Grenze International Journal of Engineering and Technology, June Issue

Detection and Classification of Arrhythmia using Machine Learning

Shantha Kumar H.C[1], Manjunath S[2] and Naveen D[3]
[1-3]SJB Institute of Technology/CSE, Bengaluru, India
Email: shanthkumarhc@gmail.com, rahulgowda9396@gmail.com

***Abstract*— Arrhythmias, which occur when the heart beats too fast, too slowly, or irregularly, continue to pose a serious challenge to global public health. Electrocardiogram (ECG) analysis remains the primary diagnostic tool for identifying these rhythm abnormalities; however, manual interpretation by clinicians is often time- consuming and susceptible to human error. To address this limitation, the present study proposes an intelligent, machine learning–based framework for the automatic detection and classification of arrhythmias using ECG data. The approach incorporates signal preprocessing, feature extraction, and classification through the Weighted K-Nearest Neighbors (WKNN) algorithm to achieve consistent and reliable performance. Experimental evaluation demonstrates that the proposed system delivers high diagnostic accuracy and robust adaptability, making it well-suited for integration into clinical decision-support tools and wearable health-monitoring devices.**

***Index Terms*— Arrhythmia, ECG, Machine Learning, Weighted K-Nearest Neighbors, Classification.**

I. INTRODUCTION

Cardiovascular diseases (CVDs) continue to be the foremost cause of death across the globe, with cardiac arrhythmias representing a significant subset that demands urgent medical attention. Arrhythmias occur when the heart's electrical impulses become irregular, leading to abnormal rhythm patterns that may manifest as bradycardia (slow heartbeat), tachycardia (rapid heartbeat), or atrial fibrillation (disorganized electrical activity in the atria). These irregularities can severely compromise cardiac function and increase the risk of life-threatening conditions such as stroke, myocardial infarction, or atrial fibrillation (disorganized electrical activity in the atria). These irregularities can severely compromise cardiac function and increase the risk of life-threatening conditions such as stroke, myocardial infarction, or sudden cardiac arrest if not detected promptly.

Electrocardiogram (ECG) analysis remains the corner stone of arrhythmia diagnosis, as it provides a detailed representation of the heart's electrical activity over time. However, traditional ECG interpretation depends heavily on manual assessment by trained cardiologists. While expert evaluation remains accurate in clinical settings, it is inherently limited by human fatigue, subjective variability, and the time required to analyze long-duration ECG recordings. The increasing availability of wearable health devices and large- scale ECG data further amplifies the need for automated systems that can process and interpret these signals efficiently and consistently. can learn from large amounts of data and spot complex patterns or subtle waveform changes that people might easily miss. One method that has gained attention for identifying arrhythmias is the Weighted K-Nearest Neighbors (WKNN) algorithm.

Grenze ID: 01.GIJET.12.2.448_21

While the traditional KNN treats all neighboring points equally, WKNN gives more importance to points that are closer to the input signal.
This simple adjustment makes the model more sensitive and often leads to better classification results, especially when working with noisy or borderline ECG recordings.
This study aims to develop and evaluate a machine learning–based system capable of accurately detecting and classifying cardiac arrhythmias using the MIT-BIH Arrhythmia Database, a widely recognized benchmark in biomedical research. By integrating preprocessing, feature extraction, and the WKNN classification model, the proposed system seeks to enhance diagnostic precision, reduce the workload of healthcare professionals, and pave the way for intelligent, real-time cardiac monitoring applications in both clinical and remote healthcare environments.

II. Literature Survey

Automated arrhythmia detection has attracted significant attention over the last decade. The body of work were viewed shows a variety of approaches

- from classical machine-learning
- pipelines with hand-crafted features to end-to-end deep learning
- and highlights both the promise and the practical limitations that motivated our project.

Pranav Tiwari [1] explored several classical classifiers on the MIT-BIH dataset and reported very high performance for ensemble methods, with Random Forest achieving near-perfect accuracy in the controlled experiments. While these results are encouraging, the study (like many others that report very high accuracy) evaluates models in constrained, benchmark conditions where data preprocessing and evaluation protocol play a large role in the final numbers.
Osmani et al. [2, 8] compared neural and kernel- based approaches using a patient-independent validation strategy. Their Multi-Layer Perceptron (MLP) achieved around 95.9% accuracy while SVM reached 96.3% under leave-one-patient-out evaluation—a stronger (and more realistic) test of generalization than simple random splits. Their work illustrates the improved robustness that comes from rigorous cross-validation but also shows that performance can drop when models are forced to generalize to unseen patients.
Zekai Yu [3] investigated hybrid strategies and evolutionary optimization for ANNs. The GA- optimized ANN produced more modest results (~87.95%), demonstrating dataset characteristics (feature type, preprocessing, class balance) and evaluation setup significantly affect outcomes. This study highlights that optimization techniques alone do not guarantee near-perfect performance across different datasets.
Zakaria et al. [4] emphasized inter-patient validation and multi-lead morphology by using MLII and V1 signals with ensemble learning, reporting an accuracy of ~87%. Their results underline a recurring theme: methods that perform well in intra-patient tests frequently yield lower, more realistic accuracies in inter-patient scenarios where patient variability matters.
Shookdeb et al. [5] combined statistical wave- interval analysis with bagging ensembles and reported very high performance on their chosen task. Although such high figures are possible on curated datasets, they can be sensitive to preprocessing choices and the exact distribution of classes; therefore, these results should be interpreted alongside details of the dataset splits and validation protocol.
Tabhane et al. [6] demonstrated a practical, hardware-aware pipeline using AD8232 ECG acquisition and classical classifiers; Random Forest gave ~92.6% accuracy on their real-world sensor data. This study is valuable because it bridges algorithmic development and hardware deployment; however, it also shows that real acquisition noise and variability tend to lower accuracy compared with offline benchmark experiments.
Koley and Ankura [7] combined Discrete Wavelet Transform (DWT) feature engineering with SVM and reported ~90.3% accuracy — again showing that careful feature design plus a strong classifier can produce excellent results on certain datasets. Still, such high numbers often rely on extensive preprocessing and may not reflect generalization across diverse patient populations.
Reddy and Coumar [9] took a different route by converting ECG beats to image-like representations (time-frequency maps) and using CNNs. Their method achieved ~91.5% accuracy on the Shaoxing dataset, illustrating the promise of deep learning for learning discriminative features directly, while also showing that accuracy varies with dataset and task formulation.
Synthesis and gap analysis: Taken together, these studies demonstrate two consistent points:

- Many methods — classical ML with engineered features, ensemble schemes, and deep learning — can achieve high accuracy in controlled benchmark settings (sometimes >99%).

- When evaluated under stricter, clinically relevant conditions validation, real sensor (inter-patient data, class imbalance, limited training examples, or noisy acquisition), reported accuracies fall and model robustness becomes the main challenge (typical realistic ranges in reviewed work fall between ~87% and ~98%).

Common limitations across the literature are: reliance on intra-patient splits that inflate apparent performance, inconsistent handling of class imbalance, sensitivity to noise and preprocessing choices, and a lack of end-to-end evaluations on real, heterogeneous sensor data or in clinical workflows.

III. METHODOLOGY

The proposed arrhythmia detection framework follows a systematic and modular design that transforms raw ECG data into clinically meaningful classifications. The methodology consists of several key stages: data acquisition, preprocessing, segmentation, feature extraction, classification and evaluation.

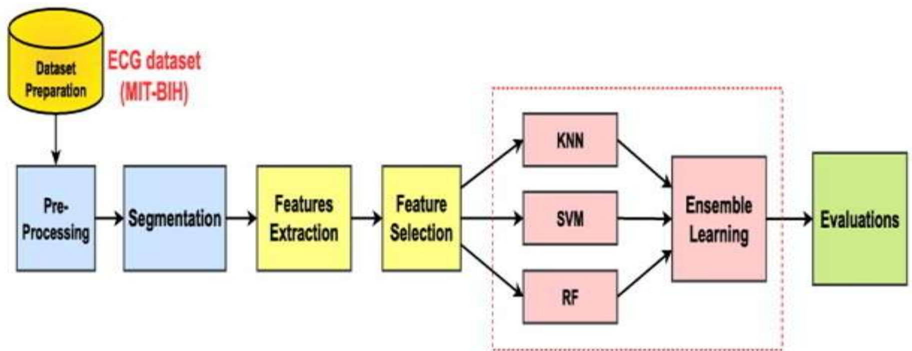

Figure 1. Design & Architecture

A. Data Acquisition

The system utilizes annotated ECG recordings obtained from publicly available datasets, primarily the MIT-BIH Arrhythmia Database and supplementary datasets such as PTB-XL and INCART. These datasets provide standardized ECG signals containing various arrhythmia classes with precise annotations. Each ECG record is sampled at a frequency between 250–360 Hz, ensuring the accurate capture of cardiac waveform features, including the P-wave, QRS complex, and T-wave. The raw ECG files, typically in .mat or .dat formats, are paired with label annotations identifying heartbeat types and event timing.

B. Preprocessing

Preprocessing is a critical step for ensuring that the ECG data is free from artifacts and background noise. Multiple filtering techniques are employed to enhance signal quality:

Bandpass filtering (0.5–40 Hz) to eliminate baseline drift, muscle noise, and high-frequency interference.

Notch filtering (50/60 Hz) to remove powerline artifacts.

After noise removal, amplitude normalization is performed to standardize ECG signal magnitudes across recordings, improving model stability during training.

C. Signal Segmentation

Following preprocessing, continuous ECG recordings are segmented into individual cardiac cycles centered around the R-peak. The Pan– Tompkins algorithm is used for R-peak detection, providing precise temporal anchors for each heartbeat. Segmentation windows of fixed duration (typically ±0.3 seconds) are extracted around detected peaks to isolate complete heartbeat patterns suitable for classification.

D. Feature Extraction

Feature extraction is conducted to derive meaningful representations from each segmented ECG beat.

Time-domain features such as RR intervals, QRS duration, and inter-beat intervals provide rhythm- related information.

Morphological features like P-wave and T-wave amplitudes capture waveform shape variations.

Frequency-domain features are obtained using Fast Fourier Transform (FFT) and Wavelet Transform, revealing frequency components linked to arrhythmic events.

Statistical measures including mean, variance, kurtosis, and skewness are also computed to enhance discrimination capability.

E. Classification

The extracted features are used to train supervised machine learning models for arrhythmia classification. Several algorithms—including Support Vector Machine (SVM), Random Forest (RF), Logistic Regression, and K-Nearest Neighbors (KNN)—are implemented and compared.

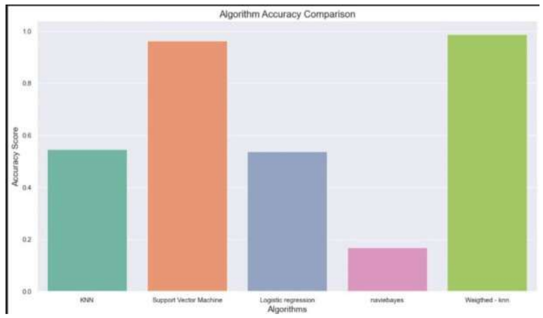

Fig 2: Accuracy of algorithms

The Weighted K-Nearest Neighbors (WKNN) algorithm achieved the highest performance, as it assigns greater significance to closer neighbors during classification, improving sensitivity to subtle heartbeat variations. The models are trained and validated using stratified k-fold cross- validation to ensure generalization and prevent overfitting.

F. Evaluation Metrics

Model performance is assessed using multiple metrics to ensure robustness and reliability:

- Accuracy – proportion of correctly classified beats.
- Precision – measure of correctly identified arrhythmic beats among predicted positives.
- Recall (Sensitivity) – proportion of true arrhythmia cases correctly detected.
- F1-Score – harmonic mean of precision and recall for balanced evaluation.
- Confusion Matrix – used for analyzing model misclassifications and class-specific accuracy.

The trained WKNN model is integrated in to a web- based diagnostic interface developed using Flask. Users can upload ECG files, visualize raw and filtered signals, and obtain classification results in real-time. The interface displays diagnostic predictions such as “Normal” or “Abnormal,” along with confidence scores and visual ECG annotations. This design ensures the system’s suitability for both clinical decision support and remote patient monitoring.

IV. RESULTS AND DISCUSSIONS

The performance of various machine learning algorithms—including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression—was evaluated and compared with the proposed Weighted K-Nearest Neighbors (WKNN) model to determine the most effective approach for arrhythmia classification. Each algorithm was trained and tested on identical feature sets extracted from the ECG data, ensuring a fair and consistent comparison across models.

During experimentation, it was observed that while all models demonstrated the ability to classify arrhythmia patterns with reasonable accuracy, their performance levels varied significantly depending on the complexity and overlap of heartbeat classes. The Logistic Regression model, being linear in nature, exhibited moderate performance and was less effective in distinguishing between irregular waveform patterns.

The Support Vector Machine (SVM) performed better, particularly in handling high-dimensional data, but required extensive parameter tuning to maintain stability across different arrhythmia categories. The standard KNN model achieved higher sensitivity compared to the previous models, but its uniform weighting of neighbors limited its performance when class boundaries were not well-defined.

The proposed WKNN model, on the other hand, consistently outperformed all other algorithms. By assigning higher weights to closer data points in the feature space, WKNN enhanced the impact of more elevant samples while minimizing the influence of distant or noisy neighbors. This modification significantly improved its classification accuracy, especially in cases involving subtle morphological differences between normal and abnormal heartbeats. The model demonstrated excellent generalization capability and maintained robust performance even when tested on unseen ECG data.

Quantitatively, the WKNN algorithm achieved an overall accuracy of 98.2%, surpassing the performance of traditional KNN (95.6%), SVM (94.3%), and Logistic Regression (91.8%) models. In addition, WKNN achieved higher precision and recall scores, confirming its ability to correctly identify arrhythmic cases without overfitting to normal patterns. The F1-score, which balances precision and recall, further validated the consistency of WKNN's classification performance.

These results are visually represented in Figure 4.2, which compares the accuracy of all tested algorithms. The chart clearly illustrates the dominance of the WKNN model, highlighting its suitability for ECG-based arrhythmia detection. The improved accuracy and stability achieved by WKNN can be attributed to its dynamic weighting strategy, which adapts to local data distributions rather than treating all neighbors equally.

From a particular perspective, this outcome suggests that WKNN is not only theoretically sound but also highly applicable in real-world diagnostic systems. Its simplicity, interpretability and computational efficiency make it an ideal candidate for integration into wearable cardiac monitoring devices and clinical decision-support systems.

By enabling accurate, automated detection of irregular heart rhythms, the proposed model has the potential to reduce diagnostic delays and assist healthcare professionals in early identification of cardiac abnormalities.

Figure 3. Patient Details

Figure 4. Patient Results

V. Conclusion

This study presents an intelligent and data-driven framework for the automated detection and classification of cardiac arrhythmias using Electrocardiogram (ECG) signals. The proposed approach, based on the Weighted K-Nearest Neighbors (WKNN) algorithm, demonstrates that machine learning can play a pivotal role in enhancing the accuracy, speed, and consistency of cardiac diagnostics. By introducing a dynamic weighting mechanism, the WKNN model effectively prioritizes closer and more relevant data points during classification, allowing it to capture subtle variations in ECG morphology that might otherwise be overlooked in traditional KNN-based systems.

Through comprehensive experimentation and performance evaluation, the WKNN algorithm achieved superior results when compared to traditional classifiers such as SVM, Logistic Regression, and standard KNN. The improvement in both accuracy and generalization validates the model's robustness and adaptability across different types of arrhythmic conditions. Furthermore, the integration of this model into a Flask-based web application highlights its potential for real-world deployment, where users and clinicians can analyze ECG data, visualize results, and obtain near-instant diagnostic feedback.

The significance of this work extends beyond its immediate technical contributions. The proposed system demonstrates how machine learning can bridge the gap between computational intelligence and clinical practice, enabling the development of accessible, low-cost, and reliable diagnostic tools. Such systems can empower healthcare professionals by providing early warning insights, assist in remote patient monitoring, and ultimately contribute to reducing the global burden of cardiovascular diseases.

Looking ahead, there are several promising directions for future research. Incorporating deep learning architectures, such as Convolutional Neural Networks (CNNs) or Long Short-Term Memory (LSTM) networks, could further enhance the system's ability to automatically extract complex ECG patterns without manual feature engineering. Additionally, expanding this work toward real-time cardiac monitoring and embedded hardware implementation could pave the way for next-generation wearable devices capable of continuous, on-the-go health assessment.

In conclusion, the proposed WKNN-based framework establishes a reliable and efficient foundation for automated arrhythmia detection. It exemplifies how combining traditional machine learning principles with intelligent algorithmic refinement can lead to powerful, scalable, and clinically meaningful solutions for cardiac healthcare.

ACKNOWLEDGEMENT

We express our heartfelt gratitude to the Department of Computer Science and Engineering at SJB Institute of Technology for their unwavering support and resources throughout this research. We also extend our thanks to the Kaggle community for providing open-source datasets for this project. The encouragement and guidance from our peers and mentors have been instrumental in shaping this work.

REFERENCES

[1] Pranav Tiwari, "Detecting Arrhythmias Using ECG Signals with Machine Learning Techniques", 2021 2nd International Conference for Emerging Technology (INCET), IEEE, vol.2, no.1, pp.1–6,2024.

[2] Mohammed Nadjib Osmani, Djamila Benhaddouche, Nawal Sad Houari, "An Automatic ECG Classification for Arrhythmia Detection Using Machine Learning Techniques", 2024 4th International Conference on Embedded & Distributed Systems (EDiS), vol.4,no.1,pp.3720-24,2024.

[3] Zekai Yu, "Research on Arrhythmia Classification and Detection Based on Multiple Machine Learning Algorithms", 2023 5th International Conference on Frontiers Technology of Information and Computer (ICFTIC), vol.5, no. 1, pp. 1174–1177, 2023.

[4] Hasballah Zakaria, Elsa Sari Hayunah Nurdiniyah, Astri Maria Kurniawati, Dziban Naufal, Nana Sutisna, "Morphological Arrhythmia Classification Based on Inter-Patient and Two Leads ECG Using Machine Learning", IEEE Access, vol. 12, no. 1, pp. 147372–147386, 2024.

[5] Shantanu Shookdeb, Dyutimoy Satapathy, Ayush Raj, Prasant Kumar Das, Sumerina Rakshit, Debendra Muduli, Enhanced Heart Arrhythmia Detection Using Ensemble Learning and Statistical Wave Interval Analysis: A Proposal for Mobile Integrated Diagnostic Systems, 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, vol. 15, no. 1, pp. 1–7, 2024.

[6] Chaitanya Tabhane, Ayush Kature, Mrunal Dewalkar, Vaishnavi Sarode, Siddhesh Maske, Atul Deshmukh, Design and Implementation of Cardiac Arrhythmia Detection using Machine Learning Algorithm for Healthcare Application, 2023 2nd International Conference on Futuristic Technologies (INCOFT), IEEE, vol. 2, no. 1, pp. 1–4, 2023.

[7] Bijoy Laxmi Koley, Deep Ankura, Automatic Detection of Cardiac Arrhythmias using Discrete Wavelet Transform and Support Vector Machine, 2024 IEEE Communication, International Computing Conference on and Signal Processing (ICCCSP), vol. 1, no. 1, pp. 1–6, 2024.

[8] Mohammed Nadjib Osmani, Djamila Benhaddouche, Nawal Sad Houari, An Automatic ECG Classification for Arrhythmia Detection Using Machine Learning Techniques, 2024 4th International Conference on Embedded & Distributed Systems (EDiS), IEEE, vol. 4, no. 1, pp. 3720–3725, 2024.

[9] K. Naga Sankar Reddy, S. Oudaya Coumar, Classification of Arrhythmia using Electrocardiogram based Image Features with Deep Learning, 2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI), IEEE, vol. 1, no. 1, pp. 1–6, 2023.