Maharshi Karve Stree Shikshan Samstha's
Cummins College of Engineering for Women, Pune
(An autonomous Institute affiliated to Savitribai Phule Pune University)
Department of Computer Engineering

A presentation on

PulseX: Automated Arrhythmia Detection

Submitted to

Dept. Of Computer Engineering, Cummins College of Engineering for Women, Pune, India

In the Partial Fulfillment of the requirements of the 23PCCE501L

Artificial Intelligence and Machine Learning Laboratory

TY B. Tech Semester-I 2025-26

Submitted by

UCE2023415 Bhakti Chaudhari

UCE2023420 Samruddhi Dargode

UCE2023421 Shravni Garde

Batch: A2

Arrhythmia in India: A Hidden but Growing Health Risk

Cardiovascular diseases are a leading cause of mortality. Cardiac arrhythmias, or irregular heart rhythms, require immediate detection and intervention.

AF(Atrial Fibrillation) detected in **17.4%** of high-risk Indian adults during 24-hour Holter monitoring. PMC

PIn community screening, AF prevalence is **0.196%**, many with rheumatic heart disease. PMC

Indian AF patients are relatively young (average ~55 yrs), and **almost half** have **valvular (rheumatic) AF**. PubMed

Many patients are **under-treated**: in one survey, **18.4%** of AF patients had no therapy despite high stroke risk. PMC

From a pan-India registry: **15% arrhythmia patients** had AF, **10%** had other supraventricular arrhythmias, and **4.5%** had dangerous ventricular arrhythmias. PMC

PulseX – Automated Arrhythmia Detection

A Signal Processing and Machine Learning Approach

An Electrocardiogram (ECG) is a recording of the electrical activity of the heart. It reflects how the heart muscles contract and relax during each heartbeat.

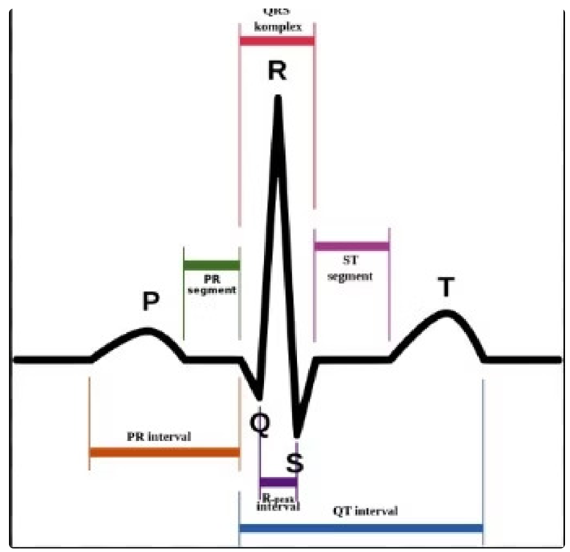
Clinical Applications

- Monitor heart rhythm and rate.
- Detect abnormalities in heart function.
- Diagnose arrhythmias and cardiac diseases.

> The high-fidelity ECG signal forms the foundation for advanced, computer-based arrhythmia detection systems that assist clinicians.

Understanding the ECG Waveform Components

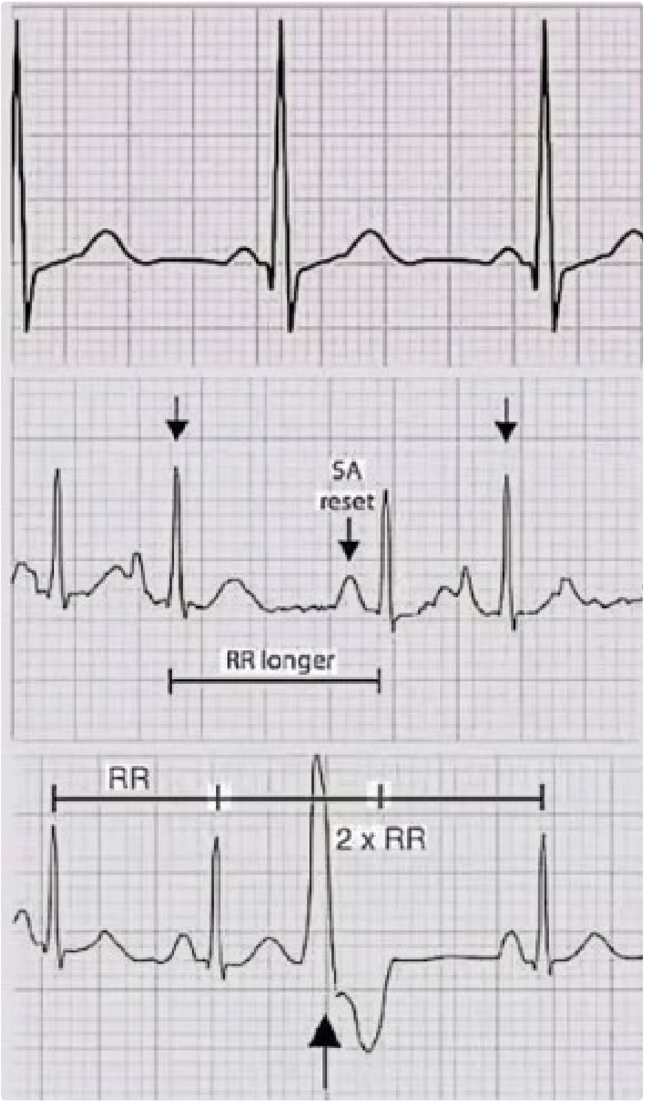
The ECG waveform is characterized by distinct waves and intervals that correspond to specific phases of the cardiac cycle. Interpreting these components is crucial for rhythm analysis.



Component	Meaning	Represents
P Wave	Atrial depolarization	Atria contract
QRS Complex	Ventricular depolarization	Ventricles contract (systole)
T Wave	Ventricular repolarization	Ventricles relax (diastole)
PR Interval	Time between atrial and ventricular activation	Electrical conduction delay
QT Interval	Total time of ventricular activity	Contraction → Relaxation cycle

Normal Rhythms vs. Arrhythmic Patterns

Differences in morphology and timing are the primary visual cues for identifying arrhythmias, forming the basis for automated classification models.



Normal Sinus Beat (N)

Characterized by consistent R-R intervals and a regular, narrow QRS complex shape. The rhythm is stable and predictable.

Key: Consistent spacing and regular morphology.

Premature Atrial Contraction (PAC)

An Arrhythmic Beat (V) that occurs earlier than expected

Visual Difference: Normal QRS complex, but appears early with an abnormal/early P wave.

Premature Ventricular Contraction (PVC)

An Arrhythmic Beat (V) that occurs earlier than expected and features a highly abnormal QRS waveform.

Visual Differences: Wider, taller, and distorted QRS shape, disrupting the regular rhythm.

Our Comprehensive Approach

We address key challenges in ECG-based arrhythmia detection through a robust methodology.

1

Data Integration

Combined MIT-BIH and Kaggle datasets for a unified corpus of 909,963 heartbeats.

2

Preprocessing

Rigorous pipeline including filtering, baseline wander removal, and feature scaling.

3

Algorithm Evaluation

Systematic assessment of eleven machine learning algorithms.

4

Class Imbalance

Addressed severe class imbalance using SMOTE.

Datasets and Preprocessing

Our unified dataset combines MIT-BIH and Kaggle data, totaling 909,963 heartbeats.

1

MIT-BIH Arrhythmia Database

Gold standard with 48 half-hour ECG recordings, expert-annotated.

2

Kaggle Classification Dataset

Augments MIT-BIH with 170,000 heartbeats and 34 pre-computed features.

3

Preprocessing Pipeline

Bandpass filtering, baseline wander removal, missing value imputation, and feature scaling.

4

R-Peak Detection

Pan-Tompkins algorithm for robust R-peak identification and heartbeat segmentation.

Class Distribution

Class	Count	Percentage	Clinical Significance
N (Normal)	719,957	98.9%	Healthy baseline rhythm
S (Supraventricular)	2,225	0.3%	Requires monitoring
V (Ventricular)	5,788	0.8%	Potentially life-threatening

Feature Extraction

We extract 40 features from each heartbeat across four domains to capture comprehensive information.

Temporal Features

10 features: amplitude and timing statistics (mean, std dev, range, energy).

Morphological Features

10 features: shape descriptors (R-peak amplitude, QRS width, area under curve).

Statistical Features

8 features: distribution characteristics (variance, skewness, kurtosis, entropy).

Frequency Features

12 features: spectral characteristics (total energy, dominant frequency, power bands).

Feature Extraction

Temporal Features

Feature	Description	Clinical Relevance
mean	Average amplitude	Overall signal strength
std	Standard deviation	Signal variability
median	Median amplitude	Central tendency
min / max	Extreme values	Peak detection
range	Max - Min	Dynamic range
ptp	Peak-to-peak	QRS amplitude
rms	Root mean square	Energy measure
energy	Sum of squares	Total signal energy
mad	Mean absolute deviation	Variability measure

Morphological Features

Feature	Description	Clinical Relevance
r_amplitude	R-peak height	Ventricular contraction strength
r_position	R-peak location (normalized)	Timing of contraction
q_amplitude	Q-wave depth	Initial depolarization
qr_amplitude	Q to R difference	Depolarization magnitude
s_amplitude	S-wave depth	Post-depolarization
rs_amplitude	R to S difference	Repolarization pattern
qrs_width	QRS complex duration	**Critical for V vs S classification**
area	Signal integral	Total electrical activity
abs_area	Absolute area	Magnitude of activity
symmetry	Left-right symmetry	Waveform regularity

Made with GAMMA

Feature Extraction

Statistical Features

Feature	Description	Clinical Relevance
variance	Signal variance	Spread of values
skewness	Asymmetry	Distribution shape
kurtosis	Tail heaviness	Outlier presence
entropy	Randomness	Signal irregularity
zero_crossing_rate	Baseline crossings	Oscillation frequency
peak_count	Number of peaks	Multi-peak detection
mean_crossing_rate	Mean value crossings	Variability measure

Frequency Features

Feature	Description	Clinical Relevance
spectral_energy	Total power	Overall frequency content
dominant_freq	Peak frequency	Primary oscillation
lf_power	Low frequency (0-5Hz)	Baseline wander
hf_power	High frequency (15-45Hz)	Muscle noise, artifacts

Machine Learning Classification

We evaluated eleven algorithms, addressing severe class imbalance with SMOTE.

1

Basic Models

Logistic Regression, Naive Bayes.

2

Intermediate Models

K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest.

3

Advanced Models

XGBoost, LightGBM.

XGBoost achieved superior performance with 99.87% accuracy without balancing.

Training Configuration

- All models were trained using a stratified 80-20 train-test split (`random_state=42`) to ensure representative datasets.
- Hyperparameters were configured with a focus on production deployment rather than exhaustive grid search, balancing performance with practical application.
- Prioritization was given to training efficiency while maintaining competitive performance across all evaluated models.
- Class weights were employed where applicable (e.g., Logistic Regression, Support Vector Machine) to inherently handle class imbalance within the models.

Evaluation Metrics

Given the class imbalance, multiple metrics are reported to ensure a comprehensive evaluation of our models' performance.

Accuracy

Overall correct classification rate

Precision

Positive predictive value per class

Recall

Sensitivity per class

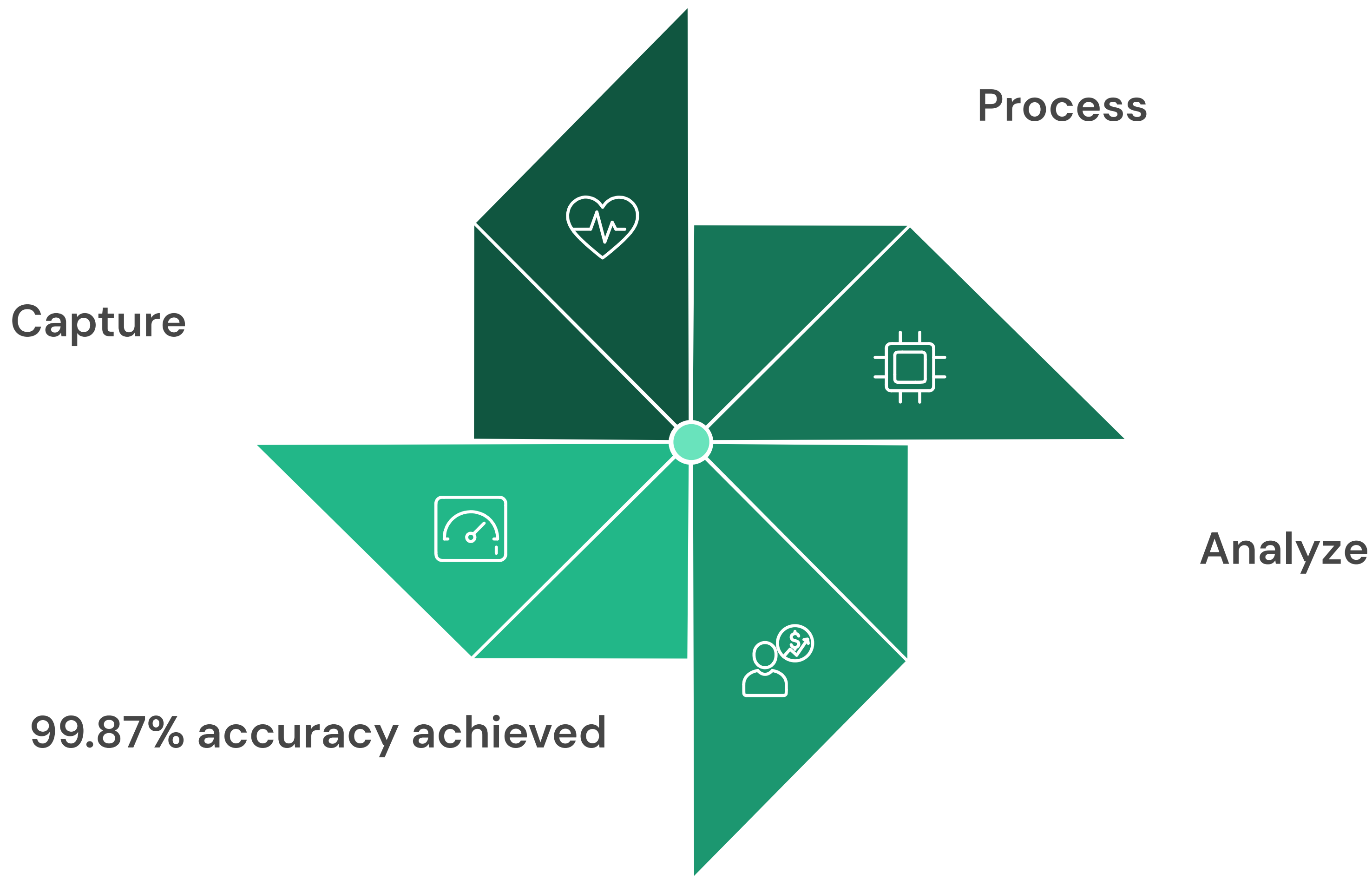
F1-score

Harmonic mean of precision and recall

Per-class performance

Individual class precision, recall and F1-score

Clinical Performance



Overall Performance Comparison

Our comprehensive evaluation of eleven algorithms reveals significant differences in performance, highlighting the strengths of advanced ensemble methods.

XGBoost: Leading Performance

Achieved **99.87% accuracy** and **99.86% F1-score** without class balancing, demonstrating superior classification capability.

Random Forest: Strong Contender

Followed closely with **99.25% accuracy**, confirming the effectiveness of ensemble learning for this task.

Naive Bayes: Lowest Performance

Exhibited the lowest accuracy at **89.97%**, confirming predictions regarding its feature independence assumption violations with ECG data.

These results showcase XGBoost's exceptional ability to accurately detect arrhythmias in raw, imbalanced datasets.

Conclusion and Future Work

XGBoost demonstrates clinical viability for automated arrhythmia screening, achieving 99.87% accuracy.

Deep Learning Integration

Explore CNN and LSTM for end-to-end learning from raw signals.

Multi-Database Validation

Evaluate generalization across diverse patient populations.

Rare Event Detection

Implement specialized techniques for minority classes.

Real-time Optimization

Adapt system for streaming ECG analysis with sub-second latency.