Computers in Biology and Medicine 166 (2023) 107503

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

MVKT-ECG: Efficient single-lead ECG classification for multi-label arrhythmia by multi-view knowledge transferring

Yuzhen Qin [a,1], Li Sun [b,1], Hui Chen [b], Wenming Yang [a], Wei-Qiang Zhang [b], Jintao Fei [c], Guijin Wang [b,*]

[a] Shenzhen International Graduate School, Tsinghua University, Shenzhen 518071, China
[b] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[c] Beijing Tsinghua Changgung Hospital, Beijing 102218, China

ARTICLE INFO

Keywords:
Single-lead Electrocardiogram
Arrhythmia classification
Deep learning
Knowledge transfer
Neural network

ABSTRACT

Electrocardiogram (ECG) is a widely used technique for diagnosing cardiovascular disease. The widespread emergence of smart ECG devices has sparked the demand for intelligent single-lead ECG-based diagnostic systems. However, it is challenging to develop a single-lead-based ECG interpretation model for multiple disease diagnosis due to the lack of some key disease information. We aim to improve the diagnostic capabilities of single-lead ECG for multi-label disease classification in a new teacher-student manner, where the teacher trained by multi-lead ECG educates a student who observes only single-lead ECG We present a new disease-aware *Contrastive Lead-information Transferring* (CLT) to improve the mutual disease information between the single-lead-based ECG interpretation model and multi-lead-based ECG interpretation model. Moreover, We modify the traditional Knowledge Distillation into *Multi-label disease Knowledge Distillation* (MKD) to make it applicable for multi-label disease diagnosis. The whole knowledge transferring process is inter-lead **M**ulti-**V**iew **K**nowledge **T**ransferring of ECG (MVKT-ECG). By employing the training strategy, we can effectively transfer comprehensive disease knowledge from various views of ECG, such as the 12-lead ECG, to a single-lead-based ECG interpretation model. This enables the model to extract intricate details from single-lead ECG signals and enhances the model's capability of diagnosing and identifying single-lead signals. Extensive experiments on two commonly used public multi-label datasets, ICBEB2018 and PTB-XL demonstrate that our MVKT-ECG yields exceptional diagnostic performance improvements for single-lead ECG. The student outperforms its baseline observably on the PTB-XL dataset (1.3 % on PTB.super, and 1.4 % on PTB.sub), and on ICBEB2018 dataset (3.2 %).

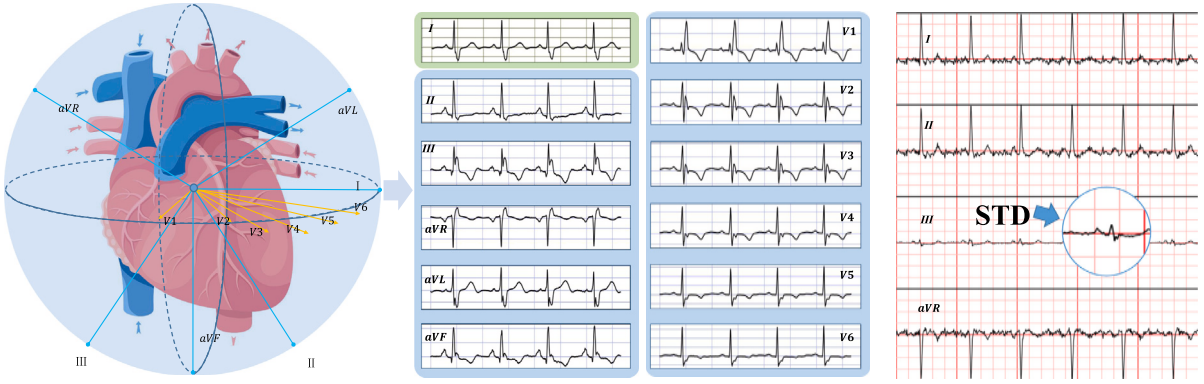
1. Introduction

Electrocardiogram (ECG) is a commonly used, non-invasive, and convenient diagnostic method for detecting cardiac arrhythmias and other cardiovascular diseases [1]. It can be roughly divided into two categories: multi-lead (12-lead in clinical) ECG and single-lead ECG. Multi-lead ECG is acquired by multiple electrodes placed on the patient's chest wall and limbs, which can be seen as visualizing the heartbeat signals from multiple different views (Fig. 1(a)) [2,3]. Under the strong ability of deep learning, many multi-lead-based deep ECG interpretation models are widely used in practice. Previous models focused on extracting discriminative features [4–6]. Subsequently, the residual connection has been widely used in deep convolutional network structures to enhance the expressiveness of the model [5]. To exploit the properties of different leads, a series of deep neural networks are incorporated with an attention mechanism to establish inter-lead relationships [7,8]. Considering the scarcity of high-quality ECG datasets, many self-supervised approaches focus on mining effective information from unlabeled data [9–11]. However, multi-lead ECG data is only provided by an environment with strong observation equipment, such as a hospital [12]. These methods did not perform specific analysis on the features of single-lead ECG, nor did they thoroughly investigate the relationship among different leads. Therefore, the performance of previous methods was unsatisfactory when diagnosing multiple diseases from single-lead ECG.

Single-lead ECG signals have attracted much attention recently due to the emergence of smart ECG devices, such as the Apple Watch [13] and Alivecor [14], which bring great convenience to people. In

* Corresponding author.
E-mail address: wangguijin@tsinghua.edu.cn (G. Wang).
[1] These authors contributed equally to this work as first authors.

https://doi.org/10.1016/j.compbiomed.2023.107503
Received 6 June 2023; Received in revised form 14 August 2023; Accepted 15 September 2023
Available online 19 September 2023



(a) 12-lead ECG from 12 different viewpoints

(b) An ECG waveform with arrhythmia STD (ST-segment depression)

Fig. 1. (a): A standard 12-lead ECG signal by depicting heartbeat from 12 different views. Multi-lead ECG can record the heart's activity more comprehensively, as it includes rich disease information from various perspectives. (b): Lead III of this ECG signal shows the disease STD, while other leads are all normal (four leads are shown).

recent years, many single-lead ECG signals have been produced with the popularization of these ECG devices, raising high expectations for robust intelligent single-lead ECG diagnostic systems. However, the poor observation equipment always provides low-quality and single-view ECG signals [12]. Due to the lack of views, the research on the intelligent diagnosis of single-lead ECG still focuses on a few diseases [15,16], such as atrial fibrillation. To the best of our knowledge, few previous studies have addressed the multi-class multi-label classification problem of single-lead ECG. Single-lead ECG records the heart's electrical activity from only one perspective, which can be seen as a partial view of a 12-lead ECG. A large gap in arrhythmia diagnosis performance still subsists between single-lead ECG signals and multi-lead ECG signals, indicating that the number of leads plays a crucial role in achieving better results (see Table 2). In this study, we aim to address the multi-class multi-label diagnosis task of single-lead ECG signals.

Recently, the Knowledge Distillation [17] method has been studied in the field of single-lead ECG recognition. [18] proposed a method based on knowledge distillation to compress the single-lead model, but the nature of inter-lead information transferring was not explored, and this method can only diagnose ECG with a single label, unable to solve multi-label disease diagnostic questions. Rich Data Helps Poor Data via Imitation (RDPD) [12] can enhance the model trained on poor data using a teacher model trained on rich, private data. Since the student is trained to imitate the output space and internal attention maps of its teacher, maintaining strict consistency of the internal feature maps to be processed is essential, which makes the network structure lack flexibility. In addition, it is noted that mutual information plays a key role in improving single-lead ECG's potential. So lack of consideration of the information relationship between single-lead ECG and multi-lead ECG, just forcing the same output and internal feature maps or attention maps of the student with the teacher, does not guarantee the robustness of the student.

In this paper, to efficiently stimulate the potential of single-lead ECG signals, we propose a novel and efficient architecture to transfer multi-view information of ECG for multi-label disease classification — Multi-View Knowledge Transferring of ECG (MVKT-ECG). We adopt the teacher-student paradigm and extend it to a more general inter-lead knowledge transferring approach. Since mutual information is the key to improving single-lead ECG's ability, we formulate a new objective to maximize the mutual information of the feature representations between teacher and student, namely Contrastive Lead-information Transferring (CLT). The CLT can maximize a lower bound to the mutual information between the teacher and student representations. Furthermore, We modify traditional Knowledge Distillation to Multi-label Knowledge Distillation (MKD), intending to make it applicable to the context of multiple labels. To verify the effectiveness of our method, we conduct extensive experiments on two commonly used public datasets, PTB-XL [19] and ICBEB2018 [20]. The experimental results demonstrate the effectiveness and robustness of multi-view knowledge transferring in the single-lead ECG classification task. The contributions of this paper include the following:

- We propose a more general and efficient framework MVKT-ECG to transfer multi-view information from the multi-lead-based ECG interpretation model to the single-lead-based ECG interpretation model, which can effectively improve the network's diagnostic performance of single-lead ECG.
- We explore the nature of inter-lead knowledge transferring and design a novel inter-lead information transferring objective—Contrastive Lead Transferring (CLT), which is effective in maintaining the stability of the diagnosis with single-lead ECG signals and reducing the probability of misdiagnosis of single-lead ECG signals.
- We formulate Multi-label Knowledge Distillation (MKD) to fit the context of more general multi-label disease detection, which is an important component of MVKT-ECG.
- We perform intensive experiments using various advanced backbones on two freely accessible 12-lead ECG-waveform datasets and compare various performance metrics. The results are also compared with several state-of-the-art single-lead ECG classification methods, verifying our proposed approach's effectiveness.

2. Related work

2.1. Deep learning on arrhythmias

Recently, deep learning has become the mainstream method of arrhythmia classification task research, in which a large majority of methods focus on designing the network structure suitable for ECG signals to extract more expressive ECG representations. [4] applied the adaptive one-dimensional Convolutional Neural Network (CNN) to the heartbeat classifier for the first time, which combined feature extraction and classification into one learning body. [21] firstly proposed a comprehensive evaluation of an end-to-end deep neural network for ECG analysis across various diagnostic classes, demonstrating that an end-to-end deep learning approach can classify a broad range of distinct arrhythmias from single-lead ECGs with high diagnostic performance similar to that of cardiologists. Subsequently, residual connection [22]

has been widely used in deeper convolutional network structures to increase the expressiveness of the model [5]. The above studies used one-dimensional CNNs, [23] converted one-dimensional ECG into two-dimensional images and used a two-dimensional convolution network for feature extraction.

Since ECGs are essentially sequential signals, Recurrent Neural Network (RNN) can model the sequential information of ECG signals. [6] used a Long short-term Memory module (LSTM) to extract rhythm information from Long ECG fragments. Meanwhile, many works [24–26] combined CNN and RNN to synthesize the advantages of both structures. As many types of arrhythmias are episodic, showing abnormalities only on local ECGs, the Attention Mechanism can selectively focus on inputs that are more relevant to the task while attenuating others. Therefore, there are also many studies [27–29] integrated attention mechanism in networks to capture abnormal waveforms more accurately and improve diagnostic performance.

However, these studies did not enhance the intelligent diagnosis of single-lead electrocardiograms according to their features, nor did they investigate the relationship among different leads thoroughly. Therefore, the performance of previous methods was unsatisfactory when diagnosing multiple diseases from single-lead ECG.

2.2. Research on single-lead ECG

Due to the limited information provided by single lead ECGs, all kinds of smart ECG devices can currently only support a few types of detection. For example, the Apple Watch can only detect sinus rhythm(SR) and atrial fibrillation(AF) [15], and most academic research has also focused on that. Existing studies have achieved high accuracy in detecting AF [16,30–32]. Some studies have shown neural networks can diagnose myocardial infarction disease from single-lead ECG [33,34]. However, these studies are limited to the diagnosis of a single disease by single-lead ECG. Limited by input information, the single-lead-based model performs poorly in classifying other arrhythmias.

Many models that achieved top scores in ECG competitions can also serve as classifiers for single-lead ECGs. Hannun's CNN-based model [21] achieved the best score in the Physionet Challenge 2017 competition. The DenseCNN_Wang [35], and Resnet1d_Liu [36] achieved the top performance in China ECG AI Contest 2019 competition. [37] proposed a large kernel size model SEresnet_Zhao based on SE-block [38], achieving second place in the PhysioNet 2020 competition [39]. However, these methods still did not adapt better to the defects of single-lead ECG, and the performance was still unsatisfactory when diagnosing multiple diseases from single-lead electrocardiograms.

To address this problem of lacking ECG views, people proposed the method of reconstruction and synthesis of ECG. [40] restored another nine leads based on noisy three-lead signals through LSTM. [41] transformed the reconstruction of missing leads into solving the least squares problem by mapping the known ECG leads to the corresponding Koopman space and using the theory that the Koopman operator is linear. [2] proposed the concept of ECG panorama and the corresponding generation network Nef-Net and further proposed disease-aware synthesis method — ME-GAN, which achieves panoptic electrocardio representations conditioned on heart diseases. These methods can restructure or synthesize multi-view ECG signals. However, the lack of evaluation on downstream classification tasks makes it difficult to guarantee the quality of the reconstructed signals for disease detection.

There are also studies based on knowledge distillation. [18] used the traditional method of KD, forcing the student to imitate the output logits and internal features of the teacher. However, it cannot handle the multi-label classification task. Different from this work, we propose a new distillation framework for multi-label classification (MKD) to address this deficiency. To deploy models in the poor-data environment without requiring direct access to multi-modal data acquired from a rich-data environment, [12] proposed Rich Data Helps Poor Data via Imitation (RDPD) to enhance a predictive model trained on poor data using knowledge distilled from a high-complexity model trained on rich, private data. Since the student is trained to imitate the output space and internal attention maps of its teacher, internal feature maps to be processed must maintain strict consistency, which makes the network structure lack flexibility. Different from this work, our proposed method can align teacher-student features of different dimensions through MLP and normalization layers, making knowledge transfer more applicable. In addition, these methods lack consideration of the information relationship between single-lead ECG and multi-lead ECG, just forcing the same output and internal features or attention maps of the student with the teacher, which does not guarantee the robustness of the student. So we explore the nature of inter-lead information transferring and formulate a novel inter-lead information transferring objective — Contrastive Lead Transferring(CLT).

3. Method

3.1. Architecture

Here we describe the framework of our method and introduce the notations used in this paper. Given a 12 leads ECG signal $\mathbf{X} = \left[\mathbf{x^{(1)}}, \mathbf{x^{(2)}}, \ldots, \mathbf{x^{(12)}}\right]$ or a single lead ECG signal $\boldsymbol{x} = \mathbf{x^{(i)}}$, where $\mathbf{X} \in \mathbb{R}^{12\times L}$ and $\boldsymbol{x} \in \mathbb{R}^{L}$ and L is the length of each data. Our models aim to learn a function $\mathcal{F}_\theta(\mathbf{X})$, mapping a set of ECGs $\left[\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_N}\right]$ to a representative feature space, where θ is network parameters. Then the classifier $\mathcal{C}_\theta(\cdot)$ projects the above features into category $\hat{Y} \in \mathbb{R}^{C}$, where C is the number of class. Considering the medical fact that sometimes patients with heart disease often have multiple arrhythmias, our task is a multi-class multi-label classification task. Thus the *sigmoid* function is used as the final activation function. Here, we hope that this kind of representation can (1) minimize the information gap between single-lead ECG and multi-lead ECG signals and (2) be robust to the lead reduction in the single-lead ECGs. As shown in Fig. 2, our proposed method is a two-stage procedure.

3.2. Training teacher model

We train the teacher network by standard 12 leads ECGs $\mathbf{X} = \left[\mathbf{x^{(1)}}, \mathbf{x^{(2)}}, \ldots, \mathbf{x^{(12)}}\right]$. The module $\mathcal{F}_\theta^T(\boldsymbol{X}) : \mathbb{R}^{12\times L} \rightarrow \mathbb{R}^{D}$ will map each ECG signal $\mathbf{X_i}$ to a fixed-size representation (in $xresnet1d101$ [42] $D = 2048$). Because ECG signals are time-series signals, we chose residual 1D CNN with mean pooling as the encoder on each sample $\mathbf{X_i}$ as given in Eq. (1)

$$\mathcal{F}_\theta(\mathbf{X_i}) = Pooling(CNN_{1d}(\mathbf{X_i})), \tag{1}$$

Then, the above features are projected into category $\hat{Y} \in \mathbb{R}^{C}$, as given in Eq. (2), where C is the number of classes. In order to make the information transferring path between teachers and students more flexible. The representations of teacher and student will be projected by MLP and a normalization layer to same dimension disease information carrier vectors $\mathbf{G}^T(\mathbf{X_i})$, as given in Eq. (3).

$$\hat{\boldsymbol{y}} = \mathcal{C}_\theta(\mathcal{F}_\theta(\mathbf{X_i})), \tag{2}$$

$$\mathbf{G}^T(\mathbf{X_i}) = Norm(MLP(\mathcal{F}_\theta(\mathbf{X_i}))), \tag{3}$$

We train the teacher network-which will observe 12-lead ECG using the multi-label classification term — Binary Cross Entropy Loss ($\mathcal{L}_{\mathrm{BCE}}$), which can be formulated as:

$$\mathcal{L}_{\mathrm{BCE}} = -\left[\boldsymbol{y}\log\hat{\boldsymbol{y}} + (1-\boldsymbol{y})\log(1-\hat{\boldsymbol{y}})\right], \tag{4}$$

where $\boldsymbol{y}$ represents the one-hot labels and $\hat{\boldsymbol{y}}$ represents the predicted classes probability.

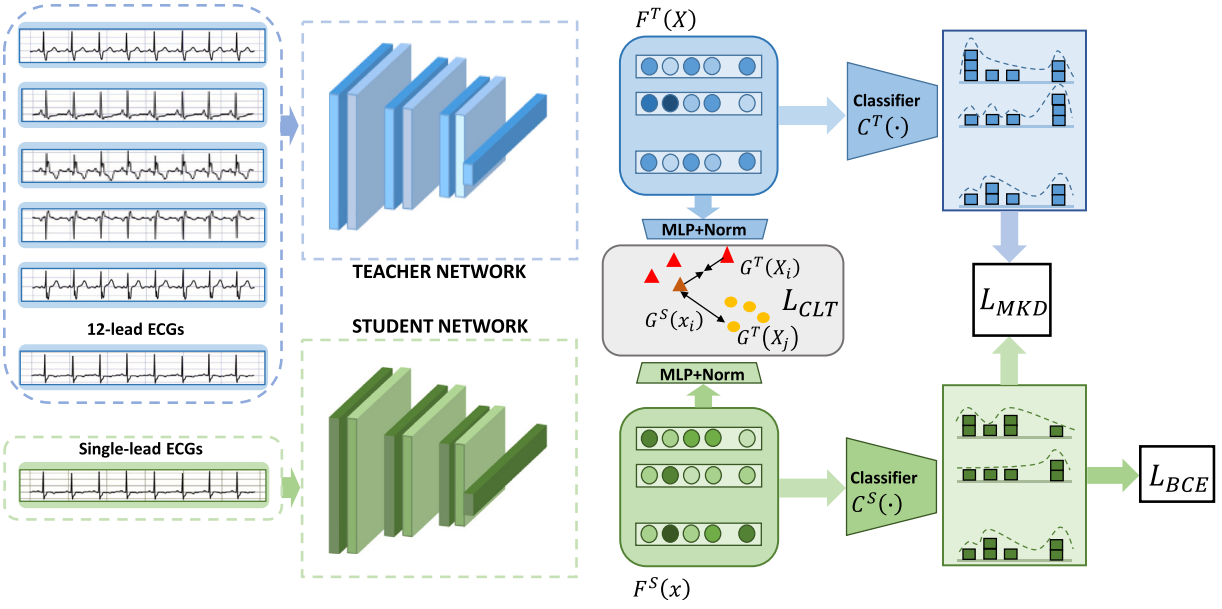


Fig. 2. An overview of multi-view knowledge transferring: The student observing single-lead ECG signals is optimized by the teacher observing 12-lead ECG signals and ground truth. The teacher network has already been trained with 12-lead ECGs, and the student network receives the multi-view knowledge in two ways: (a) minimizing $\mathcal{L}_{\text{CLT}}$ to improve the inter-lead disease mutual information and (b) matching the logits of the teacher by minimizing the $\mathcal{L}_{\text{MKD}}$.

3.3. Knowledge transferring

We use the teacher network trained in the first step to educate the new network, which only observes single-lead ECG signals, and the parameters of the teacher are frozen. During the knowledge transferring period, the representations of teacher and student will be projected by MLP and a normalization layer to same dimension vectors $\mathbf{G}^T(\mathbf{X})$ and $\mathbf{G}^S(\mathbf{X})$, as given in Eq. (3). We hope the representation ability of the network, which only observes single-lead ECG signals, can be improved. In order to achieve this, we aim at the information we can gather from different leads, depicting the same heart condition under different views. When faced with a 12-lead ECG classification task, various lead viewpoints can be exploited to provide a variety of appearances for a target disease. We aim to teach the network to recover as much full-lead information as possible from single-lead ECG signals. Although some diseases cannot be inferred from single-lead ECG signals, our goal is to shorten the information gap between the single-lead-based ECG interpretation model and multi-lead-based ECG interpretation model as much as possible, encouraging the student to focus more on key details of some particular diseases and maximize the mutual disease information between single-lead-based ECG interpretation model and multi-lead-based ECG interpretation model, further reducing the incidence of misdiagnosis. Typically, the teacher network is always more powerful than the student network due to the information gap between 12-lead ECG signals and single-lead ECG signals. This asymmetry between the teacher and the student can produce a distillation objective different from the one caused by the difference in model complexity. MVKT-ECG improves the mutual information between the feature representations of teacher and student by the Contrastive Lead-information Transferring, which is achieved by shortening the student model and teacher model's representation of the positive pairs and pushing apart the representation between "negative" pairs. In the process, we also allow the student to imitate the teacher's output by our Multi-label diseases Knowledge Distillation.

3.3.1. Contrastive lead-information transferring

To improve the inter-lead mutual information, we propose the disease-aware Contrastive Lead-information Transferring Loss, which can transfer useful disease information by maximizing the mutual information between single-lead ECG signals and multi-lead ECG signals.

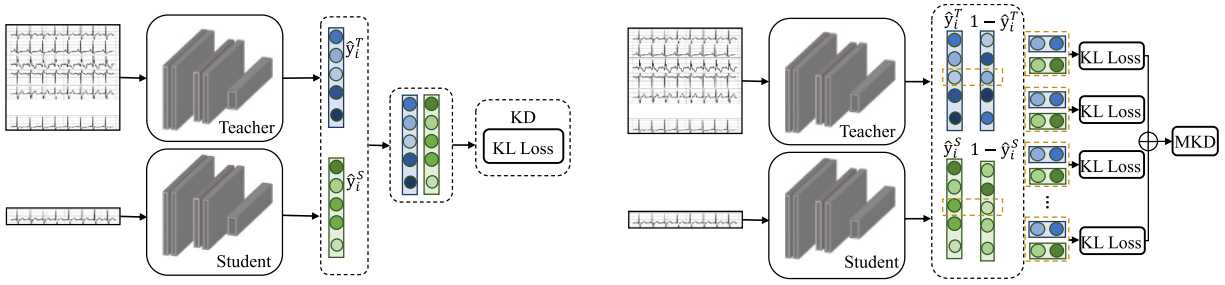
As we all know, due to the lack of data dimensions, the disease information extracted by a neural network from single-lead ECG signals is less compared with 12-lead ECG signals, but it still contains part disease information. We refer to this information as the mutual information between single-lead ECG and 12-lead ECG. In another aspect, the absence of available data for single-lead ECG signals will also trigger a false diagnosis. For example, abnormal T-wave morphology in any lead would be clinically diagnosed as a T-wave change. In the case of abnormal T-waves in other leads, judgment based on lead I alone is likely to lead to missed detection. We refer to this information as misleading information. We know the lower limit of mutual information [43]:

$$
\begin{aligned}
&I(\mathbf{G}^T(\mathbf{X}), \mathbf{G}^S(\mathbf{X})) \\
&= \sum_{i,j} p(\mathbf{G}^T(\mathbf{X_i}), \mathbf{G}^S(\mathbf{X_j})) \log \frac{p(\mathbf{G}^T(\mathbf{X_i}), \mathbf{G}^S(\mathbf{X_j}))}{p(\mathbf{G}^T(\mathbf{X_i})) \cdot p(\mathbf{G}^S(\mathbf{X_j}))} \\
&\geq \log(N) + \mathbb{E}\left[\log \frac{f(\mathbf{G}^T(\mathbf{X_i}), \mathbf{G}^S(\mathbf{X_j}))}{f(\mathbf{G}^T(\mathbf{X_i})) \cdot f(\mathbf{G}^S(\mathbf{X_j}))}\right] \\
&= \log(N) + \mathbb{E}\left[\log \frac{exp\left[\mathbf{G}^T(\mathbf{X_i})\mathbf{G}^S(\mathbf{X_j})/\tau\right]}{\sum_{i=1}^{N} exp\left[(\mathbf{G}^T(\mathbf{X_i})\mathbf{G}^S(\mathbf{X_j}))/\tau\right]}\right] \\
&= \log(N) - \mathcal{L}_{\text{CLT}}.
\end{aligned}
\tag{5}
$$

In detail, for samples X_i and X_j, MVKT-ECG shortens the student model and teacher model's representation of the same sample $\mathbf{G}^T(\mathbf{X_i})$ and $\mathbf{G}^S(\mathbf{X_j})$ and further the feature representation of the two models for different samples. This can be achieved by minimizing:

$$
\mathcal{L}_{\text{CLT}} = -\mathbb{E}\left[\log \frac{exp\left[(\mathbf{G}^T(\mathbf{X_i})\mathbf{G}^S(\mathbf{X_j}))/\tau\right]}{\sum_{i=1}^{N} exp\left[(\mathbf{G}^T(\mathbf{X_i})\mathbf{G}^S(\mathbf{X_j}))/\tau\right]}\right], \tag{6}
$$

where N is the number of positive and negative sample pairs; the larger, the better. τ is the temperature coefficient. Then, a memory buffer is used to store the features of all training samples in this study. Since the teacher has observed multi-view ECGs, we believe the different distances between "positive" and "negative" pairs yield a strong description of the corresponding disease identities. During the training, N samples are randomly selected to calculate the loss to meet the model's demand for the number of negative samples. The expression of each sample in the memory will carry on momentum updates according to the latest feature extracted from the network.



(a) Classical Knowledge Distillation (KD). (b) Multi-label diseases Knowledge Distillation (MKD).

Fig. 3. Illustration of Classical Knowledge Distillation (KD) and our Multi-label diseases Knowledge Distillation (KD). We separate the output probability $\hat{y}_i$ with $1-\hat{y}_i$, and for each sample, we calculate the fore-and-aft softmax by temperature coefficient τ_{kd}.

3.3.2. Multi-label diseases knowledge distillation

Traditional KD focuses on single-label classification problems, where the output logits of the classifier are softened by the temperature coefficient τ with softmax, which can only generate one target class Fig. 3(a). Therefore, this method only works for single-label classification tasks. Arrhythmia patients often suffer from multiple diseases. The arrhythmia detection task can often be abstracted as a multi-label classification task, so we need the sigmoid function as the output of the network's last layer. Therefore, the traditional distillation paradigm cannot be applied to multi-label classification. Considering the case that the output contains multiple target classes, we propose a method to soften the classification probability longitudinally, MKD (Fig. 3(b)), a new knowledge distillation objective that can process multi-label disease problems. Especially, we view a multi-label classification task as a combination of multiple binary classification tasks and separate the output probability $\hat{y}_i$ with $1-\hat{y}_i$, and for each sample, we calculate the fore-and-aft softmax by temperature coefficient τ_{kd},

$$q_i = \frac{e^{\hat{y}_i/\tau_{kd}}}{e^{\hat{y}_i/\tau_{kd}} + e^{(1-\hat{y}_i)/\tau_{kd}}}, \tag{7}$$

$$\mathcal{L}_{\mathrm{MKD}} = \tau_{kd}^2 \sum_i \mathrm{KL}\left(q_i^T \parallel q_i^S\right), \tag{8}$$

where τ_{kd} represents the temperature coefficient of multi-label disease knowledge distillation loss.

3.4. Joint optimization

In summary, the overall objective of MVKT-ECG combines the distillation terms ($\mathcal{L}_{\mathrm{MKD}}$), the Contrastive Lead-information Transferring terms ($\mathcal{L}_{\mathrm{CLT}}$), and the ones optimized by label-$\mathcal{L}_{\mathrm{BCE}}$, which provide a higher conditional likelihood w.r.t. ground truth labels. To sum up, MVKT aims to boost single lead ECG's representation ability by the following optimization problem:

$$\mathrm{argmin}_{\theta_S}\ \mathcal{L}_{\mathrm{MVKT}} = \mathcal{L}_{\mathrm{BCE}}^S + \alpha\mathcal{L}_{\mathrm{MKD}} + \beta\mathcal{L}_{\mathrm{CLT}}, \tag{9}$$

where θ_S represents the trainable parameters of the student model and α, β are hyperparameters balancing the contributions of $\mathcal{L}_{\mathrm{MKD}}$ and $\mathcal{L}_{\mathrm{CLT}}$ to the total loss $\mathcal{L}_{\mathrm{MVKT}}$. Regarding the student initialization, we found that the CLECG's self-supervised strategy [10] is beneficial, which will be mentioned in Section 5.5. Algorithm 1 displays the complete training procedure of our proposed method.

3.5. Implementation details

We adopt several models as the backbone of the teacher and student, including CNN_Hannun [21], ResNet1d34 [22], ResNet1d_wang [44], inception1d [45], and XresNet1d101 [46]. The input channels for teacher and student are 12 and 1, respectively. The dimensions of the feature vectors of CNN_Hannun, ResNet1d34, ResNet1d_wang, inception1d, and XresNet1d101 are 256, 512, 128, 128, and 2048, respectively. The dimension of the MLP output is 128, and the MLP consists of two fully connected layers. In each training round, 1024 samples are randomly selected from the memory storage area to calculate the loss-$\mathcal{L}_{\mathrm{CLT}}$. The negative samples are extracted with the student output and the teacher output as the anchor points, respectively. The loss is composed of two symmetrical parts. The temperature coefficient of knowledge distillation τ_{KD} is set as 1.5 because the dataset we use does not contain many categories, and the network prediction scores for each category are not concentrated. The temperature coefficient in the loss of comparative representation τ is set as 0.07, and the dimension of teacher and student representation d is 128 in all structures. The coefficient α and β are set to 1 and 0.8, respectively.

4. Dataset

To build robust and efficient single-lead ECG interpretation models in a multi-label disease context, we conduct experiments on two freely accessible multi-label datasets.

ICBEB2018 dataset [20] contains 6877 12-lead ECG recordings, each ranging in length from 6 to 60 s. The dataset considers one normal and eight abnormal arrhythmia categories, including Atrial fibrillation (AF), First-degree atrioventricular block (I-AVB), Left bundle branch block (LBBB), Right bundle branch block (RBBB), Premature atrial contraction (PAC), Premature ventricular contraction (PVC), ST-segment depression (STD), and ST-segment elevated (STE). Each record may have more than one label. This paper follows the processing method in [20] to divide the dataset into ten folds by stratified sampling, and the original label distribution is maintained in each fold. Among them, the first eight are used as training sets, while the 9th and 10th are used as validation and test sets, respectively. Then, we preprocess them to get equal-length data.

PTB-XL dataset [19] is the to-date largest freely accessible clinical 12-lead ECG-waveform dataset comprising 21,837 10-second records from 18,885 patients. The original dataset division setting in [19] is dividing the dataset into ten class-balanced folds, with the first eight folds as the training set and the 9th and 10th folds as the validation set and test set, respectively. In addition, this fold assignment also respects the underlying patient assignments, which avoids data leakage arising from having ECG signals from the same patient in different folds. Many works [47,48] are based on this dataset, so for a fair comparison, we also use this dataset. We use all signals downsampled at 100 Hz as the labeled data source, and the signal length of each sample is 1000 points. This dataset provides rich multi-level annotations, which, in terms of diagnosis, include superclasses of 5 classes, containing Normal ECG (Norm), Conduction Disturbance (CD), Myocardial Infarction (MI), Hypertrophy (HYP), ST/T change (STTC) and subclasses of 24 classes.

Algorithm 1 Complete training procedure of our MVKT-ECG

Input: $\mathbf{X} = \left[\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(12)}\right]$, $\boldsymbol{x} = \mathbf{x}^{(k)}$
true label: $\boldsymbol{y}$
representations: $\mathcal{F}_\theta^T(\boldsymbol{X}), \mathcal{F}_\theta^S(\boldsymbol{x}), \mathbf{G}^T(\mathbf{X}), \mathbf{G}^S(\mathbf{x})$
output logits: $z^T = C(\mathcal{F}_\theta^T(\boldsymbol{X})), z^S = C(\mathcal{F}_\theta^S(\boldsymbol{x}))$
output probabilities: $\hat{\boldsymbol{y}} = sigmoid(z^S)$
repeat
  calculate $\mathcal{L}_{\text{BCE}} = -\left[\boldsymbol{y}\log\hat{\boldsymbol{y}} + (1-\boldsymbol{y})\log(1-\hat{\boldsymbol{y}})\right]$
  calculate $\mathcal{L}_{\text{CLT}} = -\mathbb{E}\left[\log\frac{exp[(G^T(X_i)G^S(X_j))/\tau]}{\sum_{i=1}^{N} exp[(G^T(X_i)G^S(X_j))/\tau]}\right]$
  calculate $\mathcal{L}_{\text{MKD}} = \tau_{kd}^2 \sum_i \text{KL}\left(q_i^T || q_i^S\right)$
  Take gradient descent step on
  $\nabla_\theta \mathcal{L}_{\text{MVKT}} = \mathcal{L}_{\text{BCE}}^S + \alpha\mathcal{L}_{\text{MKD}} + \beta\mathcal{L}_{\text{CLT}}$
until converged

5. Experiments

5.1. Evaluation metrics

Below, we report performance in terms of the area under the Receiver Operating Characteristic (ROC-AUC), Accuracy, and the F1 score.

Accuracy.

Accuracy reflects the ability of the classifier or model to judge the correctness of the overall sample. Specifically, it is the ratio of the number of correctly classified samples to the total number of samples. It can be calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{10}$$

ROC-AUC.

Receiver Operator Characteristic Curve (ROC) is a widely used performance metric for binary classification models. It measures the ability of a model to distinguish between the positive and negative classes and provides an aggregate measure of its predictive performance across various classification thresholds.

AUC (Area Under the Curve): AUC is the area under the ROC curve. It represents the overall performance of a classification model. The AUC value ranges between 0 and 1, with 0.5 indicating a random classifier and 1 indicating a perfect classifier. The higher the AUC, the better the model can distinguish between positive and negative instances.

F1 Score.

It represents the balance between precision and recall. Specifically, it can be represented as the harmonic mean of precision and recall. Mathematically, it can be calculated as follows:

$$Precision = \frac{TP}{TP + FP}, \tag{11}$$

$$Recall = \frac{TP}{TP + FN}, \tag{12}$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall}. \tag{13}$$

5.2. Experimental setups

We implemented the MVKT algorithm in the framework of Pytorch. We used a server with an Intel Xeon CPU E5-2640 v4 and 2 1080Ti GPU for the experiment. During the training, the model is trained according to the two-stage teacher-student training process, and the 12-lead ECG signals are first used to train the teacher model. All the teacher networks are trained for 100 epochs with Adam Optimization [49]. During the information transferring period, the parameters of the teacher model are frozen, and only the student model is updated. We feed 12-lead ECG signals to the teacher and single-lead ECG signals to the student. The size of the training batch is 32.

5.3. Comparison with state-of-the-art single lead ECG's classification methods

In this section, we compare the performance of our method, MVKT-ECG, with several state-of-the-art single-lead ECG classification methods on lead I of the CPSC2018 and PTB-XL datasets. Table 1 reports a thorough comparison with current state-of-the-art (SOTA) methods across datasets. In the Physionet Challenge 2017 competition [30], Hannun's CNN-based model [21] achieved the best score. The DenseCNN_Wang [35], and Resnet1d_Liu [36] achieved the top performance in China ECG AI Contest 2019 competition. [37] proposed a large kernel size model SEresnet_Zhao based on SE-block [38], achieving second place in the PhysioNet 2020 competition [39]. [18] (KD+FitNet) and [12] (KD + attention) also use the distillation idea to bridge the gap between the model with multi-lead ECG signals and single-lead ECG signals by trying to make the student's output and internal features the same as the teacher's. Since KD loss is only applicable to single-label classification, we use the KL divergence when computing KD loss on these two datasets. Compared with KD, Our proposed method performs better and outperforms other SOTA single-lead models. This result is fully consistent with our goal of robustness when only a single-lead ECG is provided as a query.

5.4. MVKT-ECG on different backbones

To further explore the impact of our method on the single-lead-based ECG interpretation network, we also compare several state-of-the-art networks for arrhythmia detection from ECG signals. We compare the performance of these networks before and MVKT. The 5 models: CNN_Hannun [21], ResNet1d34 [22], ResNet1d_wang [44], inception1d [45], XresNet1d101 [46] are set as our backbones. The performance of the backbones on independent single-lead ECG signals is set as our baseline. We indicate the baseline and teacher model with the name of the backbone and append "MVKT" for the model after the MVKT (e.g., ResMVKT34). We first benchmark 5 SOTA models' performance when they observe 12-lead ECG signals and single-lead ECG signals, respectively. Table 2 reports the comparison for different backbones. In detail, the student XresMVKT1d101 outperforms its baseline observably on the PTB-XL dataset (1.3% on PTB.super, and 1.4% on PTB.sub), and ResMVKT34 outperforms its baseline observably on ICBEB2018 dataset (3.2% on ICBEB2018).

The detailed experimental results in Tables 3 and 4 demonstrate the detailed promotion of each disease. Here we report the F1 score of each disease on different backbones. We also try to use different frameworks for teachers and students. We can see from Tables 3 and 4 that MVKT is beneficial in most cases for the diagnosis of any disease. Notably, the students' detection performance of AF can surpass their teacher for CNN_Hannun, which shows that in the field of diagnosis of some dependent diseases, single leads have more advantages than

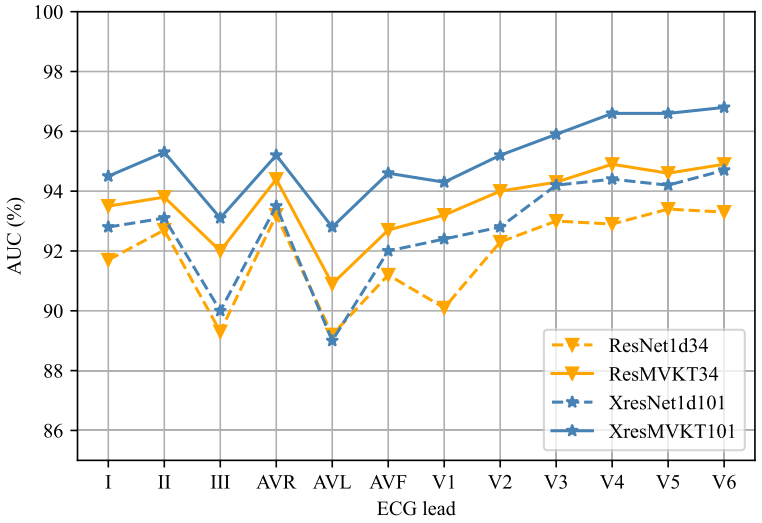


Fig. 4. Performance (AUC) at evaluation time when changing the lead of ECG in ICBEB2018 dataset.

Table 1
The comparison of MVKT's performance with State-of-the-Art single-lead ECG interpretation methods on public datasets.

Method	ICBEB2018			PTBXL.subdiagnostic			PTBXL.superdiagostic		
	AUC	ACC	F1-score	AUC	ACC	F1-score	AUC	ACC	F1-score
Resnet1d_Liu [36]	90.2	92.2	65.6	84.9	88.5	48.8	82.6	80.2	60.9
DenseCNN_Wang [35]	88.7	90.6	61.7	84.5	88.3	47.9	82.7	81.1	61.7
SEresnet18 [39]	93.5	94.4	72.5	85.0	88.8	48.5	82.7	81.3	61.6
SEresnet34 [39]	93.7	94.4	73.1	84.4	89.3	47.7	82.9	81.8	62.3
CNN_Hannun [21]	94.0	94.7	74.6	84.9	89.2	48.9	83.1	80.8	61.4
KD+FitNet [18]	94.0	94.8	74.8	83.6	87.2	46.4	83.8	76.5	60.1
KD+AT [12]	92.8	93.8	71.4	82.0	86.1	44.2	80.0	78.5	59.8
MVKT-ECG	**95.7**	**95.7**	**78.0**	**86.1**	**88.3**	**58.3**	**84.3**	**82.2**	**62.6**

multiple leads. However, there is poor performance in identifying ST-segmentation elevation with single-lead ECG. A previous study showed that the key feature of ST-segment abnormalities is the small-amplitude difference (0.2 mV) between the J node and the PR node, which might be filtered out by the DNN models through multiple convolution operations in deep layers [50]. For this hard-to-detect disease, MVKT still improves its detection accuracy. Based on this, we draw the following conclusions: (1) according to the objective which the student seeks to optimize, our method can gain significant improvement when single-lead ECG signals are available; (2) compared to the baseline, the students do gain improvement, but it is hard for them to exceed the teacher's overall performance. The single-lead ECG signals contain only a large fraction of the 12-lead ECG signals' information that the MVKT can elicit to the greatest extent, but the information dropped by single-lead ECG signals can never be recovered.

As additional proof, the plots from Fig. 4 show a comparison between models before and after information transferring with each lead. MVKT-ECG improves the performance considerably on the ICBEB2018 dataset for each selected lead. Surprisingly, the lead II, AVR, and limb lead (especially v3–v6) always perform better. What is more, MVKT-ECG can make lighter single-lead-based ECG interpretation networks superior to more complex models several times deeper: for example, ResMVKT34 scores better than even XresNet101 on ICBEB2018, regardless of the selected lead of ECGs.

5.5. *Analysis on* **MVKT-ECG**

Ablation study.

We conducted a thorough ablation study for the loss terms. Without loss of generality, we focus our analysis on XresMVKT101 on the ICBEB2018 dataset. Table 5 reports the result of the ablation study.

Table 2
Experiment results on different datasets, settings, and architectures before and after MVKT.

Number of leads	Backbone	AUC (%)		
		PTB.sub	PTB.super	ICBEB2018
12	CNN_Hannun [21]	92.3	92.0	97.0
1	CNN_Hannun	84.8	83.3	94.0
	CNN_Hannun_MVKT	**85.6**	**83.4**	**95.7**
12	ResNet1d34 [22]	92.3	92.8	94.5
1	ResNet1d34	84.5	82.6	90.7
	ResMVKT34	**85.2**	**83.8**	**93.9**
12	ResNet1d_wang [44]	92.7	92.1	91.0
1	ResNet1d_wang	84.9	83.1	87.3
	ResMVKT_Wang	**85.8**	**84.0**	**89.3**
12	Inception1d [45]	93.2	92.4	93.8
1	Inception1d	85.2	83.2	88.1
	InceptionMVKT	**86.1**	**84.3**	**90.7**
12	XresNet1d101 [46]	92.1	93.7	95.3
1	XresNet1d101	83.9	82.7	92.6
	XresMVKT101	**85.3**	**84.0**	**94.6**

Among the three losses, $\mathcal{L}_{\mathrm{CLT}}$ plays an important role. Although we only use the hard label and $\mathcal{L}_{\mathrm{CLT}}$, it can also achieve a significant result, demonstrating the effectiveness of our proposed role of information transferring between leads. And as expected, the most outstanding performance of AUC is obtained with all the losses.

Significant impact of MKD.

As mentioned above, we propose a new knowledge distillation objective that can handle multi-label disease problems. Here, to illustrate the effect of our proposed MKD, we compare it with the KL divergence. Table 6 shows the result of the comparison on ICBEB2018 datasets. We

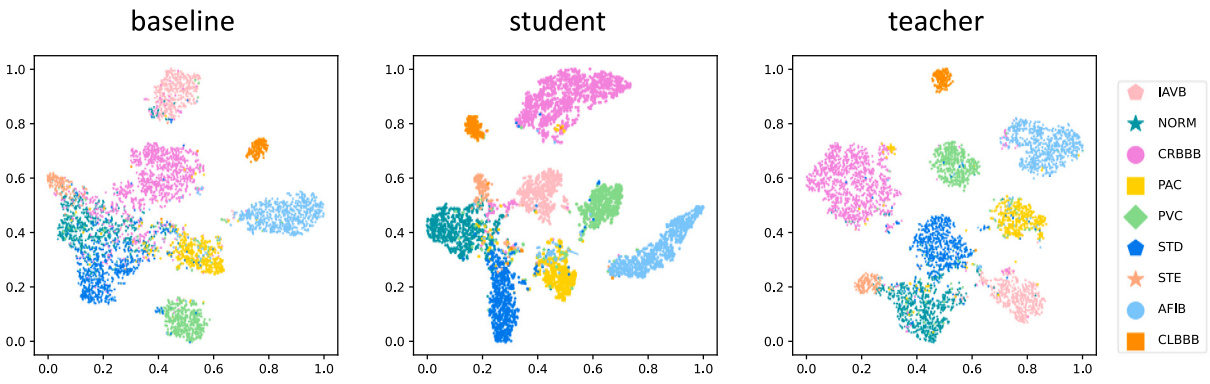


Fig. 5. Feature visualization. T-SNE plot of the feature distribution of baseline, student, and teacher.

Table 3
Detailed results of each disease on ICBEB2018 dataset.

Number of leads	Backbone	F1_score (%)									
		NORM	AF	I-AVB	LBBB	RBBB	PAC	PVC	STD	STE	Average
12	CNN_Hannun	75.8	92.0	87.0	88.0	91.7	71.3	86.7	74.0	68.1	81.6
1	CNN_Hannun	65.2	95.2	84.8	**91.3**	85.0	66.7	**86.5**	66.0	30.8	74.6
	CNN_Hannun_MVKT	**73.9**	**95.9**	**85.5**	87.5	**86.9**	**70.8**	82.2	**72.0**	**47.4**	**78.0**
12	XresNet1d101	73.7	95.2	85.3	88.9	92.5	60.2	89.6	77.0	66.7	81.0
1	XresNet1d101	65.8	89.8	83.8	87.0	85.2	63.1	80.0	65.9	32.3	72.5
	XresMVKT101	**68.1**	**95.2**	**82.5**	**91.3**	**86.8**	**67.3**	**83.0**	**68.8**	**36.4**	**75.5**
	ResNet1d34	63.9	92.1	84.5	87.5	85.5	33.8	60.0	65.1	26.9	66.6
	ResMVKT34	**69.1**	**93.3**	**86.3**	**87.5**	**87.3**	**50.0**	**74.5**	**66.3**	**30.8**	**71.7**

Table 4
MVKT's results of different diseases on PTB-XL dataset.

Number of leads	Backbone	F1_score					
		NORM	MI	STTC	CD	HYP	Average
12	CNN_Hannun	74.8	60.5	73.2	86.2	76.0	74.2
1	CNN_Hannun	58.9	45.8	56.3	79.7	66.5	61.4
	CNN_Hannun_MVKT	**60.1**	**46.4**	**57.4**	**80.3**	**67.8**	**62.4**
12	ResNet1d34	74.5	58.6	75.1	85.8	76.8	74.2
1	ResNet1d34	59.2	45.0	55.4	80.1	66.4	61.2
	ResMVKT34	**61.3**	**47.3**	**56.7**	**80.6**	**68.7**	**62.9**
12	XresNet1d101	74.8	53.8	72.7	85.9	75.6	72.6
1	XresNet1d101	58.2	45.5	55.6	80.5	66.6	61.3
	XresMVKT101	**59.8**	**47.2**	**57.3**	80.4	**67.5**	**62.4**

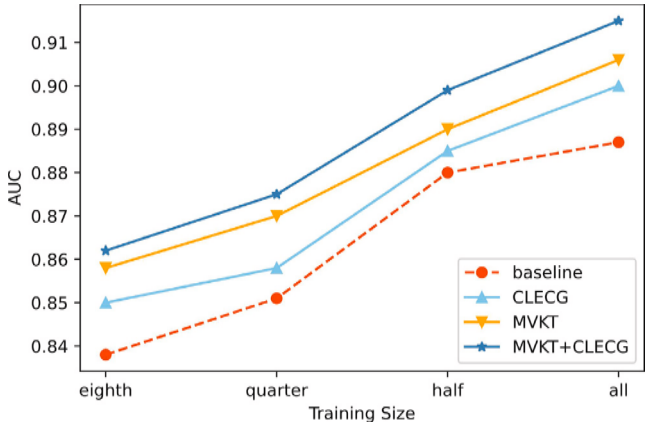


Fig. 6. Combine MVKT-ECG with CLECG. The pre-training procedure is based on the PTB-XL dataset, and the finetune procedure is based on the ICBEB2018 dataset.

Table 5
Ablation study in terms of the impact of each loss term.

	$\mathcal{L}_{BCE}$	$\mathcal{L}_{MKD}$	$\mathcal{L}_{CLT}$	AUC (%)
XresNet101 (Teacher)				95.3
XresMVKT101 (Student)	✓			92.6
	✓	✓		93.7
	✓		✓	93.9
	✓	✓	✓	**94.6**

evaluate three different results for the following settings: training with single-lead ECG signals independently, learning from the teacher with CLT and KL, and learning from the teacher with CLT and MKD. From the results, we can see that the proposed MKD can further improve the performance, which is better than directly minimizing the KL divergence between the output probabilities of the teacher and student.

Visualization of the class distribution.

To visually assess the differences between the baseline and student, we use the T-SNE graphs to highlight the feature distribution. Fig. 5 depicts the impact of MVKT between different backbones on ICBEB2018. As we can see, the distribution of features of normal (Norm), atrial

Table 6
Ablation study in terms of the impact of MKD.

	CNN_Hannun [21]	ResNet1d34 [22]	ResNet1d_Wang [44]	XresNet101 [46]
Ind	94.0	90.7	87.3	92.6
CLT+KL	95.2	92.8	87.4	94.1
CLT+MKD	**95.7**	**93.9**	**89.3**	**94.6**

premature beat (PAC), ST-segment depression (STD), and ST-segment elevation (STE) in the baseline network is confused to some extent. After MVKT, although the four categories are still relatively concentrated in the same area, the feature distribution of the distilled model in all nine categories became denser. As for the issue of category overlap, the status is greatly improved. This suggests that the information conveyed by other leads can compensate for the lack of single leads for various arrhythmias.

Compatibility.

Faced with the shortage of ECG datasets, many self-supervised learning methods have been proposed to take advantage of unlabeled data. So the compatibility with self-supervised learning methods of ECG signals is a vital evaluation indicator for MVKT-ECG. As for the initialization of the student network, [10] introduced a SOTA unsupervised pre-training program — CLECG, to mine adequate information from unlabeled data. During the pre-training, CLECG encourages the representations of different augmented views of the same signal to be similar and increases the distance between the representations of augmented views from the different signals.

To evaluate the compatibility of our method with self-supervised pre-training, we combine MVKT-ECG with CLECG on the Inception1d backbone. As the PTB-XL dataset is the to-date largest freely accessible clinical ECG-waveform dataset, we first pre-train the network with CLECG on the PTB-XL dataset, then use MVKT-ECG to finetune. The different proportions (eighth, quarter, half, all) of the ICBEB2018 training set data are used in the finetune procedure. The AUC indicators of random initialization and self-supervised pre-training are compared, and the results are shown in Fig. 6.

It can be seen that at different data sizes, both MVKT-ECG and CLECG can improve the performance compared to the baseline model and even surpass the performance of the baseline model trained with total data when using only half of the data. On this basis, using MVKT-ECG on the parameters obtained from external data of CLECG can further improve the performance of the student model, and the AUC metric increases by 0.9% when using all training data, which shows that our method performs well when it is compatible with other methods. Moreover, this result suggests that the arrhythmia detection model can improve performance by accepting knowledge transferred both from unlabeled external data and other leads.

6. Conclusion

An effective ECG smart device detection algorithm requires higher accuracy and robustness for single-lead ECG signals. To achieve this goal, we propose MVKT-ECG. This is a teacher-student information transferring approach in which the student observes only a single lead of the input ECG signals. This strategy encourages students to find better representations and to be closer to the teacher in performance through the knowledge transferring of the 12-lead ECG processing network. Notably, MVKT-ECG shows robustness in different datasets and different backbones. Experimental results show that the proposed algorithm is insensitive to the classification granularity and specific categories of the datasets, and the accuracy of the student model in the detection of multiple arrhythmias is greatly improved. The visualization results also show that multi-view knowledge transferring can guide the student model to simulate the expression of the teacher model, which compensates for the deficiency of single-lead ECGs in differentiating some specific categories.

MVKT-ECG can effectively improve the performance of the model for diagnosing single-lead ECG. However, some diseases, such as ST-segmentation abnormalities, still cannot be well recognized by neural networks with high performance. This is because the features of ST-segmentation abnormalities will be diminished by convolution layers. Therefore, our training process must include other lead electrocardiograms to train the teacher network. Therefore, our future work will focus on using self-supervised methods to mine more powerful information from limited lead signals, especially single-lead ECG, and strive to build a Foundation Mode of single-lead ECG intelligent recognition for more and broader downstream tasks.

Declaration of competing interest

There is no conflict of interest in the paper.

Acknowledgement

This research was supported by the National Natural Science Foundation of China under Grant No. 62276153.

References

[1] Holger Holst, Mattias Ohlsson, Carsten Peterson, Lars Edenbrandt, A confident decision support system for interpreting electrocardiograms, Clin. Physiol. 19 (5) (1999) 410–418.
[2] Jintai Chen, Xiangshang Zheng, Hongyun Yu, Danny Z. Chen, Jian Wu, Electrocardio panorama: Synthesizing new ECG views with self-supervision, 2021, http://dx.doi.org/10.48550/ARXIV.2105.06293.
[3] Jintai Chen, Kuanlun Liao, Kun Wei, Haochao Ying, Danny Z Chen, Jian Wu, ME-GAN: Learning panoptic electrocardio representations for multi-view ECG synthesis conditioned on heart diseases, in: Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, Sivan Sabato (Eds.), Proceedings of the 39th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 162, PMLR, 2022, pp. 3360–3370.
[4] Serkan Kiranyaz, Turker Ince, Moncef Gabbouj, Real-time patient-specific ECG classification by 1-D convolutional neural networks, IEEE Trans. Biomed. Eng. 63 (3) (2016) 664–675.
[5] Antônio H Ribeiro, Manoel Horta Ribeiro, Gabriela MM Paixão, Derick M Oliveira, Paulo R Gomes, Jéssica A Canazart, Milton PS Ferreira, Carl R Andersson, Peter W Macfarlane, Wagner Meira Jr., et al., Automatic diagnosis of the 12-lead ECG using a deep neural network, Nat. Commun. 11 (1) (2020) 1–9.
[6] Guijin Wang, Chenshuang Zhang, Yongpan Liu, Huazhong Yang, Dapeng Fu, Haiqing Wang, Ping Zhang, A global and updatable ECG beat classification system based on recurrent neural networks and active learning, Inform. Sci. 501 (2019) 523–542.
[7] Zhourui Xia, Zhenhua Sang, Yutong Guo, Weijie Ji, Chenguang Han, Yanlin Chen, Sifan Yang, Long Meng, Automatic multi-label classification in 12-lead ECGs using neural networks and characteristic points, in: Hongen Liao, Simone Balocco, Guijin Wang, Feng Zhang, Yongpan Liu, Zijian Ding, Luc Duong, Renzo Phellan, Guillaume Zahnd, Katharina Breininger, Shadi Albarqouni, Stefano Moriconi, Su-Lin Lee, Stefanie Demirci (Eds.), Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting, Springer International Publishing, Cham, 2019, pp. 80–87.
[8] Yang Liu, Qince Li, Kuanquan Wang, Jun Liu, Runnan He, Yongfeng Yuan, Henggui Zhang, Automatic multi-label ECG classification with category imbalance and cost-sensitive thresholding, Biosensors 11 (11) (2021).
[9] Dani Kiyasseh, Tingting Zhu, David A. Clifton, Clocs: Contrastive learning of cardiac signals across space, time, and patients, in: International Conference on Machine Learning, PMLR, 2021, pp. 5606–5615.
[10] Hui Chen, Guijin Wang, Guodong Zhang, Ping Zhang, Huazhong Yang, CLECG: A novel contrastive learning framework for electrocardiogram arrhythmia classification, IEEE Signal Process. Lett. 28 (2021) 1993–1997.
[11] Pritam Sarkar, Ali Etemad, Self-supervised ECG representation learning for emotion recognition, IEEE Trans. Affect. Comput. (2020) 1.

[12] Shenda Hong, Cao Xiao, Trong Nghia Hoang, Tengfei Ma, Hongyan Li, Jimeng Sun, RDPD: rich data helps poor data via imitation, 2018, arXiv preprint arXiv:1809.01921.

[13] Joel M Raja, Carol Elsakr, Sherif Roman, Brandon Cave, Issa Pour-Ghaz, Amit Nanda, Miguel Maturana, Rami N Khouzam, Apple watch, wearables, and heart rhythm: where do we stand? Ann. Transl. Med. 7 (17) (2019).

[14] Felix K Wegner, Simon Kochhäuser, Christian Ellermann, Philipp S Lange, Gerrit Frommeyer, Patrick Leitz, Lars Eckardt, Dirk G Dechering, Prospective blinded evaluation of the smartphone-based AliveCor kardia ECG monitor for atrial fibrillation detection: The PEAK-AF study, Eur. J. Intern. Med. 73 (2020) 72–75.

[15] Konstantinos D. Rizas, Luisa Freyer, Nikolay Sappler, Lukas von Stülpnagel, Peter Spielbichler, Aresa Krasniqi, Michael Schreinlechner, Felix N. Wenner, Fabian Theurl, Amira Behroz, Elodie Eiffener, Mathias P. Klemm, Annika Schneidewind, Martin Zens, Theresa Dolejsi, Ulrich Mansmann, Steffen Massberg, Axel Bauer, Smartphone-based screening for atrial fibrillation: a pragmatic randomized clinical trial, Nat. Med. 28 (9) (2022) 1823–1830.

[16] Shenda Hong, Cao Xiao, Tengfei Ma, Hongyan Li, Jimeng Sun, MINA: multilevel knowledge-guided attention for modeling electrocardiography signals, 2019, arXiv preprint arXiv:1905.11333.

[17] Geoffrey Hinton, Oriol Vinyals, Jeff Dean, et al., Distilling the knowledge in a neural network, 2015, arXiv preprint arXiv:1503.02531, 2(7).

[18] Majid Sepahvand, Fardin Abdali-Mohammadi, A novel method for reducing arrhythmia classification from 12-lead ecg signals to single-lead ECG with minimal loss of accuracy through teacher-student knowledge distillation, Inform. Sci. 593 (2022) 64–77.

[19] Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I Lunze, Wojciech Samek, Tobias Schaeffter, PTB-XL, a large publicly available electrocardiography dataset, Sci. Data 7 (1) (2020) 1–15.

[20] Feifei Liu, Chengyu Liu, Lina Zhao, Xiangyu Zhang, Xiaoling Wu, Xiaoyan Xu, Yulin Liu, Caiyun Ma, Shoushui Wei, Zhiqiang He, Jianqing Li, Eddie Ng, An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection, J. Med. Imag. Health Inform. 8 (2018) 1368–1373.

[21] Awni Y Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H Tison, Codie Bourn, Mintu P Turakhia, Andrew Y Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nat. Med. 25 (1) (2019) 65–69.

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[23] Mohamad Mahmoud Al Rahhal, Yakoub Bazi, Haidar Almubarak, Naif Alajlan, Mansour Al Zuair, Dense convolutional networks with focal loss and image generation for electrocardiogram classification, IEEE Access 7 (2019) 182225–182237.

[24] Pengwei Xie, Guijin Wang, Chenshuang Zhang, Ming Chen, Huazhong Yang, Tingting Lv, Zhenhua Sang, Ping Zhang, Bidirectional recurrent neural network and convolutional neural network (BiRCNN) for ECG beat classification, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 2555–2558.

[25] Artzai Picon, Unai Irusta, Aitor Álvarez-Gila, Elisabete Aramendi, Felipe Alonso-Atienza, Carlos Figuera, Unai Ayala, Estibaliz Garrote, Lars Wik, Jo Kramer-Johansen, et al., Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia, PLoS One 14 (5) (2019) e0216756.

[26] Runnan He, Yang Liu, Kuanquan Wang, Na Zhao, Yongfeng Yuan, Qince Li, Henggui Zhang, Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional LSTM, IEEE Access 7 (2019) 102119–102135.

[27] Supreeth P. Shashikumar, Amit J. Shah, Gari D. Clifford, Shamim Nemati, Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 715–723.

[28] Shenda Hong, Cao Xiao, Tengfei Ma, Hongyan Li, Jimeng Sun, MINA: Multilevel knowledge-guided attention for modeling electrocardiography signals, 2019, http://dx.doi.org/10.48550/ARXIV.1905.11333.

[29] Tianqi Fan, Sen Qiu, Zhelong Wang, Hongyu Zhao, Junhan Jiang, Yongzhen Wang, Junnan Xu, Tao Sun, Nan Jiang, A new deep convolutional neural network incorporating attentional mechanisms for ECG emotion recognition, Comput. Biol. Med. 159 (2023) 106938.

[30] Gari D Clifford, Chengyu Liu, Benjamin Moody, Li-wei H. Lehman, Ikaro Silva, Qiao Li, A E Johnson, Roger G. Mark, AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017, in: 2017 Computing in Cardiology (CinC), 2017, pp. 1–4.

[31] Abhimanyu Singh Udawat, Pushpendra Singh, An automated detection of atrial fibrillation from single-lead ecg using HRV features and machine learning, J. Electrocardiol. 75 (2022) 70–81.

[32] Hai Yue Gu, Jun Huang, Xu Liu, Shu Qian Qiao, Xi Cao, Effectiveness of single-lead ECG devices for detecting atrial fibrillation: An overview of systematic reviews, Worldviews Evid.-Based Nurs. (2023).

[33] Wenqiang Li, Yuk Ming Tang, Kai Ming Yu, Suet To, SLC-GAN: An automated myocardial infarction detection model based on generative adversarial networks and convolutional neural networks with single-lead electrocardiogram synthesis, Inform. Sci. 589 (2022) 738–750.

[34] C Michael Gibson, Sameer Mehta, Mariana RS Ceschim, Alejandra Frauenfelder, Daniel Vieira, Roberto Botelho, Francisco Fernandez, Carlos Villagran, Sebastian Niklitschek, Cristina I Matheus, et al., Evolution of single-lead ECG for STEMI detection using a deep learning approach, Int. J. Cardiol. 346 (2022) 47–52.

[35] Chunli Wang, Shan Yang, Xun Tang, Bin Li, A 12-lead ECG arrhythmia classification method based on 1D densely connected CNN, in: Hongen Liao, Simone Balocco, Guijin Wang, Feng Zhang, Yongpan Liu, Zijian Ding, Luc Duong, Renzo Phellan, Guillaume Zahnd, Katharina Breininger, Shadi Albarqouni, Stefano Moriconi, Su-Lin Lee, Stefanie Demirci (Eds.), Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting, Springer International Publishing, Cham, 2019, pp. 72–79.

[36] Yang Liu, Runnan He, Kuanquan Wang, Qince Li, Qiang Sun, Na Zhao, Henggui Zhang, Automatic detection of ECG abnormalities by using an ensemble of deep residual networks with attention, in: Hongen Liao, Simone Balocco, Guijin Wang, Feng Zhang, Yongpan Liu, Zijian Ding, Luc Duong, Renzo Phellan, Guillaume Zahnd, Katharina Breininger, Shadi Albarqouni, Stefano Moriconi, Su-Lin Lee, Stefanie Demirci (Eds.), Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting, Springer International Publishing, Cham, 2019, pp. 88–95.

[37] Zhibin Zhao, Hui Fang, Samuel D Relton, Ruqiang Yan, Yuhong Liu, Zhijing Li, Jing Qin, David C Wong, Adaptive lead weighted ResNet trained with different duration signals for classifying 12-lead ECGs, in: 2020 Computing in Cardiology, 2020, pp. 1–4.

[38] Jie Hu, Li Shen, Gang Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[39] Erick A Perez Alday, Annie Gu, Amit J Shah, Chad Robichaux, An-Kwok Ian Wong, Chengyu Liu, Feifei Liu, Ali Bahrami Rad, Andoni Elola, Salman Seyedi, et al., Classification of 12-lead ecgs: the physionet/computing in cardiology challenge 2020, Physiol. Meas. 41 (12) (2020) 124003.

[40] Qingxue Zhang, Kyle Frick, All-ECG: A least-number of leads ECG monitor for standard 12-lead ecg tracking during motion, in: 2019 IEEE Healthcare Innovations and Point of Care Technologies, (HI-POCT), 2019, pp. 103–106.

[41] Tomer Golany, Kira Radinsky, Daniel Freedman, Saar Minha, 12-Lead ECG reconstruction via koopman operators, in: Marina Meila, Tong Zhang (Eds.), Proceedings of the 38th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 139, PMLR, 2021, pp. 3745–3754.

[42] Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, Wojciech Samek, Deep learning for ECG analysis: Benchmarks and insights from PTB-XL, IEEE J. Biomed. Health Inf. 25 (5) (2021) 1519–1528.

[43] Aaron van den Oord, Yazhe Li, Oriol Vinyals, Representation learning with contrastive predictive coding, 2018, arXiv preprint arXiv:1807.03748.

[44] Zhiguang Wang, Weizhong Yan, Tim Oates, Time series classification from scratch with deep neural networks: A strong baseline, in: 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, 2017, pp. 1578–1585.

[45] Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F Schmidt, Jonathan Weber, Geoffrey I Webb, Lhassane Idoumghar, Pierre-Alain Muller, François Petitjean, Inceptiontime: Finding alexnet for time series classification, Data Min. Knowl. Discov. 34 (6) (2020) 1936–1962.

[46] Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, Wojciech Samek, Deep learning for ECG analysis: Benchmarks and insights from PTB-XL, IEEE J. Biomed. Health Inf. 25 (5) (2020) 1519–1528.

[47] Sandra Śmigiel, Krzysztof Pałczyński, Damian Ledziński, ECG signal classification using deep learning techniques based on the PTB-XL dataset, Entropy 23 (9) (2021) 1121.

[48] Krzysztof Pałczyński, Sandra Śmigiel, Damian Ledziński, Sławomir Bujnowski, Study of the few-shot learning for ECG classification based on the PTB-XL dataset, Sensors 22 (3) (2022) 904.

[49] Diederik P. Kingma, Jimmy Ba, Adam: A method for stochastic optimization, 2014, http://dx.doi.org/10.48550/ARXIV.1412.6980.

[50] Jian Ni, Yingtao Jiang, Shengjie Zhai, Yihan Chen, Sijia Li, Amei Amei, Dieu-My. T. Tran, Lijie Zhai, Yu Kuang, Multi-class cardiovascular disease detection and classification from 12-lead ECG signals using an inception residual network, in: 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), 2021, pp. 1532–1537.