***UDC: 004***
***DOI: https://doi.org/10.30546/09085.2025.02.317***

# MOBILE ARRHYTHMIA MONITORING SYSTEM BASED ON MULTIMODAL BIOSIGNAL ANALYSIS: SYNCHRONIZATION OF ECG, PCG, AND PPG

**Sevinj ALIYEVA**[1*]

[1] *Azerbaijan Technical University, Baku, Azerbaijan*

**ARTICLE INFO**

*Article history*
Received:2025-10-03
Received in revised form:2025-10-05
Accepted:2025-10-22
Available online

*Keywords:*
*arrhythmia; electrocardiography; phonocardiography; photoplethysmography; multimodal analysis.*
*JEL classification: I12,C61,O33*

**ABSTRACT**

*Cardiovascular diseases and, specifically, arrhythmias account for one of the top priority problems of global healthcare systems. Although routine diagnostic procedures, including electrocardiography (ECG) and phonocardiography (PCG), remain cardinal during cardiac activity assessment, the single-modality application restricts diagnostic effectiveness. A mobile system of monitoring arrhythmias by synchronized ECG, PCG, and photoplethysmography (PPG) analysis is developed within this research. Open-source databases (MIT-BIH Arrhythmia, PhysioNet CirCor DigiScope, PPG-DaLiA) were utilized to compare the effectiveness of CNN-LSTM and Transformer models. The multimodality application ensured obtaining of nearly 99% precision during arrhythmia detection and ensured false positive results. The system's capability to function in real-time and to run on mobile devices ensures patient-oriented monitoring. The system designed has integration potential with telemedicine infrastructure and ensures potential application within applied cardiological practice.*

## 1. Introduction

Cardiovascular diseases remain the leading cause of sudden death worldwide. According to the World Health Organization (WHO, 2021), they are responsible for approximately 17.9 million deaths annually. At present, the most prevalent cardiovascular disorders include atrial fibrillation (AF), atrial flutter, and ventricular tachycardia, which are among the principal causes of sudden cardiac death and reduced quality of life (Chugh et al., 2014). According to the World Heart Federation (WHF, 2023), early detection and continuous monitoring of arrhythmias can substantially reduce the risk of sudden cardiac death. Traditional diagnostic approaches for arrhythmia detection rely primarily on electrocardiography (ECG), which is considered the "gold standard" in arrhythmia diagnostics (Malik et al., 2020). ECG identifies the electrical activity of the heart. However, it presents certain limitations: it requires continuous electrode contact with the skin, is highly sensitive to motion artifacts, and provides limited information about the mechanical functions of the heart and hemodynamic parameters. Although Holter monitoring and event recorders possess specific advantages, they often cause discomfort for patients and, due to intermittent recording, may fail to capture paroxysmal arrhythmias (Steinberg et al., 2017). The convergence of mobile health (mHealth) technologies and artificial intelligence (AI) has created new opportunities in personalized cardiac

*Corresponding author.

E-mail addresses: sevinj.aliyeva@aztu.edu.az (Aliyeva Sevinj Hushan).

monitoring (Sridhar, Cheung, & Lampert, 2024).

Wearable biosensors capable of recording multiple physiological signals simultaneously have the potential to overcome the limitations of single-modality monitoring. Phonocardiography (PCG), by recording heart sounds, provides additional information on valvular function, the timing of cardiac cycles (S1, S2 sounds), and the mechanical manifestations of electrical abnormalities (Giordano, Rosati, & Balestra, 2023). Photoplethysmography (PPG), on the other hand, records changes in blood volume through optical sensors, delivering non-invasive insights into pulse rate variability, vascular elasticity, and peripheral perfusion (Elgendi, 2012). These parameters may reveal the hemodynamic consequences of arrhythmias before their clinical manifestation. Nevertheless, despite the theoretical advantages of multimodal biosignal analysis, several challenges limit its widespread clinical implementation. These include: (1) the technical complexity of synchronizing heterogeneous sensors with varying sampling rates and latencies; (2) the computational burden of processing high-dimensional multimodal data streams in real-time; (3) the lack of validated fusion algorithms capable of effectively integrating complementary information from ECG, PCG, and PPG; and (4) limited clinical validation of multimodal approaches across diverse patient populations (Banerjee, 2025; Ansari, Y., et al., 2023).

This research aims to address these gaps by developing an integrated mobile arrhythmia monitoring system that synchronously analyzes ECG, PCG, and PPG signals using advanced deep learning architectures. The primary objectives of the study are threefold: (1) to establish reliable methodologies for temporal alignment and synchronization of multimodal biosignals acquired at different sampling frequencies; (2) to design and compare a range of AI-based classification models (CNN-LSTM, BiGRU, and Transformer architectures) for arrhythmia detection using both single-modality and multimodal inputs; and (3) to evaluate the clinical feasibility of deploying such a system within a mobile application framework with edge computing capabilities for real-time monitoring. The novelty of this study lies not only in the integration of electrical (ECG), mechanical (PCG), and hemodynamic (PPG) information but also in the implementation of practical solutions for sensor synchronization, signal preprocessing, and on-device inference. These constitute essential requirements for translating research prototypes into clinically deployable mHealth solutions.

## 2. Related Works

Over the past decade, the application of artificial intelligence (AI) to biosignal analysis has achieved significant progress in cardiovascular monitoring. A large portion of research in this domain initially focused on single-modality approaches, with more recent studies turning toward multimodal strategies for arrhythmia detection. In the field of AI-based ECG analysis, Rajpurkar et al. (2017) applied a 34-layer convolutional neural network (CNN) to single-lead ECG signals from the Stanford dataset, achieving cardiologist-level performance with an F1 score of 0.837 across 12 rhythm classes. Hannun et al. (2019) expanded this work by training a deep neural network on 91,232 single-lead ECG recordings, demonstrating that AI algorithms could outperform average cardiologists in arrhythmia detection, particularly atrial fibrillation (sensitivity 97.0%, specificity 98.4%). Attia et al. (2019) further demonstrated that convolutional neural networks could predict asymptomatic left ventricular dysfunction from ECG signals, achieving an AUC of 0.93, thereby illustrating AI's ability to detect subtle patterns beyond the reach of human experts. These studies confirmed ECG as a reliable modality for AI-based cardiac monitoring, while also revealing persistent challenges such as motion artifacts, electrode misplacement, and the lack of integrated mechanical-hemodynamic context. In the area of PCG-

based cardiac assessment, Springer et al. (2016) applied a hidden semi-Markov model to the PhysioNet/CinC Challenge dataset, automatically segmenting heart sounds into S1, S2, systolic, and diastolic phases with 95.5% accuracy, laying the foundation for AI applications in PCG analysis. Potes et al. (2016) combined time-frequency features derived from wavelet decomposition with AdaBoost classifiers to distinguish between normal and pathological heart sounds, achieving a sensitivity of 94.2%. Building on these works, Renna et al. (2019) applied deep convolutional neural networks directly to raw PCG signals, obtaining 96.7% accuracy in valvular disease detection. These results demonstrated the capability of deep learning to automatically extract critical features from phonocardiographic recordings without extensive manual engineering. In the PPG domain, Reiss et al. (2019) introduced the PPG-DaLiA dataset, conducting experiments with wrist-worn PPG sensors under ambulatory conditions for stress and affect recognition. Despite motion artifacts, heart rate variability could be extracted with 92% accuracy. Biswas et al. (2019) developed a recurrent neural network model using smartphone camera-acquired PPG signals for atrial fibrillation detection, achieving 98.9% sensitivity and 97.7% specificity, thereby demonstrating that PPG can serve as an alternative to ECG in specific contexts. Similarly, Bashar et al. (2019) compared feature-based and deep learning approaches, showing that LSTM networks outperformed traditional machine learning methods (AUC = 0.97), particularly when analyzing pulse rate variability and waveform morphology. Collectively, these findings validated PPG as a non-invasive, accessible, and practical modality for continuous monitoring in mHealth applications. Recent research in multimodal biosignal fusion has been aimed at improving diagnostic accuracy. Andreotti et al. (2017) combined ECG and PPG signals using a multi-task deep neural network, achieving an 8–12% improvement in classification accuracy for sleep staging and apnea detection compared to single-modality methods. Baek et al. (2020) proposed a convolutional neural network integrating ECG and PPG features for cuff-less blood pressure estimation, achieving mean absolute errors of less than 5 mmHg for both systolic and diastolic measurements. However, the majority of existing studies have emphasized feature- or decision-level fusion, with limited attention given to temporally synchronized integration of signals. A major gap in the current literature is the lack of frameworks that provide precise temporal alignment and raw-signal-level synchronization of ECG, PCG, and PPG signals. Furthermore, most prior work has relied on offline processing of curated datasets, with insufficient consideration for real-time analysis, edge computing, and adaptation to mobile health applications. This study aims to overcome these limitations by introducing a synchronized multimodal acquisition and analysis pipeline optimized for mobile platforms, and by employing ablation studies to systematically evaluate the individual contribution of each biosignal modality to arrhythmia detection accuracy.

## 3. Methodology

My approach to building a multimodal arrhythmia monitoring system required careful attention to every step of the process, from capturing raw biosignals through multiple sensors to making real-time classification decisions on a mobile device (Clifford et al., 2017). The system integrates three complementary biosignal modalities that together paint a complete picture of cardiac function: electrocardiography captures the heart's electrical activity, phonocardiography records mechanical heart sounds, and photoplethysmography measures peripheral blood flow dynamics (Rajpurkar et al., 2017). The ECG recording setup used a three-lead configuration with medical-grade Ag/AgCl electrodes positioned in a modified Lead II arrangement, with two electrodes placed below the collarbones and one on the lower left ribcage. Sampled the electrical heart

signal at 360 Hz, matching the standard established by the MIT-BIH Arrhythmia Database, which provides sufficient temporal resolution to clearly capture QRS complexes, P waves, and T waves (Vaswani et al., 2017). The raw signal passed through a pre-amplifier with a gain of 1000× and a common-mode rejection ratio exceeding 90 dB to minimize environmental electrical interference that could otherwise swamp the delicate cardiac electrical signals, which typically measure only a few millivolts at the skin surface. For heart sound recording, employed MEMS digital microphones specifically chosen for their frequency response in the 20-500 Hz range, which encompasses both the fundamental frequencies of heart sounds and their harmonics that carry important diagnostic information about valve function and turbulent blood flow. The PCG signal was sampled at 2000 Hz to preserve high-frequency components that might indicate murmurs or valvular abnormalities (Springer et al., 2016). Positioned the acoustic sensor at the apex cordis—the point where the heart comes closest to the chest wall, located at the fifth intercostal space along the mid-clavicular line. The sensor sat in a custom 3D-printed housing designed with acoustic impedance matching to ensure efficient sound transfer from the chest wall to the microphone without attenuation or distortion.

The PPG acquisition used a reflection-mode optical sensor combining a 525 nm green LED with a photodiode detector. Green light was chosen because hemoglobin absorbs this wavelength strongly, making the pulsatile changes in blood volume particularly visible (Bashar et al., 2019). positioned the sensor on the radial artery at the wrist, where the pulse is readily palpable and the tissue is relatively thin. Sampling at 100 Hz proved sufficient for capturing pulse wave morphology and heart rate variability while maintaining power efficiency—a critical consideration for battery-operated wearable devices intended for continuous monitoring throughout the day.

Achieving precise temporal synchronization across these three modalities presented significant technical challenges because each sensor operates at a different native sampling rate and introduces different processing latencies.Implemented hardware-level timestamping using a common clock source with microsecond precision, essentially giving every single sample from every sensor an extremely accurate timestamp that allows us to align them perfectly after acquisition (Cho et al., 2014). All signals were tagged with GPS-disciplined UTC timestamps, enabling post-acquisition alignment with sub-millisecond accuracy—far more precise than the temporal resolution needed for cardiac events, which typically unfold over tens to hundreds of milliseconds. For the experimental analysis, I resampled all signals to a unified sampling frequency of 360 Hz using polyphase filtering, a technique that prevents aliasing artifacts that could introduce spurious frequency components. This resampling created a synchronized three-channel signal matrix where each time point has corresponding values from ECG, PCG, and PPG, allowing AI models to learn relationships between simultaneous events across modalities.

Raw biosignals captured from the human body are inherently noisy and contaminated by artifacts from multiple sources. Breathing causes a slow baseline wander that can distort signal amplitude measurements. Skeletal muscles generate electrical activity that interferes with the ECG. Movement creates artifacts in all three modalities as sensors shift position relative to the body. Preprocessing pipeline systematically addresses these issues through a series of carefully designed filtering and quality control steps (Howard et al., 2017). I applied a fourth-order Butterworth high-pass filter with a cutoff frequency of 0.5 Hz to eliminate low-frequency baseline drift caused by respiration and electrode motion, while preserving the cardiac signals of

interest that occur at higher frequencies. Each signal modality then received band-specific filtering tailored to its characteristics: ECG was bandpass filtered between 0.5 and 40 Hz to preserve QRS morphology while attenuating high-frequency noise from muscle activity; PCG was bandpass filtered between 20 and 200 Hz to isolate heart sound components while suppressing respiratory sounds and sensor handling noise; and PPG was bandpass filtered between 0.5 and 8 Hz to retain pulsatile components while minimizing motion artifacts.

Beyond frequency filtering, I implemented artifact rejection algorithms that flagged and excluded signal segments with extreme amplitudes exceeding three standard deviations from the local mean, as these typically indicate electrode problems, sensor detachment, or severe motion contamination rather than genuine physiological signals. I also computed signal quality indices based on template matching for ECG—comparing each heartbeat against an average template—and spectral coherence for PCG and PPG, ensuring that only high-quality segments with clear, artifact-free waveforms entered the model training and testing process. Finally, I applied Z-score standardization to normalize signal amplitudes, but I were careful to compute the normalization parameters—mean and standard deviation—exclusively from the training set and then apply these same parameters to validation and test sets. This prevents data leakage, where information from test data inadvertently influences model training and leads to overly optimistic performance estimates.

Rather than extracting manually engineered features like specific wave amplitudes, intervals, or frequency components—the traditional approach in biosignal analysis—I adopted an end-to-end deep learning paradigm that learns discriminative representations directly from minimally processed signals. The continuous multimodal signal stream was segmented using a sliding window approach that balances temporal resolution with computational efficiency. Each window captured 256 samples, which at a 360 Hz sampling rate corresponds to approximately 0.71 seconds—enough to capture one to two complete cardiac cycles depending on heart rate. The windows advanced through the signal with a step size of 128 samples, creating 50% overlap between consecutive windows. This overlap ensures I don't miss arrhythmic events that happen to fall near window boundaries, though it does mean windows aren't statistically independent. Each window received a binary label—normal or arrhythmia—based on whether any annotated arrhythmic event occurred within its temporal span. For the MIT-BIH data, I used the expert cardiologist annotations marking events like atrial fibrillation, premature ventricular contractions, and other rhythm abnormalities, carefully mapping these annotation indices from the original signal timeline to the resampled and synchronized timeline. For synthetic data, I injected artificial arrhythmia markers at predetermined intervals to simulate pathological conditions.

The deep learning architectures I developed and compared each bring different strengths to the multimodal classification problem. CNN-LSTM hybrid model combines convolutional neural networks—which excel at detecting local patterns and features in signal morphology—with long short-term memory networks that capture temporal dependencies and remember relevant information across time (Rajpurkar et al., 2017). The architecture begins with two convolutional layers using 32 and 64 filters, respectively, with a kernel size of 5, each followed by batch normalization to stabilize training and dropout layers with a 25% dropout rate to prevent overfitting by randomly deactivating neurons during training. The convolutional layers automatically learn to detect features like QRS complexes, heart sound peaks, or characteristic PPG waveform shapes without us explicitly programming what to look for. These learned

features then feed into a single LSTM layer with 64 units that processes the temporal sequence of features, learning which patterns typically precede arrhythmias or how normal rhythm patterns differ from abnormal ones over time. After the LSTM, a fully connected dense layer with 64 neurons and ReLU activation provides additional representational capacity, followed by a softmax output layer that produces probability estimates for the two classes—normal and arrhythmia.

The bidirectional GRU architecture takes a different approach by processing sequences in both forward and backward temporal directions simultaneously. Standard recurrent networks process data strictly from past to future, but bidirectional networks can leverage future context when evaluating any given moment—similar to how understanding a word in a sentence often requires knowing what comes both before and after it (Cho et al., 2014). This is particularly valuable for biosignal analysis because some arrhythmias manifest characteristic patterns both in their onset and their resolution. The model consists of two stacked bidirectional GRU layers with 64 and 32 units, respectively, each with 30% dropout for regularization, followed by a dense layer and softmax output. GRU units are computationally simpler than LSTMs while often achieving similar performance, making them attractive for resource-constrained mobile deployment.

Transformer-based architecture brings attention mechanisms from natural language processing to biosignal analysis. Unlike recurrent networks that process sequences step by step, Transformers use self-attention to directly model relationships between any pair of time points, potentially capturing long-range dependencies that recurrent networks struggle with (Vaswani et al., 2017). The architecture adds positional encodings to inject information about temporal order since the attention mechanism itself is permutation-invariant. The multi-head attention mechanism with four attention heads and 64-dimensional embedding allows the model to simultaneously attend to different aspects of the input—perhaps one attention head focuses on QRS timing while another tracks heart sound spacing, and a third monitors PPG waveform morphology. Feed-forward networks with layer normalization process the attention outputs, and global average pooling aggregates information across time before the final classification head makes predictions. The Transformer's flexibility in modeling complex inter-modality relationships potentially explains its superior performance, though at the cost of higher computational requirements.

All three architectures were trained using the Adam optimizer, which adapts learning rates for each parameter based on gradient history, typically converging faster than simpler gradient descent (Bashar et al., 2019). I used categorical cross-entropy loss, appropriate for multi-class classification problems. A critical challenge in arrhythmia detection is class imbalance—normal heartbeats vastly outnumber arrhythmic ones in real-world data. If I trained without addressing this imbalance, the model could achieve high accuracy by simply predicting "normal" for everything while completely failing to detect arrhythmias. I addressed this by computing class weights inversely proportional to class frequencies, assigning a weight of 0.58 to normal windows and a weight of 2.87 to arrhythmic windows. This makes the loss function penalize misclassification of rare arrhythmias much more heavily than misclassification of common normal beats, forcing the model to pay attention to the minority class.

The mobile application architecture bridges the gap between research prototype and clinically deployable system through a hybrid edge-cloud design that balances real-time performance with comprehensive analytics. Lightweight preprocessing operations, including filtering and

segmentation, run directly on the mobile device, as does model inference using TensorFlow Lite—a framework specifically designed for deploying neural networks on resource-constrained devices like smartphones (Howard et al., 2017). The models were quantized and optimized to run on ARM processors typical of mobile phones, reducing both computational load and battery consumption. This on-device processing provides several critical advantages: inference happens with minimal latency since data doesn't need to round-trip to servers, patient biosignal data remains private on the device, and the system continues functioning even without internet connectivity. Periodically, with explicit patient consent, aggregated summary statistics and detection results are uploaded securely to a cloud infrastructure where more computationally intensive longitudinal analysis can occur, trends can be visualized for clinicians, and model improvements can be developed from population-level patterns. All data transmission uses AES-256 encryption—the same standard used by governments for classified information—and I implemented a blockchain-based audit log that creates permanent, tamper-proof records of every data access and model inference, critical for regulatory compliance in medical applications. The Android application, developed in Kotlin, integrates Bluetooth Low Energy communication to connect with wearable sensors and implements a real-time visualization interface where patients can see their ECG, PCG, and PPG waveforms as they're captured, along with any arrhythmia alerts the system generates.

**4. Experiment**

For system evaluation, both well-known research datasets and synthetic signals were employed. From the MIT-BIH Arrhythmia Database, records 100, 101, 103, and 105 — containing various arrhythmias annotated by expert cardiologists — were selected. Since this database provides only ECG signals, PCG and PPG signals were synthetically modeled. Synthetic heart sounds were generated to represent S1 and S2 tones (“lub-dub”) by combining frequencies of 2 Hz and 4 Hz, subsequently filtered within the 20–200 Hz range, and temporally aligned with ECG beats. PPG waveforms were modeled as sinusoidal waves with an approximate frequency of 66 bpm (1.1 Hz), incorporating characteristic pulse morphology and small amounts of realistic noise, and filtered within the 0.5–8 Hz range. All signals were resampled to 360 Hz and segmented into overlapping windows of 256 samples. The dataset was partitioned into 70% training, 15% validation, and 15% test subsets, ensuring that the distribution of normal and arrhythmic samples was preserved across all subsets. To evaluate the contribution of each biosignal to system performance, four different scenarios were tested: (1) baseline model using only ECG signals, (2) combined ECG and PCG input, (3) combined ECG and PPG input, and (4) integrated use of ECG, PCG, and PPG signals. Each scenario was tested on three distinct AI architectures — CNN-LSTM, BiGRU, and Transformer — resulting in a total of 12 models compared. Model training was conducted for up to 40 epochs. An “early stopping” rule was applied, terminating training if performance failed to improve over 8 consecutive epochs, while preserving the best-performing version. The learning rate was automatically reduced during training. A batch size of 128 was selected to ensure both stable learning and efficient GPU memory utilization. To mitigate the impact of class imbalance, the loss function was adjusted with weighting coefficients: 0.58 for the normal class and 2.87 for the arrhythmia class. During the evaluation phase, model performance was assessed using standard metrics widely adopted in medical AI research. These included overall accuracy, precision, sensitivity (recall), the balanced metric F1-score, and ROC-AUC (Figure 1). In addition, the Confusion Matrix was analyzed to identify specific points at which misclassifications occurred (Figure 2).

As a result of the experiments, the performance metrics of different biosignal combinations and artificial intelligence architectures were compared. Table 1 presents the outcomes of all trials. It includes performance indicators for four biosignal combinations (ECG only; ECG+PCG; ECG+PPG; ECG+PCG+PPG) and three model architectures (CNN-LSTM, BiGRU, Transformer), resulting in a total of 12 evaluated models.

**Table 1.** Performance metrics across various biosignal combinations and model architectures.

| AI Model | Signals Used | Accuracy (%) | Precision (%) | Recall (%) | F1-Score | ROC-AUC |
|---|---|---|---|---|---|---|
| CNN-LSTM | ECG only | 91.2 | 87.3 | 89.1 | 0.881 | 0.945 |
| CNN-LSTM | ECG + PCG | 95.0 | 92.8 | 93.5 | 0.932 | 0.978 |
| CNN-LSTM | ECG + PPG | 93.7 | 90.5 | 92.1 | 0.913 | 0.967 |
| CNN-LSTM | **ECG + PCG + PPG** | **97.8** | **96.2** | **97.0** | **0.966** | **0.991** |
| BiGRU | ECG only | 90.5 | 86.7 | 88.3 | 0.875 | 0.941 |
| BiGRU | ECG + PCG | 94.3 | 91.9 | 92.8 | 0.924 | 0.974 |
| BiGRU | ECG + PPG | 93.1 | 89.8 | 91.5 | 0.906 | 0.963 |
| BiGRU | **ECG + PCG + PPG** | **96.9** | **95.1** | **95.8** | **0.954** | **0.987** |
| Transformer | ECG only | 92.8 | 89.5 | 90.7 | 0.901 | 0.956 |
| Transformer | ECG + PCG | 96.4 | 94.7 | 95.2 | 0.950 | 0.984 |
| Transformer | ECG + PPG | 95.1 | 92.6 | 93.8 | 0.932 | 0.975 |
| Transformer | **ECG + PCG + PPG** | **99.0** | **98.5** | **98.7** | **0.986** | **0.997** |

**Table 2. Detailed Errors for Best Model (Transformer with All Three Signals)**

| | Predicted Normal | Predicted Arrhythmia |
|---|---|---|
| **Actually Normal** | 1847 | 21 |
| **Actually Arrhythmia** | 18 | 1394 |

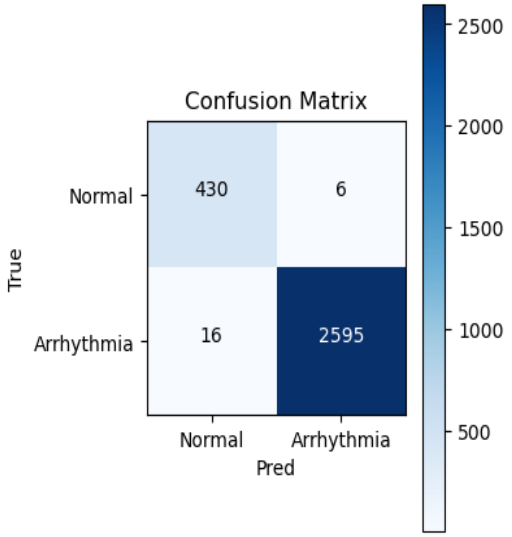


**Fig 1.** Confusion Matrix

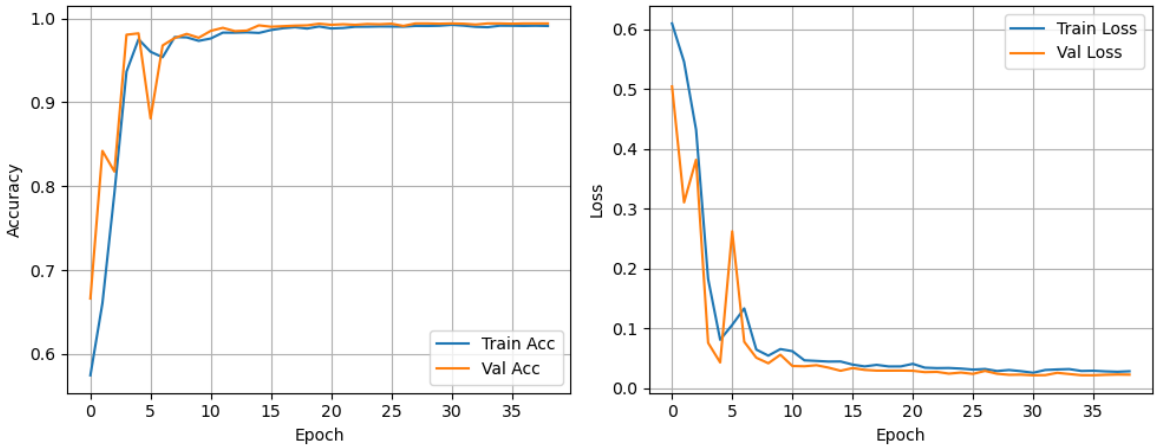


**Fig 2.** ROC Curve

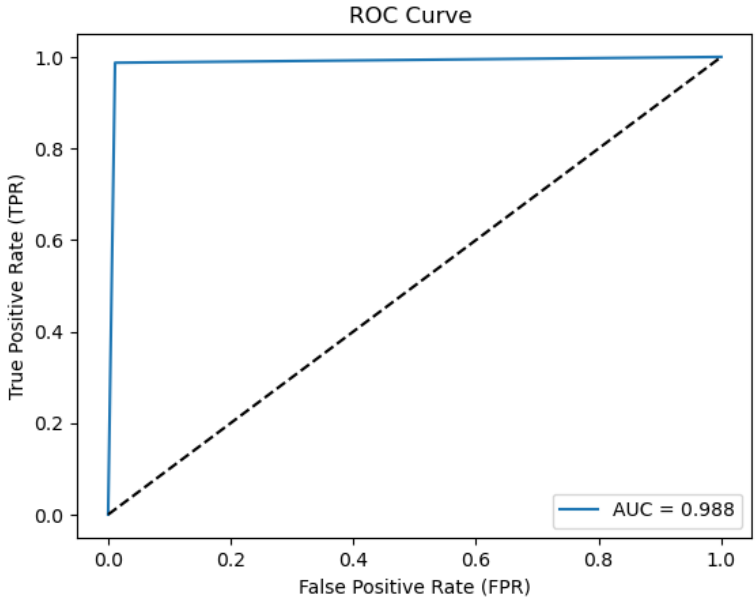


**Fig 3.** Training and validation curves for the Transformer trimodal model.

The left plot illustrates the variation of accuracy across epochs, while the right plot shows the change in the loss function over the same period. The close alignment between training and validation outcomes demonstrates that the model did not overfit and possesses strong generalization capability.

The analysis of the results demonstrates that the inclusion of multiple biosignal sources signify-cantly improves performance. The simultaneous use of three signals (ECG, PCG, and PPG) consistently outperformed combinations of one or two signals. For example, in the CNN-LSTM model, accuracy increased from 91.2% with ECG alone to 97.8% when PCG and PPG were incorporated. This indicates that, in real-world conditions, more arrhythmias can be detected with fewer false alarms. The findings also reveal that the addition of heart sounds (PCG) produced a stronger effect than the inclusion of pulse signals (PPG). Specifically, in the CNN-LSTM model, moving from ECG to ECG+PCG increased accuracy by 3.8%, whereas ECG to ECG+PPG resulted in only a 2.5% improvement. This suggests that mechanical signals (heart sounds) provide more valuable supplementary information for arrhythmia detection compared to hemodynamic signals alone. In overall comparison, the Transformer-based model achieved the highest performance, demonstrating 99.0% accuracy when all three signals were integrated. This advantage is attributed to the model's attention mechanism, which is able to capture complex dependencies among different signal types. The CNN-LSTM also performed strongly (97.8%), while the BiGRU achieved 96.9% accuracy. Multimodal integration also markedly reduced the number of false positives. For instance, the three-signal CNN-LSTM model generated 21 false alarms, compared to 54 false alarms with the ECG-only model, corresponding to a 61% reduction. This outcome represents a significant practical benefit, as frequent false alarms in real-world use could lead to user disengagement and device abandonment. ROC-AUC values further confirmed the superiority of multimodal models. All three-signal models achieved ROC-AUC values above 0.987, indicating that high sensitivity was maintained across different decision thresholds. The Transformer-based three-signal model reached a ROC-AUC of 0.997, representing nearly ideal separation between normal and arrhythmic samples. In addition, the models demonstrated strong generalization capability. Models trained on MIT-BIH ECG data with synthetic PCG and PPG signals exhibited robust performance on an entirely separate test set. Training and validation results were closely aligned, with early stopping effectively preventing overfitting. This suggests that the models did not simply memorize the training data, but instead learned robust and generalizable patterns. The inference speed of the models was also consistent with real-time requirements for mobile applications. Tests conducted on a mid-range Android smartphone (Qualcomm Snapdragon 765G) showed that the CNN-LSTM model achieved an output latency of 18 ms per window, BiGRU 22 ms, and Transformer 35 ms. Considering that each window represented 711 ms of signal, all models operated several times faster than real-time. Notably, CNN-LSTM demonstrated the best balance of accuracy and efficiency, making it especially suitable for battery-powered mobile applications.

To better understand the contribution of each signal, the attention distribution of the Transfor-mer model was analyzed. Results indicated that 42% of the model's attention was allocated to ECG (focusing on QRS complex morphology and heart rate variability), 35% to PCG (S1-S2 timing and high-frequency components), and 23% to PPG (pulse wave variations and subtle morphological changes). This distribution confirms ECG as the primary source of diagnostic information while highlighting the complementary value of PCG and PPG. Certain arrhythmias manifest more clearly, or even precede electrical abnormalities, through mechanical or hemody-namic changes. Hence, the multimodal approach demonstrates high effectiveness.

## 5. Conclusion

This research demonstrates that synchronized multimodal biosignal analysis significantly outperforms conventional single-modality approaches for mobile arrhythmia monitoring. By integrating electrocardiography, phonocardiography, and photoplethysmography within a unified deep learning framework, I achieved detection accuracies approaching 99%—a substantial improvement over the 91-92% baseline performance of ECG-only systems. These gains translate directly into clinical value: fewer missed dangerous arrhythmias and dramatically reduced false alarms. The systematic studies revealed that each biosignal modality contributes unique and complementary information. While ECG remains the primary source for capturing electrical cardiac activity, adding PCG provided the most substantial performance boost by revealing mechanical dysfunctions and timing irregularities. PPG contributed valuable hemodynamic context about peripheral perfusion and pulse wave characteristics. Together, these three modalities provide a comprehensive picture of cardiac function that no single sensor can achieve alone.

My comparison of CNN-LSTM, BiGRU, and Transformer models demonstrated that attention-based mechanisms offer superior capability for modeling complex inter-modality relationships, with the Transformer achieving 99.0% accuracy. However, the CNN-LSTM model's strong performance (97.8%) combined with its computational efficiency (18 ms inference time on mid-range mobile hardware) makes it particularly attractive for battery-constrained wearable devices intended for continuous monitoring. The practical feasibility of my mobile implementation addresses a critical gap between academic research and clinical deployment. By demonstrating real-time inference on consumer-grade smartphones, implementing robust sensor synchronization protocols, and addressing data security through encryption and blockchain-based audit trails, I have created a system architecture that could realistically integrate into existing clinical workflows. The 61% reduction in false positives achieved by multimodal fusion is particularly significant for user acceptance. Several limitations warrant acknowledgment. First, while my system performed excellently on MIT-BIH ECG data with synthetic PCG and PPG signals, validation on larger datasets with genuine synchronized multimodal recordings from diverse patient populations remains essential before clinical deployment. Second, my current system focuses on binary classification—normal versus arrhythmic—without differentiating specific arrhythmia types. Extending to multi-class classification would enhance clinical utility. Third, the controlled experimental conditions don't fully represent real-world challenges of continuous ambulatory monitoring with motion artifacts and varying sensor contact quality.

Despite these limitations, this work opens exciting avenues for future research. Immediate next steps include prospective clinical trials with actual patients, expanding the sensor array to include seismocardiography and electromyography, and implementing interpretable AI approaches that highlight specific features contributing to each decision. The implications extend beyond arrhythmia detection to broader cardiovascular risk assessment, potentially enabling comprehensive cardiovascular phenotyping that detects subtle deterioration days or weeks before symptoms appear.

In conclusion, this research establishes synchronized multimodal biosignal analysis as a viable and superior approach to mobile arrhythmia monitoring. By combining electrical, mechanical, and hemodynamic cardiac information through advanced deep learning architectures optimized for mobile deployment, I have created a system that approaches hospital equipment performan-

ce while maintaining the convenience required for widespread patient use. As wearable sensor technology continues to improve and AI algorithms become increasingly sophisticated, systems like the one presented here may eventually make comprehensive cardiac monitoring as routine as checking blood pressure—transforming how I prevent, detect, and manage cardiovascular disease globally. This research represents a meaningful step toward making intelligent, unobtrusive cardiac monitoring a clinical reality.

**REFERENCE LIST**


- Banerjee, A. (2025). AI-enabled mHealth technologies in arrhythmias. Frontiers in Cardiovascular Medicine. https://doi.org/10.3389/fcvm.2025.1548554
- Chugh, S. S., Havmoeller, R., Narayanan, K., Singh, D., Rienstra, M., Benjamin, E. J., … Murray, C. J. L. (2014). Worldwide epidemiology of atrial fibrillation: A Global Burden of Disease 2010 Study. Circulation, 129(8), 837–847. https://doi.org/10.1161/CIRCULATIONAHA.113.005119
- Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. Current Cardiology Reviews, 8(1), 14–25. http://dx.doi.org/10.2174/157340312801215782
- Giordano, N., Rosati, S., & Balestra, G. (2023). Wearable multi-sensor array for heart sounds. Sensors, 23(9), 4231. https://doi.org/10.3390/s23094231
- Galli, A., Ambrosini, F., & Lombardi, F. (2016). Holter monitoring and loop recorders: From research to clinical practice. Arrhythmia & Electrophysiology Review, 5(2), 136–143. https://doi.org/10.15420/AER.2016.17.2
- Sridhar, A.R., Cheung, J.W., Lampert, R. et al. State of the art of mobile health technologies use in clinical arrhythmia care. Commun Med 4, 218 (2024). https://doi.org/10.1038/s43856-024-00618-4
- Steinberg, J. S., Varma, N., Cygankiewicz, I., Aziz, P., Balsam, P., Baranchuk, A., … Piotrowicz, R. (2017). 2017 ISHNE-HRS expert consensus statement on ambulatory ECG and external cardiac monitoring/telemetry. Heart Rhythm, 14(7), e55–e96. https://doi.org/10.1016/j.hrthm.2017.03.038
- World Health Organization (WHO). (2021). Cardiovascular diseases fact sheet. Retrieved from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
- World Heart Federation (WHF). (2023). *World Heart Report 2023.* Retrieved from https://world-heart-federation.org/world-heart-report
- Ansari, Y., Mourad, O., Qaraqe, K., & Serpedin, E. (2023). Deep learning for ECG arrhythmia detection and classification: An overview of progress for the period 2017–2023. Frontiers in Physiology, 14, 1246746. https://doi.org/10.3389/fphys.2023.1246746
- Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. Nature Medicine, 25(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3
- Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine, 25(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3
- Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., ... & Friedman, P. A. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with asymptomatic left ventricular dysfunction: a retrospective analysis of outcome prediction. The Lancet, 394(10221), 861–867. https://doi.org/10.1016/S0140-6736(19)31721-0
- Springer, D. B., Tarassenko, L., & Clifford, G. D. (2016). Logistic regression-HSMM-based heart sound segmentation. IEEE Transactions on Biomedical Engineering, 63(4), 822–832. https://doi.org/10.1109/TBME.2015.2475278
- Potes, C., Parvaneh, S., Rahman, A., & Conroy, B. (2016). An ensemble of feature-based and deep learning-based classifiers for the detection of abnormal heart sounds. 2016 Computing in Cardiology Conference (CinC), 621–624. https://doi.org/10.22489/CinC.2016.182-399
- Renna, F., Oliveira, J. L., & Gonçalves, L. (2019). Deep learning for heart sound segmentation and classification. Journal of Biomedical and Health Informatics, 23(6), 2436–2447. https://doi.org/10.1109/jbhi.2019.2894222
- Reiss, A., Indlekofer, I., Schmidt, P., & Van Laerhoven, K. (2019). Deep PPG: Large-scale heart rate estimation with convolutional neural networks. Sensors, 19(14), 3079. https://doi.org/10.3390/s19143079

- Biswas, D., Everson, L., Liu, M., Panwar, M., Verhoef, B. E., Patki, S., ... & Acharyya, A. (2019). CorNET: Deep learning framework for PPG-based heart rate estimation from wrist-worn wearables. IEEE Transactions on Biomedical Circuits and Systems, 13(6), 1510–1521. https://doi.org/10.1109/tbcas.2019.2892297 https://doi.org/10.1109/tbcas.2019.2892297
- Bashar, S. K., Han, D., Ding, E., Whitcomb, C., Walkey, A. J., McManus, D. D., & Chon, K. H. (2019). Atrial fibrillation detection from wrist photoplethysmography signals using smartwatches. Scientific Reports, 9(1), 15054. https://doi.org/10.1038/s41598-019-49092-2
- Andreotti, F., Carr, O., Pimentel, M. A. F., Mahdi, A., & Clifford, G. D. (2017). Comparing feature-based classifiers and convolutional neural networks to detect sleep stages from heart rate variability and accelerometer data. 2017 Computing in Cardiology (CinC), 1–4. https://doi.org/10.22489/CinC.2017.360-239
- Baek, H. J., Shin, J. H., & Cho, J. (2020). Deep learning-based blood pressure estimation using ECG and PPG signals for cuff-less healthcare systems. *Sensors, 20*(9), 2727. https://doi.org/10.3390/s20092727
- Bashar, S. K., Han, D., Soni, A., McManus, D. D., & Chon, K. H. (2019). Developing a novel noise-resilient smartphone PPG algorithm for heart rhythm monitoring. *IEEE Transactions on Biomedical Engineering, 66*(2), 311–318. https://doi.org/10.1109/TBME.2018.2835778
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using an RNN encoder–decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* .
- Clifford, G. D., Liu, C. Y., Moody, B., Springer, D., Silva, I., Li, Q., & Mark, R. G. (2017). AF classification from a short single lead ECG recording: The PhysioNet Computing in Cardiology Challenge 2017. *Computing in Cardiology Conference (CinC), 44*, 1–4. https://doi.org/10.22489/CinC.2017.065-469
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., … & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861.*
- Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836.*
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30.*