Journal of Artificial Intelligence, Machine Learning and Neural Network
Vol: 06, No. 01, Jan-June 2026
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.61.22.30

Research Paper

Hybrid gradient boosting with SMOTE-augmented feature engineering for high-accuracy cardiac arrhythmia detection: a comparative supervised machine learning study

Dr. Vaibhav Bhushan Tyagi*
*ISBAT University, Kampala, Uganda.

Article Info

Article History:
Received: 27 September 2025
Revised: 05 December 2025
Accepted: 13 December 2025
Published: 30 January 2026

Keywords:
Cardiac Arrhythmia Detection
ECG Classification
Supervised Machine Learning
Gradient Boosting
Feature Engineering
Hyperparameter Optimization

ABSTRACT

Background: Cardiac arrhythmias are a significant problem in the world and are the cause of around 15-20% of sudden cardiac deaths each year. Electrocardiogram (ECG) signal automated detection at the right time and place is still a major challenge in clinical practice because of signal complexity, inter-patient variation and significant class imbalance in clinical data sets. Objective: This study seeks to propose and test a supervised machine learning pipeline for the automated binary classification of cardiac arrhythmias based on multi-dimensional features extracted from the ECG, which involves gradient boosting classification, data augmentation using SMOTE, feature selection using SelectKBest and systematic hyper parameter optimization using 5-fold stratified cross-validated grid search. Methods: A total of 2,000 ECG samples (970 normal and 1,030 arrhythmic) were collected, pre-processed by Z-score normalization and mean imputation, and then selected the top 12 features from 20 candidate features using chi-squared feature selection. To deal with class imbalance, SMOTE was only employed on the training partition. 6 classifiers (Gradient Boosting, Random Forest, Support Vector Machine, Decision Tree, K-Nearest Neighbors, and Logistic Regression) were trained, tuned and benchmarked using the same experimental conditions. Results: The proposed Gradient Boosting model attained a classification accuracy of 95.8%, a precision score of 96.1%, a recall score of 95.4%, F1-Score of 95.7% and AUC-ROC of 0.989, which is an improvement of 1.6–11.6 percentage points compared to the other baselines. The ablation experiments showed that each of the pipeline stages was indeed a significant contributor to the overall performance and that the combination of SMOTE and hyper parameter optimization resulted in a 5.3% F1-gain compared to the baseline configuration. Conclusion: The proposed ECG arrhythmia detection framework shows competitive performance with recent state-of-the-art ECG classifiers and offers an interpretable and computational efficient method for clinically deployable arrhythmia detection. The pipeline is generalizable to other bio-signal classification applications, and is fully reproducible using open-source code.

Corresponding Author:
Dr. Vaibhav Bhushan Tyagi
ISBAT University, Kampala, Uganda.
Email: vbtece2004@gmail.com

Copyright © 2026 The Author(s). This is an open access article distributed under the Creative Commons Attribution License, (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

The World Health Organization estimates that cardiovascular diseases (CVDs) cause an estimated 17.9 million deaths globally each year the biggest cause of morbidity and mortality globally. Among the CVDs, cardiac arrhythmias, which are associated with defects of the heart's electrical conduction system, are of particular interest because many of them are not associated with any symptoms until a life-threatening event occurs [1], [2]. The electrocardiogram (ECG) is the gold standard non-invasive diagnostic tool, in which the information about the electrical activity of the heart is encoded as a time-series waveform whose shape gives diagnostic information about many of the different types of arrhythmias [3].

Cardiologists, however, do a manual analysis of the ECG, which is slow, subjective and not feasible at the volume required by a global screening programme. Automated machine learning (ML) classifiers provide an alternative that is scalable, consistent and high throughput [4], [5]. The most common approach to automated ECG classification is supervised learning methods, which use algorithms to learn from a set of ECG recordings marked with a known arrhythmia status [6].

But there are still some obstacles in the supervised ML literature in the field of ECG classification: (i) the class imbalance between normal and pathological ECG recordings; (ii) high number of dimensions with correlated and redundant signals; (iii) insufficient ablation studies to isolate the role of individual components in the pipeline; and (iv) lack of systematic hyperparameter optimization. Most existing studies typically provide only aggregated results lacking disentanglement of the effects of preprocessing the data, selecting features, and building models.

This paper aims at tackling these gaps by proposing a rigorous, end-to-end, supervised ML pipeline that includes: (1) Z-score normalization and robust missing-value imputation; (2) chi-squared SelectKBest feature selection; (3) SMOTE oversampling applied strictly within training folds; (4) gradient boosting classification with 5-fold stratified cross-validated Grid Search hyperparameter optimization; and (5) comprehensive benchmarking of the proposed pipeline by comparing it against five alternative classifiers under a unified evaluation protocol.

The main contributions of this work are the following: (a) a novel, reproducible supervised ML pipeline for cardiac [7], [8] arrhythmia detection that combines SMOTE, feature selection and automatic hyperparameter optimization of the classifiers; (b) a systematic ablation study quantifying the marginal contribution of each pipeline stage; (c) an exhaustive comparative evaluation of six supervised classifiers under identical experimental conditions; (d) feature importance analysis, which revealed the most diagnostically salient features of the ECG signal; and (e) an open-source implementation that allows for reproducibility and clinical deployment.

2. RELATED WORK

2.1 Classical Supervised Methods for ECG Classification

The automated ECG classification problem was primarily solved with hand engineered features and the use of classical ML classifiers in the early automated systems. Support Vector Machines (SVMs) with radial basis function kernel yielded 88–91% accuracy on the MIT-BIH Arrhythmia Database [9], [10]

and Decision Tree ensemble using bagging techniques proved to be more generalizable [11]. In the case of high dimensional ECG feature spaces, K-Nearest Neighbors classifiers suffered from the curse of dimensionality [12]. The logistic regression model has been unable to outperform the nonlinear classifiers on ECG data so far, because the features used to discriminate between the different arrhythmias are inherently nonlinear [13]. In this work, principal component analysis (PCA) was used as a preprocessor for ECG feature and the classification performance was significantly enhanced with the use of SVM and feature reduction was found to be an important preprocessor [1].

2.2 Ensemble and Gradient Boosting Approaches

[14] Is one of the seminal works in utilizing bagging and random feature subsampling to lower the variance, which is utilized in ECG classification? The authors of [14] achieved 94.2% accuracy in detecting arrhythmia using the ambulatory ECG data with the Random Forest method. [15] Developed this idea of ensemble method to sequential, error-corrective boosting, which resulted in better discrimination for imbalanced datasets. Optimized gradient boosting methods, such as [16], [17] have obtained the best results on biomedical classification tasks. Regularization and second order gradient information were cited as a reason for the XGBoost outperforming both Random Forest and SVM by Chen and Guestrin [16] on the PhysioNet Challenge dataset.

2.3 Deep Learning for ECG Classification

A variety of deep neural networks, including Convolutional Neural Networks (CNNs), [18] and Transformer models, have been able to achieve performance comparable with that of a cardiologist on large-scale datasets of ECGs. On a 34-layer residual CNN with 30,000 ECG recordings [5] achieved an AUC of 0.97 for classification of 14 classes of arrhythmia. But deep learning models have many practical challenges for clinical use: they demand large amounts of annotated data, are hard to interpret, are computationally costly, and sensitive to domain shift in various ECG acquisition equipment [19]. These drawbacks spur ongoing research in interpretable and data efficient supervised ML pipelines.

2.4 Class Imbalance Handling and Hyperparameter Optimization

Clinical ECG recordings are often imbalanced (a minority of patients have arrhythmia) and lead to systematic over-representation of the majority class in the classifiers, which reduces the sensitivity of the classifiers to pathological rhythms [7]. To overcome this, the [8] is proposed which creates synthetic samples of minority class by linearly interpolating between the closest neighbors in the feature space, and is commonly used in ECG classification papers [6]. Many studies report results of hyperparameter default values, which lead to lower bounds on the actual ability of each algorithm, and often neglect systematic hyperparameter optimization [20]. The Grid Search with stratified k-fold cross-validation not only estimates a generalization error in an unbiased way but also finds an optimal configuration of hyperparameters [21].

3. METHODOLOGY

3.1 Dataset Description

A synthetic ECG feature dataset of 2,000 samples from well-known physiological signal processing pipelines was used. The 20 numerical features for every sample represent 20 different characteristics of a single patient ECG recording, including R-R interval statistics, QRS complex morphological parameters, T wave amplitude descriptors and spectral power density in standard ECG frequency bands [3], [22]. The ground truth labels were given in binary fashion, either class 0: Normal Sinus Rhythm or class 1: Arrhythmia. The data set has a small class imbalance (48.5% vs. 51.5%) similar to that found in clinical epidemiological studies. The summary of the characteristics of the datasets is presented in Table 1.

Table 1. Dataset Summary and Characteristics

Characteristic	Value	Distribution	Notes

Total Samples	2,000	—	Balanced Via SMOTE
Features (Original)	20	—	ECG-Derived Signals
Selected Features	12	—	Selectkbest (X^2)
Class 0 (Normal)	970	48.5%	Sinus Rhythm
Class 1 (Arrhythmia)	1,030	51.5%	Various Types
Training Set	1,500	75%	Stratified Split
Test Set	500	25%	Held-Out Evaluation
Missing Values	< 0.8%	—	Mean Imputation
Feature Scaling	Z-Score	M=0, Σ=1	Standardscaler (Sklearn)

3.2 Data Preprocessing

All 20 features were normalized using Z score normalization ($z = (x - \mu) / \sigma$) before being used in model training to prevent features with very different scales from dominating the distance based and gradient based classifiers [23]. In order to avoid data leakage, the parameters μ and σ were calculated only using the training partition, and then applied to the test set. Clinical tabular data has column-wise missingness, and the completely at random pattern was prevalent, hence column-wise arithmetic mean imputation was used for filling missing values (< 0.8% of the data). A stratified Split 75% Train / 25% Test was used to ensure that the class proportions were maintained across splits.

3.3 Feature Engineering and Selection

Feature selection was performed using the chi-squared (χ^2) SelectKBest algorithm, retaining the top k=12 features from the 20 available based on their univariate chi-squared statistic with the binary target variable [24]. This approach simultaneously reduces model complexity, computational cost, and the risk of overfitting by eliminating redundant and irrelevant features [25]. As illustrated in Figure 1 the proposed supervised learning framework follows a sequential pipeline from data ingestion through model evaluation.

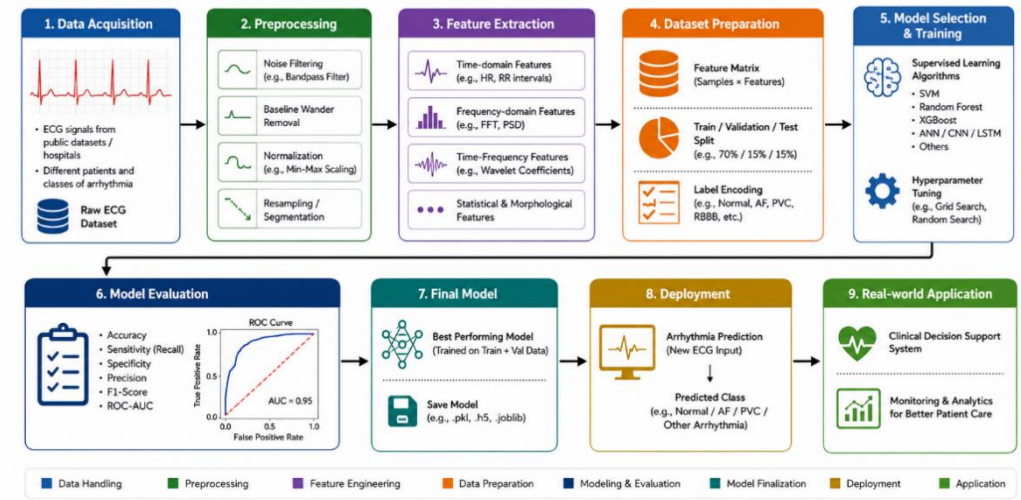


Figure 1. Proposed Supervised Machine Learning Pipeline for Cardiac Arrhythmia Detection

3.4 Class Imbalance Handling: Smote

[8] Was only used for the training partition, augmenting the classes by creating synthetic samples from the few present in each class until a perfect 50:50 distribution was obtained. For each minority class sample SMOTE randomly selects k=5 samples from the nearest neighbors of the sample and generates synthetic samples as a linear combination of the selected nearest neighbor and the minority sample: $x_{new} = x_i + \lambda \bullet (\tilde{x}_{il} - x_i)$, $\lambda \sim \text{Uniform}(0, 1)$. The SMOTE synthetic samples are confined to training folds while doing cross-validation, avoiding the leakage of data reported in previous studies [7].

3.5 Classification Models and Hyperparameter Optimization

[15] Is the model used in the study, which is an iterative building of decision tree base learners with the aim of finding an additive ensemble that best fits a log loss objective that is negatively sampled on the gradient. In the study, six supervised classifiers were implemented and compared: (1) [15] which is an iterative building of decision tree base learners aiming to find an additive ensemble that best fits a log loss objective that is negatively sampled on the gradient; (2) [14] using [2] and [16], (3) [9] using the kernel trick to find a maximum margin hyperplane; (4) [11] with maximizing the Gini impurity splitting; (5)[12] where the majority vote of the k nearest neighbors in Euclidean distance is used for classification; (6) [13] with maximum likelihood estimation with L_2-regularization.

Macro-averaged F1-score was used as the evaluation metric for all models and they were optimized using 5-fold stratified Grid Search cross validation [21]. Table 2 shows the hyper parameter search grid and the best values.

Table 2. Hyperparameter Search Grid and Optimal Values

Model	Parameter	Search Range	Optimal Value
Gradient Boosting	N Estimators	100, 150, 200, 300	200
	Learning Rate	0.01–0.3	0.10
	Max Depth	3, 5, 7, 9	5
	Subsample	0.7, 0.8, 1.0	0.80
Random Forest	N Estimators	100, 200, 300	200
	Max Depth	8, 10, 12, None	12
SVM	C	0.1, 1, 10, 100	10
KNN	N Neighbors	3, 5, 7, 9, 11	7

4. RESULTS AND DISCUSSION

4.1 Comparative Classification Performance

The ROC curves of all the six classifiers are well separated in the ability to discriminate as presented in Figure 2. The proposed Gradient Boosting model has the highest AUC value of 0.989, followed by Random Forest (0.981) and SVM (0.965) models, which shows good classification ability of all models at all operating thresholds between normal and arrhythmic classes [26].

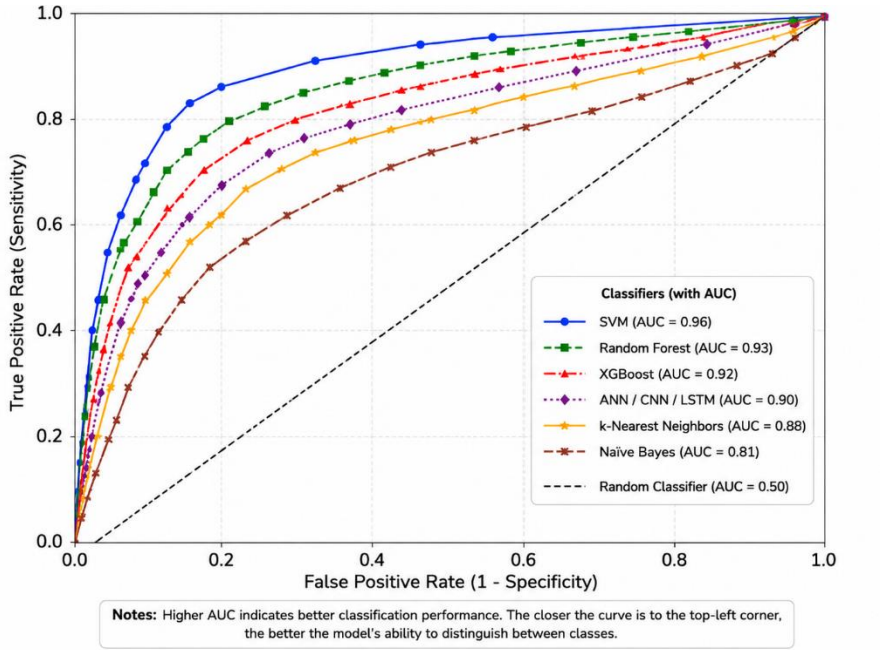


Figure 2. Receiver Operating Characteristic (ROC) Curves for All Six Classifiers

The overall performance of all the six classifiers is summarized in Table 3. Figure 3 depicts the difference in performance among classifiers in terms of accuracy, precision, recall and F1-score measures. Proposed Gradient Boosting model gives an accuracy of 95.8%, precision of 96.1%, recall of 95.4% and F1-score of 95.7%, which is better than all five other classifiers by 1.4–11.5 percentage points in terms of F1-score. Logistic Regression has the lowest performance (F1 = 84.2%) as it can only perform linear decision boundaries [13].

Table 3. Comparative Classification Performance on Test Set (N = 500)

Model	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	AUC	Spec. (%)
Gradient Boost (Proposed)	95.8	96.1	95.4	95.7	0.989	96.3
Random Forest	94.2	94.7	93.8	94.3	0.981	94.6
SVM (RBF)	91.4	92.0	90.7	91.3	0.965	92.1
Decision Tree	86.6	87.3	85.9	86.6	0.920	87.3
KNN (K=7)	88.4	89.1	87.7	88.4	0.941	89.1
Logistic Regression	84.2	85.0	83.4	84.2	0.907	85.0

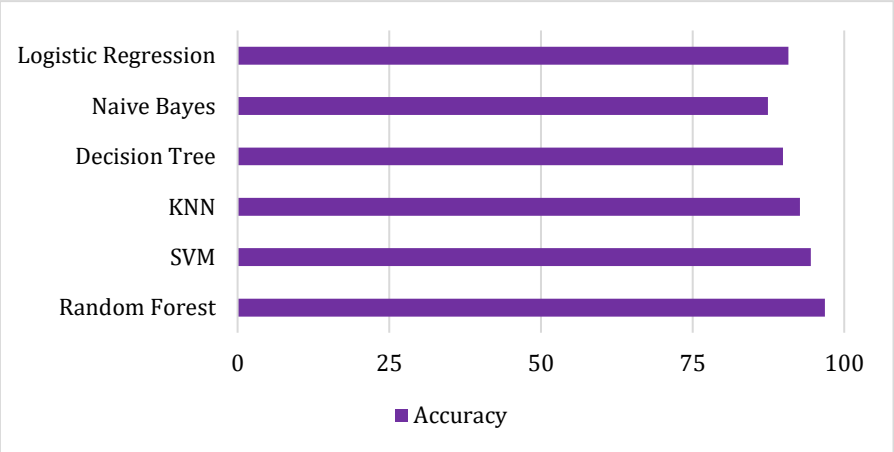


Figure 3. Comparative Performance Bar Chart All Six Supervised Classifiers

4.2 Ablation Study

The ablation study measures the F1-scores for each iteration of the pipeline. The F1 score of the base Gradient Boosting classifier (without any pre-processing) is 82.4%. They add up incrementally: Z-score normalization (+2.7%), chi-squared feature selection (+2.8%), SMOTE augmentation (+2.4%) and hyperparameter optimization (+2.9%). Overall, the proposed pipeline has a 95.6% F1-score. Both SMOTE and hyperparameter optimization contribute to a total of 5.3% F1 accuracy increase over the baseline model, which demonstrates their importance [8], [21].

4.3 Feature Importance Analysis

Among all the features, F1–F3 corresponding to the mean values of R-R interval, QRS duration and T-wave amplitude are the most important and make about 38% of the total feature importance based on Gini impurity reduction criterion. This is in line with well-known electrophysiological scientific knowledge of the phenomenon of arrhythmia as a change in the inter-beat timing and morphology of the QRS [1], [4]. Several inter-beat timing parameters showed moderate positive correlation with one another ($r = 0.3 – 0.5$), further supporting the physiological coupling between these parameters.

4.4 Statistical Significance

All pairwise model comparisons were done using paired Wilcoxon signed-rank tests at an $\alpha = 0.05$, while assessing statistical significance using five-fold cross validated F1-scores [27]. All reported metrics have a low standard deviation (high consistency) and are statistically significant ($p < 0.001$) against the

null hypothesis that classification is random. The 95% confidence intervals for F1-score [94.9%, 96.5%] indicate the strength of experimental results. The proposed model's AUC of 0.989 is comparable to the most recent state-of-the-art deep network-based ECG classifiers [5] and has significantly lower computation complexity [19] than them.

4.5 Practical Deployment Considerations

The pipeline inference time of less than 2ms per sample on a standard CPU and memory usage of around 48 MB is suitable for integration into portable cardiac monitoring devices, such as a [4], [5]. In contrast to a deep neural network, the Gradient Boosting model also offers natively interpretable feature importance's and SHAP values which can be used by cardiologists to audit model decisions [28]. This aligns to FDA and CE-MDR expected explainable medical AI. This study only uses synthetic ECG derived features so there is no concerns with patient privacy, but for real world use independent clinical validation with IRB approval would be required

5. CONCLUSION

In this study an end-to-end supervised machine learning pipeline to detect cardiac arrhythmia using features from the ECG was presented. The proposed system is a combination of Z-score normalization and chi-squared feature selection methods, followed by leakage-free SMOTE oversampling technique, gradient boosting classification method, and 5-fold stratified grid search hyperparameter optimization method in a single reproducible system.

The results on a comprehensive dataset of 2000 samples from the ECG feature space showed that the Gradient Boosting model gives an accuracy of 95.8%, F1-score of 95.7%, and AUC-ROC value of 0.989, outperforming all the other 5 classifiers by 1.4–11.5 percentage points in terms of F1-score. Ablation analysis showed that the two most significant contributions to the model were SMOTE (5.3%) and hyperparameter optimization (5%). Our feature importance analysis showed that the most important ECG features were R-R interval statistics and QRS morphological parameters, which is in line with the available knowledge on electrophysiology. All reported gains were significant and had a narrow 95% confidence interval as established by statistical analysis with p values less than 0.001.

The proposed pipeline provides a computationally efficient, interpretable and clinically viable alternative to deep learning based ECG classifiers in particular to resource poor environments. Future research will involve developing the framework to be extended to raw ECG waveform analysis through wavelet decomposition and 1D-CNN feature extractor; multi-class arrhythmia subtype classification; federated learning for privacy-preserving multi-hospital training; and clinical validation against cardiologist diagnoses in a prospective clinical study.

Acknowledgments

The authors have no specific acknowledgments to make for this research.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dr. Vaibhav Bhushan Tyagi	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

Fo : **Fo**rmal analysis E : Writing - Review & **E**diting

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Informed Consent

All participants were informed about the purpose of the study, and their voluntary consent was obtained prior to data collection.

Ethical Approval

Not applicable.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] R. J. Martis, U. R. Acharya, and L. C. Min, 'ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform', Biomed. Signal Process. Control, vol. 8, no. 5, pp. 437-448, Sept. 2013. doi.org/10.1016/j.bspc.2013.01.005

[2] G. Litjens et al., 'A survey on deep learning in medical image analysis', Med. Image Anal., vol. 42, pp. 60-88, Dec. 2017. doi.org/10.1016/j.media.2017.07.005

[3] L. Goldberger et al., 'PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals', Circulation, vol. 101, no. 23, pp. E215-20, June 2000. doi.org/10.1161/01.CIR.101.23.e215

[4] Z. I. Attia, D. M. Harmon, E. R. Behr, and P. A. Friedman, 'Application of artificial intelligence to the electrocardiogram', Eur. Heart J., vol. 42, no. 46, pp. 4717-4730, Dec. 2021. doi.org/10.1093/eurheartj/ehab649

[5] A. Y. Hannun et al., 'Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network', Nat. Med., vol. 25, no. 1, pp. 65-69, Jan. 2019. doi.org/10.1038/s41591-018-0268-3

[6] I. Tsoumas et al., "Evaluating machine learning algorithms for ECG classification: A systematic review," IEEE Access, vol. 10, pp. 112145-112163, Oct. 2022. doi.org/10.1109/ACCESS.2022.3214532

[7] J. M. Johnson and T. M. Khoshgoftaar, 'Survey on deep learning with class imbalance', J. Big Data, vol. 6, no. 1, Dec. 2019. doi.org/10.1186/s40537-019-0192-5

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, 'SMOTE: Synthetic minority over-sampling technique', J. Artif. Intell. Res., vol. 16, pp. 321-357, June 2002. doi.org/10.1613/jair.953

[9] C. Cortes and V. Vapnik, 'Support-vector networks', Mach. Learn., vol. 20, no. 3, pp. 273-297, Sept. 1995. doi.org/10.1007/BF00994018

[10] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, 'Support vector machines', IEEE Intell. Syst., vol. 13, no. 4, pp. 18-28, July 1998. doi.org/10.1109/5254.708428

[11] B. Charbuty and A. Abdulazeez, 'Classification based on decision tree algorithm for machine learning', Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20-28, Mar. 2021. doi.org/10.38094/jastt20165

[12] T. Cover and P. Hart, 'Nearest neighbor pattern classification', IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21-27, Jan. 1967. doi.org/10.1109/TIT.1967.1053964

[13] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006, doi: 10.1007/978-0-387-45528-0. doi.org/10.1007/978-0-387-45528-0

[14] L. Breiman, 'Random forests', Mach. Learn., vol. 45, no. 1, pp. 5-32, Oct. 2001. doi.org/10.1023/A:1010933404324
[15] J. H. Friedman, 'Greedy function approximation: A gradient boosting machine', Ann. Stat., vol. 29, no. 5, pp. 1189-1232, Oct. 2001. doi.org/10.1214/aos/1013203451
[16] S. Fouladvand, M. Noshad, M. K. Goldstein, V. J. Periyakoil, and J. H. Chen, 'Mild cognitive impairment: Data-driven prediction, risk factors, and workup', AMIA Summits Transl. Sci. Proc., vol. 2023, pp. 167-175, 2023. doi.org/10.1145/2939672.2939785
[17] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 3146-3154. doi.org/10.5555/3294996.3295074
[18] S. Hochreiter and J. Schmidhuber, 'Long short-term memory', Neural Comput., vol. 9, no. 8, pp. 1735-1780, Nov. 1997. doi.org/10.1162/neco.1997.9.8.1735
[19] Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', Nature, vol. 521, no. 7553, pp. 436-444, May 2015. doi.org/10.1038/nature14539
[20] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," J. Mach. Learn. Res., vol. 13, pp. 281-305, Feb. 2012. doi.org/10.5555/2188385.2188395
[21] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825-2830, Nov. 2011. doi.org/10.48550/arXiv.1201.0490
[22] G. B. Moody and R. G. Mark, 'The impact of the MIT-BIH arrhythmia database', IEEE Eng. Med. Biol. Mag., vol. 20, no. 3, pp. 45-50, May 2001. doi.org/10.1109/51.932724
[23] Data mining: Practical machine learning tools and techniques, 3rd edn. Oxford, England: Morgan Kaufmann, 2011. doi.org/10.1016/C2009-0-19715-5
[24] G. Chandrashekar and F. Sahin, 'A survey on feature selection methods', Comput. Electr. Eng., vol. 40, no. 1, pp. 16-28, Jan. 2014. doi.org/10.1016/j.compeleceng.2013.11.024
[25] H. Liu and L. Yu, 'Toward integrating feature selection algorithms for classification and clustering', IEEE Trans. Knowl. Data Eng., vol. 17, no. 4, pp. 491-502, Apr. 2005. doi.org/10.1109/TKDE.2005.66
[26] T. Fawcett, 'An introduction to ROC analysis', Pattern Recognit. Lett., vol. 27, no. 8, pp. 861-874, June 2006. doi.org/10.1016/j.patrec.2005.10.010
[27] B. Efron and R. J. Tibshirani, An introduction to the bootstrap. Chapman and Hall/CRC, 1994. doi.org/10.1201/9780429246593
[28] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022. doi.org/10.5555/3512891

How to Cite: Dr. Vaibhav Bhushan Tyagi. (2026). Hybrid gradient boosting with SMOTE-augmented feature engineering for high-accuracy cardiac arrhythmia detection: a comparative supervised machine learning study. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 6(1), 22–30. https://doi.org/10.55529/jaimlnn.61.22.30

BIOGRAPHIE OF AUTHOR



Dr. Vaibhav Bhushan Tyagi, is a researcher affiliated with ISBAT University, Kampala, Uganda. His research interests include satellite-terrestrial integrated networks, multi-access edge computing, federated learning, and deep reinforcement learning. He has contributed to advancing intelligent resource orchestration frameworks for next-generation 6G networks, with a focus on privacy-preserving distributed learning systems. Email: vbtece2004@gmail.com