2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

CDAD: A Novel High-Efficiency Cross-age Domain Adaptation ECG Diagnosis Algorithm for Adolescents

Yiyang Li[1], Zhenghan Zhang[2], Dejing Zhang[5] ,JieJia Chen[3], Zhenkun Cai[6] and Tang Tang[4,✉]

{[1] College of Computer Science and Technology，[2]The Key Laboratory of Road and Traffic Engineering, Ministry of Education, College of Transportation Engineering，[3]College of Design and Innovation，[4]School of Mechanical Engineering }，Tongji University，Shanghai 201804, China，[5]University of Shanghai for Science and Technology, Shanghai 200093, China，[6]Fudan University, Shanghai 200433, China

tangtang@tongji.edu.cn

Abstract. Cardiovascular diseases are rapidly becoming one of the major health problems in adolescents. With the advancement of deep learning techniques, smart ECG diagnostic tools based on these techniques show great potential for application in real-world healthcare settings. However, the scarcity of ECG data in adolescents compared to older adults is a key challenge for deep learning techniques, the accuracy of which relies on extensive labeled training data. In this paper, we propose a Crossage Domain Adaptation Diagnosis (CDAD) approach and introduce a domain adaptation network, Squeeze and Excitation Widekernel Neural Network (SEWNN), aiming to alleviate the constraints imposed by unlabeled data and cross-domain diagnosis. Firstly, Largescale labeled ECG data from elderly individuals are employed for feature extraction and model training. Subsequently, an adversarial learning approach is employed to enhance the model's cross-domain transfer capabilities. In addition, Ensemble learning techniques that consider information from multiple cues to improve prediction accuracy are applied. In this study, we validate the effectiveness of the proposed method by applying it to three public ECG diagnostic datasets and evaluating its applicability from the elderly to adolescents. By comparing the experimental results with other methods, we demonstrate the validity of the method in adolescents diagnosing ECG, as well as its robustness in cross-dataset diagnosis.

Keywords: ECG Diagnosis, Domain Adaptation, Adversarial Learning, Ensemble Learning.

1 Introduction

According to the Heart Disease and Stroke Statistics 2023 released by The American Heart Association and the National Institutes of Health [28]. Cardiovascular diseases (CVDs) are among the leading causes of mortality globally [19, 24]. Arrhythmia is a

frequent cardiac abnormality, which is often linked to sudden cardiac events. Electrocardiography (ECG), serving as a simple and non-invasive diagnostic method, assesses the cardiac health status by recording around ten seconds of cardiac electrical activity. Experienced physicians can usually diagnose heart disease in patients by analyzing the ECG.

With advancements in machine learning [11] and deep learning [14], numerous computational methods for automated ECG diagnosis have been developed, significantly enhancing the efficiency of ECG diagnosis [20, 21, 5, 4]. Hou et al. [22] proposed an ECG classification method based on Long Short-Term Memory (LSTM). Li et al. [16] employed 1-D Convolutional Neural Networks (CNNs) for the classification and diagnosis of ECG. Obeidat et al. [23] introduced a hybrid algorithm known as 1-D CNN-LSTM, which achieves commendable classification performance while maintaining a lightweight structure. However, in classification tasks, different age groups often demonstrate distinct distributions of symptom labels. Moreover, as age increases, publicly available datasets with age information [1, 30, 36] often lack adequate data on arrhythmias in adolescents. Diseases affecting adolescents differ from those impacting other age groups, and deep learning models often exhibit poor performance on datasets with different label distributions [14, 37]. Hence, research on ECG diagnosis under limited data and across diverse age groups becomes particularly significant. This approach is referred to as cross-domain fewshot learning. The differences between the source and target domains, known as domain shifts, significantly negatively impact the model's performance. Transfer learning, as a method for solving cross-domain problems, aims to transfer a model that performs well on the source domain to the target domain at a minimal cost of loss, so that it maintains a good generalization ability. The main challenges faced in studies of arrhythmia classification across domains are the lack of heartbeat data with labels and the variability in data distribution between domains. The current strategies for addressing these issues primarily include fine-tuning techniques and domain adaptation methods. For example, researchers such as Li[17] proposed to train patient-specific classifiers by fine-tuning a general-purpose convolutional neural network into a task-specific CNN. Domain adaptation, as an important branch of transfer learning, aims to learn shared features between two domains. Wang et al. [31] further proposed an ECG heart rate classification model combining CNN and unsupervised domain adaptation techniques to reduce distributional bias. Although domain adaptation methods have been widely used in the field of computer vision, their application in ECG diagnosis is still relatively limited, especially in solving the problem of sparse categories in the diagnostic samples of ECGs of adolescents, for which there is a lack of specialized solutions. Moreover, a standard ECG usually consists of 12-lead ECG signal data. When using deep learning methods for diagnosis, the direct correlation between the 12 leads cannot be ignored. However, existing work directly inputs the 12-lead data into a neural network for training, which not only fails to fully capture the significant features of each lead, but also consumes a large amount of computational resources.

To address these challenges, we propose a Cross-age Domain Adaptation Diagnosis method (CDAD) to alleviate the problem of limited data and labeling in adolescent ECG diagnosis. First, we denoise the ECG data and designate the rich elderly ECG data

as the source domain while the limited adolescent ECG data as the target domain. We propose a new network, Squeeze and Excitation Widekernel Neural Network (SEWNN), which integrates a channel attention mechanism and a CNN with a wide first-layer kernel as the backbone network. And a multi-adversarial domain adaptation approach is used to align the source and target domains, which improves the diagnostic performance of adolescent ECG data. In addition, we propose an ensemble method based on the self-attention mechanism that groups together multiple cues extracted from different body locations, and integrates the predictions of these groups to improve the diagnostic performance. The main contributions of this study are as follows:

- We propose a Cross-age Domain Adaptation Diagnosis (CDAD) method that utilizes a multi-adversarial domain adaptation approach to transfer the elderly ECG data classifier to the diagnosis of adolescent ECG data, thereby addressing the lack of data as well as distributional bias in adolescent ECG diagnosis.
- We introduce a novel network, Squeeze and Excitation Wide-kernel Neural Network (SEWNN), which combines deep neural networks with wide kernels and channel attention mechanism to better learn domain invariant features between source and target domains.
- We design an ensemble learning method based on a self-attention mechanism, which improves the diagnostic performance by integrating the diagnostic results from different cues to capture the salient features of the leads in each ECG signal. In addition, the parallel computing capability of our method greatly reduces the computational time overhead, making less computational resources required.
- Our experiments on three public ECG diagnostic datasets reached the stateof-the-art, demonstrating the validity of the method in ECG diagnosis of adolescents.

2 Related works

2.1 Domain Adaptation

Domain Adaptation (DA) is an important sub-direction in transfer learning that aims to facilitate the seamless transfer of knowledge from the source domain to the target domain, especially when the data distributions of the two domains are significantly different [25]. Ganin et al. [6] pioneered the concept of adversarial learning in domain adaptation and introduced the Domain Adversarial Neural Network (DANN). This approach integrates the concept of generative adversarial networks (GAN) [7] into domain adaptation, where the feature extractor and the domain discriminator engage in an adversarial process.

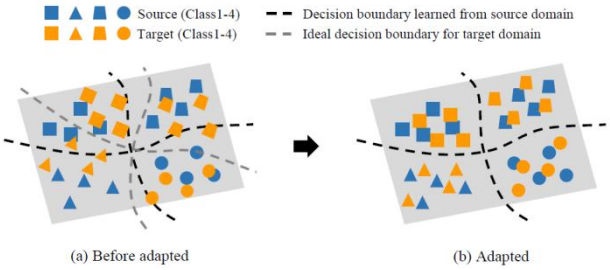


Fig. 1. Instructions for Domain adaptation. There is an obvious domain shift between the source domain and the target domain before adaptation.

2.2 CNN with wide first-layer kernels

In general classification tasks, networks with wide first-layer convolutional kernels are often considered to compromise computational efficiency. However, in diagnosis and classification problems involving longer-time signals, small convolutional kernels may not sufficiently capture signal features. In bearing fault diagnosis, networks utilizing wide first-layer kernels are already extensively employed in vibration data diagnosis [15]. WDCNN[34] is a deep convolutional neural network with wide first-layer kernels. Its network structure, depicted in Figure 2, is designed to extract features from signals and suppress high-frequency noise using wide kernels in the first convolutional layer. Moreover, WDCNN integrates Adaptive Batch Normalization (AdaBN) [18], bolstering the neural network with robust domain adaptation capabilities. In this paper, we employ CNN with wide first-layer kernels and implement corresponding enhancements to improve model performance.

Table 1. Some notation and explanation used throughout this paper.

Terms	Explanation
Source Domain	The domain which is initially trained.
Target Domain	The domain which needs to be transferred.
Domain Shift	Refers to the difference in data distributions between the source and target domains.
Unsupervised DA	Domain adaptation scenario where labeled data is available only in the source domain.
Transferability	Evaluation metric assessing how well the adapted model performs.
Feature Extractor	Extracts relevant features from input data to facilitate domain adaptation.
Domain Discriminator	Identifies the source or target domain.
Adversarial Learning	The feature extractor and domain discriminator are trained adversarially.

2.3 Ensemble Learning and Self-Attention

Ensemble learning [27] is a technique that integrates the results of multiple learners to achieve superior performance compared to any individual learner. By incorporating diagnostic results from various leads, ensemble learning can enhance diagnostic accuracy while conserving computational resources. Selfattention was initially proposed by Vaswani et al. [29] and was first utilized in the Transformer model for machine translation tasks, yielding state-of-the-art results at that time.

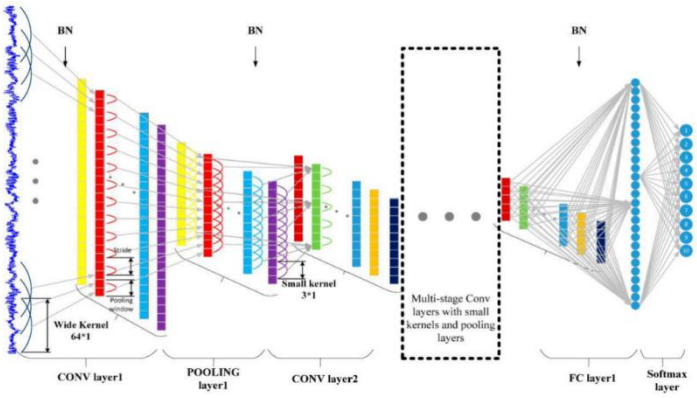


Fig. 2. The Structure of WDCNN.

3 Proposed arrhythmia diagnosis method

The structure of our proposed Cross-age Domain Adaptation Diagnosis Method (CDAD) is illustrated in Figure 3(a). It primarily comprises three modules: a backbone network (SEWNN), adversarial transfer learning blocks, and ensemble learning blocks. Pre-processed data are input into CDAD for diagnosis.

3.1 Data pre-processing

ECG signals typically consist of 12 leads, representing a 12-dimensional timeseries signal. During the acquisition of the ECG signal, it is influenced by various types of noise, primarily including baseline drift [2], power line interference[8], and muscle contraction[12]. Firstly, a Butterworth low-pass filter is employed to eliminate signals above a frequency of 50 Hz (the normal frequency range of ECG signals is between 0.5 Hz and 50 Hz). Subsequently, the ECG signal undergoes processing using Empirical Mode Decomposition (EMD) to eliminate baseline drift. Finally, the residual noise is eliminated using the Non-Local Means (NLM) technique. The denoising procedure is depicted in Figure 4.

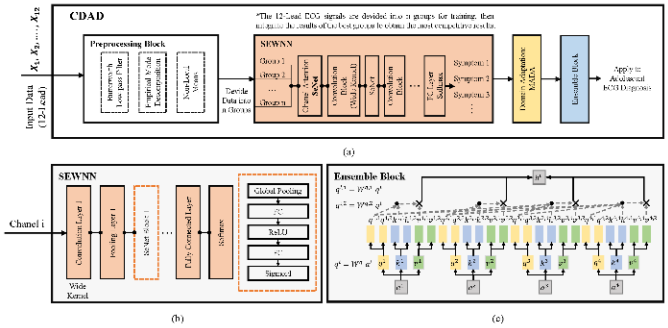


Fig. 3. The Structure of proposed CDAD method. (a) The overall framework of CDAD, including processing block, SEWNN, MADA block and ensemble block. (b) The structure of SEWNN, the part enclosed by the dotted line is the implementation details of SeNet. (c) The implementation procedure of proposed ensemble block.

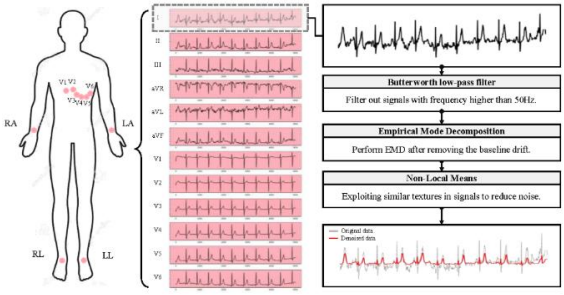


Fig. 4. The procedure of ECG signals denoising.

Table 2. The parameters of the convolution module

Layer	Layer type	Kernel size/stride	Num of kernels	Output size (W*H)	Zero-padding
1	Conv1d	64×1/16×1	16	128×16	√
2	MaxPool1d	2×1/2×1	16	64×16	
3	Conv1d	3×1/3×1	32	64×32	√
4	MaxPool1d	2×1/2×1	32	32×32	
5	Conv1d	3×1/3×1	64	32×64	√
6	MaxPool1d	2×1/2×1	64	16×64	
7	Conv1d	3×1/3×1	64	16×64	√
8	MaxPool1d	2×1/2×1	64	8×64	
9	Conv1d	3×1/3×1	64	8×64	√
10	MaxPool1d	2×1/2×1	64	6×64	
11	Conv1d	5×1/5×1	128	6×64	
12	MaxPool1d	2×1/2×1	128	3×64	
13	FC	100	1	100	
14	Softmax	4	1	4	

3.2 SEWNN

The model architecture of SEWNN (Squeeze and Excitation Wide-Kernel Neural Network) is depicted in Figure 3(b). The network contains six layers in total. Each layer contains a convolution module and a channel attention module. The parameters of each convolution module are shown in Table 2. The convolutional module utilizes a wide convolution kernel (kernel size = 64) in the first layer to extract features from the signal and suppress high-frequency noise. Subsequently, it employs multiple small kernels to capture low-frequency information from the signal. Batch normalization (BN) layers are incorporated after the convolutional layer and fully connected layer to mitigate internal covariance shift and expedite training. Dropout modules are introduced in the first convolutional module to enhance the model's generalization ability. The channel attention module uses SeNet[10], and its structure is showed in the SeNet Block in Figure 3(b). The calculation formula is given as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j) \tag{1}$$

where z_c is the output associated with the c-th channel, and u is the convolution kernel with a fixed height H and width W.

Excitation is similar to the gating mechanism in recurrent neural networks. It generates weights for each feature channel using parameters, where the parameters are learned to explicitly model the correlation between feature channels. The calculation formula is as follows:

$$\begin{aligned} s_c &= F_{ex}(\mathbf{z}, \mathbf{W}) = Sigmoid(g(\mathbf{z}, \mathbf{W})) \\ &= Sigmoid\left(\mathbf{W}_2 \operatorname{Re} Lu\left(\mathbf{W}_1 \mathbf{z}\right)\right) \end{aligned} \tag{2}$$

where $g(\ ,\ )$ is the gating network, which learns the weights for each channel through a two-layer neural network. The activation functions chosen are ReLU and Sigmoid. $W_1 = R^{\frac{C}{r} \cdot C}$ acts to reduce the dimensionality, and $W_2 = R^{C \cdot \frac{C}{r}}$ is used to restore the dimensionality.

Scale considers the weights of the output from Excitation as the importance of each feature channel after feature selection, assigning different weights to different channels. The calculation formula is as follows:

$$x_c = F_{scale}\left(\mathbf{u_c}, \mathbf{s_c}\right) = \mathbf{s_c} \cdot \mathbf{u_c} \tag{3}$$

The classification stage comprises two fully connected layers. In the output layer, the softmax function is applied to transform the logits of the neurons to comply with the probability distribution of various ECG diagnostic results.

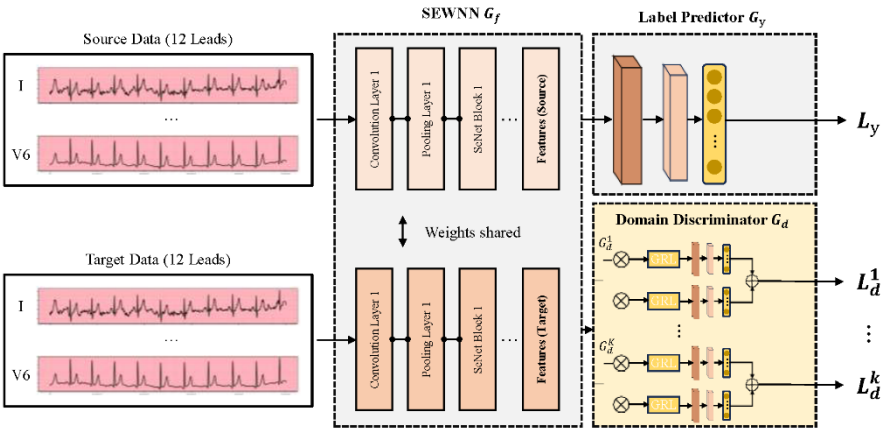


Fig. 5. The structure of domain adaptation method. G_f refers to the proposed feature extractor SEWNN, G_y refers to the label predictor, and G_y refers to the domain discriminator. L_y represents G_y' ys cross-entropy loss.

3.3 Domain adaptation method

Due to the different distributions of extracted electrocardiogram (ECG) signals from different age groups, using the same feature extractor may result in features with varying data distributions, significantly reducing the performance of the classifier. Domain adversarial networks address this issue by extracting transferable features that minimize the distribution shift between the source and target domains. The structure of the domain adaptation method is shown in Figure 5. This involves a trained domain discriminator G_d, which distinguishes between the source and target domains, and a fine-tuned feature extractor G_f , which confuses the domain discriminator. In the process of transferring from a data-rich older age group to a data-scarce younger age group in ECG diagnosis, the data distribution often exhibits multimodal characteristics. To avoid positive or negative transfer, a multi-adversarial domain adaptation method (MADA) is employed.

First, the domain discriminator G_d is decomposed into multiple discriminators G_d^k. The importance of each data point xi to K domain discriminators G_d^k, $k=1,2,\ldots,K$ is represented by the probabilities y_i. Calculate the domain loss as follows:

$$L_d = \frac{1}{n}\sum_{k=1}^{K}\sum_{x_i \in \mathcal{D}_f \cup \mathcal{D}_\square} L_d^k\left(G_d^k\left(\hat{y}_i^k\left(x_i\right)\right), d_i\right) \tag{4}$$

where G_d^k represents the k-th domain discriminator, L_d^k represents its crossentropy loss, and di denotes the domain label of point xi.

The objective of MADA is:

$$C\left(\theta_f,\theta_y,\theta_d^k\big|_{k=1}^K\right)=\frac{1}{n_s}\sum_{x_i\in\mathcal{D}_f}L_y\left(G_y\left(G_f\left(x_i\right)\right),y_i\right) -\frac{\lambda}{n}\sum_{k=1}^{K}\sum_{\mathbf{x_i}\in\mathcal{D}}L_d^k\left(G_d^k\left(\hat{y}_i^k G_f\left(\mathbf{x_i}\right)\right),d_i\right) \tag{5}$$

where $n=n_s+n_t$ is the sum of the number of source domains and target domains, $\mathcal{D}=\mathcal{D}_\rho\cap\mathcal{D}_t$ represents the intersection of the source and target domains, and λ is a balancing parameter between the two objectives in the learning process. The optimization problem aims to find a solution that satisfies:

$$\left(\theta_f,\theta_y\right)=\arg\min_{\theta_f,\theta_y}C\left(\theta_f,\theta_y,\theta_d^k\big|_{k=1}^K\right) \tag{6}$$

$$\left(\theta_d^1,\ldots,\theta_d^K\right)=\arg\max\theta_{\theta_d^1,\ldots,\theta_d^K}C\left(\theta_f,\theta_y,\theta_d^k\big|_{k=1}^K\right) \tag{7}$$

where $\theta_f,\theta_y,\theta_d^1,\ldots,\theta_d^K$ represents the learnable parameters in $G_f,G_y,G_d^k\big|_{k=1}^K$.

3.4 Self-Attention-Based Ensemble Model

The proposed approach in this paper introduces an ensemble module based on self-attention, as illustrated in Figure 3(c). Firstly, three lead combinations from the human body (Chest leads: V1-V6, Limb leads: I, II, III, Augmented limb leads: avR, avL, avF) are selected, and one lead's data from each combination is extracted to form a group. The diagnostic process is then applied to this group of ECG data using the processing steps, resulting in the network's output. Considering the computational time requirements for specific tasks, several groups with higher accuracy are selected as the output of the base-learners. The multi-head attention module is treated as the meta-learner, which integrates the results.

For any attention head, its output is computed as follows:

$$\mathbf{H}=\mathbf{AV} \tag{8}$$

where $\mathbf{A}=Soft\max\left(\frac{\mathbf{QQ^T}}{\sqrt{d_k}}\right),\mathbf{V}=\mathbf{W_V X},\mathbf{Q}=\mathbf{W_Q X},\mathbf{K}=\mathbf{W_K X},Q,K,$ and V represent the Query, Key, and Value, respectively, while $\mathbf{W}$ denotes the weight matrix.

Multi-head attention allows the model to jointly attend to information from different representations. It linearly projects the query, key, and value h times, with separate linear projections for the dimensions d_q, d_k, and dv, respectively. This generates the query, keys, and values for each projection, which are then processed in parallel using

the attention function to produce outputs of dimension d_v. These values are concatenated and projected again to generate the final output. The calculation formula is as follows:

$$\text{MultiHead}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \left(\text{head}_1 + \text{head}_2 + \ldots + \text{head}_h\right)\mathbf{W}_O \tag{9}$$

where $head_i = Attention\left(\mathbf{QW_i^Q}, \mathbf{KW_i^K}, \mathbf{VW_i^V}\right)$, $\oplus$ represents vector concatenation. The aforementioned results are fed into the integration module for training. After applying the Softmax operation, the integrated result is obtained. This result takes into account the correlations between multiple leads. Additionally, the integration learning strategy helps avoid excessive computational resource consumption associated with large-scale training.

Table 3. Basic information of Chapman University, Shaoxing People's Hospital, and PTB-XL datasets. Abbreviations: AFIB = Atrial Fibrillation, GSVT = General Supraventricular Tachycardia, SB = Sinus Bradycardia, SR = Sinus Rhythm, NORM= Normal, CD = Cardiomyopathy, STTC = ST-T changes, HYP = Hypertrophy, MI= Myocardial Infarction.

Database	Total	Age Group	Number (%)	Disease Distribution
Chapman University and Shaoxing People's Hospital Dataset	10646	Adolescent (12–19)	164 (1.5%)	AFIB (0, 0%), GSVT (52, 31.7%), SB (28, 17.1%), SR (84, 51.2%)
		Old (age > 50)	7879 (74.0%)	AFIB (2132, 27.1%), GSVT (1482, 18.9%), SB (2987, 37.8%), SR (1278, 16.2%)
PTB-XL Dataset	21837	Adolescent (12–19)	459 (2.1%)	NORM (404, 87.8%), CD (27, 5.9%), STTC (14, 3.1%), HYP (7, 1.6%), MI (7, 1.6%)
		Old (age > 50)	16098 (73.7%)	NORM (5595, 34.7%), CD (2005, 12.5%), STTC (2459, 15.3%), HYP (1100, 6.8%), MI (4939, 30.7%)

4 Experimental setup

4.1 Datasets

To evaluate the performance of the algorithm, we conduct experiments on two public datasets containing 12-lead data with age information. These datasets have been widely used in arrhythmia classification tasks and are beneficial for comparing the proposed method with other methods. The basic information of these two databases, the proportion of young people (12-19 years old) and the elderly (over 50 years old) and the distribution of diseases are shown in the Table 3.

Chapman University and Shaoxing People's Hospital Dataset The Chapman University and Shaoxing People's Hospital Dataset consists of 10,646 patients' 12-lead electrocardiograms (ECG) sampled at a rate of 500 Hz. The dataset includes 11 common arrhythmias and 67 additional cardiovascular diseases. Based on the recommendations

of cardiac experts, several rare cases were grouped together under higher-level arrhythmia types. As a result, the 11 arrhythmias were merged into 4 groups, namely Sinus Bradycardia (SB), Atrial Fibrillation (AFIB), General Supraventricular Tachycardia (GSVT), and Sinus Rhythm (SR). The image of the above disease in the time and frequency domains is shown in Figure 6.

PTB-XL Dataset The PTB-XL ECG dataset is a large-scale dataset consisting of 21,799 clinical 12-lead electrocardiograms (ECG) from 18,869 patients. Each ECG recording has a length of 10 seconds. The dataset's value lies in its comprehensive collection of various comorbidities and a substantial number of healthy control samples. The diagnoses in the dataset are distributed as follows: for simplicity, diagnostic statements have been aggregated into super classes including normal (NORM), myocardial infarction (MI), ST-T wave changes (STTC), cardiomyopathy (CD), and hypertrophy (HYP). The image of the above disease in the time and frequency domains is shown in Figure 7.

4.2 Experiment procedure

The CDAD model proposed in this paper primarily comprises three modules, delineated in Sections 3.2, 3.3, and 3.4. Initially, the time-domain vibration signals consisting of 5000 data points are denoised and then input into the model, with the output size determined by the number of symptom categories in the dataset.

The feature extractor in the feature extraction module is the SEWNN proposed in this paper. Both the classifier and the domain discriminator utilize two fully connected (FC) layers, with softmax deployed for multi-class classification in the final layer. Dropout after each layer is used to reduce the risk of overfitting. All experiments utilize Adam optimizer to train the network structure. The learning rate is set to 0.001, with a momentum of 0.9, and a step size of 0.5. The cross-entropy loss function is used as the objective function. Each experiment is trained for 100 epochs, with ten repeated experiments conducted. The datasets from each domain are partitioned into training and testing sets at an 8:2 ratio. Lastly, an ensemble learning method based on self-attention is employed to integrate the results of various lead combinations. To further assess the effectiveness of the proposed CDAD model, six different models are compared using two datasets.

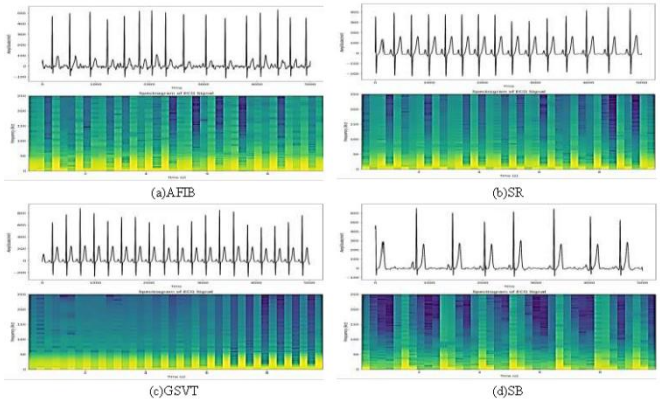


Fig. 6. Image of CUSPH dataset in the time and frequency domains.

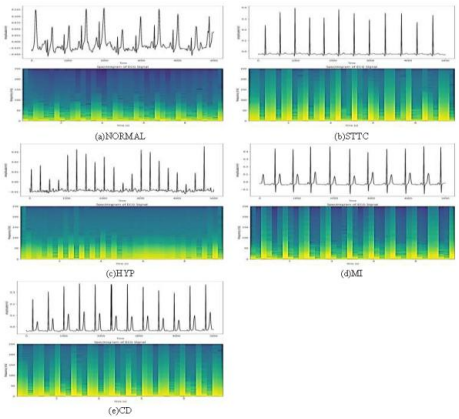
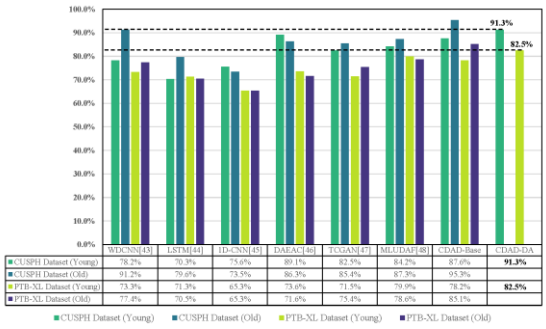


Fig. 7. Image of PTB-XL dataset in the time and frequency domains.



	WDCNN[43]	LSTM[44]	1D-CNN[45]	DAEAC[46]	TCGAN[47]	MLUDAP[48]	CDAD-Base	CDAD-DA
CUSPH Dataset (Young)	78.2%	70.3%	75.6%	89.1%	82.5%	84.2%	87.6%	**91.3%**
CUSPH Dataset (Old)	91.2%	79.6%	73.5%	86.3%	85.4%	87.3%	95.3%	
PTB-XL Dataset (Young)	73.3%	71.3%	65.3%	73.6%	71.5%	79.9%	78.2%	**82.5%**
PTB-XL Dataset (Old)	77.4%	70.5%	65.3%	71.6%	75.4%	78.6%	85.1%	

Fig. 8. The results on CUSPH and PTB-XL datasets. After domain adaptation, the test accuracy and F1 score of the proposed method significantly increased.

The experiments are divided into two parts: In the first part, we validate the performance of CDAD on a single dataset. In the second part, we evaluate CDAD's ability to diagnose across datasets. Initially, we acquire a pre-trained model, CDAD-Base, trained solely on labeled data in the source domain (elderly data), serving as the baseline. Subsequently, we employ the proposed algorithm to transfer the model from labeled source data (elderly data) to unlabeled target data (adolescent data), resulting in the transferred model, referred to as CDADDA. To assess the transferability of the proposed method, we conducted a crossdataset experiment in the second part. Initially, the model is trained on the PTB-XL dataset and subsequently transferred to the CPCS 2020 Dataset [3].

5 Experimental results

5.1 Experiment I (results on single dataset)

In this section, we compare our method with other state-of-the-art deep learning models for adolescent ECG diagnosis on the CUSPH Dataset and PTB-XL Dataset, including references[35],[26],[13], and [32]. References [33] and [9] employ the same domain adaptation concept as ours, but they do not propose a diagnosis method specifically for adolescents. We assess the accuracy of each method. Since each category is not uniform, we consider incorporating the F1 Score. The calculation formula is provided as follows:

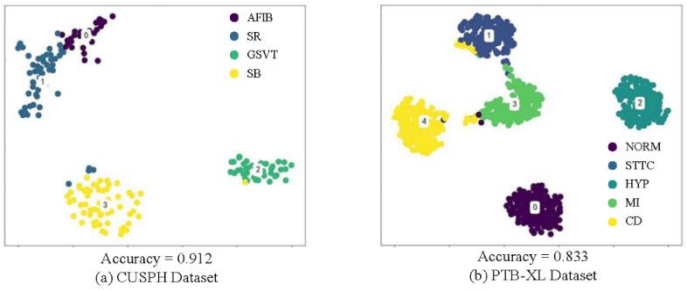


Fig. 9. t-SNE Plot on CUSPH Dataset and PTB-XL Dataset.

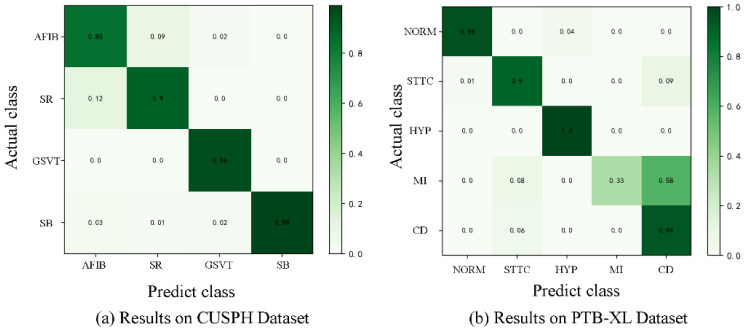


Fig. 10. Confusion Matrix of CDAD-DA on CUSPH Dataset and PTB-XL Dataset.

Table 4. The test results on CUSPH Dataset and PTB-XL Dataset (including F1 Score).

Datasets	CUSPH Dataset				PTB-XL Dataset			
	Young		Old		Young		Old	
	F1	Acc	F1	Acc	F1	Acc	F1	Acc
WDCNN[43]	0.778	0.782	0.901	0.912	0.722	0.733	0.753	0.774
LSTM[44]	0.696	0.703	0.788	0.796	0.701	0.713	0.713	0.705
1D-CNN[45]	0.705	0.756	0.712	0.735	0.665	0.653	0.688	0.653
DAEAC[46]	0.844	0.891	0.852	0.863	0.754	0.736	0.703	0.716
TCGAN[47]	0.836	0.825	0.844	0.854	0.722	0.715	0.765	0.754
MLUDAF[48]	0.856	0.842	0.857	0.873	0.801	0.799	0.763	0.786
CDAD-Base	0.863	0.876	0.944	0.953	0.776	0.782	0.845	0.851
CDAD-DA	0.902	0.913	-	-	0.809	0.825	-	-

$$\mathrm{F1} = 2 \times \frac{\mathrm{Acc} \times \mathrm{Recall}}{\mathrm{Acc} + \mathrm{Recall}} \tag{10}$$

The F1 score considers both Precision and Recall indicators, enabling a more comprehensive evaluation of the classification model's performance and mitigating the bias of a single indicator. As depicted in Figure 8 and 4, our baseline model CDAD-Base has already attained comparable performance across all categories (with better F1 scores and overall accuracy). Following adaptation, CDAD-DA achieves substantial performance enhancement and attains the highest F1 scores. The representation of CDAD-DA on various disorders is illustrated as a t-SNE plot in Figure 9. In comparison to two specific patient algorithms that train individual models for each patient, our method only requires training a global model without necessitating additional storage or computing resources. Additionally, we generated the confusion matrix of CDAD-DA on the two datasets, depicted in Figure 10. Observations on the CUSPH dataset reveal that CDAD-DA exhibits superior performance in various symptom, demonstrating strong diagnostic capabilities. Conversely, on the PTB-XL dataset, CDAD-DA demonstrates subpar diagnostic performance for MI diseases. This could be attributed to the limited number of MI cases, while exhibiting better performance on other diseases.

5.2 Experiment II (results on cross-datasets tests)

In order to prove that our method still performs well in the face of different data, rather than being singularly applicable to a certain dataset, we conducted cross-dataset tests. We used the elderly data on CPCS 2020 Dataset (The reason why we don't use CUSPH Dataset is that, CUSPH Dataset contains only the unnormal ECG signals) as the source domain and the young people data on PTB-XL Dataset as the target domain for this test. Since the disease categories are different in these 2 datasets, we consider only a dichotomous problem, i.e., whether this ECG is normal or not, and the test results are shown in Figure 11. It shows that the cross-dataset test still performs competitively comparing with the single-dataset test conducted on PTB-XL.

5.3 Ablation experiments and sensitivity analysis

Since our proposed CDAD method consists of three different parts, it's essentially important to verify the validity of each part. Therefore, we conducted ablation experiments on CUSPH Dataset, and the experimental results are shown in Table 5. Meanwhile, in the transfer learning task, the influence of the parameters cannot be ignored, so we performed sensitivity analysis on the parameter λ in Eq. (7), and the results are shown in Figure 12 as a way of verifying that our proposed method is insensitive to the parameters.

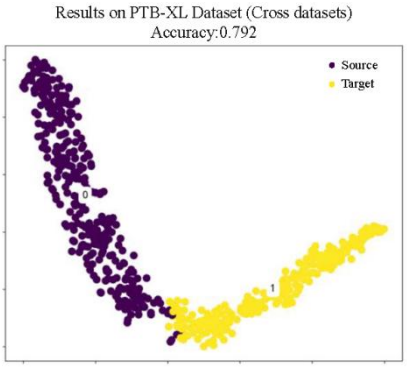


Fig. 11. Results on PTB-XL Dataset (Cross datasets case, where 0 stands for unnormal and 1 stands for normal).

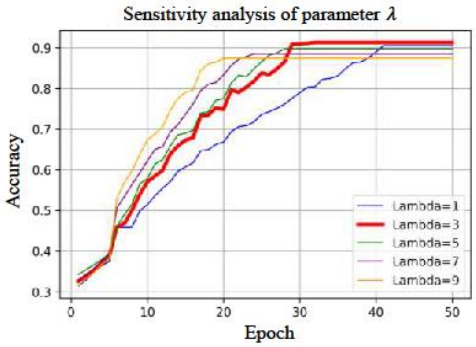


Fig. 12. Results of ablation experiments on CUSPH Dataset (Block 1: SeNet, Block 2: Domain Adaptation, Block 3: Ensemble Learning based on self attetntion).

6 Conclusion

An ECG diagnosis method incorporating a wide convolutional kernel backbone network, adversarial transfer learning method, and ensemble learning based on self-attention is proposed. Among them, the wide convolutional kernel network facilitates the extraction of features of ECG signals as well as the suppression of low-frequency noise, the strategy of adversarial transfer learning provides a solution strategy for the small amount of data and the lack of data labels existing in the ECG diagnosis of adolescents, and the method of ensemble learning makes it possible for us to take the features of each lead into account fully to improve the efficiency as well as the accuracy of diagnosis. In this paper, we conducted single dataset tests on CUSPH dataset and PTB-XL dataset, and made crossdataset attempts, and the experimental results show that the CDAD method has great potential in solving the problem of diagnosis of adolescent cardiac disorders with little data and few labels. We also analyzed and validated the validity of each module of the CDAD method and its sensitivity to parameters.

Table 5. Results of ablation experiments on CUSPH Dataset (Block 1: SeNet, Block 2: Domain Adaptation, Block 3: Ensemble Learning based on self attention)

Method	Block1	Block2	Block3	Acc	F1
(a)				0.782	0.778
(b)	√			0.832	0.811
(c)		√		0.796	0.783
(d)			√	0.841	0.839
(e)	√	√		0.876	0.863
(f)	√	√	√	**0.913**	**0.902**

References

1. Apandi, Z.F.M., Ikeura, R., Hayakawa, S.: Arrhythmia detection using mit-bih dataset: A review. In: 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA). pp. 1–5. IEEE (2018)
2. Blanco-Velasco, M., Weng, B., Barner, K.E.: Ecg signal denoising and baseline wander correction based on the empirical mode decomposition. Computers in Biology and Medicine 38, 1–13 (2008)
3. Cai, Z.P., Liu, C.Y., Gao, H.X., et al.: An open-access long-term wearable ecg database for premature ventricular contractions and supraventricular premature beat detection. Journal of Medical Imaging and Health Informatics 10, 2663–2667 (2020)
4. Chen, T.M., Huang, C.H., Shih, E.S.C., et al.: Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model. iScience 23, 100886 (2020)
5. Datta, S., Puri, C., Mukherjee, A., et al.: Identifying normal, af and other abnormal ecg rhythms using a cascaded binary classifier. In: 2017 Computing in Cardiology (CinC). pp. 1–4. IEEE (2017)
6. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International Conference on Machine Learning. pp. 1180–1189. PMLR (2015)
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial networks. Communications of the ACM 63, 139–144 (2020)
8. Hamilton, P.S.: A comparison of adaptive and nonadaptive filters for reduction of power line interference in the ecg. IEEE Transactions on Biomedical Engineering 43, 105–109 (1996)
9. He, Z., Chen, Y., Yuan, S., et al.: A novel unsupervised domain adaptation framework based on graph convolutional network and multi-level feature alignment for inter-subject ecg classification. Expert Systems with Applications 221, 119711 (2023)
10. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
11. Jordan, M.I., Mitchell, T.M.: Machine learning: Trends, perspectives, and prospects. Science 349, 255–260 (2015)
12. Kher, R.: Signal processing techniques for removing noise from ecg signals. J. Biomed. Eng. Res 3, 1–9 (2019)
13. Kiranyaz, S., Ince, T., Gabbouj, M.: Real-time patient-specific ecg classification by 1-d convolutional neural networks. IEEE Transactions on Biomedical Engineering 63, 664–675 (2016)
14. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521, 436–444 (2015)
15. Li, B., Chow, M.Y., Tipsuwan, Y., Hung, J.C.: Neural-network-based motor rolling bearing fault diagnosis. IEEE transactions on industrial electronics 47(5), 1060– 1069 (2000)
16. Li, D., Zhang, J., Zhang, Q., et al.: Classification of ecg signals based on 1d convolution neural network. In: 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom). pp. 1–6. IEEE (2017)
17. Li, Y., Pang, Y., Wang, J., et al.: Patient-specific ecg classification by deeper cnn from generic to dedicated. Neurocomputing 314, 336–346 (2018)
18. Li, Y., Wang, N., Shi, J., Liu, J., Hou, X.: Revisiting batch normalization for practical domain adaptation. arXiv preprint arXiv:1603.04779 (2016)
19. Lin, L., Liu, J., Zhou, M.: Top 10 causes of death and the most growing causes during the chinese spring festival holiday—china, 2017–2021. China CDC Weekly 5, 68–69 (2023)

20. Liu, X., Wang, H., Li, Z., et al.: Deep learning in ecg diagnosis: A review. Knowledge-Based Systems 227, 107187 (2021)
21. Murat, F., Yildirim, O., Talo, M., et al.: Application of deep learning techniques for heartbeats detection using ecg signals-analysis and review. Computers in Biology and Medicine 120, 103726 (2020)
22. Niu, L., Chen, C., Liu, H., Zhou, S., Shu, M.: A deep-learning approach to ecg classification based on adversarial domain adaptation. In: Healthcare. vol. 8, p. 437. MDPI (2020)
23. Obeidat, Y., Alqudah, A.M.: A hybrid lightweight 1d cnn-lstm architecture for automated ecg beat-wise classification. Traitement du Signal 38(5) (2021)
24. Organization, W.H.: World health statistics 2023: monitoring health for the SDGs, Sustainable Development Goals (2023)
25. Patel, V.M., Gopalan, R., Li, R., et al.: Visual domain adaptation: A survey of recent advances. IEEE Signal Processing Magazine 32, 53–69 (2015)
26. Saadatnejad, S., Oveisi, M., Hashemi, M.: Lstm-based ecg classification for continuous monitoring on personal wearable devices. IEEE Journal of Biomedical and Health Informatics 24, 515–523 (2020)
27. Sagi, O., Rokach, L.: Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8, e1249 (2018)
28. Tsao, C.W., Aday, A.W., Almarzooq, Z.I., et al.: Heart disease and stroke statistics—2023 update: a report from the american heart association. Circulation 147, e93–e621 (2023)
29. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł., Polosukhin, I.: Attention is all you need in advances in neural information processing systems, 2017. Search PubMed pp. 5998–6008
30. Wagner, P., Strodthoff, N., Bousseljot, R.D., et al.: Ptb-xl, a large publicly available electrocardiography dataset. Scientific Data 7, 154 (2020)
31. Wang, G., Chen, M., Ding, Z., Li, J., Yang, H., Zhang, P.: Inter-patient ecg arrhythmia heartbeat classification based on unsupervised domain adaptation. Neurocomputing 454, 339–349 (2021)
32. Wang, G., Chen, M., Ding, Z., et al.: Inter-patient ecg arrhythmia heartbeat classification based on unsupervised domain adaptation. Neurocomputing 454, 339–349 (2021)
33. Xia, Y., Xu, Y., Chen, P., et al.: Generative adversarial network with transformer generator for boosting ecg classification. Biomedical Signal Processing and Control 80, 104276 (2023)
34. Zhang, W., Peng, G., Li, C., Chen, Y., Zhang, Z.: A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors 17(2), 425 (2017)
35. Zhang, W., Peng, G., Li, C., Chen, Y., Zhang, Z.: A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors 17(2), 425 (2017)
36. Zheng, J., Zhang, J., Danioko, S., et al.: A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Sci Data 7, 48 (2020)
37. Zhuang, F., Qi, Z., Duan, K., et al.: A comprehensive survey on transfer learning. Proceedings of the IEEE 109, 43–76 (2020)