Research on Electrocardiogram Abnormality Detection Method Based on Integrated Learning Models

Miaomiao Gao *

Department of Information Engineering, Harbin Institute of Petroleum University, Harbin, China

* Corresponding Author Email: 2105714318@qq.com

Abstract. For the early detection and prompt treatment of arrhythmias and other related problems, ECG anomaly detection is essential to the diagnostic process of cardiovascular illnesses. But for ECG anomaly detection applications, conventional single models frequently face drawbacks such overfitting and low robustness to abnormalities. To enhance the accuracy and robustness of ECG anomaly detection, this research emphasizes the application of ensemble learning techniques, particularly by constructing an ensemble model that employs the AdaBoost algorithm in conjunction with Random Forest as the base classifier. The experimental outcomes demonstrate that the ensemble model achieves outstanding performance in various assessment metrics, including precision, sensitivity, and F1 score. We validated our findings using the reputable MIT-BIH Arrhythmia Database. In particular, the model's overall accuracy was 99.53%, its recall rate for ventricular ectopic beats (V) was 99.57%, and its overall F1 score was 97.7%. The ensemble model suggested in this study exhibits notable benefits over conventional machine learning algorithms and deep methods of learning, offering strong technical support and a reference for clinical ECG diagnosis decision-making.

Keywords: Electrocardiogram (ECG), AdaBoost-RF Algorithm, Deep Learning, Heartbeat Classification.

1. Introduction

Globally, cardiovascular diseases have become a major health challenge. According to statistics, approximately 46.66% of the population in China will be affected by arrhythmias to varying degrees throughout their lifetime, imposing a heavy medical and economic burden on society and families [1]. In the realm of cardiovascular disease management, the Electrocardiogram (ECG) is a vital instrument, crucial for the early detection, accurate diagnosis, and effective treatment of arrhythmias and other heart conditions.[2]. In clinical practice, ECG can effectively reveal the presence of arrhythmias based on its unique waveform characteristics. Nevertheless, as medical technology progresses swiftly and the volume of medical data surges exponentially, conventional approaches for detecting ECG abnormalities have increasingly revealed their limitations, struggling to satisfy the heightened requirements for efficiency and precision in contemporary clinical diagnostics.[3].

Traditional machine learning algorithms rely heavily on artificially designed features when processing ECG data. Take the KNN algorithm as an example. When dealing with ECG data, professional personnel need to manually select features such as RR intervals and QRS complex morphology based on rich professional knowledge [4]. The method selected features are often limited by human experience and cognitive level, and may not fully and accurately reflect the complex and changeable nature of ECG signals [5]. Although deep learning models have shown strong capabilities in automatic feature extraction, they also face many challenges. Taking deep convolutional neural networks (CNN) as a typical example, their complex multi-layer convolutional and fully connected layer architectures require powerful computing resources to support training [6]. When processing large-scale ECG datasets, the training time often lasts for several hours or even days, which is a huge obstacle for clinical applications.

In view of the dilemmas of traditional methods, ensemble learning models have brought new hope and light to the field of ECG abnormality detection. The fundamental principle of ensemble learning lies in the strategic integration of strengths from multiple distinct models. By working in concert, these models can significantly bolster the overall generalization capability and mitigate the likelihood of overfitting [7]. When facing complex and changeable data sources like ECG signals, different learners can mine and analyze the features of ECG signals from their unique perspectives.

2. Current Research Status

2.1. Conventional Machine Learning Techniques

Conventional Machine Learning Techniques have achieved certain results in ECG abnormality detection. As the scale of data and the intricacy of models continue to rise, these methods are gradually reaching their limits, highlighting the need for more advanced approaches. For example,traditional Machine Learning Algorithms are inefficient in high-dimensional ECG data and have difficulty capturing key feature[8]. Support Vector Machine (SVM) segments samples by constructing hyperplanes. When dealing with large-scale ECG data, it has high computational costs and lacks a unified standard for kernel function selection [9]. Decision trees have been used for false alarm detection in ICUs, but their generalization ability is limited [10].

2.2. Deep learning methods

Convolutional neural networks (CNNs) extract ECG features through convolutional layers and pooling layers, and are good at capturing spatial features. Wang et al. [11] proposed an elevated CNN-based method for automatic heartbeat classification. Chen et al.Automatic arrhythmia classification was achieved in [12] through the integration of CNN and LSTM networks. LSTM and its variants focus on time series data, making them suitable for long-term monitoring, but they have low interpretability and require a large amount of data and parameter optimization [13]. BERT faces adaptation and accuracy challenges in ECG analysis [14]. In addition, deep learning methods also involve some well-known issues, such as low interpretability, the need for a more extensive dataset, and the need to optimize a large number of parameters.

2.3. Ensemble learning methods

Ensemble learning enhances model efficacy, mitigates overfitting, and bolsters robustness through the integration of several base models. Bagging enhances model accuracy by minimizing variance [15]. As reported in [16], Dogan and colleagues introduced a weighted ensemble voting approach for classification tasks. Kiranyaz et al.[17]utilized 1-D convolutional neural networks to achieve real-time, patient-specific ECG classification, highlighting the potential of combining deep learning techniques with ensemble approaches to enhance classification accuracy. Ghiasi and others [18] used decision tree models to diagnose coronary artery disease, further illustrating the benefits of ensemble learning when managing intricate medical data. These studies offer significant guidance and inspiration for our research, suggesting that by amalgamating diverse learning techniques, the accuracy of cardiac anomaly detection can be substantially enhanced.

3. Method

3.1. Introduction to the dataset.

This study makes use of the well-known MIT-BIH Arrhythmia Database, which was created in collaboration between Beth Israel Hospital (BIH) and the MIT Laboratory for Computational Physiology. Since the 1970s, this public dataset has been a standard in the discipline of ECG data processing. It includes ECG data from 47 patients and consists of 48 records, each monitored for about 30 minutes at an average rate of sampling of 360 Hz. This experiment extracted 109,454 heartbeat feature data points. The dataset was divided into three portions: 70% allocated for the training set, 10% for the validation set, and the remaining 20% for the test set.

While Table 1 summarizes the heartbeat statistics. Class N account for 82.77% of the total, whereas abnormal types—class S, class V,class F, and class Q—constitute only 17.23% combined, with proportions of 2.54%, 6.61%, 0.73%, and 7.35%, respectively.

Table 1. Statistics of Experimental Data

Type	Training Set	Validation Set	Test Set	Number of Beats
N	63417	9059	18119	90,595
S	1947	278	556	2,781
V	5065	723	1447	7,235
F	562	80	160	802
Q	5629	804	1608	8,041
Total	76620	10944	21890	109,454

3.2. Construction of Ensemble Learning Models

In the in-depth exploration of ECG anomaly detection technology based on ensemble learning models, this study fully considers the complexity and high variability of ECG signals, as well as the potential limitations of single models in tackling such complex tasks. In light of this, we have carefully constructed the methodological framework for this study.

3.2.1. Selection of Base Classifiers

This study decided to adopt Random Forest as the base classifier in the AdaBoost algorithm. Random Forest only utilizes a subset of features when constructing each decision tree, a characteristic that makes the model insensitive to variations in individual features, thereby greatly enhancing the model's robustness. It is these advantages that make Random Forest an ideal choice for ECG anomaly detection tasks.

3.2.2. Design of Ensemble Strategy

The core concept of ensemble learning lies in combining multiple classifiers with relatively average performance into a stronger classifier with superior performance through specific ensemble strategies. Random Forest is an optimization of the Bagging algorithm and adopts CART (Classification and Regression Trees) decision trees as its base classifiers. CART decision trees are highly regarded for their ability to effectively prevent overfitting, their strong generalization capabilities, and their insensitivity to feature variations [19].

The Gini coefficient is an indicator used to measure the uncertainty of a sample set. In classification problems, if the sample set D contains N categories and the probability that a sample point belongs to the K-th category is Pk, then the probability that the sample point is misclassified is $1-P_k$. The Gini coefficient is defined as follows:

$$\mathrm{Gini(P)} = \sum_{K=1}^{\mathrm{K}} p_K(1-p_K) = 1-\sum_{K=1}^{K} p_K^2 \tag{1}$$

For a given sample set D, let C_k be the subset of D that belongs to the k-th class, and let K be the number of classes. The Gini coefficient for the sample set D is calculated as follows:

$$\mathrm{Gini(D)} = 1-\sum_{K=1}^{\mathrm{K}} \left(\frac{|\mathrm{C}_K|}{|\mathrm{D}|}\right)^2 \tag{2}$$

If the sample set D is split into two parts D_1 and D_2 based on whether the feature A takes a certain possible value a, is denoted as $D_1 = \{(x, y) \in D \mid A(x) = a\},\ D_2 = D - D_1$. the Gini coefficient of the set D under the condition of feature A is defined as:

$$\text{Gini}(D,A)=\frac{|D_1|}{|D|}Gini(D_1)+\frac{|D_2|}{|D|}Gini(D_2) \tag{3}$$

The Gini coefficient, Gini(D), is used to measure the degree of uncertainty in a set D, while Gini (D, A) represents the uncertainty obtained after splitting the set D based on feature A.

To formally define the AdaBoost adaptive boosting model, assume the training dataset is $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$, where each sample consists of an instance and a label. The instance $x_i \in X \subseteq R^n$, and the label $y_i \in Y$, with X being the instance space and Y being the set of labels. The algorithm is defined as follows:

(1) Initialize the weight distribution of the training data,

$$D_1=\left(w_{11},\cdots,w_{1i},\cdots,w_{1N}\right),\quad w_{1i}=\frac{1}{N},\quad i=1,2,\cdots,N \tag{4}$$

(2) AdaBoost iteratively learns base classifiers, performing the following operations in each round $m=1,2,3....M$ in sequence:

(a) Learn using the training dataset with weight distribution D_m to obtain the base classifier $G_m(x)$;

$$G_m(x): X \rightarrow Y \tag{5}$$

(b) Calculate the classification error rate of $G_m(x)$ on the training set,

$$e_m=P\left(G_m(x_i)\neq y_i\right)=\sum_{i=1}^{N}w_{mi}I\left(G_m(x_i)\neq y_i\right) \tag{6}$$

In the exploration of decision tree algorithms in the field of machine learning, we introduce the variable w_{mi} to represent the weight of the i-th sample in the m-th iteration. These weights play a crucial role in the weighted training dataset, as the classification error rate of the decision tree is directly reflected by the sum of the weights of the misclassified samples. It is worth noting that the total sum of all sample weights is normalized to 1 (*i.e.*, $\sum w_{mi}=1$), which ensures that the calculation of the error rate is based on a standardized weight distribution.

(c) Calculate the coefficient α_m of the base classifier $G_m(x)$,

$$\alpha_m=\frac{1}{2}\log\frac{1-e_m}{e_m} \tag{7}$$

From this formula, it can be known that when $e_m \leq \frac{1}{2}$, $\alpha_m \geq 0$, and α_m increases as e_m decreases.

(d) Update the weight distribution of the training dataset in preparation for the next round

$$D_{m+1}=\left(w_{m+1,1},\text{L}\ ,w_{m+1,i},\text{L}\ ,w_{m+1,N}\right) \tag{8}$$

$$w_{m+1,i}=\frac{w_{mi}}{Z_m}\exp\left(-\alpha_m y_i G_m(x_i)\right),\quad i=1,2,\text{L}\ ,N \tag{9}$$

$$Z_m=\sum_{i=1}^{N}w_{mi}\exp\left(-\alpha_m y_i G_m(x_i)\right) \tag{10}$$

Here, Z_m is the normalization factor that makes D_{m+1} a probability distribution. It can be seen that the weights of samples misclassified by the base classifier $G_m(x)$ are increased, while the weights of correctly classified samples are decreased. Comparing the two, the weights of misclassified samples are amplified by a factor of $e^{2\alpha_m} = \frac{e_m}{1-e_m}$

(3) Construct the linear combination of base classifiers

$$f(x) = \sum_{m=1}^{M} \alpha_m G_m(x) \tag{11}$$

Obtain the final classifier

$$G(x) = \mathrm{sign}(f(x)) = \mathrm{sign}\left(\sum_{m=1}^{M} \alpha_m G_m(x)\right) \tag{12}$$

$f(x)$ implements a weighted vote of M base classifiers, where the coefficient α_m represents the importance of the base classifier $G_m(x)$.

3.2.3. Settings of Model Parameters

In this study, we set the following key parameters: the number of base classifiers is set to 100 decision trees, which ensures the performance of the model while avoiding overfitting; the learning rate is configured at 0.1, a relatively small value that helps the model learn more information in each round of training and avoids instability caused by excessive updates; the proportion of feature selection is set to 70%, a balanced ratio that ensures the diversity of the model while preventing information loss due to too few feature selections.

4. Presentation and Interpretation of Experimental Findings

4.1. Experimental Environment

We detail the experimental results of the model in detecting arrhythmia rhythm categories in ECG signals. All experiments were conducted using the PyTorch 2.3.0 framework to build the model, with the programming environment being Python 3.12. The system specifications included two NVIDIA RTX 3080 GPUs, each with 20GB of memory, and a 12-core Intel Xeon Platinum 8352V CPU with 48GB of RAM. The primary libraries used for executing the model included: numpy, pandas, matplotlib, pytorch, etc.

4.2. Evaluation Metrics

In the performance evaluation of machine learning models, we adopt accuracy (Acc), sensitivity (Se), specificity (Spe), and F1-score as measurement standards. The evaluation metrics are shown in the following formulas:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

$$Recall = Sen = \frac{TP}{TP+FN} \tag{14}$$

$$Spe = \frac{TN}{TN+FP} \tag{15}$$

$$Ppv = Pre = \frac{TP}{TP+FP} \tag{16}$$

$$F1 = 2 \times \frac{Sen \times Pre}{Sen+Pre} = \frac{2TP}{2TP+FP+FN} \tag{17}$$

Here, TP stands for true positives, FP for false positives, TN for true negatives, and FN for false negatives.

We adopt the following four main indicators: Accuracy, Recall (also known as Sensitivity), Precision, and F1 Score.

4.3. Experimental Results

The detailed experimental results are presented in Table 2. A review of the table reveals that the ensemble learning model significantly outperforms both traditional machine learning and deep learning methods across all performance metrics.Particularly notable are the recall rate and F1 score, where the ensemble learning model achieved excellent results of 99.57% and 97.7% respectively, demonstrating a significant difference compared to other models. This fully proves that the ensemble learning model has higher accuracy and stronger robustness in the field of ECG anomaly detection.

Table 2. Comparison of Performance Results with Different Models

Model	Accuracy	Recall	Precision	F1 Score
Support Vector Machine (SVM) Model	91.51%	90.29%	91.45%	62.88%
Convolutional Neural Network (CNN) Model	98.25%	98.33%	98.20%	90.78%
AdaBoost-RF Ensemble Learning Model	99.53%	99.57%	99.55%	97.7%

Compared to conventional machine learning models, ensemble learning models significantly elevate the generalization and stability of the model by leveraging the strengths of multiple weak classifiers. In contrast to deep learning approaches, while deep learning models excel at automatic feature extraction, their complex network structures and substantial computational resource requirements limit their practicality in clinical settings. Our proposed ensemble learning model, based on the AdaBoost algorithm combined with Random Forest, demonstrates notable advantages in the field of electrocardiogram (ECG) anomaly detection, as shown in Table 3.

When tested on the MIT-BIH Arrhythmia Database, the AdaBoost-RF model achieved an overall accuracy of 99.53%. Especially in the detection of V, its recall rate even reached 99.57%, and the overall F1 score also attained a high level of 97.7%.

Table 3. Classification Performance of the AdaBoost-RF Algorithm Model

		Evaluation Criteria (%)		
	Se	***Sp***	**+*p***	***Acc***
N	99.03	92.41	97.71	98.42
S	90.29	99.51	95.98	98.33
V	99.42	99.20	97.70	99.62
F	90.70	99.94	98.28	99.40
Q	98.59	99.93	99.80	99.84

As illustrated in Table 4. The results of this comparison indicate that the AdaBoost-RF model excels across all metrics, achieving superior classification accuracy.

Table 4. Comparison with Classification Methods of Other Researchers

Algorithm	Feature Relationships	Performance
BiLSTM-CNN [20]	Morphology of Heartbeats	$Acc = 99.54\%$ $Sen = 99.87\%; Spn = 98.54\%; +Pn = 99.69\%$ $Ses = 95.58\%; Sps = 99.93\%; +Ps = 97.57\%$ $Sev = 98.63\%; Spv = 99.92\%; +Pv = 98.90\%$ $Sef = 89.02\%; Spf = 99.94\%; +Pf = 92.41\%$
CNN-BiLSTM [21]	Morphological Features	$Acc = 97{,}29\%$ $Sen = 98.57\%; Spn = 93.62\%; +Pn = 98.81\%$ $Ses = 92.00\%; Sps = 95.81\%; +Ps = 96.80\%$ $Sev = 95.81\%; Spv = 96.80\%; +Pv = 96.80\%$ $Sef = 80.55\%; Spf = 85.22\%;$
AdaBoost-RF	Morphology of Heartbeats	$Acc = 98.91\%$ $Sen = 99.03\%; Spn = 92.41\%; +Pn = 97.71\%$ $Ses = 90.29\%; Sps = 99.51\%; +Ps = 95.98\%$ $Sev = 99.42\%; Spv = 99.20\%; +Pv = 97.70\%$ $Sef = 90.70\%; Spf = 99.94\%; +Pf = 98.28\%$

5. Summary

This study successfully devised and validated an ensemble learning model for electrocardiogram (ECG) anomaly detection, incorporating the AdaBoost algorithm with Random Forest. Utilizing the MIT-BIH Arrhythmia Database, our method exhibited remarkable enhancements in key performance indicators, including accuracy, recall, and F1 score. Precisely, we achieved an overall accuracy of 99.53%, a recall rate of 99.57% for ventricular ectopic beats (V), and an overall F1 score of 97.7%, showcasing the model's exceptional ability to accurately pinpoint ECG anomalies. The primary contributions of our research are outlined below:

1. We introduce an innovative ensemble learning framework that adeptly addresses the shortcomings of traditional single models in ECG anomaly detection, such as susceptibility to overfitting and limited robustness against anomalies.
2. Our experimental results underscore the model's substantial superiority over conventional machine learning techniques and deep learning methodologies, particularly in managing class imbalance and maintaining consistent high performance.
3. This study offers a dependable technical solution and a valuable decision-support tool for clinical ECG diagnosis, ultimately facilitating more efficient and accurate diagnostic processes.

References

[1] Wang Z W, Hu S S. Key Points Interpretation of "China Cardiovascular Health and Disease Report 2019"[J]. Chinese Journal of Cardiology, 2020, 25 (05): 401 - 410.

[2] Y. Birnbaum, K. Nikus, P. Kligfield, M. Fiol, J. A. Barrabés, A. Sionis, O. Pahlm, J. G. Niebla, and A. B. de Luna, ‘‘The role of the ECG in diagnosis, risk estimation, and catheterization laboratory activation in patients with acute coronary syndromes: A consensus document,’’ Ann. Noninvasive Electricidal., vol. 19, no. 5, pp. 412 – 425, Sep. 2014.

[3] Li, Hui, et al. "Clinical knowledge-based ECG abnormalities detection using dual-view CNN-Transformer and external attention mechanism." Computers in Biology and Medicine 178 (2024): 108751.

[4] Saini, Indu, Dilbag Singh, and Arun Khosla. "QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases." Journal of advanced research 4.4 (2013): 331 - 344.

[5] Pham, Huy, et al. "Machine learning-based detection of cardiovascular disease using ECG signals: performance vs. complexity." Frontiers in Cardiovascular Medicine 10 (2023): 1229743.

[6] Liu, X., Wang, H., Li, Z., & Qin, L. (2021). Deep learning in ECG diagnosis: A review. Knowledge-Based Systems, 227, 107187.

[7] Singh, Anupreet Kaur, and Sridhar Krishnan. "ECG signal feature extraction trends in methods and applications." BioMedical Engineering OnLine 22.1 (2023): 22.

[8] Indu Saini, Dilbag Singh, Arun Khosla, QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases, Journal of Advanced Research, Volume 4, Issue 4, 2013, Pages 331 - 344, ISSN 2090 - 1232.

[9] Hamza, Sihem, and Yassine Ben Ayed. "Svm for human identification using the ECG signal." Procedia computer science 176 (2020): 430 - 439.

[10] Manna, Tishya, Aleena Sweta Padma, and Moloud Abdar. “Decision Tree Predictive Learner-Based Approach for False Alarm Detection in ICU.” Journal of Medical Systems 43, no. 7 (May 21, 2019): 191.

[11] Wang, Haoren, Haotian Shi, Xiaojun Chen, Liqun Zhao, Yixiang Huang, and Chengliang Liu. “An Improved Convolutional Neural Network Based Approach for Automated Heartbeat Classification.” Journal of Medical Systems 44, no. 2 (December 18, 2019): 35.

[12] Chen Chen, Zhengchun Hua, Ruiqi Zhang, Guangyuan Liu, and Wanhui Wen. “Automated Arrhythmia Classification Based on a Combination Network of CNN and LSTM.” Biomedical Signal Processing and Control 57 (March 1, 2020): 101819.

[13] Karim, F., Majumdar, S., Darabi, H., & Harford, S. (2019). Multivariate LSTM-FCNs for time series classification. Neural networks, 116, 237 - 245.

[14] Hannun, Awni Y., Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H. Tison, Codie Bourn, Mintu P. Turakhia, and Andrew Y. Ng. “Cardiologist-Level Arrhythmia Detection and Classification in Ambulatory Electrocardiograms Using a Deep Neural Network.” Nature Medicine 25, no. 1 (January 1, 2019): 65 – 69.

[15] Sexton, Joseph, and Petter Laake. "Standard errors for bagged and random forest estimators." Computational Statistics & Data Analysis 53.3 (2009): 801 - 811.

[16] Dogan, Alican, and Derya Birant. "A weighted majority voting ensemble approach for classification." 2019 4th International Conference on Computer Science and Engineering (UBMK). IEEE, 2019.

[17] S. Kiranyaz, T. Ince and M. Gabbouj, "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks," in IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 664 - 675, March 2016.

[18] Ghiasi, Mohammad M., Sohrab Zendehboudi, and Ali Asghar Mohseni pour. "Decision tree-based diagnosis of coronary artery disease: CART model." Computer methods and programs in biomedicine 192 (2020): 105400.

[19] Loh, W. Y. (2011). Classification and regression trees. Wiley interdisciplinary reviews: data mining and knowledge discovery, 1 (1), 14 - 23.

[20] Chen, C., Hua, Z., Zhang, R., Liu, G., & Wen, W. (2020). Automated arrhythmia classification based on a combination network of CNN and LSTM. Biomedical Signal Processing and Control, 57, 101819.

[21] Xu, X., Jeong, S., & Li, J. (2020). Interpretation of electrocardiogram (ECG) rhythm by combined CNN and BiLSTM. Ieee Access, 8, 125380 - 125388.