# CardioForest: An Explainable Ensemble Learning Model for Automatic Wide QRS Complex Tachycardia Diagnosis from ECG

Vaskar Chakma[†a,d,e], Ju Xiaolin[†a], Cao Heling[b], Xue Feng[c], Ji Xiaodong[c], Pan Haiyan[c,*], Gao Zhan[a,*]

[a]*School of Artificial Intelligence and Computer Science, Nantong University, Nantong, Jiangsu, China*
[b]*College of Information Science and Engineering, Henan University of Technology, Zhengzhou, China*
[c]*Affiliated Hospital of Nantong University, Nantong, Jiangsu, China*
[d]*Intelligent Networking Lab (INL), Chung-Ang University, Seoul, Republic of Korea*
[e]*Department of Computer Science and Engineering, College of Software, Chung-Ang University, Seoul, Republic of Korea*

---

**Abstract**

**Objective:** This study aims to develop and evaluate an ensemble machine learning-based framework for the automatic detection of Wide QRS Complex Tachycardia (WCT) from ECG signals, emphasizing diagnostic accuracy and interpretability using Explainable AI.

**Method:** The proposed system integrates ensemble learning techniques, i.e., an optimized Random Forest known as CardioForest, and models like XGBoost and LightGBM. The models were trained and tested on ECG data from the publicly available MIMIC-IV dataset. The testing was carried out with the assistance of accuracy, balanced accuracy, precision, recall, F1 score, ROC-AUC, and error rates (RMSE, MAE) measures. In addition, SHAP (SHapley Additive exPlanations) was used to ascertain model explainability and clinical relevance.

**Results:** The *CardioForest* model performed best on all metrics, achieving a test accuracy of 94.95%, balanced accuracy of 88.31%, and high precision and recall metrics. SHAP analysis confirmed the model's ability to rank the most relevant ECG features—such as

*Corresponding authors at: School of Artificial Intelligence and Computer Science, Nantong University, Nantong 226019, Jiangsu, China & Affiliated Hospital of Nantong University.
E-mail addresses: gaozhan@ntu.edu.cn (G. Zhan), dr.phy@ntu.edu.cn (P. Haiyan).
[1†] These authors contributed equally.

QRS duration—in accordance with clinical intuitions, thereby fostering trust and usability in clinical practice.

**Conclusion:** The findings recognize *CardioForest* as an extremely dependable and interpretable WCT detection model. Being able to offer accurate predictions and transparency through explainability makes it a valuable tool to help cardiologists make timely and well-informed diagnoses, especially for high-stakes and emergency scenarios.
*Keywords:* Wide QRS Complex Tachycardia (WCT), ECG Analysis, Ensemble Machine Learning, Explainable AI, Artificial Intelligence in Healthcare

---

## 1. Introduction

Wide QRS Complex Tachycardia (WCT) is a severe and potentially lethal cardiac condition characterized by an exceedingly rapid heartbeat in combination with a widened QRS complex on the electrocardiogram (ECG) [1, 2, 3]. Normally, the QRS complex—a short, spiky waveform—registers the process of ventricular depolarization, whereby the ventricles of the heart contract and effectively pump blood [4, 5]. A regular narrow QRS complex indicates typical electrical conduction through the heart's normal pathways [6, 7]. However, if the QRS complex is wide, then this is an indication of a disruption in electrical propagation [8], typically as a result of underlying structural disease, electrolyte imbalance, or an inherited disorder. Untreated, WCT can significantly weaken the heart's function to circulate blood effectively, causing symptoms that range from palpitations, dizziness, and chest pain to, in extreme cases, sudden cardiac arrest [9, 10]. As such, the early and correct diagnosis of WCT is not only critical—it is a matter of life and death. Diagnosis of WCT has traditionally depended to a large extent on manual ECG interpretation by experienced cardiologists. While still the gold standard, this process is time-consuming, labor-intensive, and subject to considerable variability [11, 12]. Individual cardiologists may disagree in borderline or uncertain cases, postponing diagnosis and treatment [13, 14]. In high-pressure clinical environments where minutes matter, delays can be detrimental. As healthcare systems globally face rising

demands, the demand for faster, more accurate diagnostic support that augments, rather than replaces, clinical judgment is pressing. In the past few years, Artificial Intelligence (AI)-driven models have been demonstrated to achieve stellar performance in ECG interpretation with accuracy and speed [15]. Among them, deep learning methods—particularly Convolutional Neural Networks (CNNs)—have worked incredibly well in identifying complex patterns within ECG signals that are not easily visible to the naked eye [16, 17, 18]. However, for all the high-accuracy deep learning models claim, they tend to behave like "black boxes" with little description of decision-making. This absence of transparency has been a significant barrier to clinical adoption because cardiologists and clinicians need not only accuracy but also transparency to trust AI recommendations. For AI to be successfully integrated into clinical practice, especially in life-critical conditions like WCT, interpretability is equally as important as accuracy [19]. Cardiologists must understand the rationale for AI predictions—seeing not just the output, but the supporting evidence, e.g., what ECG features led to a specific classification. Without this transparency, clinicians will remain unconvinced and reluctant to trust AI for decision-making, especially when a patient's life hangs in the balance. Given these challenges, this research introduces a novel solution: **CardioForest**, an interpretable ensemble-based AI model for WCT detection. Based primarily on Random Forest architecture—augmented with techniques such as XGBoost, LightGBM, and Gradient Boosting—CardioForest leverages the strengths of ensemble machine learning to achieve both high diagnostic precision and clear interpretability [20]. Unlike deep neural networks, Random Forest models are natively explainable through feature importance rankings and decision-tree visualization. This dual advantage ensures clinicians can rely on and interpret the model outputs. Our experiment design is focused not only on the evaluation of diagnostic accuracy but also on increasing clinical trust by employing explainable AI (XAI) methods, for example, SHAP (SHapley Additive exPlanations) values and feature attribution analysis [21]. CardioForest bridges the crucial gap between AI's computational power and the clinician's need for interpretability by highlighting which features most significantly

influenced the model's decisions. Preliminary results suggest that CardioForest outperforms traditional manual approaches and competes favorably with state-of-the-art deep learning models while offering superior transparency, an essential quality for clinical adoption. For all these developments, we acknowledge that challenges remain. Future research should also explore continuous learning frameworks, where AI models learn incrementally from new data, thereby improving their diagnostic acumen over time without compromising explainability. This work presents *CardioForest* as a pioneering WCT detection solution—diagnostic performance coupled with interpretability, a union necessary for real-world clinical practice. By providing cardiologists with speedy, interpretable, and reliable AI support, we hope to enhance cardiac diagnosis, reduce diagnostic latency, and ultimately save more lives. Looking ahead, we envision the expansion of explainable ensemble model applications beyond WCT toward general arrhythmia detection with the inclusion of real-time ECG monitoring for preemptive cardiac health management.

## 2. Related Works

Accurate and timely prediction of Wide Complex Tachycardia (WCT) remains a major focus in cardiovascular research, driven by the need to distinguish between ventricular tachycardia (VT) and supraventricular tachycardia (SVT) with aberrant conduction. Machine learning (ML) and deep learning (DL) techniques have gained prominence in this domain, offering new avenues for improved diagnostic performance compared to traditional criteria-based methods.

Li et al. [22] proposed a Gradient Boosting Machine (GBM) model for differentiating VT from SVT using surface ECG features. Their approach leveraged a rich set of ECG-derived parameters, leading to an impressive classification performance with an overall accuracy of 91.2%, sensitivity of 89.5%, specificity of 92.8%, and an area under the ROC curve (AUC) of 0.94. This study highlighted the importance of carefully selected ECG features in enhancing machine learning model performance. Building on the trend of AI-driven diagnosis, Chow et al. [23] developed a specialized AI model to interpret WCT directly from ECGs. Their

system, designed with clinical applicability in mind, demonstrated an overall accuracy of 93%, with sensitivity and specificity exceeding 91%. This work showcased the potential of deep learning models in outperforming traditional rule-based algorithms for arrhythmia classification, particularly for ambiguous WCT cases.

Focusing on high-risk populations, Bhattacharya et al. [24] introduced the HCM-VAr-Risk model, which applies machine learning techniques to predict ventricular arrhythmias in patients with hypertrophic cardiomyopathy (HCM). Their model achieved a C-index of 0.83, reflecting strong discriminative ability. The study underscored the utility of ML for risk stratification in structurally abnormal hearts, offering a more individualized approach to arrhythmia prediction. Hong et al. [25] provided a broader perspective by conducting a systematic review of deep learning applications for ECG analysis, including arrhythmia detection and classification tasks. The review covered a range of architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models, illustrating the high accuracy and generalizability of DL models when trained on large, diverse ECG datasets. Their findings support the growing consensus that deep learning can significantly enhance the detection of complex arrhythmias, including WCT.

Addressing diagnostic challenges from a different angle, May et al. [26] introduced the QRS Polarity Shift (QRS-PS) method, which focuses on changes in QRS polarity between baseline ECGs and WCT episodes. By simplifying the interpretation to polarity shifts, their algorithm achieved AUC values ranging from 0.90 to 0.93. This technique provides a pragmatic and explainable tool that can be readily integrated into clinical workflows, assisting clinicians in making rapid and accurate diagnoses. Machine learning classification models have also shown remarkable potential in SVT detection. Howladar and Sahoo [27] developed a decision-tree-based model specifically tailored for SVT identification. Their model attained a striking 97% accuracy, demonstrating the effectiveness of even relatively simple ML algorithms when paired with relevant feature selection. Deep learning models have further pushed the boundaries of arrhythmia prediction. Rajpurkar et al. [28] designed a CNN-

based model, trained on a large annotated ECG dataset, that achieved recall and precision rates exceeding those of board-certified cardiologists. Their work set a new benchmark for DL-based arrhythmia detection and provided strong evidence for adopting AI-assisted ECG interpretation tools in clinical practice.

Table 1: Summary of Diverse Studies on WCT Prediction

| **Study** | **Model/Method** | **Key Features** | **Performance Metrics** | **References** |
|---|---|---|---|---|
| Li et al. (2024) | Gradient Boosting Machine (GBM) | Differentiates VT from SVT using surface ECG features | Accuracy: 91.2%, Sensitivity: 89.5%, Specificity: 92.8%, AUC: 0.94 | [22] |
| Chow et al. (2024) | Artificial Intelligence (AI) Model | AI algorithm interpreting WCT ECGs | Accuracy: 93%, Sensitivity: 91.9%, Specificity: 93.4% | [23] |
| Bhattacharya et al. (2021) | HCM-VAr-Risk Model | ML model for ventricular arrhythmias in HCM patients | Sensitivity: 73%, Specificity: 76%, C-index: 0.83 | [24] |
| Hong et al. (2020) | Deep Learning Review | Systematic review of DL methods for ECG data | Various models achieving high accuracy and AUC | [25] |
| May et al. (2024) | QRS Polarity Shift (QRS-PS) | Algorithm based on QRS polarity shifts between WCT and baseline ECGs | AUC: 0.90–0.93 | [26] |
| Howladar & Sahoo (2021) | Classification Model | ML-based SVT detection & decision-tree-based model for SVT classification | Accuracy: 97% | [27] |
| Rajpurkar et al. (2017) | CNN-based Model | Deep learning model for arrhythmia detection | Exceeds average cardiologist performance in recall and precision | [28] |
| Frausto-Avila et al. (2024) | Compact Neural Network | ANN with feature enhancement for arrhythmia classification | Accuracy: 97.36% | [29] |

In addition, Frausto-Avila et al. [29] presented a compact neural network architecture enhanced with advanced feature engineering techniques. Their model achieved an accuracy of 97.36% in arrhythmia classification tasks, suggesting that lightweight models can maintain high predictive performance while offering advantages in computational efficiency, making

them suitable for deployment in real-time or resource-constrained environments. The reviewed studies demonstrate that both machine learning and deep learning approaches have significantly advanced WCT prediction and arrhythmia classification. The diversity of methods—from feature-driven models like GBM and decision trees to sophisticated deep learning architectures like CNNs—reflects the rich potential of AI technologies to improve clinical outcomes. A detailed comparison of these related works, including their methodologies, key innovations, and achieved performance metrics, is presented in Table 1.

## 3. Diagnostic Data Resources

This study utilizes the MIMIC-IV-ECG [30] Module (statistics summary of the dataset has been shown in Table 2), a comprehensive database of diagnostic electrocardiogram (ECG) waveforms [31] integrated with the broader MIMIC-IV Clinical Database. The dataset contains approximately 800,000 ten-second-long 12-lead ECG recordings sampled at 500 Hz, collected from nearly 160,000 unique patients. Each electrocardiogram (ECG) record is stored in the standard WaveForm DataBase (WFDB)[2] format, which includes a header file (.hea) and a binary data file (.dat). The records are organized in a structured directory hierarchy based on the subject identifier, allowing for efficient data retrieval. For example, a subject with ID 10001725 would be stored under the path files/p1000/p10001725/, with each diagnostic study within a subdirectory labelled by a randomly generated study ID. Approximately 55% of the ECGs in the dataset overlap with a hospital admission and 25% with an emergency department visit, while the remaining records were collected outside traditional inpatient or emergency settings. This diversity in acquisition context allows for a wide range of use cases, from acute event analysis to routine monitoring assessments. However, it is important to note that the ECG timestamps are derived from the internal clock of the acquisition device and are not synchronized with the hospital's clinical information systems.

[2]https://physionet.org/lightwave/

Table 2: Statistics of the MIMIC-IV real-world datasets used in this paper

| **Metric** | **Value** |
|---|---:|
| Total Records | 800,035 |
| Unique Subjects | 161,352 |
| Unique Studies | 800,035 |
| Unique Carts | 156 |
| Average RR Interval (ms) | 865.60 |
| Average QRS Duration (ms) | 108.51 |
| Average P Onset (ms) | 4,702.88 |
| Average P End (ms) | 8,745.08 |
| Average QRS Onset (ms) | 283.42 |
| Average QRS End (ms) | 391.66 |
| Average T End (ms) | 688.65 |
| Average P Axis (degrees) | 4,973.35 |
| Average QRS Axis (degrees) | 107.37 |
| Average T Axis (degrees) | 192.55 |

Consequently, temporal alignment between the ECGs and clinical events in the MIMIC-IV database may require additional validation. Each ECG waveform is accompanied by machine-generated summary measurements [30] stored in the machine_measurements.csv file. These include standard parameters such as RR interval, QRS onset and end, and filter settings, along with textual machine-generated interpretation notes across columns report_0 to report_17. The accompanying data dictionary in machine_measurements-_data_dictionary.csv describes the technical and clinical meaning of each column. Each record includes a subject_id, study_id, and ecg_time, enabling direct linkage to clinical data in the MIMIC-IV hospital and emergency department modules. Cardiologist interpretations are also available for over 600,000 ECG studies. These free-text reports are stored in the MIMIC-IV-Note module and are linked to the ECG waveforms via the waveform_note_links.csv file. Each entry in this linkage file includes the subject ID, study ID, waveform path, and a note_id that can be used to retrieve the corresponding cardiologist report. A sequential integer (note_seq) is also provided to determine the order of ECG collection for individual patients. This linkage enables researchers to perform comparative analyses between machine-generated and clinician-interpreted findings. To support large-

scale analysis, key metadata from record_list.csv, machine_measurements.csv, and waveform_note_links.csv has been made available through Google BigQuery. This facilitates efficient querying and integration with other clinical tables in the MIMIC-IV ecosystem. As a practical illustration, using BigQuery, a researcher can identify a patient's hospital admissions and correlate them with the timing of their ECGs, determine whether a given ECG occurred during a hospital stay, and check for the presence of associated cardiologist notes [32], [33]. For waveform visualization and signal processing, the dataset supports standard PhysioNet WFDB toolkits in Python, MATLAB, and C. Researchers can read and visualize ECG waveforms using the wfdb Python package [34]. For instance, using wfdb.rdrecord() and wfdb.plot_wfdb(), one can extract and display the raw signal for any given ECG study. This compatibility makes the dataset highly accessible for both signal processing and clinical informatics researchers. Despite its richness, the dataset has some limitations. Notably, the ECG device timestamps may be inaccurate due to a lack of clock synchronization. Additionally, some ECGs were recorded outside the hospital or emergency department, limiting direct temporal correlation with clinical events [35]. Nonetheless, MIMIC-IV-ECG is invaluable for studying cardiac health, machine learning applications in ECG interpretation, and cross-modal linkage with comprehensive clinical records.

## 4. Data Preparation and Processing Pipeline

### *4.1. Data Cleaning and Preprocessing Techniques*

Duplicate entries were identified using `subject_id` and `study_id`, ensuring each ECG was uniquely represented. Pandas' duplicated() function [36] detected redundant records, which were subsequently removed. Post-cleaning verification confirmed the dataset (Fig. 1) retained its integrity, with zero duplicate records in the final 800,035 entries. Biologically implausible values (e.g., negative RR intervals) were corrected using interpolation [37], while extreme outliers were adjusted or removed. Visualizations like boxplots and histograms validated the corrections, showing normalized distributions for key features such as RR intervals and QRS durations. Timestamps (ecg_time_x and ecg_time_y) were converted to

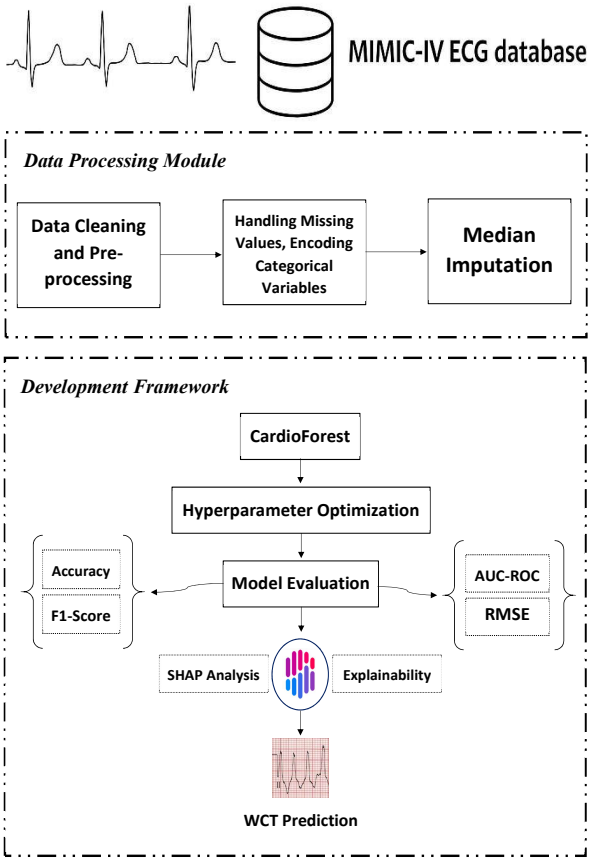


Figure 1: An overview of the WCT prediction system using the MIMIC-IV ECG database, featuring pre-processing, ensemble machine learning models, cross validation, and final prediction.

Figure 2: Temporal dynamics of ECG features showing rolling statistics (mean, standard deviation, and skewness) for RR interval and QRS duration across the time sequence.

a uniform format using Python's datetime module, ensuring consistency for time-series analysis. This step addressed discrepancies arising from unsynchronized ECG machine clocks. Categorical variables (e.g., wct_label) were encoded numerically using label encoding, and floating-point precision errors were truncated. This ensured compatibility with machine learning algorithms and improved computational efficiency.

### *4.2. Data Merging, Feature Selection, and Extraction*

Data merging, feature selection, and extraction represent a critical phase in transforming the cleaned MIMIC-IV-ECG dataset [30] into a format optimized for machine learning analysis. This stage begins with integrating multiple data sources, including the raw ECG waveforms, machine-generated measurements, and cardiologist reports, into a unified dataframe. The merging process leverages key identifiers such as `subject_id` and `study_id` to ensure accurate alignment of records across different tables. Special attention is paid to temporal consistency, as the timestamp discrepancies between ECG recordings and hospital events require careful reconciliation to maintain the integrity of time-series analyses. Feature selection constitutes the next crucial step, where we systematically evaluate the clinical relevance and statistical properties of each potential predictor [38]. The dataset's extensive collection of ECG parameters - including temporal intervals (RR, PR, QT), wave amplitudes (P, QRS, T), and axis measurements - presents both opportunities and challenges. We employ correlation analysis (Fig. 3) to identify redundant features, using heatmap visualizations to detect strong linear relationships between variables. For instance, the analysis revealed

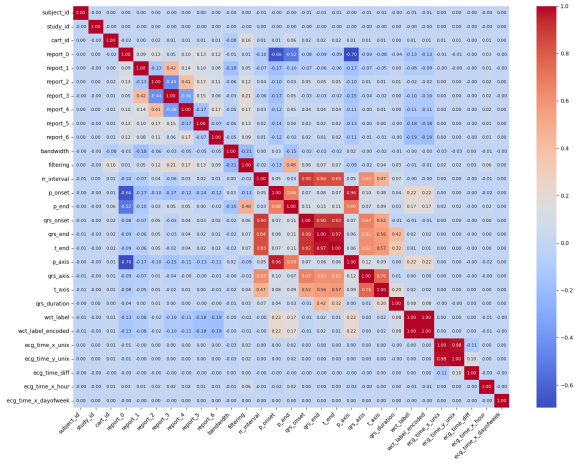

Figure 3: Initialization parameters and preprocessing metadata for ECG signal analysis, showing default values (0.00-0.01) for subject identifiers, report fields, filtering parameters, and waveform annotation markers (P-onset, QRS complex). The WCT (Wide Complex Tachycardia) label indicators suggest the beginning of arrhythmia classification preprocessing.

a high correlation between specific lead-specific measurements, prompting the removal of redundant leads to reduce dimensionality while preserving diagnostic information. Distribution plots for key features like RR interval and QRS duration provide insights into their statistical properties, highlighting skewness that may require transformation (Fig. 2). Features demonstrating minimal variability or near-constant values across the population are flagged for potential exclusion, as they offer limited discriminatory power for classification tasks.

The feature extraction phase employs advanced techniques to derive more informative representations of the raw data. Principal Component Analysis (PCA) proves particularly valuable for condensing the multidimensional ECG features into a smaller set of orthogonal components that capture the majority of variance in the data [39]. Prior to PCA application (Fig. 4), we standardize all features to zero mean and unit variance to prevent variables with larger scales from dominating the component calculation. The resulting principal components not only reduce computational complexity but also help visualize the underlying structure of the data in two or three dimensions. Boxplot analyses complement this approach by comparing feature distributions across different clinical conditions, such as

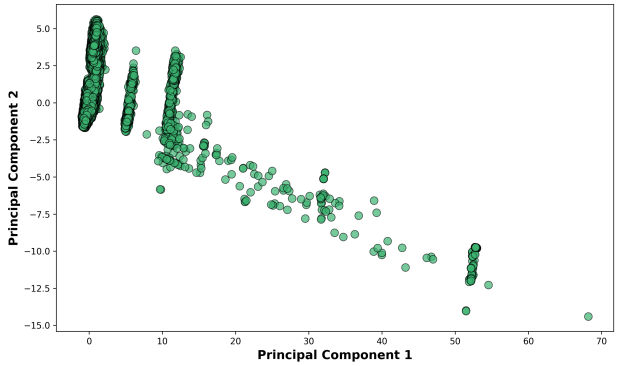


Figure 4: Relationship between Principal Component 1 (x-axis) and Principal Component 2 (y-axis). The axis scaling (0-70) indicates the relative variance explained by each component in this dimensionality reduction visualization.

normal sinus rhythm versus wide complex tachycardia [40]. These visualizations help identify features that show significant separation between classes, making them prime candidates for inclusion in predictive models. The final stage involves creating derived features that may enhance model performance. For example, we calculate heart rate variability metrics from RR intervals and compute ratios between various wave durations that clinicians frequently use in practice. The feature engineering process shown in Fig. 5 remains grounded in clinical knowledge to ensure the biological plausibility of all derived measures. Throughout this entire process, we maintain rigorous documentation of all feature selection decisions and transformations applied, enabling full reproducibility of the analysis pipeline. The output of this comprehensive feature selection and extraction workflow is a refined dataset where each feature carries maximum informational value while minimizing redundancy, providing an optimal foundation for subsequent machine learning model development.

### *4.3. Handling Missing Values and Categorical Variables*

In this section, two crucial data preprocessing tasks have been focused on and shown in Fig. 6: handling missing values and encoding categorical variables, both of which are vital steps to ensure that the dataset is suitable for machine learning models. Missing values are a common issue in many real-world datasets. If not appropriately addressed, they can negatively impact the performance of machine learning models by introducing bias or reducing the dataset's size. We chose median imputation as the strategy to handle missing

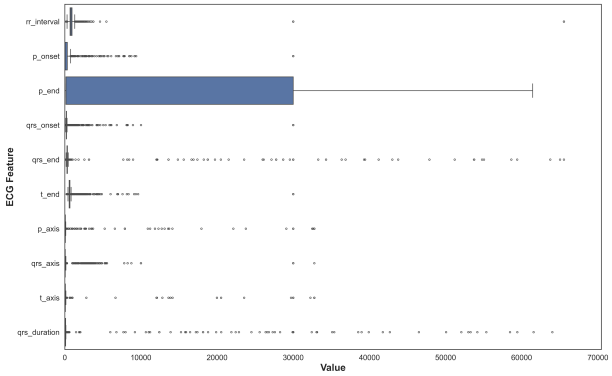


Figure 5: This boxplot illustrates the statistical distribution of QRS complex durations across all ECG recordings, showing median values, interquartile ranges, and outliers. The visualization helped validate measurement quality and identify extreme values requiring clinical review before feature selection.

values in the dataset. Median imputation involves replacing missing values with the median value of a column [41]. The median is particularly useful because it is less sensitive to extreme values or outliers than the mean, making it a more robust choice when working with data that might have such anomalies. For example, in the ECG dataset, some numerical columns, such as the ‘rr_interval‘ or ‘qrs_duration‘, may contain missing values for various reasons, such as data collection issues or measurement errors. Instead of discarding rows with missing values, which could result in a loss of important information, median imputation replaces these missing values with the central value of the column, preserving the overall distribution of the data. This approach helps maintain the integrity of the dataset, ensuring that the analysis and modelling processes are not disrupted by missing entries. We used Scikit-learn’s ‘SimpleImputer‘ with the ’median’ strategy to perform this imputation across all relevant numerical columns in the dataset [42].

Once the missing values were handled, the next step was to address the categorical variables in the dataset. Many machine learning algorithms require numerical inputs, so categorical data, often represented as text labels, needs to be converted into a numerical format. In this case, the dataset contains several categorical columns, such as ‘report_0‘, ‘report_1‘, ‘report_2‘, ‘report_3‘, ‘report_4‘, ‘report_5‘, ‘report_6‘, ‘filtering‘, and ‘wct_label‘, all of which contain textual labels that represent different categories or classifications of the data.

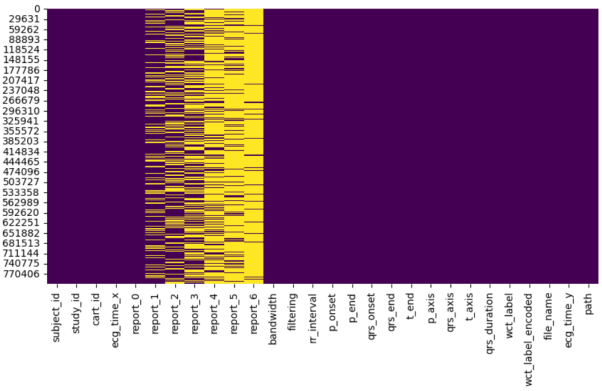


Figure 6: It displays the organized dataframe containing 800,035 ECG records and 28 clinically relevant features after preprocessing and highlights preserved temporal measurements, electrical axis values, and metadata for machine learning applications.

Label encoding was chosen to convert these categorical text values into numerical labels. Label encoding assigns a unique integer to each category within a given column. For instance, if the 'report_0' column contains the values 'Normal', 'Abnormal', and 'Pending', the Label Encoder would transform these labels into numerical values like 0, 1, and 2. This transformation makes the data usable by machine learning algorithms, which can only process numerical inputs. Scikit-learn's 'LabelEncoder' has been used for this task, applying it to each categorical column in the dataset [43]. Both preprocessing steps—median imputation for handling missing values and label encoding for categorical variables—ensure the dataset is ready for analysis and modelling. Properly handling missing data avoids biases, maintains the integrity of the dataset, and prevents the loss of valuable information.

## 5. Methods, Experiments and Results

### *5.1. Model Selection and Justification*

The analysis of electrocardiogram (ECG) [44] signals demands robust methodologies capable of navigating noise, patient-specific variability, and subtle morphological changes. These challenges are particularly acute when diagnosing life-threatening arrhythmias like Wide Complex Tachycardia (WCT) [45]. Although deep learning methods, especially Convolutional Neural Networks (CNNs), have demonstrated significant accuracy, their opaque decision-making and high computational requirements limit their deployment in real-time,

resource-constrained clinical settings [46]. To address these concerns, this study introduces a specialized Random Forest model named **CardioForest**, tailored for predicting Wide Complex Tachycardia (WCT) events. CardioForest is benchmarked against other ensemble methods, including Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), balancing performance, interpretability, and computational efficiency—crucial attributes for clinical ECG analysis [47].

*5.1.1. CardioForest: Model Formulation*

CardioForest is built upon the Random Forest (RF) framework, enhanced with hyperparameters tuned specifically for ECG feature characteristics and arrhythmic prediction. Random Forest aggregates multiple decision trees trained on bootstrap samples to minimize variance and prevent overfitting [48]. Given a dataset $D$, each tree $t$ is trained on a subset $D_t$ sampled with replacement:

$$D_t = (x_i, y_i) \mid (x_i, y_i) \sim D,\ |D_t| = |D| \tag{1}$$

At each split node, CardioForest selects a random subset $m$ of features (where $m \ll$ total features) to introduce tree diversity. The prediction for an input $x$ is made via majority voting across all $T$ trees:

$$\hat{y} = \text{mode}\left(f_t(x)_{t=1}^{T}\right) \tag{2}$$

where $f_t(x)$ is the prediction of the $t$-th decision tree. Each split in a tree aims to minimize the Gini impurity:

$$G(N) = 1 - \sum_{k=1}^{K} p_k^2 \tag{3}$$

where $p_k$ is the proportion of samples belonging to class $k$ at node $N$. Here, $K = 2$ for binary classification (WCT vs. non-WCT). CardioForest also incorporates cost-complexity pruning with parameter $\alpha$ to penalize overly complex trees, improving generalization:

$$R_\alpha(T) = R(T) + \alpha|\text{leaves}(T)| \tag{4}$$

where $R(T)$ is the empirical risk (e.g., misclassification rate) of tree $T$.

*5.1.2. Comparison Models: GBM, XGBoost, LightGBM*

In addition to CardioForest, we compared three gradient-boosting-based models:

*Gradient Boosting Machine (GBM).* GBM constructs an additive model:

$$F_M(x) = \sum_{m=1}^{M} \gamma_m h_m(x) \tag{5}$$

where $h_m(x)$ represents the weak learner at iteration $m$, and $\gamma_m$ is its associated weight. Each $h_m$ approximates the negative gradient of the loss function $L$:

$$h_m(x) \approx -\nabla_{F_{m-1}} L(y, F_{m-1}(x)) \tag{6}$$

The model is updated iteratively using a learning rate $\vartheta$:

$$F_m(x) = F_{m-1}(x) + \vartheta \gamma_m h_m(x) \tag{7}$$

*Extreme Gradient Boosting (XGBoost).* XGBoost refines GBM by incorporating regularization into the objective function:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} L(y_i, \hat{y}i) + \sum k = 1^K \Omega(f_k) \tag{8}$$

where $\Omega(f_k) = \alpha|\omega|_1 + \frac{1}{2}\lambda|\omega|2^2$ penalizes model complexity through $L_1$ and $L_2$ norms. Optimization is performed using a second-order Taylor approximation:

$$\mathcal{L}^{(t)} \approx \sum i = 1^n \left[ g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{9}$$

where $g_i$ and $h_i$ are the first and second derivatives of the loss function with respect to $\hat{y}^{(t-1)}$.

*Light Gradient Boosting Machine (LightGBM).* LightGBM accelerates XGBoost's design using two key strategies:

- **Histogram-based Feature Binning**: Discretizes continuous feature values to reduce memory and computation.

Table 3: Hyperparameter Specifications for Machine Learning Models

| **Models** | **Parameters** | **Optimal Values** |
|---|---|---|
| CardioForest [Proposed] | n_estimators | 1000 |
| | max_depth | 20 |
| | min_samples_split | 5 |
| | min_samples_leaf | 2 |
| | max_features | 0.6 |
| | class_weight | balanced |
| | random_state | 42 |
| | n_jobs | -1 |
| | bootstrap | True |
| | oob_score | True |
| | ccp_alpha | 0.01 |
| XGBoost | n_estimators | 10 |
| | max_depth | 2 |
| | learning_rate | 0.5 |
| | subsample | 0.4 |
| | colsample_bytree | 0.2 |
| | random_state | 42 |
| | gamma | 3 |
| | reg_alpha | 2 |
| | reg_lambda | 2 |
| LightGBM | n_estimators | 5 |
| | max_depth | 1 |
| | learning_rate | 0.8 |
| | subsample | 0.3 |
| | colsample_bytree | 0.1 |
| | random_state | 42 |
| | min_child_samples | 50 |
| | reg_alpha | 3 |
| | reg_lambda | 3 |
| Gradient Boosting | n_estimators | 3 |
| | max_depth | 2 |
| | learning_rate | 0.4 |
| | subsample | 0.5 |
| | min_samples_split | 9 |
| | min_samples_leaf | 10 |
| | max_features | 0.3 |
| | validation_fraction | 0.1 |
| | n_iter_no_change | 2 |

- **Gradient-based One-Sided Sampling (GOSS)**: Retains instances with large gradients and randomly samples small-gradient instances to speed up the training without significantly losing accuracy.

*5.2. Hyperparameter Tuning for Experimental Setup*

To ensure optimal model generalization while preserving clinical relevance, a systematic hyperparameter tuning process [49] was employed across all classifiers (Table 3). Each model underwent a comprehensive grid search procedure, constrained within physiologically plausible and empirically supported parameter ranges.

For the proposed **CardioForest** model, key parameters were tuned to balance complexity and stability: 1,000 decision trees (`n_estimators=1000`) with a maximum depth of 20 (`max_depth=20`) were used to capture meaningful ECG patterns without overfitting. Splits required at least 5 samples (`min_samples_split=5`), and each leaf node required at least 2 samples (`min_samples_leaf=2`). A feature subset of 60% (`max_features=0.6`) was randomly selected at each split to promote tree diversity. Balanced class weights were used to address potential label imbalance, and out-of-bag (OOB) evaluation (`oob_score=True`) enhanced model validation. A pruning penalty (`ccp_alpha=0.01`) was applied to simplify the final trees.

**XGBoost**, a highly regularized shallow structure, was adopted: 10 estimators (`n_estimators=10`) with a maximum depth of 2 (`max_depth=2`) ensured rapid and cautious learning. A relatively high learning rate (`learning_rate=0.5`) expedited convergence, while strong regularization parameters ($\gamma = 3$, `reg_alpha=2`, `reg_lambda=2`) minimized overfitting. Feature and instance subsampling ratios (`subsample=0.4`, `colsample_bytree=0.2`) further contributed to model robustness. **LightGBM** was configured with extreme minimalism: only 5 estimators (`n_estimators=5`) with a single-level depth (`max_depth=1`), using a high learning rate (`learning_rate=0.8`) for rapid adaptation. Regularization was reinforced (`reg_alpha=3`, `reg_lambda=3`), with a minimum of 50 samples per leaf (`min_child_samples=50`) to maintain generalization. Subsampling strategies (`subsample=0.3`, `colsample_bytree=0.1`) con-

Table 4: Highlights the aggregate performance across all simulations, demonstrating CardioForest's superior mean values and tighter confidence intervals compared to alternatives

| Model | Fold | Accuracy | Balanced Accuracy | Precision | Recall | F1 | ROC_AUC | RMSE | MAE |
|---|---|---|---|---|---|---|---|---|---|
| CardioForest | 1 | 0.9484 | 0.8831 | 0.9511 | 0.7758 | 0.8546 | 0.8844 | 0.2513 | 0.1909 |
| | 2 | 0.9520 | 0.8851 | 0.9509 | 0.7793 | 0.8566 | 0.8913 | 0.2468 | 0.1893 |
| | 3 | 0.9472 | 0.8792 | 0.9350 | 0.7706 | 0.8449 | 0.8722 | 0.2544 | 0.1926 |
| | 4 | 0.9558 | 0.8906 | 0.9493 | 0.7902 | 0.8625 | 0.8893 | 0.2426 | 0.1885 |
| | 5 | 0.9478 | 0.8790 | 0.9468 | 0.7682 | 0.8482 | 0.8802 | 0.2522 | 0.1914 |
| | 6 | 0.9516 | 0.8894 | 0.9399 | 0.7903 | 0.8586 | 0.8899 | 0.2490 | 0.1903 |
| | 7 | 0.9568 | 0.9017 | 0.9545 | 0.8126 | 0.8778 | 0.9007 | 0.2415 | 0.1871 |
| | 8 | 0.9450 | 0.8775 | 0.9369 | 0.7673 | 0.8437 | 0.8728 | 0.2567 | 0.1928 |
| | 9 | 0.9544 | 0.8940 | 0.9474 | 0.7981 | 0.8664 | 0.9009 | 0.2447 | 0.1887 |
| | 10 | 0.9532 | 0.8925 | 0.9481 | 0.7951 | 0.8649 | 0.8843 | 0.2462 | 0.1890 |
| XGBoost | 1 | 0.8668 | 0.6847 | 0.8510 | 0.3859 | 0.5310 | 0.8429 | 0.3124 | 0.2069 |
| | 2 | 0.8910 | 0.7215 | 0.9085 | 0.4533 | 0.6048 | 0.8576 | 0.2941 | 0.1942 |
| | 3 | 0.8838 | 0.7105 | 0.8843 | 0.4341 | 0.5823 | 0.8389 | 0.3067 | 0.2054 |
| | 4 | 0.8966 | 0.7340 | 0.8689 | 0.4835 | 0.6212 | 0.8537 | 0.2929 | 0.1988 |
| | 5 | 0.8814 | 0.7102 | 0.8803 | 0.4341 | 0.5815 | 0.8416 | 0.3081 | 0.2029 |
| | 6 | 0.8872 | 0.7233 | 0.8704 | 0.4624 | 0.6039 | 0.8552 | 0.3009 | 0.2009 |
| | 7 | 0.8862 | 0.7189 | 0.9106 | 0.4482 | 0.6007 | 0.8691 | 0.2960 | 0.2000 |
| | 8 | 0.8720 | 0.6966 | 0.8501 | 0.4105 | 0.5537 | 0.8389 | 0.3124 | 0.2078 |
| | 9 | 0.8916 | 0.7257 | 0.9068 | 0.4622 | 0.6123 | 0.8661 | 0.2950 | 0.1979 |
| | 10 | 0.8812 | 0.7096 | 0.8702 | 0.4342 | 0.5793 | 0.8473 | 0.3040 | 0.2032 |
| LightGBM | 1 | 0.8372 | 0.6377 | 0.6840 | 0.3101 | 0.4268 | 0.7717 | 0.3515 | 0.2461 |
| | 2 | 0.8504 | 0.6503 | 0.6946 | 0.3337 | 0.4508 | 0.7801 | 0.3399 | 0.2367 |
| | 3 | 0.8334 | 0.6250 | 0.6121 | 0.2926 | 0.3959 | 0.7660 | 0.3530 | 0.2459 |
| | 4 | 0.8522 | 0.6415 | 0.6651 | 0.3170 | 0.4293 | 0.7904 | 0.3355 | 0.2345 |
| | 5 | 0.8390 | 0.6348 | 0.6651 | 0.3056 | 0.4188 | 0.7655 | 0.3501 | 0.2442 |
| | 6 | 0.8464 | 0.6481 | 0.6776 | 0.3323 | 0.4459 | 0.7891 | 0.3418 | 0.2387 |
| | 7 | 0.8420 | 0.6432 | 0.6837 | 0.3215 | 0.4373 | 0.7881 | 0.3449 | 0.2424 |
| | 8 | 0.8390 | 0.6451 | 0.6709 | 0.3289 | 0.4414 | 0.7747 | 0.3497 | 0.2451 |
| | 9 | 0.8524 | 0.6574 | 0.7061 | 0.3477 | 0.4660 | 0.8013 | 0.3355 | 0.2348 |
| | 10 | 0.8408 | 0.6423 | 0.6573 | 0.3238 | 0.4339 | 0.7732 | 0.3473 | 0.2441 |
| GradientBoosting | 1 | 0.9300 | 0.8600 | 0.8782 | 0.7451 | 0.8062 | 0.8763 | 0.2757 | 0.1965 |
| | 2 | 0.9520 | 0.8851 | 0.9509 | 0.7793 | 0.8566 | 0.8882 | 0.2379 | 0.1631 |
| | 3 | 0.8802 | 0.6893 | 0.9349 | 0.3848 | 0.5452 | 0.8430 | 0.2952 | 0.2119 |
| | 4 | 0.9486 | 0.8701 | 0.9467 | 0.7491 | 0.8364 | 0.8934 | 0.2423 | 0.1657 |
| | 5 | 0.8486 | 0.6193 | 0.8404 | 0.2497 | 0.3851 | 0.7931 | 0.3436 | 0.2549 |
| | 6 | 0.9514 | 0.8888 | 0.9398 | 0.7892 | 0.8580 | 0.8890 | 0.2484 | 0.1767 |
| | 7 | 0.9568 | 0.9017 | 0.9545 | 0.8126 | 0.8778 | 0.8996 | 0.2250 | 0.1532 |
| | 8 | 0.9450 | 0.8775 | 0.9369 | 0.7673 | 0.8437 | 0.8770 | 0.2346 | 0.1434 |
| | 9 | 0.8862 | 0.7036 | 0.9364 | 0.4136 | 0.5738 | 0.8663 | 0.3054 | 0.2234 |
| | 10 | 0.9532 | 0.8925 | 0.9481 | 0.7951 | 0.8649 | 0.8887 | 0.2280 | 0.1514 |

trolled variance during training. **Gradient Boosting**, a compact architecture was utilized: only 3 trees (`n_estimators=3`) with a maximum depth of 2 (`max_depth=2`). The learning rate was moderately high (`learning_rate=0.4`) to favor quick learning. A minimum of 9 samples was required to split internal nodes (`min_samples_split=9`), and at least 10 samples were mandated per leaf (`min_samples_leaf=10`), preserving robustness. Only 30% of features (`max_features=0.3`) were considered at each split. Subsampling (`subsample=0.5`) and early stopping after 2 rounds of no improvement (`n_iter_no_change=2`) were employed to further stabilize learning.

All hyperparameter tuning outlined in Table 3 was performed using stratified cross-validation, ensuring robust performance estimation under varying data partitions. Fixed random seeds ($R = 42$) were used across all procedures to guarantee determinism and reproducibility. Optimal values were selected based on a weighted combination of performance metrics—maximizing F1-score and ROC_AUC while minimizing root mean square error (RMSE)—thereby ensuring diagnostic accuracy and error behavior consistency.

*5.3. Performance Metrics Overview*

The performance evaluation (Table 4) of various models through 10 CV [50] revealed that all the classifiers performed well, four machine learning models—**CardioForest**, **XGBoost**, **LightGBM**, and **Gradient Boosting**—were compared, but CardioForest stood out as the most reliable and consistent for WCT detection. Several metrics were recorded: Accuracy, Balanced Accuracy, Precision, Recall, F1-Score, ROC_AUC, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). CardioForest was superior to all the other models in almost all the folds throughout, with an average accuracy of almost 95%, high balanced accuracy ($\sim 0.88$), good precision (greater than 0.93), and good recall ($\sim 0.77$–$0.81$). Its ROC_AUC scores were highly significant, indicating excellent classification ability, and its RMSE and MAE scores remained the lowest for all the models, reflecting high overall stability and prediction accuracy. Conversely, XGBoost performed fairly well but with a clear deterioration compared to CardioForest. Average accuracy ranged from 88%–89%, whereas

Table 5: Model Error Metrics Comparison

| **Model** | **RMSE** | **MAE** |
|---|---|---|
| CardioForest | 0.2532 | 0.1944 |
| XGBoost | 0.3003 | 0.2008 |
| LightGBM | 0.3471 | 0.2424 |
| GradientBoosting | 0.2637 | 0.1910 |

balanced accuracy ranged from 0.71 to 0.73. Precision remained strong (approximately 0.87–0.91), although recall values were significantly lower ($\sim$ 0.43–0.48), demonstrating that the model performed worse at identifying positive cases. Values for RMSE and MAE were higher, indicating higher prediction errors. LightGBM did the worst of all. It had accuracy scores of 83%–85%, with balanced accuracy below 0.66 on average. Precision and recall were lower compared to the rest of the models, which resulted in lower F1-scores and lower ROC_AUC scores. RMSE and MAE were also highest across all models, indicating that LightGBM's predictive ability on this data was weaker. Gradient Boosting performed both well and poorly. It possessed some of the highest accuracy levels (up to 95.6%) in some folds but also showed instability, particularly in folds 3, 5, and 9, where its performance became extremely poor. Its precision and recall values jumped widely between folds, affecting global stability. Still, Gradient Boosting maintained high ROC_AUC scores, showing a perfect trade-off between sensitivity and specificity when performance was consistent.

*5.4. Error Analysis and Model Precision*

Fig. 7 and Table 5 present an evaluation of error metrics, which offer a more meaningful interpretation of the performance behavior of the various models. Among all models compared, CardioForest had the lowest RMSE of 0.2532, outperforming XGBoost (0.3003), LightGBM (0.3471), and GradientBoosting (0.2637). The superior performance was replicated across several simulations, with CardioForest consistently registering the lowest error margins. Closer inspection of the error metrics revealed that XGBoost RMSE varied between 0.300 and 0.312, while CardioForest errors were all less than 0.3 for all simulations. GradientBoosting had the widest error extremes, where RMSE went up to 0.2637 for one

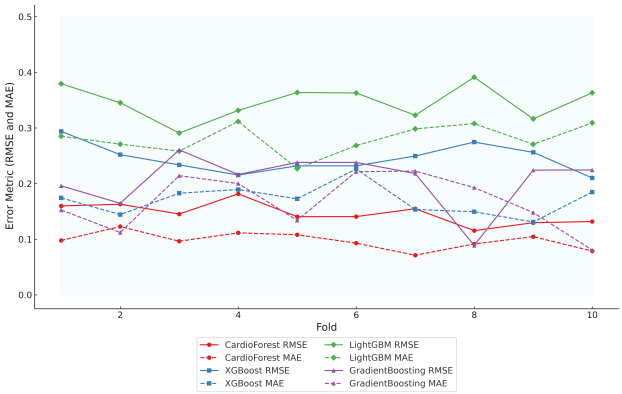


Figure 7: Error metric evaluation reveals that CardioForest consistently achieved the lowest maximum RMSE (0.2532), outperforming XGBoost and LightGBM across all simulations.

simulation. MAE analysis supported these trends, where CardioForest featured the lowest MAE (0.1944), followed by GradientBoosting (0.1910), XGBoost (0.2008), and LightGBM (0.2424). This indicates that CardioForest demonstrated the most stable and reliable error behavior across simulations, performing better than the other models consistently in both RMSE and MAE.

### *5.5. Consistency and Model Fit*

Radar plot analysis (Fig. 8) highlighted substantial differences in model fitting and performance stability. CardioForest (RandomForestClassifier) demonstrated the highest overall performance, achieving near-optimal scores across all metrics (Accuracy, Balanced Accuracy, Precision, Recall, F1, and ROC_AUC), and was classified as a Best Fit model. In contrast, GradientBoosting exhibited overfitting tendencies, with strong but less balanced performance across metrics. Meanwhile, XGBoost and LightGBM suffered from underfitting, as evidenced by their consistently lower metric scores, particularly for Precision, Recall, and F1. Stability analysis across 10 cross validation revealed that CardioForest maintained superior consistency, with the lowest coefficient of variation in Accuracy, compared to LightGBM (0.89%), GradientBoosting (1.71%), and XGBoost (2.31%). Furthermore, CardioForest achieved the narrowest range, underscoring its robustness in variable clinical environments. LightGBM exhibited intermediate stability, while XGBoost and GradientBoosting showed wider performance fluctuations, potentially compromising reliability across diverse patient cohorts.

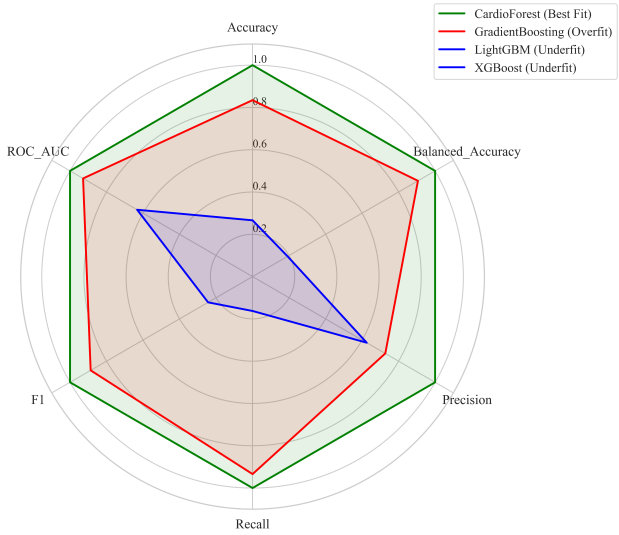


Figure 8: This figure illustrates the performance stability of the models, highlighting model fitting illustration.

### *5.6. WCT Detection Prediction*

Here in the Fig. 9, we performed WCT (Wide Complex Tachycardia) prediction detection using the CardioForest model, a Random Forest-based ensemble method optimized for clinical ECG data. The dataset used, `MIMIC-IV dataset` [30], included significant cardiac features such as `rr_interval`, `p_onset`, `p_end`, `qrs_onset`, `qrs_end`, `t_end`, `p_axis`, `qrs_axis`, `t_axis`, and `qrs_duration`. The target label, `wct_label_encoded`, was a binary value where 0 represented a normal rhythm and 1 represented the presence of WCT. Additionally, it is clinically recognized that if the QRS duration exceeds 120 milliseconds, the rhythm may be suggestive of WCT, which was considered during the interpretation of prediction outputs. The CardioForest model, with 1000 estimators, a maximum depth of 20, a minimum samples split of 5, class balancing enabled, and other parameters, has been described in Table 3, tuned for robust out-of-bag (OOB) estimation. Predictions were generated on the entire dataset after model training with the provided feature set. The study revealed that the CardioForest model identified WCT cases with an efficiency of 15.46% and detected 123,653 WCT occurrences out of a total of 800,035 samples. Normal rhythms

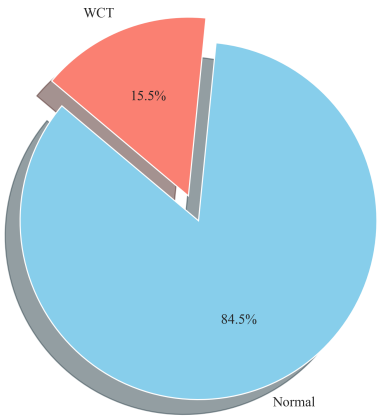


Figure 9: Prediction distribution by the CardioForest model: 15.46% WCT cases and 84.54% Normal rhythms identified from 800,035 samples, highlighting the model's ability to detect clinically significant arrhythmias.

accounted for 84.54% of the database, with 676,382 being correctly identified occurrences. These findings reflect the relatively low rate but high clinical significance of precisely identifying WCT events, as they are associated with potentially fatal arrhythmias.

Table 6: Overall Comparative Performance of Machine Learning Models

| **Model** | **Accuracy** | | **Balanced Accuracy** | | **Precision** | | **Recall** | | **F1 Score** | | **ROC AUC** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Train | Test | Train | Test | Train | Test | Train | Test | Train | Test | Train | Test |
| CardioForest | 0.9495 | 0.9495 | 0.8832 | 0.8832 | 0.9488 | 0.9489 | 0.7762 | 0.7762 | 0.8539 | 0.8538 | 0.8867 | 0.8833 |
| XGBoost | 0.8843 | 0.8835 | 0.7181 | 0.7168 | 0.8849 | 0.8810 | 0.4500 | 0.4479 | 0.5966 | 0.5937 | 0.8523 | 0.8495 |
| LightGBM | 0.8422 | 0.8422 | 0.6436 | 0.6435 | 0.6784 | 0.6787 | 0.3233 | 0.3229 | 0.4379 | 0.4374 | 0.7812 | 0.7806 |
| GradientBoosting | 0.9179 | 0.9179 | 0.8008 | 0.8006 | 0.9265 | 0.9282 | 0.6120 | 0.6113 | 0.7213 | 0.7214 | 0.8574 | 0.8544 |

Note: All values represent averages across 10-fold cross-validation.

## 6. Discussion

The Table 6 comparison clearly shows that **CardioForest** is the best-performing model for WCT detection when both performance indexes and interpretability are considered. It effectively leverages prominent ECG features—e.g., features of the QRS complex and other suitable indicators—to provide excellent diagnostic performance with excellent prediction stability. Table 6 highlights the relative performance of all the proposed and current models, further reinforcing the dominance of CardioForest in achieving a good balance between accuracy, robustness, and reliability. Most importantly, Fig. 10 plots the SHAP value analysis,

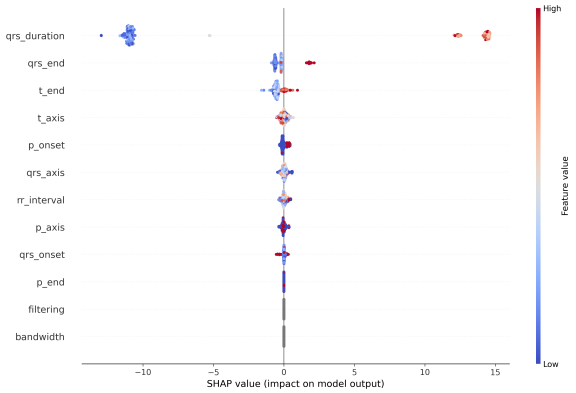


Figure 10: SHAP summary plot illustrating the contribution of each ECG feature to the CardioForest model's predictions for WCT detection. The analysis highlights the most influential features driving the model's decision-making, enhancing interpretability and building trust in AI-assisted clinical diagnosis.

providing valuable explainability insights into the model's behavior. The analysis highlights how different factors lead to WCT predictions and most prominently outlines the most important features. For example, the QRS duration remains the most significant predictor, perfectly aligning with common clinical knowledge.

One of the most significant advantages of CardioForest is its ability to consistently generate feature importance rankings in different evaluations with minimal variability. The consistency of the rankings assures healthcare professionals in relying on the model's output. Furthermore, it is capable of accurately classifying actual WCT cases from other arrhythmias with a best sensitivity-specified balance, a crucial requirement for emergency and clinical applications. The SHAP analysis and model assessment combined strengthen the essential cross-link between artificial intelligence and human clinical decision-making. AI models such as CardioForest can be relied upon as accurate assistants to physicians, freeing up precious time for cardiologists and enhancing the speed and accuracy of WCT diagnosis. Since WCT is a potentially harmful cardiac condition if not diagnosed early and correctly, integrating explainable AI tools into clinical workflows is a significant step in safeguarding patient health.

## 7. Conclusion

In this study, we explored how AI can predict Wide QRS Complex Tachycardia (WCT) more accurately and efficiently and specifically a model called CardioForest. Our results are that the model works well in making good predictions while giving easy-to-interpret results—a very important factor for doctors to make quick decisions, particularly in emergency treatment. Much scope still exists for further improving the system. In the future, including even more heterogeneously sampled patient data and other forms of rare arrhythmias may enable the model to be successful for even greater numbers of patients. We believe there is an enormous opportunity to combine CardioForest's explainable decision-making with deep learning's ability to find hidden patterns in raw ECG signals. Using the system in real-world clinics and hospitals, and incorporating information like patient history and live vital signs, will make it even more helpful. Refining and extending this approach further, we can develop a tool that doctors can rely on—one that saves time, improves accuracy, and helps deliver improved care to patients.

### CRediT authorship contribution statement

**Vaskar Chakma :** Data curation, Methodology, Software, Data generation, Validation, Writing–original draft, Investigation. **Ju Xiaolin :** Supervision, Conceptualization, Methodology, Writing–editing, Resources. **Cao Heling :** Methodology, Data curation, Investigation. **Xue Feng :** Resources, Writing - Review & Editing. **Ji Xiaodong :** Writing–editing, Data generation. **Pan Haiyan :** Validation, Formal analysis, Investigation, Writing - Review & Editing. **Gao Zhan :** Supervision, Conceptualization, Methodology, Writing–editing, Project administration.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data used in this study are derived from the MIMIC-IV-ECG: Diagnostic Electrocardiogram Matched Subset (version 1.0), which is publicly available through PhysioNet at https://doi.org/10.13026/4nqg-sb35. The corresponding author can make further data or processing scripts available upon reasonable request.

## Acknowledgment

The authors would like to express their sincere gratitude to the MIT Laboratory for Computational Physiology for providing access to the MIMIC-IV-ECG dataset that made this research possible. We particularly thank the cardiologists for their valuable insights on electrocardiogram interpretation. We also acknowledge the constructive feedback from our supervisors in the School of Artificial Intelligence and Computer Science at Nantong University that helped improve this work. Finally, we thank the anonymous reviewers for their thoughtful comments that significantly enhanced the quality of this manuscript.

## References

[1] L. Alblaihed, T. Al-Salamah, Wide complex tachycardias, Emergency Medicine Clinics 40 (4) (2022) 733–753.

[2] B. G. Song, Electrocardiographic differential diagnosis of narrow qrs and wide qrs complex tachycardias, in: Clinical Use of Electrocardiogram, IntechOpen, 2022.

[3] N. Fayyazifar, G. Dwivedi, D. Suter, S. Ahderom, A. Maiorana, O. Clarkin, S. Balamane, N. Saha, B. King, M. S. Green, et al., A novel convolutional neural network structure for differential diagnosis of wide qrs complex tachycardia, Biomedical Signal Processing and Control 81 (2023) 104506.

[4] S. Kurl, T. H. Mäkikallio, P. Rautaharju, V. Kiviniemi, J. A. Laukkanen, Duration of qrs complex in resting electrocardiogram is a predictor of sudden cardiac death in men, Circulation 125 (21) (2012) 2588–2594.

[5] E. Silvetti, O. Lanza, F. Romeo, A. Martino, E. Fedele, C. Lanzillo, C. Crescenzi, F. Fanisio, L. Calò, The pivotal role of ecg in cardiomyopathies, Frontiers in Cardiovascular Medicine 10 (2023) 1178163.

[6] J. Hampton, J. Hampton, D. Adlam, The ecg made easy e-book: The ecg made easy e-book, Elsevier Health Sciences, 2024.

[7] K. Badura, D. Buławska, B. Dąbek, A. Witkowska, W. Lisińska, E. Radzioch, S. Skwira, E. Młynarska, J. Rysz, B. Franczyk, Primary electrical heart disease—principles of pathophysiology and genetics, International Journal of Molecular Sciences 25 (3) (2024) 1826.

[8] M. Jastrzebski, P. Kukla, D. Czarnecka, K. Kawecka-Jaszcz, Comparison of five electrocardiographic methods for differentiation of wide qrs-complex tachycardias, Europace 14 (8) (2012) 1165–1171.

[9] O. E. Osadchii, Role of abnormal repolarization in the mechanism of cardiac arrhythmia, Acta physiologica 220 (2017) 1–71.

[10] G. Tse, Mechanisms of cardiac arrhythmias, Journal of arrhythmia 32 (2) (2016) 75–81.

[11] E. Adib, Generating Synthetic Electrocardiograms Using Deep Generative Algorithms, The University of Texas at San Antonio, 2023.

[12] X. Yang, T. Li, Q. Su, Y. Liu, C. Kang, Y. Lyu, L. Zhao, Y. Nie, Y. Pan, Application of large language models in disease diagnosis and treatment, Chinese Medical Journal 138 (02) (2025) 130–142.

[13] K. S. Adewole, H. A. Mojeed, J. A. Ogunmodede, L. A. Gabralla, N. Faruk, A. Abdulkarim, E. Ifada, Y. Y. Folawiyo, A. A. Oloyede, L. A. Olawoyin, et al., Expert system and decision support system for electrocardiogram interpretation and diagnosis: review, challenges and research directions, Applied Sciences 12 (23) (2022) 12342.

[14] A. N. Meyer, T. D. Giardina, L. Khawaja, H. Singh, Patient and clinician experiences of uncertainty in the diagnostic process: current understanding and future directions, Patient Education and Counseling 104 (11) (2021) 2606–2615.

[15] M. A. Muzammil, S. Javid, A. K. Afridi, R. Siddineni, M. Shahabi, M. Haseeb, F. Fariha, S. Kumar, S. Zaveri, A. J. Nashwan, Artificial intelligence-enhanced electrocardiography for accurate diagnosis and management of cardiovascular diseases, Journal of Electrocardiology 83 (2024) 30–40.

[16] M. B. Abubaker, B. Babayiğit, Detection of cardiovascular diseases in ecg images using machine learning and deep learning methods, IEEE transactions on artificial intelligence 4 (2) (2022) 373–382.

[17] A. A. Ahmed, W. Ali, T. A. Abdullah, S. J. Malebary, Classifying cardiac arrhythmia from ecg signal using 1d cnn deep learning model, Mathematics 11 (3) (2023) 562.

[18] M. Ramkumar, C. Ganesh Babu, A. Manjunathan, S. Udhayanan, M. Mathankumar, R. Sarath Kumar, A graphical user interface based heart rate monitoring process and detection of pqrst peaks from ecg signal, in: Inventive Computation and Information Technologies: Proceedings of ICICIT 2020, Springer, 2021, pp. 481–496.

[19] R. Brisk, Towards broader application of deep learning methods to the automated analysis of electrocardiograms, Ph.D. thesis, Ulster University (2023).

[20] N. Nissa, S. Jamwal, M. Neshat, A technical comparative heart disease prediction framework using boosting ensemble techniques, Computation 12 (1) (2024) 15.

[21] Z. Sadeghi, R. Alizadehsani, M. A. Cifci, S. Kausar, R. Rehman, P. Mahanta, P. K. Bora, A. Almasri, R. S. Alkhawaldeh, S. Hussain, et al., A review of explainable artificial intelligence in healthcare, Computers and Electrical Engineering 118 (2024) 109370.

[22] Z.-Z. Li, W. Zhao, Y. Mao, D. Bo, Q. Chen, P. Kojodjojo, F. Zhang, A machine learning approach to differentiate wide qrs tachycardia: distinguishing ventricular tachycardia from supraventricular tachycardia, Journal of Interventional Cardiac Electrophysiology 67 (2024) 1391–1398. doi:10.1007/s10840-024-01743-9.

[23] B. J. Chow, N. Fayyazifar, S. Balamane, N. Saha, M. Farooqui, B. A. Hasan, O. Clarkin, M. Green, A. Maiorana, M. Golian, G. Dwivedi, Interpreting wide-complex tachycardia with the use of artificial intelligence, Canadian Journal of Cardiology 40 (2024) 1965–1973. doi:10.1016/j.cjca.2024.03.027.

[24] M. Bhattacharya, D. Lu, S. M. Kudchadkar, G. V. Greenland, P. Lingamaneni, C. P. Corona-Villalobos, Y. Guan, J. E. Marine, J. E. Olgin, S. L. Zimmerman, T. P. Abraham, H. Shatkay, M. R. Abraham, Identifying ventricular arrhythmias and their predictors by applying machine learning methods to electronic health records in patients with hypertrophic cardiomyopathy(hcm-var-risk model), CoRR abs/2109.09210 (2021). arXiv:2109.09210.
URL https://arxiv.org/abs/2109.09210

[25] S. Hong, Y. Zhou, J. Shang, C. Xiao, J. Sun, Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review, Computers in Biology and Medicine 122 (2020) 103801. doi:10.1016/j.compbiomed.2020.103801.

[26] A. M. May, B. B. Katbamna, P. A. Shaikh, S. LoCoco, E. Deych, R. Zhou, L. Liu, K. M. Mikhova, R. Ghadban, P. S. Cuculich, D. H. Cooper, T. M. Maddox, P. A. Noseworthy, A. Kashou, Automated differentiation of wide qrs complex tachycardia using qrs complex polarity, Communications Medicine 4 (2024) 282. doi:10.1038/s43856-024-00725-2.

[27] P. Howladar, M. Sahoo, Supraventricular tachycardia detection and classification model

of ecg signal using machine learning (2021). arXiv:2112.12953.
URL https://arxiv.org/abs/2112.12953

[28] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, A. Y. Ng, Cardiologist-level arrhythmia detection with convolutional neural networks, CoRR abs/1707.01836 (2017). arXiv:1707.01836.
URL http://arxiv.org/abs/1707.01836

[29] M. Frausto-Avila, J. P. Manriquez-Amavizca, A. U'Ren, M. A. Quiroz-Juarez, Compact neural network algorithm for electrocardiogram classification (2024). arXiv:2412.17852.
URL https://arxiv.org/abs/2412.17852

[30] B. Gow, T. Pollard, L. A. Nathanson, A. Johnson, B. Moody, C. Fernandes, N. Greenbaum, J. W. Waks, P. Eslami, T. Carbonati, A. Chaudhari, E. Herbst, D. Moukheiber, S. Berkowitz, R. Mark, S. Horng, Mimic-iv-ecg: Diagnostic electrocardiogram matched subset (version 1.0), PhysioNet (10 2023). doi:https://doi.org/10.13026/4nqg-sb35.
URL https://physionet.org/content/mimic-iv-ecg/1.0/

[31] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, H. E. Stanley, Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. circulation [online]. doi:https://doi.org/10.13026/b95v-ff39.

[32] H. Li, R. Monger, E. Pishgar, M. Pishgar, Icu readmission prediction for intracerebral hemorrhage patients using mimic iii and mimic iv databases, medRxiv (2025) 2025–01.

[33] D. Yoon, C. Han, D. W. Kim, S. Kim, S. Bae, J. A. Ryu, Y. Choi, Redefining health care data interoperability: empirical exploration of large language models in information exchange, Journal of Medical Internet Research 26 (2024) e56614.

[34] D. Sharma, N. Kohli, Wfdb software for python: A toolkit for physiological signals, in: 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC), IEEE, 2023, pp. 86–92.

[35] A. J. Goodwin, D. Eytan, W. Dixon, S. D. Goodfellow, Z. Doherty, R. W. Greer, A. McEwan, M. Tracy, P. C. Laussen, A. Assadi, et al., Timing errors and temporal uncertainty in clinical databases—a narrative review, Frontiers in Digital Health 4 (2022) 932599.

[36] P. Gupta, A. Bagchi, Data manipulation with pandas, in: Essentials of Python for Artificial Intelligence and Machine Learning, Springer, 2024, pp. 197–235.

[37] D. Morelli, A. Rossi, M. Cairo, D. A. Clifton, Analysis of the impact of interpolation methods of missing rr-intervals caused by motion artifacts on hrv features estimations, Sensors 19 (14) (2019) 3163.

[38] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, J. M. O'Sullivan, A review of feature selection methods for machine learning-based disease risk prediction, Frontiers in Bioinformatics 2 (2022) 927312.

[39] L. Sharma, S. Dandapat, A. Mahanta, Multichannel ecg data compression based on multiscale principal component analysis, IEEE Transactions on Information technology in Biomedicine 16 (4) (2012) 730–736.

[40] P. Melzi, R. Tolosana, A. Cecconi, A. Sanz-Garcia, G. J. Ortega, L. J. Jimenez-Borreguero, R. Vera-Rodriguez, Analyzing artificial intelligence systems for the prediction of atrial fibrillation from sinus-rhythm ecgs including demographics and feature visualization, Scientific reports 11 (1) (2021) 22786.

[41] F. Ochieng'Odhiambo, Comparative study of various methods of handling missing data, Mathematical Modelling and Applications 5 (2) (2020) 87.

[42] T. Y. Deo, A. Sanju, Data imputation and comparison of custom ensemble models with existing libraries like xgboost, catboost, adaboost and scikit learn for predictive equipment failure, Materials Today: Proceedings 72 (2023) 1596–1604.

[43] R. Garreta, G. Moncecchi, Learning scikit-learn: machine learning in python, Vol. 2013, Packt Publishing Birmingham, 2013.

[44] K. C. Siontis, P. A. Noseworthy, Z. I. Attia, P. A. Friedman, Artificial intelligence-enhanced electrocardiography in cardiovascular disease management, Nature Reviews Cardiology 18 (7) (2021) 465–478.

[45] F. D. Regoli, M. Cattaneo, F. Kola, A. Thartori, H. Bytyci, L. Saccarello, M. Amoruso, M. Di Valentino, A. Menafoglio, Management of hemodynamically stable wide qrs complex tachycardia in patients with implantable cardioverter defibrillators, Frontiers in Cardiovascular Medicine 9 (2023) 1011619.

[46] J. Mulo, H. Liang, M. Qian, M. Biswas, B. Rawal, Y. Guo, W. Yu, Navigating challenges and harnessing opportunities: Deep learning applications in internet of medical things, Future Internet 17 (3) (2025) 107.

[47] H. Khan, A. Bilal, M. A. Aslam, H. Mustafa, Heart disease detection: A comprehensive analysis of machine learning, ensemble learning, and deep learning algorithms., Nano Biomedicine & Engineering 16 (4) (2024).

[48] K.-L. Du, R. Zhang, B. Jiang, J. Zeng, J. Lu, Foundations and innovations in data fusion and ensemble learning for effective consensus, Mathematics 13 (4) (2025) 587.

[49] G. S. Manivannan, H. Rajaguru, S. V. Talawar, et al., Cardiovascular disease detection from cardiac arrhythmia ecg signals using artificial intelligence models with hyperparameters tuning methodologies, Heliyon 10 (17) (2024).

[50] J. H. B. d. Silva, P. C. Cortez, S. K. Jagatheesaperumal, V. H. C. de Albuquerque, Ecg measurement uncertainty based on monte carlo approach: an effective analysis for a successful cardiac health monitoring system, Bioengineering 10 (1) (2023) 115.