Multi-Label Classification of Arrhythmia for Long-Term Electrocardiogram Signals With Feature Learning

Yuwen Li, *Member, IEEE*, Zhimin Zhang, *Member, IEEE*, Fan Zhou, Yantao Xing, Jianqing Li, *Member, IEEE*, and Chengyu Liu, *Senior Member, IEEE*

***Abstract*— Arrhythmia is a kind of cardiovascular disease that seriously threats human health. Intelligent analysis of electrocardiogram (ECG) is an effective method for the early prevention and precise treatment to arrhythmia. In clinical ECG waveforms, it is common to see the multi-label phenomenon that one patient would be labeled with multiple types of arrhythmia. However, the current research is mainly to use the multiclass methods to solve the multi-label problem, ignoring the correlations between diseases and causing information loss. Therefore, this article aims: 1) to propose a multi-label feature selection method based on ECG (MS-ECG) and design an evaluation criterion of ECG features based on kernelized fuzzy rough sets so as to choose the optimal feature subset and optimize ECG feature space and 2) to propose the multi-label classification algorithm of arrhythmia based on ECG (MC-ECG) by establishing a multiobjective optimization model. This algorithm based on sparsity constraint explores the correlations between arrhythmia diseases and analyzes the mapping relationship between ECG features and arrhythmia diseases, so that one ECG signal would be automatically and accurately given multiple labels. Through sufficient experiments to prove the feasibility of our methods, we obtain the selected feature subset composed of 23 ECG features by MS-ECG. For the six evaluation criterions of MC-ECG, average precision is 0.8462, hamming loss is 0.1041, ranking loss is 0.1313, one-error is 0.2023, coverage is 0.4015, and micro-F1 is 0.6088. The outcome presents optimal to the current algorithms.**

***Index Terms*— Arrhythmia, biomedical information processing, electrocardiogram (ECG) signals, feature selection, multi-label classification.**

I. Introduction

ARRHYTHMIA is one of the most common and extremely high incidence cardiovascular diseases, which even causes syncope and sudden death. Electrocardiogram (ECG) signals intuitively reflect the generation and conduction process of cardiac electrical excitation in the conduction system, and it can reflect the physiological conditions of the heart to a certain extent [1]. ECG is one of the most common methods for the detection of arrhythmia. With the rise of wearable ECG monitoring devices [2], more and more ECG machines come into being. The automatic analysis of ECG signals can not only help doctors improve work efficiency, but also be the key to early warning and early prevention of abnormal ECG.

In recent years, machine learning (ML) made a great breakthrough in the automatic diagnosis of ECG signals [3]–[5]. Domestic and foreign researchers hope to complete automatic diagnosis of ECG signals with high accuracy and high speed. Therefore, it is an urgent problem to establish an intelligent classification model for arrhythmias using ML methods to improve the accuracy of ECG classification [7]. Although the development of ML algorithms has brought new development to the traditional classification algorithm of arrhythmia, there are two important problems need to be solved.

1) How to solve the problem of multi-label with ECG data? One patient may correspond to multiple arrhythmia types. This is known as "multi-label phenomenon." Multi-label problem is common in clinical ECG database. There are obvious differences between the multi-label problem and the multiclass problem. In multiclass problem, the relationship between the patient and the label is one-to-one. In multi-label problem, the relationship between the patient and the label is one-to-many, as shown in Fig. 1. Therefore, how to build a ML model to deal with the multi-label problem in arrhythmia classification task needs to be further studied.
2) How to deal with the high-dimensional problem of ECG features? When ECG signals are mapped to the feature

Manuscript received January 17, 2021; revised March 29, 2021; accepted April 16, 2021. Date of publication June 8, 2021; date of current version August 23, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62001111; in part by the Aviation Science Foundation Project under Grant 20200029069001; in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20200364, Grant BK20190014, and Grant BK20192004; in part by the National Natural Science Foundations of China under Grant 81871444 and Grant 62071241; and in part by the National Key Research and Development Program of China under Grant 2019YFE0113800. The Associate Editor coordinating the review process was Dr. Octavian Adrian Postolache. *(Corresponding authors: Zhimin Zhang; Chengyu Liu.)*

Yuwen Li is with the School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China, and also with the Aviation Key Laboratory of Science and Technology on Life-Support Technology, Xiangyang 441000, China (e-mail: liyuwen@seu.edu.cn).

Zhimin Zhang is with the Science and Technology on Information Systems Engineering Laboratory, The 28th Research Institute of CETC, Nanjing 210007, China (e-mail: zmzsdu@163.com).

Fan Zhou, Yantao Xing, Jianqing Li, and Chengyu Liu are with the School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: fanz1996@seu.edu.cn; 230198304@seu.edu.cn; ljq@seu.edu.cn; chengyu@seu.edu.cn).

Digital Object Identifier 10.1109/TIM.2021.3077667

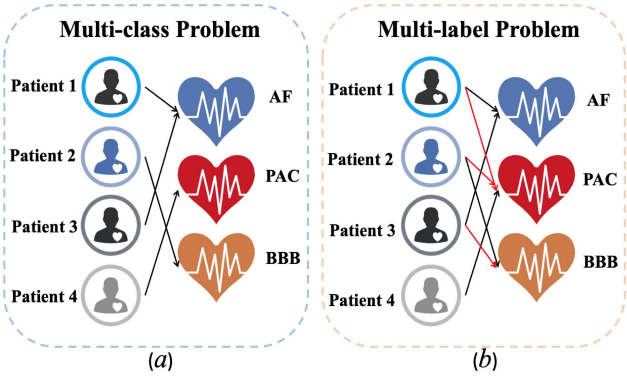


Fig. 1. Difference between (a) multiclass problem and (b) multi-label problem.

space, they often still have high dimension. In addition, the original ECG features obtained by different ECG feature extraction methods may be redundant or irrelevant for the arrhythmia classification task. Redundant features may lead to high computational complexity and the curse of dimensionality. Irrelevant features may confuse classification algorithms and reduce learning performance. Therefore, it is necessary to carry out feature selection before the classification of arrhythmias.

This article presents a framework (Fig. 2) for multi-label feature selection and multi-label classification of Arrhythmia. This work solves the problem of multiple arrhythmia labels in clinical ECG signals by proposing multi-label feature selection and multi-label classification framework, based on which ECG signal could be automatically and accurately given multiple labels. This framework can realize synchronous processing of multi-label problem and high-dimensional problem. Furthermore, the practicability and effectiveness of the proposed framework are proved to be superior than other ECG classification methods. The contributions of this article are summarized as follows.

1) Proposing an integrated framework. An effective framework was proposed for multi-label classification of Arrhythmia with feature selection and multi-label classification integrated.
2) Obtaining disease-specific features. Specific features related to each disease were analyzed and selected so that discriminative information could be acquired and the features were highly interpretable.
3) Incorporating diseases correlations. Considering diseases are not independent and related to each other, we tried to mine this correlation and incorporate it into our model.

II. Related Work

Arrhythmia classification algorithm based on ML is a hot research topic [8]–[14]. These algorithms can be divided into two categories: 1) classification algorithms based on traditional ML [8]–[11] and 2) classification algorithms based on deep learning (DL) [12]–[14].

For traditional ML classification algorithms, Raj and Ray [8] extracted time-frequency features by discrete cosine transform, and classified the ECG signals using support vector machine (SVM). Avdelazez *et al.* [9] utilized the wavelet transform, empirical mode decomposition, discrete cosine transforms, and statistical methods to obtain ECG features, and detected atrial fibrillation by random forest. Sadhukhan *et al.* [10] considered the changes of ECG waveform are reflected in the phase distribution pattern of the Fourier harmonics and conducted harmonic phase distribution pattern of ECG data to identify myocardial infarction. Banerjee and Mitra [11] analyzed ECG data to explore the spectral differences and used cross wavelet transform for the classification of ECG signals. For the classification algorithms based on DL, Hou *et al.* [12] used long–short-term memory (LSTM) method to conduct the classification of five heartbeats types, including normal, atrial premature complexes, left bundle branch block, right bundle branch block, and premature ventricular contractions. Taji *et al.* [13] proposed deep belief networks method to detect atrial fibrillation from poor-quality ECG signals, and the accuracy increased to 81%. Feng *et al.* [14] extracted features by frequential stacked sparse autoencoder with unsupervised learning, and then they proposed time-dependent cost-sensitive classification model to analyze ECG signals. In 2019, Hannun *et al.* [1] published a study in *Nature Medicine*. His team proposed a DL model about the residual neural network for accurately to detect ten types of arrhythmia, sinus rhythm, and noise by a single lead ECG. It can be seen from the literature survey that the ECG classification models based on ML have excellent performance, especially the models based on DL have reached the most advanced level in the classification learning task. However, DL is widely criticized due to the lack of interpretability. The main reason is that there are parameter sharing and complex feature extraction in the DL model, as shown in Fig. 3. Interpretability is crucial in the analysis of ECG signals. The task of ECG classification is not only to obtain high classification accuracy, but also to assist doctors to understand the ECG features. Further, ECG classification model could allow doctors to more transparently understand why this model makes such decisions and what features play an important role in this decision. Therefore, it is very important to study multi-label feature learning and multi-label classification of arrhythmias.

This article is an extension of the proceeding article [15]. In our previous article [15], we designed multi-label feature selection algorithm of ECG signals (MS-ECG), as shown in Fig. 2. The performance of MS-ECG had been preliminarily vitrificated by the experiments. We pushed forward the study and have presented an improved algorithm about multi-label classification. There are four major innovations and improvements compared to [15].

First, we design a comprehensive framework for long-term ECG signals, mainly composed of multi-label feature selection and multi-label classification, to realize synchronously handling the problem of multi-label and high-dimensional of ECG features. In comparison, only multi-label feature selection algorithm (MS-ECG) was preliminary performed to handle high-dimensional problem in [15].

Second, we further establish an effective multi-label classification model of arrhythmia for ECG signals (MC-ECG).

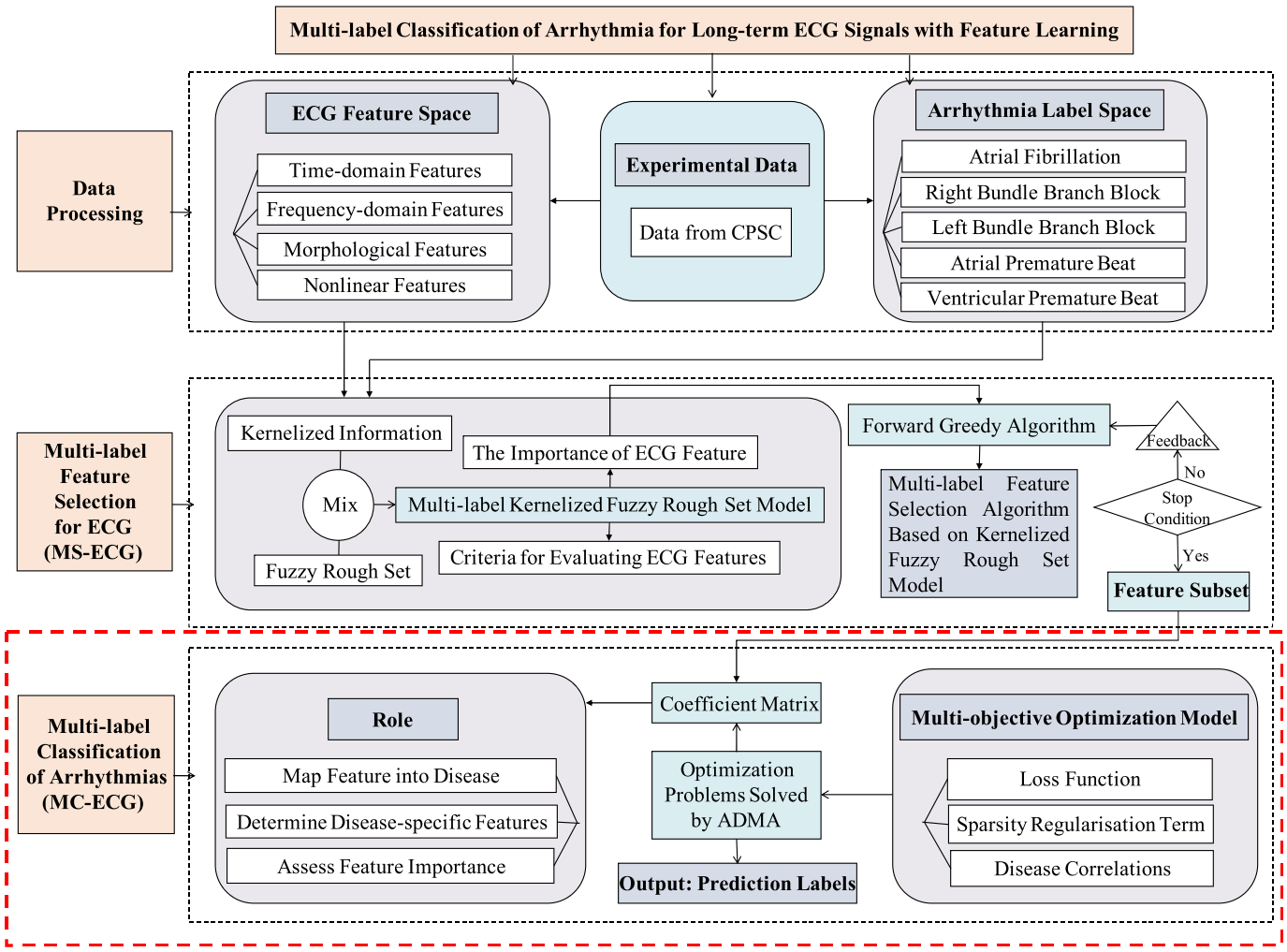


Fig. 2. Proposed framework for multi-label feature selection and multi-label classification of Arrhythmia.

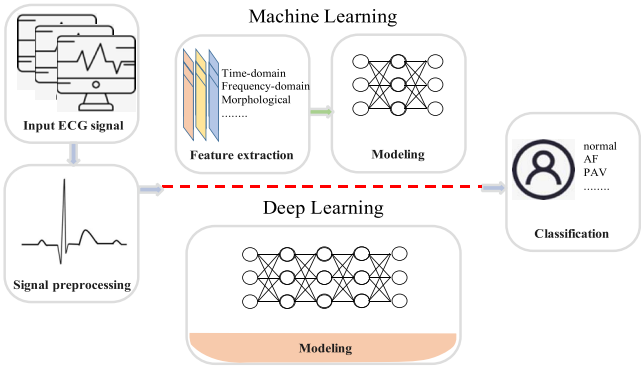


Fig. 3. Traditional ML and DL modeling steps of ECG signals.

MC-ECG is constructed by multiobjective optimization based on sparse regularization factor. First of all, each sub-objective in the classification task was determined before analyzing the relationships between sub-objectives. Then, we design the optimal function model of each sub-objective, and the correlations between abnormal ECG diseases is mined, so that the multiobjective optimization function model could be constructed. Finally, the sparse coefficient matrix is obtained by solving the optimization problem. Because the sparse coefficient matrix indicates the mapping relationship between ECG features and the corresponding diseases, MC-ECG can realize the multi-label prediction for a long-term ECG signal.

Third, although we obtained the optimal feature subset by MS-ECG in [15], the specific features corresponding to each arrhythmia type have not been studied. Therefore, another extended work in this article is that the specific ECG features corresponding to each arrhythmia type are reselected in the sparse coefficient matrix obtained from MC-ECG.

At last, extensive and comparative experiments are further performed to test the proposed framework. In addition to the results of the previous MS-ECG, the performance evaluation of the proposed MC-ECG is extended by comparing with other multi-label classification algorithms.

In summary, the framework presented in this work has been significantly improved in the classification of arrhythmia compared to the previous work [15]. This framework has promising potential for arrhythmia classification.

III. Database

An independent, open access dataset from 11 different hospitals was provided by the China Physiological Signal Challenge (CPSC) [16]. This database contained two parts:

TABLE I
RATE OF RECORDINGS FOR EACH LABEL TYPE

Type	Training Rate / Test Rate
Normal	18%
AF	24%
LBBB	5%
RBBB	36%
PAC	12%
PVC	14%

1) CPSC 2018 Training Set (open online) and 2) CPSC 2018 Test Set (unopen). Twelve leads ECG recordings sampled as 500 Hz last from 6 s to just 60 s. In this article, we use the first lead recordings. This dataset is multi-label data that includes nine types of ECG signals, and a few recordings have up to three labels. In this, we research multi-label classification for six types of ECG signals, including normal type and five types of common arrhythmia, such as atrial fibrillation (AF), right bundle branch block (RBBB), left bundle branch block (LBBB), premature atrial contraction (PAC), and premature ventricular contraction (PVC). In addition to the normal types, five types of arrhythmia all have high prevalence, among which the prevalence rate of AF is 11%∼15%, LBBB and RBBB are 5%∼7%, PAC is 5%∼7%, and PVC is 14%∼16% [2].

This article adopts 5078 ECG recordings with normal type or five types of arrhythmia from CPSC 2018 Training Set as training dataset. The rate of recordings for each label in training set is shown in Table I. This rate is obtained by calculating the proportion of the number of recordings for each type in this set to the total number in this set.

This article also adopts 2175 ECG recordings with normal type or five types of arrhythmia from CPSC 2018 Test Set as test dataset. The rate for each label in test set is also shown in Table I. As is presented, the rate is the same in both training set and test set. This is because the CPSC dataset was divided into training and test sets with a random 70–30 training-test split. It is noted that the sum of all rates is greater than 100% due to the presence of multi-label phenomenon.

IV. METHOD OF MULTI-LABEL FEATURE SELECTION OF ARRHYTHMIA FOR ECG SIGNALS

In this section, we design multi-label feature selection algorithm based on ECG signals (MS-ECG) [15] to solve high-dimensional problem in intelligent annotation of ECG. MS-ECG can effectively select the optimal ECG feature subset by evaluating the importance of features.

A. Feature Extraction

Before feature extraction, the QRS-wave, P-wave, and T-wave positions of all ECG signal are first detected. In this article, the algorithm proposed by Shang *et al.* [17] and Suárez-León *et al.* [18] is used to detect the position of T-wave, Q-wave, S-wave, P-wave, and R-wave, and the open source algorithm is prepared by Datta *et al.* [19]. In addition, ECG signal with less than 12 RR intervals was excluded.

In this article, 117 features are extracted according to the position of P-wave, Q-wave, R-wave, S-wave, and T-wave [20]–[24]. These features are divided into four types, 27 time-domain features, 34 frequency-domain features, 30 morphological features, and 26 nonlinear features.

1) Time-Domain Features: Time domain features are statistical features obtained by the RR interval of ECG signal, including the maximum and minimum value of the RR interval, the median of heart rate, the root mean square of the difference between adjacent RR intervals, etc.

2) Frequency-Domain Features: Frequency-domain features are mainly based on the ECG signals with windows. Then the spectral parameters of the window signal are calculated, including the spectral center, the frequency of the center of mass, the wavelet transform coefficients, normalized low-frequency power and normalized high-frequency power, etc.

3) Morphological Features: Morphological features are mainly based on the position and amplitude of P-wave, Q-wave, R-wave, S-wave, and T-wave to calculate the depths of S-wave and Q-wave related to R-wave, ST slope, and the width of QRS, etc.

4) Nonlinear Features: Nonlinear features are computed by nonlinear method, such as, sample entropy, approximate entropy, fuzzy entropy, etc.

B. Problem Description

The above 117 ECG features form the ECG feature set $\mathcal{A} = \{a_1, a_2, \ldots, a_{117}\}$. All the ECG signals constitute the sample set $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ of the multi-label classification algorithm, and corresponding to it is the label set $\mathcal{D} = \{d_1, d_2, \ldots, d_6\}$ composed of six arrhythmia labels. In the training set $\mathcal{D} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)|1 \leq i \leq n\}$, $\boldsymbol{x}_i \in \mathbb{R}^{117}$ is a 117-dimension feature vector, and $\boldsymbol{y}_i \in \{0.1\}^6$ is 6-D binary label vector. If x_i has the label d_j, then $y_{ij} = 1$ for the vector $\boldsymbol{y}_i$, otherwise $y_{ij} = 0$.

C. MS-ECG

For ECG feature space, we use Gaussian kernel matrix [25] $\boldsymbol{M}_{\{a_i\}}$ to evaluate the similarity between ECG signals under the ECG feature $a_i \in \mathcal{A}$ $(i = 1.2, \ldots, 117)$. When other ECG feature $a_{j,j\neq i}$ is added, the combined way of $\boldsymbol{M}_{\{a_i\}}$ and $\boldsymbol{M}_{\{a_j\}}$ is minimum strategy. For label space, we use Match kernel matrix [25] $\boldsymbol{M}'_{\{d_i\}}$ to evaluate the label overlap ratio between ECG signals under the arrhythmia label $d_i \in \mathcal{D}$ $(i = 1.2, \ldots, 6)$. When other label $d_{j,j\neq i}$ is added, the combined way of $\boldsymbol{M}'_{\{d_i\}}$ and $\boldsymbol{M}'_{\{d_j\}}$ is sum strategy.

1) Combination Strategy of Two Spaces:

Definition 1: For a multi-label decision system $\langle \mathcal{X}, \mathcal{A} \cup \mathcal{D} \rangle$, $\boldsymbol{M}_F$ is the combined kernel matrix about ECG feature subset $F \subseteq \mathcal{A}$, and $\boldsymbol{M}_L$ is the combined kernel matrix in arrhythmia label space L. Let $\lambda \in [0.1)$ be the penalty factor. The combination of kernel matrix $\boldsymbol{M}^{\{F,L\}}$ between F and L is defined as

$$\boldsymbol{M}^{\{F,L\}} = \boldsymbol{M}_F - \lambda \tilde{\boldsymbol{M}}_L. \tag{1}$$

In (1), $\tilde{\boldsymbol{M}}_L = \boldsymbol{M}_L/6$. It is clear that all elements of $\boldsymbol{M}_F$ are in [0,1], while all elements of $\boldsymbol{M}_L$ are in [0,6]. Aiming to

achieve dimensionless, we use $\tilde{\boldsymbol{M}}_L = \boldsymbol{M}_L/6$ to amend $\boldsymbol{M}_L$. $\tilde{\boldsymbol{M}}_L$ represents the label overlap ratio between ECG signals in the arrhythmia label space L. $\boldsymbol{M}^{\{F,L\}}$ indicates the probability of becoming the nearest different classes' sample. $S(x_i)$ and $T(x_i)$ are proposed to be the sets of kernel different classes' samples and kernel same class's samples, respectively. $S(x_i)$ as well as $T(x_i)$ can be calculated by $\boldsymbol{M}^{\{F,L\}}$ [26] when setting the parameter μ.

2) Construction Multi-Label Kernelized Fuzzy Rough Set Model:

Definition 2: For $\mathcal{X}(x \in \mathcal{X})$, $S(x)$ and $T(x)$ are known. The lower and upper approximations are analyzed by

$$\underline{K^R}D(x) = \underset{u\in S(x)}{\text{mean}}\{1 - \boldsymbol{M}_F(x,u)\} \tag{2}$$

$$\overline{K}^R D(x) = \underset{u\in S(x)}{\text{mean}}\{\boldsymbol{M}_F(x,u)\}. \tag{3}$$

3) Importance of ECG Feature: For $F \subseteq \mathcal{A}$, $\underline{K_F^R}(L)(x)$ is assumed to be the lower approximation of $x \in \mathcal{X}$. Then the multi-label fuzzy positive region $P_F(L)$ can be evaluated as

$$P_F^R(L) = \sum_{x\in\mathcal{X}} \underline{K_F^R}(L)(x). \tag{4}$$

By the above definition of fuzzy positive region, we can express the multi-label fuzzy dependence function as

$$\gamma_F^R(L) = \frac{P_F^R(L)(x)}{|\mathcal{X}|} = \frac{\sum_{x\in\mathcal{X}} \underline{K_F^R}(L)(x)}{|\mathcal{X}|}. \tag{5}$$

From the following (6), we can calculate the importance of the any feature $f \in \mathcal{A} - F$

$$\text{Sig}^R(f,F,L) = \gamma_F^R(L) - \gamma_{F\cup f}^R(L). \tag{6}$$

4) Algorithm: In general, a feature selection algorithm contains two crucial factors, which are feature evaluation and search strategy. In this article, the significance of each feature is measured by (6), and the forward greedy algorithm is applied as the search strategy in MS-ECG.

V. Method of Multi-Label Classification of Arrhythmia for ECG Signals

In this section, we develop a multi-label classification method of arrhythmia for long-term ECG signals combining ECG feature learning (MC-ECG) to achieve the goal of one ECG signal output multiple disease labels at the same time. MC-ECG solves the multi-label problem that traditional ML algorithms cannot handle in the process of intelligent ECG classification.

A. Establish Multi-Label Classification Model Based on Multiobjective Optimization Method

It is assumed that the optimal low-dimensional feature subset $\mathcal{A}' = \{a'_1, a'_2, \ldots, a'_{23}\}$ can be obtained by the above MS-ECG. The reconstruction of ECG signal feature space is $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]^T \in \mathbb{R}^{n\times 23}$. The corresponding multi-label space of arrhythmia is $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n]^T \in \{0.1\}^{n\times 6}$.

By learning the coefficient matrix $\boldsymbol{C} = [\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_6] \in \mathbb{R}^{23\times 6}$ between ECG feature space $\boldsymbol{X}$ and multi-label arrhythmia space $\boldsymbol{Y}$, the mapping relationship between ECG features and the types of arrhythmia is explored. The coefficient matrix $\boldsymbol{C}$ should have three characteristics.

1) $\boldsymbol{C}$ can reveal the mapping relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$.
2) The specificity of the disease can be captured by the non-zero term of $\boldsymbol{C}$. $\boldsymbol{C}$ can generate disease-specific features of each type of arrhythmia. In the multi-label classification task of arrhythmias, different diseases have different characteristics. Each type of arrhythmia is highly dependent on only a few ECG features. Therefore, compared with the feature set $\mathcal{A}'$, the disease-specific features for each disease are sparse.
3) $\boldsymbol{C}$ should include the correlation between diseases.

In view of the comprehensive consideration of the above three characteristics, the following multiobjective optimization model is proposed:

$$\min_{\boldsymbol{C}} \text{loss}(\boldsymbol{C}) + \varphi\Phi(\boldsymbol{C}) + \psi\Psi(\boldsymbol{C}) \tag{7}$$

where $\text{loss}(\cdot)$ is a loss function, $\Phi(\cdot)$ is disease correlation, $\Psi(\cdot)$ is sparse regularization, and $\varphi > 0$, $\psi > 0$ are tradeoff parameters.

In the coefficient matrix $\boldsymbol{C}$, if the element $c_{ij} = 0$, it indicates that the ith ECG feature has no discrimination on the jth arrhythmia type. On the contrary, if $c_{ij} \neq 0$, it indicates that the ith ECG feature has discrimination on the jth arrhythmia type. It also indicates that the ith ECG feature is the disease-specific feature of the jth arrhythmia type. The higher the value of element c_{ij}, the more important the ith ECG feature is to the jth arrhythmia type. The smaller the value of element c_{ij}, the less important the ith ECG feature is to the jth arrhythmia type.

B. Embed Disease Correlation

The types of arrhythmias are not independent of each other. Two strongly related diseases share more features than two weakly related or unrelated diseases. For example, both AF and PVC present a certain degree of arrhythmias in the waveform. These two types of arrhythmia are easy to be misjudged. They are strongly related diseases and share the features about ECG rhythm.

Therefore, the multiobjective optimization model should be embedded the constraint condition of disease correlation. $\Phi(\boldsymbol{C})$ is designed as

$$\Phi(\boldsymbol{C}) = \frac{1}{2}\sum_{j=l}^{n}\sum_{i=l}^{n} b_{ij}\boldsymbol{c}_i^T\boldsymbol{c}_j = \frac{1}{2}\text{Tr}\left(\boldsymbol{B}\boldsymbol{C}^T\boldsymbol{C}\right) \tag{8}$$

where $\boldsymbol{B} = [b_{ij}]_{6\times 6}$ indicates the correlation information between two disease labels d_i and d_j. $\boldsymbol{B} \in \mathbb{R}^{6\times 6}$ is a symmetric matrix, in which b_{ij} is the entry of $\boldsymbol{B}$ that reflects the relationships between two arbitrary diseases. In the label space, cosine similarity was applied to assess the correlations between different diseases. $\boldsymbol{z}_i \in \mathbb{R}^n$ was denoted as the i-th column of $\boldsymbol{Y}$ so as to define the cosine similarity between two diseases. Thus, $\boldsymbol{Y} = [z_1, z_2, \ldots, z_6]$. It is noted that z_i indicates the distribution of the i-th disease over the training data. Consequently, the entry of the affine matrix could be

defined as follows:

$$b_{ij} = \cos(z_i, z_j) = \frac{\langle z_i, z_j \rangle}{|z_i| \cdot |z_j|} \tag{9}$$

where $i, j \in \{1.2, \ldots, 6\}$, and b_{ij} could be regarded as a weight in the regularization term. According to (8), a higher value of b_{ij} means highly correlated between the i-th and the j-th diseases. According to (7), more punishment will be given to $\boldsymbol{c}_i^T \boldsymbol{c}_j$ if b_{ij} is higher, and optimization will make $\boldsymbol{c}_i^T$ and $\boldsymbol{c}_j$ become more similar in columns and sparse in rows. To fully employ this constraint, the second term in (7) goes over the entire affinity matrix of the disease correlation. In this way, the correlations between different diseases are incorporated into the framework, helping to improve the multilabel classification.

Combined with (7) and (8), the optimization problem can be written as

$$\min_{\boldsymbol{C}} \text{loss}(\boldsymbol{C}) + \frac{\varphi}{2}\text{Tr}\big(\boldsymbol{B}\boldsymbol{C}^T\boldsymbol{C}\big) + \psi\Psi(\boldsymbol{C}). \tag{10}$$

C. Solve Multiobjective Optimization Problem

When solving the multiobjective sparse convex optimization problem, the traditional method often requires strong conditions to ensure the convergence of the iterative sequence. The convergence speed is slow and the time cost is high. Therefore, for the multiobjective optimization model, how to find the global optimal solution while ensuring the convergence speed is a challenge. In this article, we conduct the alternating direction multiplier method (alternating direction method of multipliers, ADMMs) [27], [28] to solve a convex optimization problem. By decomposing the coordination process, ADMM transforms the original multiobjective optimization problem into a global consistency problem. The global solution is obtained by solving subproblems alternately, and the convergence speed is fast.

First, the least square loss function and $\boldsymbol{l_1}$-norm are assumed as

$$\text{loss}(\boldsymbol{C}) = \frac{1}{2}\|\boldsymbol{X}\boldsymbol{C} - \boldsymbol{Y}\|_F^2 \tag{11}$$

$$\Psi(\boldsymbol{C}) = \|\boldsymbol{C}\|_1. \tag{12}$$

Therefore, combined with (11) and (12), (10) can be written as

$$\min_{\boldsymbol{C}} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{C} - \boldsymbol{Y}\|_F^2 + \frac{\varphi}{2}\text{Tr}(\boldsymbol{B}\boldsymbol{C}^T\boldsymbol{C}) + \psi\|\boldsymbol{C}\|_1. \tag{13}$$

Second, two auxiliary variables $\boldsymbol{G}$ and $\boldsymbol{H}$ are introduced. Equation (13) can be written as

$$\min_{\boldsymbol{G},\boldsymbol{H},\boldsymbol{C}} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{C} - \boldsymbol{Y}\|_F^2 + \frac{\varphi}{2}\text{Tr}(\boldsymbol{B}\boldsymbol{G}^T\boldsymbol{G}) + \psi\|\boldsymbol{C}\|_1 \quad \text{s.t. } \boldsymbol{G} = \boldsymbol{C}, \quad \boldsymbol{H} = \boldsymbol{C}. \tag{14}$$

Third, we transform (14) into its augmented Lagrangian function form (15)

$$\min_{\boldsymbol{G},\boldsymbol{H},\boldsymbol{C}} \text{loss}(\boldsymbol{C}) + \frac{\varphi}{2}\text{Tr}(\boldsymbol{B}\boldsymbol{G}^T\boldsymbol{G}) + \psi\Psi(\boldsymbol{H}) + \langle \boldsymbol{L}_1, \boldsymbol{G} - \boldsymbol{C}\rangle + \langle \boldsymbol{L}_2, \boldsymbol{H} - \boldsymbol{C}\rangle + \frac{\varepsilon}{2}\big(\|\boldsymbol{G} - \boldsymbol{C}\|_F^2 + \|\boldsymbol{H} - \boldsymbol{C}\|_F^2\big). \tag{15}$$

By solving the above problem step by step, in each iteration, $\boldsymbol{G}$, $\boldsymbol{H}$, and $\boldsymbol{C}$ will alternate one by one (updating one of the unknowns while keeping the others fixed), in the following order:

$$\boldsymbol{C} \to \boldsymbol{G} \to \boldsymbol{H}.$$

[Update $\boldsymbol{C}$]: In the $k+1$ iteration, $\boldsymbol{C}$ is described as

$$\min_{\boldsymbol{C}} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{C} - \boldsymbol{Y}\|_F^2 + \big\langle \boldsymbol{L}_1^k, \boldsymbol{G}^k - \boldsymbol{C}\big\rangle + \big\langle \boldsymbol{L}_2^k, \boldsymbol{H}^k - \boldsymbol{C}\big\rangle + \frac{\varepsilon^k}{2}\big(\|\boldsymbol{G}^k - \boldsymbol{C}\|_F^2 + \|\boldsymbol{H}^k - \boldsymbol{C}\|_F^2\big). \tag{16}$$

The above problem (16) can be solved by taking the gradient of its objective function with respect to $\boldsymbol{C}$ and setting it to zero. $\boldsymbol{C}^{k+1}$ is

$$(\boldsymbol{X}^T\boldsymbol{X} + 2\varepsilon^k\boldsymbol{I})^{-1}\big(\boldsymbol{X}^T\boldsymbol{Y} + \boldsymbol{L}_1^k + \boldsymbol{L}_2^k + \varepsilon^k(\boldsymbol{G}^k + \boldsymbol{H}^k)\big). \tag{17}$$

[Update $\boldsymbol{G}$]: $\boldsymbol{G}$ is described as

$$\min_{\boldsymbol{G}} \frac{\varphi}{2}\text{Tr}(\boldsymbol{B}\boldsymbol{G}^T\boldsymbol{G}) + \big\langle \boldsymbol{L}_1^k, \boldsymbol{G} - \boldsymbol{C}^{k+1}\big\rangle + \frac{\varepsilon^k}{2}\|\boldsymbol{G} - \boldsymbol{C}^{k+1}\|_F^2. \tag{18}$$

The above problem (18) can be solved by taking the gradient of its objective function with respect to $\boldsymbol{G}$ and setting it to zero

$$\boldsymbol{G}^{k+1} = (\varphi\boldsymbol{B} + \varepsilon^k\boldsymbol{I})^{-1}\big(\varepsilon^k\boldsymbol{C}^{k+1} - \boldsymbol{L}_1^k\big). \tag{19}$$

[Update $\boldsymbol{H}$]: $\boldsymbol{H}$ is described as

$$\min_{\boldsymbol{G}} \psi\|\boldsymbol{H}\|_1 + \langle \boldsymbol{L}_2^k, \boldsymbol{H} - \boldsymbol{C}^{k+1}\rangle + \frac{\varepsilon^k}{2}\|\boldsymbol{G} - \boldsymbol{C}^{k+1}\|_F^2. \tag{20}$$

$\boldsymbol{H}$ is

$$\boldsymbol{H}^{k+1} = W_{\frac{\psi}{\varepsilon^k}}\left[\boldsymbol{C}^{k+1} - \frac{\boldsymbol{L}_2^k}{\varepsilon^k}\right] \tag{21}$$

where W is a soft threshold operator.

[Update $\boldsymbol{L}_1, \boldsymbol{L}_2$]:

$$\boldsymbol{L}_1^{k+1} = \boldsymbol{L}_1^k + \varepsilon^k(\boldsymbol{G}^{k+1} - \boldsymbol{W}^{k+1}) \tag{22}$$

$$\boldsymbol{L}_2^{k+1} = \boldsymbol{L}_2^k + \varepsilon^k(\boldsymbol{H}^{k+1} - \boldsymbol{W}^{k+1}). \tag{23}$$

[Update ε]:

$$\varepsilon^{k+1} = \min(\varepsilon^{\max}, \rho\varepsilon^k), \quad \rho > 1. \tag{24}$$

By the above steps, the sparse coefficient matrix $\boldsymbol{C}$ is obtained to realize simultaneous prediction of multiple labels for unknown ECG signals. $\boldsymbol{C}$ indicates the mapping relationship between ECG features and the types of arrhythmia. Therefore, for an ECG signal whose label information is unknown, the labels of arrhythmia type can be automatically assigned after feature subset space $\boldsymbol{X}$ obtained through MS-ECG. In addition, disease-specific features of each type of arrhythmia can be obtained from $\boldsymbol{C}$.

VI. Experiments

In this article, the experiments can be divided into two parts: the verification of the effectiveness about MS-ECG and MC-ECG.

1) For MS-ECG, we compare it with the other three multi-label feature selection algorithms, namely MLNB [29], MDDM [30], and PMU [31]. In the meantime, MLKNN ($K = 10$) [32] is used to assess the feature selection performances of these algorithms. Moreover, to stay in consistency with these algorithms, the training and testing set are also utilized. The specific experimental steps are as follows. First, 117 features were taken as the input of four multi-label feature selection algorithms under training set. The output result is the optimal feature subset for each algorithm. Second, the four kinds of feature subsets are respectively taken as the input of MLKNN under test set. The output results are the occurrence probability corresponding to each label as the basis for label prediction. Finally, various metrics are used to evaluate the performance of the multi-label feature selection algorithms.
2) For MC-ECG, we compare it with the following state-of-the-art multi-label classification algorithms to illustrate the performance of classification, including MLKNN [32], BR [33], CC [34], RankSVM [35], and LIFT [36]. The specific experimental steps are as follows. First, we take the feature subset obtained by MS-ECG as the input of six multi-label classification algorithms. Second, these six algorithms train the mapping relationship between features and labels under training set, respectively. For example, the sparse coefficient matrix $\boldsymbol{C}$ is obtained by MC-ECG. Then, under test set, the output results of six multi-label classification algorithms are the label prediction. Finally, various metrics are used to evaluate the classification performance of the MC-ECG algorithms.

A. Training Setup

For the four multi-label feature selection algorithms, MS-ECG and MLNB are able to get the feature subset directly. Nevertheless, MDDM and PMU can only get the feature selection output by obtaining the ranking list of features. Hence, to gain comparable results, the feature rank results will be cut according to the length of feature subset of MS-ECG. For the comparative methods, the parameter values of each algorithm are used as the default settings according to the corresponding literatures.

The values of λ and μ are set from 0.1 to 0.5 with a step of 0.01 for our proposed MS-ECG, so as to prove that the performance varies with λ and μ. The experiments imply that when $\lambda = 0.03$ and $\mu = 0.38$, the classification performance achieved the best for all evaluation metrics. For MC-ECG, we tune parameters φ, ψ in $\{2^{-10}, 2^{-9}, \ldots, 2^{10}\}$ and $\rho > 1$ as the default settings and report the best results.

B. Evaluation Metrics

Referring to the existing evaluation criteria for multi-label learning performance, the differences between the predicted labels ($y' \in Y_i'$) and the labels of ground-truth ($y \in Y_i$) of each sample ($\boldsymbol{x}_i$) are checkout respectively. The most classic six multi-label evaluation indicators are adopted [37], including Hamming loss, ranking loss, one-error, coverage, average precision, and micro-F1. These six indicators can be calculated by the following formulas, in which n represents the total number of samples and q represents the total number of labels.

1) Hamming Loss:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{q}|h(\boldsymbol{x}_i)\Delta Y_i| \tag{25}$$

where Δ is used to measure the symmetry difference between two sets. For example, the symmetry difference between $\{1, 2, 3\}$ and $\{3, 4\}$ is $\{1, 2, 4\}$. This index indicates the misclassification of the sample on a single label, i.e., the absence of relevant labels in the predicted label set or the presence of irrelevant labels in the predicted label set.

2) Ranking Loss:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{|Y_i||\bar{Y}_i|}|\{(y', y'')|f(\boldsymbol{x}_i, y') \le f(\boldsymbol{x}_i, y'')| \tag{26}$$

where $\bar{Y}_i$ is the complementary set of Y_i in the label space, and $(y', y'') \in Y_i \times \bar{Y}_i$. This index is used to investigate the occurrence of sorting errors in the sorting sequence of category labels of samples, that is, irrelevant labels are placed before relevant labels in the sorting sequence. The smaller this index, the better the system performance will be, and the optimal value is 0.

3) One-Error:

$$\frac{1}{n}\sum_{i=1}^{n}|[\text{argmax}_{y\in Y_i} f(\boldsymbol{x}_i, y)] \notin Y_i| \tag{27}$$

where $|\cdot|$ stands for conditional judgment. If the condition is satisfied, it is 1. While if not, it is 0. The index is used to investigate the situation that the label at the front of the sequence does not belong to the related label set in the sorting sequence of category label of samples. The smaller this index, the better the performance will be, and the optimal value is 0.

4) Coverage:

$$\frac{1}{n}\sum_{i=1}^{n}\max_{y\in Y_i}\text{rank}_f(\boldsymbol{x}_i, y) - 1 \tag{28}$$

where $\text{rank}_f(\cdot, \cdot)$ is the sorting function corresponding to the real-valued function $f(\cdot, \cdot)$. This index is used to investigate the search depth required to cover all related labels in the sorting sequence of category labels of the sample. The smaller this index, the better the performance will be.

5) Average Precision:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{|Y_i|}\sum_{y\in Y_i}\frac{|\{y' \,|\, \text{rank}_f(\boldsymbol{x}_i, y') \le \text{rank}_f(\boldsymbol{x}_i, y), y' \in Y_i\}|}{\text{rank}_f(\boldsymbol{x}_i, y)}. \tag{29}$$

This index is used to investigate the situation that the label before the relevant label is still the relevant label in the sorting sequence of category labels of samples. The larger the value of

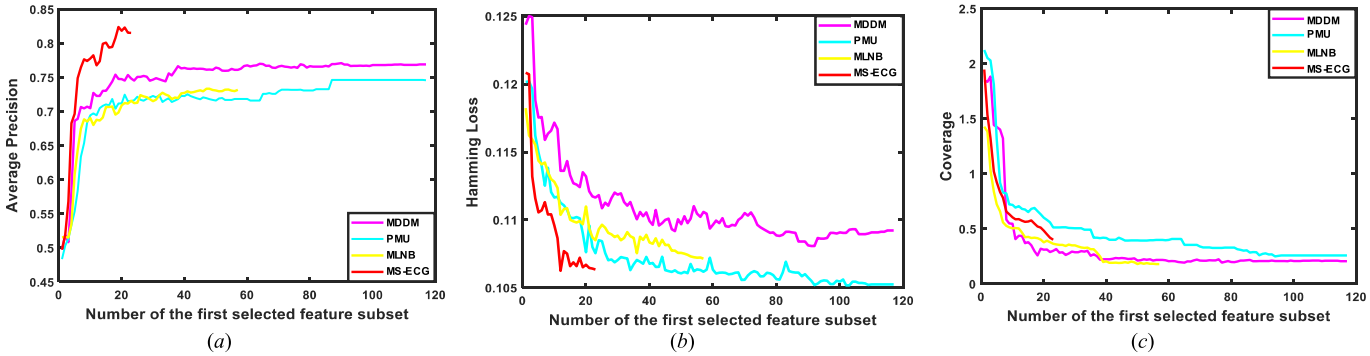


Fig. 4. Performance variations for the four algorithms based on (a) average precision, (b) Hamming loss, and (c) coverage.

this index, the better the performance will be, and the optimal value is 1.

6) Micro-F1:

$$\frac{2\sum_{l=1}^{q}\mathrm{TP}_l}{2\sum_{l=1}^{q}\mathrm{TP}_l+\sum_{l=1}^{q}\mathrm{FN}_l+2\sum_{l=1}^{q}\mathrm{FP}_l} \tag{30}$$

where TP_l represents the number of samples that belong to the lth label and are also correctly classified as the lth label. FP_l is the number of samples that do not belong to the lth label but are classified as the lth label. FN_l is the number of samples that belong to the lth label but are not classified as the lth label. The larger the evaluation criterion, the better the system performance. When the value is 1, the system performance will reach the optimal.

VII. Results

A. MS-ECG

For the comparability of performances among all multi-label feature selection algorithms, the feature subset is fed to MLKNN as input. Fig. 4 shows the variation trend with the number of selected features through the six evaluation criteria. We take average precision, hamming loss, and coverage as examples for show. In Fig. 4, horizontal axis of each graph indicates the number of the chosen features, while vertical axis of each graph indicates classification evaluation metrics. The proposed algorithm MS-ECG is the red line. As shown in Fig. 4, MS-ECG achieves the optimal classification performance with the increasing number of selected features.

Notice that MS-ECG has a certain number of features that makes it delivers better performance, and it is in line with the actual situation. We select the top 23 features and form the optimal feature subset [15]. These 23 ECG features include four features in time-domain, three in frequency-domain, 12 in nonlinear, and four in morphological domain.

To prove the validity of MS-ECG, a series of experiments are conducted which make quantitative comparisons among MS-ECG, MLNB, MDDM, and PMU. As is shown in Table II, various evaluation metrics are illustrated with the optimal results being emphasized in bold. Moreover, "↓" means "the smaller the better," while "↑" conveys that "the larger the better," respectively. From Table II, it is clearly illustrated that MS-ECG outperforms other comparative algorithms in five evaluation metrics except coverage.

TABLE II
Comparison Among Four Multi-Label Feature Selection Algorithms

Evaluation Metrics	Algorithms			
	MDDM	***PMU***	***MLNB***	***MS-ECG***
Average Precision (↑)	0.7483	0.7155	0.7159	**0.8053**
Hamming Loss (↓)	0.1116	0.1075	0.1085	**0.1063**
Ranking Loss (↓)	0.1401	0.1523	0.1472	**0.1366**
One-error (↓)	0.2219	0.2342	0.2558	**0.2021**
Coverage (↓)	**0.3001**	0.5131	0.3646	0.4018
Micro-F1 (↑)	0.5621	0.5633	0.5125	**0.5874**

B. MC-ECG

To verify the effectiveness of MC-ECG, we take the feature subset obtained from MS-ECG as the input of six multi-label classification algorithms in Table III, including MLKNN, BR, CC, RankSVM, LIFT, and MC-ECG. The classification performance is assessed in six evaluation metrics as shown in the Table III. It is clearly illustrated that the proposed MC-ECG is superior in four metrics, including average precision, Hamming loss, ranking loss, and micro-F1, except that the best One-error is given by MLKNN and the best Coverage is given by LIFT.

In fact, these six metrics measure the experimental results from different aspects. Hamming loss and micro-F1 reflect the prediction of labels. Average precision, ranking loss, coverage, and one-error reflect the ranking of labels. It is difficult for an algorithm to achieve optimal for all metrics [37]. To consider various evaluation indexes, we draw a spider web to present the comprehensive performance of different multi-label classification algorithms for the six criteria, as is shown in Fig. 5. Due to the large differences in predictive classification performance using different evaluation metrics, the results in Table III are normalized to 0.1∼0.5, and all metrics are transformed to achieve the best performance at 0.5. When the spider web diagram shows rounder and bigger, it means that the comprehensive performance of the corresponding classification algorithm is better. Specifically, the performance of MC-ECG is represented by the red

TABLE III
COMPARISON AMONG SIX MULTI-LABEL CLASSIFICATION ALGORITHMS

Evaluation Metrics	Algorithms					
	MLKNN	***BR***	***CC***	***RankSVM***	***LIFT***	***MC-ECG***
Average Precision (↑)	0.8053±0.0042	0.7258±0.0076	0.7404±0.0045	0.7491±0.0066	0.8175±0.0087	**0.8462±0.0051**
Hamming Loss (↓)	0.1063±0.0016	0.1090±0.0021	0.1087±0.0010	0.1083±0.0019	0.1065±0.0024	**0.1041±0.0011**
Ranking Loss (↓)	0.1366±0.0067	0.1405±0.0053	0.1389±0.0064	0.1371±0.0053	0.1402±0.0091	**0.1313±0.0087**
One-error (↓)	**0.2021±0.0054**	0.2452±0.0110	0.2539±0.0078	0.2545±0.0061	0.2531±0.0104	0.2023±0.0041
Coverage (↓)	0.4018±0.0082	0.4281±0.0058	0.4222±0.0065	0.4367±0.0069	**0.4009±0.0070**	0.4015±0.0065
Micro-F1 (↑)	0.5874±0.0098	0.4176±0.0102	0.5502±0.0074	0.5173±0.0101	0.5921±0.0091	**0.6088±0.0076**

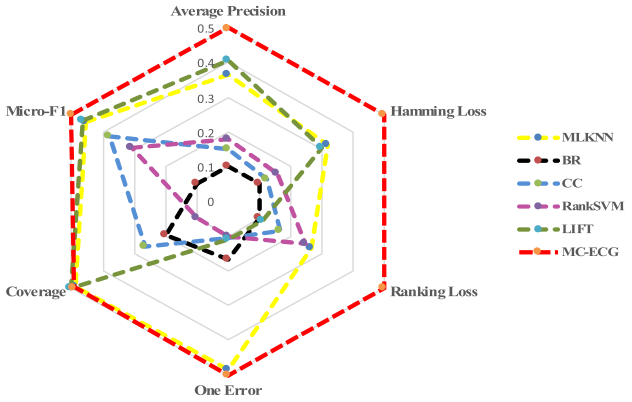


Fig. 5. Spider web diagram showing the comprehensive performance with different evaluation metrics.

dotted line. We can tell from Fig. 5 that MC-ECG is superior to the other methods from a comprehensive point of view.

VIII. DISCUSSION

A. Analysis of Superiority of MS-ECG

In MS-ECG, we used multiple kernel method to deal with multi-label problem. It can adaptively establish the relationship between the input and output of training samples. Moreover, multiple kernel learning can map different data by different kernel functions and different kernel parameters. Therefore, we constructed multi-label kernelized fuzzy rough set model to explore the association pattern between ECG features and the types of arrhythmia. By MS-ECG, we defined the importance of ECG feature and measured the degree of interdependence between features and labels to build a consistent and low-dimensional feature subset.

B. Analysis of Superiority of MC-ECG

The essence of MC-ECG is to reveal different physiological and pathological mechanisms corresponding to different ECG features. It explores the mapping relationship between ECG features and the types of arrhythmia. When the unknown ECG signal is used as the input, the ECG signal can be automatically labeled through the mapping relationship that has been learned. In MC-ECG, we established a multiobjective optimization model for long-term ECG signals. Besides, the correlations between types of arrhythmia are discovered to learn extra latent information and model a practical and stable relationship. Finally, we achieved the goal of simultaneously outputting multiple predicted labels for an ECG signal.

C. Comparison With DL Model

Based on CPSC 2018, DL algorithms [38]–[40] are efficient algorithms to classify ECG automatically. He *et al.* [38] used deep neural network (DNN) and LSTM to extract features from raw ECG signals. The extracted features are concatenated to form a feature vector which is trained to do the final classification. Wang *et al.* [39] used convolutional neural network (CNN) for arrhythmia detection. Chen *et al.* [40] also developed a CNN model to detect and classify cardiac arrhythmias. DL algorithms have high classification accuracy. However, comparing with DL algorithms, the proposed framework has two advantages: 1) it performs well in dealing with multi-label classification problem by predicting multiple labels for each ECG recording simultaneously, making full use of multi-label information and 2) our algorithm conducts feature learning to analyze some specific features containing discriminative information for each disease and possesses advantage with respect to interpretability.

IX. CONCLUSION

In this article, we established a framework for multi-label classification of Arrhythmia with feature selection and multi-label classification integrated. This framework realizes synchronously handling the problem of multi-label and high-dimensional of ECG features. Moreover, this framework breaks through the technical limitations that the existing research only use multiclass method cannot give multiple labels for an ECG at the same time. After the multi-label feature selection for ECG, a multiobjective optimization classification model was established to achieve multi-label classification of ECG signals. In this article, only traditional ML algorithms are discussed and compared. However, more DL algorithms with increasing practical applications need to be verified. In the future, we will put forward a more targeted optimization to get more suitable for real-time, automatic, and accurate detection of arrhythmia in dynamic ECG feature set and multi-label algorithm.

REFERENCES

[1] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, no. 1, pp. 65–69, Jan. 2019.

[2] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, "Video-based heart rate measurement: Recent advances and future prospects," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 10, pp. 3600–3615, Oct. 2019.

[3] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019.

[4] H. Blackburn, A. Keys, E. Simonson, P. Rautaharju, and S. Punsar, "The electrocardiogram in population studies: A classification system," *Circulation*, vol. 21, no. 6, pp. 1160–1175, 1960.

[5] C. Y. Liu *et al.*, "Signal quality assessment and lightweight QRS detection for wearable ECG SmartVest system," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1363–1374, Apr. 2019.

[6] X. Wang, C. Liu, Y. Li, X. Cheng, J. Li, and G. D. Clifford, "Temporal-framing adaptive network for heart sound segmentation without prior knowledge of state duration," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 2, pp. 650–663, Feb. 2021, doi: 10.1109/TBME.2020.3010241.

[7] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Inf. Sci.*, vol. 405, pp. 81–90, Sep. 2017.

[8] S. Raj and K. C. Ray, "ECG signal analysis using DCT-based DOST and PSO optimized SVM," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 3, pp. 470–478, Mar. 2017.

[9] M. Abdelazez, S. Rajan, and A. D. C. Chan, "Detection of atrial fibrillation in compressively sensed electrocardiogram measurements," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 2502209, doi: 10.1109/TIM.2020.3027930.

[10] D. Sadhukhan, S. Pal, and M. Mitra, "Automated identification of myocardial infarction using harmonic phase distribution pattern of ECG data," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 10, pp. 2303–2313, Oct. 2018.

[11] S. Banerjee and M. Mitra, "Application of cross wavelet transform for ECG pattern analysis and classification," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 2, pp. 326–333, Feb. 2014.

[12] B. Hou, J. Yang, P. Wang, and R. Yan, "LSTM-based auto-encoder model for ECG arrhythmias classification," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1232–1240, Apr. 2020.

[13] B. Taji, A. D. C. Chan, and S. Shirmohammadi, "False alarm reduction in atrial fibrillation detection using deep belief networks," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 5, pp. 1124–1131, May 2018.

[14] K. Feng, H. Qin, S. Wu, W. Pan, and G. Liu, "A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 4000912, doi: 10.1109/TIM.2020.3017246.

[15] Y. W. Li, Z. M. Zhang, and F. Zhou, "Multi-label feature selection for long-term electrocardiogram signals," in *Proc. Int. Conf. Sens., Meas. Data Anal. era Artif. Intell. (ICSMD)*, Oct. 2020, pp. 335–340.

[16] F. Liu *et al.*, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *J. Med. Imag. Health Informat.*, vol. 8, no. 7, pp. 1368–1373, Sep. 2018.

[17] H. Shang, S. Wei, F. Liu, D. Wei, L. Chen, and C. Liu, "An improved sliding window area method for T wave detection," *Comput. Math. Methods Med.*, vol. 2019, pp. 1–11, Apr. 2019.

[18] A. A. Suárez-León, C. Varon, R. Willems, S. Van Huffel, and C. R. Vázquez-Seisdedos, "T-wave end detection using neural networks and support vector machines," *Comput. Biol. Med.*, vol. 96, pp. 116–127, May 2018.

[19] S. Datta *et al.*, "Identifying normal, AF and other abnormal ECG rhythms using a cascaded binary classifier," in *Proc. Comput. Cardiol.*, Sep. 2017, pp. 1–4.

[20] C. Liu *et al.*, "A comparison of entropy approaches for AF discrimination," *Physiol. Meas.*, vol. 39, no. 7, pp. 74002-1–74002-18, 2018.

[21] M. Bsoul, H. Minn, and L. Tamil, "Apnea MedAssist: Real-time sleep apnea monitor using single-lead ECG," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 3, pp. 416–427, May 2011.

[22] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Nat. Acad. Sci. USA*, vol. 88, pp. 2297–2301, Mar. 1991.

[23] G. Y. Bin *et al.*, "Detection of atrial fibrillation using decision tree ensemble," in *Proc. Comput. Cardiol.*, Sep. 2017, pp. 1–4.

[24] S. Sarkar, D. Ritscher, and R. Mehra, "A detector for a chronic implantable atrial tachyarrhythmia monitor," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1219–1224, Mar. 2008.

[25] B. Schlkopf and A. Smola, "Learning with kernels," in *Proc. 21th Int. Conf. Mach. Learn.*, 2001, pp. 639–646.

[26] Y. W. Li *et al.*, "Feature selection for multi-label learning based on kernelized fuzzy rough sets," *Neurocomputing*, vol. 318, pp. 217–286, Nov. 2018.

[27] Y. Li, W. Chen, D. Liu, Z. Zhang, S. Wu, and C. Liu, "IFFLC: An integrated framework of feature learning and classification for multiple diagnosis codes assignment," *IEEE Access*, vol. 7, pp. 36810–36818, 2019.

[28] C. Lu, J. Feng, S. Yan, and Z. Lin, "A unified alternating direction method of multipliers by majorization minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 527–541, Mar. 2018.

[29] M.-L. Zhang, J. M. Peña, and V. Robles, "Feature selection for multi-label naive Bayes classification," *Inf. Sci.*, vol. 179, no. 19, pp. 3218–3229, Sep. 2009.

[30] Y. Zhang and Z.-H. Zhou, "Multilabel dimensionality reduction via dependence maximization," *ACM Trans. Knowl. Discovery Data*, vol. 4, no. 3, pp. 1–21, Oct. 2010.

[31] J. Lee and D.-W. Kim, "Feature selection for multi-label classification using multivariate mutual information," *Pattern Recognit. Lett.*, vol. 34, no. 3, pp. 349–357, Feb. 2013.

[32] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007.

[33] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognit.*, vol. 37, no. 9, pp. 1757–1771, Sep. 2004.

[34] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Mach. Learn.*, vol. 85, no. 3, pp. 333–359, Dec. 2011.

[35] A. Elisseeff and J.Weston, "A kernel method for multi-labelled classification," in *Proc. 14th Int. Conf. Neural Inf. Process. Syst.*, 2001, pp. 681–687.

[36] M.-L. Zhang and L. Wu, "Lift: Multi-label learning with label-specific features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 1, pp. 107–120, Jan. 2015.

[37] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.

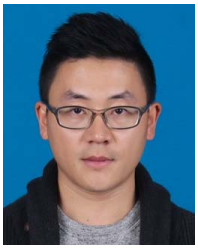
[38] R. He *et al.*, "Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional LSTM," *IEEE Access*, vol. 7, pp. 102119–102135, 2019.

[39] R. Wang, J. Fan, and Y. Li, "Deep multi-scale fusion neural network for multi-class arrhythmia detection," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2461–2472, Sep. 2020.

[40] T.-M. Chen, C.-H. Huang, E. S. C. Shih, Y.-F. Hu, and M.-J. Hwang, "Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model," *iScience*, vol. 23, no. 3, Mar. 2020, Art. no. 100886.

Yuwen Li (Member, IEEE) received the Ph.D. degree from the Department of Automation, Xiamen University, Xiamen, China, in 2019.

She is currently a Lecturer with the School of Instrument Science and Engineering, Southeast University, Nanjing, China. Her research interests include data mining, granular computing, machine learning, and big data processing for physiological signals.



Zhimin Zhang (Member, IEEE) received the B.S. and Ph.D. degrees in biomedical engineering from Shandong University, Jinan, China, in 2014 and 2020, respectively.

He is currently working at the Science and Technology on Information Systems Engineering Laboratory, The 28th Research Institute of CETC. His research topics include ECG, EEG signals for cardiovascular disease, and sleep stages classification.

Fan Zhou received the B.S. degree from Southwest Jiaotong University, Chengdu, China, in 2018. She is currently pursuing the M.S. degree with the School of Instrument Science and Engineering, Southeast University, Nanjing, China.

Her current research interests include signal processing and atrial fibrillation detection.

Yantao Xing received the B.S. degree in measurement and control technology and instrument from the School of Instrument and Electronics, North University of China, Taiyuan, China, in 2015, and the M.S. degree from the School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2018. He is currently pursuing the Ph.D. degree in instrument science and technology with Southeast University, Nanjing.

His current research interests include mHealth, physiological signal processing, and intelligent monitoring system.

Jianqing Li (Member, IEEE) received the B.S. and M.S. degrees in automatic control and the Ph.D. degree in instrument science and technology from Southeast University, Nanjing, China, in 1986, 1990, and 2000, respectively.

He is currently a Professor and the Vice Present of the School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing. He is also a Professor with the School of Instrument Science and Engineering, Southeast University. His research interests include mHealth and wireless networks.

Chengyu Liu (Senior Member, IEEE) received the Ph.D. degree in biomedical engineering from Shandong University, Jinan, China, in 2010. He has completed the post-doctoral trainings at Shandong University from 2010 to 2013; Newcastle University, Newcastle upon Tyne, U.K., from 2013 to 2014; and Emory University, Atlanta, GA, USA, from 2015 to 2017. He is currently the Director and a Professor of Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Laboratory, School of Instrument Science and Engineering, Southeast University, Nanjing, China. His main research direction include intelligent monitoring and analysis of vital signs signals (ECG, blood pressure, pulse, respiration, and heart sounds), medical big data processing and machine learning, and development of wearable active health monitoring device.

Dr. Liu is also a Federation Journal Committee Member of the International Federation for Medical and Biological Engineering (IFMBE).