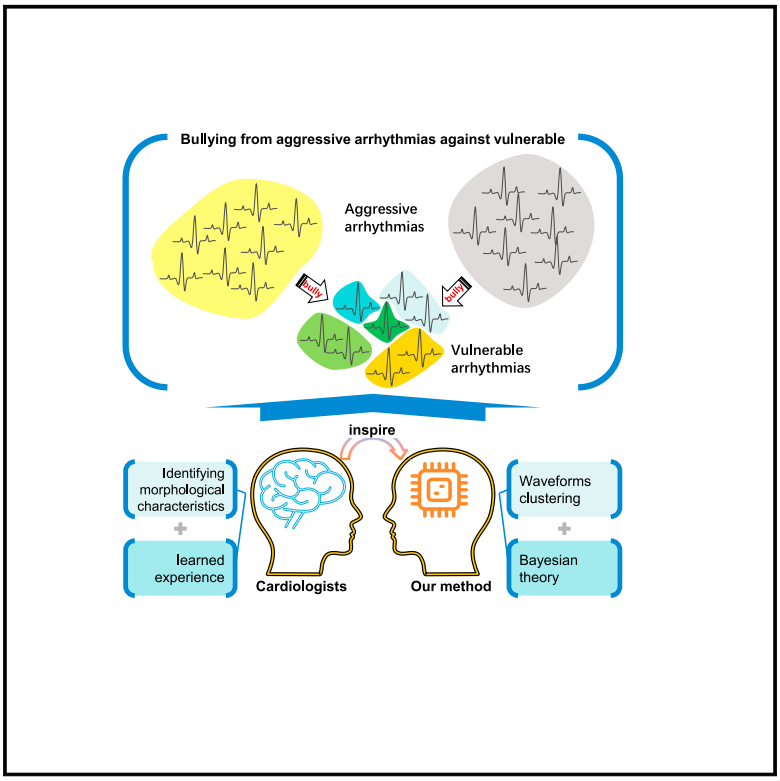
Article

Patterns

A cardiologist-like computer-aided interpretation framework to improve arrhythmia diagnosis from imbalanced training datasets

Graphical abstract



Highlights

- Bullying from aggressive arrhythmias against vulnerable arrhythmias was observed
- A cardiologist-like computer-aided interpretation framework of ECG is proposed
- Morphological-characteristic-based waveform clustering was used for ECG data encoding
- The diagnostic experience of cardiologists was simulated by Bayesian theory

Authors

Lianting Hu, Shuai Huang, Huazhang Liu, ..., Xin Li, Heng Liang, Huiying Liang

Correspondence

lianghuiying@hotmail.com

In brief

The computer-aided interpretation system of ECG is an important tool for providing decision support to cardiologists in arrhythmia diagnosis. The bullying from aggressive arrhythmias against vulnerable arrhythmias makes patients with vulnerable arrhythmias likely to be underdiagnosed. Inspired by the diagnostic thinking of cardiologists, a method that combines morphological-characteristic-based waveform clustering and Bayesian theory is proposed in this study. Our method not only achieved comparable performance on aggressive arrhythmias but also protected vulnerable arrhythmias from being bullied by aggressive arrhythmias.

Hu et al., 2023, Patterns *4*, 100795
September 8, 2023
https://doi.org/10.1016/j.patter.2023.100795

CellPress

Patterns

CellPress
OPEN ACCESS

Article

A cardiologist-like computer-aided interpretation framework to improve arrhythmia diagnosis from imbalanced training datasets

Lianting Hu,[1,2,3,6] Shuai Huang,[2,3,6] Huazhang Liu,[2,3] Yunmei Du,[2] Junfei Zhao,[1] Xiaoting Peng,[2,3] Dantong Li,[1,2,3] Xuanhui Chen,[2,3] Huan Yang,[1,2,3] Lingcong Kong,[2,3] Jiajie Tang,[4] Xin Li,[4] Heng Liang,[5] and Huiying Liang[2,3,7,*]

[1]Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Guangzhou, Guangdong 510080, China

[2]Medical Big Data Center, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong 510080, China

[3]Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Guangzhou, Guangdong 510080, China

[4]School of Information Management, Wuhan University, Wuhan, Hubei 430072, China

[5]School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

[6]These authors contributed equally

[7]Lead contact

*Correspondence: lianghuiying@hotmail.com

https://doi.org/10.1016/j.patter.2023.100795

THE BIGGER PICTURE Medical professionals are increasingly looking to AI-based methods to help interpret medical data and provide more accurate or timely diagnoses for patients. Current AI methods, however, may not always perform consistently for different patient sub-groups or disease sub-types. Underlying dataset imbalances and inhomogeneous inter-class similarity often cause these issues. Underrepresented classes may be less well learned by the AI model, and, in some cases, the presence of particular classes in a dataset can actually interfere with the AI's ability to learn other classes. In this paper, the authors study this issue in detail in the context of diagnosing heart arrhythmias, a common and sometimes life-threatening cardiac disorder, and show that an AI framework that mimics the reasoning of experienced cardiologists can better diagnose arrhythmia sub-types that are sensitive to this kind of interference.

Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

Arrhythmias can pose a significant threat to cardiac health, potentially leading to serious consequences such as stroke, heart failure, cardiac arrest, shock, and sudden death. In computer-aided electrocardiogram interpretation systems, the inclusion of certain classes of arrhythmias, which we term "aggressive" or "bullying," can lead to the underdiagnosis of other "vulnerable" classes. To address this issue, a method for arrhythmia diagnosis is proposed in this study. This method combines morphological-characteristic-based waveform clustering with Bayesian theory, drawing inspiration from the diagnostic reasoning of experienced cardiologists. The proposed method achieved optimal performance in macro-recall and macro-precision through hyperparameter optimization, including spliced heartbeats and clusters. In addition, with increasing bullying by aggressive arrhythmias, our model obtained the highest average recall and the lowest average drop in recall on the nine vulnerable arrhythmias. Furthermore, the maximum cluster characteristics were found to be consistent with established arrhythmia diagnostic criteria, lending interpretability to the proposed method.

INTRODUCTION

Millions of people around the world suffer from cardiac arrhythmias,[1] which increase the risk of complications such as stroke or heart failure and can lead to cardiac arrest, shock, and sudden death.[2] The electrocardiogram (ECG or EKG) is the simplest specific test for cardiologists to diagnose arrhythmias. Since the introduction of computer-aided interpretation over 50 years ago, it has become increasingly important in the clinical ECG workflow, providing decision support to cardiologists in many clinical scenarios.[3] Over the past decade, deep learning has been successfully applied in many fields.[4] With its powerful inferential capability for complex mappings, deep learning has become the mainstream method for computer-aided ECG interpretation in arrhythmia diagnosis.[5]

Abundant deep-learning-based models have been developed for ECG denoising, feature extraction, optimization, and classification. Hannun et al. developed a deep neural network (DNN) to classify 12 arrhythmias using 91,232 single-lead ECGs from 53,877 patients.[6] Yildirim proposed a new model for deep bidirectional long short-term memory (LSTM) network-based wavelet sequences for classifying ECG signals.[7] In addition, other types of deep learning methods, such as multilayer perceptrons,[8] deep belief networks,[9] and recurrent neural networks,[10] have also been employed in computer-aided ECG interpretation. However, variances exist in the diagnostic accuracy of different arrhythmias within computer-aided interpretation models. To facilitate clarity, we introduce two comparative concepts: ''aggressive arrhythmias'' and ''vulnerable arrhythmias.'' Within a specific system and dataset, aggressive arrhythmias exhibit a relatively high recall rate, whereas vulnerable arrhythmias demonstrate a relatively low recall rate. Regarding individual samples, patients with vulnerable arrhythmias were less likely to be correctly diagnosed than patients with aggressive arrhythmias. Moreover, in the context of a particular system, the recall rate of vulnerable arrhythmias experiences a significant decline as aggressive arrhythmias are added during training. This rapid decrease in recall is referred to as the phenomenon of ''bullying,'' whereby aggressive arrhythmias exert a suppressive influence on vulnerable arrhythmias.

The imbalanced distribution of data is considered to be a contributing factor to the prevalence of bullying from aggressive arrhythmias against vulnerable arrhythmias. In the aforementioned studies, deep learning models were trained on large and balanced ECG datasets, and minority arrhythmias were removed from the training dataset and not covered by the diagnosis field of the model. However, the idealized data environment is difficult to construct in real clinical scenarios. The heterogeneity of ECG data among different sources stifles the possibility of constructing large and balanced datasets through the cooperation of multiple centers. The ECG data of common arrhythmias can be easily collected, but rare arrhythmias, such as ventricular escape, are difficult to gather at single centers. Therefore, in real clinical scenarios, the available data for the computer-aided interpretation of ECG are likely to be large but imbalanced. Data augmentation for rare arrhythmias is a potential method to alleviate data imbalance. There are two main kinds of augmentation for time series data: one is to train a generative model from the given data[11] and the other is to increase the diversity of the data by applying augmentation operations to the data.[12,13] However, in the first approach, generative models aim to solve the issue of limited samples, but paradoxically require a substantial amount of scarce data to be effectively trained. The second approach may amplify the noise, alter the timing or duration of the signal, and the periodicity of ECG is underutilized.[14]

In deep learning models, the imbalanced distribution of data is not the only contributing factor in the diagnostic performance of each arrhythmia.[6,14,15] Saxena et al. pointed out that the classification performance of each class was also influenced by the similarity between classes.[16] The classification performance of a class is diluted by classes that are highly similar to it, and it is difficult for deep learning models to discriminate between similar classes.

Unlike deep learning models, cardiologists learn aggressive arrhythmias and vulnerable arrhythmias fairly in training. Cardiologists can learn a new arrhythmia from just one or a handful of examples.[17] Moreover, based on inductive learning, cardiologists can transfer the learned experience to similar diseases.[18,19] Identifying morphological characteristics of ECG waveforms and matching findings with previous diagnostic experience are two main steps for cardiologists to make diagnostic decisions. The presence or absence of P waves, the length of the PR interval, etc., are morphological features that cardiologists focus on.[20] Compared with less experienced cardiologists, more experienced cardiologists rely on their skills to recognize the visual signal patterns of different cardiac abnormalities, providing a more accurate ECG interpretation.[21]

Inspired by the diagnostic thinking of cardiologists, this study proposes an arrhythmia diagnosis method that combines morphological characteristic-based waveform clustering and Bayesian theory. Unsupervised clustering was utilized to simulate cardiologists to identify morphological characteristics of ECG waveforms. Waveforms with the same morphological characteristics were grouped and encoded identically. The diagnostic experience of cardiologists was simulated by the prior probability in Bayesian theory. Furthermore, heartbeat splicing was used to exhaustively enumerate the possible heartbeat combinations, which made the Bayesian model more stable. Our method was validated in the GDPH ECG-Arrhythmia Dataset, a real-world-collected dataset that covers 17 types of arrhythmias and is extremely imbalanced. We hope that the use of a cardiologist-like model framework can alleviate the bullying of vulnerable arrhythmias by aggressive arrhythmias.

RESULTS

Data distribution and inter-similarity

The GDPH ECG-Arrhythmia Dataset[22] was collected at the Guangdong Provincial People's Hospital, Guangzhou, Guangdong, China, from August 2014 to October 2021. The dataset comprises 48,063 participants (Figure 1). Each participant is labeled as normal or as having 1 of 17 kinds of arrhythmia (Table S1) based on the ECG report by a cardiologist and has been verified by another senior cardiologist. The 17 kinds of arrhythmia are premature ventricular contraction (PVC), intraventricular block (IV block), ventricular tachycardia (VT), ventricular escape (VE), atrial flutter (AFL), atrial tachycardia (AT), atrial fibrillation (AF), premature atrial contraction (PAC), premature

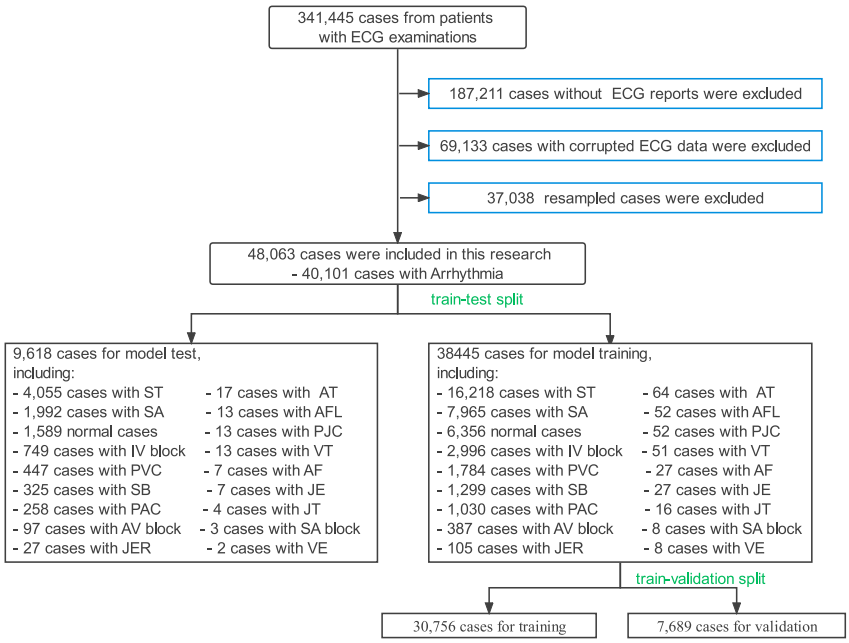


Figure 1. Flowchart of study identification and inclusion and exclusion criteria

junctional contraction (PJC), junction escape (JE), junction tachycardia (JT), junction escape rhythm (JER), atrioventricular block (AV block), sinoatrial block (SA block), sinus tachycardia (ST), sinus bradycardia (SB), and sinus arrhythmia (SA).

As a real-world dataset, the GDPH ECG-Arrhythmia Dataset is extremely imbalanced (Figure 2A). ST has the largest sample size (n = 20,273), and VE has the smallest sample size (n = 10). The standard deviation of the sample sizes for all arrhythmia categories is 5,249.63. The mean and median of the sample sizes are 2,670 and 106, respectively. The imbalance of the dataset in terms of sample size makes it difficult for the model to learn all arrhythmia categories equally. The inter-similarity matrix, as depicted in Figure 2B, is normalized to the range [0, 1] through unity-based normalization. Each element in the matrix represents the similarity score between two arrhythmia categories computed using dynamic time warping (DTW).[23] A high value in the DTW calculation indicates low similarity between the two series.[24] As illustrated in Figure 2B, SB, JER, SA, JE, AV block, PJC, PVC, AT, and AF are relatively similar, whereas the similarity score between VT and PVC is 1, indicating the lowest similarity between these two arrhythmias.

Hyperparameter optimization

The overall modeling framework is shown in Figure 3. The ECG signal is subjected to preprocessing techniques aimed at eliminating interferences. Subsequently, the P-QRS-T localization algorithm is implemented to identify the key points. Upon detection of these key points, the preprocessed ECG signal is disintegrated into multiple heartbeats, which are then further segmented into six sections. Each of these segments is encoded using the cluster number obtained from the segment clustering algorithm. The heartbeat splicing approach is adopted to increase the sample size. Finally, categorical naive Bayes is applied as a classifier to diagnose arrhythmias, based on the encoded vector.

A grid search was used in our study to perform hyperparameter optimization (Figure 4A). The macro-recall of all 18 categories guided the grid search algorithm. There were two hyperparameters in the configuration: the number of splicing heartbeats and the number of clusters. The model with five heartbeats in splicing and 20 clusters in segment clustering obtained the best performance (macro-recall, 71.55% [95% CI 68.29%–74.61%]; macro-F1 score, 54.97% [95% CI 49.63%–56.39%]; macro-precision, 52.93% [95% CI 50.11%–55.73%]). Regardless of the number of splicing heartbeats, the macro-recall initially increased with the number of clusters, but after 15 or 20 clusters, it decreased with the number of clusters. However, the three groups with different numbers of splicing heartbeats had no significant difference in the macro-recall ($p = 0.777$). The receiver operating characteristic (ROC) curve of the optimal hyperparameters is shown in Figure 4B.

Performance comparison with alternative models

In this section, our method was compared with K-nearest neighbors (KNN),[25] random forest (RF),[26] extreme gradient boosting (XGBoost),[27] one-dimensional convolutional neural network (1D CNN),[6] and LSTM[7] in performance (Figure 5A). The macro-recall of our method was higher than those of alternative models (1D CNN, 38.42% [95% CI 36.43%–42.65%], LSTM, 26.33% [95% CI 25.21%–29.57%], KNN, 22.44% [95% CI 20.14%–24.33%], XGBoost, 18.10% [95% CI 18.81%–19.74%], RF, 17.09% [95% CI 15.01%–21.51%]). The alternative models did not perform poorly for all arrhythmias. For example, all alternative models achieved over 99.93% recall for the normal class, and 1D CNN and LSTM achieved a similar recall compared with our method for ST. The advantage of our method was mainly reflected in some cases of arrhythmia that alternative models failed to recognize, such as JE, SA block, JT, AF, AT, JER, PJC, and VE. According to the recalls of the 18 categories, the first nine arrhythmias with lower recalls were defined as vulnerable arrhythmias, and the rest were defined as aggressive arrhythmias.

The correlation coefficients (CC) between the recall of the 18 categories and the data distribution (or sample size) of these categories are shown in Figure 5B. The CC of our method was 0.3626, which suggests that there was only a weak positive correlation between our method and the data distribution. However, the XGBoost had a moderate positive correlation with the data distribution (CC = 0.5561). The CCs of the other alternative models were in the range [0.6, 1], which suggested a strong positive or a very strong positive correlation. The CCs between the recall of the 18 categories and their inter-similarity are shown in Figure 5C. Our method, KNN, LSTM, and RF had similar CCs, about 0.2, suggesting that there was a weak positive

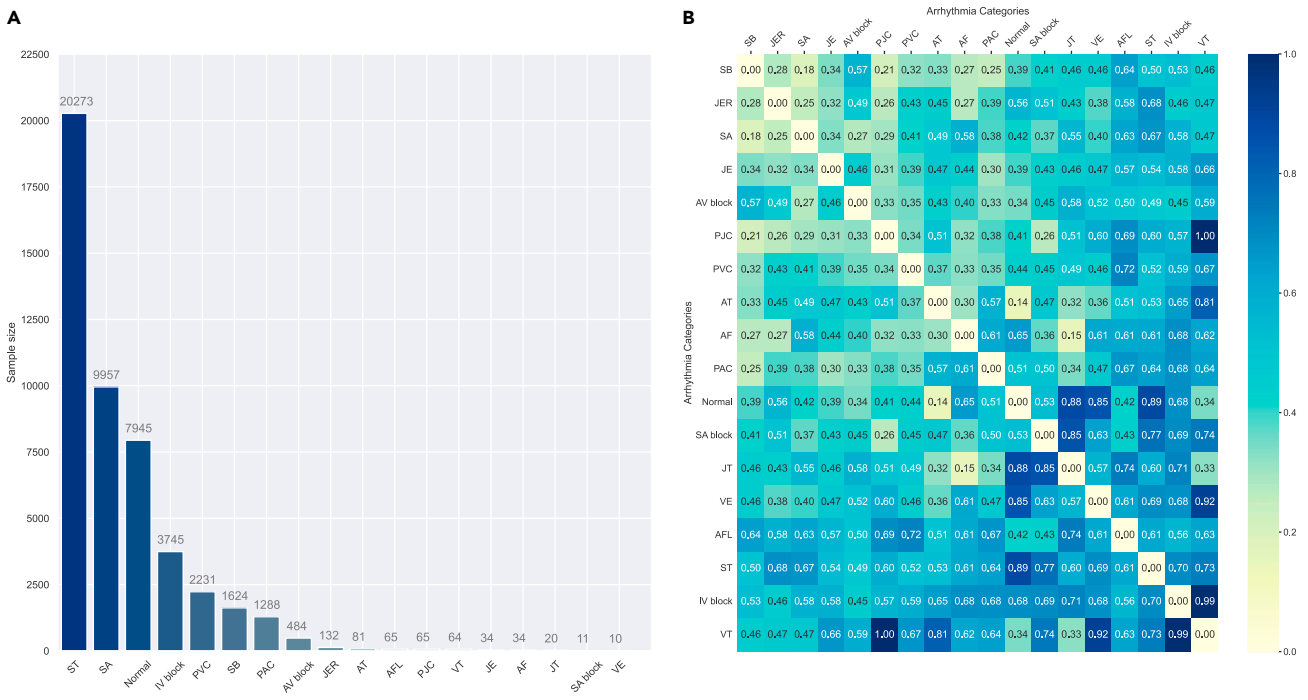


Figure 2. The GDPH ECG-Arrhythmia Dataset

(A) The data distribution (or sample size) of each arrhythmia category in the GDPH ECG-Arrhythmia Dataset.

(B) The inter-similarity matrix of arrhythmia categories. Each element in the matrix represents the similarity between two arrhythmia categories. The inter-similarity is the average similarity of all pairs of ECG signals from two arrhythmia categories. Dynamic time warping (DTW) was used to measure the similarity between two ECG signals. The inter-similarity matrix was normalized to the range [0, 1] and was sorted from minimum to maximum.

correlation between these models and the inter-similarity. In contrast, the CCs of 1D CNN and XGBoost were in the range [0.3, 0.5] and had a moderate positive correlation with the inter-similarity.

Bullying from aggressive arrhythmias against vulnerable arrhythmias

In this section, two experiments to explore our method's ability to protect vulnerable arrhythmias from being bullied by aggressive arrhythmias are described. In the first experiment, the training dataset included nine vulnerable arrhythmias and one aggressive arrhythmia, "normal" (Figure 6A). The sample size of the nine vulnerable arrhythmias did not change. The sample size of "normal" was respectively set to 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, and 6,388 during training. Regardless of the sample size of "normal," the recall of some vulnerable arrhythmias in alternative models was still 0, for example, SA block, PJC, and VE. Meanwhile, the recall of some vulnerable categories in several alternative methods gradually tended to 0 with the increase in sample size of "normal," such as the recall of JT in 1D CNN (Figure S1). In contrast, our method achieved good performance in vulnerable arrhythmias, which shows that our method had a protective effect on vulnerable arrhythmias.

In the second experiment, the models were still trained by nine vulnerable arrhythmias and aggressive arrhythmias (Figure 6B). The number of aggressive arrhythmias was set from 0 to 8, with 0 indicating that no aggressive arrhythmias were added. The order of addition was AV block, AFL, VT, PVC, IV block, SB, SA, and ST, which were sorted from the lowest to the highest level of aggressiveness. The average/average drop in recall for all models on the nine vulnerable arrhythmias were 13.21%/30.75% (1D CNN), 6.23%/50.64% (KNN), 9.66%/41.13% (LSTM), 1.68%/36.64% (RF), 4.73%/25.04% (XGBoost), and 38.25%/28.96% (our method) (Figure S1). Considering the high recall of our model, the drop in recall of our model with increasing aggressive arrhythmias was relatively low.

Interpretability

According to Bayes' theorem, the conditional distribution of segment clusters over arrhythmias is the key to the decision-making of the categorical naive Bayes. Therefore, some morphological characteristics of arrhythmias may be implied in segment clusters with large conditional probabilities. In Figure 7, the visualization of segment clusters of four arrhythmias and three examples of the maximum cluster is shown in four rows. In the AF row, three examples of the T-P interval show chaotic atrial activity; in the AV block row, compared with the normal P-Q interval, the three P-Q interval examples were prolonged by about 0.20 s; in the SA block row, the T-P interval was approximately equal to one cardiac cycle, which means one heartbeat was dropped; in the PVC row, compared with the normal QRS complexes, the three QRS complex examples were wider (>0.1 s), and the first example had a bizarre appearance. The characteristics of the maximum cluster mentioned above were consistent with the diagnostic criteria of the four arrhythmias,[28] which indicates that our method has certain interpretability.

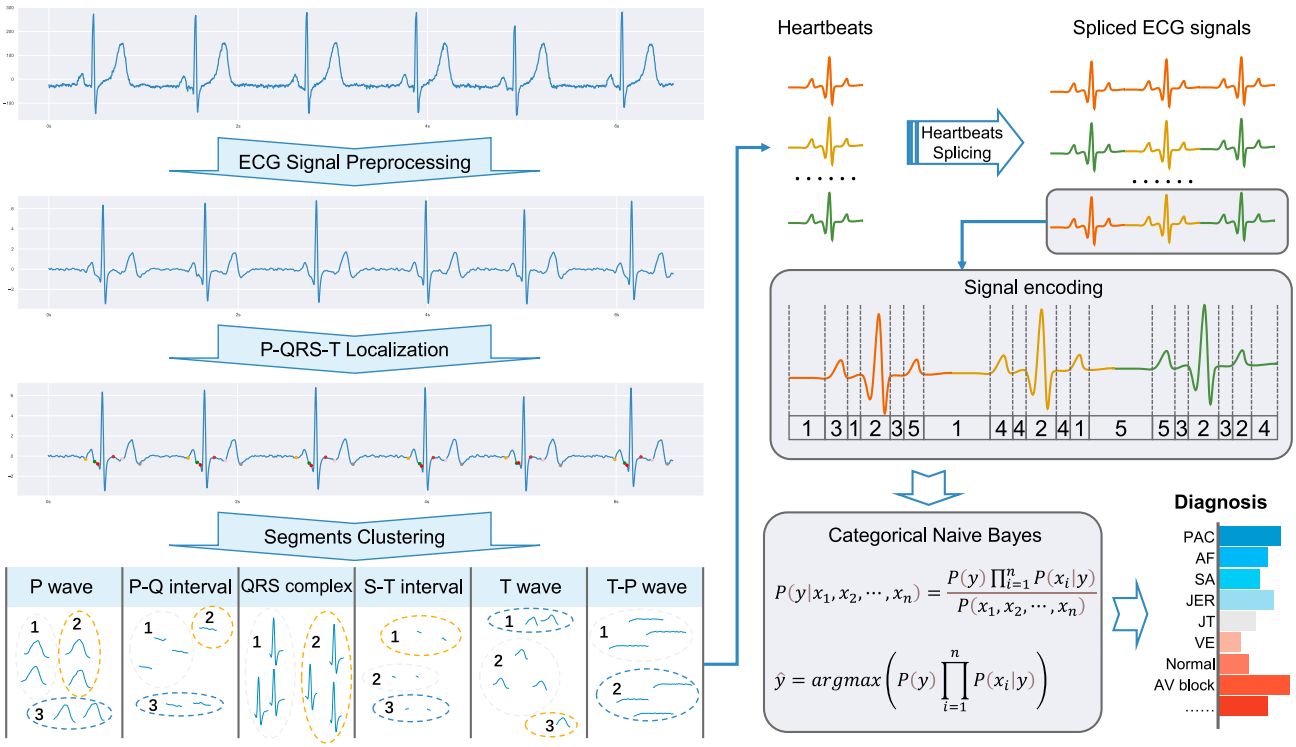


Figure 3. Overall modeling framework

The original ECG signal is preprocessed to remove interferences. Then, the P-QRS-T localization is performed. Based on those detected key points, the preprocessed ECG signal is broken down into multiple heartbeats, and heartbeats are further segmented into six segments. Each segment is encoded with the cluster number from the segment clustering. Heartbeat splicing is performed to augment the sample size. Finally, based on the encoded vector, categorical naive Bayes is used as a classifier to diagnose arrhythmias.

DISCUSSION

Due to differences in sample size and inter-similarity, arrhythmias can be divided into aggressive arrhythmias and vulnerable arrhythmias in computer-aided interpretation. The bullying from aggressive arrhythmias against vulnerable arrhythmias makes the patients with vulnerable arrhythmias likely to be underdiagnosed. To end this, and inspired by the diagnostic thinking of cardiologists, this study proposes a framework for arrhythmia diagnosis that combines ECG segment clustering and Bayesian theory. The GDPH ECG-Arrhythmia Dataset was used to validate our method. With optimization of the hyperparameters, including the number of spliced heartbeats and the number of clusters, the proposed method outperforms alternative models, achieving the best performance in terms of macro-recall (71.55%, 95% CI 68.29%–74.61%), macro-F1 score (54.97%, 95% CI 49.63%–56.39%), and macro-precision (52.93%, 95% CI 50.11%–55.73%). In addition, with increasing bullying from aggressive arrhythmias, our model obtained the highest average recall (38.25%) and the lowest average drop in recall (28.96%) on the nine vulnerable arrhythmias.

Deep learning has shown outstanding performance in the interpretation of ECG in the past few years. Its hierarchical architecture enables higher-level features to be obtained, and its strong feature extraction ability helps to fit complex mappings.[29] Due to the characteristics of vulnerable arrhythmias, it is difficult for deep learning models to treat vulnerable arrhythmias as fairly as aggressive arrhythmias during training. In this study, a possible way was explored to protect vulnerable arrhythmias from being bullied by aggressive arrhythmias. The morphological characteristics of ECG signals are key information for cardiologists to diagnose arrhythmias. Inspired by this, we used segment clustering to distinguish ECG signals with different morphological characteristics. The permutation of the ECG signal at the beat level could effectively enrich the sample size of vulnerable arrhythmias. An ECG is a type of periodic electrophysiological signal, with one heartbeat constituting one cycle.[30] As a result, splicing multiple original heartbeats does not significantly affect the morphological characteristics of the ECG. Furthermore, the typical signs of most arrhythmias are usually present within the heartbeat.[28] Our method does not damage the heartbeat during the segmentation of ECG signals, thereby effectively preserving the information related to any underlying arrhythmias. On the other hand, compared with a single heartbeat, the splicing of multiple heartbeats increases the dimension of the ECG signal, which may increase the distance of different arrhythmias in the representational space. As shown in Figures 5 and 6, our method could indeed alleviate the bullying by aggressive arrhythmias of vulnerable arrhythmias to a certain extent.

In the study, there are two main types of alternative models: feature engineering + classifier models and end-to-end models. The former utilizes commonly used feature extraction techniques[31] and three classifiers. The latter includes 1D CNN and LSTM[32] models, which have achieved state-of-the-art performance in several ECG datasets.[33–35] Our method is compared with these alternative models and is found to fail to

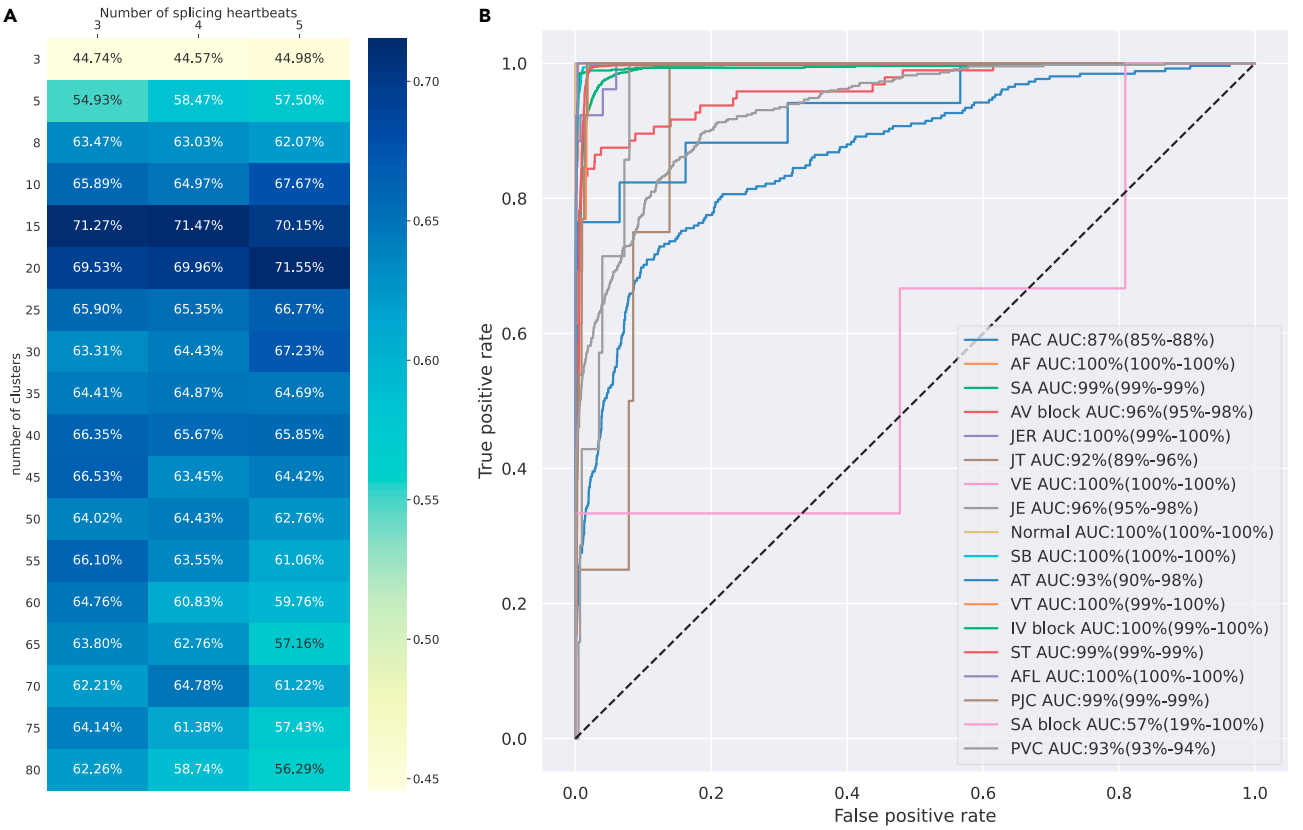


Figure 4. The hyperparameter optimization of our method and the confusion matrix of the optimal hyperparameters

(A) The hyperparameter optimization was performed by a grid search, and the performance metric was the macro-recall of all 18 categories.

(B) The receiver operating characteristic (ROC) curve of the optimal hyperparameters: five heartbeats in splicing and 20 clusters in segment clustering.

comprehensively outperform end-to-end models for diagnosing aggressive arrhythmias. However, our model exhibits superior performance in the diagnosis of vulnerable arrhythmias. Moreover, the correlation between our model's performance and data distribution and the inter-class similarity is lower than that of the alternative models. The existence of periodicity in the data is a fundamental requirement for the effective implementation of heartbeat clustering and splicing procedures. As a result, our approach has significant potential for extension to other types of periodic medical time-series data, including gait data, among others. Furthermore, our method may be potentially beneficial in mitigating bullying in clinical settings, where the aggressive category targets the vulnerable category. The application of our approach in such cases may protect the vulnerable group and potentially facilitate the development of a computer-aided diagnosis system characterized by exceptionally high diagnostic accuracy for the vulnerable population.

The computer-aided interpretation system of ECG with explanatory ability will be more reliable and acceptable to cardiologists. In our method, features are encoded from the number of segment clusters with clear practical meaning. Diagnostic decisions are made based on the conditional probabilities of the arrhythmia sub-segments, which are a simplified mathematical description of the cardiologists' previous diagnostic experience. Therefore, some morphological characteristics of arrhythmias may be implied in the segmented clusters with high conditional probabilities. Some morphological characteristics found by our model can match the current diagnostic criteria of arrhythmias, such as the typical segments for the four arrhythmias shown in Figure 7. Although there were some unmatched findings according to current diagnostic criteria, they might imply new diagnostic markers for arrhythmias. The joint conditional probabilities of multiple segments may be a feasible way to discover new diagnostic markers.

One of the main limitations of our study is that our method was not validated on multiple sites. In addition, the hyperparameters were set based solely on our dataset, which may affect the generalizability of our findings. Nonetheless, our method demonstrated good performance in protecting against vulnerable arrhythmias, while the deep learning model still holds a significant advantage in diagnosing aggressive arrhythmias. Future work could explore the potential benefits of combining the advantages of our method with deep learning technologies such as transformers and generative pretrained transformer (GPT) in timing analysis. Compression, computational efficiency, transmission, and power effectiveness with precision are the key performance indicators of computer-aided interpretation systems.[36,37] Therefore, measuring and then optimizing the efficiency of our model are also part of our future work. Moreover, cardiologists can apply their experience to new arrhythmias, and this could be simulated in our future work.

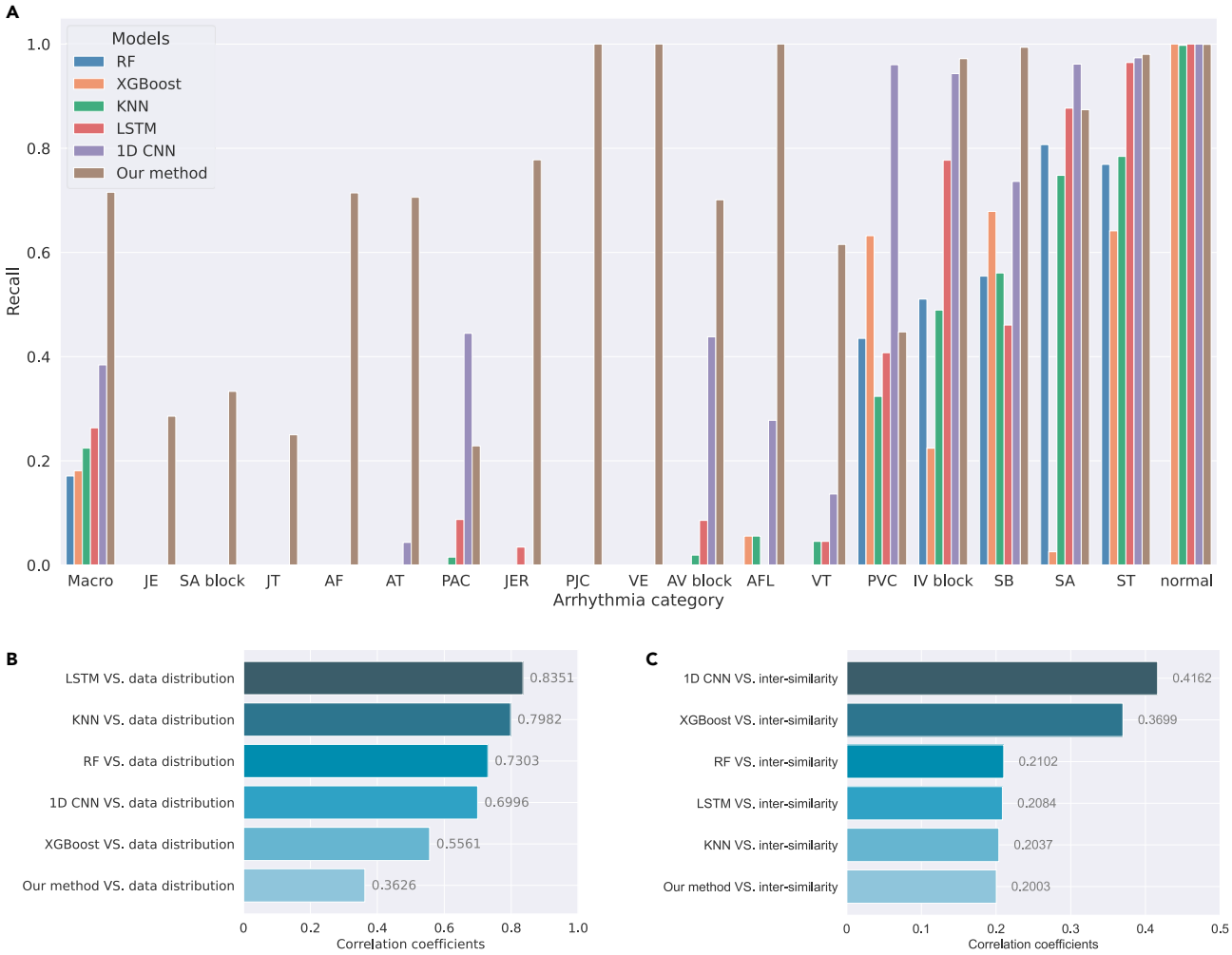


Figure 5. Performance comparison with alternative models

(A) The recalls of our method as well as of alternative models. The labels on the x axis are the 18 categories and "Macro," which indicates the macro-recall of all 18 categories. The categories are arranged from left to right based on their average recall.

(B) The CCs between the recall of the 18 arrhythmias and the data distribution of those. The labels on the x axis are the value of the CC. The string "a VS. b" indicates the CC between "a" and "b".

(C) The CCs between the recall of the 18 categories and the inter-similarity of those. The inter-similarity of a specific arrhythmia is defined as the sum of the inter-similarity between the arrhythmia and other arrhythmias.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

The lead contact for this article is Huiying Liang at lianghuiying@hotmail.com.

Materials availability

Correspondence and requests for materials should be addressed to Huiying Liang (lianghuiying@hotmail.com).

Data and code availability

The related code and some ECG data are available on Zenodo[22] (ECG data: https://zenodo.org/record/7902431#.ZFYgSnbP2Po, https://doi.org/10.5281/zenodo.7902431). Due to confidentiality agreements, the full ECG data can be made available subject to a non-disclosure agreement. For further information, you may contact Huiying Liang (lianghuiying@hotmail.com). Any data use will be restricted to non-commercial research purposes.

GDPH ECG-Arrhythmia Dataset

The ECG data utilized in this study were collected at the Guangdong Provincial People's Hospital, located in Guangzhou, Guangdong, China, from August 2014 to October 2021.[22] Due to the limited acquisition area of the chest in pediatric patients, signals were absent in leads V2, V4, and V6 for many participants, resulting in only the signal from the first nine leads being included in this investigation. The NIHON KOHDEN ECG-2550 was used as the testing ECG machine, with a sampling frequency of 500 Hz and a sampling time of 10 s. Each participant was labeled as normal or one of the other 17 kinds of arrhythmias based on the ECG report by a cardiologist and was verified by another senior cardiologist. According to the location of the occurrence, the 17 types of arrhythmias can be grouped into four super-categories: the sinus (including SA block, ST, SB, and SA), the atrial (including AFL, AT, AF, and PAC), the atrioventricular junction (including PJC, JE, JT, JER, and AV block), and the ventricular (including PVC, IV block, VT, and VE).

Some ECG cases were excluded for the following reasons: (1) ECG cases were distorted seriously due to lack of signal or excessive noise. Since the ECG records the electrical activity of the heart through electrodes placed on the skin, large movements and a noisy surrounding environment may add irremovable noise to the signal. Moreover, this kind of distortion is more common in children's ECGs. (2) The label of the ECG case was unavailable or uncertain. (3) For participants with multiple ECG tests, to avoid introducing bias in the

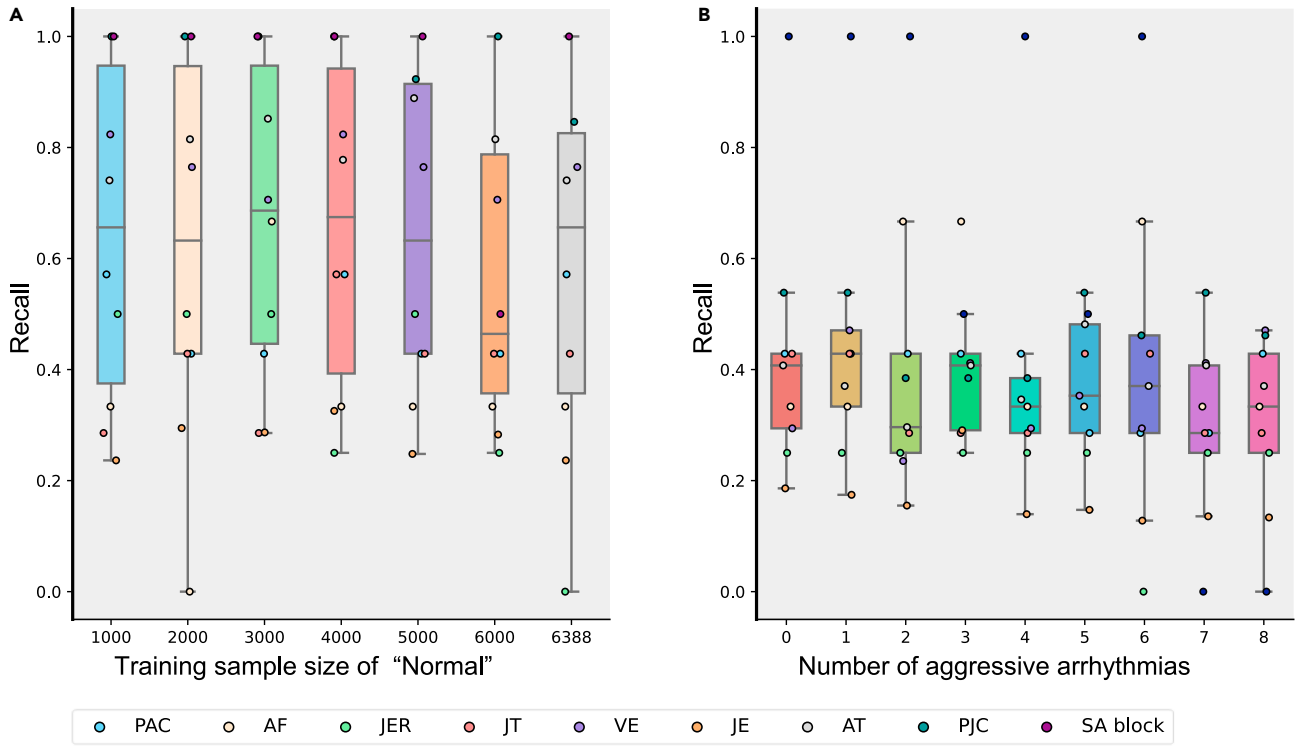


Figure 6. The performance of our model in protecting weak arrhythmias (related to Figure S1)

(A) The recalls of the nine vulnerable arrhythmias with the sample size of "normal" increasing. The model was trained by nine vulnerable arrhythmias and the aggressive class "normal." The sample size of the nine vulnerable arrhythmias did not change. The sample size of "normal" was respectively set to 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, and 6,388 during training.

(B) The recalls of the nine vulnerable arrhythmias with the number of aggressive arrhythmias increasing. The model was still trained by nine vulnerable arrhythmias and aggressive arrhythmias. The number of aggressive arrhythmias was set to 0 to 8. The order of addition was AV block, AFL, VT, PVC, IV block, SB, SA, and ST.

participant, only the last ECG test was used in our study, and other ECG tests were excluded. After exclusion, 48,063 participants remained. The whole dataset was randomly divided into the training dataset and the test dataset in a ratio of 4:1 at the level of each arrhythmia.

ECG signal preprocessing

The ECG signal is a weak physiological signal, which is easily disturbed during the acquisition process, so it is necessary to preprocess the ECG signal before analysis. The three most common interferences of ECG signals are electromyography (EMG) interference, power frequency interference, and baseline drift. Our preprocessing was performed based on the three interferences mentioned above.

EMG is also a kind of physiological signal and is the main noise in the ECG signal. The frequency of EMG is related to the type of muscle, generally in the range of 30–300 Hz, while the frequency of the ECG signal is mainly in the range of 5–20 Hz. Therefore, the EMG signal can overlap with the ECG signal. In our study, the Butterworth low-pass filter was used to remove EMG signals.[38] The Butterworth low-pass filter has the flattest bandpass frequency response curve and gradually drops to zero as the stopband is adjusted. Moreover, the amplitude of the diagonal frequency decreases monotonically, and the higher the filter order, the faster the amplitude decay in the stopband.

Power frequency interference is ubiquitous with the existence of a power supply network, while the interference signal with a frequency of 50 Hz is the most common one. In this study, the 50 Hz finite impulse response (FIR) notch filter with the Kaiser Windows function was used to eliminate the power frequency signal.[39] FIR filters have the linear phase characteristics required for ECG signal processing and can achieve the best filtering performance with minimal waveform distortion. Taking into account the population differences of the samples, the Kaiser moving window is a window function close to the optimal structure,[40] which can adaptively adjust the parameters of the filter according to different settings.

Finally, the removal of ECG baseline drift is accomplished by an infinite impulse response zero-phase shift digital filter. As a routine preprocessing step for ECG analysis, it prevents the introduction of artifact information that could distort the true oscillatory phase. As shown in Figure S2, after the preprocessing, the main noise was removed while the key information of the ECG signal was preserved.

Inter-similarity

In our study, the similarity between two arrhythmia categories was defined as inter-similarity, which was measured using the DTW algorithm. To improve the efficiency of DTW calculation, 1,000 ECG signals were sampled repeatedly from two different arrhythmias, forming 1,000 pairs of ECG signals. In each ECG signal, the first and last 2 s were removed, and the remaining signal was downsampled to 100 Hz.

DTW is an algorithm used to measure the similarity between two series. It can handle sequences of different lengths and speeds by warping the time axis to find the best match. The DTW algorithm calculates similarity by constructing a cost matrix and finding the minimum cumulative cost path through dynamic programming. Given two series, $X = \{x_1, x_2, \cdots, x_n\}$ and $Y = \{y_1, y_2, \cdots, y_m\}$, the steps to calculate DTW similarity are as follows.

Construct a cost matrix C with the shape $n \times m$, where $C[i,j]$ is the distance between the ith element of sequence X and the jth element of sequence Y. Euclidean distance is commonly used as the distance metric:

$$C[i,j] = (x_i - y_j)^2 \quad \text{. (Equation 1)}$$

Initialize the dynamic programming matrix D, setting $D[0,0]$ as $C[0,0]$. Fill in each element of matrix D using dynamic programming. For $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, m$:

Figure 7. The visualization of segment clusters in four arrhythmias

The first column is the histogram of 20 segment clusters, and the maximum cluster (segment cluster with the greatest number of instances) is shown in salmon pink. The second to fourth columns show three examples of the maximum cluster. The first to fourth rows show the clusters of the T-P interval of lead II in AF participants, the clusters of the P-Q interval of lead I in AV block participants, the clusters of the T-P interval of lead II in SA block participants, and the clusters of QRS complexes of lead II in PVC participants, respectively.

$$D[i,j] = C[i,j] + min(D[i-1,j], D[i,j-1], D[i-1,j-1]) \quad \text{. (Equation 2)}$$

Here, $min(D[i-1,j], D[i,j-1], D[i-1,j-1])$ represents finding the path with the smallest cumulative cost. Calculate the final DTW similarity, which is usually the square root of $D[n,m]$.

P-QRS-T localization in ECG data

In our study, an adaptive and time-efficient R-peak detection algorithm was used to determine the location of the R peak.[41,42] Then, the Q peak and S peak were found by searching on both sides of the R peak. Due to the rare occurrence of multiple peaks in the QRS complexes, a moving window of 250 ms was used to iteratively query both sides. The minimum value in the first window on the left is the position of the Q peak. The minimum value in the first window on the right is the S peak.

Detection of the P and T waves requires time window traversal starting from the QRS complexes. Therefore, a detection algorithm for the start and end boundaries of the wave is needed. For this purpose, an improved boundary detection algorithm based on the local distance transform was applied in our study.[43] The local distance transform finds the start and end points of a wave by calculating the maximum distance between the start and the end points of the auxiliary line segment at each point on the signal. From the view of morphology, this point is the maximum curvature point, which is in line with the doctor's subjective judgment.

A 200 ms time window was established before the Q start point and a 400 ms time window after the S end point. The P peak and the T peak were detected by the same detection algorithm used in the R-peak detection. The local distance transform was also used here to determine the start point and the end point of the P and T waves. A sample is shown in Figure S2. A total of 11 points were identified in each heartbeat. The division of each heartbeat and the division of the waves within the heartbeat were based on this.

Segment clustering

Based on the important areas identified in the previous step, the ECG signal was divided into several periodic heartbeats. From the entire dataset, N heartbeats (referred to as the beat-level dataset) were generated (Table S1). Each heartbeat was then divided into six different types of segments, creating a database of $6 \times N$ segments: P wave, P-Q interval, QRS complexes, S-T interval, T wave, and T-P interval. The signal between the P start point and the P end point, for instance, is described as the P wave. See Figure S3 for more segment definitions. Assume that the six waveforms mentioned above are used to split $6 \times N$ segments into six segment-level sub-datasets. There are undoubtedly N segments in these sub-datasets.

Consider observing a set of N segments $\{x^1(l_1), x^2(l_2), \ldots, x^N(l_N)\}$ in the specific segment-level dataset, and $x^n \in R^{L_n}$ represents its corresponding feature vector. Specifically, $x^i(l_i) \in \{x_i^1, x_i^2, \ldots, x_i^{l_i}\}$ denotes the ith segment, which retrieved the ith beat signal from the beat-level dataset, and $l_i \in \{1, 2, \ldots, l_i\}$ denotes the length of the segment. The following equation can be used to determine the average length $\bar{l}$:

$$\bar{l} = \frac{1}{N}\sum_{i=1}^{N} l_i \quad \text{. (Equation 3)}$$

Control points and their related indices are retrieved as a representation of the segments for ease of presentation. Empirically, a segment with more control points has a longer section. The number of control points N_c can be obtained:

$$N_c = \frac{1}{\gamma}\bar{l} \quad \text{, (Equation 4)}$$

where γ denotes simulation outline parameters and $\bar{l}$ denotes the average length of all segments in the corresponding segment dataset. To be more specific, a low value of γ reserves the segment a more detailed outline.

In the segment dataset, control points are retrieved at regular intervals for each segment. Consequently, the following equation can be used to determine the value of the jth control point of the ith segment:

$$v(i,j) = x_i^{\frac{l_i}{N_c}*j} \quad \text{, (Equation 5)}$$

where $v(i,j)$ denotes the control point value, x_i^* denotes the feature of the index $*$ in the ith segment, and $\frac{l_i}{N_c}*j$ is the corresponding index of the jth control point. Figure S4 displays the length distribution in each sub-dataset. The feature details extracted from each segment are shown in Table S2. Accordingly, comparable features were extracted for each of the nine leads in the ECG data.

Based on the features of the instances, the k-means was performed to cluster the segments. Euclidean distance was used to measure the distance between the instance and the centroid of the cluster. The algorithm stops when the centroids do not change significantly in iterations. Clustering needs to be implemented 54 (= 6 segments × 9 leads) times for six segments and nine leads (Figure S5). For all other clustering distributions under all leads with different clustering configurations, see Data S3.

Heartbeat splicing and encoding

In our study, multiple original heartbeats were spliced together as a spliced ECG signal. The spliced ECG signal can be regarded as permutations of the original heartbeats, and repetition is allowed in the permutation[44,45] (Figure S6). The spliced ECG signal is formed by original heartbeats from the same participant, so the spliced ECG signal can be assigned the same label as the original heartbeat. The integration of spliced ECG signals into the training dataset is capable of simulating ECG signals that may not exist within the original dataset. This procedure is advantageous as it serves to enhance the comprehensiveness of the training dataset, thereby potentially reducing the overfitting associated with our model.

Assume that $N_m^h (h = 1,2,\cdots,H; m = 1,2,\cdots,M_h)$ indicates the number of original heartbeats of the mth ECG signal in the hth class, where H is the number of classes and M_h is the number of ECG signals in the hth class. N_o is the number of original heartbeats in the splicing. For the mth ECG signal in the hth class, the number of spliced ECG signals SN_m^h can be obtained by the following equation:

$$SN_m^h = \left(N_m^h\right)^{N_o} \quad \text{. (Equation 6)}$$

Therefore, the number of all spliced ECG signals SN is:

$$SN = \sum_{h=1}^{H}\sum_{m=1}^{M_h} SN_m^h \quad \text{. (Equation 7)}$$

Substituting Equation 6 into Equation 7, we can get:

$$SN = \sum_{h=1}^{H}\sum_{m=1}^{M_h} \left(N_m^h\right)^{N_o} \quad \text{. (Equation 8)}$$

Based on the cluster number of the segment, the spliced ECG signal can be encoded as a vector (Figure S6). Each heartbeat has six segments in each lead. Each segment is assigned a cluster number in the unsupervised clustering step. The heartbeat can be encoded as a feature vector with 54 (= 6 segments × 9 leads) dimensions. Therefore, the feature vector of the spliced ECG signal has $54\times N_o$ dimensions.

Categorical naive Bayes

Based on the encoded features, categorical naive Bayes (CNB) is used to classify the spliced ECG signal. Assume that $V = (v_1, v_2, \cdots, v_{54\times N_o})$ is the feature vector of the spliced ECG signal and y is the class variable. According to the property of conditional probabilities, we can get:

$$P(v_1, v_2, \cdots, v_{54\times N_o})P(y|v_1, v_2, \cdots, v_{54\times N_o}) = P(y)P(v_1, v_2, \cdots, v_{54\times N_o}|y) \quad \text{, (Equation 9)}$$

which can be transformed to:

$$P(y|v_1, v_2, \cdots, v_{54\times N_o}) = \frac{P(y)P(v_1, v_2, \cdots, v_{54\times N_o}|y)}{P(v_1, v_2, \cdots, v_{54\times N_o})} \quad \text{. (Equation 10)}$$

Based on the "naive" assumption that features are conditional independent of each other, we can obtain:

$$P(v_i|y, v_1, \cdots, v_{i-1}, v_{i+1}, \cdots, v_{54\times N_o}) = P(v_i|y) \quad \text{. (Equation 11)}$$

Therefore, Equation 10 is simplified to:

$$P(y|v_1, v_2, \cdots, v_{54\times N_o}) = \frac{P(y)\prod_{i=1}^{54\times N_o} P(v_i|y)}{P(v_1, v_2, \cdots, v_{54\times N_o})} \quad \text{. (Equation 12)}$$

Because $P(v_1, v_2, \cdots, v_{54\times N_o})$ is constant among all in Equation 12 with different classes, the classification rule can be written as:

$$\hat{y} = argmax\left(P(y)\prod_{i=1}^{54\times N_o} P(v_i|y)\right) \quad \text{, (Equation 13)}$$

where $\hat{y}$ is the predicted label of the ECG signal. $P(v_i|y)$ can be calculated by the following equation:

$$P(v_i = a|y = h) = \frac{N_{iah}}{SM_h} \quad \text{, (Equation 14)}$$

where N_{iac} indicates the number of spliced ECG signals, with the ith feature equaling a in the hth class, and SM_h indicates the number of spliced ECG signals in the hth class. The predicted label of the original ECG signal is the average predicted label of spliced ECG signals.

Alternative models

In our study, the pattern of "feature engineering + classifier" was used to construct alternative models. In total, 114 features were extracted from the ECG signal. Ten features were defined on the nine individual lead signals (e.g., the mean value of P wave apex values), and thus 90 (10 × 9) scalar features were obtained. The remaining 24 features were associated with the recording time (e.g., position in the temporal dimension) when the key points of segments occurred. Feature selection was not performed in alternative models. One hundred fourteen features were followed by three classic classifiers, KNN, RF, and XGBoost. Default configuration was used for the three classifiers. Two deep-learning-based models were also used as alternative models, where the original ECG signal was inputted. The first one was the 1D CNN, which achieved state-of-the-art performance in the multiclassification of arrhythmia sub-types. The second one was the LSTM, which is proposed for classifying ECG signals.

Quantification and statistical analysis

In our study, recall (sensitivity) was used to measure the diagnostic performance of models on a certain arrhythmia. It is defined as the ratio of the true positives (TP) to the sum of the true positives and false negatives (FN):

$$Recall = \frac{TP}{TP+FN} \quad \text{. (Equation 15)}$$

The recall score ranges from 0 to 1, with 1 indicating perfect recall.

Moreover, the overall diagnostic performance of the models was measured by macro-recall, which was defined as the mean of recall over all individual classes. In macro-recall, each class has the same weight in calculating the average recall, instead of setting the weight according to the sample size of each class. Therefore, macro-recall is fair at the class level and is more sensitive to performance in the minority class. Recall at the 95% CI was calculated using non-parametric bootstrapping with 1,000 iterations. The linear correlation between the recall of the 17 types of arrhythmias and the data distribution or the inter-similarity of those was measured by the CC.

Experimental setup

In our study, the simulation outline parameter γ was set to 5 to maintain the contour of the ECG while reducing noise interference. Another two hyperparameters were determined by the hyperparameter optimization. The number of clusters in

the segment clustering was the first hyperparameter, which had 18 options: 3, 5, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, and 80. The number of original heartbeats in the splicing N_o was another hyperparameter to control the *SN*, which had four options: 2, 3, 4, and 5. Values of *SN* in the training dataset corresponding to the four options of N_o were 2,869,966, 3,021,293, 11,486,438, and 25,508,046. Analyzing such a large number of ECG signals is beyond the capabilities of our computational resources, so we set the upper limit of the number of spliced ECG signals in each arrhythmia to 1,500,000 in the training dataset. In the test dataset, for each original ECG signal, 20 spliced ECG signals were selected randomly, and the summary of the prediction of the 20 spliced ECG signals was used as the final prediction of the original ECG signal.

ETHICS APPROVAL

The GDPH ECG-Arrhythmia Dataset was collected at Guangdong Provincial People's Hospital, Guangzhou, Guangdong, China, from August 2014 to October 2021. Ethics committee approval was obtained, and consent was obtained from all participants or parents of the included children (Ethics no. KY-Q-2022-144-01).

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2023.100795.

ACKNOWLEDGMENTS

This study was funded by the National Key Research and Development Program of China (2019YFB1404803 to Huiying Liang), the General Program of National Natural Science Foundation of China (62076076 to Huiying Liang), the Excellent Young Scientists Fund (82122036 to Huiying Liang), the Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application (2022B1212010011), the Basic and Applied Basic Research Foundation of Guangdong Province (2022A1515110722 to L.H.), and the Guangdong Provincial Medical Science and Technology Research Fund Project (A2023011 to L.H.).

AUTHOR CONTRIBUTIONS

Study conception and design, L.H., S.H., Huiying Liang, and Heng Liang; data collection and analysis, L.H., S.H., Y.D., and H. Liu; interpretation of results, S.H., L.H., H. Liu, X.P., and Huiying Liang; manuscript preparation, L.H., S.H., Y.D., and J.Z.; manuscript review, D.L., X.C., H.Y., L.K., J.T., and X.L. All authors reviewed the results and approved the final version of the manuscript. L.H. and S.H. contributed equally to this work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

ChatGPT, Grammarly, and Google Translate were used in the writing process to improve the readability and language of the work. Those AI-assisted technologies were applied with the oversight and control of our authors, and we carefully reviewed and edited the generated result.

Received: November 23, 2022
Revised: March 6, 2023
Accepted: June 16, 2023
Published: July 12, 2023

REFERENCES

1. Lippi, G., Sanchis-Gomar, F., and Cervellin, G. (2021). Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. Int. J. Stroke *16*, 217–221. https://doi.org/10.1177/1747493019897870.
2. Srinivasan, N.T., and Schilling, R.J. (2018). Sudden cardiac death and arrhythmias. Arrhythm. Electrophysiol. Rev. *7*, 111–117. https://doi.org/10.15420/aer.2018.15.2.
3. Schläpfer, J., and Wellens, H.J. (2017). Computer-Interpreted Electrocardiograms: Benefits and Limitations. J. Am. Coll. Cardiol. *70*, 1183–1192. https://doi.org/10.1016/j.jacc.2017.07.723.
4. Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444. https://doi.org/10.1038/nature14539.
5. Ebrahimi, Z., Loni, M., Daneshtalab, M., and Gharehbaghi, A. (2020). A review on deep learning methods for ECG arrhythmia classification. In Expert Systems with Applications: X, *7*. https://doi.org/10.1016/j.eswax.2020.100033.
6. Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., and Ng, A.Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat. Med. *25*, 65–69. https://doi.org/10.1038/s41591-018-0268-3.
7. Yildirim, Ö. (2018). A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification. Comput. Biol. Med. *96*, 189–202. https://doi.org/10.1016/j.compbiomed.2018.03.016.
8. Sannino, G., and de Pietro, G. (2018). A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. Future Generat. Comput. Syst. *86*, 446–455. https://doi.org/10.1016/j.future.2018.03.057.
9. Sayantan, G., Kien, K.P., and Kadambari, V.K. (2018). Classification of ECG beats using deep belief network and active learning. Med. Biol. Eng. Comput. *56*, 1887–1898. https://doi.org/10.1007/s11517-018-1815-2.
10. Wang, G., Zhang, C., Liu, Y., Yang, H., Fu, D., Wang, H., and Zhang, P. (2019). A global and updatable ECG beat classification system based on recurrent neural networks and active learning. Inf. Sci. *501*, 523–542. https://doi.org/10.1016/j.ins.2018.06.062.
11. Golany, T., and Radinsky, K. (2019). PGANs: Personalized generative adversarial networks for ECG synthesis to improve patient-specific deep ECG classification. In 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019. https://doi.org/10.1609/aaai.v33i01.3301557.
12. Nonaka, N., and Seita, J. (2022). RandECG: Data Augmentation for Deep Neural Network Based ECG Classification. In Advances in Intelligent Systems and Computing, vol 1423. https://doi.org/10.1007/978-3-030-96451-1_16.
13. Raghu, A., Shanmugam, D., Pomerantsev, E., Guttag, J., and Stultz, C.M. (2022). Data Augmentation for Electrocardiograms. In Proceedings of the Conference on Health, Inference, and Learning Proceedings of Machine Learning Research, G. Flores, G.H. Chen, T. Pollard, J.C. Ho, and T. Naumann, eds. (PMLR), pp. 282–310.
14. Hou, B., Yang, J., Wang, P., and Yan, R. (2020). LSTM-Based Auto-Encoder Model for ECG Arrhythmias Classification. IEEE Trans. Instrum. Meas. *69*, 1232–1240. https://doi.org/10.1109/TIM.2019.2910342.
15. Salem, M., Taheri, S., and Yuan, J.S. (2018). ECG Arrhythmia Classification Using Transfer Learning from 2- Dimensional Deep CNN Features. In 2018 IEEE Biomedical Circuits and Systems Conference, BioCAS 2018 - Proceedings. https://doi.org/10.1109/BIOCAS.2018.8584808.
16. Saxena, R., Shobe, J.L., and McNaughton, B.L. (2022). Learning in deep neural networks and brains with similarity-weighted interleaved learning. Proc. Natl. Acad. Sci. USA *119*, e2115229119.
17. Lake, B.M., Salakhutdinov, R., and Tenenbaum, J.B. (2015). Human-level concept learning through probabilistic program induction. Science *350*, 1332–1338. https://doi.org/10.1126/science.aab3050.

18. Tenenbaum, J.B. (2000). Rules and similarity in concept learning. In Advances in Neural Information Processing Systems.
19. Weitnauer, E., Carvalho, P.F., Goldstone, R.L., and Ritter, H. (2013). Grouping by Similarity Helps Concept Learning. In Proceedings of the Thirty-Fifth Annual Conference of the Cognitive Science Society.
20. Dupre, A., Vincent, S., and Iaizzo, P.A. (2005). Basic ECG theory, recordings, and interpretation. In Handbook of Cardiac Anatomy, Physiology, and Devices. https://doi.org/10.1007/978-1-59259-835-9_15.
21. Tahri Sqalli, M., Al-Thani, D., Elshazly, M.B., Al-Hijji, M., Alahmadi, A., and Sqalli Houssaini, Y. (2022). Understanding Cardiology Practitioners' Interpretations of Electrocardiograms: An Eye-Tracking Study. JMIR Hum. Factors *9*, e34058. https://doi.org/10.2196/34058.
22. Hu, L., Huang, S., Liu, H., Du, Y., Zhao, J., Peng, X., Li, D., Chen, X., Yang, H., Kong, L., et al. (2022). Cardiologists-like computer-aided interpretation framework. https://doi.org/10.5281/zenodo.7902431.
23. Sakoe, H., and Chiba, S. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. IEEE Trans. Acoust. *26*, 43–49. https://doi.org/10.1109/TASSP.1978.1163055.
24. Keogh, E., and Ratanamahatana, C.A. (2005). Exact indexing of dynamic time warping. Knowl. Inf. Syst. *7*, 358–386. https://doi.org/10.1007/s10115-004-0154-9.
25. Fix, E., and Hodges, J.L. (1989). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. Int. Stat. Rev. *57*, 238. https://doi.org/10.2307/1403797.
26. Ho, T.K. (1995). Random decision forests. In Proceedings of the International Conference on Document Analysis and Recognition (ICDAR). https://doi.org/10.1109/ICDAR.1995.598994.
27. Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. https://doi.org/10.1145/2939672.2939785.
28. Jones, S.A. (2016). ECG Notes: Interpretation and Management Guide, 3rd Edition.
29. Liu, X., Wang, H., Li, Z., and Qin, L. (2021). Deep learning in ECG diagnosis: A review. Knowl. Based. Syst. *227*, 107187. https://doi.org/10.1016/j.knosys.2021.107187.
30. Ahmed, W., and Khalid, S. (2017). ECG signal processing for recognition of cardiovascular diseases: A survey. In 2016 6th International Conference on Innovative Computing Technology, INTECH 2016. https://doi.org/10.1109/INTECH.2016.7845089.
31. Kaplan Berkaya, S., Uysal, A.K., Sora Gunal, E., Ergin, S., Gunal, S., and Gulmezoglu, M.B. (2018). A survey on ECG analysis. Biomed. Signal Process Control *43*, 216–235. https://doi.org/10.1016/j.bspc.2018.03.003.
32. Lu, W., Jiang, J., Ma, L., Chen, H., Wu, H., Gong, M., Jiang, X., and Fan, M. (2021). An arrhythmia classification algorithm using C-LSTM in physiological parameters monitoring system under internet of health things environment. J. Ambient Intell. Humaniz. Comput. https://doi.org/10.1007/s12652-021-03456-7.
33. Weimann, K., and Conrad, T.O.F. (2021). Transfer learning for ECG classification. Sci. Rep. *11*, 5251. https://doi.org/10.1038/s41598-021-84374-8.
34. Mathews, S.M., Kambhamettu, C., and Barner, K.E. (2018). A novel application of deep learning for single-lead ECG classification. Comput. Biol. Med. *99*, 53–62. https://doi.org/10.1016/j.compbiomed.2018.05.013.
35. Saadatnejad, S., Oveisi, M., and Hashemi, M. (2020). LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices. IEEE J. Biomed. Health Inform. *24*, 515–523. https://doi.org/10.1109/JBHI.2019.2911367.
36. Qaisar, S.M., Khan, S.I., Dallet, D., Tadeusiewicz, R., and Pławiak, P. (2022). Signal-piloted processing metaheuristic optimization and wavelet decomposition based elucidation of arrhythmia for mobile healthcare. Biocybern. Biomed. Eng. *42*, 681–694. https://doi.org/10.1016/j.bbe.2022.05.006.
37. Mian Qaisar, S., I Khan, S., Srinivasan, K., Krichen, M., and Krichen, M. (2023). Arrhythmia classification using multirate processing metaheuristic optimization and variational mode decomposition. Journal of King Saud University - Computer and Information Sciences *35*, 26–37. https://doi.org/10.1016/j.jksuci.2022.05.009.
38. Rastogi, N., and Mehra, R. (2013). Analysis of Butterworth and Chebyshev Filters for ECG Denoising Using Wavelets. IOSR J. Electron. Commun. Eng. *6*, 37–44.
39. Chandrakar, B., Yadav, O.P., and Chandra, V.K. (2013). a Survey of Noise Removal Techniques for Ecg Signals. Ijarcc *2*.
40. Ranjeet, K., Kuamr, A., and Pandey, R.K. (2012). ECG signal compression using optimum wavelet filter bank based on kaiser window. Procedia Eng. *38*, 2889–2902. https://doi.org/10.1016/j.proeng.2012.06.338.
41. Qin, Q., Li, J., Yue, Y., and Liu, C. (2017). An Adaptive and Time-Efficient ECG R-Peak Detection Algorithm. J. Healthc. Eng. *2017*, 5980541. https://doi.org/10.1155/2017/5980541.
42. Kew, H.P., and Jeong, D.U. (2011). Variable threshold method for ECG R-peak detection. J. Med. Syst. *35*, 1085–1094. https://doi.org/10.1007/s10916-011-9745-7.
43. Wang, Y., Wang, L., Chen, X., and Zhu, W. (2016). P Wave Detection and Delineation Based on Distances Transform. In 2016 IEEE Trustcom/BigDataSE/ISPA, pp. 2197–2201.
44. Hu, L., Liang, H., and Lu, L. (2021). Splicing learning: A novel few-shot learning approach. Inf. Sci. *552*, 17–28.
45. Hu, L., Liang, H., Tang, J., Li, X., Huang, L., and Lu, L. (2022). Cutting-splicing data augmentation: a novel technology for medical image segmentation. Preprint at arXiv. https://doi.org/10.48550/arXiv.2210.09099.