IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 November 2023, accepted 16 December 2023, date of publication 25 December 2023,
date of current version 4 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3346997

RESEARCH ARTICLE

ECG Biometrics Based on Attention Enhanced Domain Adaptive Feature Fusion Network

PAN YI[1], YUJUAN SI[2,1], WEI FAN[1], AND YANG ZHANG[1]
[1]College of Communication Engineering, Jilin University, Changchun 130012, China
[2]School of Electronic and Information Engineering, Zhuhai College of Science and Technology, Zhuhai 519041, China

Corresponding author: Yujuan Si (siyj@jlu.edu.cn)

This work was supported in part by the Natural Science Foundation of Guangdong Province under Grant 2023A1515011302, in part by the Guangdong Key Disciplines Project under Grant 2022ZDJS140, and in part by the Featured Innovation Projects of the Guangdong Universities under Grant 2022KTSCX189.

ABSTRACT In recent years, there has been an increasing focus on privacy and security among individuals. Biometric systems are highly regarded for their strong resistance to counterfeiting. Among various biometric features, ECG signals are difficult to falsify and are less attack-prone. However, as the time interval between signal acquisitions increases, the dissimilarities between individual ECG signals become more pronounced, making it difficult for many studies to achieve satisfactory recognition results in multi-session recognition. Furthermore, some studies encountered challenges in extracting crucial features from the ECG signal, which posed difficulties for identification experiments. To address the above challenge, this study proposes a novel attention-enhanced domain adaptive feature fusion network. Firstly, the network employs a multi-branch architecture to extract essential features from various dimensions of the ECG signal. Secondly, it incorporates the proposed weight fusion adaptive attention mechanism to further emphasize the features of heartbeats that contribute to recognition. Additionally, domain adaptive technology is employed to mitigate differences in feature distribution of ECG signals across sessions, thereby enhancing the model's generalization capability. Finally, four well-known databases, the ECG-ID database, PTB database, CYBHi database, and Heartprint database, were utilized to evaluate the performance of the model. These databases predominantly comprise individuals with multiple records, fulfilling the prerequisites for multi-session recognition. The model achieved recognition results of 96.31%, 73.79%, 77.80%, and 54.78% on these four databases, respectively, surpassing related studies in multi-session recognition scenarios.

INDEX TERMS Electrocardiogram, biometrics, multi-branch architecture, weight fusion adaptive attention mechanism, domain adaptation.

I. INTRODUCTION

In recent years, personal identification technology has gained significant prevalence in various fields such as finance, medical care, security monitoring, and data privacy, thanks to the rapid advancement of information technology. Traditional identification technology has several drawbacks and shortcomings that pose challenges in ensuring high security. For example, credentials and oral commands are vulnerable to duplication and theft, whereas complex passwords are prone to be forgotten. As a result, biometric identification systems have emerged. These systems employ biometric or behavioral characteristics to identify or verify individuals, offering significant convenience. The commonly used biometric features encompass fingerprints [1], face [2], iris [3], voice [4], and electroencephalogram (EEG) [5].

However, it is important to acknowledge that even commonly used biometric features have their limitations and deficiencies. Specifically, it is worth noting that fingerprints and vocal characteristics can be replicated, facial recognition is sensitive to changes in lighting and proximity, iris recognition necessitates a significant financial investment, and acquiring EEG signals is relatively more challenging compared to other technologies. Therefore, it is critical to

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.

explore a biometric feature that is both more reliable and easier to obtain. In recent years, electrocardiogram (ECG) signals have been widely utilized in the field of identity recognition. It offers several distinct advantages compared to other biometric features. Firstly, the ECG signal is universal as it is present in every individual. Furthermore, it is unique to each person due to variations in heart size, the orientation of the heart muscle, electric conductivity, and the order of activation of cardiac muscle [6]. Secondly, it is generally observed that the ECG signal of an individual remains relatively stable within a certain time range unless the individual has heart disease. Moreover, acquiring ECG signals is convenient, with various methods available to choose from. In hospital settings, specialized equipment is used for ECG signal acquisition, while individuals can utilize wearable devices for the same purpose. Lastly, the device used to collect the ECG signal must be securely attached to the body, significantly reducing the risks of theft and duplication. In comparison to other biometric features, ECG signals offer a higher level of security. The existence of these advantages has led to the enhanced utilization of ECG signals for both identification [42], [43] and verification [44], [45], having a promising research prospect.

In recent years, remarkable progress in deep learning has resulted in the development of several outstanding neural network architectures. Prominent examples include convolutional neural networks (CNNs) [7], [8], long short-term memory networks (LSTMs) [9], [10], as well as advanced variations of CNNs such as residual networks (ResNets) [11], [12] and dense convolutional networks (DenseNets) [11], [13]. These networks have exhibited exceptional recognition performance in identity recognition tasks. In the field of ECG identification, the process can be categorized into multi-session identification and single-session identification based on the presence or absence of a time interval between the acquisition of ECG signals in the training and testing sets. Currently, the majority of research in ECG identity recognition has primarily concentrated on single-session recognition, with less emphasis on multi-session recognition. Despite some studies attempting multi-session recognition experiments, the ultimate recognition outcomes remain unsatisfactory, indicating the substantial potential for improvement. Meanwhile, research [8] has indicated that an individual's ECG signal exhibits variations due to various physiological and psychological factors, which become more pronounced over time. Consequently, conducting experiments and achieving better recognition results in multi-session identification can highlight the effectiveness and practicality of identity recognition systems. Existing studies in ECG identity recognition encounter several challenges, including: (1) Some studies struggle to effectively extract the crucial features present in the ECG signal, hampering the model's ability to discriminate between different individuals. (2) The problem of differences in feature distribution among ECG signals from various sessions has hindered many studies from achieving satisfactory results in multi-session recognition. To address the aforementioned limitations, this study makes the following contributions:

(1) To address the challenge of low accuracy in multi-session recognition, this study proposes a novel attention-enhanced domain adaptive feature fusion network (ADAFFN).

(2) The feature extraction module of the network incorporates a multi-branch architecture, enabling comprehensive extraction of key features from ECG signals across multiple dimensions.

(3) The weight fusion adaptive attention (WFAA) mechanism is proposed to further emphasize the features of heartbeats that contribute to recognition.

(4) Domain adaptive techniques are utilized to reduce differences in the distribution of ECG signal features between different sessions.

(5)To avoid the limitations of the ordinary Softmax Loss, a Normalized version of the Softmax Loss (NSL) is used to train the model.

The structure of this study is organized as follows: Section II provides an introduction to the related research on ECG recognition. Section III describes the methodology employed in this study. Section IV presents the specific experimental results of ECG recognition. Finally, Section V concludes the study and outlines future research directions.

II. RELATED WORK

Biel et al. [14] proposed the first application of ECG signals for identification purposes by extracting features from 12-lead ECG signals based on fiducial points. Since then, ECG identification has gained significant attention and has been widely studied. Boujnouni et al. [15] employed a capsule neural network (CapsNet) for individual identification, achieving impressive results of 98.8% accuracy on the PTB database. Fatimah et al. [16] employed Fourier decomposition (FDM) and phase transform (PT) techniques to process the ECG signal and extract relevant features. Subsequently, a random forest (RF) classifier was utilized, achieving notable results with 97.92% accuracy in identifying individuals on the ECG-ID database. Moreover, Ciocoiu and Cleju [17] employed a modified version of the Continuous Wavelet Transform known as the S-Transform, along with the Gramian Angular Field, recurrence plot, and state-space representations, to convert one-dimensional signals into two-dimensional or three-dimensional images. Their approach achieved an impressive identification accuracy of 98.6% on the CTBHi short-term database. All of the mentioned studies are focused on single-session recognition and have achieved excellent recognition results, leaving limited room for further improvement in that particular area.

The outcomes of multi-session recognition demonstrate a notable decrease in comparison to the results of single-session recognition. For instance, Ibtehaz et al. [18] proposed a network composed of MultiRes blocks and spatial pyramid pooling (SPP), which achieved a final identification accuracy

of 96.25% when single-session identification was performed on the ECG-ID database. However, this accuracy was reduced to 92.17% when multi-session identification was conducted. Similarly, Belo et al. [19] employed a temporal convolutional neural network (TCNN) to perform single-session identification in the CYBHi long-term database and obtained a final identification accuracy of 100%. However, when performing multi-session identification, the accuracy was significantly decreased to 60.65%. Ciocoiu [20] used discrete cosine transform (DCT), discrete wavelet transform (DWT), and random projections (RP) to extract heartbeat features, which resulted in a final identification accuracy of 94% when single-session identification was performed on the CYBHi short-term database. However, this accuracy dropped to 60% when conducting multi-session identification on the long-term database. The studies mentioned earlier faced challenges in extracting critical features from ECG signals and were impacted by variations in feature distribution among ECG signals from different sessions. Consequently, they achieved lower experimental results in multi-session recognition experiments. Given the significance of multi-session recognition in highlighting the effectiveness of ECG identity recognition, it is crucial to explore more effective models or algorithms to improve its performance.

Incorporating attention mechanisms into the model is an effective measure for improving its overall performance. The advantage of this approach lies in its ability to accentuate features that contribute to the ultimate recognition process, consequently yielding elevated recognition outcomes. Hammad et al. [12] introduced a residual network with an attention mechanism, resulting in identification accuracies of 98.85% and 99.27% on the PTB and CYBHi databases, respectively. Sun et al. [21] employed a CNN combined with a channel attention module (CAM) to identify ten individuals from the PTB database, achieving an identification accuracy of 94%. Additionally, Feng et al. [22] proposed a multi-perspective adaptive feature fusion (MPAFF) module to enhance the accuracy of disease classification between patients, yielding an accuracy of 95.7% on the MIT-BIH Arrhythmia Database (ARDB). The aforementioned study augmented the original model with attentional mechanisms, resulting in further improvement in the final experimental results.

Both multi-session recognition experiments and inter-patient disease classification encounter similar challenges. From what we know, the introduction of domain adaptive techniques into the model can lead to improved classification results specifically for inter-patient disease classification. For example, Wang et al. [23] proposed a novel domain-adaptative ECG arrhythmia classification (DAEAC) model based on a convolutional neural network and unsupervised domain adaptation (UDA). Experiments conducted on the MIT-BIH database for inter-patient classification achieved a classification accuracy of 97.59%. Similarly, Feng et al. [22] proposed an unsupervised semantic-aware adaptive feature fusion network (USAFFN). By mitigating the semantic distribution differences between the feature spaces of the two domains, they obtained a final classification accuracy of 95.7% in the inter-patient experiment of ARDB. Compared to prior research, the final classification accuracy of the aforementioned two has been significantly improved. Drawing inspiration from the fruitful utilization of domain adaptation technology in disease classification, we are inclined to implement this approach in ECG identity recognition. The primary objective is to diminish the disparities in feature distribution among ECG signals collected during various sessions.

Ultimately, we summarize the research content of the related work, and the specific results are presented in TABLE 1. Recognizing the existing room for improvement in the recognition outcomes of relevant studies on multi-session recognition, this study introduces a novel approach. An attention-enhanced domain adaptive network is proposed for recognition experiments. The network embraces a multi-branch architecture, facilitating the comprehensive extraction of ECG signal features across diverse dimensions. The introduction of the WFAA mechanism further enhances the network's capability to accentuate features conducive to final recognition. Additionally, domain adaptive techniques are employed to mitigate the disparities in feature distribution among ECG signals from various sessions, thus enhancing the model's generalization capacity. Finally, considering the limitations of ordinary Softmax Loss, this study employs the Normalized version of the Softmax Loss to train the model. The subsequent chapters of this study will provide a comprehensive description and analysis of the proposed network.

III. MATERIALS AND METHODS

The main focus of this study revolves around the development of an ECG identification system, which aims to accurately identify an individual's true identity using their ECG signal. The ECG identification process is composed of three primary steps: signal pre-processing, feature extraction, and identification.

A. SIGNAL PRE-PROCESSING

To facilitate the subsequent recognition tasks, certain pre-processing steps are required for the collected original ECG signal. The pre-processing process comprises four essential steps: denoising, R peak detection, heartbeat segmentation, and normalization.

1) DENOISING

Noise in raw ECG signals typically encompasses baseline wander, power-line interference, and muscle artifact. Baseline wander is commonly caused by human respiration, with a frequency typically below 0.5 Hz [24]. Power-line interference primarily arises from the power supply, with frequencies typically at 50 Hz or 60 Hz [25]. In addition, the muscle artifact is caused by the contraction of muscles other than the heart and is a random noise spread over the entire frequency range [26]. Taking into account the various

TABLE 1. The detailed literature for biometric authentication using ECG signals.

Reference	Dataset	Subjects	Technique	Method	Accuracy(%)	Remarks
[12]	PTB CYBHi	290 65	ResNet-Attention	single-session	98.85 99.27	Multi-session recognition experiments were not performed
[14]	Private	20	SIMCA	single-session	85.40	The number of subjects tested is less and accuracy is also low
[15]	PTB MIT-BIH STDB MIT-BIH NRSDB MIT-BIH Arrhythmia	290 28 18 47	CapsNet	single-session	98.80 98.20 100.0 98.20	Multi-session recognition experiments were not performed
[16]	CYBHi ECG-ID MIT-BIH Arrhythmia	65 89 47	Fourier Decomposition	single-session	99.45 97.92 97.92	The number of subjects tested is less and multi-session recognition experiments were not performed
[17]	UofTDB CYBHi	52 65	2D/3D spatial represent and 2D CNN	single-session	98.60 98.60	The number of subjects tested is less and multi-session recognition experiments were not performed
[42]	PTB CYBHi ECG-ID UofTDB	290 63 90 1020	CNN and LSTM	single-session	99.62 94.86 99.49 95.36	The generalization ability of the network is strong, but the accuracy needs to be improved
[43]	ECG-ID USSTDB MIT-BIH Arrhythmia	90 60 47	Segmentation Maximum Pooling	single-session	97.14 95.54 97.66	Multi-session recognition experiments were not performed
[44]	ECG-ID	90	CNN	single-session	91.00	The accuracy is low
[45]	PTB ECG-ID	290 90	Ensemble Siamese network	single-session	93.60 96.80	Multi-session recognition experiments were not performed and the accuracy needs to be improved
[18]	ECG-ID CYBHi	90 63	EDITH	multi-session	92.17 72.76	The generalization ability of the network is strong, but the accuracy needs to be improved
[19]	CYBHi	63	TCNN	multi-session	60.65	The number of subjects tested is less and accuracy is also low
[20]	CYBHi	63	DCT、DWT and RP	multi-session	60.00	The number of subjects tested is less and accuracy is also low
[21]	PTB ECG-ID	50+	CNN and CAM	multi-session	56.93 85.94	No specific number of people identification and accuracy is low

frequencies associated with different types of noise, this study utilizes a finite impulse response (FIR) filter for denoising the ECG signal. The lower cut-off frequency and the upper cut-off frequency of this filter are set at 3 Hz and 45 Hz, respectively. This selection helps in effectively eliminating baseline wander and power-line interference, while also enhancing the suppression of muscle artifacts within the ECG signal. The denoising outcome is visualized in Fig. 1. Upon comparing the images before and after denoising, it becomes evident that a substantial portion of the noise in the ECG signal has been effectively eliminated.

2) HEARTBEAT SEGMENTATION

Heartbeat segmenting techniques are commonly categorized into two main categories: R-peak point-based segmentation and blind segmentation. In the case of R-peak point-based segmentation, the initial step involves locating the R-peak point, followed by utilizing it as the reference point and selecting several pre-defined points before and after it for heartbeat segmentation. In contrast, blind segmentation solely requires an appropriate initial point and a fixed length to segment long ECG signals. Although blind segmentation eliminates the need to locate peaks and avoids errors caused by incorrect positioning, it introduces discrepancies in the resulting segmented heartbeats. Some heartbeats are complete, while others are incomplete, which hinders the final identification process. Consequently, this study adopts the segmentation method based on R-peak points. The specific implementation process is as follows: Firstly, the R-peak is located using the Hamilton detection algorithm [27]. Fig. 2 showcases the results of R-peak detection in the ECG signal. Secondly, a complete single heartbeat signal typically has a duration of 0.6 seconds, given a sampling frequency of 500 Hz. Considering the different segments in the heartbeat signal, 119 points before the R-peak point and 180 points after the R-peak point were ultimately selected for

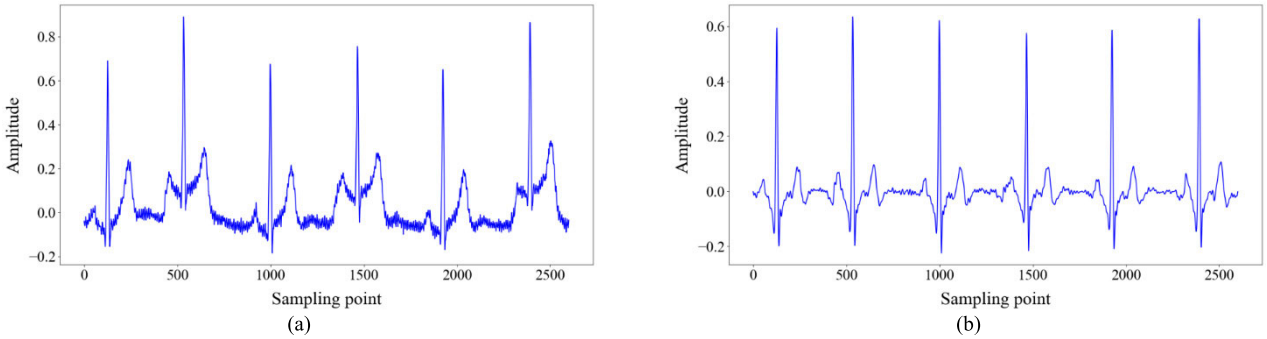

FIGURE 1. The results of denoising the ECG signal. (a) Raw ECG signal; (b) Denoised ECG signal.

heartbeat segmentation. Finally, to account for databases with different sampling frequencies, it is essential to resample the data to a standardized sampling frequency of 500 Hz before conducting R-peak detection and heartbeat segmentation. This ensures consistency and facilitates accurate analysis across different datasets.

3) MIN-MAX NORMALIZATION

The normalization operation is applied to standardize the scale of different heartbeats, which aids in the computation and convergence of the network training process. In this study, the segmented heartbeats are processed using Min-Max normalization, which is expressed by:

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

where x is the input heartbeat vector, $\min(x)$ is its minimum value, $\max(x)$ is its maximum value, and x_{norm} is the normalized heartbeat vector.

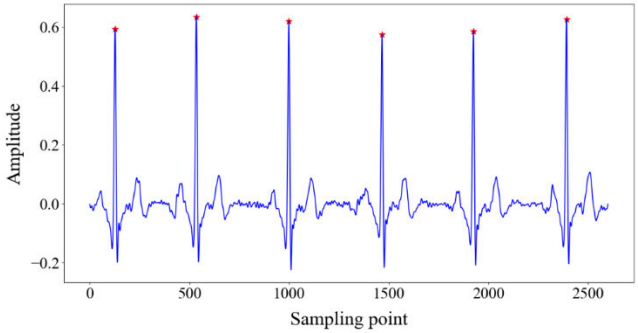

FIGURE 2. The results of R-peak detection in the ECG signal.

B. FEATURE EXTRACTION

1) DOMAIN ADAPTATION

Domain adaptation [28] represents a distinctive subset of transfer learning. The core concept entails the mapping of data features originating from diverse domains onto a unified feature space. By reducing the disparities in feature distributions between domains, models with better recognition performance in certain domains can be applied to experiments in other domains. Unsupervised domain adaptation has gained significant popularity due to the challenges associated with obtaining labeled data. Specifically, the unsupervised domain adaptation involves two domains, namely the source domain D_S and the target domain D_T. Source domain samples are usually labeled and defined as $D_S = \{X_S, Y_S\} = \{(x_S^i, y_S^i)\}_{i=1}^{N_S}$. X_S represents the data of the samples in the source domain, and Y_S denotes the corresponding identity labels. x_S^i and y_S^i represent the data and identity label, respectively, of the i-th individual in the source domain. The variable N_S denotes the total number of individuals in the source domain. Target domain samples are usually unlabeled and defined as $D_T = \{X_T\} = \{(x_T^i)\}_{i=1}^{N_T}$. X_T represents the data of the samples in the target domain. x_T^i represents the data of the i-th individual in the target domain, and N_T denotes the total number of individuals in the target domain. Although the source domain and the target domain are not identical, there exists a certain correlation between them. The goal of domain adaptation is to align the data distributions of these two domains, which have distinct distributions, into a common feature space. By defining an appropriate metric criterion, the aim is to minimize the discrepancy between them. This optimization process enables the model trained on the source domain $\{X_S, Y_S\}$ to accurately predict the labels of the target domain $\{X_T\}$. Fig. 3 presents the comparison of the results before and after domain adaptation. In the absence of domain adaptation, there is a phenomenon called domain drift, where there are notable differences between the feature distributions of the source and target domains. Consequently, using the decision boundary derived from the source domain directly to classify samples from the target domain introduces bias. However, after applying domain adaptation techniques, the feature distributions of the two domains become more aligned, eliminating the domain drift. As a result, the decision boundary learned from the source domain becomes capable of accurately distinguishing the samples from the target domain, resulting in improved classification performance.

Person re-identification (re-ID) has become increasingly popular in the community. When given video sequences of a person, Person Re-ID recognizes if that person has been in another camera to compensate for device limitations. [29] conducted experiments using three distinct

character re-identification datasets as diverse domains. The implementation of adversarial domain adaptation led to notable enhancements in the recognition outcomes within the target domain. The main focus of this study is multi-session identification, where the data from two sessions are considered as the source and target domains, respectively. To address the issue of differences in feature distribution between sessions, this study will employ the adversarial domain adaptive approach proposed in the aforementioned research for subsequent recognition experiments. Compared to conventional unsupervised domain adaptation methods, this technique incorporates the concept of adversarial training. The proposed domain-adversarial neural network (DANN) comprises a feature extractor, a domain classifier, and a label predictor. The feature extractor and domain classifier are trained in an adversarial manner to enhance recognition outcomes in the target domain.

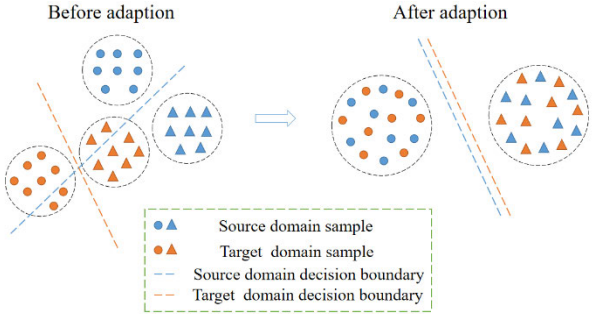

FIGURE 3. The comparison of the results before and after domain adaptation.

2) NETWORK STRUCTURE

The proposed attention-enhanced domain adaptive feature fusion network, as depicted in Fig. 4, comprises three key components: a feature extractor, a label predictor, and a domain classifier. The feature extractor in this study employed a multi-branch architecture for feature extraction. Following the findings of Li et al. [30], who observed that 3×1 kernels with varying dilation sizes tend to outperform larger kernels with the same receptive field but no dilation, the convolutional layers in the feature extractor were designed to utilize 3×1 convolutional kernels with different dilation sizes. In addition, the findings of [31] demonstrate that the degradation of the model becomes more pronounced with an increase in the number of model layers. However, by incorporating a residual structure into the model, not only does the number of model layers increase, allowing for the extraction of deeper features from the input signal, but it also prevents network degradation and preserves the original features. Hence, the residual block is introduced to the feature extractor, and its specific network structure is illustrated in Fig. 5. Simultaneously, this study incorporates a designed attention mechanism into the feature extractor to enhance recognition accuracy and further improve the final recognition results.

The MPAFF mechanism introduced by [22] showed improvements in the classification results of inter-patient disease classification. However, the sub-attention module employed the SENet mechanism, and the findings of a related study [32] demonstrated that the dimensionality reduction in SENet is unnecessary and even detrimental. This study proposes a novel weight fusion adaptive attention mechanism that combines the strengths of the attention mechanisms in [22] and [32]. The WFAA mechanism not only avoids dimensionality reduction but also reduces the number of training parameters, enhancing the overall training efficiency of the model. Moreover, the weighting parameters of the WFAA mechanism are automatically adjusted during the training process, eliminating the need for manual intervention. Fig. 6 illustrates the specific structure of this attention mechanism.

This is followed by a description of the specific workflow of the WFAA mechanism. First, the input is subjected to global average pooling (GAP) to realize the squeezing operation, which is represented as F_{sq}. The channel weights of the input features are obtained by applying N sub-attention mechanisms after the squeezing operation. Each sub-attention mechanism comprises a convolutional layer followed by a Sigmoid function, denoted as F_{eca}. The size of the convolutional kernel in the convolutional layer is determined based on the approach described in [32], as expressed by:

$$k = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{2}$$

where k represents the kernel size of convolution, $|t|_{odd}$ denotes the nearest odd number to t, γ and b are constants assigned to 2 and 1, respectively, and C represents the channel number size.

The i-th sub-attention module calculates the channel weights of the input features as follows:

$$U^i = F_{eca}(F_{sq}(U)) \in \mathbb{R}^{C \times 1} \tag{3}$$

where $U \in \mathbb{R}^{C \times L}$ is the input feature.

After obtaining all the U^i, the final output of the first branch is obtained as follows:

$$U^{all} = [U^1, U^2, \ldots, U^N] \in \mathbb{R}^{C \times N} \tag{4}$$

In the second branch, the input features undergo a squeezing operation through global average pooling to reduce their dimensionality. Following this, a fully connected (FC) layer and a Softmax function are utilized to calculate the weights for the N sub-attention modules in the first branch, as expressed by:

$$A_{weight} = Softmax(W_1(F_{sq}(U))) \in \mathbb{R}^{N \times 1} \tag{5}$$

where $W_1 \in \mathbb{R}^{N \times C}$ is the weight of the fully connected layer.

The results obtained from the two branches are multiplied together to acquire the final weighted results on the input features, as expressed by:

$$W_U = U^{all} \cdot A_{weight} \in \mathbb{R}^{C \times 1} \tag{6}$$

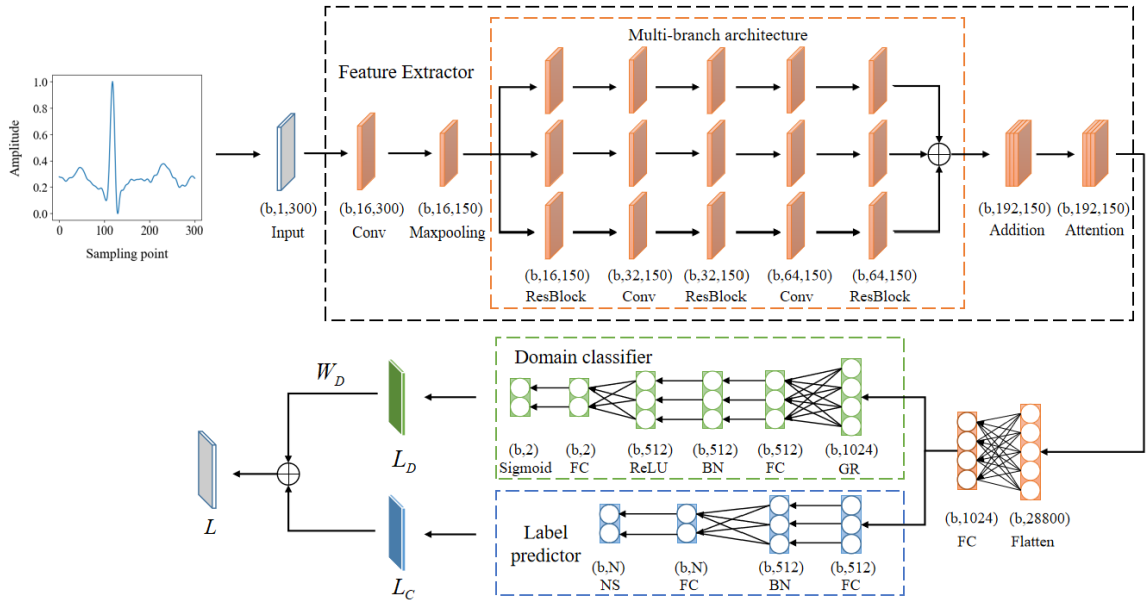

FIGURE 4. The specific structure of ADAFFN. b is the batch size, Conv is the convolutional layer, BN is the batch normalization layer, GR is the gradient reversal layer, NS is the normalized Softmax function, Attention is the WFAA, N is the total number of individuals, and ResBlock is the residual block.

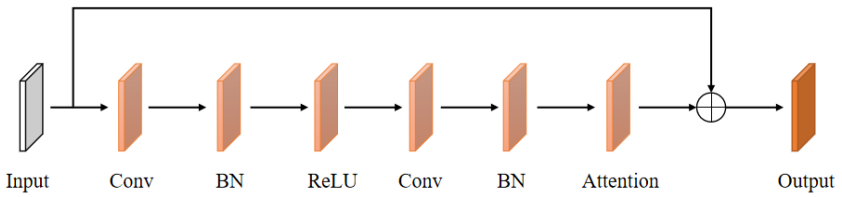

FIGURE 5. The specific structure of the ResBlock.

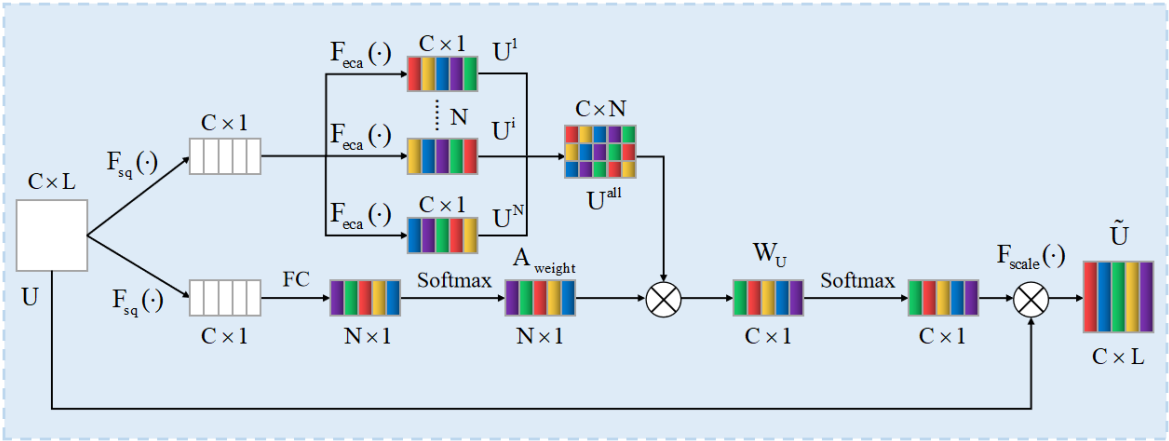

FIGURE 6. The specific structure of the attention mechanism.

Finally, in the third branch, the input features are multiplied by the final weighting result, which is then applied to the different channels of the input features to obtain the final result:

$$\tilde{U} = F_{scale}(U, Softmax(W_U)) \in \mathbb{R}^{C \times L} \tag{7}$$

where F_{scale} is the two feature vectors multiplied according to the channels.

Once the input ECG signal is processed by the feature extractor, the extracted features are subsequently fed into both the domain classifier and the label predictor. The primary

function of the domain classifier is to discriminate the source of the data. It incorporates a gradient reversal layer, which plays a crucial role in facilitating adversarial training between the domain classifier and the feature extractor. The loss function utilized in this part is the domain discrimination loss, denoted as L_D. The specific form of the domain discrimination loss is as follows:

$$L_D = -\frac{1}{M}\sum_{i=1}^{M}[d_i \log(p_i) + (1-d_i)\log(1-p_i)] \quad (8)$$

where d_i is the domain label and the source and target domain labels are 1 and 0, respectively, p_i denotes the predicted probability of the source domain for the i-th sample and M is the total number of adaptive samples.

The domain classifier is trained normally to enhance its domain discrimination performance by minimizing the domain discrimination loss. The inclusion of a gradient reversal layer prompts the feature extractor to shift in a direction that leads to an increase in the domain discrimination loss, continually extracting features that possess reduced distinctiveness between the two domains. With the progression of the training process, when the domain classifier becomes incapable of discerning the features extracted by the feature extractor, it signifies the completion of the domain adaptation process. At this stage, the discrepancies in the feature distribution among ECG signals across sessions have been successfully reduced.

The label predictor is responsible for predicting the true identity of the sample. According to Wang et al. [33], the boundary of Softmax Loss depends on both the magnitudes of weight vectors and the cosine of angles, leading to overlapping decision regions in the cosine space. The Normalized version of Softmax Loss operates by setting the weight vector to 1 through L2 regularization. Subsequently, the feature vector is normalized to s to remove variations in radial directions. This allows the resulting model to learn features that are distinguishable in the angular space. Ultimately, NSL achieves perfect classification of testing samples in the cosine space. Thus, to achieve improved recognition outcomes, this study employs the NSL function for training the model. The Normalized version of Softmax Loss is dependent on a single variable, the angle θ. The magnitude of this angle influences the computed loss value. Similarly, it can be deduced that the cosine value of the angle directly dictates the resulting loss value. So this normalization procedure strengthens the constraint imposed by the loss function on the cosine value, emphasizing the reliance of the optimization process on the cosine value to extract discriminative features. The NSL can be expressed as follows:

$$L_C = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos(\theta_{yi},i)}}{\sum_{j=1}^{n}e^{s\cos(\theta_j,i)}} \quad (9)$$

subject to

$$W = \frac{W^*}{||W^*||}, x = \frac{x^*}{||x^*||}, \cos(\theta_j, i) = W_j^T x_i \quad (10)$$

where N is the total number of training samples, n is the number of categories, s is a constant, W_j is the weight vector of the j-th class, x_i is the i-th feature vector corresponding to the ground-truth class of y_i, θ_j is the angle between W_j and x_i.

The total loss function L is obtained by adding the classification loss L_C and the domain discriminant loss L_D, as expressed by:

$$L = L_C + W_D \cdot L_D \quad (11)$$

where the parameter W_D is mainly used to adjust the magnitude of the domain discrimination loss to avoid its value being too large and affecting the classification of the label predictor.

C. IDENTIFICATION

In addition to single-session recognition and multi-session recognition, ECG identity recognition can be further classified into single-heartbeat recognition and multi-heartbeat recognition based on the number of input beats. Specifically, multi-heartbeat identification can be categorized into two distinct types: (1) In heartbeat segmentation, several successive single-heartbeat signals are segmented to obtain a long signal, which is then fed into the model for training. This can increase the similarity of the heartbeats from the same body, contributing to greater accuracy in the final classification. However, more complex network training is required to achieve better recognition results. (2) Multiple single-heartbeat signals from the same individuals are simultaneously fed into the model for training, and the corresponding prediction labels are obtained. Finally, the final predicted label of the individual is determined by voting. For example, suppose three single-heartbeat signals from the same individual are fed into the model for prediction. The first and second predictions are labeled "A", while the third prediction is labeled "B". Then the final label for that individual will be "A". In contrast to the method (1), which directly obtains recognition outcomes, method (2) necessitates the collection of recognition results from several heartbeats to finalize predictions through a voting process. However, method (2) is exclusively employed during the testing phase, exerting no influence on the training phase, and applies to a wide range of models. To enhance the recognition outcomes, this study will conduct recognition experiments utilizing multiple heartbeats. Since enough data is required for training and the model used in this study is relatively simple, method (2) is chosen for multi-heartbeat identification. The selection of the number of voting heartbeats adhered to [46], where three single-heartbeat signals were utilized for the recognition experiments.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. ECG DATASETS

This study focuses on investigating identity identification systems based on ECG signals. The majority of the public ECG signal databases used in this research are sourced from the PhysioBank physiological signal database [34]. Given the primary emphasis of this study on multi-session recognition, the recognition experiment necessitates the availability of two records. Evaluating various databases in consideration of their data content, the following four databases fulfill this prerequisite: the ECG-ID database [35], the PTB Diagnostic ECG database [36], the Check Your Biosignals Here initiative (CYBHi) database [37], and the Heartprint database [41]. Table 2 summarises the basic information about each database. The databases utilized are subsequently introduced in detail.

TABLE 2. The basic information about each database.

Dataset	Persons	Sampling Rate	Health Condition	Activity	Electrode Placement
ECG-ID	90	500 Hz	Healthy	Sitting	Wrist
PTB	290	1000 Hz	Mixed	Resting	Chest+Limbs
CYBHi	63	1000 Hz	Healthy	Sitting	Finger
Heartprint	199	250 Hz	Healthy	Resting and Reading	Finger

ECG-ID database: A total of 310 ECG signals were collected from 90 healthy individuals in the ECG-ID database, comprising 44 males and 46 females aged between 13 and 75 years. The number of records for each individual varies, ranging from 2 (collected during one day) to 20 (collected periodically over 6 months). Each signal had a duration of 20 seconds and was sampled at 500 Hz. The ECG signals consisted of two types of data: raw data and filtered data. The raw ECG signal exhibited significant noise in both high and low frequencies. Therefore, in this study, the subsequent experiments were conducted using the filtered data. Although the filtered data had undergone denoising, some residual noise remained, necessitating additional denoising during the pre-processing stage.

PTB Diagnostic ECG database: The database comprised 549 ECG recordings obtained from 290 subjects, including 209 males and 81 females, with ages ranging from 17 to 87 years. The dataset encompassed both healthy individuals and those with various diseases, with 52 subjects classified as healthy and 148 diagnosed with myocardial infarction, which represented the largest category of diseases. Each individual in the database has between one and five records, with a total of 113 individuals having more than two records. Each recorded data consisted of 15 ECG signals, including the traditional 12-lead and 3 Frank leads, with a sampling frequency of 1000 Hz. For the majority of individuals, the maximum interval between two records was less than one month. However, there were also individuals with intervals exceeding one year. In this study, only the I-lead from each individual will be utilized for the identification task.

CYBHi database: The database is divided into two datasets: the short-term dataset and the long-term dataset. The short-term dataset involves the collection of ECG recordings from 65 participants using Ag/AgCl sensors at a resolution of 12 bits. This dataset consists of multiple recordings from each individual under three different situations: neutral discussion, exposure to a low-arousal video, and exposure to a high-arousal video. The long-term dataset, on the other hand, was gathered from the same group of subjects over 3 months, totaling 63 subjects. Before data collection, all volunteers were informed about the purpose of the study and the experimental materials involved. Additionally, they were asked to provide informed consent by signing a consent form.

Heartprint database: The database serves as a comprehensive biometric repository housing multi-session ECG signals. It encompasses a collection of 1539 records, meticulously captured from the fingers of 199 healthy subjects. Each recording is acquired within 15 seconds, capturing both resting and reading conditions. This database has been amassed over ten years, spanning multiple sessions. Notably, the average interval between the first session (S1) and the third session (S3L) spans 1572.2 days. Furthermore, the average interval between the first session (S1) and the second session (S2) extends to 47.5 days. The dataset is also distinguished by its inclusion of various demographic categories, such as genders, ethnicities, and age groups.

B. EVALUATION INDICATORS

To quantitatively analyze the experimental results, the identification performance is evaluated using four standard indicators: Accuracy (Acc), Recall (Re), Precision (Pr), and F1-Score (F1). These metrics are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{15}$$

where TP (True Positive) is the number of correctly identified positive samples, TN (True Negative) is the number of correctly identified negative samples, FP (False Positive) is the number of incorrectly identified positive samples, and FN (False Negative) is the number of incorrectly identified negative samples.

C. RELATED PARAMETER SETTINGS

In this study, the optimal parameters for subsequent experiments were determined through a series of optimization-seeking experiments. Specifically, the dilation rate of the first

convolutional layer was set to 5, while the dilation rates of each branch in the multi-branch architecture were set to 3, 5, and 9, respectively. The integer s of the classification loss L_C is set to 5, the number of sub-attention modules of WFAA is set to 3, and the W_D of the discrimination loss of the control domain is set to 0.1. The domain adaptation process requires the use of part of the target domain data, and the top 20% of the data for each individual is selected for the domain adaptation experiments by considering relevant factors. Moreover, the training process is optimized using the Adam optimization algorithm. The batch size is set to 64, and the training epoch is set to 30. The learning rate is adjusted based on the specific database used during training. This study uses Pytorch for the deep learning framework and Python for programming. The evaluation of the model was conducted on a server equipped with an Intel Silver 4210 2.2G CPU, GeForce RTX1080Ti GPU, and 16GB RAM.

D. EXPERIMENTAL RESULTS

To assess the model's recognition performance on ECG signals with longer intervals, this study conducted a multi-session recognition experiment using the ADAFFN. In the ECG-ID database, all individuals, except for individual No.74, possessed two or more records. For individual No.74, we partitioned their data into two segments, allocating one to the training set and the other to the testing set. Subsequently, two records were randomly selected for training and testing from the remaining individuals. Regarding the CYBHi database, the identification experiments utilized data from the long-term dataset. Each individual in this dataset contributed two records, each assigned to the training and testing sets, respectively. In the PTB database, only 113 individuals had multiple records. To compare with the experimental results of [21], which employed data with an average interval of 63 days between acquisitions for identification, we selected the two records with the longest interval among these 113 individuals. Ultimately, to comprehensively evaluate the model's recognition performance across a broader spectrum of individuals, this study employed data from 199 individuals extracted from the Heartprint database for recognition experiments. Successively, the two records, S1 and S2, for each individual were employed as the training set and the testing set.

After preprocessing the data from the mentioned database, this study conducted a 2-fold cross-validation experiment under two recognition modes: single-heartbeat recognition and multi-heartbeat recognition. Table 3 presents the recognition results of the model in this study on the ECG-ID database, achieving average recognition accuracies of 94.26% and 96.31% for the two recognition modes, respectively. For the PTB database, the two records with the longest interval were selected, resulting in significant differences between the records. However, from Table 4, it can be seen that the model in this study still achieved average recognition accuracies of 73.53% and 73.79% for both recognition modes. Despite the presence of noise in the CYBHi database, which presents challenges for accurate recognition, the proposed model in this study achieved satisfactory results. Table 5 demonstrates that the model obtained recognition accuracies of 75.31% and 77.80% in both recognition modes. Finally, to assess the recognition performance of the model across a larger spectrum of individuals, the Heartprint database was employed for experimentation in this study. Table 6 demonstrates that

TABLE 3. The multi-session recognition results of the research model on the ECG-ID database.

Identification method	2-Fold	Re(%)	Pr(%)	F1(%)	Acc(%)
Single-heartbeat identification	Fold 1	94.50	94.94	94.72	94.18
	Fold 2	94.37	94.70	94.54	94.34
	Average	94.44	94.82	94.63	94.26
Multi-heartbeat identification	Fold 1	96.84	97.17	97.00	95.68
	Fold 2	97.72	97.98	97.85	96.93
	Average	97.28	97.58	97.43	96.31

TABLE 4. The multi-session recognition results of the research model on the PTB database.

Identification method	2-Fold	Re(%)	Pr(%)	F1(%)	Acc(%)
Single-heartbeat Identification	Fold 1	75.49	72.11	73.76	74.51
	Fold 2	74.84	69.69	72.18	72.55
	Average	75.17	70.90	72.97	73.53
Multi-heartbeat Identification	Fold 1	75.95	72.97	74.43	74.57
	Fold 2	75.60	71.21	73.34	73.00
	Average	75.78	72.09	73.89	73.79

TABLE 5. The multi-session recognition results of the research model on the CYBHI database.

Identification method	2-Fold	Re(%)	Pr(%)	F1(%)	Acc(%)
Single-heartbeat Identification	Fold 1	75.22	73.28	74.24	74.33
	Fold 2	76.61	75.29	75.95	76.28
	Average	75.92	74.29	75.10	75.31
Multi-heartbeat Identification	Fold 1	81.28	79.24	80.25	76.98
	Fold 2	80.85	80.26	80.55	78.61
	Average	81.07	79.75	80.40	77.80

TABLE 6. The multi-session recognition results of the research model on the HEARTPRINT database.

Identification method	2-Fold	Re(%)	Pr(%)	F1(%)	Acc(%)
Single-heartbeat Identification	Fold 1	52.90	44.92	48.58	52.16
	Fold 2	55.64	48.13	51.61	56.50
	Average	54.27	46.53	50.10	54.33
Multi-heartbeat Identification	Fold 1	56.13	47.61	51.52	52.54
	Fold 2	58.80	50.80	54.51	57.02
	Average	57.47	49.21	53.02	54.78

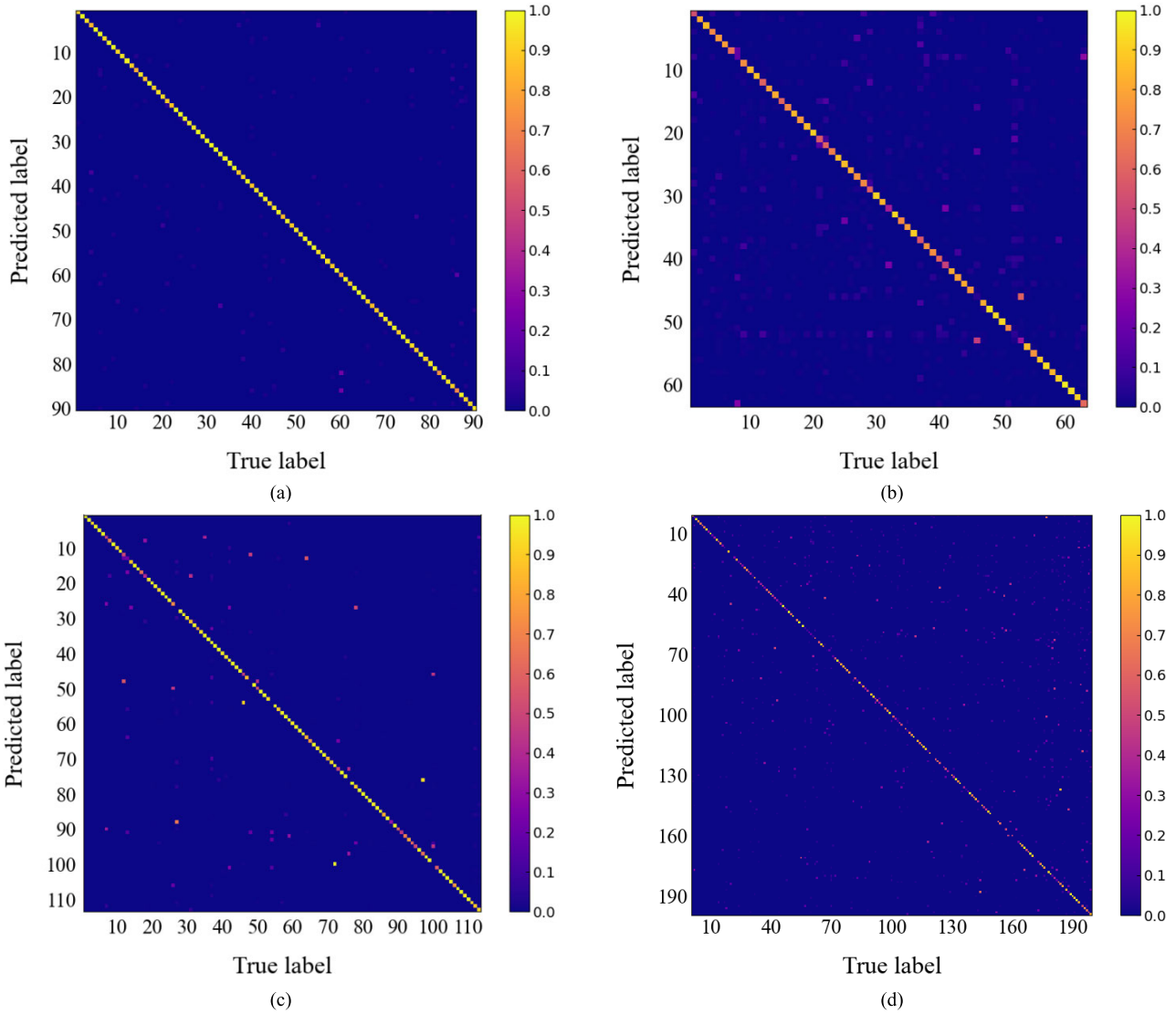

FIGURE 7. The confusion matrix for the multi-heartbeat recognition experiment. (a) ECG-ID database; (b) CYBHi database; (c) PTB database; (d) Heartprint database.

the model obtained recognition accuracies of 54.33% and 54.78% in both recognition modes. These experimental results demonstrate that ADAFFN can accurately identify the true identity, even when using ECG signals with longer intervals for recognition.

E. DISCUSSION

1) CONFUSION MATRIX

This study provides the confusion matrix for the multi-heartbeat recognition experiment. Correctly identified samples are represented by the numbers on the diagonal. The color of the confusion matrix diagonal progressively shifts to gold as the number of accurately classified samples increases. From Fig. 7, it is evident that the confusion matrices for the ECG-ID database and PTB database primarily consist of golden boxes along the diagonal, with only a small portion of red boxes. This indicates that the majority of individuals can be accurately identified using the model. Secondly, the confusion matrix obtained for the CYBHi database indicates a prevalence of red boxes along the diagonal. The result demonstrates that the identification of individuals in the CYBHi database was satisfactory, indicating that the proposed model achieved accurate recognition despite the presence of noise in the data. Lastly, the confusion matrix

TABLE 7. Comparative experimental results of different attentional mechanisms.

Attention	Parameters(B)	Time(ms)	Accuracy(%)
MPAFF	135903	4.348	93.39
ECA	6	4.259	93.20
CBAM	45390	4.314	92.74
SENet	45000	4.313	93.16
WFAA	**909**	**4.300**	**94.18**

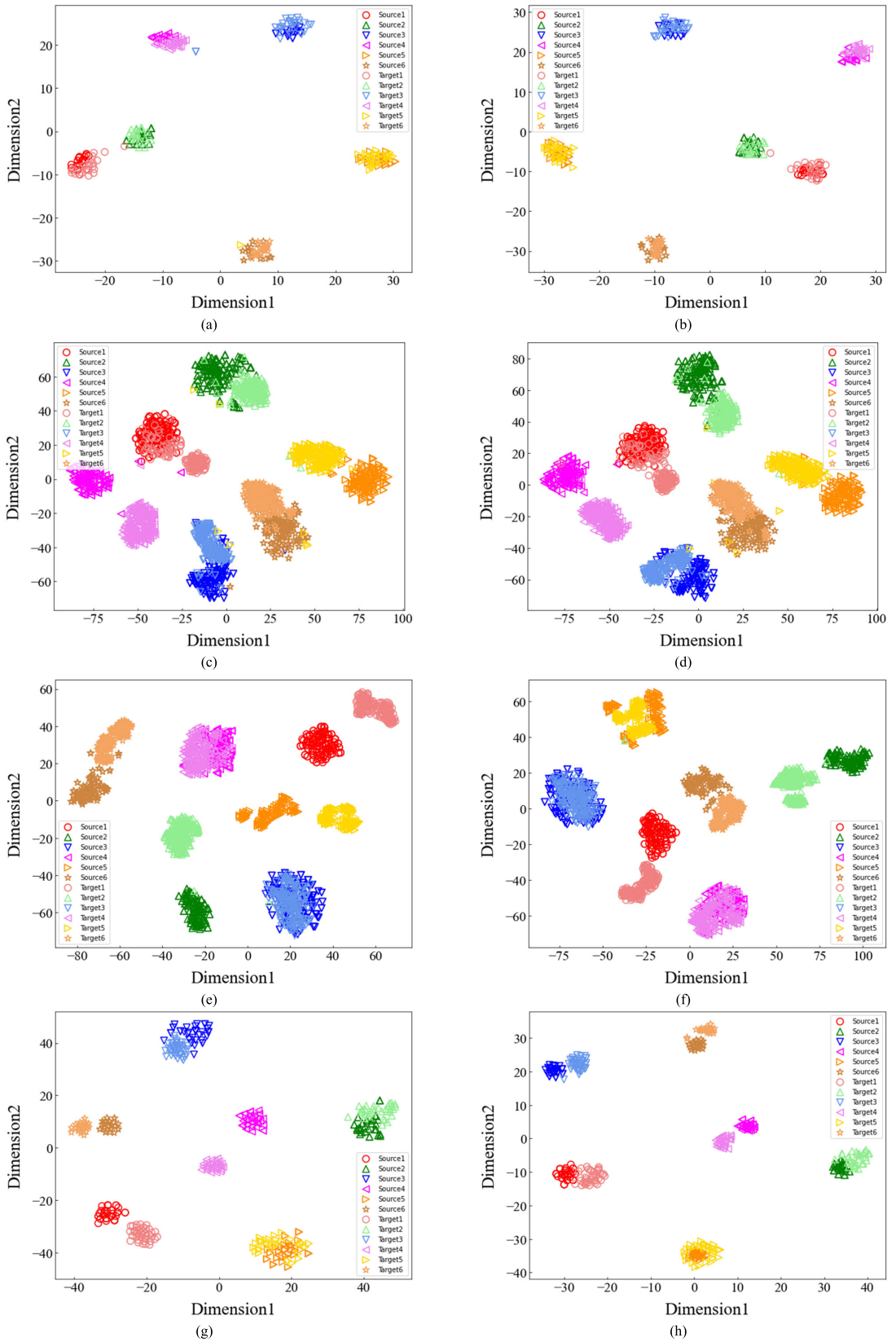

FIGURE 8. Comparison results before and after domain adaptation. (a) ECG-ID (before domain adaptation); (b) ECG-ID (after domain adaptation); (c) CYBHi (before domain adaptation); (d) CYBHi (after domain adaptation); (e) PTB (before domain adaptation); (f) PTB (after domain adaptation);(g) Heartprint (before domain adaptation); (h) Heartprint (after domain adaptation).

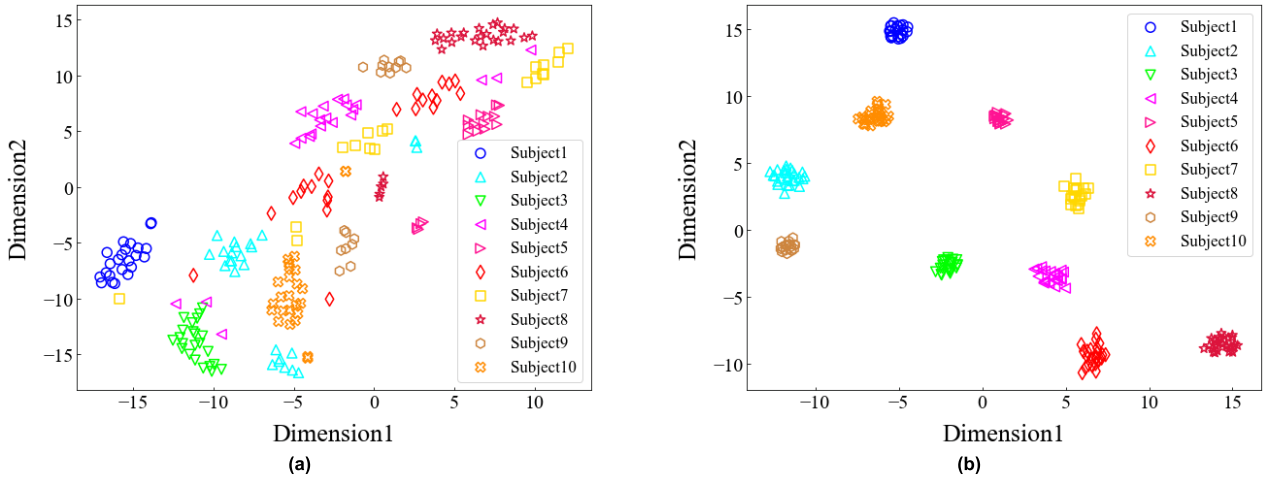

FIGURE 9. Comparison results of NSL and Softmax Loss distinguishing different individuals. (a) Softmax Loss; (b) NSL.

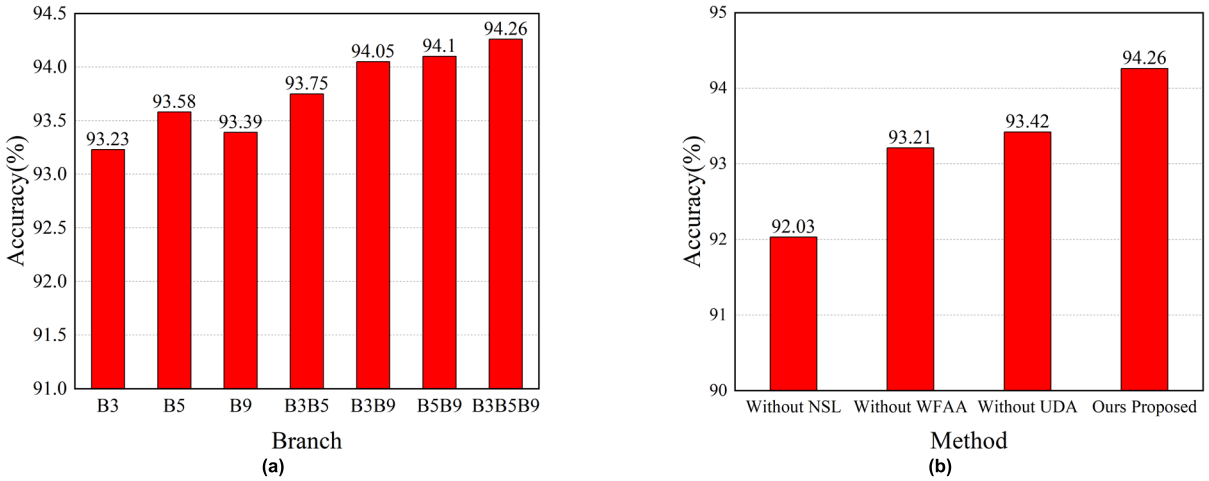

FIGURE 10. Ablation experiment results. (a) The effect of the number of network branches on the recognition results; (b) The impact of different methods on the recognition results.

obtained from the Heartprint database reflects a similar preponderance of red boxes along the diagonal. A subset of individuals displayed lower identification results. However, given the substantial number of individuals, the obtained results can still be deemed satisfactory.

2) COMPARATIVE RESULTS OF ATTENTION MECHANISMS

To highlight the advantages of the proposed attention mechanism in this study, comparative experiments were conducted with several commonly used attention mechanisms, including MPAFF [22], ECA [32], CBAM [38], and SENet [39]. The specific comparison results are presented in Table 7. Firstly, an input channel size of 300 was used to compare the parameters of the attention mechanisms. It was observed that the number of parameters in the proposed WFAA mechanism was lower than the other attention mechanisms, except for the ECA mechanism. Since the WFAA mechanism includes multiple ECA mechanisms, it is expected to have higher parameters than the ECA mechanism. Subsequently, a recognition experiment was performed using the ECG-ID database, employing the same data processing method as the multi-session recognition experiment. The WFAA mechanism demonstrated a recognition time of only 4.3 milliseconds per sample, surpassing the other attention mechanisms except for the ECA mechanism. Although the WFAA mechanism slightly lagged behind the ECA mechanism in terms of the number of parameters and single-sample recognition time, the final recognition accuracy comparison revealed that the WFAA mechanism achieved the highest recognition result. This indicates that the WFAA mechanism possesses a stronger ability to improve the accuracy of ECG recognition, ultimately leading to superior recognition outcomes.

3) DOMAIN ADAPTIVE RESULT VISUALIZATION

To provide a clearer understanding of the application of domain adaptation technology, this study selects multiple

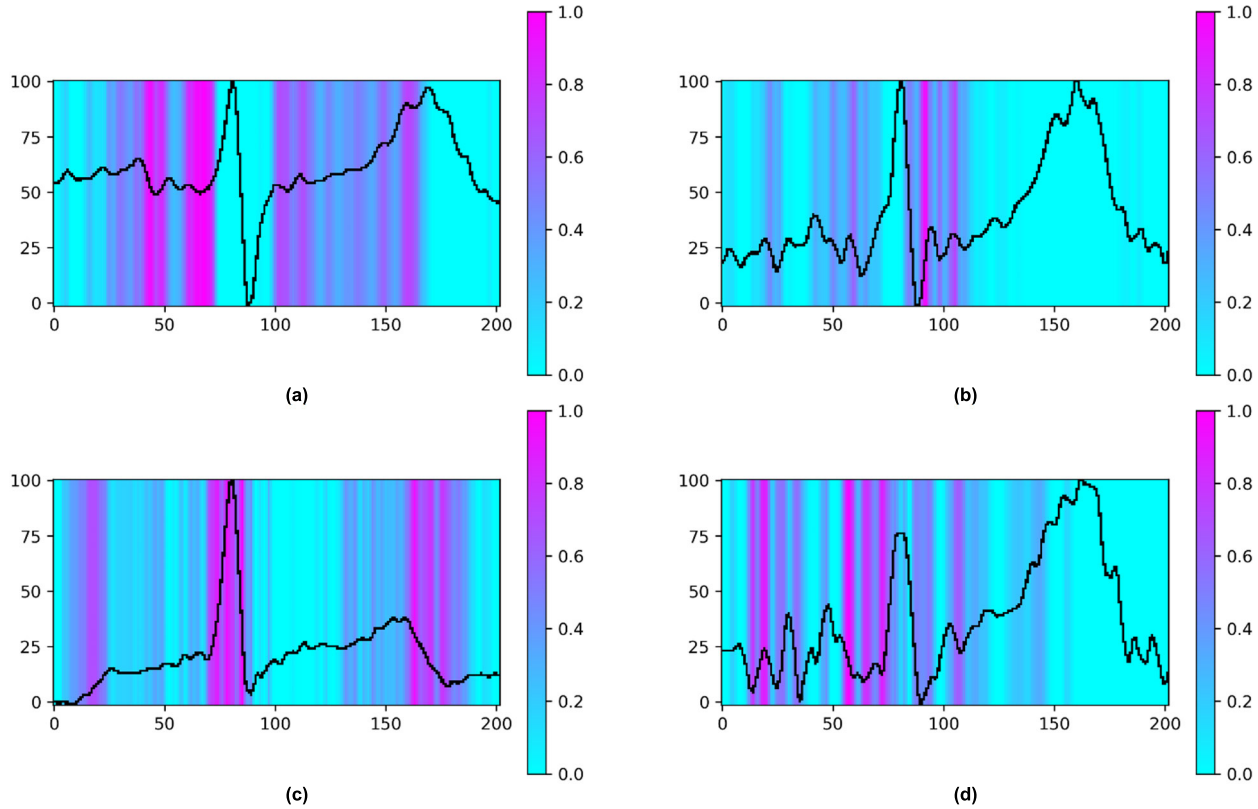

FIGURE 11. Visualization results of features extracted by the network. (a) Visualization results for Individual 1; (b) Visualization results for Individual 2; (c) Visualization results for Individual 3; (d) Visualization results for Individual 4.

individuals from each database and utilizes all their data to compare the results before and after domain adaptation. The features of the individuals are initially extracted using the trained model. Subsequently, dimensionality reduction is applied to these features using T-distributed Stochastic Neighbor Embedding (T-SNE) to obtain the visualization results. Fig. 8 illustrates that domain adaptation effectively mitigates the disparities in feature distribution among ECG signals across sessions. The feature distributions of the same categories in both the source and target domains exhibit greater clustering, while the distributions of features with distinct categories manifest increased dispersion.

4) COMPARISON RESULTS OF LOSS FUNCTIONS

To validate the effectiveness of the NSL in the final recognition experiment, this study conducted a comparison with the conventional Softmax Loss. First, this study trained two models on the ECG-ID database using different loss functions: NSL and Softmax Loss. Subsequently, we selected the top 10 individuals from the ECG-ID database, extracted features from them using the two models, and applied T-SNE for dimensionality reduction. The results of the dimensionality reduction were plotted and visualized in Fig. 9. In comparison to the traditional Softmax Loss, the NSL demonstrated improved differentiation between individuals of different types. Additionally, it exhibited a more pronounced correlation between the same individuals, resulting in a more compact distribution. Hence, the NSL is deemed more suitable for ECG identification experiments.

5) ABLATION EXPERIMENT

To evaluate the impact of each module on ECG identification, this study conducted ablation experiments, and the specific results are presented in Fig. 10. The experiments were conducted using the ECG-ID database, following the data processing method employed in the multi-session recognition experiment. Fig. 10(a) depicts the effect of different network branches on recognition accuracy, with B3, B5, and B9 representing branches with dilation rates of 3, 5, and 9, respectively. It is evident that as the number of branches increases, the recognition accuracy improves significantly, reaching its peak when all three branches are utilized. Fig. 10(b) illustrates the effect of different methods on recognition accuracy. Employing all three methods simultaneously yields the highest accuracy while removing any of the methods results in a decrease in accuracy. Notably, the NSL has the greatest impact on recognition accuracy, as substituting it with the ordinary Softmax Loss leads to a 2.23% decrease in accuracy. This finding further confirms the suitability of the NSL for ECG identification.

TABLE 8. Comparison results of multi-session recognition experiments.

Related Works	Dataset	Persons	Heartbeat Length(s)	Heartbeat Number	Accuracy(%)
Ibtehaz *et al.* (2021) [18]	ECG-ID	89	0.5	1	SI = 92.17
AlDuwaile *et al.* (2021) [8]		90	0.5	1	SI = 94.18
Bento *et al.* (2020) [13]		89	4	1	MI = 73.54
Jyotishi *et al.* (2020) [10]		89	2	1	MI = 86.24
Chee *et al.* (2022) [40]		89	3	1	MI = 92.70
Our Proposed		**90**	**0.6**	**1**	**SI = 94.26**
				3	**MI = 96.31**
Sun *et al.* (2019) [21]	PTB	50+	4	1	MI = 56.93
Chee *et al.* (2022) [40]		112	3	1	MI = 64.16
Our Proposed		**113**	**0.6**	**1**	**SI = 73.53**
				3	**MI = 73.79**
Belo *et al.* (2020) [19]	CYBHi	63	0.5	1	SI = 60.65
Ibtehaz *et al.* (2021) [18]		63	0.5	1	SI = 72.76
Ciocoiu *et al.* (2017) [20]		63	3	1	MI = 60.00
Our Proposed		**63**	**0.6**	**1**	**SI = 75.31**
				3	**MI = 77.80**
Islam et al. (2022) [41]	Heartprint	199	1	1	SI = 52.27
Our Proposed		**199**	**0.6**	**1**	**SI = 54.33**
				3	**MI = 54.78**

6) GRAD-CAM VISUALIZATION

Given the inherent opaqueness of deep learning models, direct observation of their extracted features becomes unfeasible. Various visualization methods have emerged to enhance our understanding of deep learning networks, providing visual explanations, and aiding in informed decision-making about these models. One of the well-known techniques in this regard is Gradient-weighted Class Activation Mapping (Grad-CAM), which we utilized to visualize the features extracted from the attention layer of the network. This method acquires feature weights through the computation of the global average pooling of the gradients and accentuates locations with a positive impact on classification by employing the ReLU activation function. In this study, four individuals were randomly selected from the ECG-ID database, and the Grad-CAM was harnessed to pinpoint the specific heartbeat features that play a crucial role in authenticating the identity of an individual during the final identification process. The specific visualization results are depicted in Fig. 11. The figure demonstrates the varying weights assigned to individual ECG signal features in the final recognition task. To enhance discrimination between different individuals, the network concentrates on extracting the most distinguishing features from individual ECG signals. As a consequence of the variations in ECG signals among different individuals, the model's attention differs for each individual. This experimental outcome provides additional evidence supporting the viability of utilizing ECG signals for identification purposes.

7) COMPARISON WITH RELATED RESEARCH RESULTS

To demonstrate the superiority of the proposed model in this study, a comparative analysis was conducted with related studies, and the results are presented in Table 8. The comparison was performed for both single-heartbeat identification (SI) and multi-heartbeat identification (MI) modes. The findings indicate that the research model achieves superior recognition results in both modes. Specifically, in the ECG-ID database, the multi-heartbeat recognition results obtained in this study are 3.61% higher than those reported in [40], and in the PTB database, the multi-heartbeat recognition results are 9.63% higher than [40]. Moreover, the multi-heartbeat recognition results in the CYBHi database are 17.8% higher than [20]. Finally, the single-heartbeat recognition results in the Heartprint database are 2.06% higher than [41]. These comparisons further highlight the advantages of the ADAFFN proposed in this study for multi-session recognition experiments. It demonstrates that

even when using ECG signals with long intervals as recognition signals, superior recognition results can be achieved.

V. CONCLUSION

The existing research models have shown poor recognition results in multi-session recognition experiments due to their inability to effectively extract key features from ECG signals. To address the limitation, this study proposes an attention-enhanced domain adaptive feature fusion network. Firstly, the network employs a multi-branch structure to comprehensively extract crucial features from the ECG signal across various dimensions. Additionally, the network incorporates a weight fusion attention mechanism designed in this study to enhance the recognition ability of the model. Furthermore, domain adaptive technology is employed to reduce the differences in feature distribution of ECG signals between sessions, thereby enhancing the model's generalization ability and improving recognition results in multi-session experiments. Finally, the model is trained to employ a Normalized version of the Softmax Loss instead of the traditional Softmax Loss. This decision is made to address the issue of overlapping decision boundaries. The experimental results demonstrate that the proposed ADAFFN achieves better recognition results in the two recognition modes across four public databases. Ablation experiments and visualization results further confirm the effectiveness of each component of the network for ECG recognition. Simultaneously, the GRAD-CAM visualization serves as further evidence of the feasibility of employing ECG signals for identification. Finally, a comparison of the experimental results with related research demonstrates that ADAFFN exhibits superior recognition ability in multi-session recognition experiments. It effectively utilizes ECG signals with longer intervals to achieve more accurate identification of an individual's true identity.

In comparison to related research, the present study has achieved improved recognition results in multi-session recognition experiments. However, there is still potential for further enhancement in the final recognition accuracy. Future work will focus on exploring and developing related algorithms or models to continue improving the overall recognition performance. Furthermore, despite utilizing the Heartprint database in this study for recognition experiments, there remains a necessity to encompass a broader range of individuals for enhanced recognition. A more extensive cohort of subjects is conducive to optimizing the performance of the detection model, and the quest for datasets encompassing a greater number of subjects for forthcoming experiments will persist. Finally, the voting recognition method applied in this study is well-suited for the model proposed in this study. It's worth noting that this method might result in longer recognition times during the testing phase compared to the long heartbeat recognition method. In the future, we will explore more advanced methods to enhance the recognition results.

REFERENCES

[1] Y. Sun, Y. Tang, and X. Chen, "A neural network-based partial fingerprint image identification method for crime scenes," *Appl. Sci.*, vol. 13, no. 2, pp. 1188–1201, Jan. 2023, doi: 10.3390/app13021188.

[2] P. Contardo, P. Sernani, S. Tomassini, N. Falcionelli, M. Martarelli, P. Castellini, and A. F. Dragoni, "FRMDB: Face recognition using multiple points of view," *Sensors*, vol. 23, no. 4, pp. 1939–1959, Feb. 2023, doi: 10.3390/s23041939.

[3] S. M. Arnoos, A. M. Sahan, A. H. O. Ansaf, and A. S. Al-Itbi, "An intelligent iris recognition technique," in *Proc. Int. Conf. Gener. Internet Things. (ICNGIoT)*, vol. 445, Feb. 2023, pp. 207–217.

[4] Y. Ren, Z. Zheng, S. Xu, and H. Li, "User identification leveraging whispered sound for wearable devices," *IEEE Trans. Mobile Comput.*, vol. 22, no. 3, pp. 1841–1855, Mar. 2023, doi: 10.1109/TMC.2021.3112528.

[5] J. Cui, L. Su, H. Hu, G. Li, Z. Chang, and R. Wei, "EEG pattern identification for motor imagery based on 1DCNN-GRU," *Multimedia Tools Appl.*, vol. 82, no. 13, pp. 20605–20620, Jan. 2023, doi: 10.1007/s11042-023-14380-7.

[6] R. Hoekema, G. J. H. Uijen, and A. van Oosterom, "Geometrical aspects of the interindividual variability of multilead ECG recordings," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 5, pp. 551–559, May 2001, doi: 10.1109/10.918594.

[7] J. Wu, C. Liu, Q. Long, and W. Hou, "Research on personal identity verification based on convolutional neural network," in *Proc. IEEE 2nd Int. Conf. Inf. Comput. Technol. (ICICT)*, Mar. 2019, pp. 57–64.

[8] D. A. AlDuwaile and M. S. Islam, "Using convolutional neural network and a single heartbeat for ECG biometric recognition," *Entropy*, vol. 23, no. 6, pp. 733–754, Jun. 2021, doi: 10.3390/e23060733.

[9] D. Jyotishi and S. Dandapat, "Person identification using spatial variation of cardiac signal," in *Proc. IEEE Appl. Signal Process. Conf. (ASPCON)*, Oct. 2020, pp. 168–172.

[10] D. Jyotishi and S. Dandapat, "An LSTM-based model for person identification using ECG signal," *IEEE Sensors Lett.*, vol. 4, no. 8, pp. 1–4, Aug. 2020, doi: 10.1109/LSENS.2020.3012653.

[11] R. Srivastva, A. Singh, and Y. N. Singh, "PlexNet: A fast and robust ECG biometric system for human recognition," *Inf. Sci.*, vol. 558, pp. 208–228, May 2021, doi: 10.1016/j.ins.2021.01.001.

[12] M. Hammad, P. Pławiak, K. Wang, and U. R. Acharya, "ResNet-Attention model for human authentication using ECG signals," *Expert Syst.*, vol. 38, no. 6, pp. 12757–12773, Sep. 2021, doi: 10.1111/exsy.12547.

[13] N. Bento, D. Belo, and H. Gamboa, "ECG biometrics using spectrograms and deep neural networks," *Int. J. Mach. Learn. Comput.*, vol. 10, no. 2, pp. 259–264, Feb. 2020, doi: 10.18178/ijmlc.2020.10.2.929.

[14] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: A new approach in human identification," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 3, pp. 808–812, Jun. 2001, doi: 10.1109/19.930458.

[15] I. El Boujnouni, H. Zili, A. Tali, T. Tali, and Y. Laaziz, "A wavelet-based capsule neural network for ECG biometric identification," *Biomed. Signal Process. Control*, vol. 76, Jul. 2022, Art. no. 103692, doi: 10.1016/j.bspc.2022.103692.

[16] B. Fatimah, P. Singh, A. Singhal, and R. B. Pachori, "Biometric identification from ECG signals using Fourier decomposition and machine learning," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, Aug. 2022, doi: 10.1109/TIM.2022.3199260.

[17] I. B. Ciocoiu and N. Cleju, "Off-person ECG biometrics using spatial representations and convolutional neural networks," *IEEE Access*, vol. 8, pp. 218966–218981, 2020, doi: 10.1109/ACCESS.2020.3042547.

[18] N. Ibtehaz, M. E. H. Chowdhury, A. Khandakar, S. Kiranyaz, M. S. Rahman, A. Tahir, Y. Qiblawey, and T. Rahman, "EDITH : ECG biometrics aided by deep learning for reliable individual authentication," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 4, pp. 928–940, Aug. 2022, doi: 10.1109/TETCI.2021.3131374.

[19] D. Belo, N. Bento, H. Silva, A. Fred, and H. Gamboa, "ECG biometrics using deep learning and relative score threshold classification," *Sensors*, vol. 20, no. 15, pp. 4078–4098, Jul. 2020, doi: 10.3390/s20154078.

[20] I. B. Ciocoiu, "Comparative analysis of bag-of-words models for ECG-based biometrics," *IET Biometrics*, vol. 6, no. 6, pp. 495–502, Nov. 2017, doi: 10.1049/iet-bmt.2016.0177.

[21] H. Sun, Y. Guo, B. Chen, and Y. Chen, "A practical cross-domain ECG biometric identification method," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[22] P. Feng, J. Fu, Z. Ge, H. Wang, Y. Zhou, B. Zhou, and Z. Wang, "Unsupervised semantic-aware adaptive feature fusion network for arrhythmia detection," *Inf. Sci.*, vol. 582, pp. 509–528, Jan. 2022, doi: 10.1016/j.ins.2021.09.046.

[23] G. Wang, M. Chen, Z. Ding, J. Li, H. Yang, and P. Zhang, "Inter-patient ECG arrhythmia heartbeat classification based on unsupervised domain adaptation," *Neurocomputing*, vol. 454, pp. 339–349, Sep. 2021, doi: 10.1016/j.neucom.2021.04.104.

[24] R. Jane, P. Laguna, N. V. Thakor, and P. Caminal, "Adaptive baseline wander removal in the ECG: Comparative analysis with cubic spline technique," in *Proc. Comput. Cardiol.*, Oct. 1992, pp. 143–146.

[25] A. A. Date and R. B. Ghongade, "Performance of wavelet energy gradient method for QRS detection," in *Proc. 4th Int. Conf. Intell. Adv. Syst. (ICIAS)*, vol. 2, Jun. 2012, pp. 876–881.

[26] M. Rakshit and S. Das, "An efficient ECG denoising methodology using empirical mode decomposition and adaptive switching mean filter," *Biomed. Signal Process. Control*, vol. 40, pp. 140–148, Feb. 2018, doi: 10.1016/j.bspc.2017.09.020.

[27] P. Hamilton, "Open source ECG analysis," in *Proc. Comput. Cardiol.*, Sep. 2002, pp. 101–104, doi: 10.1109/CIC.2002.1166717.

[28] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, Oct. 2018, doi: 10.1016/j.neucom.2018.05.083.

[29] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 59, pp. 1–35, Apr. 2016, doi: 10.48550/arXiv.1505.07818.

[30] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 510–519.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[32] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.

[33] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "CosFace: Large margin cosine loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5265–5274.

[34] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. 215–220, Jun. 2000, doi: 10.1161/01.cir.101.23.e215.

[35] A. P. Nemirko and T. S. Lugovaya, "Biometric human identification based on electrocardiogram," in *Proc. 12th Russian Conf. Math. Methods Pattern Recognit.*, 2005, pp. 387–390.

[36] R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung der EKG-signaldatenbank CARDIODAT der PTB über das Internet," *Biomedizinische Technik/Biomed. Eng.*, pp. 317–318, Jul. 2009, doi: 10.1515/bmte.1995.40.s1.317.

[37] H. P. da Silva, A. Lourenço, A. Fred, N. Raposo, and M. Aires-de-Sousa, "Check your biosignals here: A new dataset for off-the-person ECG biometrics," *Comput. Methods Programs Biomed.*, vol. 113, no. 2, pp. 503–514, Feb. 2014, doi: 10.1016/j.cmpb.2013.11.017.

[38] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Aug. 2018, pp. 3–19.

[39] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[40] K. J. Chee and D. A. Ramli, "Electrocardiogram biometrics using transformer's self-attention mechanism for sequence pair feature extractor and flexible enrollment scope identification," *Sensors*, vol. 22, no. 9, pp. 3446–3474, Apr. 2022, doi: 10.3390/s22093446.

[41] M. S. Islam, H. Alhichri, Y. Bazi, N. Ammour, N. Alajlan, and R. M. Jomaa, "Heartprint: A dataset of multisession ECG signal with long interval captured from fingers for biometric recognition," *Data*, vol. 7, no. 10, pp. 141–156, Oct. 2022, doi: 10.3390/data7100141.

[42] A. J. Prakash, K. K. Patro, M. Hammad, R. Tadeusiewicz, and P. Pławiak, "BAED: A secured biometric authentication system using ECG signal based on deep learning techniques," *Biocybernetics Biomed. Eng.*, vol. 42, no. 4, pp. 1081–1093, Oct. 2022, doi: 10.1016/j.bbe.2022.08.004.

[43] X. Wang, W. Cai, and M. Wang, "A novel approach for biometric recognition based on ECG feature vectors," *Biomed. Signal Process. Control*, vol. 86, pp. 104922–104929, Sep. 2023, doi: 10.1016/j.bspc.2023.104922.

[44] A. J. Prakash, K. K. Patro, S. Samantray, P. Pławiak, and M. Hammad, "A deep learning technique for biometric authentication using ECG beat template matching," *Information*, vol. 14, no. 2, pp. 65–80, Jan. 2023, doi: 10.3390/info14020065.

[45] M. Hazratifard, V. Agrawal, F. Gebali, H. Elmiligi, and M. Mamun, "Ensemble Siamese network (ESN) using ECG signals for human authentication in smart healthcare system," *Sensors*, vol. 23, no. 10, pp. 4727–4740, May 2023, doi: 10.3390/s23104727.

[46] D. Wang, Y. Si, W. Yang, G. Zhang, and J. Li, "A novel electrocardiogram biometric identification method based on temporal-frequency autoencoding," *Electronics*, vol. 8, no. 6, pp. 667–690, Jun. 2019, doi: 10.3390/electronics8060667.

PAN YI received the bachelor's degree in automation from the Hunan Institute of Engineering, in 2021. He is currently pursuing the master's degree with the College of Communication Engineering, Jilin University, China. His research interests include biomedical signal processing and recognition. His current research focuses on ECG biometric identification and recognition.

YUJUAN SI received the master's and Ph.D. degrees in engineering from the Jilin University of Technology, in 1988 and 1996, respectively. She is currently a Professor with the Institute of Communication Engineering, Jilin University. Her research interests include embedded systems and biomedical signal processing and recognition.

WEI FAN received the bachelor's and Ph.D. degrees in signal and information processing from Jilin University, Changchun, China, in 2018 and 2023, respectively. His research interests include biomedical signal processing and recognition. His current research focuses on ECG biometric identification and recognition.

YANG ZHANG received the bachelor's degree in signal and information processing from Northeast Electric Power University, in 2021. He is currently pursuing the master's degree with the College of Communication Engineering, Jilin University, China. His research interests include biomedical signal processing and recognition. His current research focuses on ECG biometric identification and recognition.