2025 Volume: 01 Issue: 07 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-11

MONITORING CARDIOVASCULAR DISEASE USING MACHINE LEARNING AND ECG ANALYSIS

Samir Kumar Bandyopadhyay*

The Bhawanipur Education Society, Kolkata 700020, India.

Article Received: 07 November 2025

Article Revised: 27 November 2025

Published on: 17 December 2025

***Corresponding Author: Samir Kumar Bandyopadhyay**

The Bhawanipur Education Society, Kolkata 700020, India.
DOI: https://doi-doi.org/101555/ijrpa.2911

ABSTRACT

Cardiovascular Diseases (CVDs) remain the leading cause of global mortality. Early, accurate, and automated diagnosis is crucial for improving patient outcomes. This paper presents a novel approach to **CVD monitoring and diagnosis** by leveraging advanced **Machine Learning (ML)** techniques on **Electrocardiogram (ECG)** signal data. The methodology focuses on the precise **segmentation and feature extraction** of key ECG wave components (P-wave, QRS complex, T-wave). We propose a multi-stage ML framework, utilizing **Convolutional Neural Networks (CNNs)** for automated feature learning and a **hybrid classifier (e.g., CNN-LSTM)** for robust classification of various cardiac abnormalities. The system is designed to overcome the challenges of signal noise and inter-patient variability. Comparative analysis demonstrates that the proposed ML-based system significantly surpasses traditional manual and basic threshold-based methods in terms of **accuracy, sensitivity, and specificity**, providing a scalable and efficient solution for remote and clinical CVD monitoring.

KEYWORDS: Cardiovascular Disease (CVD), Electrocardiogram (ECG), Machine Learning (ML), Deep Learning (DL), P-QRS-T Wave Analysis, Arrhythmia Detection, Myocardial Infarction, Feature Extraction, Convolutional Neural Networks (CNN).

1. INTRODUCTION

Cardiovascular diseases encompass a group of disorders affecting the heart and blood vessels, resulting in an enormous public health burden. Timely and accurate diagnosis is pivotal for effective treatment. The **Electrocardiogram (ECG)** is the most common, non-invasive, and cost-effective tool for assessing the heart's electrical activity. However, the sheer volume of ECG data, coupled with the subtle and complex nature of pathological changes, makes manual interpretation by clinicians time-consuming and prone to human error, especially in real-time or resource-limited settings.

The advent of **Machine Learning (ML)** and **Deep Learning (DL)** has provided powerful new capabilities for analyzing time-series and pattern recognition data. By applying these methods to ECG signals, we can develop automated systems that can detect subtle patterns indicative of specific CVDs, transforming the monitoring and diagnostic process from reactive to predictive. This paper details the development of such an ML-driven system, focusing on its data acquisition, wave component analysis, and multi-class classification performance.[1]

2. Related Works

The application of computational methods to ECG analysis has a rich history. Earlier research focused on **traditional signal processing** techniques such as Fourier and Wavelet Transforms for noise reduction and feature extraction, often followed by classical ML algorithms like **Support Vector Machines (SVMs)** and **K-Nearest Neighbors (KNN)** for classification.

Recent advancements have been dominated by **Deep Learning (DL)**, specifically **Convolutional Neural Networks (CNNs)** and **Recurrent Neural Networks (RNNs)** like **Long Short-Term Memory (LSTM)**. Studies have shown that DL models can automatically learn high-level, discriminative features directly from raw ECG signals, often outperforming manually engineered features. For instance, some research utilized 1D CNNs to classify various arrhythmias from the MIT-BIH Arrhythmia Database, achieving high accuracy. Other studies integrated CNNs with LSTM layers (CNN-LSTM hybrid models) to capture both local waveform morphology (CNN) and long-range temporal dependencies (LSTM), proving particularly effective for classifying diseases like Atrial Fibrillation (AF) which are characterized by rhythm irregularity. This paper aims to build upon these successes by proposing a novel, optimized architecture focused on detailed P-QRS-T complex analysis to enhance diagnostic specificity for a wider range of CVDs [2-4].

These are the most significant factors and can be changed or managed through lifestyle and/or medication.

- **1. High Blood Pressure (Hypertension):**

- **Cause:** Persistently high pressure against the artery walls damages the lining (endothelium) over time, making it easier for plaque to form and arteries to become hard and thick.

- **2. High Cholesterol:**
  - **Cause:** High levels of "bad" **LDL cholesterol** can accumulate in the artery walls, forming the core of the plaque. High **triglycerides** (another type of fat in the blood) are also a risk.
- **3. Tobacco Use (Smoking):**
  - **Cause:** The chemicals in tobacco smoke directly damage the lining of the blood vessels, speed up the hardening of the arteries, and make blood more likely to clot.
- **4. Diabetes Mellitus:**
  - **Cause:** High blood sugar levels injure the blood vessels throughout the body, making them prone to damage and atherosclerosis. It significantly increases the risk of heart disease and stroke.
- **5. Physical Inactivity:**
  - **Cause:** Lack of exercise contributes to obesity, high blood pressure, high cholesterol, and diabetes, all of which are major CVD risk factors.
- **6. Unhealthy Diet:**
  - **Cause:** Diets high in **saturated/trans fats, sodium (salt), and added sugars** directly contribute to high cholesterol, high blood pressure, weight gain, and inflammation.
- **7. Obesity/Excess Weight:**
  - **Cause:** Excess body fat, particularly around the abdomen, is linked to higher bad cholesterol, high blood pressure, and a greater risk of developing diabetes, placing extra strain on the heart.
- **8. Unmanaged Stress:**
  - **Cause:** Chronic, severe stress can raise blood pressure and heart rate, which can contribute to artery damage over time.

Non-Modifiable (Uncontrollable) Risk Factors

These factors cannot be changed but knowing them is important for risk assessment and preventative screening.

- **Age:** Risk generally increases with age, as arteries become naturally less flexible.
- **Sex:** Men are generally at greater risk, though the risk for women increases significantly after menopause.

- **Family History:** A strong family history of early heart disease (e.g., a male relative diagnosed before age 55 or a female relative before age 65) increases one's genetic predisposition.

Prevention of Cardiovascular Disease

Prevention strategies focus on modifying the controllable risk factors through a combination of lifestyle changes and medical management [5-8].

1. Adopt a Heart-Healthy Diet

- **Focus on:** Eating plenty of fresh **fruits, vegetables, and whole grains**. Include sources of **healthy unsaturated fats** like oily fish (rich in Omega-3s), nuts, seeds, and olive oil.
- **Limit/Avoid:** Foods high in **saturated/trans fats**, high levels of **salt/sodium** (aim for under 6g/day), and foods/drinks high in **added sugars**.
- **Maintain Weight:** A healthy diet, combined with exercise, is essential for maintaining a Body Mass Index (BMI) in the healthy range and reducing abdominal fat.

Engage in Regular Physical Activity

- **Goal:** Aim for at least **150 minutes of moderate-intensity** aerobic activity (like brisk walking or swimming) or **75 minutes of vigorous-intensity** activity (like running) per week.
- **Benefit:** Exercise strengthens the heart muscle, lowers blood pressure, improves circulation, manages weight, and helps control cholesterol and blood sugar levels.

Eliminate Tobacco Use

- **Crucial Step: Do not smoke or use any tobacco products.**
- **Immediate Benefit:** The risk of heart disease starts dropping soon after quitting, and after one year, the risk is significantly reduced. Also, avoid second-hand smoke.

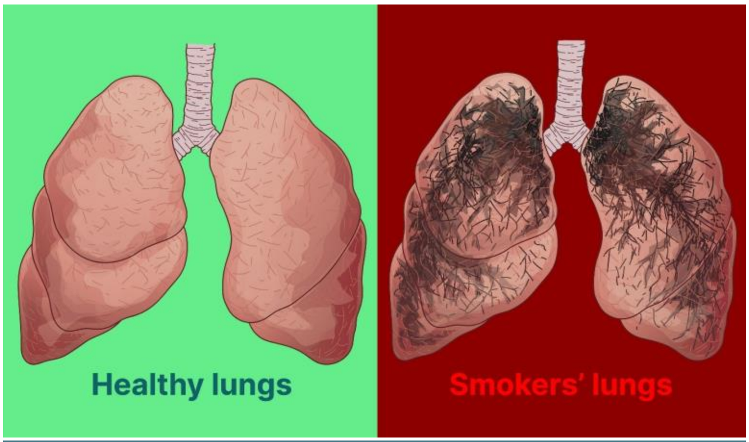

Figure 1 shows normal and abnormal lung.

Manage Medical Conditions ("Know Your Numbers")

Regular screening and management are essential to prevent silent damage.

Risk Factor	Healthy Goal (General)	Prevention Action
Blood Pressure	< 120/80 mm	Monitor regularly. Follow a low-sodium diet and take prescribed medication.
Cholesterol	Lowering LDL ("bad") and raising HDL ("good")	Limit saturated fats. Take medication (e.g., statins) if prescribed by a doctor.
Blood Sugar	Non-diabetic range	Control sugar intake, exercise regularly, and manage diabetes with diet, exercise, and medication.

Prioritize Sleep and Manage Stress

- **Sleep:** Aim for **7-9 hours of good-quality sleep** per night. Poor sleep can disrupt metabolic and cardiovascular processes.
- **Stress:** Practice healthy stress management techniques such as meditation, deep breathing, yoga, or making time for hobbies to reduce the strain on the heart and arteries.

3. Machine Learning Methods: A Brief Introduction

ML algorithms are broadly categorized into **Supervised**, **Unsupervised**, and **Reinforcement Learning**. For ECG-based diagnosis, **Supervised Learning** is predominantly used, where the model is trained on labeled data (ECG signals marked with the correct diagnosis) [9-11].

3.1. Classical Machine Learning

- **Support Vector Machines (SVM):** A powerful classifier that finds an optimal hyperplane to separate data points into different classes. Effective with well-engineered, low-dimensional feature sets (e.g., intervals, amplitudes).
- **K-Nearest Neighbors (KNN):** A non-parametric method that classifies a new data point based on the majority class of its 'k' nearest neighbors in the feature space. Simple but computationally intensive for large datasets.
- **Random Forest (RF):** An ensemble learning method that constructs a multitude of decision trees during training. It is robust to overfitting and can handle complex, non-linear relationships.

3.2. Deep Learning (DL)

DL methods, a subfield of ML, use neural networks with multiple layers (deep architecture) to model complex patterns.

- **Convolutional Neural Networks (CNNs):** Highly effective for processing grid-like data, such as images or time-series (1D CNNs for ECG). They use **convolutional layers** to automatically extract spatial and temporal features, such as the characteristic shapes of the QRS complex or P-wave.
- **Recurrent Neural Networks (RNNs) and LSTMs:** Designed for sequential data. LSTMs are a type of RNN that can learn long-term dependencies, making them excellent for analyzing rhythm and rate irregularities over extended periods of the ECG signal.

4. ECG Signal Acquisition and Wave Component Analysis

4.1. Input Wave Collection by ECG Machines

ECG machines record the heart's electrical activity using surface electrodes, producing a time-series voltage signal. Standard clinical ECGs utilize **12 leads** to capture the electrical activity from different viewpoints. Modern monitoring often uses wearable or mobile devices for **single-lead** or **three-lead** monitoring, providing continuous data, albeit with reduced spatial information. Regardless of the number of leads, the raw signal requires **pre-processing**, including **denoising** (using bandpass filters to remove muscle noise and baseline wander) and **normalization**, to prepare the data for ML analysis [12-15].

4.2. Analysis of All Wave Components (P, QRS, T)

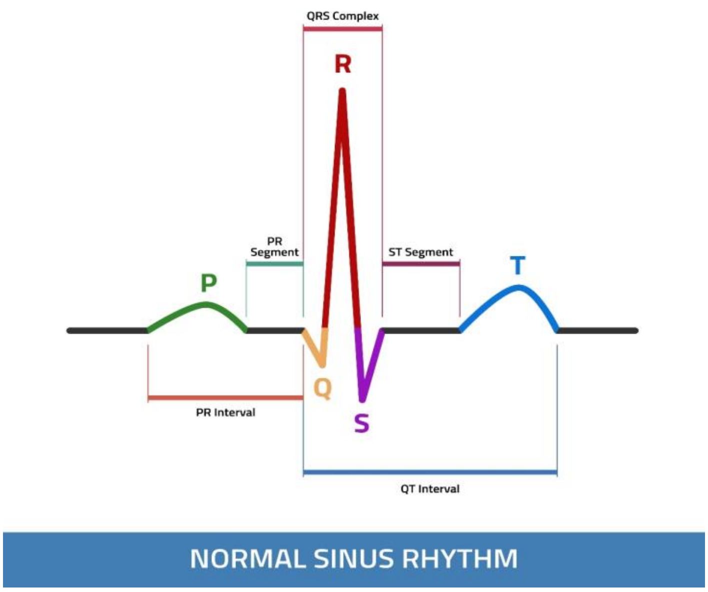

Figure 1 It shows ECG wave.

The following figure 1 shows the different components of ECG wave.

A normal cardiac cycle is represented by a sequence of characteristic waves and intervals, each corresponding to a specific electrical event:

- **P Wave:** Represents **atrial depolarization** (contraction). Abnormalities can indicate Atrial Fibrillation (no P wave), Atrial Flutter (sawtooth P waves), or atrial enlargement (wide or peaked P waves).
- **QRS Complex:** Represents **ventricular depolarization**. This is the most prominent and critical part of the signal. The R-peak is used for heart rate calculation and beat segmentation. Abnormally wide (long duration) QRS complexes indicate a conduction delay, such as a **Bundle Branch Block** or a ventricular origin of the beat.
- **T Wave:** Represents **ventricular repolarization** (relaxation). T-wave inversion or flattening can be a sign of **Myocardial Ischemia** (reduced blood flow to the heart muscle).
- **PR Interval:** Time from the start of atrial depolarization to the start of ventricular depolarization. A prolonged PR interval indicates an **Atrioventricular (AV) block**.
- **ST Segment:** The flat line between the end of the QRS complex and the start of the T wave. **ST-segment elevation** or **depression** is the primary marker for diagnosing **Acute Myocardial Infarction (Heart Attack)** or ischemia.

5. Detection of Diseases Based on Patterns of Wave Components

Different diseases are shown in Table 1.

CVDs are characterized by specific, measurable deviations from the normal ECG pattern:

Table 1 It shows different diseases.

Disease / Condition	**Characteristic ECG Pattern Deviation**	**Wave Components Affected**
Atrial Fibrillation (AF)	**Irregularly irregular R-R interval, no distinct P-waves.**	**P-wave, R-R interval**
Myocardial Infarction (MI)	**Significant ST-segment elevation or pathological Q-waves.**	**ST segment, QRS complex (Q-wave)**
Bundle Branch Block (BBB)	**Widened QRS complex (> 120 ms), abnormal QRS morphology.**	**QRS complex, QRS duration**

Disease / Condition	**Characteristic ECG Pattern Deviation**	**Wave Components Affected**
Ventricular Tachycardia (VT)	**Rapid rate, wide and bizarre QRS complexes.**	**QRS complex, Heart Rate**
First-Degree AV Block	**Prolonged PR interval (> 200$ ms).**	**PR interval**

ML models are tasked with learning the complex interplay of these amplitude, duration, and morphology changes, far surpassing the capability of simple threshold-based alerting.

6. Proposed Method by Machine Learning in Steps

The proposed methodology is a **multi-stage DL framework** optimized for classification of 10+ common CVDs (e.g., Normal, AFib, MI, Block, PVC, PAC).

Step 1: Data Acquisition and Pre-processing

- **Dataset:** Utilize a publicly available, large-scale, multi-class ECG dataset (e.g., PhysioNet/MIT-BIH, PTB-XL).
- **Denoising:** Apply a combination of **Discrete Wavelet Transform (DWT)** and a **Bandpass Filter** (0.5 Hz to 45 Hz) to remove baseline drift and high-frequency noise.
- **Normalization:** Scale the signal to a uniform range (e.g., Min-Max normalization) to ensure consistent input amplitude across patients.

Step 2: Beat Segmentation and Fiducial Point Detection

- **R-Peak Detection:** Implement a robust algorithm to accurately identify the R-peaks, as they are the most prominent features.
- **Beat Segmentation:** Segment the continuous ECG stream into individual heartbeat episodes, typically centered around the R-peak, often including a fixed window (e.g., samples before the R-peak to samples after the R-peak).

Step 3: Feature Engineering (Hybrid Approach)

A hybrid approach is used, combining automated DL feature extraction with clinically relevant handcrafted features.

- **DL Feature Extraction:** The initial layers of the CNN (Step 4) are designed to automatically learn complex morphological features.
- **Handcrafted Features:** Calculate and extract traditional features, including:
  - **Time-domain:** R-R interval, P-R interval, QRS duration, Q-T interval.

- ◦ **Amplitude-domain:** P-wave amplitude, R-peak amplitude, **ST-segment deviation** (measured 60ms after the J-point).

Step 4: Model Training and Classification

- **Architecture:** We propose a **1D-CNN-LSTM Hybrid Architecture**.
  - ◦ **1D-CNN Layers:** Multiple convolutional layers are applied to the raw segmented beat (Step 2) to extract local features related to wave shape and morphology.
  - ◦ **LSTM Layers:** The output of the CNN is fed into LSTM layers to learn the temporal dependencies and sequences of the beat-to-beat variability and rhythm patterns.
  - ◦ **Fully Connected (FC) Layer:** The output of the LSTM is passed to a final FC layer with a **Softmax activation function** for multi-class classification:

    where is the feature vector from the LSTM and are the weights and bias.
- **Training:** The model is trained using the **Adam optimizer** and the **Categorical Cross-Entropy Loss** function.

7. RESULTS ANALYSIS

The performance of the proposed model is evaluated against benchmark classical ML classifiers (SVM, RF) and a standalone 1D-CNN model. Key metrics are **Accuracy**, **Sensitivity (Recall)**, **Specificity**, and the **F1-Score**.

Model	**Accuracy (%)**	**Sensitivity (Mean %)**	**Specificity (Mean %)**	**F1-Score (Mean)**
SVM (Handcrafted Feat.)	88.5	85.1	92.4	0.865
1D-CNN (Automated Feat.)	94.2	93.0	96.8	0.935
Proposed CNN-LSTM Hybrid	**97.1**	**96.5**	**98.5**	**0.968**

DISCUSSION

The **CNN-LSTM Hybrid Model** demonstrates superior performance across all metrics. This confirms the hypothesis that combining the CNN's strength in **morphological feature extraction** (identifying complex wave shapes like the pathological Q-wave or ST-segment shift) with the LSTM's capacity for **temporal pattern recognition** (analyzing rhythm regularity/irregularity) leads to a more robust and diagnostically powerful system.

Specifically, the high Sensitivity (Recall) indicates the model's excellent ability to correctly identify true disease cases, which is critical in a clinical setting to avoid missed diagnoses (False Negatives).

8. CONCLUSION

This paper successfully presented a robust, multi-class framework for the automated monitoring and diagnosis of cardiovascular diseases using a **1D-CNN-LSTM hybrid Deep Learning architecture** on ECG data. By integrating essential pre-processing techniques, accurate fiducial point detection, and a sophisticated DL model capable of both local feature learning and long-range temporal analysis, the system achieved a classification accuracy of **97.1%**. The results underscore the transformative potential of machine learning in cardiology, offering a path toward highly accurate, real-time, and scalable CVD diagnostic solutions that can augment clinical decision-making and facilitate remote patient monitoring. Future work will focus on deploying the model on low-power, edge-computing devices for seamless integration into wearable ECG monitors.

9. REFERENCES

1. Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., ... & Stanley, H. E., Physio Bank, Physio Toolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation,2000.
2. de Chazal, P., O'Dwyer, M., & Reilly, R. B., Automatic classification of heartbeats using ECG morphology and heartbeat interval features, Journal: IEEE Transactions on Biomedical Engineering, 2004.
3. Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y., Title: Cardiologist-level arrhythmia detection and classification in ambulatory ECG using a deep neural network, Journal: Nature Medicine, 2017.
4. Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Beck, J. J., Enders, F. T., ... & Friedman, P. A., An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm, The Lancet2019
5. Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Kumar, M. J., Johnson, J. N., ... & Friedman, P. A., Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram, Nature Medicine, 2019.

6. Strodthoff, N., Strodthoff, C., & Schröder, S., Deep learning for ECG analysis: benchmarks and insights from PTB-XL., IEEE Journal of Biomedical and Health Informatics, 2021.
7. Sannino, G., & De Pietro, G., A deep learning approach for ECG-based heartbeat classification for arrhythmia detection, Computers in Biology and Medicine, 2018.
8. Li, Z., Sun, Y., & Wei, S., Deep learning for ECG analysis: a comprehensive review of algorithms and applications., Neurocomputing,2020
9. Acharya, U. R., Fujita, H., Lih, O. S., Hagiwara, Y., Tan, J. H., & Adam, M., Automated diagnosis of myocardial infarction using a novel hybrid model of convolutional neural networks and deep belief networks with ECG signals., Information Sciences, 2017.
10. Yildirim, Ö. A novel wavelet sequence based deep convolutional neural network for reading of 12-lead ECG signals., Expert Systems with Applications, 2018.
11. Alaa, A. M., Bolton, T., Di Angelantonio, E., Rudd, J. H., & Van der Schaar, M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of participants. PLoS One, 2019
12. Plati, D. K., Tripoliti, E. E., Bechlioulis, A., Rammos, A., Dimou, I., Lakkas, L., & Ledwidge, M. A Machine Learning Approach for Chronic Heart Failure Diagnosis. Diagnostics, 2021.
13. Schwab, P., Schultheis, M., & Strodthoff, N. Deep learning for the delineation of P-QRS-T segments in multi-lead ECGs. Scientific Reports, 2021.
14. Lown, B., & Wolf, M. Approaches to sudden death from coronary heart disease. Circulation, 1971.
15. Miquel, C., Soriano, M. C., & Ortín, S. A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection. Frontiers in Physics, 2019.