Knowledge-Based Systems 264 (2023) 110323

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Semantic-aware alignment and label propagation for cross-domain arrhythmia classification

Panpan Feng [a,c], Jie Fu [a,d], Ning Wang [a,c], Yanjie Zhou [b,*], Bing Zhou [a,c], Zongmin Wang [c]

[a] School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou Henan, 450000, China
[b] School of Management, Zhengzhou University, Zhengzhou Henan, 450000, China
[c] Cooperative Innovation Center of Internet Healthcare, Zhengzhou University, Zhengzhou Henan, 450000, China
[d] Pattern Recognition Laboratory of Institute of Automation, Chinese Academy of Sciences, Beijing, 100000, China

ARTICLE INFO

Article history:
Received 22 July 2022
Received in revised form 3 January 2023
Accepted 17 January 2023
Available online 20 January 2023

Keywords:
Arrhythmia classification
Semi-supervised domain adaptation
Domain alignment
Pseudo label learning

ABSTRACT

Cross-domain arrhythmia classification (CAC) aims to transfer the model trained on a label-sufficient source domain to a label-scarce target domain. To the best of our knowledge, almost all existing CAC models focus on the unsupervised setting, where no labeled target samples are available. However, in most practical scenes, acquiring limited annotated target samples is feasible, which can provide reliable target semantic information for model learning directly. Consequently, we first propose a more realistic semi-supervised CAC setting, where only the source samples and extremely limited target samples are annotated. Most previous CAC models realize cross-domain learning by aligning the feature distributions of source and target domains coarsely and globally, where the semantic invariance within each class is not taken into consideration during the domain alignment process. Additionally, the semantic information contained in the feature space is not fully utilized for target pseudo label mining. To address the above two issues, a unified framework containing the semantic-aware feature alignment (SAFA) and prototype-based label propagation (PBLP) modules is proposed. In the proposed framework, SAFA and PBLP are complementary to each other. Specifically, SAFA provides more robust prototypes for PBLP by performing semantic-aware feature alignment, and PBLP offers more reliable target pseudo labels for more effective semantic-aware feature alignment learning. Comprehensive qualitative and quantitative experimental results on different benchmarks verify the effectiveness of the proposed method.

1. Introduction

Electrocardiogram (ECG) records the electric activity of the heart, and the electrical signal of each heartbeat has been widely used for detecting heart diseases, especially arrhythmias [1,2]. Recently, the heartbeat-based arrhythmia classification models trained in a fully supervised manner have made great success, which hinges on the large-scale annotated heartbeats. In addition, the supervised models assume that test samples share the same distribution with the training samples [3]. However, in most practical scenarios, obtaining large-scale annotated heartbeats is difficult due to requiring enormous capital and time investment [4]. Meanwhile, the heartbeats of different individuals or databases are unfortunately varied, meaning that there are distributional shifts between the training data (source domain) and test data (target domain). Thus, the well trained model in a domain typically does not generalize to a new domain due to the distributional shifts [5]. Consequently, the scarcity of heartbeat annotations in a new domain and distributional shifts are two fundamental challenges in cross-domain arrhythmia classification (CAC).

To cope with the above two challenges of CAC, existing works [6–11] focus on improving the performance of a target model with label-scarce data by using the knowledge from a label-sufficient source domain. These existing methods can be divided into two categories: fine-tuning methods [6,7] and deep domain adaptation methods [8–11]. The fine-tuning methods usually first train a model with label-sufficient source samples, and then fine tune the model with labeled target samples. However, the issue of over fitting may occur when label-scarce target samples are used to fine tune the model. The deep domain adaptation approaches focus on achieving domain alignment between the source domain and target domain. To the best of our knowledge, almost all existing methods focus on the unsupervised scenario, where no labeled target samples are accessible during the training process. However, in most practical scenes, acquiring limited labeled target data is feasible, which can provide reliable target

* Corresponding author.
E-mail address: ieyjzhou@zzu.edu.cn (Y. Zhou).

https://doi.org/10.1016/j.knosys.2023.110323
0950-7051/

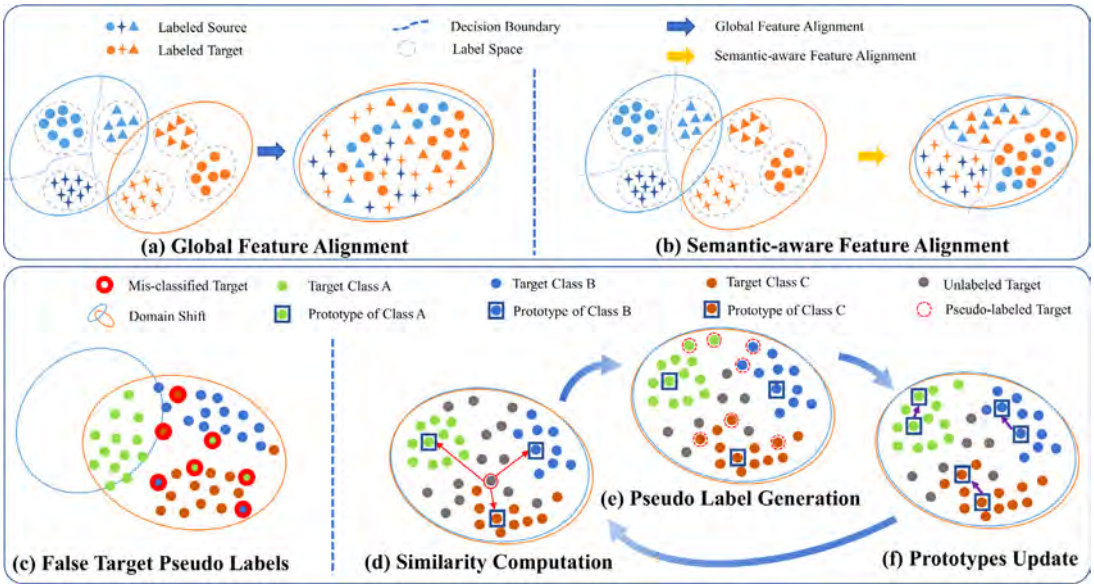

Fig. 1. The illustration of our motivations. (1) The representations obtained by the global feature alignment on the source domain and the target domain cannot be well discriminated, where the representations of several categories can be mixed (see Fig. 1(a)). In contrast, semantic-aware feature alignment considers the semantic invariance within each class between the two domains during the domain alignment process (see Fig. 1(b)). (2) The target pseudo labels obtained by the predictions of the model trained on the source domain inevitably contain false pseudo labels, due to the distributional shift between the source domain and the target domain (see Fig. 1(c)). In contrast, prototype-based label propagation is introduced to assign reliable target pseudo labels by utilizing the similarities between the representations of samples belonging to the same class in the target domain and class-wise prototypes. (see Fig. 1(d–f)).

information for model learning directly [12]. Consequently, this paper first focuses on the semi-supervised CAC scenario, where only the source samples and extremely limited target samples are annotated. In fact, existing CAC methods achieve domain alignment by reducing the gap between the feature distribution of two domains coarsely and globally, where the fine-grained semantic category information is not taken into consideration during the domain alignment process. However, intuitively, the semantic consistency of the same category in two domains should be maintained, and the semantic discrepancy between different categories should be large. Additionally, some methods [9–11] do suggest that mining pseudo labels from unlabeled target data to regularize model training is helpful to promote model performance during the domain alignment process. For these methods, the pseudo labels in the target domain are obtained by using the predictions of the model trained on the source domain. Nevertheless, the obtained target pseudo labels are inevitably noisy since there are the distributional shifts between the target domain and source domain. Actually, the semantic information contained in the feature space is helpful for target pseudo label learning, but the existing methods ignore this point.

As mentioned above, two observations are summarized: (1) A consideration of the semantic invariance within each class between the source domain and the target domain is required during domain alignment process. Intuitively, ignoring the fine-grained semantic category information could result in poor distinguishability, as illustrated in Fig. 1(a). Namely, the representations of different categories from the source domain and the target domain are mixed together, which makes the representations indistinguishable. (2) The reliability of target pseudo labels needs to be considered. Prior pseudo label learning methods utilize the model trained on the source domain to obtain target pseudo labels. However, due to the distributional shifts between the two domains, there are inevitably false target pseudo labels, as illustrated in Fig. 1(c). In light of these observations, the semantic consistency of the same category in two domains should be maintained, and the semantic discrepancy between different categories should be large. Furthermore, reliable target pseudo labels are required. Consequently, we consider aligning the feature distributions of samples from the same category across domains, and utilizing semantic information within the class to propagate label information. These considerations naturally lead to a principled way of learning semantic-aware feature alignment and prototype-based label propagation to address the above issues simultaneously. Consequently, a unified framework containing the semantic-aware feature alignment (SAFA) and prototype-based label propagation (PBLP) modules is proposed. SAFA is designed to perform semantic-aware feature alignment and learn the semantic-aware feature representations, as illustrated in Fig. 1(b). PBLP is introduced to obtain reliable target pseudo labels by calculating the similarities between the representations of samples belonging to the same class in the target domain and class-wise prototypes, as illustrated in Fig. 1(d-f). In the proposed framework, SAFA and PBLP can enhance each other. Specifically, SAFA provides more robust prototypes for PBLP by performing semantic-aware feature alignment, and PBLP offers more reliable pseudo target labels for more effective semantic-aware feature alignment learning.

The contributions of our work are summarized as follows.

- We first explore a realistic semi-supervised CAC scene and analyze existing two issues of the global feature alignment and noisy pseudo labels. A unified framework is introduced to jointly address these problems for accurate arrhythmia classification in the target domain.
- We design a SAFA module and a PBLP module to address the global feature alignment and noisy pseudo labels issues simultaneously. In specific, SAFA performs semantic-aware feature alignment and learns the semantic-aware feature representations by exploiting the semantic category information across domains. PBLP obtains reliable target pseudo labels by leveraging the similarities between the representations of samples from the same class and class-wise prototypes as reliable supervision information.
- In the proposed unified framework, SAFA and PBLP modules can enhance each other. Specifically, performing semantic-aware feature alignment by the former is beneficial for

providing more robust prototypes, and the latter generates more reliable target pseudo labels to facilitate more effective semantic-aware feature alignment learning.
- Comprehensive quantitative and qualitative experiments are constructed on four ECG benchmarks. The experimental results are consistent with the theoretical analysis, indicating the effectiveness and generalization of the proposed method.

In what follows, Section 2 reviews related work and introduces background. Section 3 describes the details of the proposed method. The experimental setups and results are presented in Section 4 and Section 5, respectively. Finally, the conclusions are presented in Section 6.

2. Related work and background

In this section, the existing research and background related to our work are summarized. First, we describe the related research ares, including ECG classification, deep domain adaptation and pseudo label learning. Second, we introduce the background of the proposed method: domain adversarial network and prototype learning.

2.1. ECG classification

Given the attracted representation learning capability and remarkable model performance of deep learning, significant progress has been made in ECG classification based on deep neural networks recently [13]. Various networks [14–18] have been investigated and modified to achieve the state-of-the-art performance of ECG classification. The impressive classification performance typically depends on large-scale annotated training data. Meanwhile, these models assume that the training samples follow the same distribution as test samples. Nevertheless, in real-world applications, acquiring large-scale annotated samples is difficult due to the large amount of capital and time investment. In addition, the assumption constraint can be easily violated since the ECG signals of different individuals or databases are varied, resulting in different distributions of the training samples and test samples. To resolve these dilemmas, the transfer learning technique has recently been successfully deployed to ECG classification, aiming to transfer knowledge from a label-sufficient source domain to a label-scarce target domain, avoiding labor-intensive data annotations. The existing methods mainly include two branches: fine-tuning methods based on parameter transfer [6,7,19–22] and deep domain adaptation methods [8–11,23–25]. The former methods utilize annotated target samples to fine-tune the models trained with labeled source data. The articles [19–22,26] first trained the model with label-sufficient source samples, and then fine tuned the model with limited labeled target samples. Al Rahhal et al. [6] and Wang et al. [7] used the active learning technology to annotate ECG samples from specific patients, and further utilized these labeled data to fine tune the pre-trained model trained with labeled source data. However, annotating data by the active learning technique requires the assistance of clinicians. Meanwhile, the pre-trained models easily lead to over fitting if the number of fine tuning data is small. The latter deep domain adaptation methods usually focus on alleviating the distribution discrepancy between the source domain and the target domain, as to improve the model performance in the target domain. The studies [10,23–25] mainly focused on the unsupervised setting, where no labeled target samples are available. For example, Deng et al. [25] proposed a multi-source unsupervised domain adaptive model for ECG classification. In the article [23], an asymmetric domain adaptive model was presented to address the issue of the distributional shifts for inter-patient atrial fibrillation detection. Jin et al. [24] introduced a domain adaptive residual network, which combined multi-kernel maximum mean discrepancy to detect atrial fibrillation across domains. Li et al. [10] proposed a mix-up asymmetric tri-training model to improve the generalization ability of the model under domain shifts. Among these methods, they generally assume that no labeled target samples are available. In contrast, this paper first explores another practical yet under-investigated semi-supervised CAC scenario, where unlabeled target samples, extremely limited labeled target samples and label-sufficient source samples are available.

2.2. Deep domain adaptation

Recently, the success of deep learning methods and the demand for label-sufficient data have made deep domain adaptation receive significant attention. The existing approaches are mainly divided into unsupervised and semi-supervised domain adaptation methods according to the quantity of labeled data in the target domain. Most researches focus on the unsupervised setting, where no labeled samples are accessible in the target domain. As a pioneering approach for unsupervised domain adaptation, Ganin and Lempitsky [27] were inspired by the generative adversarial network [28] and developed an adversarial training method to align the global feature distribution of the source domain and the target domain. The core mechanism of this adversarial training has been widely used for CAC [8,29,30]. For instance, Niu et al. [8] proposed a deep adversarial network with multi-scale feature fusion for CAC by aligning the global feature distribution of two domains. Nevertheless, these methods ignore the fine-grained semantic category information during the domain alignment process, which may result in the mixture of the category distribution of different domains. Consequently, Pei et al. [31] incorporated a separate discriminator for each class and used the task classification probability to weight the loss from each discriminator during domain alignment process. However, the task classification probability inevitably contains false semantic information since there are the distributional shifts between the source domain and the target domain. Recently, a more practical and realistic semi-supervised domain adaptation setting has been proposed to solve the domain adaptation problem. In the setting, a small amount of labeled samples are provided for the target domain. As a pioneering semi-supervised domain adaptation method, Saito et al. [32] proposed a deep neural network model via optimizing the minimax loss on the conditional entropy of the target samples. In addition, the articles [32–34] indicated that the performance of the model can be further increased via additional supervision training on a small amount of labeled target data, which implied the importance of labeled target data. Besides, another self-training mechanism was applied by training the model with obtained pseudo labeled target samples [35]. In light of the above methods, we observe that the contributions of the labeled target samples may be significantly diluted, since the utilization of these data is only restricted to optimizing ordinary supervised loss during the training process. Besides, the accessibility of pseudo labels requires careful treatment, since incorrect pseudo labels can severely degrade the model performance. Inspired by the above observations, intuitively, the semantic consistency of the same category in two domains should be maintained, and the semantic discrepancy between different categories should be large during domain alignment. In the paper, we take the semantic information into consideration during the domain alignment process, and the labeled target samples are further utilized to obtain reliable pseudo labels for unlabeled target data. To obtain reliable pseudo labels in the target domain, we introduce a prototype-based label propagation mechanism by exploiting the labeled and the unlabeled target data.

2.3. Pseudo label learning

Learning robust models requires label-sufficient training samples. However, obtaining label-sufficient samples is difficult in most practical applications. In the context of domain adaptation, several studies [9–11,36–38] attempted to mine pseudo labels for adapting the pre-trained model to the target domain, where the pre-trained model is trained with the label-sufficient source samples. For example, Zou et al. [37,38] introduced the confidence regularization to fight against overconfident pseudo labels in the training process of cross-domain semantic segmentation task. Wang et al. [9] obtained target pseudo labels by the predictions of the source classifier, where the source classifier is trained with the source labeled data. Li et al. [10] developed an asymmetric self-training scheme for arrhythmia detection, where two separate classifiers were trained with source data to obtain target pseudo labels, and the other classifier used these pseudo labeled target samples to retrain the model. However, due to the distributional shifts between the source domain and the target domain, false target pseudo labels based on the pre-trained models trained on the source domain inevitably exist. Unlike these methods, we propose a novel label propagation mechanism, in which the target pseudo labels are assigned by calculating the similarities between the unlabeled target representations and all class prototypes.

2.4. Domain adversarial network

Domain adversarial training method [27] has shown that the model can learn domain-invariant representations between the source domain and the target domain by embedding the domain discriminator on the basis of deep neural network during the domain adversarial training process. Due to the ingenious design and impressive performance improvement, domain adversarial training has attracted more and more attention. The common approach [27] is to align the feature distribution of two domains based on the adversarial training, which borrows the idea of the generative adversarial network [28]. In the process of the adversarial training, samples from different domains are encouraged to be non-discriminative with respect to domain labels. Specifically, the model includes a feature extractor f_F, a task classifier f_C, and a domain discriminator f_D. Through the adversarial training, the f_F is trained to deceive the f_D, making the feature representations of the target domain and the source domain indistinguishable. The adversarial training process is realized by inserting the gradient reversal layer between the f_F and the f_D. In addition, we need to simultaneously minimize the ordinary supervised loss l_C of the f_F and f_C with the available labeled samples. To simplify the notations, the domain discrimination loss is presented as l_D. Formally, the ultimate goal of the model is to optimize the following objective,

$$L(\theta_F, \theta_C, \theta_D) = \mathrm{E}_{(x_s, y_s)\in D_s} l_C\left(f_C\left(f_F\left(x_s\right)\right), y_s\right) - \lambda \mathrm{E}_{x_i \in D_s \cup D_t} l_D\left(f_D\left(f_F\left(x_i\right)\right), d_i\right), \quad (1)$$

where D_s and D_t represent the samples of source domain and target domain, respectively. θ_F, θ_C and θ_D denote the parameters of f_F, f_C and f_D, respectively. d_i means the domain label of the sample x_i. λ is the trade-off parameter between the two items in the objective.

2.5. Prototype learning

The core mechanism of prototype learning is to use prototypes to represent each category. A prototype usually refers to the most representative point, so it can also be considered as the anchor point or the exemplar. After obtaining the prototypes, for the given sample x, the probability value of x that belongs to a certain category can be obtained by the similarity between the sample and the prototypes. The similarity can be measured by a metric function d. Formally, the prototype μ_k of class k is represented as follows,

$$\mu_k = \frac{1}{n_k} \sum_{j=1}^{n_k} h_k^j, \quad (2)$$

where h_k^j represents the embedding of the sample belonging to category k. n_k indicates the number of samples belonging to category k. For a given sample x, the class distribution is shown as follows,

$$p_\theta(y = k|x) = \frac{\exp(-d[f_\theta(x), c_k])}{\sum_j \exp(-d[f_\theta(x), c_j])}, \quad (3)$$

where $f_\theta(x)$ represents the embedding of the sample x.

Recently, Snell et al. [39] proposed a prototypical network to address the challenge of limited training samples. The network learns the metric space and classifies the embedding of samples by calculating the shortest distance between embedding and different prototypes. To the best of our knowledge, this is the first time that prototype learning and domain adversarial network are combined for CAC.

3. Proposed methodology

In Section 3.1, we first describe the problem formulation and the overview of the proposed framework. The proposed framework mainly includes SAFA and PBLP modules. The two modules are elaborated in Sections 3.2 and 3.3 respectively. Finally, the overall training scheme is detailed in Section 3.4 and the feasibility of the proposed method is analyzed in Section 3.5.

3.1. Problem formulation and framework

Problem formulation. The goal of the semi-supervised CAC is to learn a model that preforms well in the target domain by using datasets in both domains. In the source domain, the source data and the corresponding labels $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{\mathrm{N}_s}$ are given. In the target domain, we are given unlabeled samples $D_t^u = \{x_i^{t_u}\}_{i=1}^{\mathrm{N}_u}$ and extremely limited labeled data and the corresponding labels $D_t^l = \{(x_i^{t_l}, y_i^{t_l})\}_{i=1}^{\mathrm{N}_l}$. In addition, the label space is shared between the source domain and the target domain. In view of the above notations, we aim to learn a heartbeat classification model that is trained on D_s, D_t^l and D_t^u, and tested on D_t^u.

Framework. The outline workflow of the proposed framework is shown in Fig. 2. The model in this framework consists of three components: a feature extractor f_F, a heartbeat classifier f_C and multiple category-level domain discriminators $f_{D_k}|_{k=1}^K$, where the K denotes the number of all categories. The framework containing SAFA and PBLP modules is composed of four stages. During the model training process, for the pre-training stage, the proposed model (composed of f_F, $f_{D_k}|_{k=1}^K$ and f_C) is trained with the available labeled data D_s and D_t^l. The first stage is implemented by the SAFA module, which aligns semantic-aware feature representations of the two domains and provides robust initial prototypes for the PBLP module. Then, in the pseudo labeling stage, pseudo labeled target data $D_t^p = \left\{(x_i^{t_u}, y_i^p)\right\}_{i=1}^{N_u}$ for unlabeled target samples D_t^u are obtained. Meanwhile, in the re-training stage, the additional pseudo-labeled data and available labeled data are further utilized to train the model by the SAFA module. The pseudo labeling stage and re-training stage are conducted iteratively until the total loss of the model converges. Finally, the heartbeat classification model (composed of f_F and f_C) is applied to predict the target test samples. The following subsections describe SAFA and PBLP modules in detail.

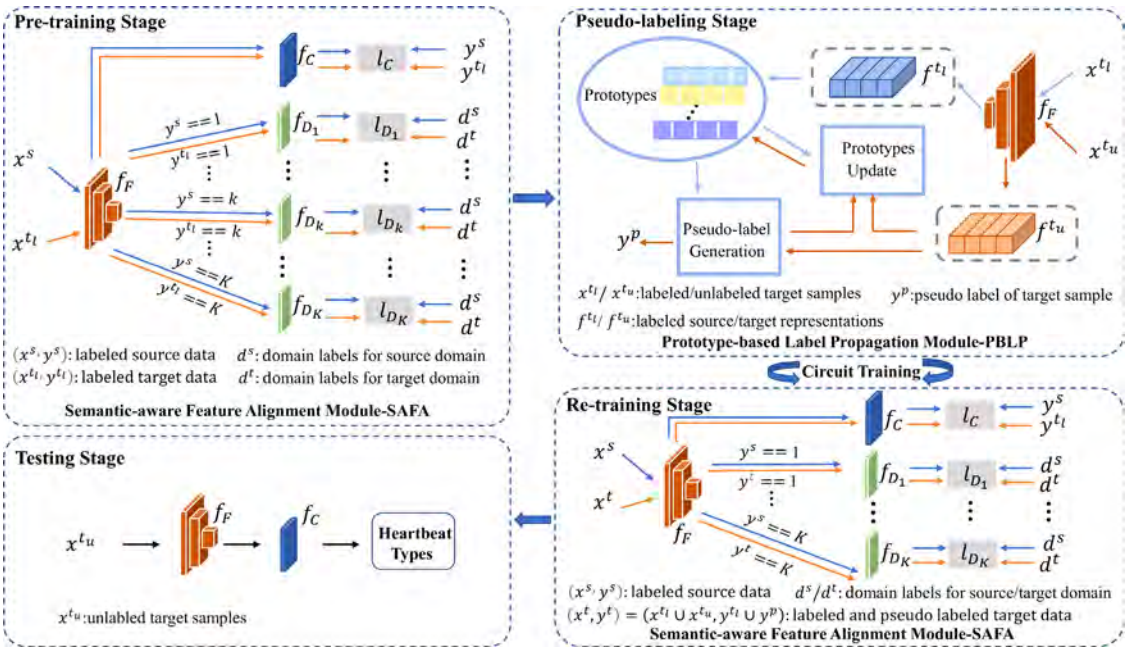

Fig. 2. The overview workflow of the proposed framework. The framework comprises of SAFA and PBLP modules to achieve four stages: pre-training, pseudo-labeling, re-training, and testing. SAFA module aims to alleviate inter-domain discrepancy via learning semantic-aware feature alignment. PBLP module aims to obtain reliable target pseudo labels for more effective semantic-aware feature alignment learning. The different colored solid lines with arrows denote the flow of different types of data. The symbols are consistent with those defined in Section 3.

3.2. Semantic-aware feature alignment module

In the process of domain alignment, intuitively, the feature representations of the samples belonging to the same category in different domains are pushed close, and the distribution gaps between different categories are enlarged. In addition, each heartbeat sample only belongs to a certain semantic category for CAC. Inspired by these observations, we propose SAFA module to perform semantic-aware feature alignment to alleviate inter-domain discrepancies of CAC. Specifically, the proposed model (composed of f_F, $f_{D_k}|_{k=1}^{K}$ and f_C) is based on the adversarial learning. During the adversarial learning process, the f_F is trained to deceive each $f_{D_k}|_{k=1}^{K}$, making the feature representations of samples belonging to the same category in different domains indistinguishable by each domain discriminator $f_{D_k}|_{k=1}^{K}$, in which each domain discriminator is used to distinguish whether the feature representations come from the source domain or the target domain. Namely, the parameters θ_F of f_F are learned by maximizing the domain discrimination loss l_{D_k} of each $f_{D_k}|_{k=1}^{K}$, while the parameters $\theta_{D_k}|_{k=1}^{K}$ are learned by minimizing each loss l_{D_k} of each $f_{D_k}|_{k=1}^{K}$. In addition, the parameters θ_F of f_F and θ_C of f_C are learned by minimizing the ordinary supervised loss of a heartbeat classification model (composed of f_F and f_C) simultaneously.

Using the notations in the above definitions, the supervised classification loss $L_C(\theta_F, \theta_C)$ is formulated as follows,

$$L_C(\theta_F, \theta_C) = E_{(x,y)\sim D_s \cup D_t^l} l_C(f_C(f_F(x)), y). \tag{4}$$

The domain discrimination loss of all domain discriminators $f_{D_k}|_{k=1}^{K}$ is defined as follows,

$$L_D\left(\theta_F, \theta_{D_k}|_{k=1}^{K}\right) = \sum_{k=1}^{K} E_{(x,y)\sim D_s \cup D_t} l_{D_k}\left(f_{D_k}(f_F(x)|y=k), d\right), \tag{5}$$

where $D_t = D_t^l \cup D_t^p$. As above the outline workflow, in the pre-training stage, $D_t^p = \phi$. In the re-training stage, the newly obtained pseudo labeled target data are utilized to perform semantic-aware feature alignment to alleviate inter-domain discrepancies. l_C is the weighted cross-entropy loss. l_{D_k} denotes the binary cross-entropy loss of f_{D_k}. d represents the domain label of the sample x, where d equals to 0 when x comes from the source domain and d equals to 1 when x comes form the target domain.

Overall, the objective of the SAFA module is the following function,

$$L(\theta_F, \theta_C, \theta_{D_k}|_{k=1}^{K}) = L_C(\theta_F, \theta_C) - \lambda L_D(\theta_F, \theta_{D_k}|_{k=1}^{K}), \tag{6}$$

where the hyper-parameter λ controls the trade-off between the two terms during adversarial training. In light of the adversarial learning process, after optimizing the above objective functions, the parameters $(\hat{\theta}_F, \hat{\theta}_C, \hat{\theta}_{D_k}|_{k=1}^{K})$ are learned.

$$(\hat{\theta}_F, \hat{\theta}_C) = \arg\min_{\theta_F, \theta_C} L(\theta_F, \theta_C, \hat{\theta}_{D_k}|_{k=1}^{K}), \tag{7}$$

$$\hat{\theta}_{D_k}|_{k=1}^{K} = \arg\max_{\theta_{D_1}, \ldots, \theta_{D_K}} L(\hat{\theta}_F, \hat{\theta}_C, \theta_{D_k}|_{k=1}^{K}). \tag{8}$$

3.3. Prototype-based label propagation module

During the domain alignment process, the semantic-aware category feature representations from the target domain are insufficient since only extremely limited labeled target samples are available. To address the issue, we resort to providing more reliable target pseudo labels for more effective semantic-aware feature alignment learning. Inspired by the fact that the feature representations of the same category are similar in the same domain, we attempt to make full use of the semantic information contained in the feature space to obtain target pseudo labels. Specifically, a PBLP module is proposed to obtain target pseudo labels by using the intra-class semantic label propagation in the target domain. First of all, the initial class-wise prototypes in the target domain are obtained by using the available labeled target data, where the class-wise prototypes denote the average feature representations of the samples associated with its class. Then, the

similarities between the unlabeled target feature representations and all class-wise prototypes in the target domain are estimated, in which the similarities can be measured by a distance function. Finally, the target pseudo labels are obtained by selecting the most likely ones, and the selected pseudo labels are used to retrain the model by the SAFA module.

Formally, we first obtain the initial prototype of class k,

$$\mu_k^{init} = \frac{\sum_{(x^{t_l}, y^{t_l}) \in D_t^l} \mathrm{I}(\tilde{y^{t_l}}, y^{t_l}) * \mathrm{I}(y^{t_l}, k) * f_F(x^{t_l})}{\sum_{(x^{t_l}, y^{t_l}) \in D_t^l} \mathrm{I}(\tilde{y^{t_l}}, y^{t_l}) * \mathrm{I}(y^{t_l}, k)}, \tag{9}$$

where $\tilde{y^{t_l}}$ represents the pseudo label of the sample x^{t_l} predicted by the pre-trained model (composed of f_F and f_C). If the pseudo label is predicted true, the corresponding target data is selected. f_F represents the initially trained feature extractor obtained through the pre-training stage by the SAFA module. $\mathrm{I}(a, b)$ denotes a binary indicator function, where the function is only activated when a and b are equal. The class-wise prototypes characterize the semantic distribution information of different categories in the target domain. If the feature representations of the unlabeled target sample x^{t_u} are the most similar to the prototype of class k, the label of x^{t_u} is assigned to category k. In order to measure the similarities, a metric function is introduced. Formally, the pseudo label y^p of unlabeled data x^{t_u} can be obtained via maximizing the similarity measurement,

$$y^p = \arg\max_k Mea(f_F(x^{t_u}), \mu_k^{init}), k \in \{1, 2, \ldots, K\}, \tag{10}$$

where $Mea(a, b)$ denotes the cosine similarity function between a and b. Afterward, we compute the class-wise prototypes μ_k' based on newly obtained pseudo labels of the current training batch $\mathcal{X}^{t_u}$, update the class-wise prototypes and obtain the updated pseudo labels. The calculation process is as follows,

$$\mu_k' = \frac{\sum_{x^{t_u} \in \mathcal{X}^{t_u}} \mathrm{I}(y^p, k) * f_F(x^{t_u})}{\sum_{x^{t_u} \in \mathcal{X}^{t_u}} \mathrm{I}(y^p, k)}, \tag{11}$$

$$\mu_k = \alpha \mu_k^{init} + (1 - \alpha)\mu_k', \tag{12}$$

$$y^p = \arg\max_k Mea(f_F(x^{t_u}), \mu_k), \tag{13}$$

where α represents the linear weight value of the prototype update scheme. The smaller the value, the faster the update rate. The pseudo label of unlabeled target data is generated by calculating the similarity between the corresponding feature representation and the class-wise prototypes obtained in an updated manner.

3.4. Overall training scheme

As illustrated in Fig. 2, the proposed framework is based on SAFA and PBLP modules to perform four stages: pre-training, pseudo-labeling, re-training, and testing. The pre-training stage aims to learn an initial feature extractor for capturing robust target representations, which is achieved by the SAFA module. Then, the pseudo-labeling stage and re-training stage are performed iteratively. Specifically, the pseudo-labeling stage leverages the PBLP module to obtain reliable target pseudo labels, and the re-training stage leverages the SAFA module to perform more effective semantic-aware feature alignment learning. Note that performing semantic-aware feature alignment by the SAFA module is beneficial for providing more robust prototypes, and the PBLP module obtains more reliable target pseudo labels to facilitate more effective semantic-aware feature alignment learning. Consequently, the SAFA and PBLP modules can benefit from each other. Finally, the testing stage uses the learned heartbeat classification model to evaluate the model performance. The overall training procedure is summarized in Algorithm 1.

Algorithm 1 : The training scheme of the proposed framework

Input: D_s, D_t^l, D_t^u, prototype generation function: *Prototypes*, pseudo-labeling generation function: *Labeling*, prototype update function: *PrototypesUpdate*, the number of training iterations: *iter*

Output: model parameters $\theta_F, \theta_C, \theta_{D_k}|_{k=1}^K$

```
D_s, D_t^l, D_t^u, pseudo-labeled target data D_t^p = φ
for i = 1 to iter do
    train f_F, f_C, and f_{D_k}|_{k=1}^K with mini-batches from D_s, D_t^l
end for
for j = 1 to iter do
    μ_k^init = Prototypes(f_F, f_C, D_t^l)
    D_t^p = Labeling(f_F, μ_k^init, D_t^u)
    μ_k' = Prototypes(f_F, D_t^p)
    μ_k = PrototypesUpdate(μ_k^init, μ_k')
    D_t^p = Labeling(f_F, μ_k, D_t^u)
    train f_F, f_C, and f_{D_k}|_{k=1}^K with mini-batches from D_s, D_t^l ∪ D_t^p
end for
```

3.5. Theoretical analysis

In this subsection, we provide interpretations from a theoretic perspective to analyze the reasons why the proposed framework is effective in semi-supervised CAC. Ben-David et al. [40] proposed a domain adaptation theory, which stated that the expected loss $R_{\mathcal{T}}(h)$ for the target domain is bounded by three terms: the expected error for the source domain $R_{\mathcal{S}}(h)$, the domain divergence between the two domains $\mathcal{D}(\mathcal{S}, \mathcal{T})$, and the shared minimum error value of the ideal joint hypothesis $\mathcal{C}$. Formally, given the source domain $\mathcal{S}$ and the target domain $\mathcal{T}$, let $\mathcal{H}$ be the hypothesis class. For any hypothesis $h \in \mathcal{H}$, we have

$$R_{\mathcal{T}}(h) \leq R_{\mathcal{S}}(h) + \mathcal{D}(\mathcal{S}, \mathcal{T}) + \mathcal{C}. \tag{14}$$

In the proposed setting, the source labels are available, so the error $R_{\mathcal{S}}(h)$ can be easily minimized by optimizing a model trained in a fully supervised manner. In inequality (14), the second term $\mathcal{D}(\mathcal{S}, \mathcal{T})$ can be minimized by the global feature alignment between the source domain and the target domain. However, the third term $\mathcal{C}$ can easily become large when the representations of samples belonging to the same category cross domains are not explicitly aligned. Therefore, semantic-aware feature alignment needs to be considered during the domain alignment process. Since the available labels of target data are limited, we resort to the pseudo labels of unlabeled target data. Assume that $\hat{D}_t$ denotes the pseudo-labeled target data and labeled target data. $R_{\mathcal{T}'}(\bullet)$ is the expected risk on the dataset $\hat{D}_t$. The excepted labeling function of the source (target) domain is denoted as $f_{\mathcal{S}}$ ($f_{\mathcal{T}}$). The pseudo labeling target function is denoted as $f_{\hat{\mathcal{T}}}$. Then, the $\mathcal{C}$ is bound by the following four terms.

$$\mathcal{C} \leq \min_{h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}\left(h, f_{\hat{\mathcal{T}}}\right) + 2R_{\mathcal{T}'}\left(f_{\mathcal{S}}, f_{\hat{\mathcal{T}}}\right) + R_{\mathcal{T}'}\left(f_{\mathcal{T}}, f_{\hat{\mathcal{T}}}\right). \tag{15}$$

Proof. The proof of inequality (15) relies on the triangle inequality, where for any prediction functions f_1, f_2, and f_3, we have $R(f_1, f_2) \leq R(f_1, f_3) + R(f_2, f_3)$. Then,

$$\mathcal{C} = \min_{\forall h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}(h, f_{\mathcal{T}}), \tag{16}$$

$$\leq \min_{\forall h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}(f_{\mathcal{S}}, f_{\mathcal{T}}), \tag{17}$$

$$\leq \min_{\forall h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}\left(h, f_{\hat{\mathcal{T}}}\right) + R_{\mathcal{T}'}\left(f_{\mathcal{S}}, f_{\hat{\mathcal{T}}}\right) + R_{\mathcal{T}'}\left(f_{\mathcal{S}}, f_{\hat{\mathcal{T}}}\right) + R_{\mathcal{T}'}\left(f_{\mathcal{T}}, f_{\hat{\mathcal{T}}}\right), \tag{18}$$

$$= \min_{\forall h \in \mathcal{H}} R_{\mathcal{S}}(h, f_{\mathcal{S}}) + R_{\mathcal{T}'}\left(h, f_{\hat{\mathcal{T}}}\right) + 2R_{\mathcal{T}'}\left(f_{\mathcal{S}}, f_{\hat{\mathcal{T}}}\right) + R_{\mathcal{T}'}\left(f_{\mathcal{T}}, f_{\hat{\mathcal{T}}}\right). \tag{19}$$

Therefore, the inequality (15) holds.

In light of above analysis, $R_{\mathcal{T}}(h)$ is bounded by the three terms $R_{\mathcal{S}}(h)$, $\mathcal{D}(\mathcal{S}, \mathcal{T})$ and $\mathcal{C}$. Since the first two small terms do not guarantee small $\mathcal{C}$, we need to prove the third term $\mathcal{C}$ tends to be small. In addition, from the inequality (15), a suitable function h in $\mathcal{H}$ is easily obtained to minimize the errors $R_{\mathcal{S}}(h, f_{\mathcal{S}})$ and $R_{\mathcal{T}'}\left(h, f_{\hat{\mathcal{T}}}\right)$ by using available labeled source data and pseudo labeled target data. Besides, the semantic-aware feature representations of source domain and target domain can be progressively aligned by the proposed SAFA module, then the semantic-aware domain-invariant feature representations are obtained. Thus, the term $R_{\mathcal{T}'}\left(f_{\mathcal{S}}, f_{\hat{\mathcal{T}}}\right)$ is expected to be minimized. Meanwhile, the reliable pseudo labeled target samples are obtained by the proposed PBLP module, and these pseudo labeled data can be utilized to minimize $R_{\mathcal{T}'}\left(f_{\mathcal{T}}, f_{\hat{\mathcal{T}}}\right)$. To sum up, the term $\mathcal{C}$ is expected to be small. Consequently, the proposed framework in theory is effective in the semi-supervised CAC task.

4. Experimental setup

This section presents the experimental details of the semi-supervised CAC task, including datasets, competitors, network architecture, implementation details and evaluation indicators.

4.1. Datasets

We evaluate the proposed framework over three challenging semi-supervised CAC tasks based on four public databases from PhysioNet [41]. These databases contain heartbeat type information marked and verified by independent experts. According to the AAMI EC57:1998 standard [42], there are five heartbeat types: *N* (normal or bundle branch block), S (supraventricular ectopic beat), V (ventricular ectopic beat), F (fusion beat), and Q (unassigned beat). This paper focuses on four heartbeat types: N, V, S, and F. While the Q is ignored because the data of the heartbeat type Q in these databases is extremely few. The details of the four datasets are summarized in Table 1. Specifically, the MIT-BIH arrhythmia database [41] (ARDB[1]) includes 48 recordings from 47 patients. Each record has two lead signals, and each is band-pass filtered at 0.1–100 Hz and sampled at 360 Hz. According to AAMI convention [42], four paced records (102,104,107,217) are excluded in evaluating model performance. Each record is approximately longer than 30 min. The St. Petersburg institute of cardiological technics 12-lead arrhythmia database [41] (INCART[2]) contains 75 records of 30 min of 12 leads. The sampling rate of each record is 257 Hz. The MIT-BIH supraventricular arrhythmia database [41] (SVDB[3]) consists of 78 recordings with two leads. Each record is approximately 30 min and sampled at 128 Hz. The MIT-BIH long-term ECG database (LTDB[4]) [41] contains 7 long-term records. Each record contains approximately 14 to 20 h of recording with two leads, and each record has a sampling rate of 128 Hz. Besides, Fig. 3 displays one sample of each heartbeat type from the four datasets, indicating that there are distributional shifts among the datasets.

Table 1
The number of heartbeat types in the four databases.

Database	The number of heartbeat types				Number of records
	N	S	V	F	
ARDB	89 705	2766	6982	802	44
INCART	152 348	1944	19 863	219	75
SVDB	161 901	12 177	9922	23	78
LTDB	600 110	1499	64 081	2906	7

4.2. Competitors

We compare with the following approaches, source + target (S + T) [34], domain adversarial neural network (DANN) [27], multi-adversarial domain adaptation (MADA) [31], classification and contrastive semantic alignment (CCSA) [43], entropy (ENT) [33] and minimax entropy (MME) [32]. These methods are either specifically designed or tailored to deal with the domain adaptation problem. In addition, several state-of-the-art researches on the CAC task [6,7,9,11] are compared. We also report the experimental results of the baseline model 'Source Only' and the full supervised (Full-T) model.

(1) 'Source Only': The model is only trained on the labeled source data, and then directly applied to the target domain for classification.

(2) S + T [34]: The S + T model trains the model using the labeled source and target data. It is a fine-tuning model based on parameter transfer.

(3) DANN [27]: The DANN is a common unsupervised domain adaptation method based on the domain adversarial training. In the semi-supervised CAC setting, limited labeled target data is available and they are used to train the DANN.

(4) MADA [31]: The MADA utilizes multiple domain discriminators and pseudo labels to perform fine-grained feature representation alignment between the two domains, where the pseudo labels are obtained by the predictions of the model trained on the source domain.

(5) CCSA [43]: The CCSA introduces the classification and contrastive semantic alignment loss to learn an embedding function and a classification function from the source domain to the target domain.

(6) ENT [33]: The ENT is trained with available samples by using standard entropy minimization.

(7) MME [32]: The MME is an adaptive domain model based on the class-wise prototypes, which optimizes the minimax loss of conditional entropy on the unlabeled target data.

(8) Proposed Method: The proposed framework comprises SAFA and PBLP modules. The two modules benefit from each other and perform together to achieve the semi-supervised CAC task.

(9) Full-T: The model is implemented in full supervised practice to obtain the upper limit of model performance.

4.3. Network architecture and implementation details

For the semi-supervised CAC task, the proposed model is based on the unsupervised semantic-aware adaptive feature fusion network (USAFFN) [11], including a feature extractor, multiple category-level domain discriminators, and a heartbeat classifier. The network structures of the heartbeat classification model and each domain discriminator are the same as those of the classification model and the global domain discriminator in USAFFN. In order to combine the morphological and rhythmic information of ECG signals, the 3-D input is constructed as the same preprocessing pipelines in USAFFN. In addition, the feature extractor includes multiple atrous spatial pyramid pooling modules and

1 https://www.physionet.org/content/mitdb/1.0.0/.
2 https://www.physionet.org/content/incartdb/1.0.0/.
3 https://www.physionet.org/content/svdb/1.0.0/.
4 https://www.physionet.org/content/ltdb/1.0.0/.

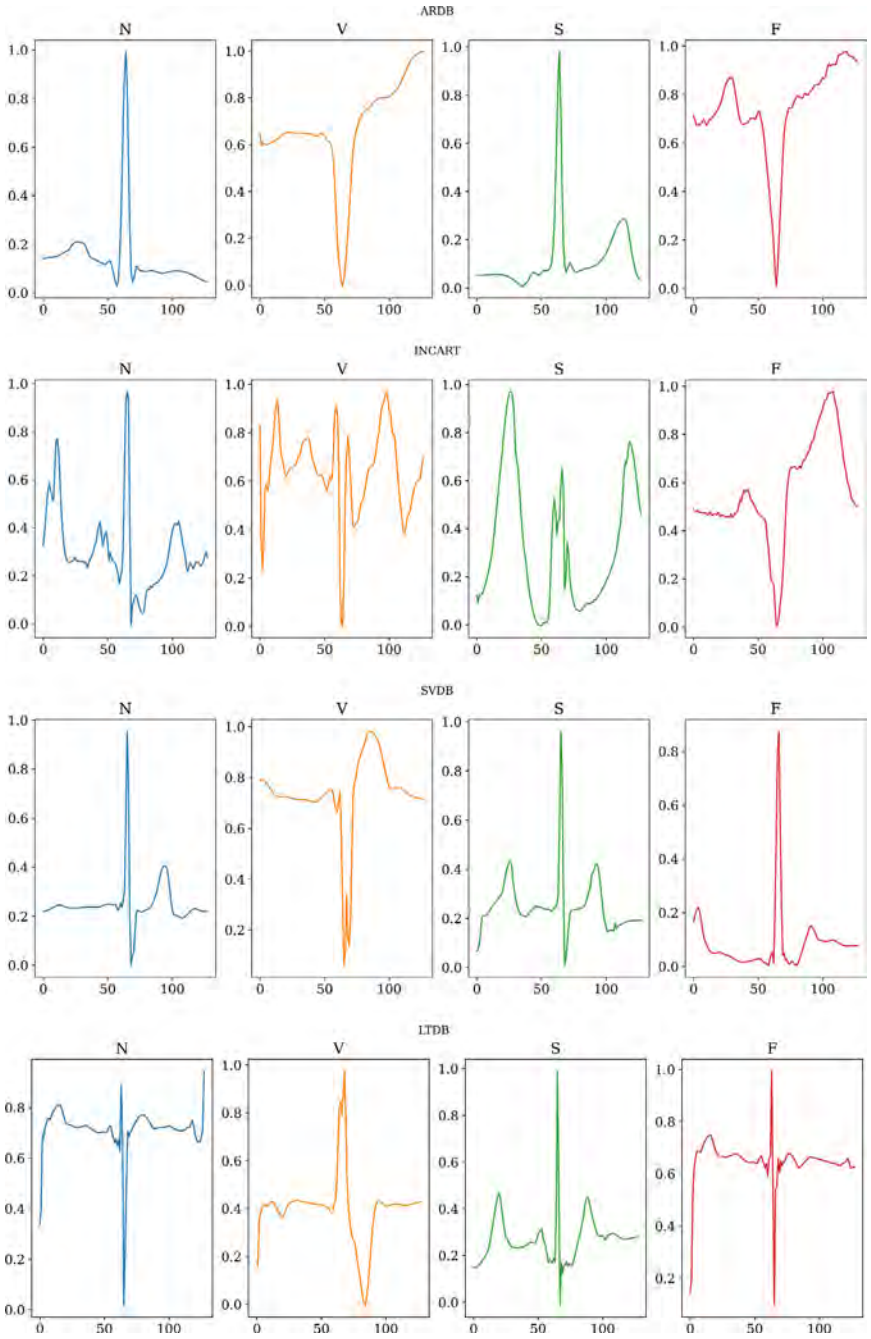

Fig. 3. Heartbeat examples for each class in different ECG databases.

multiple multi-perspective adaptive feature-fusion modules at different levels, aiming to extract multi-view features and reduce the redundancy of these features.

All the experiments in this study are implemented in Pytorch [44]. In the experiments, 50 annotated samples of each category in the target domain are available, and we take 90% of the remaining samples as unlabeled training data and 10% as the validation data. The optimizer is Adam, and the batch size is set to 128. We set the balancing hyper-parameters λ and α to 0.01 and 0.99, respectively. The training stages can be termed as two processes. In the first training process, the proposed model is trained with available labeled data for 100 epochs. Afterward, the pseudo-labeling stage and re-training stage are iteratively trained, and the maximum number of training during the process is 200. In the testing stage, the heartbeat classification model (composed of the feature extractor and the heartbeat classifier) is used to evaluate the model performance with the unlabeled target samples. Comprehensive qualitative and quantitative experiments are performed, and the corresponding results are described in Section 5. For the quantitative evaluation, we repeat the experiment 5 times to calculate the mean and standard deviation of these results.

4.4. Evaluation metrics and visual analysis

In this paper, the quantitative evaluation of model performance is based on several common evaluation indicators: sensitivity (Sen), positive productivity (Ppr), F1 score (F1), F1-macro score, the overall classification accuracy (ACC), and the area under the curve of ROC (AUC). Sen represents the rate of correctly

Table 2
Comparison results of each category on INCART.

Approach	N (%)			V (%)			S (%)			F (%)		
	Sen	Ppr	F1	Sen	Ppr	F1	Sen	Ppr	F1	Sen	Ppr	F1
Full-T	99.9 ± 0.00	99.8 ± 0.06	99.9 ± 0.03	99.6 ± 0.11	99.9 ± 0.00	99.8 ± 0.05	95.0 ± 2.87	99.9 ± 0.00	96.7 ± 2.18	73.3 ± 4.5	99.9 ± 0.00	84.7 ± 2.9
Rahhal et al. [6][a]	–	–	–	75.1	37.6	50.0	15.6	2.5	1.0	–	–	–
Wang et al. [7][a]	–	–	–	83.9	93.0	88.0	78.3	22.6	35.0	–	–	–
S + T [34][a]	**99.9 ± 0.03**	85.4 ± 0.17	92.1 ± 0.10	99.4 ± 0.05	78.3 ± 1.49	87.6 ± 0.92	12.8 ± 2.60	92.7 ± 2.86	22.5 ± 4.0	1.0 ± 0.09	**89.8 ± 0.70**	1.9 ± 0.14
Wang et al. [9][b]	–	–	–	82.5	95.2	88.0	66.7	47.6	55.0	–	–	–
Feng et al. [11][b]	98.0	97.0	98.0	85.0	95.0	89.0	70.0	44.0	54.0	0.0	0.0	0.0
DANN [27][b]	99.2 ± 0.06	94.5 ± 0.12	96.8 ± 0.09	89.8 ± 0.13	**95.1 ± 0.24**	92.4 ± 0.18	35.8 ± 0.36	54.8 ± 4.4	43.2 ± 1.18	0.8 ± 0.02	25.1 ± 1.29	1.58 ± 0.05
MADA [31][b]	98.4 ± 0.01	97.2 ± 0.01	97.8 ± 0.01	86.2 ± 0.01	92.2 ± 0.00	89.1 ± 0.00	39.3 ± 0.39	45.2 ± 0.50	42.1 ± 0.42	0.8 ± 0.00	2.1 ± 0.05	1.2 ± 0.01
CCSA [43][b]	97.2 ± 0.07	**99.7 ± 0.03**	98.4 ± 0.04	98.0 ± 0.05	84.0 ± 0.08	90.4 ± 0.06	**79.1 ± 0.34**	36.3 ± 0.17	49.8 ± 0.23	**12.4 ± 0.04**	0.6 ± 0.01	1.2 ± 0.03
ENT [33][b]	99.2 ± 0.03	95.2 ± 0.08	97.1 ± 0.03	98.9 ± 0.02	70.1 ± 0.07	82.0 ± 0.05	16.3 ± 0.03	62.1 ± 0.34	25.9 ± 0.07	0.3 ± 0.03	13.6 ± 0.38	0.6 ± 0.07
MME [32][b]	97.8 ± 0.01	99.5 ± 0.01	**98.6 ± 0.01**	89.8 ± 0.04	84.8 ± 0.06	87.3 ± 0.05	76.8 ± 0.93	19.9 ± 0.35	31.6 ± 0.52	0	0	0
Source Only	99.3 ± 0.03	80.9 ± 0.05	89.2 ± 0.03	62.1 ± 0.09	87.3 ± 0.07	72.6 ± 0.07	10.1 ± 0.03	77.1 ± 0.19	17.9 ± 0.05	0.2 ± 0.02	12.9 ± 0.10	0.5 ± 0.04
Proposed method[b]	99.3 ± 0.02	97.4 ± 0.03	98.3 ± 0.03	**99.6 ± 0.03**	92.0 ± 0.06	**95.6 ± 0.04**	40.6 ± 0.25	**93.5 ± 0.04**	**56.6 ± 0.24**	6.6 ± 0.12	89.7 ± 0.80	**12.4 ± 0.23**

[a]Parameter-based fine-tuning scheme.
[b]Domain adaptation scheme.

Table 3
Overall performance comparison on INCART.

	Full-T	S + T [34]	DANN [27]	MADA [31]	CCSA [43]	ENT [33]	MME [32]	Source Only	Proposed method
F1-macro	0.953 ± 0.78	0.510 ± 1.23	0.585 ± 0.36	0.575 ± 0.10	0.599 ± 0.05	0.514 ± 0.01	0.543 ± 0.14	0.450 ± 0.03	**0.657 ± 0.04**
ACC (%)	99.8 ± 0.05	84.9 ± 0.37	94.3 ± 0.14	94.9 ± 1.07	**97.3 ± 0.19**	94.3 ± 2.49	95.2 ± 1.29	83.7 ± 2.23	96.7 ± 0.00
AUC	0.999 ± 0.01	0.917 ± 0.60	0.931 ± 0.46	0.866 ± 0.89	0.977 ± 0.01	0.923 ± 0.02	0.856 ± 0.25	0.772 ± 0.49	**0.995 ± 0.20**

classified events among all events. Ppr is the rate of correctly classified events among all detected events. F1 is the harmonic mean of the positive productivity and sensitivity, which is used as the comprehensive criterion for each category. F1-macro is the average of the F1 score for each category. ACC is the ratio of correctly classified samples to the total classified samples. ROC represents a probability curve, and AUC denotes the degree of separability.

In addition, to qualitatively evaluate the model performance, we use t-distributed stochastic neighbor embedding (t-SNE) [45] to visualize how the model maps samples from two different domains to a feature space. Specifically, the samples from the source and target domains are compressed to their 2D features, and the visualization of these features is displayed in a 2D subspace. In this paper, we use two different shapes to represent source samples and target samples, and different colors denote different categories. Specifically, the symbol • presents the source samples, and the symbol × denotes the target samples. In addition, red, blue, green, and black are used to represent heartbeat types N, V, S, and F in the source domain. Light-coral, light-sky-blue, light-green, and gray colors are used to represent heartbeat types N, V, S, and F in the target domain.

5. Experimental results

In this section, we quantitatively and qualitatively analyze the experimental results of the proposed method in the semi-supervised CAC task. The task can be represented as ARDB → target domain, where the target domain belongs to INCART, SVDB, and LTDB, respectively. Specifically, the effectiveness of the proposed method is verified on the task: ARDB → INCART. Afterward, the effectiveness of the proposed SAFA and PBLP modules is validated, and the impact of the number of labeled target samples on the model performance is further explored. Finally, to further testify the generalization ability of our proposed method, the influence of different distributional shifts is explored on the tasks: ARDB → SVDB and ARDB → LTDB. Meanwhile, we use the t-SNE [45] to qualitatively investigate the characteristics of the features obtained by the proposed method.

5.1. Classification performance

To verify the effectiveness of the proposed method, the semi-supervised CAC task: ARDB → INCART is conducted and the proposed method is compared with the studies based on the deep domain adaptation [6,7,9,11] and several representative methods [27,31–34,43] on INCART. The results of competitors on INCART are reported in Tables 2, 3, and Fig. 4. Specifically, Tables 2 and 3 report the results of each category and overall model performance, respectively. Fig. 4 displays the AUC for each category of these competitors. From these results, several observations can be obtained. (1) The proposed method shows superior performance among the reported counterparts. The encouraging experimental results highlight the importance of semantic-aware feature alignment learning, and reliable pseudo labels mining is conducive to effective semantic-aware feature alignment learning. (2) The approaches [6,7] based on the active learning scheme require constantly interacting with experts and picking up informative samples of new records to adapt to the model. When the number of new records is large, it would introduce additional manual labeling workload. Compared with these approaches, the proposed method only needs to train the model with limited labeled samples of each category. (4) Compared with the existing algorithms [9,11,27,31–33,43] based on the deep domain adaptation scheme, the proposed method achieves the best F1-macro score, indicating that learning semantic-aware feature alignment between the two domains and semantic label propagation within the class in the target domain is valid for the semi-supervised CAC task. The performance can be attributed to the alignment of semantic-aware features between the two domains and the full supplement of semantic-aware target information.

To sum up, the proposed model shows the competitive performance among these methods. It only requires annotating extremely limited heartbeat data of the new subjects, which is relatively practical and highly operable.

5.2. Ablation study

To examine the contributions of the SAFA and PBLP modules to the semi-supervised CAC task, we first define several notations.

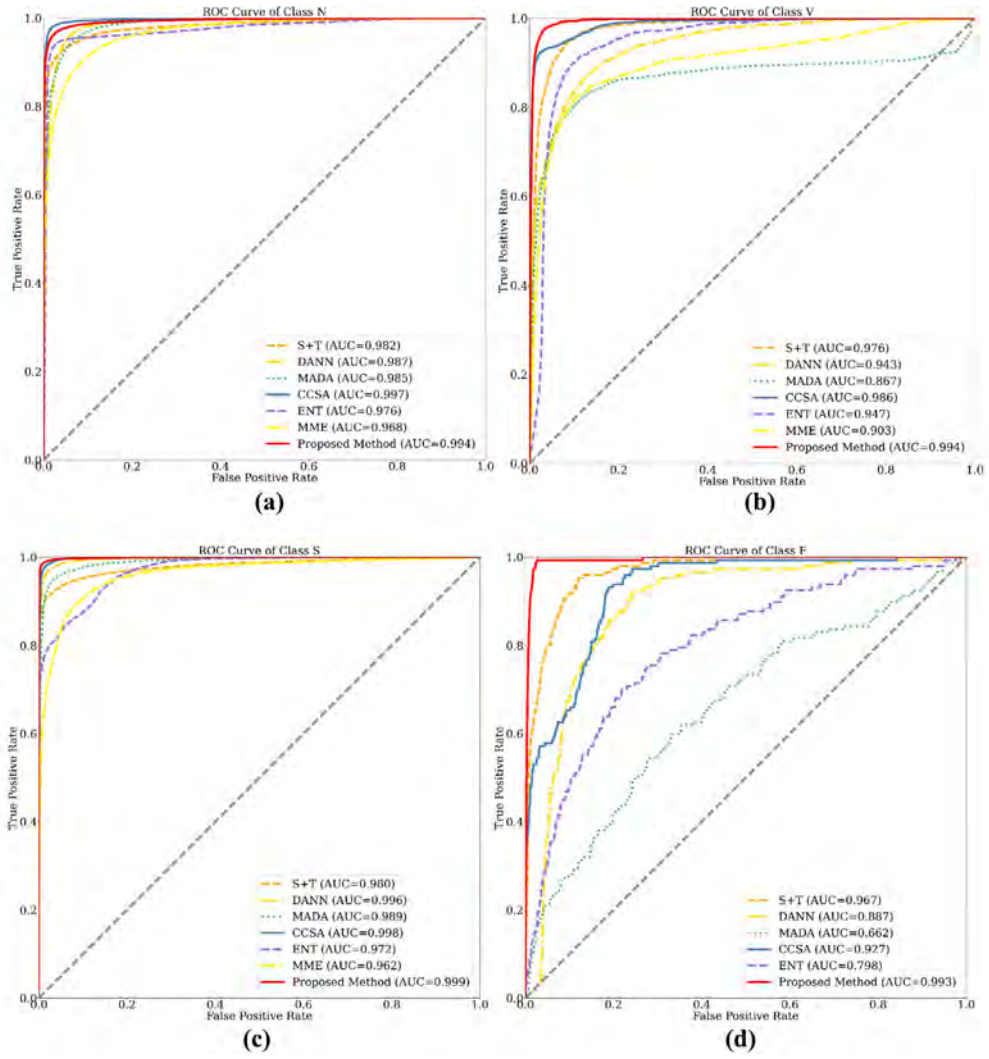

Fig. 4. ROC curve of competitors in each category on INCART.

Table 4
Ablation studies of SAFA and PBLP modules on INCART.

Method	N (%)			V (%)			S (%)			F (%)		
	Sen	Ppr	F1	Sen	Ppr	F1	Sen	Ppr	F1	Sen	Ppr	F1
Source Only	99.3 ± 0.03	80.9 ± 0.05	89.2 ± 0.03	62.1 ± 0.09	87.3 ± 0.07	72.6 ± 0.07	10.1 ± 0.03	77.1 ± 0.19	17.9 ± 0.05	0.2 ± 0.02	12.9 ± 0.10	0.5 ± 0.04
SDA	**99.9 ± 0.05**	82.9 ± 0.26	90.6 ± 0.19	88.6 ± 0.08	81.7 ± 0.64	85.0 ± 0.31	8.3 ± 0.63	97.2 ± 1.75	15.3 ± 1.11	1.6 ± 0.01	71.3 ± 0.74	3.9 ± 1.07
SAFA-label	98.4 ± 0.18	96.9 ± 0.05	97.7 ± 0.07	92.8 ± 0.01	89.3 ± 0.07	91.0 ± 0.04	30.5 ± 0.02	69.2 ± 0.03	42.3 ± 0.02	6.5 ± 0.41	28.4 ± 1.53	10.6 ± 0.72
SAFA + CBPL	90.5 ± 0.01	94.4 ± 0.04	92.4 ± 0.02	97.2 ± 0.03	**92.2 ± 0.03**	94.7 ± 0.01	19.0 ± 0.04	**98.0 ± 0.04**	31.8 ± 0.06	**7.6 ± 0.19**	55.1 ± 0.03	**13.4 ± 0.30**
SAFA + PBLP	99.3 ± 0.03	**97.4 ± 0.03**	**98.3 ± 0.03**	**99.6 ± 0.03**	**92.2 ± 0.06**	**95.6 ± 0.04**	**40.6 ± 0.25**	93.5 ± 0.04	**56.6 ± 0.24**	6.6 ± 0.12	**89.7 ± 0.80**	12.4 ± 0.23

Specifically, 'Source Only' represents that the model is trained with only labeled source data, and then directly evaluates on the target domain, which serves as the baseline method. SDA denotes that the model performs the global feature alignment between the source domain and the target domain with labeled data. SAFA-label means that the SAFA model performs semantic-aware feature alignment between the two domains with only available labeled data. SAFA + CBPL represents the model performs semantic-aware feature alignment between the two domains with labeled and pseudo labeled samples, where the pseudo labeled target data are obtained by the classifier-based pseudo labeling module (CBPL) and the classifier is trained on the source domain. SAFA + PBLP represents the proposed model. In light of the above notations and interpretations, we compare these models in the semi-supervised CAC task: ARDB → INCART, and their results on INCART are reported in Table 4 and Fig. 5. In specific, compared with the results of SDA and SAFA-label, a certain average performance improvement is achieved by the SAFA-label model. After that, compared with the results of SAFA-label, SAFA + CBPL and SAFA + PBLP, we can see that the performance of the model SAFA + PBLP outperforms the performance of SAFA-label or SAFA + CBPL. This indicates that the proposed SAFA and PBLP modules are complementary. Concretely, the SAFA focuses on aligning semantic-aware category feature representations between two domains and learning more compact semantic-aware representations. At the same time, PBLP focuses on providing more reliable target training samples to further alleviate the distribution discrepancies between the two domains, which can provide more robust target features. Therefore, when the SAFA and PBLP modules are included at the same time, the model performance performs clearly the best.

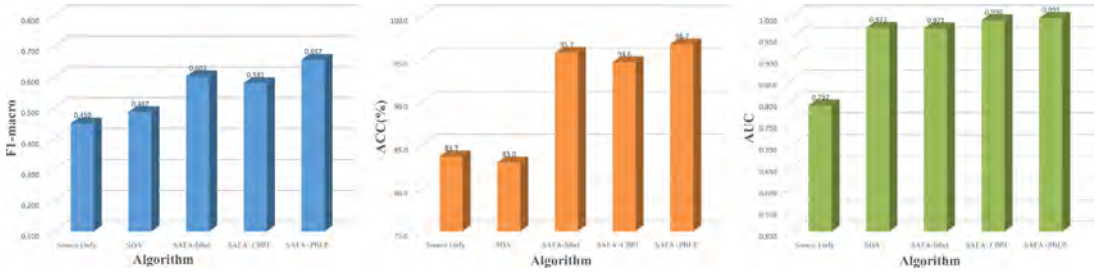

Fig. 5. Ablation studies of SAFA module and PBLP module on INCART.

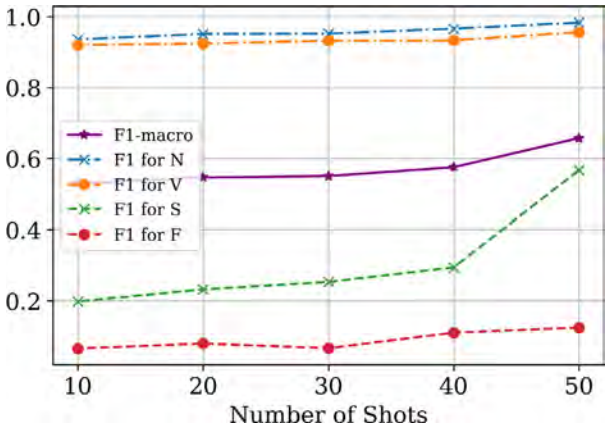

Fig. 6. The effect of the number of labeled target data on INCART.

5.3. The number of labeled target samples

We discuss the performance of the proposed method when the number of labeled target samples varies in the process of training. Specifically, five scenarios are conducted on the semi-supervised CAC task: ARDB → INCART, where the number of labeled target samples in each category, i.e., the number of shots, only varies during the training process. Five scenarios include from 10 to 50 shots. In the process of model training, labeled target data is required in two stages. Concretely, in the pre-training stage, labeled target samples are required to perform the SAFA module. In the pseudo-labeling stage, labeled target samples are required to perform the PBLP module. The effect of the number of the labeled target data on INCART is illustrated in Fig. 6. The results illustrate that the model performance can be improved consistently when more labeled target samples are available.

5.4. Generalization evaluation

To further verify the generalization ability of the proposed model, two different distributional shift tasks are conducted: ARDB → SVDB and ARDB → LTDB. Specifically, comprehensive quantitative experiments on the task ARDB → SVDB are performed and the corresponding results are reported in Tables 5 and 6, and the corresponding qualitatively visual analyses are illustrated in Fig. 8. In addition, the quantitative results on the task ARDB → LTDB are reported in Tables 7 and 8, and the qualitative analyses are illustrated in Fig. 9.

The quantitative results on the SVDB dataset are reported in Tables 5 and 6. The performance of the proposed method outperforms that of the comparison methods, especially for category V. The superior classification performance is largely attributed to the semantic-aware specific category feature alignment and the reliable representations obtained by effectively guiding the label propagation through the utilization of limited labeled target samples.

For the CAC task: ARDB → LTDB, Tables 7 and 8 show the quantitative results of each category and the overall performance on LTDB. It can be seen that the proposed method achieves competitive model performance. Specifically, our method performs much better than the model S + T when the limited labeled target samples are given, indicating that the simple fine-tuning strategy has a larger room for improvement. In addition, the performance of the proposed method in different categories is better than that of the DANN model. The excellent performance can be attributed to the specific training of semantic-aware domain discriminators and reliable feature representations. They can help reduce the category-level feature distribution gap between the source domain and the target domain, and retain semantic consistency within the class between the two domains.

In general, we can find that the performance of the proposed method is superior to that of most state-of-the-art methods on the SVDB and LTDB benchmarks, which denotes that the semantic-aware adversarial learning mechanism and prototype-based label propagation strategy are effective for the semi-supervised CAC task. In addition, we can conclude that the performance improvement of the proposed method on LTDB benchmark is not as significant as other benchmarks. An interpretation is that the number of different categories on the LTDB benchmark is highly imbalanced. In the future work, we will further make attempts to explore several effective strategies to deal with the challenge of imbalanced training samples.

5.5. Visual analysis

For intuitive qualitative analysis, t-SNE [45] is utilized to visualize the feature representations of the source domain and the target domain on three semi-supervised CAC tasks: ARDB → INCART, ARDB → SVDB, and ARDB → LTDB, and their corresponding results are illustrated in Figs. 7, 8 and 9, respectively. Specifically, from Fig. 7, the 'Source Only' or S + T model cannot well align the representations of source domain and target domain, while the DANN or MADA can better align, but the DANN or MADA cannot discriminate the categories well. Besides, SAFA-label can better align the representations of the two domains and the categories can be discriminated well, while the proposed method is evidently better than SAFA-label. Specifically, the representations of the same category between the two domains are compact, and the representations of different categories between the source and target domain lie far away from each other. In addition, comparing Figs. 7(a), 8(a), and 9(a), we can find that there are significant distribution discrepancies between the feature representation spaces of the same category in the two domains. In addition, comparing Fig. 8(a)–(c) (Fig. 9(a)–(c)), the

Table 5
Comparison results of each category on SVDB.

Algorithm	N (%)			V (%)			S (%)			F (%)		
	Sen	Ppr	F1	Sen	Ppr	F1	Sen	Ppr	F1	Sen	Ppr	F1
Full-T	99.7 ± 0.10	96.4 ± 0.75	98.1 ± 0.38	95.2 ± 0.35	94.8 ± 0.87	95.0 ± 0.31	67.6 ± 2.43	97.3 ± 1.40	79.8 ± 1.25	49.9 ± 0.03	99.9 ± 0.00	66.6 ± 0.01
Al Rahhal et al. [6]	–	–	–	65.2	9.3	16.0	8.8	14.3	11.0	–	–	
Wang et al. [7]	–	–	–	85.7	48.3	61.8	25.4	36.7	30.0	–	–	
Guo et al. [46]	–	–	–	**86.8**	58.8	70.1	7.90	64.5	14.1	–	–	
Wang et al. [9]	–	–	–	84.3	56.2	68.0	23.6	53.8	33.0	–	–	–
Li et al. [10]	–	–	–	78.5	72.4	75.3	23.8	47.2	31.6	–	–	–
S + T [34]	98.4 ± 0.11	74.6 ± 0.62	85.2 ± 0.01	33.1 ± 0.91	88.1 ± 0.66	48.4 ± 0.31	68.9 ± 0.70	34.1 ± 0.43	43.9 ± 0.53	0.01 ± 0.01	68.4 ± 0.12	0.03 ± 0.02
DANN [27]	96.8 ± 0.05	72.2 ± 0.45	82.7 ± 0.28	25.9 ± 0.97	92.2 ± 0.02	40.5 ± 1.19	26.5 ± 1.11	28.7 ± 1.08	27.6 ± 1.10	0.0 ± 0.00	6.2 ± 0.02	0.1 ± 0.01
MADA [31]	97.4 ± 0.22	92.9 ± 0.78	95.1 ± 0.30	53.8 ± 0.53	89.5 ± 1.09	67.2 ± 0.72	52.6 ± 1.00	45.3 ± 0.35	48.7 ± 0.22	0.45 ± 0.30	37.3 ± 0.12	0.88 ± 0.59
CCSA [43]	95.6 ± 0.80	96.0 ± 0.89	95.8 ± 0.10	75.5 ± 0.82	84.4 ± 1.28	77.3 ± 0.87	53.2 ± 0.15	42.7 ± 0.44	47.3 ± 0.39	0.0 ± 0.00	0.0 ± 0.00	0.0 ± 0.00
ENT [33]	**99.5 ± 0.12**	83.9 ± 0.09	91.0 ± 0.10	68.2 ± 0.22	87.2 ± 0.11	76.5 ± 0.03	30.0 ± 1.01	84.3 ± 0.44	44.3 ± 0.38	2.22 ± 0.40	87.4 ± 1.20	**4.3 ± 0.60**
MME [32]	99.1 ± 0.05	89.5 ± 0.13	94.1 ± 0.08	79.5 ± 0.41	73.7 ± 0.11	78.1 ± 0.02	40.7 ± 1.01	**87.5 ± 0.44**	55.5 ± 0.38	0.77 ± 0.37	**93.7 ± 0.52**	1.5 ± 0.24
SAFA-label	96.8 ± 0.02	94.5 ± 0.02	95.6 ± 0.02	66.2 ± 0.17	86.4 ± 0.04	75.0 ± 0.12	54.2 ± 0.05	49.9 ± 2.56	51.5 ± 0.57	**0.51 ± 0.03**	49.9 ± 0.02	1.0 ± 0.08
Source Only	97.4 ± 0.39	70.2 ± 0.77	81.4 ± 0.50	22.7 ± 0.06	**95.1 ± 0.05**	36.1 ± 0.07	26.6 ± 0.09	27.8 ± 0.08	27.0 ± 0.08	0.04 ± 0.002	6.2 ± 0.03	0.08 ± 0.04
Proposed method	98.5 ± 0.04	**97.5 ± 0.03**	**98.0 ± 0.03**	85.1 ± 0.06	92.0 ± 0.08	**88.4 ± 0.06**	**78.2 ± 0.06**	75.4 ± 0.07	**76.8 ± 0.05**	0.1 ± 0.02	12.4 ± 0.03	0.3 ± 0.003

Table 6
Overall performance comparison on SVDB.

	Full-T	S + T [34]	DANN [27]	MADA [31]	CCSA [43]	ENT [33]	MME [32]	Source Only	Proposed method
F1-macro	0.849 ± 0.40	0.419 ± 0.05	0.446 ± 0.24	0.529 ± 0.01	0.533 ± 0.10	0.540 ± 0.42	0.573 ± 0.34	0.361 ± 0.05	**0.659 ± 0.11**
ACC (%)	96.0 ± 0.84	71.4 ± 1.87	79.2 ± 0.48	89.5 ± 0.65	91.5 ± 0.30	84.1 ± 0.32	88.7 ± 0.53	70.4 ± 0.00	**96.0 ± 0.27**
AUC	0.995 ± 0.14	0.871 ± 0.70	0.759 ± 0.35	0.932 ± 0.49	0.890 ± 0.17	0.937 ± 0.03	0.951 ± 0.06	0.767 ± 1.05	**0.979 ± 0.28**

Table 7
Comparison results of each category on LTDB.

Algorithm	N (%)			V (%)			S (%)			F (%)		
	Sen	Ppr	F1	Sen	Ppr	F1	Sen	Ppr	F1	Sen	Ppr	F1
Full-T	99.9 ± 0.05	97.3 ± 0.80	98.6 ± 0.40	99.4 ± 0.08	98.5 ± 0.88	98.9 ± 0.41	9.21 ± 0.05	98.6 ± 0.39	16.8 ± 0.08	61.7 ± 0.42	95.2 ± 1.56	74.5 ± 0.94
S + T [34]	99.8 ± 0.21	62.3 ± 0.14	73.7 ± 0.43	87.7 ± 0.01	98.0 ± 0.15	92.5 ± 0.06	0.61 ± 0.52	92.3 ± 0.22	1.21 ± 0.65	39.3 ± 0.59	71.4 ± 1.80	50.7 ± 0.62
DANN [27]	98.3 ± 1.21	68.4 ± 0.24	80.7 ± 0.76	29.5 ± 0.21	76.9 ± 1.26	42.6 ± 0.74	0.43 ± 1.32	19.0 ± 0.54	0.85 ± 0.73	4.78 ± 1.35	35.3 ± 3.21	8.42 ± 1.58
MADA [31]	**99.8 ± 0.03**	92.4 ± 0.01	95.9 ± 0.01	93.0 ± 1.04	95.3 ± 1.71	**94.1 ± 0.82**	4.06 ± 2.05	94.7 ± 0.03	7.80 ± 0.97	19.6 ± 0.85	**95.4 ± 2.72**	32.5 ± 1.61
CCSA [43]	98.0 ± 0.13	83.8 ± 0.21	90.3 ± 0.11	**99.3 ± 2.51**	74.0 ± 1.63	84.8 ± 1.54	60.2 ± 2.05	61.2 ± 1.07	**60.7 ± 1.86**	3.84 ± 0.65	0.03 ± 1.12	0.07 ± 0.91
ENT [33]	94.1 ± 0.15	**98.2 ± 0.51**	96.1 ± 0.35	49.5 ± 2.13	73.1 ± 1.82	59.0 ± 1.05	**83.9 ± 1.74**	1.67 ± 0.79	3.28 ± 0.98	0.0 ± 0.00	0.0 ± 0.00	0.0 ± 0.00
MME [32]	99.8 ± 0.08	35.2 ± 0.17	52.1 ± 0.54	61.6 ± 0.54	**98.6 ± 1.74**	75.8 ± 2.24	0.40 ± 1.82	**97.2 ± 1.48**	7.59 ± 1.65	22.2 ± 1.54	77.9 ± 2.37	34.6 ± 2.02
SAFA-label	98.5 ± 0.28	95.8 ± 0.19	97.1 ± 0.31	97.7 ± 0.17	80.5 ± 0.25	88.3 ± 0.20	3.86 ± 0.06	71.2 ± 0.07	7.33 ± 0.11	44.4 ± 0.32	80.0 ± 0.52	**57.1 ± 0.36**
Source Only	98.8 ± 0.11	58.1 ± 0.09	73.2 ± 0.07	22.6 ± 0.13	81.1 ± 0.09	35.3 ± 0.16	0.52 ± 0.05	5.3 ± 0.72	0.7 ± 0.30	0.51 ± 0.007	5.3 ± 0.03	0.9 ± 0.01
Proposed method	99.3 ± 0.02	96.7 ± 0.04	**98.0 ± 0.02**	84.8 ± 0.06	94.6 ± 0.12	89.5 ± 0.03	9.46 ± 0.09	67.9 ± 0.56	16.6 ± 0.15	**47.9 ± 0.08**	39.0 ± 0.06	43.0 ± 0.02

Table 8
Overall performance comparison on LTDB.

	Full-T	S + T [34]	DANN [27]	MADA [31]	CCSA [43]	ENT [33]	MME [32]	Source Only	Proposed method
F1-macro	0.722 ± 0.29	0.545 ± 0.26	0.331 ± 0.41	0.576 ± 0.79	0.589 ± 0.82	0.396 ± 0.47	0.425 ± 0.53	0.275 ± 0.08	**0.617 ± 0.04**
ACC (%)	97.9 ± 0.52	65.8 ± 0.62	69.0 ± 0.73	92.7 ± 0.35	**96.8 ± 1.57**	90.5 ± 0.46	41.6 ± 0.21	60.0 ± 0.52	96.2 ± 0.12
AUC	0.996 ± 0.36	0.977 ± 0.42	0.843 ± 0.52	0.971 ± 0.71	**0.992 ± 0.78**	0.762 ± 0.78	0.935 ± 0.43	0.784 ± 0.54	0.970 ± 0.42

representations of the same category between the source and target domain by the proposed model in Fig. 8(c) (Fig. 9(c)) are compact, and the representations of different categories between the source and target domain lie far away from each other in the 2D subspace. An interpretation is that more compact feature learning obtained by the proposed method can help the model capture more reliable heartbeat type prototypes. Thus the prototypes-based label labeling strategy provides more reliable target training samples for further alleviating distribution discrepancies between the two domains, which can provide more robust target features.

6. Conclusions

We have explored two critical issues that hinder the performance of the semi-supervised CAC: the gaps in the distribution of semantic features across domains and insufficient semantic information in the target domain. Accordingly, we propose a unified framework to address these issues simultaneously. The framework comprises a SAFA module and a PBLP module. First, the SAFA module performs semantic-aware feature alignment to alleviate inter-domain discrepancies. Second, the PBLP module provides reliable pseudo labels in the target domain through the intra-class label propagation scheme. The two modules are complementary. The semantic-aware domain-invariant feature representations learned by the former are beneficial for extracting more accurate prototypes for the latter, and the latter provides high-quality pseudo training samples to promote more robust representation learning.

Advantages: (1) Comprehensive quantitative and qualitative experiments demonstrate the superiority of the proposed method; (2) The proposed model requires extremely limited annotated training samples from the target domain, relieving the annotation burden; (3) The SAFA can not only reduce the global gap

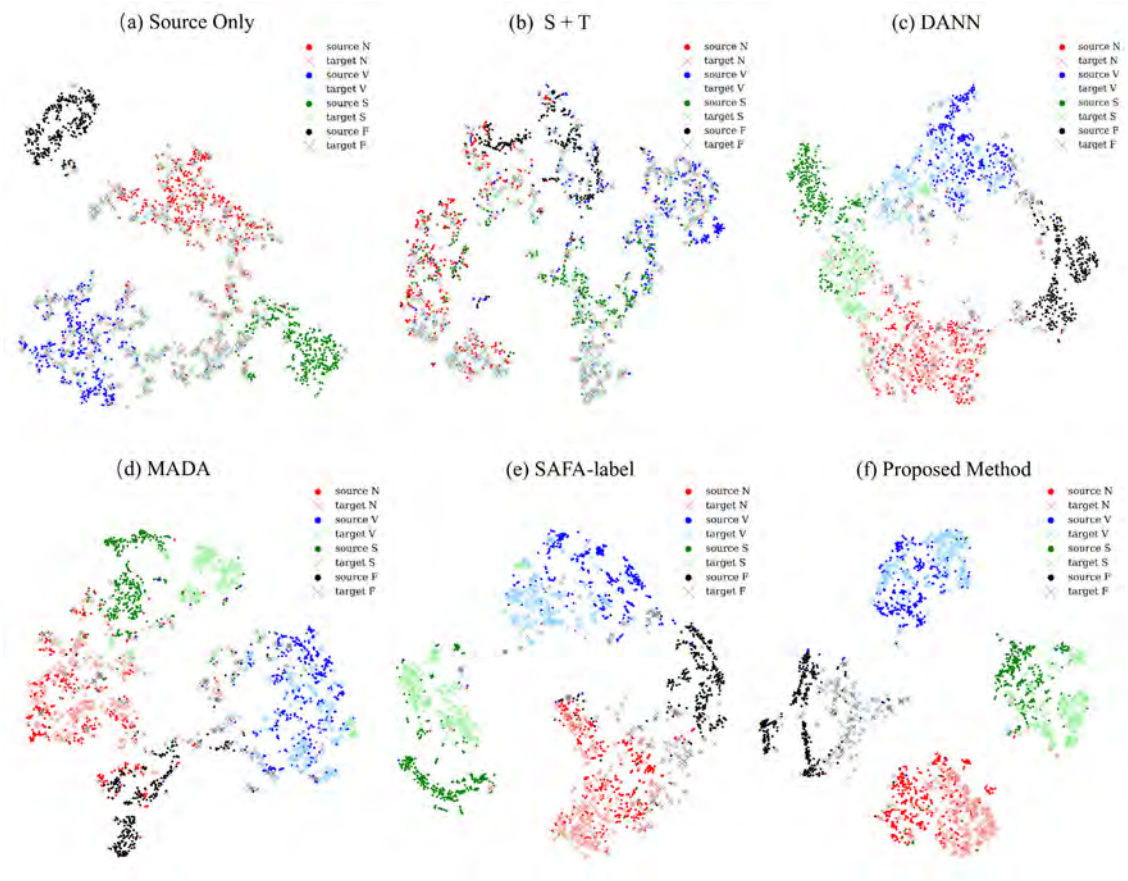

Fig. 7. ARDB → INCART. Comparisons of our method with other methods via using t-SNE [45] to visual the feature representations. (a) Source Only. (b) S + T. (c) DANN. (d) MADA. (e) SAFA-label. (f) Proposed Method.

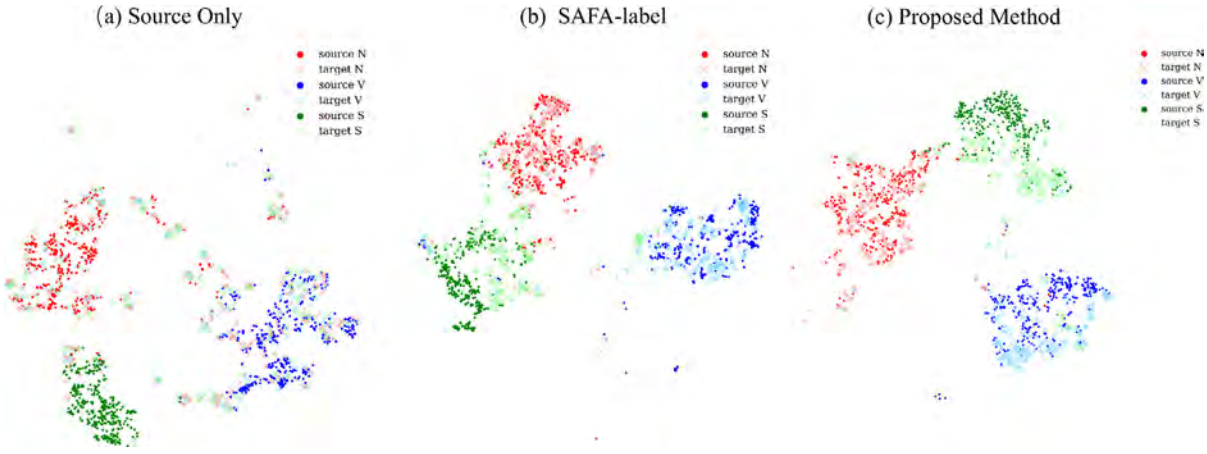

Fig. 8. ARDB → SVDB. Comparisons of our method with other methods via using t-SNE [45] to visual the feature representations. (a) Source Only. (b) SAFA-label. (c) Proposed Method.

between two domains but also retain the feature consistency of the same category in different domains, where the category-level domain discriminator specifically for each sample is trained during the adversarial training. (4) The PBLP is proposed to mine high-quality target pseudo labels by considering the semantic information, so as to improve the model performance. The semantic information here refers to the semantic similarity of the same category and the semantic discrepancies between different categories in the same domain.

Disadvantages: (1) When the number of training samples belonging to different categories is extremely imbalanced, the imbalance problem leads to performance degradation, which significantly challenges the domain adaptation classification task. (2) Granted ECG data is highly sensitive and private, and it can reveal

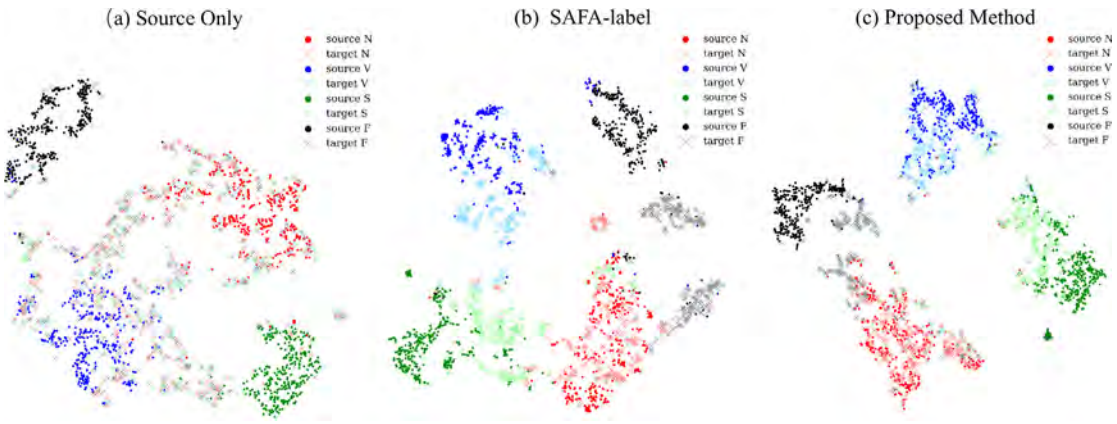

Fig. 9. ARDB → LTDB. Comparisons of our method with other methods via using t-SNE [45] to visual the feature representations. (a) Source Only. (b) SAFA-label. (c) Proposed Method.

the disease status of patients. Moreover, it can also be used for human identification. Therefore, the proposed method requires to access the source data during learning to adapt, which is not efficient for data transmission and may violate the data privacy policy.

In the future work, we will attempt to solve the data imbalance and privacy concern problems, and design a more robust and general arrhythmia classification model.

CRediT authorship contribution statement

Panpan Feng: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Visualization. **Jie Fu:** Conceptualization, Methodology, Formal analysis, Writing – review & editing. **Ning Wang:** Visualization, Investigation, Resources. **Yanjie Zhou:** Writing – review & editing, Project administration. **Bing Zhou:** Funding acquisition. **Zongmin Wang:** Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported in part by The Key Research, Development, and Dissemination Program of Henan Province (Science and Technology for the People), China under Grant 182207310002 and in part by the Young Talents Enterprise Cooperative Innovation Team grant for Zhengzhou University, China under Grant 32320424.

References

[1] Y. Hagiwara, H. Fujita, S.L. Oh, J.H. Tan, R. San Tan, E.J. Ciaccio, U.R. Acharya, Computer-aided diagnosis of atrial fibrillation based on ECG signals: A review, Inform. Sci. 467 (2018) 99–114.

[2] H. Wang, H. Dai, Y. Zhou, B. Zhou, P. Lu, H. Zhang, Z. Wang, An effective feature extraction method based on GDS for atrial fibrillation detection, J. Biomed. Inform. 119 (2021) 103819.

[3] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nature Med. 25 (1) (2019) 65–69.

[4] H. Wang, Y. Zhou, B. Zhou, X. Niu, H. Zhang, Z. Wang, Interactive ECG annotation: An artificial intelligence method for smart ECG manipulation, Inform. Sci. 581 (2021) 42–59.

[5] C. Zhang, Q. Zhao, Deep discriminative domain adaptation, Inform. Sci. 575 (2021) 599–610.

[6] M.M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, R.R. Yager, Deep learning approach for active classification of electrocardiogram signals, Inform. Sci. 345 (2016) 340–354.

[7] G. Wang, C. Zhang, Y. Liu, H. Yang, D. Fu, H. Wang, P. Zhang, A global and updatable ECG beat classification system based on recurrent neural networks and active learning, Inform. Sci. 501 (2019) 523–542.

[8] L. Niu, C. Chen, H. Liu, S. Zhou, M. Shu, A deep-learning approach to ECG classification based on adversarial domain adaptation, Healthcare 8 (4) (2020) 437.

[9] G. Wang, M. Chen, Z. Ding, J. Li, H. Yang, P. Zhang, Inter-patient ECG arrhythmia heartbeat classification based on unsupervised domain adaptation, Neurocomputing 454 (2021) 339–349.

[10] J. Li, G. Wang, M. Chen, Z. Ding, H. Yang, Mixup asymmetric tri-training for heartbeat classification under domain shift, IEEE Signal Process. Lett. 28 (2021) 718–722.

[11] P. Feng, J. Fu, Z. Ge, H. Wang, Y. Zhou, B. Zhou, Z. Wang, Unsupervised semantic-aware adaptive feature fusion network for arrhythmia detection, Inform. Sci. 582 (2022) 509–528.

[12] S. Hanneke, S. Kpotufe, On the value of target data in transfer learning, Proc. Adv. Neural Inf. Process. Syst. 32 (2019) 9871–9881.

[13] X. Liu, H. Wang, Z. Li, L. Qin, Deep learning in ECG diagnosis: A review, Knowl.-Based Syst. 227 (2021) 107187.

[14] L. Meng, W. Tan, J. Ma, R. Wang, X. Yin, Y. Zhang, Enhancing dynamic ECG heartbeat classification with lightweight transformer model, Artif. Intell. Med. 124 (2022) 102236.

[15] N. Ammour, H. Alhichri, Y. Bazi, N. Alajlan, LwF-ECG: Learning-without-forgetting approach for electrocardiogram heartbeat classification based on memory with task selector, Comput. Biol. Med. 137 (2021) 104807.

[16] Y. Liu, Y. Jin, J. Liu, C. Qin, K. Lin, H. Shi, J. Tao, L. Zhao, C. Liu, Precise and efficient heartbeat classification using a novel lightweight-modified method, Biomed. Signal Process. Control. 68 (2021) 102771.

[17] Z. Ge, X. Jiang, Z. Tong, P. Feng, B. Zhou, M. Xu, Z. Wang, Y. Pang, Multi-label correlation guided feature fusion network for abnormal ECG diagnosis, Knowl.-Based Syst. 233 (2021) 107508.

[18] P. Sharma, S.K. Dinkar, A linearly adaptive Sine–cosine algorithm with application in deep neural network for feature optimization in arrhythmia classification using ECG signals, Knowl.-Based Syst. 242 (2022) 108411.

[19] J.H. Jang, T.Y. Kim, D. Yoon, Effectiveness of transfer learning for deep learning-based electrocardiogram analysis, Healthc. Inform. Res. 27 (1) (2021) 19–28.

[20] L. Chen, G. Xu, S. Zhang, J. Kuang, L. Hao, Transfer learning for electrocardiogram classification under small dataset, in: Proceedings of the Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting, 2019, pp. 45–54.

[21] M.K. Gajendran, M.Z. Khan, M.A.K. Khattak, ECG Classification using deep transfer learning, in: Proceedings of the 4th International Conference on Information and Computer Technologies, 2021, pp. 1–5.

[22] A. Pal, R. Srivastva, Y.N. Singh, CardioNet: An efficient ECG arrhythmia classification system using transfer learning, Big Data Res. 26 (2021) 100271.

[23] N. Ammour, Atrial fibrillation detection with a domain adaptation neural network approach, in: Proceedings of the International Conference on Computational Science and Computational Intelligence, 2018, pp. 738–743.
[24] Y. Jin, C. Qin, J. Liu, K. Lin, H. Shi, Y. Huang, C. Liu, A novel domain adaptive residual network for automatic atrial fibrillation detection, Knowl.-Based Syst. 203 (2020) 106122.
[25] F. Deng, S. Tu, L. Xu, Multi-source unsupervised domain adaptation for ECG classification, in: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, 2021, pp. 854–859.
[26] M. Salem, S. Taheri, J.-S. Yuan, ECG arrhythmia classification using transfer learning from 2-dimensional deep CNN features, in: Proceedings of the IEEE Biomedical Circuits and Systems Conference, 2018, pp. 1–4.
[27] Y. Ganin, V. Lempitsky, Unsupervised domain adaptation by backpropagation, in: Proceedings of the International Conference on Machine Learning, 2015, pp. 1180–1189.
[28] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commu. ACM. 63 (11) (2020) 139–144.
[29] J. Jiao, M. Zhao, J. Lin, Unsupervised adversarial adaptation network for intelligent fault diagnosis, IEEE Trans. Ind. Electron. 67 (11) (2019) 9904–9913.
[30] X. Li, W. Zhang, N.X. Xu, Q. Ding, Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places, IEEE Trans. Ind. Electron. 67 (8) (2019) 6785–6794.
[31] Z. Pei, Z. Cao, M. Long, J. Wang, Multi-adversarial domain adaptation, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
[32] K. Saito, D. Kim, S. Sclaroff, T. Darrell, K. Saenko, Semi-supervised domain adaptation via minimax entropy, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 8050–8058.
[33] Y. Grandvalet, Y. Bengio, Semi-supervised learning by entropy minimization, in: Proceedings of the Advances in Neural Information Processing Systems, Vol. 17, 2004, pp. 529–536.
[34] W. Chen, Y. Liu, Z. Kira, Y.F. Wang, J. Huang, A closer look at few-shot classification, in: Proceedings of the International Conference on Learning Representations, 2019.
[35] D.H. Lee, et al., Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks, in: Proceedings of the Workshop on Challenges in Representation Learning, Vol. 3, 2013, p. 896.
[36] M. Chen, G. Wang, Z. Ding, J. Li, H. Yang, Unsupervised domain adaptation for ECG arrhythmia classification, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2020, pp. 304–307.
[37] Y. Zou, Z. Yu, B. Kumar, J. Wang, Unsupervised domain adaptation for semantic segmentation via class-balanced self-training, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 289–305.
[38] Y. Zou, Z. Yu, X. Liu, B. Kumar, J. Wang, Confidence regularized self-training, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 5982–5991.
[39] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, Proc. Adv. Neural Inf. Process. Syst. 30 (2017).
[40] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J.W. Vaughan, A theory of learning from different domains, Mach. Learn. 79 (1) (2010) 151–175.
[41] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.
[42] ANSI/AAMI EC57, testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms, 1998.
[43] S. Motiian, M. Piccirilli, D.A. Adjeroh, G. Doretto, Unified deep supervised domain adaptation and generalization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 5715–5725.
[44] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: an imperative style, high-performance deep learning library, Proc. Adv. Neural Inf. Process. Syst. 32 (2019) 8024–8035.
[45] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008) 2579–2605.
[46] L. Guo, G. Sim, B. Matuszewski, Inter-patient ECG classification with convolutional and recurrent neural networks, Biocybern. Biomed. Eng. 39 (3) (2019) 868–879.