

Cardiology Research: Machine Learning Approaches for Pediatric AVNRT Detection

Hunter Mena Trinity Tran

School of Computer Science, University of Oklahoma

Course Number: CS 3440 **Credits:** 3

Company & Sponsor: Data Institute for Societal Challenges (DISC), Dr. David Ebert

Project Supervisor: Dr. Gopichandh Danala

1. Introduction

Pediatric arrhythmia refers to irregular heart rhythms that occur in children, ranging from benign conditions to requiring urgent medical intervention. Among these, **Supraventricular Tachycardia (SVT)** stands out as one of the most prevalent types, characterized by an abnormally fast heart rate originating above the ventricles. SVT is also the most common arrhythmia in the pediatric population.¹ While SVT itself encompasses various subtypes, our study focuses on **Atrioventricular Nodal Reentrant Tachycardia (AVNRT)**, a particularly significant form due to its potential to impact a child's stability and quality of life.

Early and accurate identification of AVNRT is crucial for timely intervention, but diagnosis can be challenging due to overlapping features with other arrhythmic events and the dynamic nature of pediatric electrocardiograms (ECG). Recent advances in machine learning offer promising tools to improve diagnostic precision by analyzing complex patterns within ECG data that may not be easily discernible using traditional clinical methods.

In this report, we present a comprehensive study that applies a range of machine learning algorithms to classify AVNRT based on extracted ECG features. We describe the dataset used, preprocessing techniques, and the evaluation of multiple classification models. Our goal is to identify the most effective predictive model for supporting the diagnosis of AVNRT in pediatric patients and to contribute to the growing intersection of computational methods and cardiology.

2. Previous Work

Automated analysis of ECG signals for arrhythmia detection has advanced significantly with the help of machine learning and deep learning techniques. Traditional ML models, such as Support Vector Machines (SVM), Random Forests (RF), and k-Nearest Neighbors (k-NN), often rely on carefully engineered features derived from ECG morphology and heart rate variability (HRV). Although effective in some contexts, these approaches are limited by their dependence on domain-specific feature extraction and their sensitivity to noise and data imbalance.

Deep learning approaches, by contrast, extract hierarchical features directly from raw ECG signals. Hammad et al.² introduced a ResNet-LSTM deep learning framework, optimized using a Genetic Algorithm (GA), which achieved high accuracy (up to 98 percent) on general arrhythmia detection using the MIT-BIH Arrhythmia Database. Their multitier model combined convolutional

and recurrent layers to capture both spatial and temporal ECG patterns, highlighting the potential of deep learning in automated arrhythmia diagnosis.

Zvuloni et al. (2023)³ compared classical Feature Engineering (FE) methods with deep learning models for ECG-based tasks such as arrhythmia classification, atrial fibrillation risk prediction, and age estimation. They found that deep learning models generally require large datasets to outperform FE-based methods, particularly in regression tasks, while both approaches yielded similar results in classification tasks when datasets were small. Moreover, combining FE with deep learning did not significantly improve performance, suggesting redundancy in combined models.

Mousavi and Afghah (2019)⁴ developed a sequence-to-sequence deep learning model leveraging BiLSTM and SMOTE to handle class imbalance in ECG datasets. Their model achieved state-of-the-art performance in both intra- and inter-patient evaluations, demonstrating the effectiveness of sequential models for heartbeat classification. However, their work, like many others, focused on broader arrhythmia categories and did not specifically address Atrioventricular Nodal Reentrant Tachycardia (AVNRT).

3. Motive and Research Gap

Our research addresses the critical gap in automated detection of AVNRT in pediatric patients. We aim to develop a machine learning model that incorporates preprocessing steps (e.g., outlier removal, normalization, PCA-based dimensionality reduction) tailored to pediatric ECG data. Focusing specifically on AVNRT, our goal is to enhance early and accurate AVNRT detection, supporting clinical decision-making in pediatric cardiology where timely intervention can significantly impact patient outcomes.

4. Current Dataset

4.1 Data Sources

The dataset used in this study was obtained from clinical ECG recordings sourced from pediatric patients evaluated for suspected arrhythmias, specifically Atrioventricular Nodal Reentrant Tachycardia (AVNRT). The data were provided through a collaboration with the University of Oklahoma Health Sciences Center (OUHSC).

4.2 Summary of Data Features and Target Variables

Each patient record includes standard 12-lead ECG signals. The dataset comprises a diverse range of ECG features extracted from various wave components including:

- **P-wave metrics:** Duration, Amplitude, Area
- **QRS complex metrics:** Max R Amplitude, Max S Amplitude, Duration
- **ST segment and T-wave metrics:** Maximum/Minimum ST levels, T Duration, T Peak Amplitude
- **Demographic Metadata:** Age, Acquisition Date, Patient ID (removed during preprocessing)

The target variable for this study is the binary classification label indicating the presence or absence of AVNRT, denoted in the dataset as AVNRT = 1 (presence) or 0 (absence).

5. Methods

5.1 Approach

In this study, we developed a comprehensive pipeline for the classification of Atrioventricular Nodal Reentrant Tachycardia (AVNRT) using pediatric ECG data. Our approach integrates data pre-processing, feature extraction, and dimensionality reduction, followed by machine learning-based classification. Key preprocessing steps include normalization, outlier removal, and PCA-based feature reduction to retain 95 percent of the data variance. We explored multiple modeling strategies, including classical machine learning classifiers and deep learning frameworks, to evaluate the effectiveness of each in accurately detecting AVNRT.

The machine learning pipeline followed a structured, repeatable process to ensure consistency and reliability:

1. Data cleaning to remove irrelevant or inconsistent values.
2. Feature scaling and normalization.
3. Splitting into training and test sets (80/20).
4. Addressing class imbalance using SMOTE on the training set.
5. Hyperparameter tuning using RandomizedSearchCV with 5-fold cross-validation.
6. Training of the best model.
7. Evaluation of test set performance using predefined metrics.

This structured approach allows for rigorous comparison and reproducibility across all evaluated models.

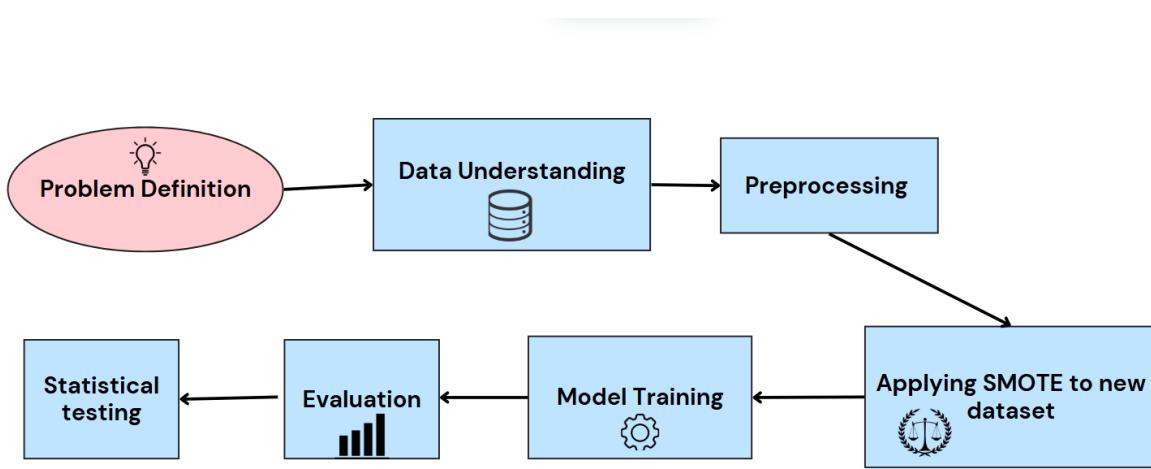


Figure 1: Research and Modeling Pipeline

5.2 Data Preprocessing

Data preprocessing is a crucial step in the data analysis and machine learning pipeline that involves transforming raw data into a clean, structured, and usable format. Real-world data is often incomplete, inconsistent, or noisy, which can negatively affect the performance of models. Preprocessing addresses these issues through tasks such as handling missing values, removing irrelevant or redundant features, normalizing scales, detecting and removing outliers, and reducing dimensionality. These steps ensure that the dataset accurately represents the underlying patterns, improves model training efficiency, and enhances the reliability and generalizability of predictive results.

1. **Removal of Irrelevant Columns:** Certain columns were excluded from the dataset to improve model focus and mitigate potential bias. Specifically, the following were removed:
 - Patient ID
 - Acquisition Date:
 - Gender
2. **Remove Missing Data:** All missing values were imputed using the mean of their respective columns. This ensured that no rows were excluded due to NaNs.
3. **Outlier Detection and Removal:** Outliers were removed using the Interquartile Range (IQR) method. Specifically, any values falling outside of $[Q_1 - 1.5 \cdot IQR, Q_3 + 1.5 \cdot IQR]$ were altered to the mean of their respective columns.
4. **Normalization:** All numeric features were normalized using Min-Max scaling to bring values into the range $[0, 1]$, ensuring equal contribution across features during model training.
5. **Dimensionality Reduction:** Principal Component Analysis (PCA) was applied to the feature set. The number of components was selected to retain 95% of the explained variance, reducing redundancy while preserving the majority of the information.

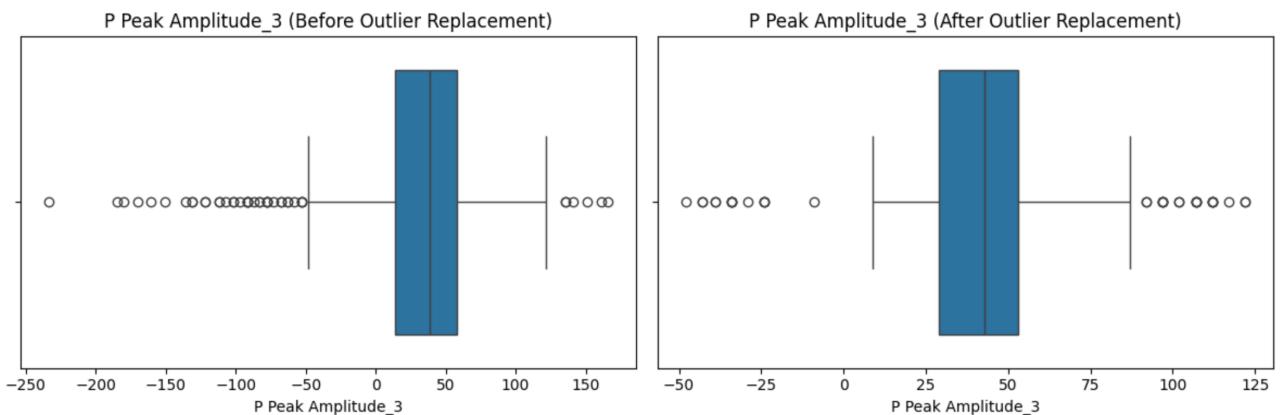


Figure 2: Before and After Outlier Removal.

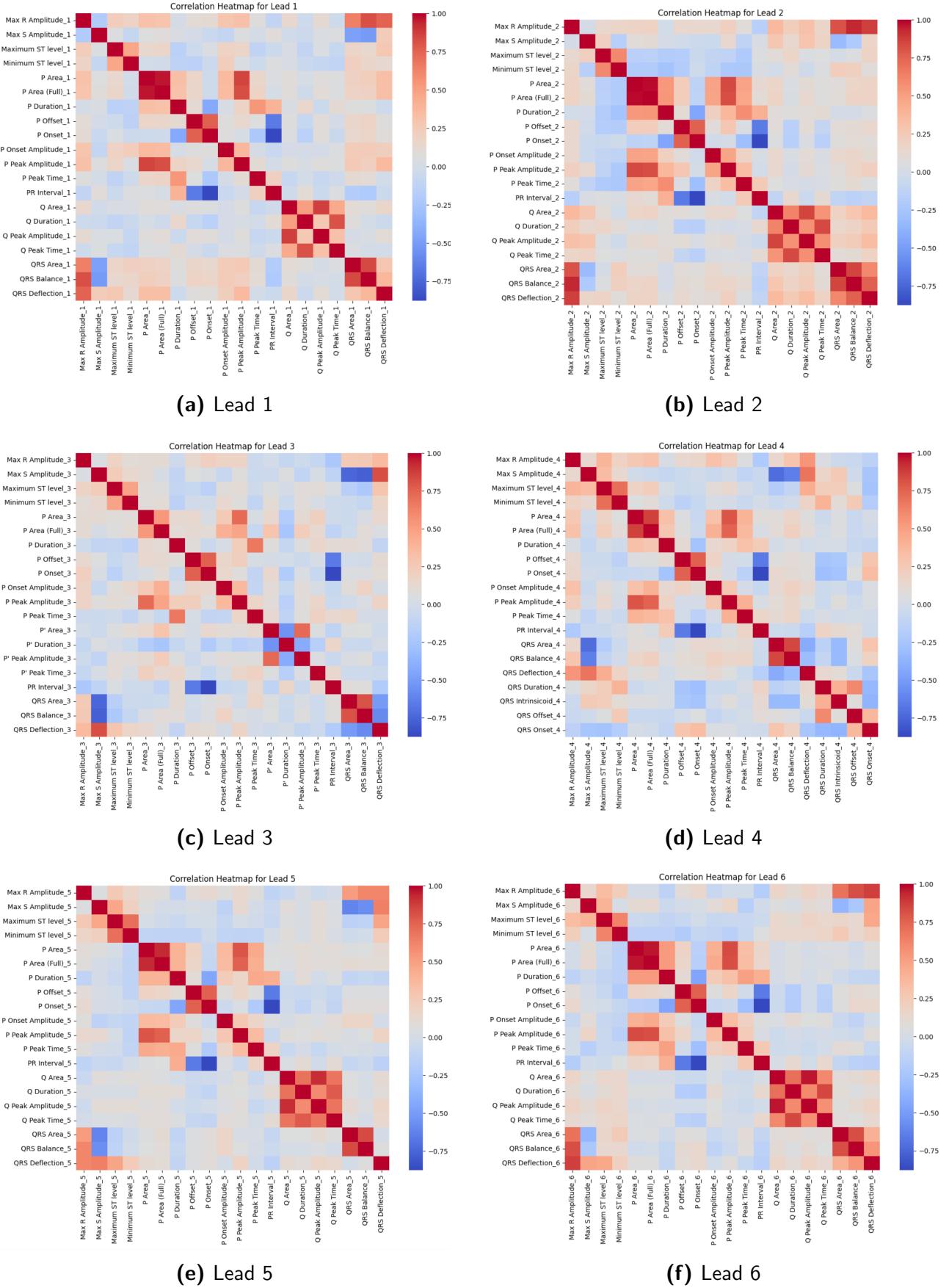


Figure 3: Correlation heatmaps for ECG leads 1–6.

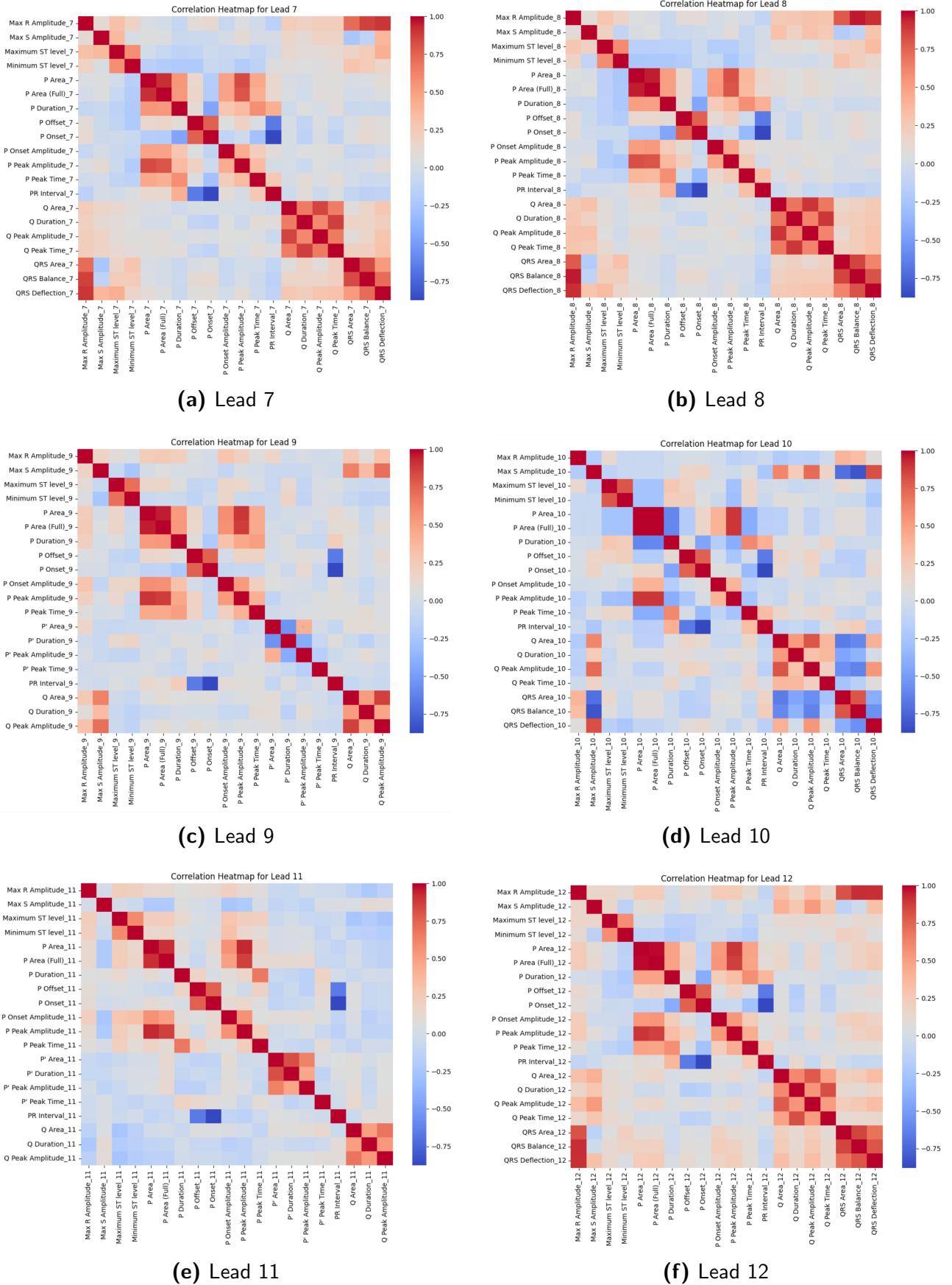


Figure 4: Correlation heatmaps for ECG leads 7–12.

5.3 Machine Learning Models

To address the classification of AVNRT in pediatric patients, we employed a variety of machine learning models. The goal was to evaluate the predictive strength and generalizability of each model on our preprocessed dataset.

The following models were implemented and tuned:

- **Logistic Regression (LR):** A baseline linear model valued for its interpretability. It was tuned using the regularization parameter C and solvers such as `liblinear` and `lbfgs`.
- **Support Vector Machine (SVM):** A kernel-based model suitable for high-dimensional data. Tuning included kernel types (`linear`, `rbf`), C , and γ .
- **K-Nearest Neighbors (KNN):** A distance-based model tuned for the number of neighbors, weight functions, and distance metrics (`euclidean`, `manhattan`).
- **Naive Bayes:** A lightweight probabilistic classifier used without hyperparameter tuning as a baseline.
- **Random Forest (RF):** An ensemble method based on bagging of decision trees, tuned for `n_estimators`, `max_depth`, and `min_samples_split`.
- **Gradient Boosting (GB):** A boosting ensemble method tuned using `learning_rate`, `n_estimators`, and `max_depth`.
- **XGBoost:** An efficient implementation of gradient boosting with regularization, tuned for `learning_rate`, `max_depth`, and `n_estimators`.
- **LightGBM:** A high-performance gradient boosting method tuned for `num_leaves`, `learning_rate`, and `max_depth`.
- **CatBoost:** A gradient boosting model optimized for categorical features, tuned using `iterations`, `depth`, and `learning_rate`.

All models were trained using an 80/20 train-test split. Hyperparameter tuning was conducted via `RandomizedSearchCV` with five-fold cross-validation. Final performance evaluation was carried out on the held-out test set.

5.4 Performance Metrics

To comprehensively evaluate model performance, we used the following classification metrics:

- **AUC (Area Under the ROC Curve):** Reflects the model's ability to distinguish between classes across all classification thresholds.
- **Accuracy:** Proportion of total correct predictions out of all predictions.
- **Precision:** Ratio of true positives to all predicted positives (positive predictive value).
- **Recall:** Ratio of true positives to all actual positives.
- **F1-Score:** Harmonic mean of precision and recall, balancing false positives and false negatives.

- **Sensitivity:** The ability of the model to correctly identify actual positives (true positive rate).
- **Specificity:** The ability of the model to correctly identify actual negatives (true negative rate).

These metrics were chosen to provide a well-rounded assessment of each model, balancing clinical relevance with statistical robustness. Special attention was given to sensitivity and specificity, which are critical in distinguishing between actual AVNRT cases and nonarrhythmic events.

6. Results

6.1 Model Performance Overview

Table 1: Model performance metrics on the test set.

Model	AUC	Accuracy	Precision	Recall	F1-score	Sensitivity	Specificity
XGBoost	0.90	0.79	0.78	0.79	0.79	0.86	0.60
Gradient Boosting	0.86	0.81	0.82	0.81	0.81	0.84	0.73
LightGBM	0.84	0.71	0.72	0.71	0.71	0.78	0.53
Logistic Regression	0.82	0.81	0.82	0.81	0.81	0.84	0.73
SVM	0.81	0.77	0.76	0.77	0.76	0.86	0.53
Random Forest	0.76	0.75	0.77	0.75	0.76	0.78	0.67
KNN	0.70	0.63	0.67	0.63	0.65	0.68	0.53
Naive Bayes	0.68	0.73	0.74	0.73	0.74	0.78	0.60

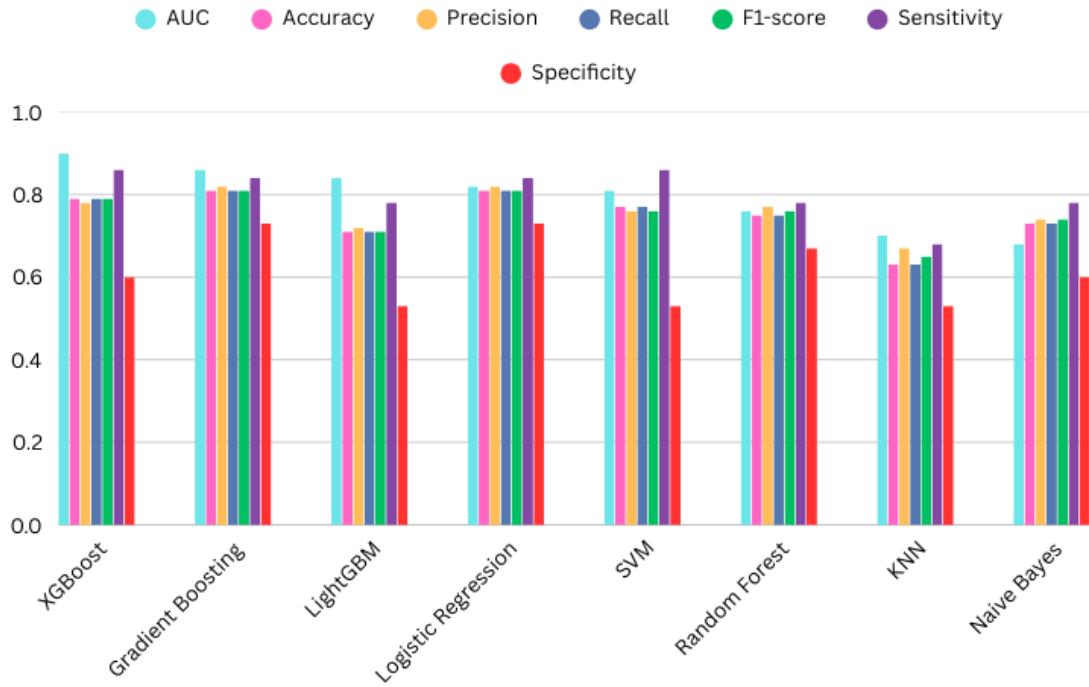


Figure 5: Bar chart comparing the performance of various models across multiple evaluation metrics. This visual summary complements the table above and aids in identifying models with balanced performance.

6.2 Paired T-Test for Model Comparison

To evaluate whether differences in predictive performance across models were statistically significant, we conducted paired t-tests on prediction score distributions using a threshold of $p < 0.05$.

Significant Comparisons

The following model pairs showed statistically significant differences:

- **XGBoost** significantly outperformed:
 - Gradient Boosting ($p = 0.02$)
 - Random Forest ($p < 0.001$)
 - KNN ($p = 0.03$)
- **Gradient Boosting** significantly outperformed:
 - SVM ($p < 0.001$)
 - Random Forest ($p = 0.02$)
- **LightGBM** significantly outperformed:
 - Random Forest ($p = 0.01$)
- **SVM** significantly outperformed:
 - KNN ($p < 0.001$)

Non-Significant Comparisons

No significant difference was observed between the following pairs:

- XGBoost and: LightGBM, Logistic Regression, SVM, Naive Bayes
- Gradient Boosting and: LightGBM, Logistic Regression, KNN, Naive Bayes
- LightGBM and: Logistic Regression, SVM, KNN, Naive Bayes
- Logistic Regression and: SVM, Random Forest, KNN, Naive Bayes
- SVM and: Naive Bayes
- Random Forest and: KNN, Naive Bayes
- KNN and: Naive Bayes

Interpretation

While models like XGBoost and Gradient Boosting demonstrated significantly better performance compared to others (notably Random Forest and KNN), many top-tier models such as Logistic Regression, LightGBM, and XGBoost did not differ significantly from each other. This suggests that multiple models may be viable, depending on constraints or computational efficiency.

7. Conclusion

Summary of Findings

This study explored multiple machine learning algorithms for arrhythmia detection using ECG signals. The evaluation across various metrics (AUC, accuracy, precision, recall, F1-score, sensitivity, and specificity) demonstrated that **XGBoost**, **Gradient Boosting**, and **Logistic Regression** consistently ranked among the top-performing models.

- **XGBoost** achieved the highest AUC (0.90) and demonstrated strong balance across all performance metrics.
- **Gradient Boosting** and **Logistic Regression** showed comparable performance, particularly in precision and recall.
- Simpler models like **Naive Bayes** and **KNN** underperformed in accuracy and generalization.

Best-Performing Model

Based on both the raw metrics and the results of the paired t-tests, **XGBoost** emerges as the best-performing model, offering statistically significant improvements over several other classifiers including Gradient Boosting, Random Forest, and KNN.

Limitations and Future Work

- Although we tuned hyper parameters and compared many models using evaluation metrics, we did not explore **model ensembling** or techniques like stacking and voting, which could improve generalization by combining model strengths.
- Due to the **use of preprocessed data**, we did not directly test with the raw ECG signals, in future implementations this could provide a deeper understanding of the data pipeline and allow end-to-end model development.
- While standard metrics and t-tests were used, a deeper evaluation with cross-validation, external datasets, and long-term studies would be necessary to fully assess model robustness and utility.

Acknowledgement

This work is supported in part by the Data Institute for Societal Challenges, University of Oklahoma, and the Department of Pediatrics at the University of Oklahoma Health Sciences Center.

References

- [1] Kafal H, Ergl AB. Approach to Tachycardia in Children. Turkish Archives of Pediatrics. 2022;57(5):506-13. Available from: <https://doi.org/10.5152/TurkArchPediatr.2022.22033>.
- [2] Hammad M, Iliyasu AM, Subasi A, Ho ESL, El-Latif AAA. A Multitier Deep Learning Model for Arrhythmia Detection. IEEE Transactions on Instrumentation and Measurement. 2020;70:1-9. Available from: <https://doi.org/10.1109/TIM.2020.3033072>.
- [3] Zvuloni E, Read J, Ribeiro AH, Ribeiro ALP, Behar JA. On Merging Feature Engineering and Deep Learning for Diagnosis, Risk Prediction and Age Estimation Based on the 12-lead ECG. IEEE Transactions on Biomedical Engineering. 2023;70(7):2227-36. Available from: <https://doi.org/10.1109/tbme.2023.3239527>.
- [4] Mousavi S, Afghah F. Inter- and Intra-Patient ECG Heartbeat Classification for Arrhythmia Detection: A Sequence to Sequence Deep Learning Approach. Proceedings of the School of Informatics, Computing and Cyber Systems, Northern Arizona University. 2019. Available at: <https://pubmed.ncbi.nlm.nih.gov/33082716/>.

Appendix

A Report on Preprocessing Code for ECG Heartbeat Classification

This is designed to prepare data for a sequence-to-sequence deep learning model. The code processes ECG signals from the MIT-BIH Arrhythmia Database, and then prepares them for both intra-patient and inter-patient evaluation paradigms.

Preprocessing Pipeline The preprocessing pipeline can be broken down into the following key stages:

A1 Remove Irrelevant Columns

Certain columns were excluded from the dataset to improve model focus and mitigate potential bias. Specifically, the following were removed:

- **Patient ID:** A unique identifier that carries no physiological meaning and could lead to data leakage or overfitting, as the model might learn patient-specific patterns rather than generalizable features.
- **Acquisition Date:** A metadata field reflecting the time of recording, which does not influence the electrical activity of the heart and hence offers no predictive value.
- **Gender:** While gender can influence ECG patterns, its removal in this context was intended to prevent demographic bias in the model and ensure a focus on purely signal-based classification. This decision also supports generalization across diverse populations, especially if the dataset is not balanced across genders.

A2 Handling Missing Data:

Prior to imputation, all column names were cleaned to remove extraneous whitespace characters. To ensure data integrity, invalid zero entries, which may represent missing or corrupted measurements in physiological signals, were replaced with NaN (Not a Number) values in all columns except for the target variable, AVNRT. This exception was made to preserve the class labels. Subsequently, missing values were imputed using the mean of each respective feature column.

A3 Outlier Detection and Removal:

Extreme values can disproportionately influence model training, especially for algorithms sensitive to scale. To mitigate this, we employed the Interquartile Range (IQR) method, identifying and then altering the data points lying outside the range to the mean of the column:

A4 Normalization:

Features were normalized using Min-Max scaling to a range of [0, 1]. This step ensures that all variables contribute equally to the learning process by eliminating differences in scale and unit.

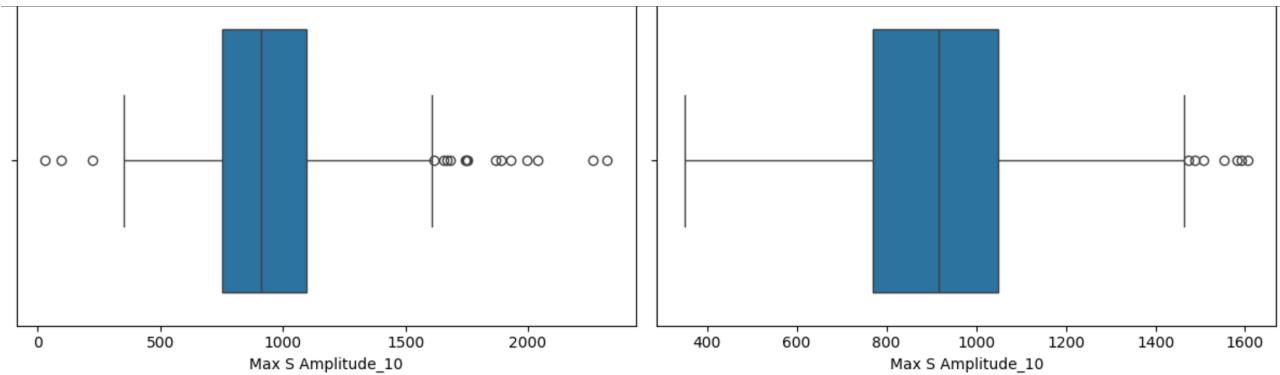


Figure 6: Boxplot of Max S Amplitude_10 before (left) and after (right) outlier removal using the IQR method.

A5 Dimensionality Reduction using PCA:

To reduce feature redundancy and improve computational efficiency, we applied Principal Component Analysis (PCA). PCA transforms the original features into a smaller set of orthogonal components that capture the maximum variance in the data. The number of principal components was determined by computing the cumulative explained variance of the dataset. We selected the smallest number of components such that at least 95% of the original variance was preserved. Specifically, the number of components k was chosen as: 101.

B Report on MIT-BIH Data Processing and Analysis

Overview

This report outlines the steps taken to process the MIT-BIH Arrhythmia Database using the WFDB Toolbox for MATLAB. Two different paradigms were analyzed: intra-patient and inter-patient paradigms.

B1 Data Preparation

- Installed the WFDB Toolbox for MATLAB.
- Executed `download_MITBIHDB.m` to download the MIT-BIH database.
- Set local paths for the database and toolbox.
- Ran `seq2seq_mitbih_AAMI.m` for intra-patient paradigm.
- Ran `seq2seq_mitbih_AAMI_DS1DS2.m` for inter-patient paradigm.

B2 Results

B3 Intra-Patient Paradigm (seq2seq_mitbih_AAMI.m)

Class	Count
F	802
N	90,502
Q	8,031
S	2,777
V	7,226
Total	109,338

Table 2: Intra-Patient Paradigm Results

B4 Inter-Patient Paradigm (seq2seq_mitbih_AAMI_DS1DS2.m)

Class	Count
F	414
N	45,798
Q	7
S	941
V	3,782
Total	50,942

Table 3: Inter-Patient Paradigm Results (DS1)

Class	Count
F	388
N	44,198
Q	7
S	1,836
V	3,217
Total	49,646

Table 4: Inter-Patient Paradigm Results (DS2)

B5 Conclusion

The processing of the MIT-BIH database using the intra-patient and inter-patient paradigms was successful. The intra-patient paradigm processed a total of 109,338 beats, whereas the inter-patient paradigm analyzed DS1 and DS2 separately, processing 50,942 and 49,646 beats, respectively. The analysis provides a structured classification of different heartbeat types, which can be used for further research in arrhythmia detection.