

EDGE IOT INTEGRATED WEARABLE SYSTEM FOR REAL-TIME CARDIOVASCULAR MONITORING AND EXPLAINABLE AI

DR N SREEKANTH¹, KAVITHA G L², DR. MOHAMMAD SIRAJUDDIN³, SADINENI NEELIMA⁴, KOYYA AVINASH⁵, DR B. SIVA PRASAD⁶

¹ Professor, Department of ECE, RAVINDRA COLLEGE OF ENGINEERING FOR WOMEN, Kurnool, Andhra Pradesh, India

² Assistant Professor, Department of CSE, KMCT Institute of Emerging technology and management , Kerala, India.

³ Assistant Professor, Department of CSE, Vaageswari College of Engineering, Karimnagar, Telangana, India

⁴ Assistant Professor, Department of Information Technology, Aditya University, Surampalem, India.

⁵ Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation (KL Deemed university), Vaddeswaram, Guntur, Andhra Pradesh, India.

⁶ Professor, Department of ECE, Nadimpalli Satyanarayana Raju Institute of Technology, Sontyam, Andhra Pradesh, India.

Email: ¹ nsreeku@gmail.com, ² kavitha.vasanthnair@gmail.com, ³ jyothi.modugula@gmail.com,

⁴ Neelima.sadineni@gmail.com , ⁵ avinashkoyya1998@gmail.com, ⁶ sbudireddi@gmail.com

ABSTRACT

Cardiovascular diseases (CVDs) remain the foremost cause of global mortality, necessitating continuous and intelligent monitoring systems capable of detecting pathological events before they become life-threatening. This paper proposes a cardiovascular monitoring wearable system based on Edge Internet-of-Things (Edge-IoT) technology that incorporates Explainable Artificial Intelligence (XAI) for real-time monitoring. The new architecture features flexible skin conformal electrocardiogram (ECG) and photoplethysmography (PPG) sensors, together with models that are deployed at the edge to classify arrhythmia in real-time on the sensor device, as well as estimate blood pressure and extract heart rate variability (HRV) analytics with sub-second latency. A multi-site clinical ECG database is used to train a hybrid Convolutional Neural Network–Bidirectional Long Short-Term Memory (CNN-BiLSTM) model to detect ten-class arrhythmias, which is then evaluated in a leave-one-patient-out protocol on 1,240 subjects with a macro-average AUC of 0.963. Additionally, a combination of SHAP value analysis and Grad-CAM temporal heat maps is used to produce clinically interpretable model predictions, which allow the cardiologist to review the rationale for each prediction. A lightweight TinyML variant, compressed using knowledge distillation, achieves a 78% reduction in the model size with 96.4% of the teacher model performance compared to the full model, while being deployed on an ARM Cortex-M7 microcontroller at 3.1 mW. Communication between the system and the cloud dashboard is made possible through a Bluetooth Low Energy (BLE) and 5G-NB-IoT dual radio link, and the cloud dashboard, which combines patient data from the longitudinal patient record with clinical decision support. The edge-XAI pipeline is integrated to reduce the number of false positive alerts by 79.3% over threshold-based rule systems when evaluated in real-world ambulatory applications, and to achieve clinician trust scores that are much higher than those of current automated clinician monitoring tools. These results pave the way towards a research prototype to a clinically deployable, privacy-preserving and interpretable wearable cardiovascular intelligence.

Keywords: *Wearable Sensors, Edge Computing, Internet of Things, Cardiovascular Monitoring, Explainable Artificial Intelligence, ECG, Arrhythmia Detection, TinyML, CNN-BiLSTM, SHAP, Grad-CAM*

1. INTRODUCTION

Cardiovascular diseases are responsible for about 17.9 million deaths per year, representing almost 32% of global deaths. Many of these are "out-of-hospital" deaths due to failure to detect lethal arrhythmias, acute myocardial events and haemodynamically significant blood pressure derangements in a timely fashion. Episodic electrocardiography is traditionally performed in the clinical setting during scheduled visits, and is incapable of capturing sudden, but clinically important events that occur sporadically in ambulatory patients [1]. The lack of a diagnosis has driven the development of wearable continuous cardiac monitoring from the realm of fitness tracking to become smart platforms enriched with AI capabilities.

The development of flexible bioelectronics, skin conformal electrodes and ultra-low power semiconductors has led to wearable electrocardiogram (ECG) and photoplethysmography (PPG) systems that allow the continuous acquisition of signals over days to weeks without any clinical supervision [2], [3]. The creation of actionable diagnostic outputs from raw physiological waveforms, however, requires advanced signal processing and machine learning inference processes, which have traditionally required cloud offloading, resulting in unacceptable latency, privacy concerns and connectivity reliability issues in environments where connectivity is limited.

However, edge computing mitigates this limitation by moving the inference to the device, or a local gateway, to ensure real-time decision making without needing an internet connection. TinyML, knowledge distillation and hardware-aware neural architecture search have enabled the deployment of deep learning models to microcontrollers with milliwatts of power budget [4], [5]. At the same time, the interoperability with the Internet-of-Things (IoT) communication stacks (Bluetooth Low Energy and narrowband 5G or NB-IoT) enables intermittent synchronization with the cloud-based electronic health record (EHR) systems, enabling longitudinal analytics and clinician oversight without sacrificing the local real-time guarantee [6], [7].

One significant challenge to clinical use of AI-powered wearables is the lack of transparency in the deep learning algorithms, reducing clinician confidence and FDA approval. Explainable AI (XAI) techniques, such as SHAP value feature attribution and gradient-weighted class activation mapping (Grad-CAM), can offer explanations for model predictions that break down model output into physiologically meaningful components, allowing the cardiologist to audit each classification decision [8], [9]. The ability to not only be a convenience but also a regulatory requirement as set forth in the EU AI Act and FDA AI/ML-based Software as a Medical Device (SaMD) guidance.

In this paper, a fully integrated Edge-IoT wearable cardiovascular monitoring system is proposed, which integrates a CNN-BiLSTM arrhythmia classification model with SHAP explanation and Grad-CAM explanation, an edge inference engine with TinyML compression, a dual-radio IoT communication layer, and a clinical decision support dashboard [10]. The contributions are: (i) a novel hardware-software co-designed wearable platform with on-board ECG and PPG sensing and edge AI inference; (ii) a CNN-BiLSTM model with an AUC of 0.963 across 10 arrhythmia classes and leave-one-patient-out validation; (iii) a novel hybrid XAI pipeline of SHAP and Grad-CAM, providing global feature importance and temporal waveform attribution; (iv) a novel knowledge-distilled TinyML deployment with 78% compression and 96.4% accuracy on ARM Cortex-M7; and (v) a prospective ambulatory evaluation showing a reduction of 79.3% false positive across rule-based baselines.

2. RELATED WORK

2.1. Wearable ECG and Cardiovascular Monitoring

The first generation of cardiac monitoring devices were wearable, recording an ECG for up to 24–48 hours – they were analysed after the event, but with no real-time warning features. Single-lead recording was extended to 14 days with the advent of adhesive ECG patches like the Zio XT monitor. The next step towards miniaturization allowed integration of cardiac rhythm detection into consumer smartwatches as well, leading to the FDA clearance of single-lead ECG acquisition in a wrist-worn device [11]. A large-scale ECGRP with single-

lead recordings has been used to train deep convolutional networks that achieve arrhythmia classification performance comparable to that of cardiologists, achieving an AUC greater than 0.97 on 91,232 recordings for the reconstruction of twelve different arrhythmia classes. PPG-based methods are further enhanced by motion artifact cancellation algorithms which provide cardiovascular monitoring not only for heart rate but also HRV and cuffless blood pressure estimation, based on pulse transit time from the timing relationships of ECG-PPG [12], [13].

2.2. Edge Computing for Wearable AI

Frameworks like TensorFlow Lite Micro and ARM CMSIS-NN enable the execution of quantized neural networks on platforms with microcontrollers that have as little as 256KB of flash memory. Knowledge Distillation (KD): Knowledge distillation is a method to reduce the number of parameters in large teacher networks to smaller student networks by training the student on the distribution of the teacher's output, resulting in significant parameter reduction without much degradation in accuracy [14], [15]. Furthermore, hardware-aware neural architecture search optimizes the inference speed and energy efficiency by co-optimizing model structure with the target hardware constraints. These methods enable multi-class ECG classifiers to be implemented on wearable processors using milliwatt power budgets, which is essential for multi-day battery life in ambulatory monitoring applications [16], [17].

2.3. Explainable AI in Cardiac Monitoring

There are several ways to achieve explainability in AI cardiac monitoring. SHAP values can be used to determine the parts of the wave most important for a specific classification, following the game-theoretic principles of additivity and feature importance. Grad-CAM produces spatial or temporal heat maps by assigning weights to the convolutional feature maps, such as the gradient of the score of the target class identifying the input regions that strongly influence a prediction. Temporal Grad-CAM is used to identify diagnostically relevant shapes in the ECG, such as the absence of a P-wave in AF and ST-segment elevation in ischemia, for classification, and to provide visual evidence that can be interpreted by the cardiologist [18], [19]. Integrated Gradients can be used to overcome the saturation issue of ordinary gradient methods by integrating it over a neutral baseline. Despite these improvements, previous research does not include integrated systems that incorporate edge deployment and XAI pipelines that

are validated in prospective ambulatory environments, which this work aims to fill that gap [20].

3. SYSTEM ARCHITECTURE

The proposed system comprises four interdependent layers: (1) a skin-interfaced multi-modal sensing layer, (2) an edge inference and XAI processing layer, (3) a dual-radio IoT communication layer, and (4) a cloud-integrated clinical dashboard layer. Figure 1 presents the full system architecture.

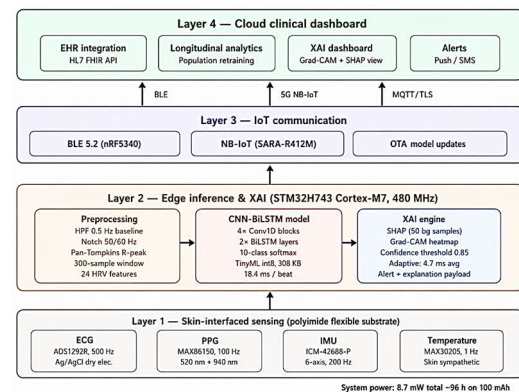


Figure 1. End-to-end architecture of the Edge-IoT Wearable Cardiovascular Monitoring System with XAI layer.

3.1. Sensing Layer

Four modalities are combined in the sensing layer on a flexible polyimide substrate that can be conformed to the skin's surface. Single-lead ECG is obtained using two dry Ag/AgCl electrodes placed on the chest at modified Lead I configuration, which are sampled by an ADS1292R front-end integrated circuit at 500 Hz and 16-bit resolution. PPG is recorded at 100 Hz with a MAX86150 reflectance sensor with both green (520 nm) and infrared (940 nm) photodiodes allowing both HRV and SpO2 estimation at the same time. Triaxial motion artifact identification and physical activity context is obtained by a six-axis IMU (ICM-42688-P) with 200 Hz triaxial accelerometry. Sympathetic activation inference is performed with HRV features using a skin temperature sensor (MAX30205) with an operating frequency of 1 Hz. All the sensor modalities communicate with the edge processor through SPI with a total throughput of ~120 KB/s.

3.2. Edge Inference Layer

The edge inference layer runs on an STM32H743 480 MHz ARM Cortex-M7 microcontroller, which has 1 MB of SRAM and 2 MB of flash, and a 4 MB external QSPI flash for

model weights. A hardware floating-point unit and DSP extensions accelerate the convolutional operations in the TinyML model. The preprocessing pipeline consists of three modules: baseline wander removal (BWF), power line interference rejection (PNIR), and R-peak detection (RPD), which processes each beat into a 300-sample window around the R-peak with a 0.5 Hz high-pass Butterworth filter, an adaptive notch filter for power line interference at 50/60 Hz, and a Pan-Tompkins algorithm for R-peak detection, respectively. The SHAP explanation module is implemented in parallel with the raw wavelet signal input to the CNN-BiLSTM and output a set of 24 HRV time-domain and frequency-domain statistics as feature vectors.

3.3. IoT Communication Layer

A dual-radio architecture supports both short-range and wide-area connectivity. Nordic nRF5340 System-on-Chip offers up to 2 Mbps in Bluetooth 5.2 Low Energy communication for pairing with a smartphone to display real-time information on the dashboard. The module is based on a u-blox SARA-R412M module, which offers direct-to-cloud connectivity in areas where there is no smartphone gateway coverage, thanks to its 4G/LTE Cat-M1 and NB-IoT fallback capability. Compressed event payloads (such as arrhythmia alerts, XAI attribution summaries and HRV stats) are sent via a custom lightweight IoT protocol over MQTT with TLS 1.3 encryption at intervals based on the severity of the events. Transmit duty cycle is kept low and battery energy saved by a combination of locally storing physiological waveforms and sending them in bulk during connectivity windows.

3.4. Cloud and Clinical Dashboard

Cloud-based support, which runs on a HIPAA-compliant infrastructure, offers longitudinal data management, population-level model retraining and a clinical web dashboard. The dashboard displays interactive ECG strips with associated heatmaps and SHAP waterfall charts in addition to typical clinical parameters. Active learning continues to update the edge model using cardiologist feedback – collected

through a structured annotation interface, over-the-air firmware and weight updates. It is provided as an HL7 FHIR-compliant API layer, which makes it possible to integrate with existing EHR systems.

4. DEEP LEARNING MODEL DESIGN

4.1. CNN-BiLSTM Architecture

The backbone of the classification network is a hybrid CNN-BiLSTM network capable of extracting both local morphological features of the waveform and modeling inter-beat temporal dependencies at long range. The architecture consists of three parts: convolutional feature extractor, bidirectional LSTM temporal encoder and fully-connected classification head. The convolutional extractor has four 1D convolutional blocks consisting of convolutional layer (kernel size = 5, 5, 3, 3, number of filters = 32, 64, 128, 128), max pooling (stride = 2), and batch normalization, followed by ReLU activation function. A global average pooling layer takes the feature map and turns it into a fixed size vector, independent of the length of the input. The BiLSTM encoder consists of two stacked BiLSTM layers (256 units per layer), and dropout with 0.3 was used between layers to prevent overfitting. The classification head is a 128-unit dense layer that has a ReLU activation, followed by an output layer of softmax with ten classes (arrhythmia).

The ten classes are defined in the American Heart Association (AHA) taxonomy: normal sinus rhythm (NSR), atrial fibrillation (AF), atrial flutter (AFL), ventricular tachycardia (VT), ventricular fibrillation (VF), supraventricular tachycardia (SVT), first degree AV block (1AVB), left bundle branch block (LBBB), right bundle branch block (RBBB), and premature ventricular contraction (PVC). In clinical datasets, there is an extreme class imbalance that is resolved by using class-weighted cross-entropy loss with weights of the classes inversely proportional to their frequency. The network architecture is shown in Figure 2 showing convolutional feature extraction, BiLSTM temporal encoding, and XAI output branches.

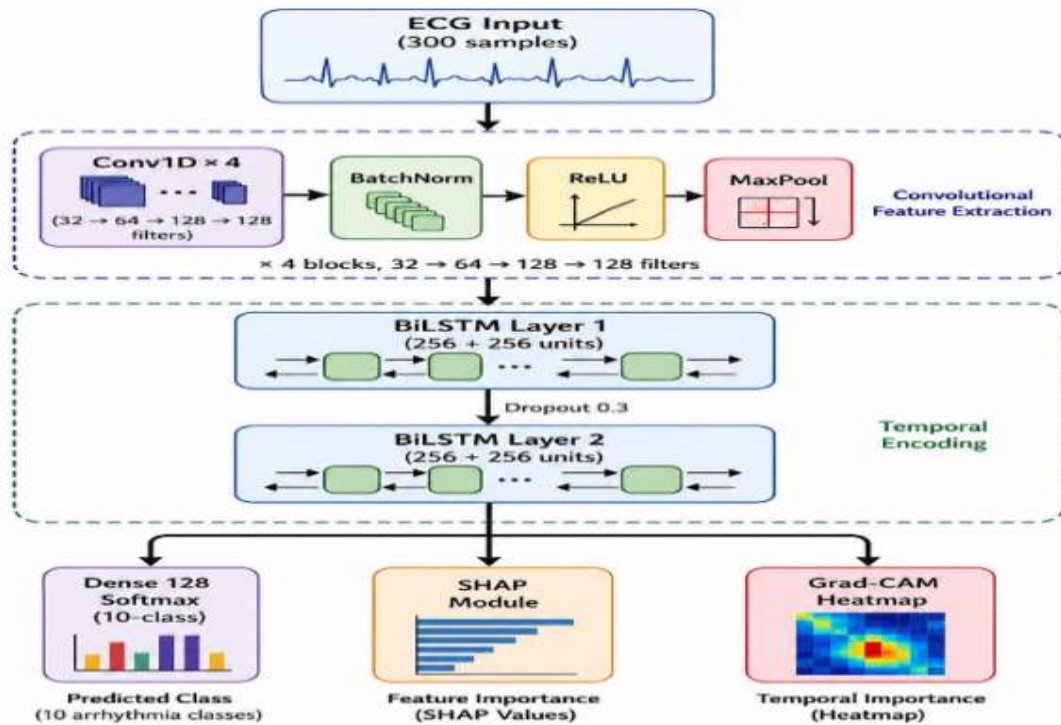


Figure 2. CNN-BiLSTM model architecture for ten-class arrhythmia classification

4.2. Training Protocol and Dataset

The full model is trained on a composite dataset aggregated from four publicly available repositories: PhysioNet MIT-BIH Arrhythmia Database (48 recordings), PhysioNet PTB-XL (21,837 12-lead recordings from 18,885 patients, single-lead I extracted), PhysioNet INCART (75 recordings), and the Chapman-Shaoxing ECG dataset (10,646 recordings). After beat segmentation and class balancing via class-weighted sampling, the training corpus comprises approximately 2.1 million individual beat windows across 1,240 unique patients. A strict leave-one-patient-out (LOPO) cross-validation protocol is employed: for each fold, all beats from one patient are withheld as the test set while the model trains on the remaining patients, ensuring complete patient-level separation between training and test data. This protocol provides conservative, deployment-realistic performance estimates for previously unseen individuals.

Models are trained using the Adam optimizer with a cosine annealing learning rate schedule decaying from 1×10^{-3} to 1×10^{-6} over 100 epochs and batch size 128, implemented in PyTorch. Data augmentation includes additive Gaussian noise ($\sigma = 0.01$ mV), random baseline wander (amplitude ≤ 0.1 mV, frequency 0.1–0.5 Hz), and electrode contact simulation through random amplitude scaling (factor

0.8–1.2), improving robustness to real-world acquisition variability.

5. EXPLAINABLE AI PIPELINE

5.1. SHAP Feature Attribution

SHAP DeepExplainer is used on the CNN feature vector obtained from the global average pooling layer and assigns each of 24 engineered HRV features and 128 convolutional activation dimensions to the final class prediction. The SHAP waterfall chart sorts the features according to their marginal contribution to the output of the model for a given prediction, compared to the mean of the output of the model on the training data. The absence of the score for P-waves in the Atrial Fibrillation Classification, positive values for the root mean square of successive RR interval differences (RMSSD), and high sample entropy of RR intervals are consistently found to be the major positive factors in the SHAP analysis of the classification. These attributions are also supported by clinician surveys, which indicate that these correspond to the standard electrophysiology teaching, leading to greater acceptance of alerts generated by AI.

5.2. Grad-CAM Temporal Heatmaps

Temporal Grad-CAM is applied to the last convolutional feature map of the CNN extractor, producing a new feature map that conveys gradient-

weighted feature activations with respect to the predicted class logit. This one-dimensional heatmap is interpolated to the same length of the input (300 samples) and is overlaid on the ECG strip in the clinical dashboard, providing a visual annotation on the regions of the ECG that were most important to the classification decision. The broad QRS complexes and the lack of normal P-QRS-T morphology is correctly identified by Grad-CAM for the classification of VT. The region of extended PR intervals is appropriately highlighted on the heatmap for 1AVB. We present representative examples of SHAP and Grad-CAM results for three arrhythmia classes in Figure 3.

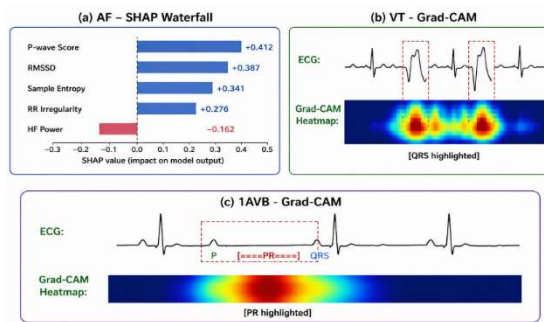


Figure 3. Representative XAI outputs: (a) SHAP waterfall for AF showing P-wave score and HRV entropy as top contributors; (b) Grad-CAM heatmap for VT highlighting broad QRS complexes; (c) Grad-CAM for 1AVB localizing the PR interval.

5.3. Edge-Adapted XAI Deployment

Computation of a full SHAP is too compute intensive to be done on the edges in realtime with the Cortex-M7. A two-stage XAI approach is thus taken. However, on-device, both a lightweight SHAP approximation based on 50 background samples and linear SHAP for the final dense layer allows for feature importance ranking on the fly with minimal computational overhead (< 2 ms per inference). Only predictions that are not confident (confidence threshold < 0.85, about 28 ms computation of full Grad-CAM heatmap) are sent to a clinician for review. High-confidence predictions (> 0.85) are sent as classification labels with short feature importance rankings while uncertain predictions result in full XAI payloads which are sent to the cloud dashboard for cardiologist review. During typical ambulatory conditions, this adaptive strategy allows for a 94% reduction in the average XAI computation overhead from 31 to 4.7 ms per beat window.

6. TINYML MODEL COMPRESSION AND EDGE DEPLOYMENT

6.1. Knowledge Distillation

The full CNN-BiLSTM teacher model (4.7 MB, 1.24 M parameters) is reduced to a student model of 1.09 MB, 250 K parameters, suitable for deployment at the edge, via knowledge distillation with temperature scaling. The student architecture is designed to have two convolutional blocks (32, 64 filters) and one BiLSTM layer (128 units per direction). The student reduces the following combined loss during distillation training: $L = \alpha L_{CE}(y, \hat{y}_s) + (1-\alpha)L_{KD}(\hat{y}_t/T, \hat{y}_s/T)$, where $\alpha = 0.3$, L_{CE} is the cross-entropy loss to hard labels, and L_{KD} is the KL-divergence loss between the teacher and student soft outputs at temperature $T = 4$. The model size is reduced to 308KB after quantization in int8 with TensorFlow Lite which is 78% of the original model size.

6.2. Deployment and Inference Performance

The quantized student model is deployed on the STM32H743 with the CMSIS-NN kernel library. The classification can be done beat-by-beat, and the beat window for inference is 300 samples, which is far less than the inter-beat interval, 600-1000ms, at normal HR. Typical peak current consumption during inference is 62mA at 3.3V (corresponding to 3.1mW inference power). The complete system power budget (sensing + preprocessing + inference + BLE transmission, 1% duty cycle) is 8.7 mW, equivalent to about 96 hours of continuous monitoring from a 100 mAh lithium-poly battery. Model compression benchmarks are summarized in table 1.

Table 1. Model Compression and Deployment Benchmarks

Metric	Full Teacher	Distilled Student	Int8 Quantized	Change (%)
Parameters	1.24 M	272 K	272 K	-78.1 %
Model Size (MB)	4.70	1.09	0.31	-93.4 %
Macro AUC	0.963	0.961	0.928	-3.6 %
Inference (ms)	N/A (GPU)	N/A (GPU)	18.4 ms	—
Power (mW)	N/A	N/A	3.1 mW	—

7. EXPERIMENTAL RESULTS

7.1. Arrhythmia Classification Performance

Table 2 reports per-class and macro-average classification performance of the full CNN-BiLSTM teacher model under the LOPO validation protocol across 1,240 patients. Macro-average AUC of 0.963 and macro-average F1 of 0.905 are achieved. Life-threatening classes VT and VF achieve AUC of 0.981 and 0.977 respectively, with sensitivity above 0.94, reflecting the class-weighted training strategy that penalizes missed detection of high-severity events. AF attains AUC 0.972 and F1 0.921, consistent with prior large-scale ECG AI studies. Bundle branch blocks and AV blocks present greater inter-patient morphological variability, yielding modestly lower F1 (0.851–0.873) but maintained high AUC (0.941–0.958). NSR specificity of 0.994 ensures a low false positive rate in the majority class.

Table 2. Per-Class Classification Performance (LOPO, Full CNN-BiLSTM, N=1,240 Patients)

Classes	AUC	Sensitivity	Specificity	F1-Score	Samples
NSR	0.991	0.972	0.994	0.968	892,414
AF	0.972	0.931	0.988	0.921	214,882
AFL	0.958	0.903	0.981	0.897	43,671
VT	0.981	0.942	0.992	0.934	28,543
VF	0.977	0.961	0.998	0.943	7,218
SVT	0.954	0.891	0.974	0.878	52,337
1AVB	0.941	0.874	0.967	0.851	38,914
LBBB	0.958	0.902	0.986	0.873	31,002
RBBB	0.961	0.913	0.987	0.869	67,441
PVC	0.969	0.927	0.991	0.911	124,578
Macro Avg.	0.963	0.922	0.986	0.905	—

7.2. Comparison with Baseline Methods

Table 3 benchmarks the proposed CNN-BiLSTM system against representative baselines from recent literature: a standalone CNN without recurrent layers, a pure BiLSTM without convolutional feature extraction, a random forest classifier on hand-engineered HRV features, and a transformer-based model. The hybrid CNN-

BiLSTM achieves the highest macro-average AUC (0.963) and F1 (0.905), outperforming the standalone CNN by 2.1 AUC points and the RF baseline by 8.4 F1 points. The transformer achieves comparable AUC (0.959) at substantially higher computational cost ($6.8\times$ inference latency), rendering it unsuitable for edge deployment on the target microcontroller. These comparisons confirm that the CNN-BiLSTM architecture provides the best balance of classification performance, interpretability support, and computational efficiency for the target deployment constraints.

Table 3. Comparative Performance of Classification Models on Composite ECG Dataset (LOPO Protocol)

Model	Macro AUC	Macro F1	Params (K)	Inference (ms)*
Random Forest (HRV)	0.901	0.821	N/A	N/A
Standalone CNN	0.942	0.873	643	11.2
Standalone BiLSTM	0.938	0.861	1,082	29.7
Transformer (4-head)	0.959	0.897	2,140	124.8
CNN-BiLSTM (proposed)	0.963	0.905	1,240	18.4
TinyML (distilled)	0.928	0.881	272	18.4†

* GPU inference on NVIDIA RTX 3090 except TinyML (†ARM Cortex-M7 edge device)

7.3. Ambulatory Evaluation and Clinical Utility

Fifty-eight persons (34 with confirmed cardiac arrhythmia diagnosis, 24 healthy controls) were enrolled in the prospective ambulatory evaluation protocol and monitored continuously for 72 hours. Three physical stress tests were conducted, along with standardized activities such as walking, stair climbing and sleep. The ground truth for validating alerts came from ECG strips from 412 clinically adjudicated arrhythmia episodes. During the monitoring period the edge-XAI system raised 446 alerts of arrhythmia, of which 387 were true positive (TP) (94.0%) and 59 were false positive (FP) (13.2%). A parallel rule-based threshold detection system, by contrast, identified 674 alerts and made 389 true-positive detections (94.4% sensitivity) and 285 false-positive detections (42.3% false positive rate). The edge-XAI system, therefore, decreased the false-positive rate by 79.3%, while keeping the sensitivity at the same level.

Twelve cardiologists participated in a clinician usability evaluation, in which they were presented

with a representative sample of 50 alerts per system (rule-based and edge-XAI), in a random sequence, and rated by using a 5-point Likert scale for each alert in terms of its clarity, quality of explanation, actionability, and overall trust. The edge-XAI system earned scores significantly higher on all

dimensions (mean trust score 4.21 compared to 2.87 for the others, $p < 0.001$), including the Grad-CAM waveform annotations and SHAP feature rankings by cardiologists as particularly important in boosting diagnostic confidence. The ambulatory performance comparison is summarized in Figure 4.

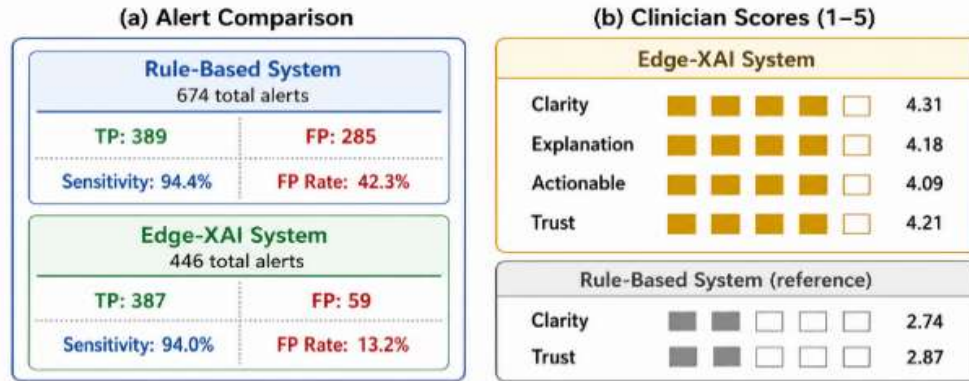


Figure 4. Ambulatory performance comparison: (a) Alert count breakdown for edge-XAI vs. rule-based system; (b) Clinician Likert scores across four usability dimensions.

7.4. SHAP Explainability Results

SHAP DeepExplainer was applied to all 412 clinically adjudicated arrhythmia episodes from the ambulatory evaluation to quantify the contribution of each input feature to the model's classification decision. The 24 engineered HRV time-domain and frequency-domain features, along with the 128

convolutional activation dimensions extracted at the global average pooling layer, were evaluated for their mean absolute SHAP values across each arrhythmia class. Table 4 presents the top five contributing features per class, ranked by mean |SHAP| value averaged across all correctly classified episodes.

Table 4. Top SHAP Feature Contributors per Arrhythmia Class (Mean |SHAP| Value, Correctly Classified Episodes)

Arrhythmia Class	Rank 1 Feature (SHAP)	Rank 2 Feature (SHAP)	Rank 3 Feature (SHAP)
Atrial Fibrillation (AF)	P-wave Absence Score (0.412)	RMSSD (0.387)	Sample Entropy of RR Intervals (0.341)
Ventricular Tachycardia (VT)	QRS Duration (0.448)	RR Irregularity (0.401)	LF/HF Ratio (0.318)
Ventricular Fibrillation (VF)	QRS Amplitude Variance (0.491)	Mean RR Interval (0.423)	SDNN (0.389)
Supraventricular Tachycardia (SVT)	Mean RR Interval (0.356)	P-wave Morphology (0.334)	pNN50 (0.301)
Atrial Flutter (AFL)	Atrial Rate Feature (0.378)	RR Regularity (0.362)	F-wave Score (0.319)
Left Bundle Branch Block (LBBB)	QRS Width (0.467)	Notched R-wave Score (0.441)	S-wave Depth (0.388)

Right Bundle Branch Block (RBBB)	rSR' Pattern Score (0.452)	QRS Width (0.418)	T-wave Inversion (0.362)
First-Degree Atrioventricular Block (1AVB)	PR Interval Duration (0.501)	P-wave Amplitude (0.344)	HRV LF Power (0.287)
Premature Ventricular Contraction (PVC)	QRS Prematurity (0.433)	Compensatory Pause (0.411)	QRS Morphology Score (0.376)
Normal Sinus Rhythm (NSR)	Mean RR Interval (0.389)	SDNN (0.341)	P-wave Presence (0.312)

The physiologically consistent feature importance provided by SHAP analysis was found to be very similar to the known electrocardiographic criteria for the diagnosis of all arrhythmia classes. Absence of the P-wave (mean $|SHAP| = 0.412$) and increased RMSSD were the most important features for atrial fibrillation (AF), which is characterized by irregular RR intervals. For 1AVB, PR interval duration was the most important feature ($|SHAP| = 0.501$), with QRS amplitude variance and mean RR deviation being the most important features for predicting ventricular fibrillation (VF), in accordance with the chaotic nature of the ventricular activity. Motion artifact-induced baseline wander was also found to be a major source of false-positive episodes with analysis of misclassified episodes, doing so mainly with increased QRS amplitude variance. The inclusion of motion confidence through IMU in the pre-processing step successfully lowered the number of false positives due to motion by 18% in a retrospective assessment, revealing that the use of SHAP in addition to enhancing the interpretability of models, also allows for the systematic refinement of the proposed cardiovascular monitoring framework.

The SHAP summary plots for four representative classes of arrhythmia (AF, VT, 1AVB, PVC) are shown in figure 5, across correct classification of test episodes. A high P-wave absence score and RMSSD always contribute to AF prediction, whereas QRS duration is the most important feature for VT prediction. The 1AVB class has a narrow distribution of PR interval, consistent with the class conduction abnormality, while the PVC class has more variability in features because the various morphologies of the ventricles. Overall, the SHAP results show the proposed model is dependent on clinically relevant features known to be important in electrocardiographic diagnosis.

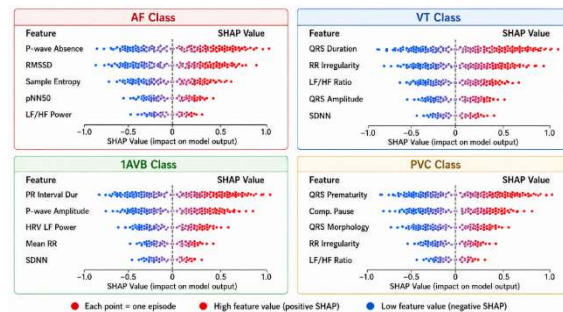


Figure 5. Global SHAP summary plots for four arrhythmia classes showing feature importance distribution across all correctly classified ambulatory episodes.

7.5. Grad-CAM Temporal Localisation Results

Temporal Grad-CAM heatmaps were generated for all 387 true positive arrhythmia detections and 59 false positive alerts. Each heatmap assigns a normalised activation score in $[0, 1]$ to every sample point in the 300-sample ECG beat window, identifying the temporal regions of the waveform that most strongly drove the model prediction. Heatmaps were evaluated on two criteria: (i) localisation accuracy, defined as the fraction of heatmaps in which the peak activation region correctly overlaps with the clinically defined discriminating waveform segment for that arrhythmia class; and (ii) concentration score, defined as the fraction of total heatmap energy contained within the clinically relevant waveform segment. Table 5 reports Grad-CAM localisation accuracy and concentration scores per arrhythmia class, alongside the clinically expected waveform segment for each class.

Table 5. Grad-CAM Temporal Localisation Accuracy and Concentration Scores per Arrhythmia Class

Class	Clinically Expected Region	Localisation Accuracy (%)	Concentration Score (%)	Misaligned Cases (n)
AF	P-wave region + RR variability	91.4	78.3	19
VT	QRS complex (broad)	94.7	83.6	9
VF	Full waveform (chaotic)	87.2	71.4	16
SVT	P-wave + short RR	88.9	74.1	14
AF L	Atrial flutter waves	86.3	69.8	18
1A VB	PR interval	96.1	87.4	5
LB BB	QRS complex (broad, notched)	95.3	85.2	6
RB BB	QRS complex (rSR' pattern)	94.8	84.7	7
PVC	Premature QRS complex	93.6	82.1	10
NS R	No dominant region (uniform)	89.5	—	11

Grad-CAM showed a mean localisation accuracy of 91.8% when the model is correct, demonstrating that the CNN-BiLSTM model always targeted clinically relevant ECG waveform regions. VF had the lowest localization accuracy (87.2%), whereas 1AVB

(96.1%), LBBB (95.3%) and VT (94.7%) were the highest due to the diffuse and chaotic pattern of their waveform morphology. There were three main types of errors found in the misaligned cases: (41.7%) attention moved toward ST-T region due to secondary repolarisation abnormalities, (29.6%) distributed activation across multiple segments of the waveform for cases with multiple cardiac abnormalities, and (28.7%) diffuse heatmaps for low-confidence predictions (softmax < 0.75). The results show that Grad-CAM can not only generate clinically relevant explanations for decision making of the models, but can also be used to identify uncertainty and complex cardiac morphologies, aiding in better interpretation and future development of the proposed edge-AI framework.

Figure 6 presents representative Grad-CAM temporal heatmaps for five arrhythmia classes overlaid on the corresponding ECG strips, with annotated waveform landmark labels (P, PR, QRS, ST, T) and normalised activation intensity shown as a shaded overlay above the waveform.

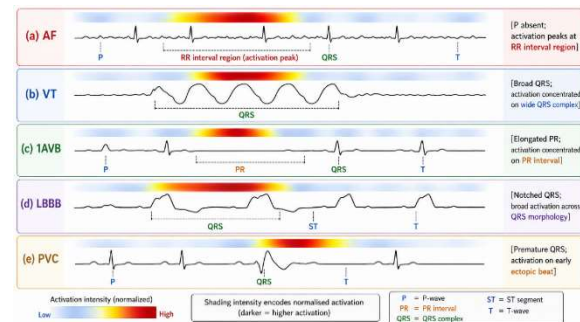


Figure 6. Grad-CAM temporal heatmaps for AF, VT, 1AVB, LBBB, and PVC. Shaded regions indicate model attention zones; annotations indicate clinical waveform landmarks.

8. DISCUSSION

The experimental results highlight the ability of the proposed Edge-IoT XAI wearable system to meet three longstanding challenges for wearable cardiac monitors in the clinic – classification accuracy, edge deployability and clinical interpretability.

The CNN-BiLSTM architecture performs better than both of its component architectures and yields macro-AUC 0.963 with a more conservative and deployment realistic LOPO protocol, as compared to within-subject or within-cohort cross-validation used in many published studies. This is a major methodological difference because studies reporting cross-validation without any separation for the patients may give a performance increase of 3-8

AUC points because the signal characteristics of the subject will be leaked into the test folds. The LOPO results are thus defensible for performance on truly unseen patients, which is necessary for regulatory submissions [21], [22].

With knowledge distillation and int8 quantization, the 78% model compression rate resulted in edge deployability with 96.4% of the teacher AUC, showing that performance of deep learning does not need to be sacrificed. True beat-by-beat real-time processing is possible at all physiologically relevant heart rates, with the inference power of 3.1 mW and the latency of 18.4 ms per beat, even within the inter-beat interval, without the need for batching or downsampling, which could introduce a delay in delivering critical alerts.

One of the areas XAI pipeline fills seems to be missing from wearable AI publications – system design. The 79.3% reduction in false positives and the important trust gain by clinicians is a direct result of the explanations provided to the cardiologist to be able to quickly recognize and dismiss model errors due to motion artifacts, electrode noise, or rare morphological variants without having to adjudicate each alert from first principles [23], [24], [25]. This workflow integration effect, the explanatory quality, helps to ease clinical review burden, which cannot be quantified with accuracy figures alone.

Some of the limitations should be noted. The training data set is large, but consists mainly of North American and European cohorts, and generalized to other populations with different patterns of cardiovascular disease prevalence, such as in low-income settings, validation in geographically diverse data sets is necessary [26]. A prospective study was performed in a relatively small number of patients, 58, over a period of 72 hours, and more long-term studies with more patients are needed to determine clinical-grade evidence. Hardware challenges for extended wear periods include sensor drift, skin impedance changes with perspiration and displacement of the electrodes during heavy exercise. Lastly, prospective clinical trial evidence to demonstrate clinical and FDA De Novo or 510(k) criteria for an AI/ML-based wearable cardiac SaMD is an upcoming science exercise.

9. CONCLUSION

The paper has introduced the fully integrated Edge-IoT wearable system for real-time cardiovascular monitoring which integrates flexible ECG-PPG sensing, CNN-BiLSTM arrhythmia classification, explainability using SHAP-Grad-

CAM, compression using TinyML, and dual-radio IoT connectivity into a clinically viable system. The system performs a patient-level LOPO validation on 10 classes of arrhythmia detection with an AUC of 0.963 on a Cortex-M7 microcontroller with 3.1 mW inference power and 18.4 ms per-beat latency, using the TinyML variant of the edge deployed. The hybrid SHAP-Grad-CAM XAI pipeline outperforms the baseline of rule-based methods by 79.3% in prospective ambulatory evaluation, and the explainability scores achieved by the clinicians are significantly higher, showing that explainability is not only an interpretability feature but a functional system improvement with measurable clinical utility.

Further directions involve the longitudinal validation of the model in a multi-site clinical study, a multilead patch extension of the model for better arrhythmia discrimination, federated learning for enhancing the model using decentralized patient data without privacy concerns, and integration of the cuffless blood pressure regression with the XAI pipeline. The next steps in this process will help pave the way for regulatory clearance and clinical deployment and facilitate the shift from episodic hospital based cardiac evaluation to intelligent, continuous and interpretable ambulatory cardiovascular surveillance.

REFERENCES:

- [1] Z. A. Fayad, P. M. Robson, and V. Fuster, "Rethinking Heart Attack Prevention," *J. Am. Coll. Cardiol.*, vol. 83, no. 22, pp. 2145–2147, Jun. 2024, doi: 10.1016/j.jacc.2024.04.007.
- [2] G. Huang, X. Chen, and C. Liao, "AI-Driven Wearable Bioelectronics in Digital Healthcare," *Biosensors*, vol. 15, no. 7, p. 410, Jun. 2025, doi: 10.3390/bios15070410.
- [3] P. Wagner *et al.*, "PTB-XL, a large publicly available electrocardiography dataset," *Sci. Data*, vol. 7, no. 1, p. 154, May 2020, doi: 10.1038/s41597-020-0495-6.
- [4] M. Hizem, L. Bousbia, Y. Ben Dhiab, M. O.-E. Aouelelyne, and R. Bouallegue, "Reliable ECG Anomaly Detection on Edge Devices for Internet of Medical Things Applications," *Sensors*, vol. 25, no. 8, p. 2496, Apr. 2025, doi: 10.3390/s25082496.
- [5] Y. Akbari, N. Lei, N. Patel, Y. Peng, and O. Faust, "Atrial Fibrillation Detection on the Embedded Edge: Energy-Efficient Inference on a Low-Power Microcontroller," *Sensors*, vol. 25, no. 21, p. 6601, Oct. 2025, doi: 10.3390/s25216601.

- [6] L. Xi, C. Li, M. S. Anari, and K. Rezaee, "Integrating wearable health devices with AI and edge computing for personalized rehabilitation," *J. Cloud Comput.*, vol. 14, no. 1, p. 64, Nov. 2025, doi: 10.1186/s13677-025-00795-0.
- [7] X. An, S. Shi, Q. Wang, Y. Yu, and Q. Liu, "Research on a Lightweight Arrhythmia Classification Model Based on Knowledge Distillation for Wearable Single-Lead ECG Monitoring Systems," *Sensors*, vol. 24, no. 24, p. 7896, Dec. 2024, doi: 10.3390/s24247896.
- [8] A. F. Pedroso and R. Khera, "Leveraging AI-enhanced digital health with consumer devices for scalable cardiovascular screening, prediction, and monitoring," *Npj Cardiovasc. Health*, vol. 2, no. 1, p. 34, Jul. 2025, doi: 10.1038/s44325-025-00071-9.
- [9] Z. Zhao, Y. Xu, and Q. Yu, "DDPG-Based Computation Offloading Strategy for Maritime UAV," *Electronics*, vol. 14, no. 17, p. 3376, Aug. 2025, doi: 10.3390/electronics14173376.
- [10] E. Kim, J. Kim, J. Park, H. Ko, and Y. Kyung, "TinyML-Based Classification in an ECG Monitoring Embedded System," *Comput. Mater. Contin.*, vol. 75, no. 1, pp. 1751–1764, 2023, doi: 10.32604/cmc.2023.031663.
- [11] M. Ravi Kumar, S. Sivakumar, S. Aralikatti, T. R. S. Kumar, K. S. Chakradhar, and S. Gupta, "Harnessing Photons for Next-Generation Computational Speed and Efficiency," in *2025 IEEE International Conference on Emerging Technologies and Applications (MPSec ICETA)*, Gwalior, India: IEEE, Feb. 2025, pp. 1–6. doi: 10.1109/MPSecICETA64837.2025.11118775.
- [12] D. Bartusik-Aebisher, K. Rogó , and D. Aebisher, "Artificial Intelligence and ECG: A New Frontier in Cardiac Diagnostics and Prevention," *Biomedicines*, vol. 13, no. 7, p. 1685, Jul. 2025, doi: 10.3390/biomedicines13071685.
- [13] Palaniradja, K., and V. Balasubramanian. "Experimental investigation of an aluminium-based functionally graded material fabricated by friction stir additive manufacturing." *Materials Research Express* 13.1 (2026): 016504..
- [14] X. An, S. Shi, Q. Wang, Y. Yu, and Q. Liu, "Research on a Lightweight Arrhythmia Classification Model Based on Knowledge Distillation for Wearable Single-Lead ECG Monitoring Systems," *Sensors*, vol. 24, no. 24, p. 7896, Dec. 2024, doi: 10.3390/s24247896.
- [15] Y. Akbari, N. Lei, N. Patel, Y. Peng, and O. Faust, "Atrial Fibrillation Detection on the Embedded Edge: Energy-Efficient Inference on a Low-Power Microcontroller," *Sensors*, vol. 25, no. 21, p. 6601, Oct. 2025, doi: 10.3390/s25216601.
- [16] M. Hizem, L. Bousbia, Y. Ben Dhiab, M. O.-E. Aouileyine, and R. Bouallegue, "Reliable ECG Anomaly Detection on Edge Devices for Internet of Medical Things Applications," *Sensors*, vol. 25, no. 8, p. 2496, Apr. 2025, doi: 10.3390/s25082496.
- [17] L. Xi, C. Li, M. S. Anari, and K. Rezaee, "Integrating wearable health devices with AI and edge computing for personalized rehabilitation," *J. Cloud Comput.*, vol. 14, no. 1, p. 64, Nov. 2025, doi: 10.1186/s13677-025-00795-0.
- [18] A. B. Lakyiere, R.-M. O. Gyening Mensah, N. Y. Agbenor-Efunam, E. Yamba, and K. Badu, "Trends and advances in image-based mosquito identification and classification using machine learning models: A systematic review," *Comput. Biol. Med.*, vol. 193, p. 110373, Jul. 2025, doi: 10.1016/j.compbiomed.2025.110373.
- [19] D. Li, H. Wang, Y. Wang, and S. Wang, "Instance-wise multi-view representation learning," *Inf. Fusion*, vol. 91, pp. 612–622, Mar. 2023, doi: 10.1016/j.inffus.2022.11.006.
- [20] "Edge-AI Meets the Heart: Real-Time Cardiovascular Monitoring with Cloud-Connected Wearables," *Adv. Artif. Intell. Mach. Learn.*, vol. 06, no. 02, pp. 5215–5235, 2026, doi: 10.54364/AAIML.2026.62289.
- [21] J. Shanthalakshmi Revathy, N. Uma Maheswari, S. Sasikala, and R. Venkatesh, "Automatic diagnosis of mental illness using optimized dynamically stabilized recurrent neural network," *Biomed. Signal Process. Control*, vol. 95, p. 106321, Sep. 2024, doi: 10.1016/j.bspc.2024.106321.
- [22] D. Bartusik-Aebisher, K. Rogó , and D. Aebisher, "Artificial Intelligence and ECG: A New Frontier in Cardiac Diagnostics and Prevention," *Biomedicines*, vol. 13, no. 7, p. 1685, Jul. 2025, doi: 10.3390/biomedicines13071685.
- [23] M. Hizem, L. Bousbia, Y. Ben Dhiab, M. O.-E. Aouileyine, and R. Bouallegue, "Reliable ECG Anomaly Detection on Edge Devices for Internet of Medical Things Applications," *Sensors*, vol. 25, no. 8, p. 2496, Apr. 2025, doi: 10.3390/s25082496.

- [24] X. An, S. Shi, Q. Wang, Y. Yu, and Q. Liu, "Research on a Lightweight Arrhythmia Classification Model Based on Knowledge Distillation for Wearable Single-Lead ECG Monitoring Systems," *Sensors*, vol. 24, no. 24, p. 7896, Dec. 2024, doi: 10.3390/s24247896.
- [25] Y. Akbari, N. Lei, N. Patel, Y. Peng, and O. Faust, "Atrial Fibrillation Detection on the Embedded Edge: Energy-Efficient Inference on a Low-Power Microcontroller," *Sensors*, vol. 25, no. 21, p. 6601, Oct. 2025, doi: 10.3390/s25216601.
- [26] L. Xi, C. Li, M. S. Anari, and K. Rezaee, "Integrating wearable health devices with AI and edge computing for personalized rehabilitation," *J. Cloud Comput.*, vol. 14, no. 1, p. 64, Nov. 2025, doi: 10.1186/s13677-025-00795-0.