Real-Time Multimodal ECG Arrhythmia Detection

Chandrapati Tejasai
Department of Computing Technologies
SRM Institute of Science and Technology
Chennai, India
tc6855@srmist.edu.in

Rajula Pratheesh Reddy
Department of Computing Technologies
SRM Institute of Science and Technology
Chennai, India
pr0510@srmist.edu.in

J Selvin Paul Peter
Department of Computing Technologies
SRM Institute of Science and Technology
Chennai, India
selvinpj@srmist.edu.in

B Priya Esther
Department of Mechatronics Engineering
SRM Institute of Science and Technology
Chennai, India
priyaesther@srmist.edu.in

***Abstract*—Cardiovascular diseases continue to be the leading cause of death worldwide, emphasizing the urgent need for ECG analysis systems that are not only accurate but also fast and easily accessible. In this work, we propose a real-time multimodal deep learning framework designed to detect arrhythmias using both 1D ECG signals and 2D clinical ECG images. The system combines two complementary neural network pipelines. The first is a lightweight three-block 1D Convolutional Neural Network (CNN) trained on the MIT-BIH Arrhythmia Database, which contains 109,438 beats categorized into five AAMI classes. This model achieves a test accuracy of 99.27% along with a weighted F1-score of 0.99. The second pipeline is a fine-tuned ResNet-18 model trained on a curated dataset of 491 clinical ECG images spanning four diagnostic categories: Normal, Abnormal Heartbeat, Myocardial Infarction (MI), and Post-MI History. Using a two-phase transfer learning strategy with class imbalance handling, this model achieves a validation accuracy of 84.8%. Additionally, a computer vision-based pipeline is introduced, which uses Otsu thresholding and connected component analysis to extract rhythm signals from clinical ECG images for beat-level inference. The complete system is deployed using a FastAPI backend with a WebSocket-based streaming interface, enabling real-time predictions with latency below 20 ms and a throughput of 24 predictions per second from USB or serial ECG devices. A user-friendly web interface is also developed to provide continuous ECG visualization, beat-wise predictions, diagnostic confidence scores, and real-time heart rate estimation in beats per minute (BPM). Overall, this framework demonstrates the practical feasibility of integrating multimodal deep learning with real-time deployment for AI-assisted cardiac monitoring.**

***Index Terms*—ECG, Arrhythmia Detection, Deep Learning, Multimodal Learning, Real-Time Systems, CNN, ResNet-18, Medical AI**

I. Introduction

Cardiovascular diseases remain a leading cause of death worldwide, highlighting the need for fast and reliable ECG analysis systems. The electrocardiogram (ECG) is widely used for non-invasive cardiac diagnosis, but manual interpretation requires expertise and can be time-consuming, especially in high-volume clinical settings [1].

Traditional approaches relied on handcrafted features such as RR intervals and waveform morphology [2]. With the rise of deep learning, particularly convolutional neural networks (CNNs), it has become possible to automatically learn meaningful patterns directly from raw ECG signals, often achieving superior performance [3], [4].

Despite these advances, many systems are limited to offline analysis and focus on either signal-based or image-based data. Additionally, class imbalance remains a major challenge, often leading to poor detection of minority classes. Furthermore, unified systems capable of handling both raw ECG signals and clinical ECG images in real time are still limited.

To address these challenges, this work proposes a real-time multimodal deep learning framework that integrates both signal-based and image-based analysis within a single system.

The main contributions are as follows:

1) A lightweight 1D CNN achieving 99.27% accuracy for real-time arrhythmia classification.
2) A ResNet-18-based model achieving 84.8% validation accuracy for ECG image classification.
3) A computer vision pipeline for extracting rhythm signals from ECG images.
4) A real-time streaming system with sub-20 ms latency.
5) A web-based interface supporting multiple input modes.

II. Related Work

Early approaches to ECG analysis focused on rule-based and signal processing methods, such as the well-known QRS detection algorithm proposed by Pan and Tompkins [2]. These methods relied on manually engineered features, including RR intervals and waveform morphology.

With the emergence of deep learning, especially convolutional neural networks (CNNs), ECG classification has significantly improved. Models proposed by Acharya *et al.* [3] and Hannun *et al.* [4] demonstrated that neural networks can automatically learn meaningful features directly from raw ECG signals, often achieving performance comparable to clinical experts.

In addition to signal-based approaches, recent studies have explored ECG interpretation from images. Raghunath *et al.*

[11] showed that deep learning models applied to ECG images can capture broader diagnostic patterns, extending beyond arrhythmia detection.

Despite these advancements, most existing systems focus on either signal-based or image-based analysis independently and are often designed for offline use. Real-time multimodal systems that combine both representations remain limited, motivating the approach proposed in this work.

III. System Architecture

The proposed system is a dual-model framework designed for real-time ECG analysis from multiple input sources. Fig. 1 illustrates the overall architecture.

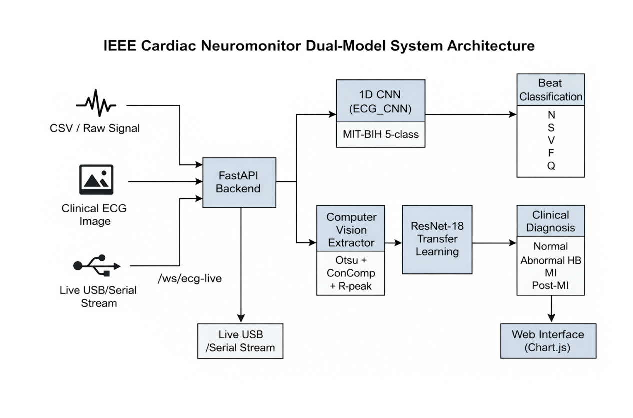

Fig. 1. Overall system architecture of the proposed multimodal ECG analysis framework.

A. Input Modalities

The system supports three input types: raw ECG signals (CSV), clinical ECG images, and live ECG streams from USB or serial-connected devices.

B. Backend

The backend is implemented using FastAPI with WebSocket support for real-time communication. Both the 1D CNN and ResNet-18 models are loaded and executed on the available compute device.

C. Live Inference

Incoming ECG samples are processed using a sliding window of 360 samples. Predictions are generated every 15 samples:

$$f_{\text{pred}} = \frac{360}{15} = 24 \text{ predictions/second}$$

D. Frontend

A web-based interface provides real-time ECG visualization, prediction results, and heart rate (BPM) estimation.

IV. Datasets

A. MIT-BIH Arrhythmia Database

The signal-based model is trained using the MIT-BIH Arrhythmia Database [6], obtained from PhysioNet [7]. This dataset consists of 48 half-hour two-lead ambulatory ECG recordings sampled at 360 Hz, with beat-level annotations provided by expert cardiologists.

To ensure a realistic evaluation and avoid data leakage, the inter-patient evaluation protocol proposed by de Chazal *et al.* [9] is followed.

According to the AAMI EC57 standard [8], heartbeat annotations are grouped into five classes:

- **N** — Normal beats, including left/right bundle branch block, atrial escape, and nodal escape beats
- **S** — Supraventricular ectopic beats (SVEB), including atrial premature and aberrant atrial contractions
- **V** — Ventricular ectopic beats (VEB), including premature ventricular contractions and ventricular escape beats
- **F** — Fusion of ventricular and normal beats
- **Q** — Unknown or paced beats

Each heartbeat is extracted as a 360-sample window centred around the annotated R-peak and subsequently normalised using z-score scaling before being fed into the CNN model.

Table I presents the class-wise distribution of the dataset used in this study.

TABLE I
MIT-BIH Dataset Class Distribution

Class	Full Dataset		Test Set	
	Count	**%**	**Count**	**%**
Normal (N)	90,582	82.8	18,117	82.8
SVEB (S)	2,781	2.5	556	2.5
VEB (V)	7,235	6.6	1,447	6.6
Fusion (F)	802	0.7	160	0.7
Unknown (Q)	8,038	7.3	1,608	7.3
Total	**109,438**	**100**	**21,888**	**100**

B. ECG Image Dataset

To support image-based classification, a curated dataset of clinical ECG images was constructed by merging two publicly available sources and removing duplicates using MD5 hash-based filtering. This process eliminated 4,388 redundant images (approximately 90% of the initial dataset), resulting in a final dataset of 491 unique images across four diagnostic categories.

An 80/20 stratified split was used to create 393 training images and 98 validation images while preserving class distribution.

Table II summarises the dataset composition.

1) Class Imbalance Handling: The dataset exhibits significant class imbalance, with myocardial infarction samples representing only 6.1% compared to 47.5% for abnormal heartbeat images. This imbalance is addressed using weighted random sampling and weighted cross-entropy loss, as discussed in Section IV-B1.

TABLE II
ECG IMAGE DATASET: CLASS DISTRIBUTION AND TRAIN/VALIDATION SPLIT

Class	Total	%	Train	Val
Normal	142	28.9	114	28
Abnormal Heartbeat	233	47.5	186	47
Myocardial Infarction	30	6.1	24	6
Post-MI History	86	17.5	69	17
Total	**491**	**100**	**393**	**98**

V. PROPOSED METHODOLOGY

A. 1D CNN for Signal-Based Arrhythmia Classification

1) Architecture: The signal classification model (`ECG_CNN`) is a lightweight three-block 1D convolutional neural network. Each block includes a convolution layer followed by batch normalization, ReLU activation, and max pooling to progressively extract meaningful features from the ECG signal.

The number of filters increases from 32 to 128 across the blocks to capture hierarchical patterns. The resulting feature maps are flattened and passed through fully connected layers with dropout regularisation, followed by a final 5-class output layer for arrhythmia classification.

2) Preprocessing: Each ECG beat is represented as a 360-sample window centred on the R-peak and normalised using z-score standardisation:

$$\hat{x}_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

where μ and σ denote the mean and standard deviation. If $\sigma = 0$, only mean subtraction is applied.

3) Live Inference: Incoming ECG samples are stored in a rolling buffer of 360 samples. Inference is triggered every $\Delta s = 15$ samples using a sliding window mechanism:

$$f_{\text{pred}} = \frac{f_s}{\Delta s} = \frac{360}{15} = 24 \text{ predictions/second} \tag{2}$$

B. ResNet-18 for Clinical ECG Image Classification

1) Architecture: The image classification module is based on ResNet-18 [5], pretrained on ImageNet. A custom classification head is added, and predictions are obtained as:

$$\hat{\mathbf{y}} = \text{softmax}(W_2\,\text{ReLU}(\text{Dropout}(f(\mathbf{x}),\, p{=}0.4))) \tag{3}$$

Input images are resized to 224×224 pixels.

2) Two-Phase Training: **Phase 1:** Train classifier head (8 epochs, LR $= 10^{-3}$).

Phase 2: Fine-tune full model (12 epochs, LR $= 10^{-4}$) with cosine annealing scheduler.

3) Class Imbalance Handling:

1) **Weighted Sampling**: Balances mini-batches.
2) **Weighted Loss**:

$$\mathcal{L} = -\sum_{j=1}^{C} w_j\, y_j \log \hat{p}_j \tag{4}$$

4) Data Augmentation: Augmentations include horizontal flips, $\pm 10°$ rotations, and brightness/contrast adjustments. Images are normalised using ImageNet statistics.

C. Computer Vision-Based Rhythm Extraction

1) Region cropping
2) Otsu thresholding
3) Connected component selection
4) Morphological processing
5) Signal reconstruction
6) R-peak detection
7) Beat segmentation

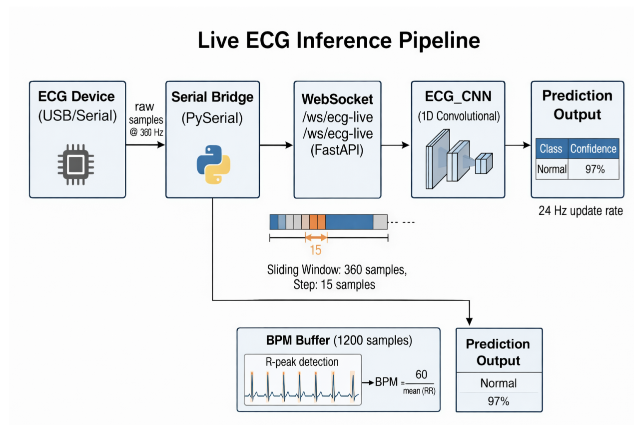

Fig. 2. Live ECG inference pipeline for real-time arrhythmia detection.

D. BPM Estimation

$$\text{BPM} = \frac{60\, f_s}{\Delta_{\text{peaks}}} \tag{5}$$

Values outside 30–250 BPM are discarded.

VI. EXPERIMENTAL RESULTS

A. 1D CNN — MIT-BIH Classification Performance

The proposed 1D CNN model was evaluated on the inter-patient test split consisting of 21,888 beats (20% of the dataset), following the protocol of de Chazal *et al.* [9]. The classification performance in terms of precision, recall, and F1-score is presented in Table III.

The model achieves an overall accuracy of **99.27%** with a **weighted F1-score of 0.99**, demonstrating strong and reliable performance across all arrhythmia classes.

The Normal class achieves perfect precision and recall due to its dominant presence in the dataset. Importantly, the clinically significant VEB class achieves a high F1-score of 0.98, indicating reliable detection of ventricular abnormalities. The Fusion class shows the lowest F1-score (0.86), which can be attributed to its limited samples and similarity with other classes.

Fig. 3 presents the confusion matrix. Most misclassifications occur between SVEB and Normal classes, and between Fusion

TABLE III
1D CNN CLASSIFICATION REPORT ON MIT-BIH TEST SET (21,888 BEATS)

Class	Support	Precision	Recall	F1
Normal (N)	18,117	1.00	1.00	1.00
SVEB (S)	556	0.95	0.92	0.93
VEB (V)	1,447	0.98	0.98	0.98
Fusion (F)	160	0.87	0.86	0.86
Unknown (Q)	1,608	1.00	1.00	1.00
Macro Avg	21,888	0.96	0.95	0.95
Weighted Avg	**21,888**	**0.99**	**0.99**	**0.99**
Overall Accuracy		**99.27%**		

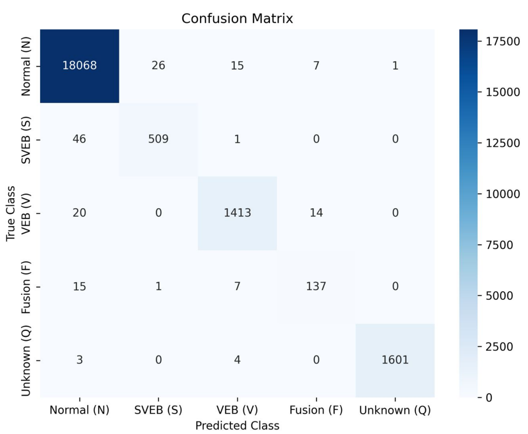

Fig. 3. Confusion matrix for the 1D CNN model. Rows represent true classes, and columns represent predicted classes. Misclassifications are primarily observed between SVEB and Normal classes, and between Fusion and VEB classes.

and VEB classes, which are also challenging distinctions in clinical practice.

Fig. 4 illustrates the training and validation curves over 15 epochs. The model converges rapidly, achieving over 99% training accuracy while maintaining stable validation performance, indicating strong generalisation and minimal overfitting.

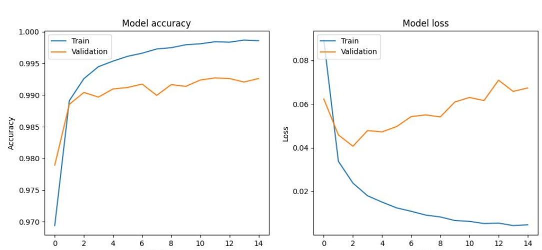

Fig. 4. Training history of the 1D CNN model. Training accuracy exceeds 99%, while validation accuracy stabilises around 99.2%, demonstrating strong generalisation.

B. ResNet-18 — Clinical Image Classification Performance

The ResNet-18 model was trained for 20 epochs (8 epochs for head training and 12 epochs for fine-tuning) using 393 training images. The best model, selected based on validation accuracy, achieved a peak performance of **84.8%**.

Table IV summarises the dataset distribution and validation results.

TABLE IV
RESNET-18 VALIDATION PERFORMANCE FOR ECG IMAGE CLASSIFICATION

Class	Train	Val	Class Weight	Imbalance Ratio
Normal	114	28	$1/114 = 0.0088$	$1.00\times$
Abnormal Heartbeat	186	47	$1/186 = 0.0054$	$1.63\times$
Myocardial Infarction	24	6	$1/24 = 0.0417$	$0.13\times$
Post-MI History	69	17	$1/69 = 0.0145$	$0.37\times$
Total	**393**	**98**		
Best Validation Accuracy			**84.8%**	

Fig. 5 presents the training curves. A noticeable improvement in validation accuracy occurs after epoch 8, corresponding to the transition from head-only training to full fine-tuning. This highlights the importance of adapting pretrained features to domain-specific data.

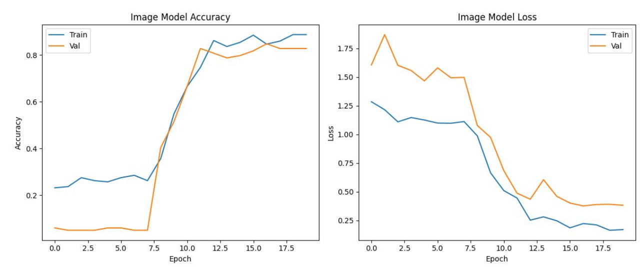

Fig. 5. Training curves for the ResNet-18 model. The transition at epoch 8 marks the start of full fine-tuning, resulting in a sharp improvement in validation accuracy.

C. Live Streaming Latency

The real-time performance of the system was evaluated on a local setup using an Apple Silicon device. Table V summarises key latency metrics.

TABLE V
LIVE ECG STREAMING PERFORMANCE

Metric	Value
ECG device sample rate	360 Hz
Sliding window length	360 samples (1.0 s)
Inference trigger step	15 samples
Prediction update rate	24 Hz
1D CNN inference latency	$<$5 ms
WebSocket round-trip latency	$<$20 ms
BPM buffer length	1,200 samples ($\approx$3.3 s)
R-peak refractory period	80 samples ($\approx$222 ms)
Serial bridge batch size	20 samples

The results confirm that the system achieves real-time performance with minimal latency, making it suitable for continuous ECG monitoring.

D. Dataset Deduplication Impact

The preprocessing pipeline removed 4,388 duplicate images (approximately 90% of the original dataset) using MD5 hash-based filtering. This resulted in a refined dataset of 491 unique images.

This step significantly improves data quality, reduces redundancy, and prevents train-test leakage, leading to more reliable model evaluation.

VII. Discussion and Analysis

The proposed system demonstrates strong performance across both signal-based and image-based tasks. The 1D CNN achieves high accuracy, particularly for dominant classes, while still maintaining reliable detection for clinically important classes such as ventricular ectopic beats.

The ResNet-18 model performs well despite limited data, showing the effectiveness of transfer learning in medical imaging tasks. The improvement observed after fine-tuning highlights the importance of adapting pretrained features to domain-specific data.

A key strength of the system lies in its dual-model design. The signal-based model enables fast, beat-level predictions suitable for real-time monitoring, while the image-based model captures broader diagnostic patterns from full ECG reports. This complementary approach improves overall system capability.

However, some limitations remain. The dataset imbalance, particularly for myocardial infarction samples, may affect generalisation. The system also depends on image quality for accurate rhythm extraction and currently processes only single-lead signals. Addressing these limitations presents opportunities for future improvement.

VIII. Conclusion

This paper presents a real-time multimodal deep learning framework for ECG analysis, combining signal-based and image-based approaches. The 1D CNN achieves 99.27% accuracy, while the ResNet-18 model achieves 84.8% validation accuracy.

The system integrates real-time streaming with low latency, demonstrating its potential for practical deployment. Overall, the framework provides an effective solution for AI-assisted cardiac monitoring.

References

[1] P. Rajpurkar *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65–69, Jan. 2019.

[2] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[3] U. R. Acharya *et al.*, "A deep convolutional neural network model to classify heartbeats," *Computers in Biology and Medicine*, vol. 89, pp. 389–396, 2017.

[4] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65–69, 2019.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.

[6] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May/Jun. 2001.

[7] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.

[8] Association for the Advancement of Medical Instrumentation, *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, ANSI/AAMI EC57:2012(R)2020, Arlington, VA, USA, 2012.

[9] P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[10] J. Schläpfer and H. J. Wellens, "Computer-interpreted electrocardiograms: benefits and limitations," *J. Am. College of Cardiology*, vol. 70, no. 9, pp. 1183–1192, 2017.

[11] S. Raghunath *et al.*, "Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network," *Nature Medicine*, vol. 26, pp. 886–891, Jun. 2020.

[12] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, Feb. 2017.

[13] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[14] L. Clifton *et al.*, "Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 3, pp. 722–730, May 2014.

[15] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[16] P. Virtanen *et al.*, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, pp. 261–272, Feb. 2020.

[17] S. Ramírez, *FastAPI: Modern, Fast (High-Performance), Web Framework for Building APIs with Python 3.6+ Based on Standard Python Type Hints*, 2018. [Online]. Available: https://fastapi.tiangolo.com

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learning Representations (ICLR)*, San Diego, CA, USA, 2015.

[19] Z. Li *et al.*, "BioSignal Copilot: Leveraging the power of LLMs in discovering label-efficient approaches for biosignal subannotation," *arXiv preprint arXiv:2309.11589*, Sep. 2023.