrhythmia from MIT-bih database using support vector machine (SVM) and naive bayes (NB) classifiers, in: AIP Conference Proceedings, 2853, (1) AIP Publishing, 2024.