

Intelligent Edge Device for Real Time Cardiac Arrhythmia Detection

Kiran Bobby¹, Linu Jose², Ridhin Thomas Alex³, Isra Emmanuel⁴, Rameesa Parveen⁵, Joel Vinod⁶

Assistant Professor Department of EEE , Mar Athanasius College of Engineering, Kothamangalam, Kerala

Assistant Professor Department of EEE , Mar Athanasius College of Engineering, Kothamangalam, Kerala

UG Student Department of EEE , Mar Athanasius College of Engineering, Kothamangalam, Kerala

UG Student Department of EEE , Mar Athanasius College of Engineering, Kothamangalam, Kerala

UG Student Department of EEE , Mar Athanasius College of Engineering, Kothamangalam, Kerala

UG Student Department of EEE , Mar Athanasius College of Engineering, Kothamangalam, Kerala

Abstract - Cardiac arrhythmias are irregular heart rhythms that may appear suddenly and remain undetected without continuous monitoring. Conventional monitoring approaches often depend on cloud-based analysis, leading to increased latency, higher power consumption, and concerns regarding data privacy. This work presents a compact intelligent edge device capable of performing real-time arrhythmia detection directly on embedded hardware. The proposed system combines a MAX30003 ECG analog front-end for accurate biopotential acquisition with an ESP32-S3 microcontroller for on device signal processing and classification. Acquired ECG signals undergo filtering, segmentation, and normalization before being analyzed using a lightweight convolutional neural network optimized for resource-constrained environments. The model identifies five classes of arrhythmia defined by AAMI standards (N, S, V, F, Q) without requiring continuous internet connectivity. A local OLED interface provides instant visual alerts, while optional wireless communication enables data transfer when needed. The implementation demonstrates that reliable cardiac monitoring and classification can be achieved using a portable, low-power architecture, making it suitable for wearable and remote healthcare applications where real-time response, energy efficiency, and data security are critical.

Key Words: ECG Monitoring, Arrhythmia Classification, Edge Computing, ESP32-S3, MAX30003, Embedded AI.

I. INTRODUCTION

Cardiovascular diseases continue to be one of the leading causes of death worldwide, and cardiac arrhythmias form a major category of these disorders. Arrhythmias are irregular heart rhythms that may occur unpredictably and therefore require continuous electrocardiogram (ECG)

monitoring for reliable detection and early medical intervention. Because these abnormalities are often intermittent, short clinical examinations may fail to capture them, making long-term monitoring essential for accurate diagnosis.

Traditional ECG analysis is typically performed manually by clinicians using hospital-based equipment. While clinically effective, this process is time-consuming, resource-intensive, and not suitable for continuous personal monitoring. The increasing volume of physiological data generated during long-duration ECG recording also makes manual interpretation challenging and prone to human error.

Recent progress in artificial intelligence (AI) has enabled automated arrhythmia detection, offering improved accuracy and faster analysis. However, many AI-based solutions rely heavily on cloud computing, where ECG data must be transmitted to remote servers for processing. This dependency introduces several limitations, including communication latency, continuous internet requirements, higher energy consumption, and concerns related to the privacy and security of sensitive medical data. These challenges reduce the practicality of such systems for real-time and wearable healthcare applications.

Consumer wearable devices have improved accessibility to heart-rate monitoring, yet most are limited to detecting only a narrow range of abnormalities [7], such as atrial fibrillation, and lack the capability for comprehensive multi-class arrhythmia classification. As a result, a clear gap exists between medical-grade monitoring systems, which provide accurate analysis but are expensive and non-portable, and consumer devices, which are convenient but diagnostically limited.

To address these challenges, there is a growing need for compact, low-power systems capable of performing cardiac analysis directly on embedded hardware without reliance on cloud infrastructure. Edge computing enables local data processing at the device level, significantly reducing latency while enhancing data privacy and allowing continuous operation.

In this context, the present work proposes an intelligent edge device for real-time cardiac arrhythmia detection that integrates ECG signal acquisition, preprocessing, and lightweight deep learning inference into a single portable platform. The system is designed to perform multi-class arrhythmia classification locally while maintaining low power consumption and cost, making it suitable for wearable and remote healthcare monitoring environments.

The remainder of this paper is organized as follows. Section II reviews related work on ECG based arrhythmia detection and embedded edge-AI healthcare systems. Section III presents the proposed signal acquisition and preprocessing methodology. Section IV describes the machine learning model design and on-device inference framework. Section V details the hardware architecture and implementation of the proposed device, Section VI evaluates the experimental results and system performance. Section VII discusses future improvements and potential extensions. Finally, Section VIII concludes the paper. evaluates the experimental results and system performance.

II. RELATED WORK

Automated detection of cardiac arrhythmias using ECG signals has been an active research area due to the increasing demand for continuous and accurate cardiac monitoring [1],[4]. Traditional signal-processing-based methods relied on handcrafted feature extraction techniques, which required expert knowledge and often struggled to generalize across diverse patient datasets. With the advancement of machine learning and deep learning, researchers have shifted toward data-driven approaches capable of learning complex temporal and morphological patterns directly from ECG signals [3].

Recent studies have demonstrated that convolutional neural networks (CNNs) and recurrent architectures can achieve high classification accuracy for arrhythmia detection by automatically extracting relevant features from raw ECG data [3]. These approaches significantly

reduce the need for manual feature engineering and improve diagnostic consistency. However, many high-performing models are computationally intensive and designed primarily for cloud or workstation environments, limiting their applicability in portable healthcare systems.

To address this limitation, edge-based healthcare solutions have been explored, where ECG processing and classification are performed locally on embedded platforms [5],[6]. Such systems aim to reduce latency, minimize bandwidth usage, and enhance data privacy by avoiding continuous transmission of sensitive physiological signals to remote servers. Research has shown that integrating optimized deep learning models with energy-efficient hardware can enable real-time inference while maintaining acceptable diagnostic performance [5].

In addition, specialized biomedical front-end integrated circuits have been developed to improve ECG acquisition quality in wearable applications [6]. These devices provide precise signal amplification, filtering, and analog-to-digital conversion, which are essential for obtaining clinically meaningful ECG data in portable environments. Combining such acquisition modules with lightweight embedded intelligence allows the development of compact systems capable of continuous monitoring outside traditional clinical settings.

Despite these advancements, achieving a balance between model accuracy, computational efficiency, and hardware constraints remains a key challenge. Many existing implementations either focus on algorithmic performance without considering deployment feasibility or emphasize hardware efficiency while supporting only limited classification capability. Therefore, there is a need for a unified approach that integrates reliable ECG acquisition with optimized on-device intelligence to enable practical real-time arrhythmia detection.

III. PROPOSED SYSTEM ARCHITECTURE

The proposed system is designed as a compact and intelligent edge-based platform capable of acquiring electrocardiogram (ECG) signals, processing them locally, and performing real-time cardiac arrhythmia classification without reliance on cloud infrastructure. The architecture integrates hardware components for signal acquisition and embedded processing with lightweight machine learning for on-device inference.

3.1 Overall System Overview

The system consists of four primary modules:

1. ECG Signal Acquisition Module
2. Analog Front-End (AFE) Processing
3. Edge Processing Unit (Microcontroller)
4. Display and Alert Interface

These modules work together to ensure accurate signal capture, preprocessing, classification, and real-time feedback.

3.2 ECG Signal Acquisition

Disposable ECG electrodes are placed on the human body to capture the electrical activity generated by cardiac muscle contractions. The detected signals are extremely low in amplitude and highly susceptible to noise from muscle movement, power-line interference, and environmental disturbances. Therefore, proper amplification and filtering are necessary before digital processing.

3.3 Analog Front-End Module

To obtain high-quality ECG signals, an analog front-end (AFE) module is used. In the final hardware implementation, the MAX30003-based ECG front-end is utilized to provide:

- Low-noise signal amplification
- Built-in filtering
- High-resolution analog-to-digital conversion
- Motion artifact reduction

The AFE ensures that the ECG waveform is clean, stable, and suitable for accurate classification. This improves diagnostic reliability and reduces the probability of false detections.

3.4 Edge Processing Unit

The core processing element of the system is a **microcontroller**, which performs all computation locally. The microcontroller is responsible for:

- Receiving digitized ECG samples from the analog front-end
- Performing signal preprocessing (noise removal and normalization)
- Segmenting ECG beats for analysis

- Executing a lightweight deep learning model for multi-class arrhythmia classification

By performing inference directly on the device, the system eliminates the need for cloud communication, thereby reducing latency and preserving patient data privacy. The

3.5 Classification Module

A lightweight convolutional neural network (CNN) model is deployed on the embedded platform to classify ECG signals into five categories based on standard arrhythmia classification guidelines:

- Normal (N)
- Supraventricular ectopic beat (S)
- Ventricular ectopic beat (V)
- Fusion beat (F)
- Unknown/Other class (Q)

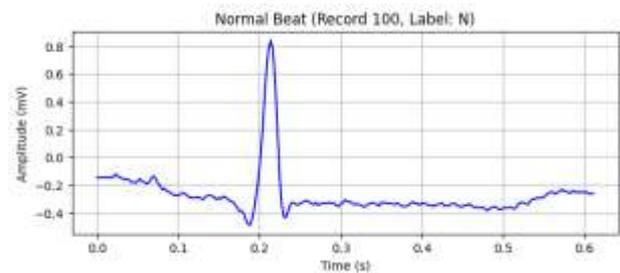


Figure 1: Waveform of normal heartbeat (N)

Fig 1 represents a **normal heartbeat (N)**. In this condition, the ECG signal shows a regular P-QRS-T sequence with consistent amplitude and periodic intervals between successive beats. The R-R interval remains relatively constant, indicating stable heart rhythm.

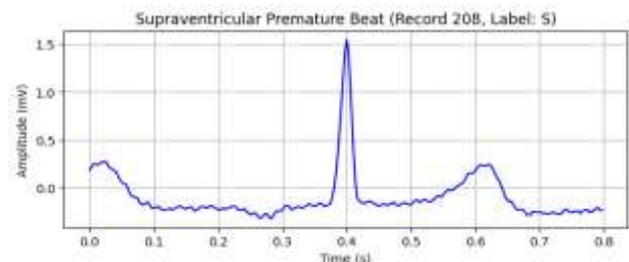


Figure 2: Waveform of supraventricular ectopic beat (S)

Fig 2 corresponds to a **supraventricular ectopic beat (S)**. In this case, the ECG signal shows irregular timing between beats due to premature electrical activity originating above the ventricles. The waveform pattern differs slightly from the normal rhythm, often showing shortened R-R intervals or altered P-wave characteristics.

When this waveform is analyzed by the proposed model, the classification module detects the abnormal timing behavior and categorizes the signal as a supraventricular ectopic beat. Supraventricular ectopic activity has also been associated with an increased risk of atrial fibrillation and stroke [2].

Fig 3 represents a **ventricular ectopic beat (V)**, which is characterized by a widened and distorted QRS complex and the absence of a preceding P-wave in many cases. Ventricular ectopic beats originate from abnormal electrical activity within the ventricles and typically appear as premature and irregular spikes in the ECG waveform

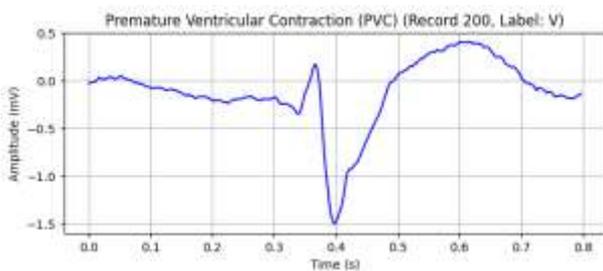


Figure 3 : waveform of a ventricular ectopic beat (V)

Ventricular ectopic beats originate from abnormal electrical activity within the ventricles and typically appear as premature and irregular spikes in the ECG waveform. During simulation, the system successfully detected these abnormal waveform patterns and classified them as ventricular ectopic beats. .

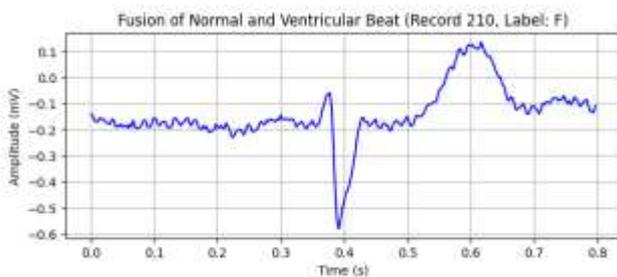


Figure 4 :Fusion beats (F)

Fig 4 occur when electrical impulses from the normal conduction pathway and an ectopic ventricular source overlap, producing a waveform that combines characteristics of both normal and abnormal beats. The resulting ECG waveform appears as a hybrid signal with intermediate morphology between normal and ventricular beats.

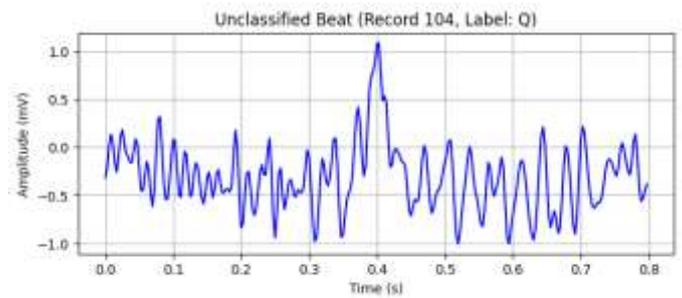


Figure 5:Unclassified beat

This class includes ECG signals that do not clearly match the patterns of the other defined categories or contain irregular features that fall outside the trained dataset. In such cases, the system labels the waveform as an unknown class to avoid incorrect classification. This approach improves the reliability of the system by preventing false predictions when unfamiliar signal patterns are encountered.

IV. MACHINE LEARNING MODEL DESIGN

4.1 CNN-Based Architecture Overview

The model development in this work employs a teacher–student knowledge distillation framework, inspired by the methodology of Wong et al. [1]. A large, high capacity ResNet-based teacher model is first trained to achieve maximum classification accuracy on the MIT-BIH Arrhythmia Database. Subsequently, a compact CNN based student model learns to replicate the teacher's behavior through knowledge distillation, while using a fraction of the computational resources. This enables deployment on resource-constrained microcontrollers without cloud connectivity. Two post-training quantization strategies dynamic range quantization and student model to produce deployment-ready compressed variants for edge inference.

4.2 Teacher Model Architecture

The teacher model Fig 6 serves as the knowledge source in the distillation framework, prioritizing accuracy over efficiency to capture complex morphological patterns across all five AAMI arrhythmia classes. The input layer receives ECG segments of 321 samples (~0.9 s at 360 Hz), centered on the R-peak with 160 samples before and after, ensuring complete P-wave, QRS complex, and T-wave coverage.

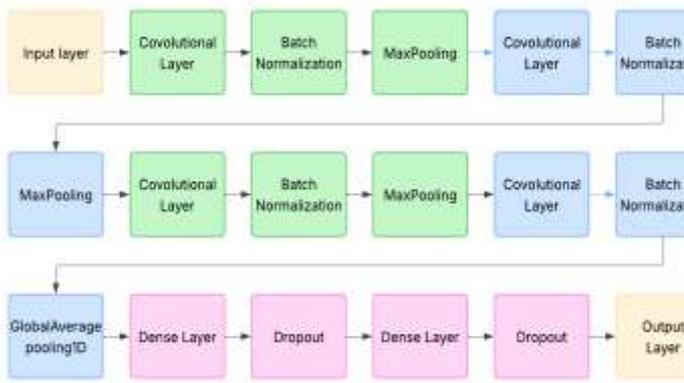


Figure 6: Teacher model Architecture

The teacher model contains **847,552 trainable parameters** with a total size of **3009.1 KB**, achieving a test accuracy of **89.43%**. Training converged at epoch 42 with minimal overfitting.

4.3 Student Model Architecture

The student model Fig 7 is specifically designed for deployment on microcontrollers with severe memory and computational constraints, achieved through a dramatically simplified architecture that retains clinically relevant accuracy through knowledge distillation with only 25,728 trainable parameters and a Float32 size of 90.8 KB, the student model is 33.1× smaller than the teacher. Despite this compression, the student

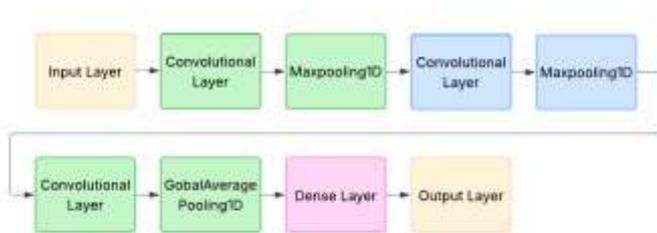


Figure 7: Student Model Architecture

achieves 85.23% accuracy retaining 95.3% of the teacher model's performance with only a 4.20% accuracy drop

4.4 Knowledge Distillation Framework

Knowledge distillation is a model compression technique where a smaller student model learns from a larger and more accurate teacher model. Instead of training only with hard labels, the student learns from the teacher's soft probability outputs, which capture relationships between different classes. These soft targets provide additional information about class similarities and uncertainties. As a result, the student model can achieve improved generalization while maintaining a significantly smaller model size suitable for edge deployment.

4.5 Quantization Results and Analysis

Two post-training quantization approaches were evaluated: dynamic range (INT8 weights) and Float16 (F16). Dynamic range quantization reduced the model size from 90.8 KB (Float32 student) to 30.0 KB, achieving approximately 100× compression relative to the 3.0 MB teacher model, with overall accuracy improving to 91.28%. Float16 quantization produced a 42.2 KB model (~71× compression) with 91.45% overall accuracy.

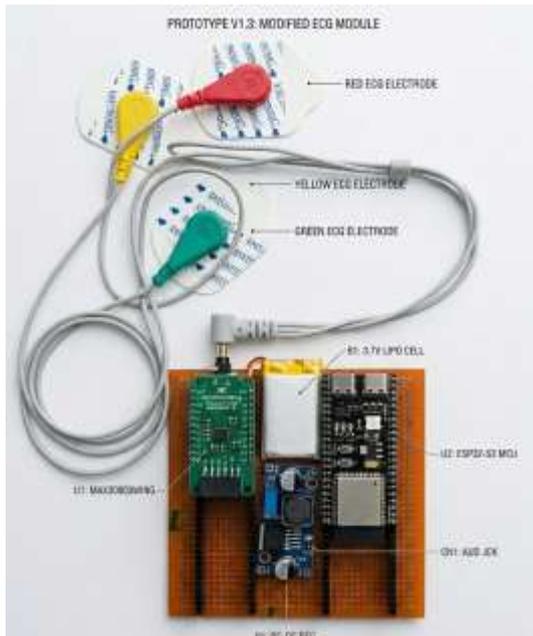
Despite strong overall accuracy, per class analysis revealed imbalance effects particularly on minority classes, with a Macro F1 of 51.91% reflecting the dominance of the Normal class in the test set. Dynamic range quantization was selected for final deployment given its superior compression at 30.0 KB.

4.6 On-Device Inference on ESP32-S3

The quantized TensorFlow Lite model is deployed on the ESP32-S3 using TensorFlow Lite for Microcontrollers (TFLM). The trained .tflite model is converted into a C header file and embedded in flash memory, while a tensor arena is allocated in SRAM to store intermediate activations during inference. The dual core architecture of the ESP32-S3 enables separation of ECG signal acquisition and inference tasks, ensuring continuous data sampling and real time arrhythmia classification.

V. HARDWARE ARCHITECTURE AND IMPLEMENTATION

The proposed intelligent edge device integrates ECG acquisition hardware with an embedded inference platform based on the ESP32-S3 microcontroller. The overall system architecture is shown in Fig 8.

**Figure 8: Hardware Implementation**

5.1 ECG Acquisition Module

ECG signals are acquired using the MAX30003WING analog front-end. The MAX30003 provides high-resolution biopotential measurement with integrated R-peak detection and a digital SPI interface, making it suitable for embedded cardiac monitoring applications. The system operates at a sampling frequency of 256 Hz, aligned with the embedded deployment configuration. ECG samples are transmitted to the ESP32-S3 via SPI and buffered for segmentation and preprocessing prior to neural network inference.

5.2 Embedded Processing Unit

The ESP32-S3 serves as the primary processing unit for signal handling and model inference. The microcontroller was selected due to its dual core architecture, sufficient SRAM for TensorFlow Lite Micro execution, integrated wireless connectivity, and suitability for low-power edge applications. The firmware is developed using the ESP-IDF framework, integrating TensorFlow Lite for Microcontrollers (TFLM) for embedded model execution.

5.3 Power Architecture

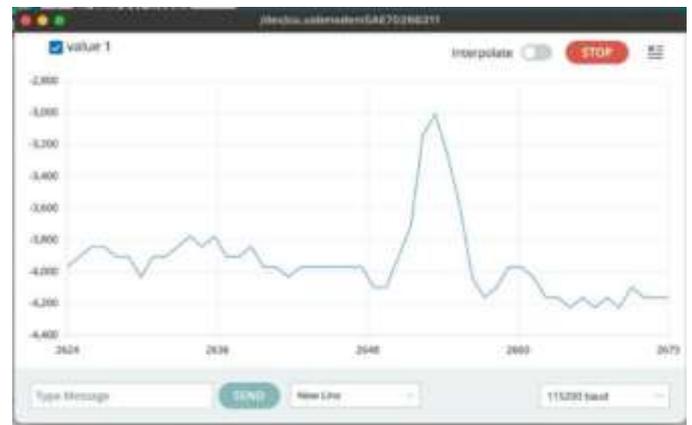
A 3.7 V LiPo battery with a suitable buck boost converter is used to power esp32s3 with a voltage of 3.3v. However, due to voltage conversion constraints and stability considerations associated with boost and buck conversion stages, regulated external power was used during system

validation. Portable power optimization is planned as a future enhancement.

VI. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

6.1 Real-Time ECG Signal Acquisition

The proposed hardware prototype was used to acquire ECG signals from human subjects using disposable electrodes. The electrodes were placed in a standard three-lead configuration to capture the electrical activity of the heart. The MAX30003 analog front-end successfully amplified and digitized the ECG signals, which were then transmitted to the ESP32-S3 for processing.

**Figure 9: Acquired ECG waveform**

The acquired ECG waveform was plotted in real time. R-R intervals were detected, and heartbeat was calculated, demonstrating the system's ability to perform continuous signal acquisition and monitoring. Distinct ECG features such as the P-wave, QRS complex, and T-wave were observed. Fig 9 shows the ECG waveform obtained from the system.

During the current stage of experimentation, the system was validated for real-time ECG signal acquisition and waveform visualization using normal physiological signals obtained from human subjects. Due to ethical and practical constraints, it is difficult to intentionally reproduce abnormal cardiac conditions in healthy individuals. Therefore, real-time testing of abnormal arrhythmia cases is currently limited. Validation of arrhythmia detection using recorded ECG datasets and controlled signal inputs is ongoing and will be included in future work.

VII. CONCLUSIONS

This work presented the design and implementation of an intelligent edge device for real-time cardiac arrhythmia detection. The proposed system integrates ECG signal acquisition, preprocessing, and lightweight deep learning inference within a compact embedded platform powered by the ESP32-S3 microcontroller.

Unlike conventional cloud-based diagnostic systems, the proposed architecture performs complete signal analysis directly on the device. This eliminates latency associated with remote processing, enhances patient data privacy, and reduces dependence on continuous internet connectivity. Such characteristics make the system particularly suitable for wearable and remote healthcare applications.

The integration of efficient signal conditioning, beat segmentation, normalization, and optimized convolutional neural network classification demonstrates that reliable multi-class arrhythmia detection can be achieved within strict memory and power constraints. The results confirm that embedded edge-AI solutions can bridge the gap between medical grade monitoring systems and consumer wearable devices.

Furthermore, the modular design of the proposed system allows future scalability, including integration with advanced noise filtering techniques, improved model compression strategies, and extended wireless connectivity options. With further validation and optimization, the system has strong potential for deployment in portable cardiac healthcare devices aimed at continuous and preventive monitoring.

In summary, this research establishes a practical framework for deploying intelligent cardiac monitoring systems on resource constrained embedded hardware, contributing to the advancement of real-time, privacy preserving, and accessible healthcare technology.

REFERENCES

1. S. Dash, K. H. Chon, S. Lu, and E. A. Raeder, "Automatic real time detection of atrial fibrillation," *Ann.Biomed. Eng.*, vol. 37, no. 9, pp. 1701–1709, Sep. 2009.
2. Z. Binici, T. Intzilakis, O. W. Nielsen, L. Køber, and A. Sajadieh, Excessive supraventricular ectopic activity and increased risk of atrial brillation and stroke, *Circulation*, vol. 121, no. 17, pp. 19041911, May 2010.
3. Kovalchuk, O.; Barmak, O.; Radiuk, P.; Klymenko, L.; Krak, I. Towards Transparent AI in Medicine: ECG-Based Arrhythmia Detection with Explainable Deep Learning. *Technologies* 2025, 13, 34.
4. H. Wellens, Cardiac arrhythmias: The quest for a cureA historical perspective, *J. Amer. College Cardiol.*, vol. 44, no. 6, pp. 11551163, Sep. 2004.
5. Wong, J., Nerbonne, J., & Zhang, Q. (2024). Ultra-Efficient Edge Cardiac Disease Detection Towards real-Time Precision Health. *IEEE Access*, 12, 9940-9951
6. Dell'Aquila, C. R., Cañadas, G. E., & Laciari, E. (2023). Embedded System for the Simultaneous Study of SAHS and Cardiac Arrhythmia. *IEEE Embedded Systems Letters*, 15(2), 53-56.
7. Apple Inc. (n.d.). *Take an ECG with the ECG app on Apple Watch*. Apple Support. Retrieved August 4, 2025.
8. Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3), 45-50.
9. *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, ANSI/AAMI Standard EC 38, American National Standards Institute, Washington, DC, USA, 1998.