INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

ECG Pattern Algorithms for Classification and Machine Learning To Acquire Arrhythmia Identification

R Sudha[1]*, **A Nithya**[2]

1 Research Scholar , Department of Computer Science, Park's College (Autonomous), Tirupur, 05, Tamil Nadu, India
2 Research Supervisor, Department of Computer Science, Park's College (Autonomous), Tirupur, 05, Tamil Nadu, India

OPEN ACCESS

Received: 08-08-2024

Accepted: 08-05-2025

Published: 30-05-2025

Editor: Special Issue Editors: Dr. N. Pothanna, Dr.T. Jayashree, Dr. V. Ganesh Kumar

Citation: Sudha R, Nithya A (2025) ECG Pattern Algorithms for Classification and Machine Learning To Acquire Arrhythmia Identification. Indian Journal of Science and Technology 18(SP1): 56-61. https://doi.org/10.17485/IJST/v18si1.icamada36

* **Corresponding author**.

sudharjayachandran@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objective: The objective of this research is to develop & judge comparative review of machine learning methods for the classification and forecast of cardiac disorders, notably arrhythmias, gleaned from ECG signals. **Methods:** This study focused on especially appealing machine learning techniques, especially for classification algorithms and detecting arrhythmia disease through ECG data. It demands the comparison of various machine learning (classification) methods based on performance parameters like accuracy. Several ML algorithms, SVM, Navie Bayes, Gradient Boosting, Extra tree classifier, SGD classifier, Random Forest was deployed to classify arrhythmia disease. This study included 2084 patients who had suffered an acute myocardial infarction. The entire dataset can be found at https://github .com/wangsuhuai/AMI-database1.git) Whether tachyarrhythmia with atrial, ventricular, and supraventricular tachycardia occurred during admission is the main outcome. The dataset is split into two parts 67:33 (test:train). Input features are mentioned as X and labels are defined as y. The random state guarantees the split is reproducible every time the code is run. **Findings:** Accuracy for Gradient Boosting algorithm 0.89%. Significantly, other models such as Decision tree (0.5336%), Random Forest (0.5413%), Navie bayes (0.5917%), Neural Network (0.5869%) SVM (0.5828%), Extra tree classifier (0.5451%), SGD Classifier (0.5451%). Comparative analysis of various machine learning algorithms is relevant to ECG data, where the Gradient Boosting algorithm significantly achieving a remarkable accuracy of 0.8965%. The contrast with other methods such as Decision Tree, Neural Network, SVM and increased lower accuracy. **Novelty:** The novelty of this work focuses on several areas, including comparative analysis, improvement in accuracy, specific neural networks applications, and real-time detection, depending on the approaches used.

Keywords: ECG; AI in Healthcare; Arrhythmias; Neural Network; Performance Metrics

1 Introduction

Cardiovascular diseases (CVDs) remain the primary cause of human mortality worldwide, creating an urgent need for correct and early detection approaches for conditions such as arrhythmias. Although recent research has applied different deep learning and machine learning techniques, especially CNN (Convolutional Neural Networks) on datasets such as PTB-XL, MIT-BIH for ECG waveform classification, these works often have some limitations. M Any previous work frequently concentrated on only one algorithm or some specific arrhythmia types, requiring extensive, comparative analysis through diverse methods and features of ECG. Moreover, difficulties in feature selection, method illustration, and real-world applicability preserve and obstruct these methods' clinical relevance. This research work shows a novel mixed model combining DNN and FNN, focused on improving the identification of precision or accuracy for detecting arrhythmia. By addressing key areas such as feature engineering and infallibility, this work credits to the real-life use of AI in healthcare and provides clinicians with reliable tools for early diagnosis.

The recent ECG arrhythmia research may concentrate on specific models. In this circumstance, the evaluation endeavored to comprise the composed space by offering a detailed survey that compasses a composed of data-driven models. This incorporates inquiry into the various types of CVD (cardiovascular diseases), the properties of the datasets, and the data input mode[(1)]. Procedure, and the utility of Deep Learning & Machine learning algorithms, explicitly obsess on deep learning[(2)]. This appraisal also surveys the exposure of reliable AI, including justifiable, bias, and Nobel. The research explores the merits and demerits of identical algorithms, and furnishes a broad summary of cutting-edge learning that has made use of these algorithms on datasets like MIT-BIH and PTB-XL(3). Significantly, CNN (Convolutional Neural Network) indicates consistently high accuracy across both datasets. This research seems to bridge this gap by encapsulating the merits, restrictions, progress of several methods, and beneficial insights that could lead to future studies[(3)]. This literature review follows a two-stage process[(4)]. Initially, articles published in various publications from 2008 to 2023 were collected. At the time of the review, plenty of critical gaps were observed and focused on strategic foresight directions, particularly in feature engineering, the reciprocities of ECG attributes and the unrelenting depletion of ECG data for superior scrutiny. It also concentrates on the T-wave Classification[(4)].

This work found a novel blended method for the early detection of arrhythmias by combining FNN (Feedforward Neural Network) with the help of DNN (Deep Neural Network)[(5)]. The blend's model efficacy is evaluated using different metrics, like accuracy. (5). Machine learning methods were tested and trained with the help of a leave-one-out-cross validation approach, which permitted a comprehensive analysis across plenty of learning groups. The DT method, a regularly used technique in this study, has revealed significant potential in assisting doctors in finding early-stage heart illness with high accuracy. Moreover, this study involves an innovative technique that integrates various classification algorithms to classify heart rhythms (heartbeats) into abnormal, normal & COVID-19 affected categories[(6)]. This procedure works with 102 simultaneous validations, resulting in more robust and reliable findings. The beneficiary operating characteristic curve was used to evaluate the overall effectiveness of every feature, with the RF (Random Forest) model showing strong accuracy[(6)].

Furthermore, the intention of one more feature of the research was to amplify the comprehension of AEs (adverse events) correlated with Anti-arrhythmic drugs by implementing knowledge extraction techniques in the US FAERS. Asymmetrical analysis techniques were used to detect signals related to anti-arthritis drug-associated adverse events[(7)].

The current research may focus on some specific classification methods, such as Decision Tree, Random Forest, SVM, Navie Bayes, Neural Networks, Gradient Boosting Algorithm, Extra Tree Classifier, and SGD Classifier, which have been compared and revealed significant potential in assisting doctors in finding early-stage heart illness with high accuracy.

The preeminent technological discontinuity in previous ECG-based arrhythmia studies encompassed the defined aim of feature engineering, credence on the illustrative number of datasets, low focus on illustration and biases in AI models. The latest models demonstrate great accuracy in some datasets but may not perform well in different professional settings. There is also a confined revolution in mixed models drawn near that connect companion algorithms. To avoid this discontinuity, this research work proposes a novel approach by fusing FNN with DNN, enhancing feature engineering, enlarging dataset employment, as well as comparable evaluation of classification techniques to remediate arrhythmia exposure too soon and sustain clinical applications.

2 Methodology

The data set is AMI-database1 [The entire dataset can be found at https://github.com/wangsuhuai/AMI-database1.git.] The contemplative studied sick persons with relevance to MI(Myocardial Infarction), determined in the cardiovascular ICU, Harbin Medical University, during the period of 2014 - 2019. Victims received three measurable cardiac ultrasound, cardiac catheterisation, and all-day and all-night dynamic ECG within 24 a day & night of admission. Conquer is defined as regardless of tachyarrhythmia or not tachyarrhythmia occurs. The database consists of the below parameters.

There are Atrial arrhythmia, fibrillation, flutter, atrial premature, ventricular arrhythmia, tachycardia, flutter, fibrillation, frequent premature ventricular. The dataset file contains two Excel files. The file named 'Raw data' is the original file of 2084 AMI patients. One more file named 'variable Conversion' is a factual exhibit dependent on the AMI's variable conversion. The dataset can be found at the following link: https://github.com/wangsuhuai/AMI-database1.git. The novelty of this study is a comparative study between the classification algorithms for the above dataset.

There are five vital steps in the suggested ECG surf classification for arrhythmia spotting[8]. Figure[1] shows the system architecture.

Step 1: Preprocess the ECG readings by removing noise and segmenting the clear signals into individual beats[9].
Step 2: Removing the Null and invalid data.
Step 3: Extracting relevant features from the ECG signals.
Step 4: Apply classification algorithms to the extracted datasets.
Step 5: Analyse and interpret the results.

2.1 Signal Preprocessing and QRS- wave Characterization

The letters P, Q, R, S and T are commonly decrepit to represent the signals of ECG[10]. The low frequency components are emblematized by the P and T, whereas those with a higher frequency are symbolized by the QRS complex.

2.1.1 P Wave

P wave shrinking the two higher most cell of the heart[11].

2.1.2 QRS Complex

QRS defines how electrical activity shifts in the two chambers, the ventricles.

2.1.3 T Wave

This is the heart's electrical order in preparation for the next cardiac cycle.

2.1.4 ST Segment

The end of ventricular shows the ST segment. Baseline drift was removed from the ECG signals using cubic splines' interpolation, low-frequency filtering at 40 Hz eliminated electric and muscular noise, and ectopic beats were removed[12]. The filtering ECG signal was used to calculate an average heartbeat. Heartbeats with a low signal-to-noise ratio were eliminated.

2.1.5 PR Interval

The PR duration refers to the length of the first departure of the P Wave and the very first bending of the facilities of QRS signals that follows.

2.2 Using the classification methods in ECG & Health care Process

This paper examines a number of supervised machine learning methods, including K-nearest neighbors (KNN), Decision Tree, SVM, Naïve Bayes and Neural Networks. Below is a succinct explanation of these classifiers:

2.2.1 KNN (K-Nearest Neighbour's)

K's nearest neighbour classification is among the most effective, straightforward, and widely used classifiers derived from sample learning[12]. For a number of reasons, including its high perceived proficiency and the requirement to develop a hypothesis based on data, KNN classification is an easy and useful method. This assigns the test sample to the class with the topmost number of votes among the K's closest neighbours.

2.2.2 Decision Tree

DT uses a collection of determination guidelines, and divides large datasets into more manageable parts of a larger set. This produces a pattern that can predict a target variable and learn simple choice rules from input features.

2.2.3 Support Vector Machine (SVM)

A more advanced classifier built from SVM is One Vs One SVM, which can be used in multi-classification situations [13]. Since SVM's inherent properties can only determine one hyperplane, each should be trained separately. The remaining step is to make a final selection for each based on the classification outcomes of all hyperplanes using a well-known SVM training tool to identify the support vectors for each hyperplane [14].

2.2.4 Naive Bayes

Statistical classifiers known as NB classifiers are able to forecast class participation potentials, for instance, the likelihood that the specified tuple is a member of a specific class [15]. Applying Bayes classifiers to extremely large databases has also made them quite proper and fast. The premise produced by naive Bayes classifiers such an attribute's value in a particular class is unaffected by the value of the remainder of the attribute.

3 Results and Discussion

This study included 2084 patients who had suffered an acute myocardial infarction [16]. (The entire dataset can be found at https://github.com/wangsuhuai/AMI-database1.git.) Whether tachyarrhythmia with atrial, ventricular, and supraventricular tachycardia occurs during admission is the main outcome. Utilize the machine learning algorithms: Neural networks (NN), random forests (RF), and decision trees (DT), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Gradient Boosting Algorithm, Extra Tree Classifier, SGD Classifier to learn the training set and create a model. Next, assess the prediction performance using the testing set and contrast it with the model created.

The results of the comparison will likely be presented in a table. The table will show the performance of each algorithm on the different metrics.

Table 1. Performance of classification algorithms

S.NO	CLASSIFICATION ALGORITHM	ACCURACY
1	Decision Tree	0.5336
2	Random Forest	0.5413
3	SVM	0.5828
4	Navie Bayes	0.5917
5	Neural Networks	0.5869
6	Gradient Boosting Algorithm	0.8965
7	Extra tree Classifier	0.5451
8	SGD Classifier	0.5451

3.1 Comparative Analysis with Previous Literature Study

- **Performance Benchmarking**

This research fulfils the different machine learning algorithms. It includes Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), Neural Network (NN), Gradient Boosting etc., to predict arrhythmia outcomes in myocardial infarction sufferers. As stated by previous studies, various ML models have been examined in the detection of arrhythmia with varying success, frequently attaining moderate accuracy levels of around 70-80% for some specific kinds of arrhythmia. In particular, research using similar algorithms, such as SVM and Random Forest, carried out reasonable accuracy for exact ECG-based classifications but failed to reach high generalizability, highlighting the high performance of gradient boosting. In distinction, this research shows that the Gradient Boosting algorithm obtains a renowned high accuracy of 89.65%, outperforming the remaining models and completely advancing upon the accuracy range seen in comparable studies. This outcome recommends that Gradient boosting may better grasp the complications in arrhythmia data, undoubtedly its approach and competence to decrease bias and variance. The high accuracy shows potential for well-improved reliability in clinical settings.

- **Comparison with Neural Network Results**

Neural Networks attained an accuracy of 58.69% in this research work, which corresponds with past literature where standard NN have shown limitations in arrhythmia spotting accuracy due to challenges in supervision in handling noisy ECG signals.

Regardless, it contributes insights into where NN and remaining algorithms, such as Random Forest (54.13%) can be superior to ensemble models.

- **Novelty of Ensemble Approach & Divergent Model Assessment**

Non-identical to innumerable prior works that either concentrate on a single classification technique or a finite number of models, this research work significantly evaluates eight algorithms for indistinguishable optimal approaches for arrhythmia prediction. The collation reinforces the vitality of using an accumulation technique, such as gradient boosting, which can influence plenty of weak classifiers to achieve stronger predictive approaches, thus framing this research work apart from previous works.

- **Innovation and future Implications**

The research work's novelty lies in its thorough comparison of a wide range of algorithms on a remarkable dataset specific to myocardial infarction cases, emphasizing the Gradient Boosting algorithm's potential for real-time, fidelity arrhythmia identification. These analogous recommend that future research can concentrate on further refining ensemble approaches to facilitate robust and widespread applications in hospital environments.

4 Conclusion

This research work highlights Gradient Boosting as a superior model for predicting arrhythmia in acute long-suffering myocardial infraction, accomplishing a precision of 89.65%, considerably superior to models such as Neural Network, Decision Tree, and Random Forest, which ranged between 53% and 59%. This approach preceded research by an identical ensemble method that extended a powerfully built and primitive arrhythmia detection prospect.

- **Strengths and Limitations**

This study's strength is in its complete comparison of Machine Learning algorithms, but it is restricted by its training on particular datasets. Neural Networks also revealed lower precision, advising areas for upcoming development.

- **Future Directions**

Prospective work will research a customized Neural Network architecture and wide datasets to improve in advance detection and universality, helping much more diagnostic applications.

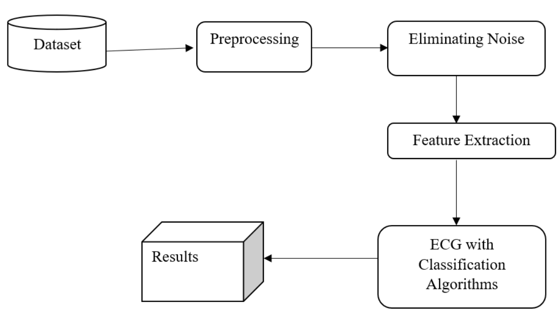

Fig 1. System Architecture

Acknowledgements

This paper has been presented in "International Conference on Applied Mathematics and Advanced Data Analytics for Industry 5.0 (ICAMADA-2024)" held during 25th to 27th April 2024 at VNR VJIET, in collaboration with APTSMS and CCOE. The authors express their sincere gratitude to the guest editors for their tireless efforts and dedication by means of comments, editing and overseeing the manuscripts to bring in this final form.

The APC is deferred partially by Indian Society for Education and Environment

References

1) Moreno-Sánchez PA, García-Isla G, Corino VDA, Vehkaoja A, Brukamp K, van Gils M, et al. ECG-based data-driven solutions for diagnosis and prognosis of cardiovascular diseases: A systematic review. *Computers in Biology and Medicine*. 2024;172. Available from: https://dx.doi.org/10.1016/j.compbiomed.2024.108235.
2) Ribeiro P, Sá J, Paiva D, Rodrigues PM. Cardiovascular Diseases Diagnosis Using an ECG Multi-Band Non-Linear Machine Learning Framework Analysis. *Bioengineering*. 2024;11(1). Available from: https://dx.doi.org/10.3390/bioengineering11010058.
3) Gour A, Gupta M, Wadhvani R, Shukla S. ECG Based Heart Disease Classification: Advancement and Review of Techniques. *Procedia Computer Science*. 2024;235:1634–1648. Available from: https://dx.doi.org/10.1016/j.procs.2024.04.155.
4) Iqbal U, Almakki R, Usman M, Altameem A, Albathan M, Jilani AK. Methodological identification of anomalies episodes in ECG streams: a systematic mapping study. *BMC Medical Research Methodology*. 2024;24(1):1–20. Available from: https://dx.doi.org/10.1186/s12874-024-02251-0.
5) Hybrid FNN-DNN Approach for Early Detection of Cardiac Arrhythmia: A Novel Framework for Enhanced Diagnosis. *VAWKUM Transactions on Computer Sciences*. 2024;12(1):49–64. Available from: https://www.vfast.org/journals/index.php/VTCS/article/view/1781/1489.
6) Banjarnahor J, Sinaga F, Sitorus DS, Sitanggang WAA, Turnip M. Application of Decision Tree Method in ECG Signal Classification For Heart Disorder Detection. *Sinkron*. 2024;8(2):1064–1072. Available from: https://dx.doi.org/10.33395/sinkron.v8i2.13596.
7) Nawaz SA, Arshad M, Aslam T, Wajid AH, Shah RA, Malik MH. ECG Based Heart Disease Diagnosis Using Machine Learning Approaches. *Journal of Computing & Biomedical Informatics*. 2024. Available from: https://www.jcbi.org/index.php/Main/article/view/338.
8) Fatimah B, Singhal A, Singh P. ECG arrhythmia detection in an inter-patient setting using Fourier decomposition and machine learning. *Medical Engineering & Physics*. 2024;124. Available from: https://dx.doi.org/10.1016/j.medengphy.2024.104102.
9) Singh VM, Saran V, Kadambi P. Autonomous Myocardial Infarction Detection from Electrocardiogram with a Multi Label Classification Approach. In: Khan E, Gönen M, et al., editors. Proceedings of Machine Learning Research;vol. 189. 2023;p. 911–937. Available from: https://proceedings.mlr.press/v189/singh23a/singh23a.pdf.
10) Appathurai A, Carol JJ, Raja C, Kumar SN, Daniel AV, Malar AJG, et al. A study on ECG signal characterization and practical implementation of some ECG characterization techniques. *Measurement*. 2019;147. Available from: https://dx.doi.org/10.1016/j.measurement.2019.02.040.
11) Kohli P. What does a an ECG for coronary artery disease show?. 2023. Available from: https://www.medicalnewstoday.com/articles/coronary-artery-disease-ecg#detect-cad.
12) Moghadas E, Rezazadeh J, Farahbakhsh R. An IoT patient monitoring based on fog computing and data mining: Cardiac arrhythmia usecase. *Internet of Things*. 2020;11. Available from: https://dx.doi.org/10.1016/j.iot.2020.100251.
13) Tseng LM, Tseng VS. Predicting Ventricular Fibrillation Through Deep Learning. *IEEE Access*. 2020;8:221886–221896. Available from: https://dx.doi.org/10.1109/access.2020.3042782.
14) Martis RJ, Acharya UR, Min LC. ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform. *Biomedical Signal Processing and Control*. 2013;8(5):437–448. Available from: https://dx.doi.org/10.1016/j.bspc.2013.01.005.
15) Padmavathi S, Ramanujam E. Naïve Bayes Classifier for ECG Abnormalities Using Multivariate Maximal Time Series Motif. *Procedia Computer Science*. 2015;47:222–228. Available from: https://dx.doi.org/10.1016/j.procs.2015.03.201.
16) Wang S, Li J, Sun L, Cai J, Wang S, Zeng L, et al. Application of machine learning to predict the occurrence of arrhythmia after acute myocardial infarction. *BMC Medical Informatics and Decision Making*. 2021;21(1):1–14. Available from: https://dx.doi.org/10.1186/s12911-021-01667-8.