PREDICTIVE PULSE: MACHINE LEARNING ECG ARRHYTHMIA CLASSIFICATION

Dr.R.K.Krishna[#1], Ringane Sanskruti [#2], Nandekar Tanvi[#3], Padole Tanvi Vinay[#4], Tarhewar Nandini[#5], Kurkhede Alok Rahul[#6], Dange Ansh Amar[#7]

[#]Department of ECE, Rajiv Gandhi College of Engineering, Research and Technology,

Chandrapur, India

Email: rkrishna40@rediffmail.com

ABSTRACT

Electrocardiogram (ECG) arrhythmia classification has become an important research area in intelligent healthcare systems due to the increasing prevalence of cardiovascular diseases and the need for early diagnosis of abnormal heart conditions. This project, titled "Predictive Pulse: Machine Learning ECG Arrhythmia Classification," aims to develop an efficient and accurate automated system for identifying different heartbeat abnormalities from ECG signals. The dataset consists of heartbeat-related features including RR intervals, ECG peak amplitudes, waveform intervals, and QRS morphological characteristics that represent the electrical activity of the heart. In the existing system, the K-Nearest Neighbor (KNN) algorithm is utilized for heartbeat classification, providing baseline performance for arrhythmia detection. To improve classification accuracy and robustness, the proposed system employs the Extra Trees Classifier (ETC), an ensemble-based machine learning algorithm capable of handling complex and high-dimensional ECG features effectively. The dataset undergoes preprocessing, feature extraction, and model training before classification. Experimental analysis shows that the proposed ETC model achieves a higher classification accuracycompared to the existing KNN approach. The proposed system demonstrates improved reliability, faster prediction capability, and better detection of abnormal heartbeats, making it suitable for real-time cardiac monitoring and intelligent healthcare applications.

Keywords: *Electrocardiogram (ECG), Arrhythmia Classification, Machine Learning, Extra Trees Classifier (ETC), K-Nearest Neighbor (KNN), Heartbeat Analysis, Cardiovascular Disease Detection, Feature Extraction, Intelligent Healthcare, Cardiac Monitoring.*

1. INTRODUCTION

Electrocardiogram (ECG) signal analysis plays a significant role in modern healthcare systems for identifying cardiovascular abnormalities and monitoring heart conditions. Heart diseases continue to be one of the major causes of mortality across the world, creating a strong need for accurate and automated diagnostic systems. Manual examination of ECG signals requires expert medical interpretation and consumes considerable clinical time, especially when handling large volumes of patient data. Machine learning techniques provide an efficient solution for analyzing ECG heartbeat patterns and detecting arrhythmias with improved speed and accuracy. ECG signals contain important waveform information such as P wave, QRS complex, T wave, RR intervals, and other cardiac features that help in recognizing abnormal heart activities. Advanced classification algorithms process these features to distinguish between normal and irregular heartbeats effectively. The proposed system utilizes heartbeat interval features, ECG peak amplitudes, waveform intervals, and QRS morphological characteristics for arrhythmia classification. In the existing approach, the K-Nearest Neighbor (KNN) algorithm performs heartbeat prediction based on similarity measures between ECG samples. To improve classification performance, the proposed system implements the Extra Trees Classifier (ETC), which provides enhanced feature handling, faster prediction capability, and better classification accuracy. The system supports early detection of cardiac abnormalities and assists healthcare professionals in clinical decision-making. Automated ECG classification systems also reduce diagnostic workload and improve patient monitoring efficiency in hospitals and healthcare centers.

A. Objective

The primary aim is to develop an intelligent ECG heartbeat classification system capable of detecting normal and abnormal heartbeats using machine learning techniques. The system focuses on analyzing ECG waveform characteristics such as RR intervals, ECG peak amplitudes, waveform durations, and QRS morphological features to identify different arrhythmia categories accurately. The implementation of the Extra Trees Classifier improves prediction accuracy, classification speed, and overall reliability when compared with the traditional K-Nearest Neighbor algorithm. The system supports automated cardiac monitoring and assists medical professionals in achieving faster and more accurate diagnosis of heart abnormalities.

B. Problem Statement

Cardiovascular diseases and heartbeat irregularities require accurate and timely diagnosis to prevent serious health complications. Traditional manual ECG analysis demands expert interpretation and becomes difficult when processing large-scale heartbeat data continuously generated from monitoring systems. Existing classification approaches often experience limitations in prediction accuracy, computational efficiency, and handling complex ECG waveform variations. Inaccurate heartbeat classification affects early diagnosis and delays clinical treatment decisions. An efficient machine learning-based ECG arrhythmia classification system is required to process heartbeat features automatically and provide reliable identification of normal and abnormal cardiac conditions with improved performance and reduced diagnostic complexity.

C. Motivation

The increasing number of cardiovascular patients has created a strong demand for automated and intelligent healthcare monitoring systems capable of detecting heart abnormalities efficiently. Hospitals and diagnostic centers require accurate ECG classification systems that reduce clinical workload and support faster patient diagnosis. Machine learning techniques provide advanced solutions for processing large-scale ECG datasets with improved prediction accuracy and reduced human intervention. The development of ensemble learning algorithms such as Extra Trees Classifier has enhanced the capability of medical data analysis by improving classification performance and computational efficiency. The selection of this research area is driven by the importance of early arrhythmia detection, the growing adoption of artificial intelligence in healthcare, and the need for reliable automated cardiac diagnostic systems that support better patient care and medical decision-making.

2. LITERATURE SURVEY

One of the most fundamental vital organs is the heart. It's the engine that pumps blood to many networks of vessels. The heart moves constantly, beating 100,000 times a day by providing oxygen and nutrients while clearing away harmful waste matter. The beating of the heart produces electrical actions measured on the body surface by an electrocardiogram recording (ECG). Skin electrodes record the electrical activity, exposing how each chamber operates in the form of PQRST waves. Therefore, the morphology and heart rate variability (HRV) extracted from the ECG signal reveal the cardiac behavior. The heart behavior analysis expressed by the electrocardiogram signal provides specific information about the heart. Thus, if the ECG is irregular or faster, or slower than normal, that means cardiac arrhythmia. Arrhythmia can cause several types of consequences; an imminent threat to a patient's life (e.g., ventricular fibrillation and tachycardia), long-term threats, or even more causing death, the thing that made it the most common leading cause of deaths in the worldNormal cardiac rhythm is occasionally interrupted by a beat that occurs before the regular time of the next sinus beat, and this is described as a premature beat or premature contraction (the terms "ectopic beat" and "extrasystole" are frequently used as synonyms). Originally, the sinus beats start from the SA node, unlike the premature beat, which is preceded by an ectopic focus that may be localized in any section of the heart other than the SA node . Thus, the premature beat is classified into two types depending on the location of the focus; Premature Atrial Contraction (PAC) (also known as an atrial premature beat (APB)) if its origin is above the ventricles, i.e., in the atria or the AV node, or Premature Ventricular Contraction (PVC) (also known as a ventricular premature beat (VPB)) if its origin is in the ventricles. We can recognize the PAC and PVC based on specific characteristics and different circumstances. The usual traditional variety of PAC is linked with an abnormal P wave morphology and a QRS complex morphology matching that of a normal sinus beat, but for the associated compensatory pause; the interval between the two sinus beats that enclose the PAC is less than the length of two normal RR intervals. Unlike the PAC, the presence of a PVC almost always prevents the occurrence of the next sinus beat. Although the SA node discharges on schedule, the impulse cannot propagate to the ventricles because the premature beat has made the tissue refractory. The pause that results between the PVC and the next sinus beat is called the compensatory pause.The recent developments in biomedical sensors, the Internet of Medical Things (IoMT), and artificial intelligence (AI)-based techniques have increased interest in smart healthcare technologies [1,2]. Microelectronics, smart sensors, AI, 5G, and IoMT constitute the cornerstone of smart healthcare [3,4]. A smart healthcare system does not suffer fatigue; hence, it can process big data at a much higher speed than humans with greater accuracy [5]. With smart healthcare systems, the diagnosis and treatment of diseases have become more intelligent. For instance, smart patient monitoring empowers the observation of a patient outside the traditional clinical settings, which offers a lower cost through reducing visits to physician offices and hospitalizations [6].

III. EXISTING SYSTEM

The existing system for the project "Predictive Pulse: Machine Learning ECG Arrhythmia Classification" is based on the K-Nearest Neighbor (KNN) algorithm for identifying and classifying abnormal heartbeats from Electrocardiogram (ECG) signals. The primary objective of the existing system is to automatically analyze ECG heartbeat patterns and categorize them into different arrhythmia classes using supervised machine learning techniques. In this approach, ECG heartbeat records are represented using multiple physiological and waveform-based features extracted from ECG signals, including RR intervals, peak amplitudes, waveform durations, and QRS morphology characteristics. These features capture important information related to the electrical activity of the heart and serve as the input for the classification process.

A. Disadvantages

The K-Nearest Neighbor (KNN) algorithm becomes computationally expensive during prediction because it calculates the distance between the test sample and all training samples, which reduces efficiency when the dataset size is large, especially in real-time ECG monitoring systems. Its performance is highly sensitive to the choice of the parameter 'k' and the quality of feature scaling, which can lead to inconsistent or suboptimal classification results for ECG arrhythmia detection if not properly tuned. KNN struggles with high-dimensional ECG feature spaces and imbalanced class distributions, which can reduce its ability to correctly identify minority arrhythmia classes such as SVEB and VEB, leading to lower overall diagnostic reliability.

IV. PROPOSED SYSTEM

The proposed system, titled "Predictive Pulse: Machine Learning ECG Arrhythmia Classification," is designed to enhance the accuracy and efficiency of automated heartbeat classification by leveraging the Extra Trees Classifier (ETC), an advanced ensemble learning technique. The system is developed to analyze ECG arrhythmia data consisting of temporal, amplitude, interval-based, and morphological features extracted from heartbeat signals. These features include RR intervals (pre-RR and post-RR), ECG peak amplitudes (pPeak, qPeak, rPeak, sPeak, tPeak), clinically significant ECG intervals (qrs_interval, pq_interval, qt_interval, st_interval), and QRS morphological descriptors (qrs_morph0 to qrs_morph4), which collectively capture the electrical and structural behavior of cardiac cycles. In the proposed approach, the dataset undergoes preprocessing steps such as handling missing values, feature scaling, normalization, and encoding of categorical labels to ensure uniformity and improve model learning efficiency. After preprocessing, the dataset is split into training and testing sets to evaluate model performance objectively.

A. System Architecture

The system architecture of the ECG Heartbeat Classification System is designed as a modular web-based machine learning pipeline that integrates data management, model training, and real-time prediction within a Flask framework. The architecture follows a layered approach consisting of a presentation layer, application layer, machine learning layer, and data storage layer. The **presentation layer** is the user interface built using Flask templates (HTML forms and dashboards), where users interact with the system. Through this layer, users can register, log in, upload ECG datasets, view exploratory data analysis (EDA) results, train models, and obtain predictions. The **application layer** is implemented using Flask routes and controllers. It manages core functionalities such as authentication, dataset handling, preprocessing, feature selection, model training orchestration, and prediction workflow. This layer acts as the bridge between the user interface and backend processing.

B. Preprocessingpipeline

It is an important stage in the ECG arrhythmia classification system because it improves data quality, removes inconsistencies, and prepares the dataset for accurate machine learning analysis. The preprocessing procedure begins with uploading the ECG heartbeat dataset in CSV format through the Flask-based web application. After loading the dataset, the system reads the data using the Pandas library and removes incomplete or missing records using the dropna() function to ensure that only valid heartbeat samples are processed during training and prediction. The dataset contains multiple ECG-related features such as pre-RR interval, post-RR interval, P wave amplitude, QRS interval, QT interval, ST interval, and QRS morphological characteristics that represent the electrical behavior of the heart. Since machine learning algorithms require numerical target values, the categorical heartbeat labels stored in the “type” column are converted into numerical form using the

LabelEncoder technique. The feature matrix X is created by selecting the first 32 ECG attributes, while the target vector y contains the encoded heartbeat classes. To handle class imbalance present in ECG heartbeat categories, the Synthetic Minority Oversampling Technique (SMOTE) is applied to generate balanced samples for minority heartbeat classes, improving the model's learning capability and classification performance.

C. Software & Hardware Requirements

Software: Windows 11, Python 3.7, TensorFlow 2.x, Keras, NumPy, Pandas, OpenCV, Scikit-learn, Matplotlib. Hardware: Intel Core i5 / Pentium IV 2.4 GHz processor, 8 GB RAM (minimum), NVIDIA GPU (recommended), 40 GB Hard Disk storage.

D. Advantages:

It provides higher classification accuracy by combining multiple decision trees and reducing prediction variance through ensemble learning. It handles high-dimensional ECG feature sets effectively, improving the model's ability to learn complex patterns from RR intervals, peaks, and waveform intervals. It offers strong robustness against overfitting and performs well even on imbalanced ECG datasets by using random feature selection and majority voting.

V. RESULTS AND DISCUSSIONS

The ECG Heartbeat Classification system successfully performs arrhythmia detection using machine learning models such as K-Nearest Neighbors (KNN) and Extra Trees Classifier. After training on the processed ECG dataset, which includes preprocessing, handling class imbalance using SMOTE, and selecting important features using SelectKBest, the models are able to classify heartbeats into categories like Normal sinus, Supraventricular ectopic, and Ventricular ectopic. The results show that the system achieves good accuracy, precision, recall, and F1-score, indicating that it can reliably distinguish between different types of heart conditions. The confusion matrix further helps in understanding the number of correct and incorrect predictions for each class, showing that most samples are classified correctly with minimal misclassification. In the discussion, it can be observed that ensemble-based Extra Trees Classifier generally performs better or more stable compared to KNN due to its ability to handle complex feature relationships, while KNN provides simpler but slightly less robust predictions depending on data distribution.

A. Classification Performance

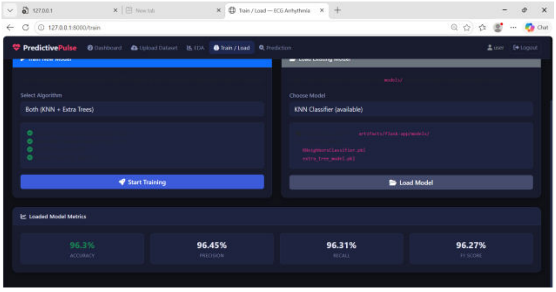

Fig 1: KNN metrics

Figure 1 illustrates the Loaded Model Metrics of the existing K-Nearest Neighbor (KNN) classifier used in the ECG Arrhythmia Classification System. The figure presents important performance evaluation parameters including Accuracy, Precision, Recall, and F1-Score, which are used to measure the effectiveness of the heartbeat classification model. The KNN classifier achieves an accuracy of 96.3%, precision of 96.45%, recall of 96.31%, and F1-score of 96.27%, indicating that the model performs well in identifying different ECG heartbeat categories. Precision measures the correctness of predicted heartbeat classes, recall evaluates the model's ability to detect actual arrhythmia cases, and the F1-score provides a balanced measure of both precision and recall. These results demonstrate that the KNN model provides reliable baseline performance for ECG arrhythmia detection; however, its performance is slightly lower compared to the proposed Extra Trees Classifier model, which achieves higher classification accuracy and robustness.

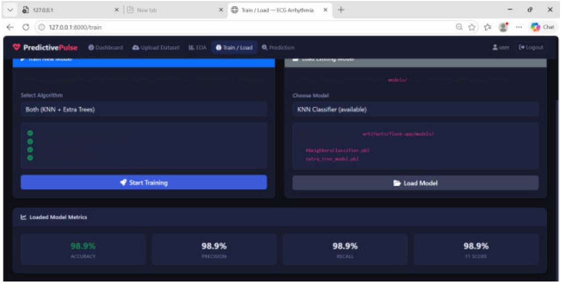

Fig2: ETC Metrics

Figure 2 illustrates the Loaded Model Metrics of the proposed Extra Trees Classifier (ETC) used in the ECG Arrhythmia Classification System. The figure displays the major performance evaluation metrics including Accuracy, Precision, Recall, and F1-Score, all achieving an impressive value of 98.9%. These results indicate that the proposed ETC model provides highly accurate and reliable heartbeat classification performance for detecting ECG arrhythmias.

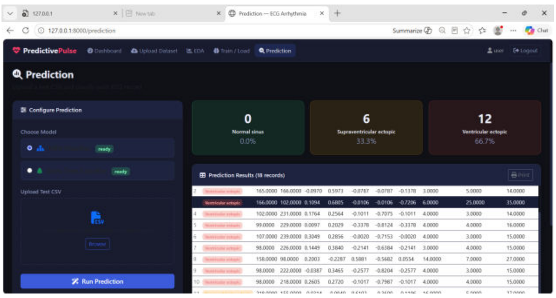

Fig3: Prediction

Figure 3 illustrates the Prediction module of the ECG Arrhythmia Classification System, where new ECG heartbeat records are classified using trained machine learning models. The interface allows users to upload a test CSV file containing ECG features similar to the training dataset and select either the KNN Classifier or the proposed Extra Trees Classifier for heartbeat prediction. The system preprocesses the uploaded test data and applies the selected trained model to classify each ECG record into heartbeat categories such as Normal Sinus, Supraventricular Ectopic Beat (SVEB), and Ventricular Ectopic Beat (VEB).

Comparative Analysis

Table 1 presents the comparative performance analysis of the existing K-Nearest Neighbor (KNN) classifier and the proposed Extra Trees Classifier (ETC) for ECG arrhythmia classification. The KNN model achieved an accuracy of 96.3% and precision of 96.45%, demonstrating strong capability in classifying heartbeat patterns. However, the proposed ETC model significantly improved the classification performance by achieving 98.9% accuracy and 98.9% precision. The superior performance of ETC can be attributed to its ensemble learning mechanism, which combines multiple randomized decision trees to reduce overfitting and improve generalization.

Algorithm	Accuracy (%)	Precision (%)
K-Nearest Neighbor (KNN)	96.30	96.45
Extra Trees Classifier (ETC)	98.90	98.90

C. Test Cases

The functional test cases are designed to verify the proper operation of the ECG Arrhythmia Classification System under different scenarios. The testing process includes uploading the ECG dataset, performing data preprocessing, generating training and testing datasets, executing the KNN and Extra Trees Classifier (ETC) algorithms, loading trained models, and evaluating performance metrics such as accuracy, precision, recall, and F1-score. Additional test cases validate the prediction module by uploading test ECG records, classifying heartbeat patterns into Normal Sinus, Supraventricular Ectopic Beat (SVEB), and Ventricular Ectopic Beat (VEB), and displaying prediction summaries and detailed classification results. If the required input is unavailable, the system does not perform the corresponding operation, ensuring reliable and error-free execution.

TABLE IV. System Test Cases for ECG Arrhythmia Classification System

S.No	Input	If Available	If Not Available
1	Upload ECG heartbeat dataset (CSV)	Dataset loaded successfully	No process
2	Generate train and test datasets	Training and testing data generated	No process
3	Run KNN classification algorithm	KNN accuracy metrics displayed	No process
4	Run Extra Trees Classifier (ETC) algorithm	ETC accuracy metrics displayed	No process
5	Perform ECG data preprocessing	Preprocessed ECG data displayed	No process

V. Conclusion

The project titled "Predictive Pulse: Machine Learning ECG Arrhythmia Classification" successfully demonstrates the development of an intelligent and automated system for detecting and classifying ECG heartbeat abnormalities using machine learning techniques. Cardiovascular diseases and arrhythmias are among the leading causes of health complications worldwide, making early diagnosis and continuous cardiac monitoring extremely important. The proposed system effectively analyzes ECG heartbeat data using important temporal, interval-based, amplitude, and morphological ECG features extracted from cardiac signals. The system integrates multiple stages including dataset uploading, preprocessing, exploratory data analysis, feature selection, model training, and real-time prediction through a Flask-based web application.In the existing system, the K-Nearest Neighbor (KNN) algorithm was utilized as a baseline classifier for arrhythmia detection. Although the KNN model achieved good classification performance, it showed certain limitations in handling high-dimensional ECG features, large datasets, and complex heartbeat patterns efficiently. To overcome these challenges, the proposed system implemented the Extra Trees Classifier (ETC), an ensemble-based machine learning algorithm capable of improving prediction accuracy and reducing overfitting through randomized tree construction and majority voting mechanisms.

References

[1] Ismail, A.R.; Jovanovic, S.; Ramzan, N.; Rabah, H. ECG Classification Using an Optimal Temporal Convolutional Network for Remote Health Monitoring. Sensors 2023, 23, 1697.

[2] Huang, Y.; Li, H.; Yu, X. A novel time representation input based on deep learning for ECG classification. Biomed. Signal Process. Control 2023, 83, 104628.

[3] Wang, W.; Kumar, N.; Chen, J.; Gong, Z.; Kong, X.; Wei, W.; Gao, H. Realizing the Potential of Internet of Things for Smart Tourism with 5G and AI. IEEE Netw. 2020, 34, 295–301.

[4] Balasundaram, A.; Routray, S.; Prabu, A.; Krishnan, P.; Malla, P.P.; Maiti, M. Internet of things (IoT) based Smart Healthcare System for Efficient Diagnostics of Health Parameters of Patients in Emergency Care. IEEE Internet Things J. 2023.

[5] Sangaiah, A.K.; Arumugam, M.; Bian, G.B. An intelligent learning approach for improving ECG signal classification and arrhythmia analysis. Artif. Intell. Med. 2020, 103, 101788.

[6] Yang, S.; Lian, C.; Zeng, Z.; Xu, B.; Zang, J.; Zhang, Z. A Multi-View Multi-Scale Neural Network for Multi-Label ECG Classification. IEEE Trans. Emerg. Top. Comput. Intell. 2023.