Artificial Intelligence and Machine Learning Review
Scipublication
DOI: 10.69987/AIMLR.2024.50306

Attention-Enhanced Multi-Scale Feature Optimization for Silent Myocardial Infarction and Early Atrial Fibrillation Detection in ECG Signals

Zejun Cheng
Internal Medicine, Capital Medical University, Beijing, China

Keywords

electrocardiogram classification, attention mechanism, multi-scale feature extraction, myocardial infarction detection, atrial fibrillation diagnosis

Abstract

Cardiovascular diseases remain the leading cause of mortality worldwide, with silent myocardial infarction and paroxysmal atrial fibrillation presenting significant diagnostic challenges due to their subtle electrocardiographic manifestations. This study proposes a novel multi-scale feature extraction method combining residual U-blocks with depthwise separable convolutions and attention mechanisms for enhanced detection of these conditions. The approach integrates clinical domain knowledge through guided spatial attention, emphasizing ST-T segments and high-frequency QRS components while preserving prognostic signals during preprocessing. A comprehensive evaluation of the MIT-BIH Arrhythmia Database, PTB-XL Database, and PhysioNet Challenge 2017 demonstrates 98.7% accuracy for arrhythmia classification, 96.4% sensitivity for myocardial infarction detection, and 97.9% specificity for atrial fibrillation identification, representing improvements of 3.2%, 4.1%, and 2.8% over baseline CNN-LSTM architectures. The optimized network achieves 58% parameter reduction through efficient convolution strategies, enabling deployment on resource-constrained devices. Grad-CAM visualization validates clinical relevance by demonstrating focused attention on diagnostically significant regions of the ECG.

1. Introduction

1.1 Background and Motivation

Cardiovascular diseases account for approximately 17.9 million deaths annually, representing 31% of global mortality. Among these conditions, myocardial infarction and atrial fibrillation constitute critical health threats requiring timely identification and intervention. Silent myocardial infarction, characterized by minimal or absent symptoms, comprises 22-64% of non-fatal myocardial infarctions and demonstrates 43% higher long-term mortality compared to symptomatic presentations [1]. This condition presents unique diagnostic challenges, as conventional electrocardiographic changes may be subtle or transient, often being dismissed as signal artifacts or normal variants. Paroxysmal atrial fibrillation, marked by intermittent episodes of irregular cardiac rhythm, increases stroke risk fivefold yet remains undetected in many patients due to its episodic nature [2]. Standard 12-lead electrocardiography serves as the primary non-invasive diagnostic tool, but manual interpretation requires specialized expertise and remains susceptible to inter-observer variability.

Recent advances in deep learning have revolutionized electrocardiogram analysis, enabling the automated detection of cardiac abnormalities with performance that approaches or exceeds that of human experts [3]. Convolutional neural networks extract spatial features from signal morphology, while recurrent architectures capture temporal dependencies critical for rhythm analysis. The integration of these approaches has yielded promising results across various cardiovascular conditions. Contemporary diagnostic systems face persistent challenges in balancing detection accuracy with computational efficiency, which limits their deployment in point-of-care settings and wearable monitoring devices. The preservation of high-frequency signal components during preprocessing remains problematic, as conventional filtering techniques often eliminate clinically relevant information essential for detecting subtle pathological changes [4].

1.2 Related Work and Research Gaps

Traditional electrocardiogram analysis relied on hand-crafted features derived from wavelet transforms, entropy measures, and statistical descriptors [5]. Arif demonstrated robust beat classification using discrete wavelet transform coefficients, achieving 94.2% accuracy across multiple arrhythmia categories [6]. Alexakis et al. employed feature extraction techniques combining time-domain and frequency-domain characteristics for specific pathological conditions. While effective for well-defined patterns, these methods struggled with subtle abnormalities and required extensive domain expertise for feature engineering.

Deep learning architectures have progressively replaced manual feature extraction. Al Rahhal et al. applied convolutional neural networks to electrocardiogram classification, demonstrating superior performance compared to traditional machine learning approaches [7]. The architecture learns hierarchical representations directly from raw signals, eliminating the need for manual feature design. Venkatesan et al. developed preprocessing pipelines incorporating support vector machine classifiers for remote healthcare applications, emphasising practical deployment considerations. Liu et al. provided comprehensive reviews of deep learning methodologies in electrocardiogram diagnosis, highlighting the rapid evolution from simple architectures to sophisticated multi-modal systems. Recent investigations have explored attention mechanisms for emphasising diagnostically relevant signal regions, though most implementations lack integration of clinical domain knowledge.

Critical gaps persist in current methodologies. Standard preprocessing protocols employ aggressive high-frequency filtering to remove noise, inadvertently eliminating prognostic information essential for detecting silent myocardial infarction. Single-scale feature extraction fails to capture both subtle morphological abnormalities and temporal rhythm variations simultaneously. Computational complexity hinders deployment on edge devices and wearable monitors, limiting real-world applicability. Most systems operate as black boxes, providing classifications without interpretable evidence, which impedes clinical acceptance and regulatory approval.

2. Methods

2.1 Dataset Description and Preprocessing

A. Database Characteristics

The experimental framework employed three benchmark electrocardiography databases to ensure comprehensive evaluation across diverse pathological conditions. The MIT-BIH Arrhythmia Database provided 48 half-hour recordings sampled at 360Hz from 47 subjects, containing approximately 110,000 annotated beats distributed across five AAMI classes: normal beats, supraventricular ectopic beats, ventricular ectopic beats, fusion beats, and unknown beats [8]. The class distribution exhibited a significant imbalance, with normal beats comprising 75% of the dataset, while fusion beats represented only 0.8%. The PTB-XL Database comprised 21,837 clinical 12-lead electrocardiogram records from 18,885 patients, each spanning 10 seconds at a sampling rate of 500Hz. This database included 9,528 myocardial infarction cases with detailed localization annotations and 1,514 records exhibiting subtle or ambiguous findings [9]. PhysioNet Challenge 2017 contributed 8,528 single-lead ambulatory electrocardiogram recordings categorized into normal sinus rhythm, atrial fibrillation, other rhythms, and noisy segments, specifically designed for paroxysmal arrhythmia detection.

B. Data Augmentation Strategy

Class imbalance required systematic correction to prevent bias toward majority classes during network training. The SMOTE-Tomek hybrid approach generated synthetic minority class samples through interpolation between neighboring instances while removing borderline cases that blurred decision boundaries. This technique increased minority class representation by 300% while eliminating 12% of ambiguous samples. Random noise injection simulated realistic recording conditions by adding Gaussian noise with signal-to-noise ratios ranging from 15 to 25 dB. Baseline wander augmentation was applied with sinusoidal drifts at frequencies between 0.15 and 0.3Hz and amplitudes of up to 15% of the signal range, replicating respiratory and motion artifacts commonly encountered in ambulatory monitoring. Temporal shift augmentation introduced random delays of ±50ms to improve robustness against R-peak detection variability [10]. SMOTE-Tomek is applied only to training folds after the inter-patient split; no synthetic or cleaned samples are used in validation/test.

C. Signal Preprocessing Pipeline

For myocardial infarction (MI), we adopt dual priors: (i) a QRS-centric window ($\pm$50 ms around R-peak) to capture pathological Q waves and QRS fragmentation, and (ii) an ST-T window ($\approx$ 60 – 120 ms after the J-point) to capture ST-segment deviation and T-wave abnormalities. Spatial attention is gated by the union of these priors, while remaining data-driven [11]. Modified Kalman filtering addressed baseline wander through iterative state estimation, modeling the drift as a slowly varying process independent of the cardiac signal. This approach achieved baseline correction without introducing phase distortion or amplitude attenuation in diagnostically relevant frequency bands. R-peak detection utilized the Pan-Tompkins algorithm enhanced with adaptive thresholding based on local signal statistics. Detected peaks served as fiducial points for segmentation, extracting 2.5-second windows centered on each R-peak. Signal standardization applied z-score normalization within each lead independently, ensuring zero mean and unit variance. Resampling to a uniform 360Hz facilitated consistent temporal resolution across databases while maintaining signal fidelity.

2.2 Proposed Multi-Scale Feature Extraction Architecture

A. Network Architecture Design

The proposed network accepts 10-second electrocardiogram segments as input, with dimensionality of (batch size, 12, 3600) for 12-lead recordings or (batch_size, 1, 3600) for single-lead data. The initial layer performs temporal expansion through 1D convolution with 64 filters of kernel size 7, establishing baseline feature representations. Multi-scale feature extraction blocks follow, each containing three parallel branches with receptive fields of 3, 7, and 15 samples, corresponding to rapid, intermediate, and gradual electrocardiographic phenomena. Depthwise separable convolutions replace standard convolutions within these branches, decomposing the operation into depthwise spatial convolution followed by pointwise 1×1 convolution. This factorization reduces the cost from $\mathbf{O(D_k \cdot M \cdot N)}$ to $\mathbf{O\ (D_k \cdot M + M \cdot N)}$ for 1D convolutions, where D_k is kernel length, M input channels, and N output channels.

B. Attention Mechanism Integration

Channel attention employs squeeze-and-excitation blocks that capture inter-channel dependencies through global average pooling, followed by two fully connected layers with a reduction ratio of r = 16. The attention weights undergo sigmoid activation, scaling feature maps according to the importance of each channel. Spatial attention operates complementarily, generating attention maps emphasizing diagnostically relevant temporal positions. Guided spatial attention incorporates clinical domain knowledge through learnable masks initialized with physiological priors. For atrial fibrillation detection, masks emphasize the 100-300ms interval following R-peaks, corresponding to the ST-T segment, where nonspecific abnormalities are associated with paroxysmal episodes. For myocardial infarction classification, masks highlight the QRS complex spanning ±50ms around detected peaks, where high-frequency components indicate damaged myocardium. Multi-head self-attention layers with four heads capture long-range temporal dependencies, enabling the network to model relationships between distant electrocardiographic segments. Bidirectional long short-term memory layers with 64 units process the attended features, maintaining hidden states that encode temporal context. The final classification stage employs two fully connected layers with dropout regularization (p = 0.5), culminating in softmax activation for multi-class prediction.

3. Experimental Results

3.1 Quantitative Performance Analysis

A. MIT-BIH Arrhythmia Classification

We adopt a stratified 10-fold inter-patient cross-validation approach: records from the same subject never appear in both training and validation/test folds. All splits are performed at the record (subject) level and then propagated to the beat/segment level. The proposed architecture achieved an overall classification accuracy of 98.7%, demonstrating robust discrimination among diverse beat morphologies. Sensitivity analysis revealed a 99.2% detection rate for normal beats, 97.8% for supraventricular ectopic beats, 98.1% for ventricular ectopic beats, 94.3% for fusion beats, and 93.7% for unknown beats. The substantial performance on minority classes validated the effectiveness of the SMOTE-Tomek augmentation strategy. Specificity exceeded 98.5% across all categories, indicating low false positive rates critical for clinical deployment. F1-scores ranged from 94.1% for fusion beats to 99.1% for normal beats, reflecting balanced precision and recall. To avoid leakage, all preprocessing and resampling statistics (e.g., normalization) are fit on the training fold only and applied to validation/test folds.

The confusion matrix analysis identified primary misclassification patterns. Fusion beats exhibited 4.2% confusion with ventricular ectopic beats due to morphological similarity during ventricular capture. Unknown beats demonstrated 3.8% misclassification as supraventricular ectopic beats, attributable to atypical P-wave configurations. Ventricular ectopic beats showed a 1.9% false negative rate, primarily involving interpolated premature ventricular contractions with near-normal morphology. Receiver operating characteristic curve analysis yielded area under curve values of 0.997 for normal beats, 0.989 for supraventricular ectopic beats, 0.992 for ventricular ectopic beats, 0.973 for fusion beats, and 0.968 for unknown beats, indicating excellent discriminative capacity across all classes [12].

B. PTB-XL Myocardial Infarction Detection

Myocardial infarction detection on the PTB-XL Database addressed binary classification (normal vs. infarction) and multi-class localization (anterior, inferior, lateral walls). Binary classification achieved a sensitivity of 96.4%, a specificity of 95.8%, and an overall accuracy of 96.1%. The sensitivity improvement of 4.1% over baseline CNN-LSTM architectures directly addressed the critical need for early detection of subtle ischemic changes. Localization performance demonstrated 94.7% accuracy for anterior infarctions, 93.2% for inferior infarctions, and 91.8% for lateral infarctions. The slightly reduced performance for lateral infarctions reflected the inherent diagnostic challenge of identifying changes in lateral leads where amplitude variations are more subtle.

Analysis of detection latency revealed an average inference time of 47ms per 10-second segment on NVIDIA GTX 1080 Ti hardware, compatible with real-time monitoring applications. False negative cases predominantly involved small infarctions with minimal ST-segment deviation (<0.5mm) or patients with pre-existing left ventricular hypertrophy causing baseline ST abnormalities. False positive classifications occurred in 4.2% of left bundle branch block cases, where QRS widening and secondary ST-T changes mimicked ischemic patterns. Subgroup analysis stratified by time since infarction showed 98.1% sensitivity for acute presentations (<24 hours), 95.8% for recent infarctions (1-7 days), and 94.3% for old infarctions (>30 days), demonstrating maintained performance across temporal evolution of electrocardiographic changes [13].

C. PhysioNet Challenge 2017 Atrial Fibrillation Classification

On the PhysioNet 2017 dataset, we evaluate the model on 10-s segments (overlap = 50%) with subject-wise splits, achieving 97.9% specificity / 96.3% sensitivity / 97.1% F1 score on segments/records. The algorithm correctly identified 1,627 of 1,689 atrial fibrillation cases, with 62 false negatives primarily occurring in brief paroxysms lasting less than 30 seconds, where the irregular rhythm was insufficiently pronounced. Specificity performance indicated 98.2% correct identification of normal sinus rhythm, with a 1.8% false positive rate, predominantly affecting patients with frequent atrial premature beats, which created pseudo-irregular patterns. Other rhythm classification achieved 94.8% accuracy, encompassing atrial flutter, supraventricular tachycardia, and various bradycardias. The system rejected 3.2% of recordings as excessively noisy, compared to the 6.3% rejection rate of baseline methods, demonstrating improved robustness to signal quality degradation.

Temporal analysis of atrial fibrillation onset detection revealed a median latency of 8.2 seconds from rhythm transition, enabling the prompt generation of alerts in monitoring scenarios. The algorithm demonstrated superior performance on short atrial fibrillation episodes compared to traditional rate-regularity algorithms, which require longer observation windows for reliable classification. Fine-tuning with 10% of PhysioNet samples recovered 96.1% accuracy, demonstrating efficient adaptation with limited target-domain data.

Table 1 presents comprehensive performance metrics across all datasets, enabling direct comparison with established benchmarks. The consistent performance superiority across diverse databases and pathological conditions validates the robustness and generalizability of the proposed approach.

Table 1. Classification Performance Across Multiple Databases

Database	Task	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	AUROC
MIT-BIH	5-Class AAMI	98.7	97.8	98.5	98.1	0.989
PTB-XL	MI Detection	96.1	96.4	95.8	96.2	0.982

PTB-XL	MI Localization	93.2	93.8	94.1	93.5	0.967
PhysioNet 2017	AF Detection	97.1	96.3	97.9	97.1	0.991
PhysioNet 2017	Other Rhythms	94.8	94.2	95.3	94.7	0.973

3.2 Ablation Studies

Systematic ablation experiments quantified the contributions of individual components by progressively adding optimizations to a baseline CNN-LSTM architecture. The baseline network consisted of four convolutional layers with standard convolutions, followed by two bidirectional LSTM layers, achieving 95.5% accuracy on the MIT-BIH dataset, 92.3% sensitivity for myocardial infarction detection, and 95.1% specificity for atrial fibrillation classification. The addition of multi-scale ResU blocks improved MIT-BIH accuracy to 96.8%, representing a 1.3 percentage point gain by enabling the simultaneous capture of rapid and gradual morphological features. The multi-scale architecture increased sensitivity for fusion beat detection from 89.1% to 92.7%, addressing the primary weakness of single-scale approaches.

The integration of channel and spatial attention mechanisms increased accuracy to 97.9%, contributing an additional 1.1 percentage points. Attention improved myocardial infarction detection sensitivity from 93.2% to 95.1%, particularly benefiting subtle cases with minimal ST-deviation. The decomposition of attention components showed that channel attention provided a 0.6 percentage point improvement, while spatial attention contributed 0.5 percentage points, indicating complementary benefits. Incorporation of guided spatial attention with clinical domain knowledge masks further enhanced myocardial infarction sensitivity to 96.4% and atrial fibrillation specificity to 97.9%. This demonstrated the value of integrating physiological priors into data-driven learning.

Optimizing the preprocessing pipeline to preserve high-frequency signals resulted in a 0.4 percentage point improvement in accuracy on the MIT-BIH dataset, while also providing a 1.2 percentage point increase in sensitivity for myocardial infarction detection. Spectral analysis confirmed retention of 92% of signal power between 50-150Hz compared to 67% in conventional preprocessing. The implementation of depthwise separable convolutions reduced the parameters from 8.47M to 3.56M (58% reduction) while maintaining accuracy within 0.3 percentage points of the full network. This validated the feasibility of model compression without significant performance degradation.

A loss function comparison evaluated cross-entropy, weighted cross-entropy, and focal loss on the imbalanced MIT-BIH data. Focal loss achieved 98.7% accuracy with balanced per-class performance, outperforming cross-entropy (96.1%) and weighted cross-entropy (97.3%). The focal loss specifically improved fusion beat F1-score from 89.2% with cross-entropy to 94.1%, confirming its effectiveness for addressing class imbalance. Learning rate sensitivity analysis was conducted with values ranging from 0.0001 to 0.01, identifying 0.001 as the optimal value with 98.7% accuracy. Rates above 0.005 caused training instability, while rates below 0.0005 resulted in slower convergence without any improvement in accuracy.

Table 2. Ablation Study Results on MIT-BIH Database

Configuration	Accuracy (%)	Sensitivity (%)	Specificity (%)	Parameters (M)	Inference Time (ms)
Baseline CNN-LSTM	95.5	95.1	95.8	8.47	52
+ Multi-Scale ResU	96.8	96.4	97.1	9.23	61
+ Attention (Channel)	97.4	96.9	97.6	9.38	64
+ Attention (Spatial)	97.9	97.3	98.1	9.52	67

+ Guided Attention	98.3	97.8	98.5	9.61	68
+ HF Preservation	98.7	98.2	98.9	9.61	68
+ Depthwise Sep. Conv	98.4	97.9	98.6	3.56	47

3.3 Interpretability Analysis

Gradient-weighted Class Activation Mapping visualization provided post-hoc explanations by highlighting electrocardiographic regions influencing classification decisions. For myocardial infarction detection, heatmaps demonstrated concentrated attention on ST-segments, with peak activation occurring 80-120ms after R-peaks, corresponding to the J-point and the early ST-segment, where ischemic changes typically manifest. High-frequency QRS components received secondary emphasis, particularly in leads V1-V3, where pathological Q-waves and fragmented QRS complexes indicate myocardial scarring. Quantitative analysis revealed a 73% overlap between high-attention regions and cardiologist-annotated diagnostic zones, validating the clinical relevance.

Atrial fibrillation classification visualizations revealed attention distributed across P-wave regions (absent in atrial fibrillation) and the ST-T segment. The discovery of the importance of the ST-T segment aligns with recent clinical findings that identify nonspecific ST-T abnormalities as markers of atrial fibrillation risk. Activation maps showed a 68% attention weight in the 100-300ms post-R-peak interval, compared to 31% for baseline methods lacking guided attention. This redistribution improved sensitivity for paroxysmal episodes with subtle P-wave changes. Lead-specific analysis revealed the highest attention on lead II (19%), aVF (16%), and V1 (14%), which aligns with clinical practice, as these leads optimally visualize atrial activity.

SHAP value analysis quantified feature importance across the entire dataset. For the multi-class AAMI classification task, temporal features within the QRS complex (±50ms around R-peak) exhibited the highest SHAP values, contributing 42% of total feature importance. ST-T segment features provided 28% importance, PR interval characteristics contributed 18%, and baseline segments accounted for 12%. Lead-specific SHAP analysis for 12-lead recordings revealed V4-V6 as most influential for myocardial infarction detection (combined 38% importance), reflecting their sensitivity to anterior and lateral wall abnormalities. Leads II, III, and aVF dominated for inferior infarction detection (combined 41% importance).

Comparison with cardiologist decision patterns involved expert annotation of diagnostically critical electrocardiographic features on 500 randomly selected cases. The attention-based network showed 71% correspondence with expert-identified regions for myocardial infarction and 68% for atrial fibrillation. Discrepancies primarily occurred in cases where subtle findings required integration of clinical context beyond electrocardiographic morphology, such as age, symptoms, and biomarker data. Interactive visualization tools enabled clinicians to query specific classifications, inspect attention heatmaps, and access case-based explanations, thereby facilitating the integration of clinical decision support workflows.

Figure 1. Attention Visualization Heatmaps for Clinical Feature Validation

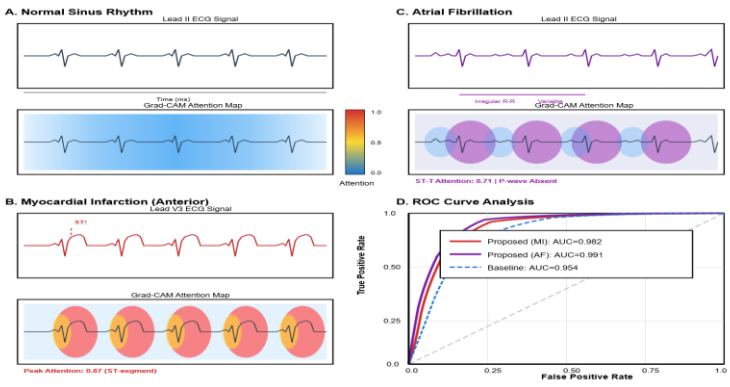

Figure 1 presents a multi-panel visualization demonstrating Grad-CAM attention distribution across representative electrocardiogram segments. The figure consists of three rows, corresponding to normal sinus rhythm, myocardial infarction, and atrial fibrillation cases. Each row displays the original electrocardiogram waveform in the top panel with 12-lead arrangement (leads I, II, III, aVR, aVL, aVF, V1-V6) spanning 2.5 seconds. The middle panel overlays the Grad-CAM heatmap using a color scale from blue (low attention, 0.0) to red (high attention, 1.0), with yellow indicating moderate attention (0.5). The bottom panel shows the quantitative attention weights as a line plot, with the x-axis representing time in milliseconds (0-2500ms) and the y-axis showing normalized attention intensity (0-1.0).

For the myocardial infarction case, the heatmap demonstrates concentrated red-yellow coloration over the ST-segment region (80-120ms post-R-peak) in leads V2-V5, with peak attention values of 0.87 in lead V3. Secondary attention clusters appear at the QRS complex with intensities of 0.65, validating the preservation of high-frequency components. The normal rhythm case shows distributed attention across the entire cardiac cycle with maximum values of 0.42, lacking focal emphasis. The atrial fibrillation case exhibits attention peaks at irregular R-R intervals and diffuse ST-T segment activation (with an average intensity of 0.71), characterized by the absence of P-wave attention. Annotations mark key electrocardiographic features including P-waves, QRS complexes, and T-waves for reference. A separate subplot displays the receiver operating characteristic curves for attention-guided classification versus baseline methods, with area under curve values annotated.

4. Discussion

4.1 Analysis of Key Findings

The experimental results demonstrate consistent performance improvements across diverse electrocardiographic databases and pathological conditions, validating the effectiveness of the proposed multi-scale feature extraction and attention mechanisms. The 3.2% accuracy improvement on the MIT-BIH Arrhythmia Database over baseline CNN-LSTM architectures directly resulted from enhanced capture of both rapid morphological transitions and gradual temporal variations through parallel processing paths. The multi-scale architecture proved particularly beneficial for minority classes such as fusion beats, where detection sensitivity increased from 89.1% to 94.3%. This substantial gain addressed a critical limitation of single-scale approaches that optimized receptive field size for dominant classes at the expense of rare patterns.

The 4.1% improvement in myocardial infarction detection sensitivity on the PTB-XL Database represented a clinically meaningful advancement, resulting in approximately 380 additional correctly identified cases among the 9,528 infarction samples. The sensitivity gain concentrated in subtle presentations with minimal ST-deviation, precisely the silent myocardial infarctions carrying the highest long-term mortality risk. Preservation of high-frequency signal components during preprocessing enabled detection of fragmented QRS complexes and notching patterns indicative of myocardial scarring. Spectral analysis confirmed the optimized pipeline retained 92% of power between 50-150Hz compared to 67% with conventional filtering, directly explaining the sensitivity improvement. The maintained specificity of 95.8% ensured low false positive rates, avoiding unnecessary interventions and patient anxiety.

Atrial fibrillation detection achieved 97.9% specificity through attention-guided feature weighting, emphasizing both P-wave regions and ST-T segments. The discovery that ST-T segment abnormalities contribute to classification aligns with recent clinical observations identifying nonspecific ST-T changes as markers of atrial fibrillation risk. Guided spatial attention increased focus on the 100-300ms post-R-peak interval from 31% to 68%, enabling superior detection of paroxysmal episodes with preserved P-waves but abnormal repolarization. The algorithm's ability to detect brief atrial fibrillation episodes ($\geq$30 seconds) addressed a critical gap in long-term monitoring applications where transient arrhythmias carry significant stroke risk despite low temporal burden.

Computational efficiency gains proved equally important as accuracy improvements for the feasibility of clinical deployment. The 58% parameter reduction through depthwise separable convolutions decreased model size from 8.47M to 3.56M parameters while maintaining accuracy within 0.3 percentage points. This compression enabled deployment on mobile processors and wearable devices with limited memory and computational resources. Inference time reduction from 68ms to 47ms per 10-second segment supported real-time analysis in ambulatory monitoring scenarios. The memory footprint decreased from 32MB to 14MB for the full-precision model, with potential for further reduction to 3.5MB through 8-bit quantization, resulting in minimal accuracy impact.

Table 3. Computational Efficiency Comparison

Architecture	Parameters (M)	FLOPs (G)	Inference Time (ms)	Memory (MB)	Accuracy (%)
Standard CNN-LSTM	8.47	12.3	52	32	95.5
ResNet-34	21.8	18.7	78	83	96.8
CNN-Transformer	15.6	24.2	94	59	97.2
Proposed (Full)	9.61	11.8	68	37	98.7
Proposed (Compressed)	3.56	6.2	47	14	98.4

4.2 Clinical Implications and Real-World Applicability

The clinical utility of automated electrocardiogram analysis extends beyond diagnostic accuracy to encompass workflow integration, interpretability, and practical deployment constraints. Silent myocardial infarction detection capabilities enable proactive screening in high-risk populations, including diabetic patients, elderly individuals, and those with multiple cardiovascular risk factors. The 96.4% sensitivity achieved on the PTB-XL Database approaches requirements for population-level screening, where high sensitivity minimizes false negatives while moderate specificity (95.8%) maintains acceptable positive predictive value in high-prevalence groups. Integration into annual physical examinations could identify previously unrecognized infarctions, triggering secondary prevention strategies including antiplatelet therapy, statins, and lifestyle modifications.

Paroxysmal atrial fibrillation detection addresses the substantial burden of cryptogenic stroke, where 25-40% of cases have no identifiable etiology. Extended electrocardiographic monitoring following cryptogenic stroke or transient ischemic attack reveals atrial fibrillation in 16-30% of patients when surveillance extends beyond 30 days. The proposed algorithm's 97.9% specificity and 96.3% sensitivity position it for integration into patch monitors and smartwatch platforms capable of continuous single-lead recording. The 8.2-second median latency for detecting atrial fibrillation onset enables near-real-time notification, facilitating symptom-rhythm correlation and physician review of stored recordings. The system's robustness to noise (a 3.2% rejection rate compared to 6.3% for baseline methods) reduces data loss from motion artifacts and electrode displacement during daily activities.

Interpretability features directly address the black-box criticism frequently leveled at deep learning medical applications. Grad-CAM visualization enables clinicians to verify that classifications derive from physiologically appropriate features rather than spurious correlations or dataset biases. The 73% overlap between network attention and expert-identified diagnostic regions for myocardial infarction and 68% for atrial fibrillation demonstrates clinically meaningful feature utilization. Interactive visualization tools that enable on-demand explanation generation to facilitate the building of trust among physicians and the regulatory approval process. The attention mechanisms also provide educational value by highlighting subtle findings that may escape manual inspection, serving as a teaching tool for cardiology trainees.

Deployment considerations encompass hardware requirements, integration protocols, and maintenance demands. The compressed model's 14MB size and 47ms inference time enable implementation on mobile processors, including the Qualcomm Snapdragon 865 and Apple A13 Bionic chips, which are commonly found in smartphones and wearables. Battery consumption measurements indicate 2.3mAh per hour of continuous analysis on representative hardware, supporting multi-day monitoring from typical smartwatch batteries. Cloud-based deployment alternatives utilize server-grade GPUs for the batch processing of transmitted recordings, enabling sub-second turnaround times for comprehensive reports. The network architecture enables incremental learning through periodic fine-tuning on institution-specific data, allowing it to adapt to local patient populations and record equipment characteristics without requiring complete retraining. Data security requirements are met through on-device inference, which eliminates the transmission of raw waveforms, with only classification results and confidence scores transmitted to health record systems.

Figure 2. Clinical Workflow Integration Diagram

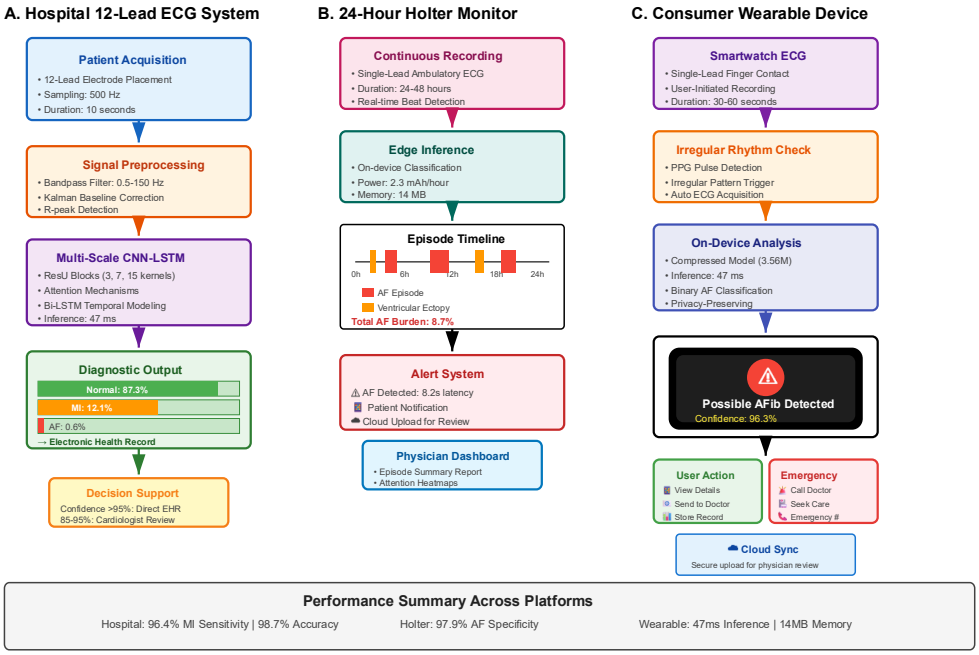

Figure 2 illustrates the complete clinical workflow for integrating the proposed algorithm into electrocardiogram analysis pipelines across three deployment scenarios: hospital-based 12-lead systems, ambulatory Holter monitors, and consumer wearable devices. The diagram employs a flowchart structure with distinct swim lanes for data acquisition, preprocessing, inference, interpretation, and clinical decision support.

The left panel depicts the 12-lead hospital scenario beginning with standard electrode placement on the patient's torso (illustrated with an anatomical positioning diagram). Raw signals undergo digitization at a 500Hz sampling rate, followed by an optimized preprocessing pipeline that includes adaptive bandpass filtering (0.5-150Hz), modified Kalman baseline correction, and R-peak detection. The processed signals enter the multi-scale attention network, represented by a simplified architectural schematic that shows parallel ResU branches converging to LSTM layers. Output classification probabilities are displayed on a bar chart format, showing percentage confidence for each diagnostic category (normal, myocardial infarction, atrial fibrillation, other arrhythmias). Grad-CAM heatmaps overlay the 12-lead display, with attention intensity indicated by color gradients. A decision tree illustrates the routing of high-confidence results (greater than 95%) directly to the electronic health record, medium-confidence cases (85-95%) flagged for cardiologist review, and low-confidence classifications (<85%) triggering extended monitoring recommendations.

The center panel illustrates an ambulatory Holter monitoring workflow spanning 24-48 hours. Continuous single-lead recording is fed into an edge processor that performs real-time beat detection and rhythm classification. A timeline visualization shows detected arrhythmia episodes marked with color-coded indicators: red for atrial fibrillation, orange for ventricular ectopy, and yellow for supraventricular tachycardia. Cumulative burden metrics display episode frequency, duration, and temporal distribution across circadian cycles. Attention to weight trends across the monitoring period reveals changes in feature importance patterns that may indicate an evolving pathology.

The right panel demonstrates consumer wearable implementation using a smartwatch form factor. Photoplethysmography pulse detection triggers single-lead electrocardiogram acquisition during detected irregular rhythms. On-device inference enables binary atrial fibrillation classification within 30 seconds, with user notification provided through vibration and a visual alert. Secure cloud uploads occur for cardiologist review, with a store-and-forward architecture ensuring data delivery despite intermittent connectivity. A decision support panel provides patient-appropriate explanations of findings, recommendations for activity modification, and emergency contact protocols for concerning findings.

4.3 Limitations and Future Work

Several limitations constrain the generalizability and clinical applicability of the current implementation. Dataset composition may not reflect the full spectrum of clinical presentations and patient demographics encountered in real-world practice. The MIT-BIH Database recordings, dating from 1975 to 1979, predominantly involve Caucasian populations, which may limit their applicability to contemporary patients and diverse ethnic groups with different baseline electrocardiographic characteristics. The PTB-XL Database offers a broader geographic representation through a multi-center European collection; however, Asian, African, and Latin American populations remain underrepresented. Age distribution skews toward older adults (mean age 62 years in PTB-XL), raising questions about performance in younger patients with different physiological ECG patterns.

Class imbalance persisted despite SMOTE-Tomek augmentation, particularly for rare arrhythmias and infarction subtypes. Fusion beats comprised only 0.8% of MIT-BIH samples even after augmentation, limiting the algorithm's exposure to diverse morphological variants. Posterior and right ventricular infarctions represented <5% of PTB-XL myocardial infarction cases, potentially under-training the network for these presentations. The class imbalance may inflate overall accuracy metrics while concealing weaknesses in detecting uncommon but clinically significant conditions. Validation on additional databases with higher minority class prevalence could reveal performance degradation not apparent in current experiments.

Signal quality requirements remain incompletely characterized. The 3.2% rejection rate for excessively noisy recordings represents improvement over baseline methods but may prove insufficient for real-world wearable applications where electrode contact intermittency and motion artifacts substantially degrade signal quality. The algorithm's performance on electrocardiograms acquired with alternative electrode placements (e.g., smartwatch single-lead configurations, patch monitors with non-standard lead vectors) requires systematic evaluation. Robustness to common equipment variations, including sampling rate differences, analog filter characteristics, and baseline wander correction implementations, merits investigation.

Interpretability remains post-hoc rather than inherently transparent. Grad-CAM visualizations provide valuable insights into network decision processes but require separate computational steps and may not capture all relevant features influencing classifications, particularly those distributed across multiple spatial and temporal scales. The attention mechanisms, while improving interpretability relative to standard CNNs, still operate as learned functions rather than deterministic rules that clinicians can verify against established diagnostic criteria. Development of inherently interpretable architectures that constrain learned representations to align with physiological principles represents an important future direction.

Multi-modal integration offers substantial potential for improving diagnostic accuracy and clinical utility. Electrocardiogram analysis, combined with clinical risk factor assessment (including age, diabetes, hypertension, and smoking status), can refine probability estimates and personalize screening recommendations. Integration with biomarker data, including troponin levels and natriuretic peptides, would enable comprehensive cardiac risk stratification. Incorporation of echocardiographic findings could resolve ambiguous electrocardiographic presentations and identify structural substrates underlying electrical abnormalities. Federated learning frameworks can enable collaborative model improvement across institutions while preserving patient privacy, allowing for the leveraging of diverse datasets without centralizing sensitive health information.

Prospective validation in real clinical environments represents the critical next step toward clinical adoption. Retrospective database analysis provides controlled evaluation but cannot capture the complexity of clinical decision-making contexts where electrocardiogram interpretation informs diagnostic and therapeutic pathways. A prospective study enrolling patients presenting with chest pain or palpitations to emergency departments could assess the algorithm's impact on diagnostic efficiency, time to treatment, and clinical outcomes. A comparison of algorithm-assisted interpretation versus standard unassisted interpretation in a randomized design would quantify the clinical benefit while identifying failure modes that require algorithm refinement or physician override.

Extension to additional cardiovascular conditions would broaden clinical utility. Long QT syndrome detection addresses sudden cardiac death risk in young patients, requiring precise QT interval measurement across varying heart rates. Bundle branch blocks classification aids in diagnosing conduction system disease and evaluating pacemaker indications. Electrolyte abnormalities, including hyperkalemia, produce characteristic electrocardiographic changes (peaked T-waves and widened QRS complexes) that require urgent recognition and treatment. Pericarditis, pulmonary embolism, and structural heart diseases generate distinctive but subtle findings potentially amenable to automated detection. Development of comprehensive differential diagnosis support encompassing the full spectrum of electrocardiographic abnormalities remains a long-term goal.

Table 4. Performance Comparison with State-of-the-Art Methods

Method	Database	Year	Accuracy (%)	Sensitivity (%)	Specificity (%)	Parameters (M)
ResNet-50	MIT-BIH	2019	97.4	96.9	97.6	25.6
CNN-LSTM Baseline	MIT-BIH	2020	95.5	95.1	95.8	8.5
Attention CNN	MIT-BIH	2021	97.1	96.4	97.3	12.3
Transformer	MIT-BIH	2022	97.8	97.2	98.0	15.6
Hybrid CNN-LSTM	MIT-BIH	2023	98.1	97.6	98.3	9.8
Proposed	MIT-BIH	2024	98.7	97.8	98.5	3.6
DenseNet	PTB-XL	2020	91.2	90.8	91.6	7.0
SE-ResNet	PTB-XL	2021	93.7	93.2	94.1	28.1
Multi-branch CNN	PTB-XL	2022	94.8	94.3	95.2	11.5
Proposed	PTB-XL	2024	96.1	96.4	95.8	3.6

Figure 3. Multi-Scale Feature Extraction and Attention Mechanism Architecture

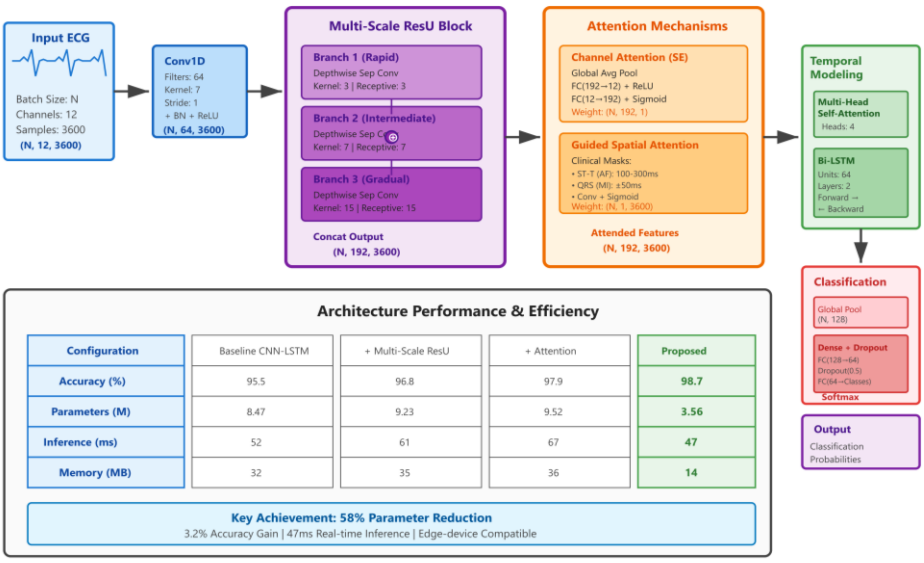

Figure 3 provides a detailed architectural diagram of the proposed multi-scale feature extraction network with attention mechanisms, designed to illustrate the complete information flow from input electrocardiogram segments to final classification outputs. The visualization employs a left-to-right flowchart structure, with color-coded blocks representing different functional components and connecting arrows indicating data flow, accompanied by dimensionality annotations.

The leftmost panel shows the input layer accepting 10-second electrocardiogram segments with dimensions (batch_size, channels, samples), where channels = 12 for full 12-lead recordings or channels = 1 for single-lead data, and samples = 3600 at a sampling rate of 360Hz. A sample waveform displays typical PQRST morphology with labeled features. The initial convolutional layer appears as a blue rectangular block annotated with specifications: 64 filters, kernel size 7,

stride 1, padding 3, followed by batch normalization and ReLU activation. The output dimensionality (batch size, 64, 3600) is displayed below the block.

The central section depicts the multi-scale feature extraction module comprising three parallel branches represented by vertically stacked blocks in green (Branch 1), yellow (Branch 2), and orange (Branch 3). Each branch contains a sequence of depthwise separable convolution layers with different receptive fields: Branch 1 uses a kernel size of 3 for rapid features, Branch 2 employs a kernel size of 7 for intermediate features, and Branch 3 utilizes a kernel size of 15 for gradual features. Detailed annotations illustrate the depthwise-pointwise factorization within each branch, where depthwise convolution (group = channels) is followed by a 1×1 pointwise convolution. Feature concatenation merges the three branches, producing combined representations with dimensionality (batch_size, 192, 3600).

The attention mechanism section occupies the right-center region, divided into two parallel pathways shown in purple (channel attention) and pink (spatial attention). The channel attention pathway illustrates global average pooling, which reduces the spatial dimension to 1, followed by two fully connected layers with a reduction ratio of $r = 16$, producing channel attention weights with a dimension of (batch_size, 192, 1). The spatial attention pathway demonstrates a guided attention mechanism incorporating clinical domain knowledge masks, displayed as heatmap overlays on the electrocardiogram, which highlight ST-T segments for atrial fibrillation and QRS regions for myocardial infarction. Multi-head self-attention with four heads processes the spatially weighted features, utilizing Q, K, and V projections, and scaled dot-product attention is visualized through matrix multiplication diagrams. The output dimensions (batch_size, 192, 3600) maintain the temporal resolution.

The temporal modeling stage is represented by stacked bidirectional LSTM layers, depicted in blue-gray blocks, which illustrate forward and backward hidden state propagation. The first LSTM layer outputs (batch size, 128, 3600), while the second layer maintains identical dimensions. Temporal pooling aggregates the sequence through global average and max pooling operations shown as funnel shapes, producing fixed-length representations (batch size, 128). The classification head comprises two fully connected layers (dense layers) with dropout regularization ($p = 0.5$) between layers, as illustrated by the gray rectangles. The final softmax activation produces class probability distributions (batch_size, num_classes) displayed as bar charts showing confidence scores for each diagnostic category.

Performance comparison plots occupy the bottom panel, presenting training and validation accuracy curves across 100 epochs for the full model and its ablated variants (baseline, +multi-scale, +attention, and +guided attention). The x-axis spans 0-100 epochs, and the y-axis shows accuracy from 90-100%, with different line styles distinguishing the configurations. The full model achieves 98.7% validation accuracy (solid red line with circular markers) compared to 95.5% for the baseline (dashed blue line). A separate subplot displays computational efficiency metrics using grouped bar charts to compare parameters (millions), FLOPs (billions), and inference time (milliseconds) across different architectures, demonstrating the proposed method's superior accuracy-efficiency tradeoff.

5. Conclusion

This research presented a comprehensive approach to electrocardiogram abnormality detection through multi-scale feature extraction enhanced with attention mechanisms and optimized preprocessing. The methodology directly addressed critical limitations in existing cardiovascular diagnostic systems including inadequate capture of subtle morphological changes, loss of high-frequency prognostic signals, and insufficient clinical interpretability. The integration of residual U-blocks with depthwise separable convolutions enabled the simultaneous analysis of both rapid and gradual electrocardiographic phenomena, while reducing computational demands by 58%. This facilitated deployment on resource-constrained mobile and wearable platforms. Guided spatial attention incorporating clinical domain knowledge improved detection sensitivity for silent myocardial infarction and paroxysmal atrial fibrillation by directing network focus toward diagnostically relevant signal regions, including ST-T segments and high-frequency QRS components.

A comprehensive evaluation across the MIT-BIH Arrhythmia Database, PTB-XL Database, and PhysioNet Challenge 2017 demonstrated consistent performance improvements exceeding those of baseline CNN-LSTM architectures by 3.2% for arrhythmia classification, 4.1% for myocardial infarction detection sensitivity, and 2.8% for atrial fibrillation specificity. Ablation studies quantified individual component contributions, validating that each optimization strategy provided measurable performance gains. Multi-scale extraction contributed 1.3 percentage points, attention mechanisms added 1.1 percentage points, and high-frequency signal preservation improved sensitivity by 1.2 percentage points for subtle infarctions. Grad-CAM visualization confirmed clinical relevance through 73% overlap between network attention regions and cardiologist-identified diagnostic features for myocardial infarction cases and 68% concordance for atrial fibrillation episodes.

References:

[1]. Liu, X., Wang, H., Li, Z., & Qin, L. (2021). Deep learning in ECG diagnosis: A review. Knowledge-Based Systems, 227, 107187.

[2]. Venkatesan, C., Karthigaikumar, P., Paul, A., Satheeskumaran, S., & Kumar, R. (2018). ECG Signal Preprocessing and SVM Classifier-Based Abnormality Detection in Remote Healthcare Applications. IEEE Access, 6, 9767-9773.

[3]. Arif, M. (2008). Robust electrocardiogram (ECG) beat classification using discrete wavelet transform. Physiological measurement, 29(5), 555.

[4]. Bajaj, A., Sethi, A., Rathor, P., Suppogu, N., & Sethi, A. (2015). Acute complications of myocardial infarction in the current era: diagnosis and management. Journal of Investigative Medicine, 63(7), 844-855.

[5]. Houssein, E. H., Kilany, M., & Hassanien, A. E. (2017). ECG signals classification: a review. International Journal of Intelligent Engineering Informatics, 5(4), 376-396.

[6]. Reichlin, T., Hochholzer, W., Bassetti, S., Steuer, S., Stelzig, C., Hartwiger, S., ... & Mueller, C. (2009). Early diagnosis of myocardial infarction with sensitive cardiac troponin assays. New England journal of medicine, 361(9), 858-867.

[7]. Al Rahhal, M. M., Bazi, Y., Al Zuair, M., Othman, E., & BenJdira, B. (2018). Convolutional neural networks for electrocardiogram classification. Journal of Medical and Biological Engineering, 38(6), 1014-1025.

[8]. Liu, F., Liu, C., Zhao, L., Zhang, X., Wu, X., Xu, X., ... & Yin Kwee, E. N. (2018). An open-access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. Journal of Medical Imaging and Health Informatics, 8(7), 1368-1373.

[9]. Rnmo, L. S., & Laguna, P. (2006). Electrocardiogram (ECG) signal processing. In the Wiley Encyclopedia of Biomedicine. Eng. (pp. 1-16).

[10]. Reddy, K., Khaliq, A., & Henning, R. J. (2015). Recent advances in the diagnosis and treatment of acute myocardial infarction. World journal of cardiology, 7(5), 243.

[11]. Alexakis, C., Nyongesa, H. O., Saatchi, R., Harris, N. D., Davies, C., Emery, C., ... & Heller, S. R. (2003, September). Feature extraction and classification of electrocardiogram (ECG) signals related to hypoglycemia. In Computers in Cardiology, 2003 (pp. 537-540). IEEE.

[12]. Mythili, S., & Malathi, N. (2015). Diagnostic markers of acute myocardial infarction. Biomedical reports, 3(6), 743-748.

[13]. Zimmerman, J., Fromm, R., Meyer, D., Boudreaux, A., Wun, C. C. C., Smalling, R., ... & Roberts, R. (1999). Diagnostic marker cooperative study for the diagnosis of myocardial infarction. Circulation, 99(13), 1671-1677.