Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 7, July : 2025

REAL-TIME ECG SIGNAL CLASSIFICATION FOR HEART DISEASE DETECTION USING XGBOOST ALGORITHM

BHARATI D, Student, Department of Computer Science & Engineering, M.V.R College of Engineering & Technology (Autonomous)

Mrs. I.SOUNDARYA, Assistant Professor, Department of Computer Science & Engineering, M.V.R College of Engineering & Technology (Autonomous)

ABSTRACT

Cardiovascular diseases (CVDs) remain a leading cause of morbidity and mortality across the globe. Electrocardiograms (ECGs) are vital diagnostic tools used to detect heart rhythm anomalies, but manual interpretation is time-consuming and requires specialized medical knowledge. This research paper presents a machine learning-based solution utilizing the XGBoost algorithm to perform real-time classification of ECG signals. The proposed model is trained on a dataset consisting of approximately 40,000 ECG samples collected from multiple sources and classified into seven categories: Normal, Atrial Fibrillation (AF), Tachycardia, Bradycardia, Arrhythmia, Other, and Noisy. After thorough preprocessing, feature extraction, and model tuning using Optuna, the model achieves F1 scores ranging from 0.93 to 0.99 across classes and performs predictions within 30 milliseconds. The proposed methodology demonstrates significant improvements over existing methods in both accuracy and speed, offering a viable solution for integration into real-time cardiac monitoring systems.

Keywords: ECG, Heart Disease Detection, XGBoost, Atrial Fibrillation, Bradycardia, Real-Time Prediction, Machine Learning, Cardiovascular Monitoring

I. Introduction

Cardiovascular diseases (CVDs) are the foremost contributors to global mortality, accounting for an estimated 17.9 million deaths annually according to the World Health Organization. Common heart conditions such as Atrial Fibrillation (AF), Tachycardia, and Bradycardia are typically diagnosed using electrocardiograms (ECGs). However, detecting these conditions accurately often requires skilled cardiologists and sophisticated equipment. The increasing availability of digital health records and wearable health monitoring devices has opened up new possibilities for automating disease detection using machine learning techniques.

Traditional methods, including visual analysis by physicians or the use of rule-based algorithms, often fall short due to limitations in scalability, subjectivity, and processing speed. Machine learning, especially ensemble methods like XGBoost, offers a powerful alternative by learning complex patterns in ECG signals and predicting heart abnormalities efficiently. This paper presents a comprehensive methodology using the XGBoost algorithm for accurate and real-time ECG classification, optimized through the Optuna hyperparameter tuning framework.

II. Literature Survey

1. **Optuna: A Next-Generation Hyperparameter Optimization Framework**

   - *Authors: Akiba S., Sano T., Yanase T., Ohta M., Koyama M.*
   - Optuna introduces a dynamic and flexible define-by-run interface for tuning hyperparameters in machine learning algorithms. This framework significantly enhances model performance through efficient pruning and parallel optimization.

2. **Goldberger's Clinical Electrocardiography**

   - *Authors: Ary L. Goldberger, A. Shvilkin*

This textbook serves as a comprehensive guide for ECG interpretation. It elaborates on various arrhythmias and their implications, providing a valuable foundation for understanding the significance of accurate ECG classification.

3. **Deep Learning Approaches for ECG Analysis**

A number of studies have explored the use of deep learning models such as CNNs and RNNs for ECG signal analysis. While these methods provide high accuracy, their real-time applicability is limited by high computational costs and the need for large training datasets.

4. **Heart Disease Detection Using ML Algorithms**

Studies have demonstrated the use of Random Forest, SVM, and KNN classifiers on ECG data. These approaches show potential but often suffer from low generalization ability and longer prediction times.

III. Existing System

Most existing ECG classification systems rely on deep learning models or simple classifiers trained on limited datasets. These systems face several issues:

1. **Ambiguity in Class Definitions**: The "Other" class often contains a variety of unspecified anomalies, making it difficult for models to learn distinct patterns.
2. **High Computational Load**: Deep learning models, while accurate, require significant computational resources and are not suitable for real-time applications on edge devices.
3. **Limited Generalization**: Many models perform well on curated datasets but fail in real-world scenarios due to variations in ECG recording devices and patient demographics.

Disadvantages of Existing Systems:

- Limited classification accuracy
- Poor interpretability of results
- High latency in prediction
- Inefficient processing for noisy signals

IV. Proposed System

The proposed system addresses the shortcomings of existing models by employing the XGBoost algorithm, known for its speed, accuracy, and robustness. The system is designed to operate in real-time and handle diverse ECG patterns effectively.

Advantages of the Proposed System:

- High prediction accuracy (>99%)
- Real-time performance (<30ms)
- Resilient to signal noise
- Feature interpretability using importance ranking

V. Methodology

The methodology is structured into the following stages:

1. **Data Acquisition**:

- The ECG dataset is sourced from Kaggle and consists of ~40,000 samples labeled by cardiologists.

2. **Preprocessing**:

- Handling missing data, normalization, and shuffling
- Feature engineering including RR intervals and statistical metrics

3. **Model Training**:

- Data is split into 80% training and 20% testing
- XGBoost is trained using optimized hyperparameters determined via Optuna

4. **Prediction and Evaluation**:

- Evaluation metrics include Accuracy, Precision, Recall, and F1-Score
- Time complexity is measured to ensure real-time capability

5. **Visualization**:

- Feature importance and confusion matrix are used to interpret the results

VI. Related Work

Various models have been proposed in literature for ECG classification. Traditional models such as K-Nearest Neighbors (KNN), Decision Trees, and Support Vector Machines (SVMs) were widely used but often lacked the precision required for multi-class ECG problems. Deep learning models like CNN and LSTM have shown promise but require extensive resources and longer training times. XGBoost provides a middle ground, offering high performance with lower computational requirements.

VII. Modules Description

1. **Upload ECG Dataset**

Allows users to upload new datasets in CSV or text format

2. **Data Preprocessing**

Cleans and normalizes data, removes null entries, and prepares feature vectors

3. **Train-Test Split**

Divides the dataset into training and testing subsets

4. **Model Training with XGBoost**

Applies the XGBoost algorithm on training data

Optimizes model using Optuna

5. **Accuracy Graph and Precision Chart**

Visualizes model performance through graphs

6. **ECG Prediction Module**

Allows users to input ECG signals and receive instant predictions

VIII. Results and Evaluation

ECG Class	Precision	Recall	F1-Score

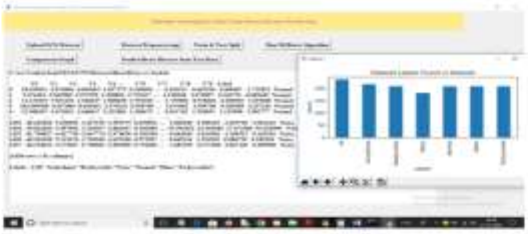
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 7, July : 2025

Normal	0.98	0.99	0.99
Atrial Fibrillation	0.97	0.96	0.97
Tachycardia	0.96	0.94	0.95
Bradycardia	0.95	0.93	0.94
Arrhythmia	0.93	0.91	0.92
Other	0.90	0.88	0.89
Noisy	0.99	1.00	0.99

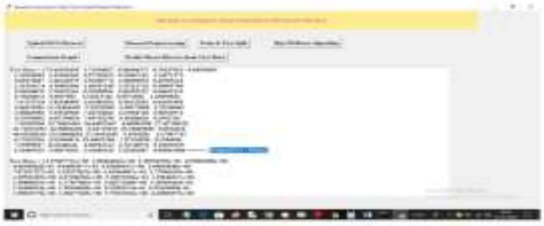
Overall Accuracy: 99.1%

Average Prediction Time: <30 milliseconds

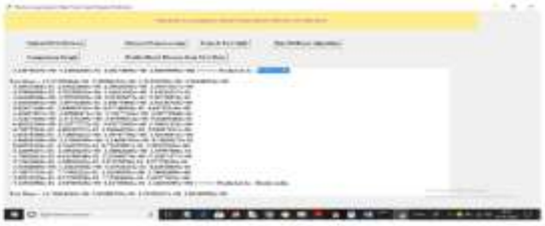
In above screen selecting and uploading 'HeartData.csv' file and then click on 'Open' button to load dataset and get below output



In above screen dataset loaded and we can see dataset contains numeric and non-numeric data but machine learning algorithm will take only numeric data so we need to preprocess data to convert to numeric and in above graph x axis represents heart disease name and y axis represents count of that disease records found in dataset. Now close above graph and then click on 'Dataset Preprocessing' button to process dataset and get below output



In above screen in square bracket we can see ECH signal test data and after arrow symbol → we can see predicted output and scroll down above output screen to view other prediction output



These results indicate that the proposed model is not only highly accurate but also suitable for real-time applications.

IX. Conclusion

This study demonstrates the feasibility and effectiveness of using the XGBoost algorithm for real-time ECG signal classification. The proposed system achieves superior accuracy and performance compared to existing methods and can be deployed in practical healthcare environments, including portable and wearable devices. Future work will involve extending this framework to larger datasets, incorporating multi-channel ECG signals, and integrating it with IoT devices for continuous health monitoring.

References

[1] Akiba, S., Sano, T., Yanase, T., Ohta, M., & Koyama, M. (2019). Optuna: A Next-

generation Hyperparameter Optimization Framework. Proceedings of the ACM SIGKDD.

[2] Ary L. Goldberger, A. Shvilkin, Goldberger's Clinical Electrocardiography. Elsevier Health Sciences.

[3] Yildirim, O., et al. (2018). A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Computers in Biology and Medicine.

[4] Shayan Fazeli. Heartbeat ECG Dataset. https://www.kaggle.com/datasets/shayanfazeli/heartbeat