Interpretable XGBoost-SHAP Model for Arrhythmic Heartbeat Classification

Rong Xiao[1], Meicheng Yang[1], Caiyun Ma [1], Lina Zhao[1], Jianqing Li[1], Chengyu Liu[1*]

[1]State Key Laboratory of Digital Medical Engineering, School of Instrument Science and Engineering, Southeast University, Nanjing, China

Abstract

With the rising incidence of cardiovascular diseases, efficient arrhythmia detection from Electrocardiogram signals is crucial. Traditional Machine Learning and Deep Learning methods for automatic ECG analysis have demonstrated promising results with high accuracy but often suffer from a lack of interpretability and challenges in correctly identifying similar beats and detecting sporadic abnormalities. We propose a novel heartbeat arrhythmia classification model that combines eXtreme Gradient Boosting with Shapley Additive explanations. Our model extracted features from single heartbeat, emphasizing the characteristics crucial for differentiating between similar heartbeat. By addressing data imbalance and selecting features, we achieved 98.91% accuracy in the 5-class classification problem devoted by the Association for the Advancement of Medical Instrumentation. The contribution of our study is to demonstrate the feasibility of directly classifying arrhythmias from single heartbeats, and incorporate interpretability into the heartbeat classification process, enabling clinicians to better understand model decision.

1. Introduction

The increasing prevalence of cardiovascular diseases, particularly arrhythmia is a growing concern, estimating it as the top mortality cause by 2035 [1]. Advancing the diagnosis and management of arrhythmia is crucial. ElectroCardiograms (ECGs), which capture the heart's electrical activity, are essential for detecting cardiac issues. Hence, the pursuit of precise, automated heartbeat classification techniques is of paramount importance in the medical field. Competitions like 2017 PhysioNet/CinC Challenge, the 2018 China Physiological Signal Challenge, and the PhysioNet Challenge in 2020 have galvanized research into automated arrhythmia detection. Each aims to refine Machine Learning (ML) and Deep Learning (DL) for accurately identifying cardiac abnormalities from ECG recordings.

The key challenges [2] in developing effective arrhythmia detection systems is the balance between interpretable feature selection and classification algorithms with high accuracy. Some studies based DL have achieved impressive accuracies, such as using DNNs [3], CNN [4], LSTM [5] for classifying arrhythmia, achieving an overall accuracy above 99%, but suffer from poor interpretability and parameter redundancy. These techniques often appear as 'black boxes,' reducing their clinical applicability. There are some approaches focus on interpretable features but are not as accurate as the previous methods. The use of heart rate variability features with support vector machines for classifying heart failure and normal sinus rhythm is proved effective [6]. And some approaches rely on entropy-based features for atrial fibrillation [7][8].

Additionally, many arrhythmia detection methods utilize segment ECG signals [3-5], advantageous for continuous monitoring and demonstrates good recognition of persistent and intermittent arrhythmias. However, it may miss sporadic arrhythmias occurring in small quantities. Heartbeat-based classification models initially segment ECG signals into single beats directing the model to focus on the morphological characteristics of each beat, mitigating the issue of missed detections. It enables precise discrimination between heartbeats with similar rhythms yet distinct morphologies, such as atrial flutter and atrial fibrillation, SVEB and VEB, enhancing the detailed and accurate diagnosis of arrhythmias.

In this study, we propose a heartbeat arrhythmia classification model using eXtreme Gradient Boosting (XGBoost) and Shapely Value (SHAP), offering feature interpretability. The extraced features of single beat we used included temporal, spectral, and statistical features. After addressing data imbalance through Synthetic Minority Over Sampling Technique (SMOTE) [9] reducing the feature dimension based on SHAP feature importance evaluation, the optimal feature subset achieved 98.91% accuracy. We analyzed the interpretability of classification features and explored classification methods based on single beat characteristics, indicating its significance for accurate detection of diseased heartbeats.

2. Methods

The methodology proposed encompasses a series of

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.186

sequential stages. First, the datasets utilized for the training and validation of the approach is described. Subsequently, the details of data preprocessing, feature extraction and selection, and the classification methods are given, respectively. The ECG signal represents the input and NSVFQ represents the output class. An overview of the proposed approach is presented in Figure 1.

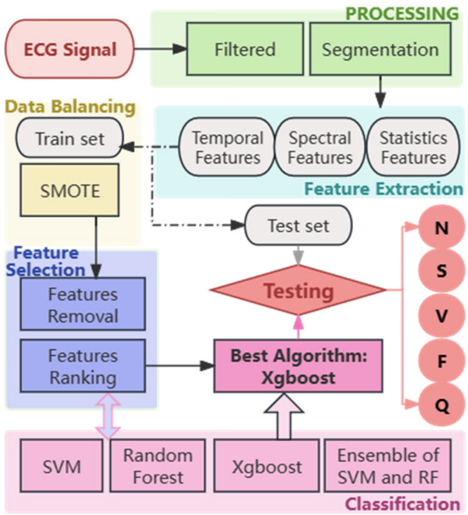

Fig 1. The framework of the proposed method

2.1. Database

ECG data is from the MIT-BIH arrhythmia database[10], consisting of 48 half-hour ECG recordings, sampled at 360 Hz, featuring a resolution of 11 bits and covering a voltage range of 10 mV. The original dataset comprises 41 heartbeat labels, with 19 distinct categories representing R-peak information. The basis for arrhythmia classified proposed by the Association for the Advancement of Medical Instrumentation (AAMI)[11], 15 recommended classes into 5 superclasses: Normal (N), Supraventricular ectopic beat (SVEB), Ventricular ectopic beat (VEB), Fusion beat (F) and Unknown beat (Q). Following the "subject-specific"[9] our approach similarly partitions the dataset into DS1 and DS2 for training and testing, respectively.

2.2. Processing

Data preprocessing entails noise reduction and heartbeat segmentation. A Butterworth bandpass filter is utilized to eliminate power frequency interference and baseline wander with minimal computational overhead. The signals are then segmented into heartbeats using the R peak annotations from the MIT-BIH dataset, with each heartbeat represented by a fixed length of 260 sampling points centered around the R-peak..

Table 1. The numbers of heartbeat segmentation

	N	*S*	*V*	*F*	*Q*	*Total*
DS1	59k	1668	4341	481	19	66k
DS2	39k	1113	2894	321	14	44k
All	98k	2781	7235	802	33	110k

2.3. Feature Extraction

ECG signals benefit from a variety of specialized software packages designed for efficient feature extraction. TSFEL[12] has a rapid feature analysis property, which analyses one feature in 0.1ms over the whole dataset[13]. It supports extraction of time-domain, spectral, and statistical features. Our methodology posits that these features are inherently adequate to encapsulate the morphological nuances of ECG signals.

2.4. Dataset Balancing with SMOTE

Data imbalance affects the performance of the classifier on minority classes. SMOTE[9] addresses this issue by synthesizing new minority instances between existing minority instances. The process of can be expressed as:

$$X_{\text{new}} = X + rand(0,1) \times \left(\tilde{X} - X\right)$$

Our approach oversamples the minority categories (S, V, F, Q) in the DS1 training set, equalizing their sample size with the majority category (N). The impact of data volume balancing is illustrated in Figure 2, demonstrating the effective enlargement of the decision region for minority categories. Q-class samples, their scarcity and dispersed distribution result in the over-sampling process further dispersing the augmented class regions. Therefore, we consider it useful to simultaneously perform under-sampling on the majority classes.

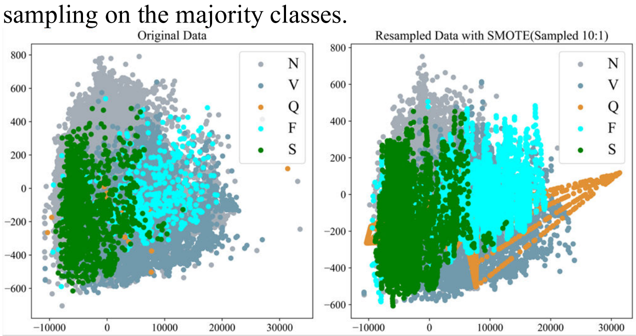

Fig 2. Balancing the minority class samples using SMOTE

2.4. Feature Selection

In our approach, after normalizing the features to a uniform scale, we employed the following feature selection techniques for reducing dimensionality:

A. *Feature Removal:* For preliminary feature set screening, we eliminate features with variances below a set threshold (threshold=0) and those with high correlations(threshold=0.9), correlation defined with Pearson correlation coefficient.

B.Recursive Feature Elimination(RFE): It has been reported that the features extracted by TSFEL contain a large amount of redundancy[13]. We employed RFE for feature selection, using XGBoost as the estimator.

C.Feature importance with SHAP: SHAP assigns a value to each feature in a prediction sample, indicating its impact on the model's prediction. These values aid in sorting and selecting critical features(reducing by 10% at each iteration) .

2.5. Classification Methods

We selected four classifiers commonly used in ECG classification tasks: Random Forest (RF), Support Vector Machine (SVM), the ensemble model of the two, and XGBoost. To assess the consistency of the feature subset across various classifiers and their performance, we conducted a 5-fold cross-validation on DS1, reported the classification results on the DS2 dataset.

2.6. Metrics for Evaluating Models

To assess the classifier's performance, a set of statistical metrics based on True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) have been selected. And the evaluation of the classification methods mainly employs the following performance measures:

$$Precision=TP/(TP+FP),$$

$$Sensitivity=TP/(TP+FN),$$

$$F1 = 2 \times \frac{Precision \times Sensitivity}{Precision+Sensitivity},$$

$$Accuracy=(TP+TN)/(TP+TN+FP+FN)$$

3. Results and Discussion

Extracting features from each heartbeat initially yielded 264 features, reduced to 132 dimensions after selection (Section 2.5). Demonstrating the method's efficacy, cross-validation with four classifiers resulted in performances of 98.00%, 98.64%, 98.78%, and 98.89% on the DS1 training set. XGBoost outperforms other methods in this classification task.

Utilizing the SHAP method discussed in Section 2.4, we ranked feature importances to eliminate weakly correlated features and decrease computational overhead. After training with four classifiers and testing on DS2, the accuracy varied with the number of features, as shown in Figure 3, with XGBoost exhibiting better classification performance.

Based on Figure 3, the optimal feature subset was identified when the feature counting 66. The XGBoost model achieved 98.91% accuracy within the subset, equivalent to 132 features, as shown in Table 2. Meanwhile, superior metrics were achieved in the Q category. In comparison, the model using 132 features yielded 50%, 28.57%, and 36.36% respectively. Accuracy of class 'V' also shows a slight improvement. When below 31 features, all classifiers showed a marked drop in performance, indicating the smallest subset crucial for effective classification.

Furthermore, examining Shap values in Figure 4, we find that for the 5-class heartbeat classification, the top 5 features are closely related to signal morphology and distribution, aligning with our emphasis on modeling heartbeat morphology. The increase in accuracy for the 'Q' classification is attributed to the model's focus on the frequency domain features of the 'Q' class, likely due to its uncategorized status in the database, leading to complex noise distribution and irregular signal morphology.

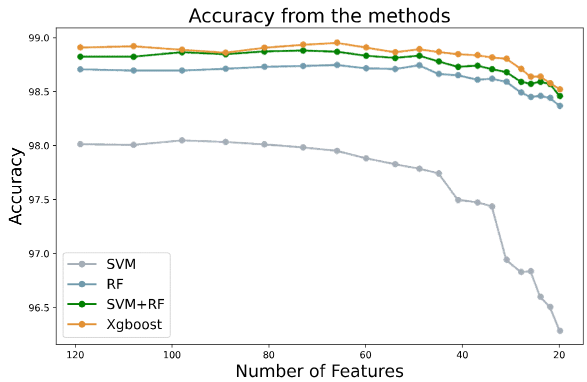

Fig 3. The accuracy of different sizes of optimal feature subsets on the four classifiers

Table 2. The result from XGBoost within 66 features

Class	*Precision* (%)	*Sensitivity*(%)	$F1-score$(%)	*Accuracy* (%)
N	99.22	99.69	99.45	
S	95.20	85.62	90.16	
V	96.96	95.85	96.40	**98.91**
F	89.90	80.37	84.87	
Q	71.43	35.71	47.62	

4. Conclusion

In this study, we introduced an interpretable XGBoost model to arrhythmic heartbeat classification. Utilizing SHAP for feature selection we achieved a classification accuracy of 98.91% with 66 feature set, outperforming the 132 feature set, especially performing better in classifying V and Q categories.

Additionally, we comprehensively compared the performance with other ML methods, SVM, RF, and ensemble models, where XGBoost consistently excelled. Importantly, the feature subset selection by our method yielded an average accuracy above 95% across all ML techniques, validating the effectiveness of our feature extraction.

Finally, we discuss the impact of features selected by Shap values on classification. While the explanations await further validation, exploring classification patterns that focus on single heartbeat features holds significance for more precise disease diagnosis.

Acknowledgement

This research was funded by the National Natural Science Foundation of China (62171123, 62201144), and the National Key Research and Development Program of China (2023YFC3603600, 2022YFC2405600).

References

[1] C. Kreatsoulas and S. S. Anand, "The impact of social determinants on cardiovascular disease", Can. J. Cardiol., vol. 26, pp. 8C-13C, 2010.

[2] Khan, Ali Haider, Muzammil Hussain, and Muhammad Kamran Malik. "Arrhythmia classification techniques using deep neural network." Complexity 2021 (2021): 1-10.

[3] Sannino, Giovanna, and Giuseppe De Pietro. "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection." Future Generation Computer Systems 86 (2018): 446-455.

[4] Al Rahhal, Mohamad M., et al. "Convolutional neural networks for electrocardiogram classification." Journal of Medical and Biological Engineering 38 (2018): 1014-1025.

[5] Faust, Oliver, et al. "Automated detection of atrial fibrillation using long short-term memory network with RR interval signals." Computers in biology and medicine 102 (2018): 327-335.

[6] Sharma, Rishi Raj, et al. “Accurate automated detection of congestive heart failure using eigenvalue decomposition based features extracted from HRV Signals.” Biocybernetics and Biomedical Engineering, vol. 39, no. 2, Apr. 2019, pp. 312–27, https://doi.org/10.1016/j.bbe.2018.10.001.

[10] Zhao, Lina, et al. “A new entropy-based atrial fibrillation detection method for scanning wearable ECG recordings.” Entropy, vol. 20, no. 12, Nov. 2018, p. 904, https://doi.org/10.3390/e20120904.

[7] Hussain, Lal, et al. “Detecting congestive heart failure by extracting multimodal features and employing machine learning techniques.” BioMed Research International, vol. 2020, Feb. 2020, pp. 1 – 19, https://doi.org/10.1155/2020/4281243.

[8] P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," IEEE Trans. Biomed. Eng., vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[9] 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, 2002.

[10] Moody, G. B. , and R. G. Mark . "The impact of the MIT-BIH arrhythmia database. " IEEE Engineering in Medicine and Biology Magazine 20.3(2002):45-50.

[11] ANSI/AAMI, Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms, American National Standards Institute, Inc. (ANSI), Association for the Advancement of Medical Instrumentation (AAMI), ANSI/AAMI/ISO EC57, 1998-(R)2008, 2008.

[12] Barandas, Marília, et al. "TSFEL: Time series feature extraction library." SoftwareX 11 (2020): 100456.

[13] Henderson, Trent, and Ben D. Fulcher. "An empirical evaluation of time-series feature sets." 2021 International Conference on Data Mining Workshops (ICDMW). IEEE, 2021.

Address for correspondence:

Chengyu Liu
Sipailou 2, Nanjing, 210096, Jiangsu Province, P.R.China
Southeast University
E-mail: chengyu@seu.edu.cn

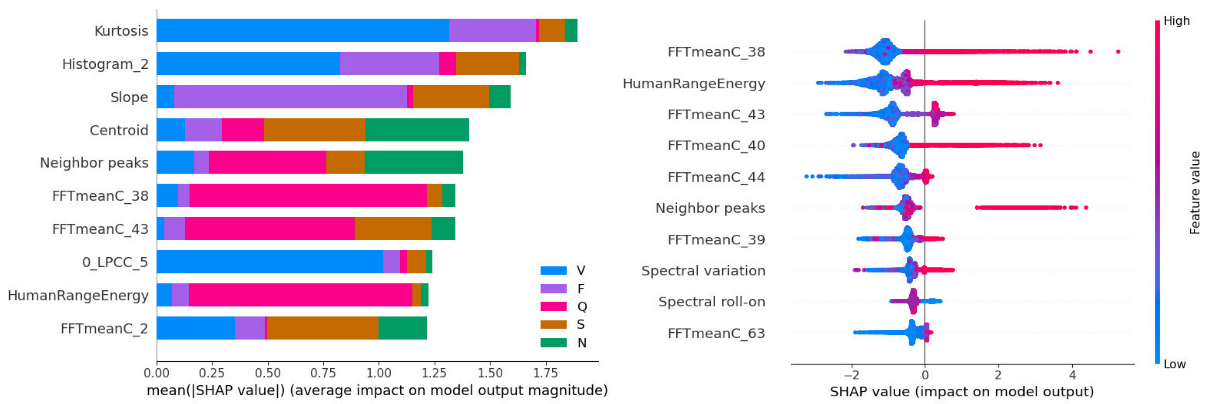

Fig 4. Feature ranking for classifier discriminating 5 categories and discerning 'Q' category (Top 10)