ECG Identity Authentication in Open-set with Multi-model Pretraining and Self-constraint Center & Irrelevant Sample Repulsion Learning

Mingyu Dong[a], Zhidong Zhao[*,a], Hao Wang[a], Yefei Zhang[a], Yanjun Deng[a]

[a]Hanghzou Dianzi University, , Hangzhou, 310018, Zhejiang, China

Abstract

Electrocardiogram (ECG) signal exhibits inherent uniqueness, making it a promising biometric modality for identity authentication. As a result, ECG authentication has gained increasing attention in recent years. However, most existing methods focus primarily on improving authentication accuracy within closed-set settings, with limited research addressing the challenges posed by open-set scenarios. In real-world applications, identity authentication systems often encounter a substantial amount of unseen data, leading to potential security vulnerabilities and performance degradation. To address this issue, we propose a robust ECG identity authentication system that maintains high performance even in open-set settings. Firstly, we employ a multi-modal pretraining framework, where ECG signals are paired with textual reports derived from their corresponding fiducial features to enhance the representational capacity of the signal encoder. During fine-tuning, we introduce Self-constraint Center Learning and Irrelevant Sample Repulsion Learning to constrain the feature distribution, ensuring that the encoded representations exhibit clear decision boundaries for classification. Our method achieves 99.83% authentication accuracy and maintains a False Accept Rate as low as 5.39% in the presence of open-set samples. Furthermore, across various open-set ratios, our method demonstrates exceptional stability, maintaining an Open-set Classification Rate above 95%.

Keywords: Electrocardiogram, Identity Authentication, Open Set, Multi-modal

1. Introduction

[1] Electrocardiogram (ECG) signals capture the electrical activity of the heart throughout its cardiac cycle, with each individual exhibiting unique physiological characteristics. Due to these inherent individual differences [9], ECG signals have gained increasing attention in recent years as a biometric modality for identity authentication [27]. Compared to well-established biometric technologies such as facial recognition and fingerprint identification [28], ECG authentication offers distinct advantages in terms of security and resilience against spoofing. The intrinsic uniqueness of ECG signals, coupled with their dynamic nature, makes them significantly more difficult to replicate or forge, positioning ECG as a promising and robust feature for biometric identification. Advancements in signal acquisition technology have significantly enhanced the quality of non-invasive data collection, enabling the capture of clear and well-defined waveforms [21]. This progress has not only improved the accuracy and reliability of acquired signals but has also facilitated seamless and unobtrusive acquisition methods. As a result, the feasibility and practicality of utilizing such signals for identity authentication have been greatly enhanced, offering a more user-friendly and efficient approach to biometric verification.

In early research on ECG identity authentication, fiducial features were introduced to determine morphological characteristics that could be used for identity differentiation [23]. A commonly employed category of these features includes the time intervals between standard medical fiducial points on the ECG waveform, such as P, Q, R, S, and T waves. In addition to these fiducial features, various non-fiducial features have been explored. Examples of such features include principal components [11], wavelet coefficients [4], and autocorrelation coefficients [30]. After extracting these features, identity authentication is performed by comparing the similarity between the extracted features and the stored reference templates in the database. This similarity assessment determines whether the input belongs to the same individual.

With the advancement of deep learning, the application of deep learning models for ECG signal classification has gradually replaced traditional identity authentication methods based on direct similarity comparison. Deep learning models, composed of multiple hidden layers, learn sample distributions through non-linear mappings and ultimately classify the extracted features. Both methods [29] and [16] utilize deep learning-based approaches, while method [2] employs graph neural networks for identity verification. Additionally, Transformer [12] and Mamba [25] models have also been explored for ECG abnormal classification, further expanding the range of deep learning techniques applied in this domain.

In the field of ECG disease diagnosis, Liu et al. proposed a dual-modal approach that integrates both ECG signals and clinical reports for disease classification [18]. This multi-modal framework leverages the complementary nature of textual and signal modalities, demonstrating that incorporating clinical text into the training process can significantly enhance the model's ability to represent ECG signals. However, in the domain of
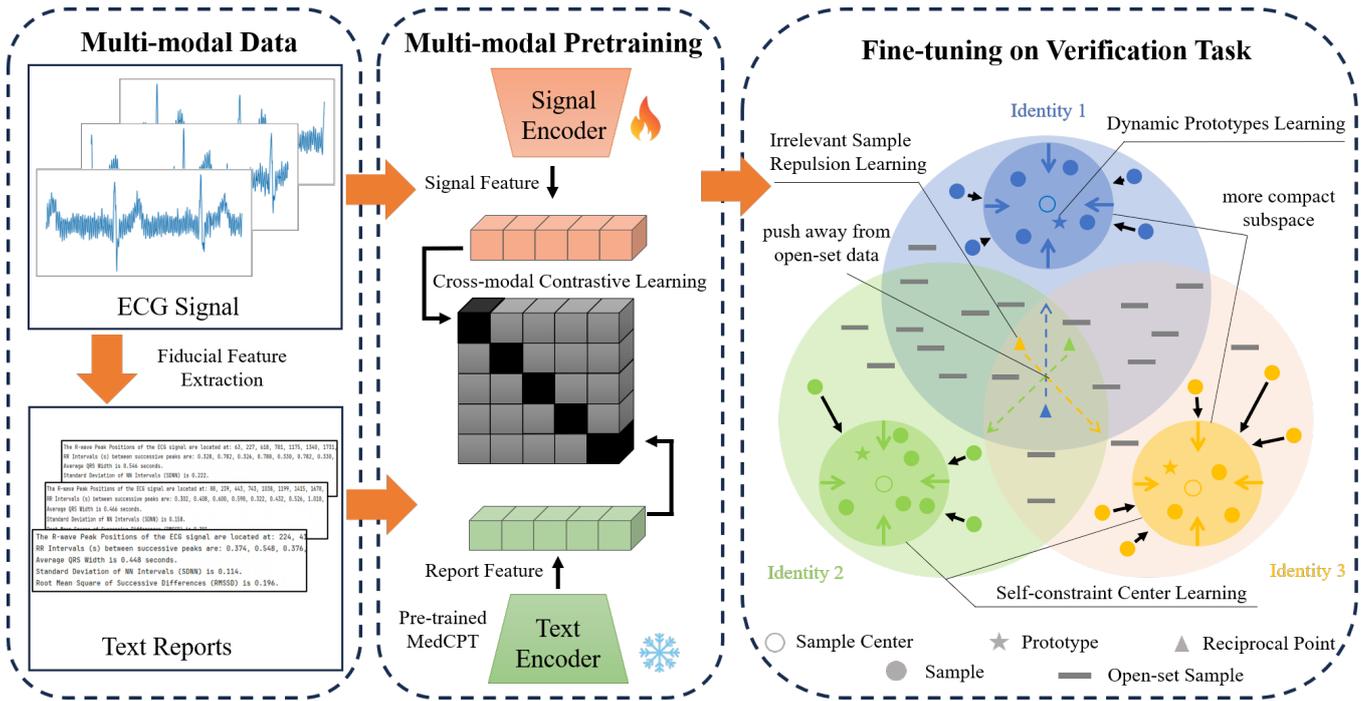
Figure 1: The proposed method is outlined in the workflow diagram, which consists of two main components: multi-modal pretraining and fine-tuning on identity authentication task. Within component B, we introduce Self-constraint Center Learning and Irrelevant Sample Repulsion Learning.

identity authentication, textual data has rarely been utilized as an additional modality in training pipelines. Given that textual information often encapsulates crucial contextual knowledge, its integration could play a vital role in improving the capability of signal encoders to capture identity-related features within ECG data. By incorporating text as an auxiliary modality, identity authentication models can achieve more robust and discriminative representations.

In real-world identity authentication scenarios, systems often face saturation attacks, where they struggle to effectively distinguish whether an input sample belongs to a registered class. Most identity authentication systems based on classification models assign a predefined label to each input signal, making them inherently vulnerable to security risks. Wu et al. were the first to highlight the issue of system stability in ECG identity authentication under open-set conditions [32]. Building upon this foundation, this study conducts a comprehensive robustness evaluation of ECG identity authentication across a broader range of open-set data.

To enhance the capacity for ECG signal representation of the model, we propose a multi-modal pretraining module that integrates both ECG signals and fiducial feature text report. During the fine-tuning phase, we introduce a novel training strategy combining Self-constraint Center Learning and Irrelevant Sample Repulsion Learning. These modules enable the model to achieve better classification capability even without additional open-set data, improving the generalization to unseen open-set samples. Our method effectively reshapes the distribution of sample features, ensuring that training samples are mapped into a more compact and well-structured subspace. Meanwhile, features extracted from open-set samples, which are not present

during training, are constrained within a predefined range, preventing them from interfering with the labeled feature space. In smaller-scale open-set scenarios, our method effectively distinguishes between known and unknown samples with high precision. This capability ensures that the model not only accurately determines whether an input sample belongs to a previously registered identity but also reliably classifies and verifies registered identities. By enhancing the model's ability to differentiate between in-distribution and out-of-distribution samples, our method strengthens the robustness of identity authentication systems. The contributions of this paper are as follows:

- **Multi-modal pretraining.** We leverage a large dataset to pre-train a multi-modal model that integrates ECG signals with fiducial feature text report. Using contrastive learning, we align the modalities, ensuring a correspondence between the information in the text and the ECG signal.

- **Self-constraint Center Learning.** The sample center is defined as the centroid of the sample feature distribution. During training, sample features are explicitly encouraged to move closer to their respective class centers. To mitigate potential biases arising from imbalanced training data distributions, we further incorporate a Dynamic Prototype Learning mechanism. This component adaptively adjusts the class prototypes throughout training, promoting a more balanced and representative feature distribution across classes.

- **Irrelevant Sample Repulsion Learning.** To approximate the distribution of open-set data without explicitly involving any open-set samples during training, we introduce the

2

concept of irrelevant sample. These samples are designed to represent the distributional characteristics of all other identity classes beyond the target identity class, within the constraints of a limited dataset. By encouraging target identity class samples to be distant from these irrelevant samples, the model is guided to form more distinct and well-separated decision boundaries.

2. Methods

This section focuses on our method of ECG identity authentication in an open-set setting. In Section 2.1, we first define the problem formally. The subsequent sections detail our proposed module, which comprises three key components: Multimodal pretraining, Self-constraint Center Learning, and Irrelevant Sample Repulsion Learning. The overall workflow is illustrated in Figure 1.

2.1. Problem Definition

Given a set of ECG signals with their identities $\mathcal{D}_L = \{(x_1, id_1), ..., ((x_n, id_n)\}$. N is the registered identities $id_i \in \{1, ..., N\}$, id_i is the identity of the signal x_i. Given another larger amount of test data $\mathcal{D}_T = \{t_1, ..., t_u\}$ where the identity of t_i belongs to $\{1, ..., N\} \cup \{N + 1, ..., N + U\}$ which contains close-set data and open-set data. The U is the number of unregistered identities in realistic scenarios. the deep embedding feature of category k is denoted by S_k and $S_k \in \mathcal{D}_L$.

For the model ψ, it is crucial to learn the feature representations of registered users from the training dataset $\mathcal{D}_L$ and establish a well-defined feature distribution. During the training phase, the model must establish clear decision boundaries between samples of different identity labels while ensuring sufficient separation R between these boundaries. To enhance the generalization to open-set scenarios despite a limited number of training samples, all samples except those belonging to a designated label id are treated as open-set data $\mathcal{D}_L^{\neq id}$. During the testing phase, the dataset includes open-set samples, introducing unseen categories that were not present during training. The model is designed to achieve two **primary objectives**: (1) accurately classify samples belonging to previously registered identities $id \in \{1, ..., N\}$, ensuring high recognition performance within known identities; and (2) effectively identify and exclude unregistered samples $id \in \{N + 1, ..., N + U\}$, minimizing false acceptances of unseen identities.

2.2. Multi-modal Pretraining

In ECG signals, fiducial features refer to key points or waveform characteristics with significant physiological relevance [3]. Prior research has demonstrated their practicality and effectiveness in ECG-based identity authentication systems [31, 17, 22]. In this work, we aim to leverage fiducial features to enhance the representational capacity of the ECG signal encoder by aligning waveform structures with physiologically meaningful cues.

To this end, we select five representative fiducial features that capture both the morphological and dynamic aspects of ECG

signals: R-wave peak positions, RR intervals, QRS widths, standard deviation of NN intervals (SDNN), and root mean square of successive differences (RMSSD). R-wave peak positions, RR intervals, and QRS widths characterize the overall morphology of the ECG waveform, while the latter two (SDNN and RMSSD) quantify signal variability, which also reflects individual-specific cardiac patterns. These extracted fiducial features are then transformed into a structured textual report to facilitate interpretability and downstream processing. The report follows a predefined template format denoted as *'The R-wave Peak Positions of the ECG signal are located at:* {}'. *RR Intervals between successive peaks are:* {}. *Average QRS Width is* {} *seconds. Standard Deviation of NN Intervals is* {}.*Root Mean Square of Successive Differences is* {}., enabling consistent alignment between waveform features and their corresponding semantic representations.

There is a corresponding relationship between ECG signals and text reports. Extracting relevant information from texts to assist model training benefits the identity authentication. Therefore, in this section, we leverage both ECG signals and their associated text reports to pre-train a model capable of effectively capturing and representing ECG signal features. This pre-trained model serves as a robust feature extractor, facilitating the downstream identity authentication task.

$\{(s_1, r_1), (s_2, r_2), ..., (s_i, r_i)\}$ donated as the ECG signals with the corresponding text reports. To independently extract features from the input data pair, separate encoders are employed for different modalities. Specifically, the report encoder $\mathcal{F}_r$ utilizes a pre-trained model and tokenizer from MedCPT [13], while the signal encoder $\mathcal{F}_s$ adopts various model configurations to process the ECG signals. Following feature extraction, the non-linear projection layers $\mathcal{P}_e$ and $\mathcal{P}_s$ are applied to unify the feature dimensions across modalities, where the features are extracted as $z_{r,i} = \mathcal{P}_r(\mathcal{F}_r(r_i))$ and $z_{s,i} = \mathcal{P}_s(\mathcal{F}_s(r_i))$. To align the representations from different modalities, contrastive learning is employed. The cosine distances between the two modalities can be denoted as $s_{i,i}^{s2r} = z_{s,i}^{\top} z_{r,i}$ and $s_{i,i}^{r2s} = z_{r,i}^{\top} z_{s,i}$, respectively. These values serve as quantitative measures of the similarity between the extracted feature representations from each modality, providing insights into their alignment and compatibility within the learned feature space. The loss function during the pretraining process can be formulated as follows:

$$\mathcal{L}_{i,j}^{s2r} = -\log \frac{\exp(s_{i,j}^{s2r}/\tau)}{\sum_{k=1}^{L} \mathbb{I}_{[k \neq i]} \exp(s_{i,k}^{s2r}/\tau)}, \qquad (1)$$

$$\mathcal{L}_{i,j}^{r2s} = -\log \frac{\exp(s_{i,j}^{r2s}/\tau)}{\sum_{k=1}^{L} \mathbb{I}_{[k \neq i]} \exp(s_{i,k}^{s2r}/\tau)}, \qquad (2)$$

$$\mathcal{L}_{\text{Contrastive}} = \frac{1}{2L} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\mathcal{L}_{i,j}^{r2s} + \mathcal{L}_{i,j}^{r2s} \right). \qquad (3)$$

where, $\mathcal{L}_{i,j}^{s2r}$ and $\mathcal{L}_{i,j}^{r2s}$ represent the signal-report and report-signal cross-modal contrastive losses, respectively. The temperature hyper-parameter, denoted as τ, is set as 0.07 in the experiment. L is the batch size per step, which is a subset of N.

2.3. Self-constraint Center Learning

The pre-trained model obtained through the aforementioned procedure demonstrates strong representational capability for ECG signal. In the subsequent stage, the pre-trained model is fine-tuned on the target dataset for identity registration. For the training dataset $\mathcal{D}_L$, the model is required to correctly assign labels in the classification task.

To address this issue, this paper introduces the concept of sample self-constraining, which aims to map the feature representations of training samples into a more compact subspace. By constraining the feature distribution, the model can better distinguish from open-set samples. Specifically, we define C as the centroid of the sample distribution, ensuring that feature representations remain clustered around a meaningful reference point, which improves both intra-class compactness and inter-class separability. For each identity category id, all samples belonging to the same identity in the $\mathcal{D}_L$ are collected. We define the class-specific sample center C by identifying the instance with the minimum total distance to all other samples within the same identity id. Specifically, for each sample, we compute the sum of pairwise distances to all other samples in the identity class, and select the one with the smallest cumulative distance as the representative center:

$$C_{id} = \arg \min_{m \in S^d} \sum_{i=1}^{n} \sum_{j \neq i}^{n} d(m_{id}^i, m_{id}^j) \tag{4}$$

here, m denotes the signal latent feature extracted by the signal encoder, while n represents the number of samples belonging to a given identity id. The pairwise distance between samples is denoted by $d(a, b) = \sqrt{(a-b)^2}$. These centers serve as key reference points for analyzing the distribution and structural characteristics of the identity representations.

During the training process, each sample in the training set is constrained by the L_2-norm to regulate the distance between its feature representation and the corresponding sample center C. The loss function is formulated as follows:

$$\mathcal{L}_{self}(x; \theta) = \frac{1}{M \cdot N} \sum_{id=1}^{M} \sum_{i=1}^{N} \left\| \mathcal{F}_r(x_i^{id}) - C_{id} \right\|_2^2 \tag{5}$$

where C_{id} represents the sample center corresponding to the specific identity label id. While training, the sample centers are dynamically updated at the beginning of each epoch, ensuring centers adapt to the evolving feature distributions.

The softmax function is commonly used in classification tasks to map feature representations to probability distributions, serving as the basis for loss computation. However, empirical analysis reveals that features transformed by softmax tend to exhibit a loosely distributed structure [6]. Therefore, in this section, we propose an alternative approach that replaces the traditional softmax-based classification loss with a distance-based metric, leveraging pairwise sample distances to enhance feature compactness.

Inspired by [33], we introduce the concept of dynamic prototypes. Unlike center C, which are typically fixed at the centroid of sample feature distributions, dynamic prototypes P^k are learnable parameters that adaptively update the positions based on the distance between the given samples and the prototypes. Sample center C, when estimated from a limited number of samples, is susceptible to shifts influenced by the distribution of outlier or atypical samples. In contrast, a dynamic prototype can effectively represent the learned distribution of a given identity by continuously adapting to the encoding space of the model. By leveraging prototype learning, we provide a dual safeguard against potential biases caused by shifts in the true distribution center. This dynamic adjustment mechanism enables samples to achieve a more balanced distribution between the sample center C and the dynamic prototypes P. In the method, the probability of features in the softmax function is replaced with the distance between samples and their respective prototypes P:

$$Prob(x \in P^{id}|x) = \frac{e^{-d(\mathcal{F}_r(x^{id}), P^{id})}}{\sum_{k=1}^{M} e^{-d(\mathcal{F}_r(x^k), P^k)}} \tag{6}$$

where $d(\mathcal{F}_r(x^{id}), P^{id}) = \left\| \mathcal{F}_r(x^{id}) - P^{id} \right\|_2^2$, M is the number of identity labels. The probability in the cross-entropy loss is represented by Equation 6. During the training process, we also use the distance $d(, \cdot,)$ to update the parameters of the prototypes. As a result, the overall loss function is defined as the sum of the modified cross-entropy loss and the prototype update loss:

$$\mathcal{L}_{proto}(x; \theta, P) = \frac{1}{M \cdot N} \sum_{id=1}^{M} \sum_{i=1}^{N} (-\log(Prob(x_i)) + d(\mathcal{F}_r(x_i^{id}), P^{id})) \tag{7}$$

2.4. Irrelevant Sample Repulsion Learning

After applying self-constrained center learning and dynamic prototype learning, the sample feature distribution is confined to a more compact subspace. However, in the absence of additional open-set datasets, effectively distinguishing whether a given sample belongs to a registered identity remains a critical challenge. Inspired by adversarial reciprocal points learning [5], we propose irrelevant sample repulsion learning. The feature representation of the reciprocal points $\mathcal{P}^{id}$ can be formulated as $\mathcal{D}_L^{\neq id} \cup \mathcal{D}_U$. The reciprocal points $\mathcal{P}^{id}$ of identity label id should be as close as possible to the feature set of non-id dataset $\mathcal{D}_L^{\neq id}$ and the open dataset $\mathcal{D}_U$.

$$\max \left(\zeta \left(\mathcal{D}_L^{\neq id} \cup \mathcal{D}_U, \mathcal{P}^{id} \right) \right) \leq R. \tag{8}$$

both $\mathcal{P}^{id}$ and R are learnable parameters. By imposing a constraint on the maximum distance between the $\mathcal{P}^{id}$ and the sample features, achieving the separation of registered samples from those in the open set.

In real-world scenarios, the training process does not involve open-set data which is often characterized by an almost infinite amount of samples and categories. Given the limited amount of training data with identity labels, where features from the open-set data and labeled data are complementary, we shift the training objective from the open-set data to the labeled data. This shift allows us to effectively leverage the complementary

nature of these two types of features. The corresponding loss function is expressed as follows:

$$\mathcal{L}_o(x; \theta, \mathcal{P}^k, R^k) = \frac{1}{M \cdot N} \sum_{id=1}^{M} \sum_{i=1}^{N} \max(d_e(\mathcal{F}_r(x_i), \mathcal{P}^{id}) - R^{id}, 0),$$
(9)

where $d_e(\mathcal{F}_s(x), \mathcal{P}^k) = \frac{1}{N} \left\| \mathcal{F}_r(x_i) - \mathcal{P}^{id} \right\|_2^2$. By imposing the L_2-norm constraint on the distance R^{id} between the training samples and the $\mathcal{P}^k$, ensuring that the samples $\mathcal{D}_L^{=id}$ are distanced from those of other identity labels $\mathcal{D}_L^{\neq id}$.

Algorithm 1 SimCLR's main learning algorithm.

Require: batch size N, num of batch L, sample $x \in \mathcal{D}_L$, dynamic prototype P, reciprocal point $\mathcal{P}$, learnable margin R, structure of encoder $\mathcal{F}_s$
1: **for** all $k \in \{1, \dots, N\}$ **do**
2:      $S_k = \mathcal{F}_s(x_{id})$
3:      $C_k = \arg \min \sum_{i=1}^{n} \sum_{j \neq i}^{n} d(S_{id}^i, S_{id}^j)$
4: **end for**
5: **for** sampled minibatch $\{x_k\}_{k=1}^{N}$ **do**
6:      **for** all $k \in \{1, \dots, M\}$ **do**
7:          $S_k = \mathcal{F}_s(x_k)$
8:          $d_{self}^k = d_e(S_{id}^k, C_{id})$
9:          $d_{proto}^k = d_e(S_{id}^k, P^{id})$
10:         $d_{reciprocal}^k = d_e(S_{id}^k, \mathcal{P}^{id})$
11:      **end for**
12:      $\mathcal{L}_{self} = \frac{1}{L} \sum^{L} d_{self}$
13:      $Prob(x \in P^{id} | x) = \frac{e^{-d_{proto}^k}}{\sum^K e^{-d_{proto}^K}}$
14:      $\mathcal{L}_{proto} = \frac{1}{L} \sum^{L} (-\log(Prob(id = k | x)) + d_{proto}^k)$
15:      $\mathcal{L}_o = \frac{1}{L} \sum^{L} (\max(d_{reciprocal}^k - R, 0))$
16:      $\mathcal{L} = \alpha \mathcal{L}_{self} + \beta \mathcal{L}_{proto} + \gamma \mathcal{L}_O$
17:      update $\mathcal{F}_s, P, \mathcal{P}$ and R to minimize $\mathcal{L}$
18: **end for**
19: **return** encoder network $\mathcal{F}_s(\cdot)$

2.5. Training Process of the Proposed Method

The overall training procedure is outlined in Algorithm 1. The ECG encoder $\mathcal{F}_s$ utilized in this process is the multimodal pre-trained model from Section 2.2. Before entering the fine-tuning loop, we compute the class centers C for all identity labels. These centers are dynamically updated in each training iteration to ensure the most accurate representation of their locations. The distance between each sample and its corresponding center $d_e(x, C)$ is used as a self-constraint center loss, encouraging samples to move closer to their respective class centers. This constraint helps refine the feature representations by minimizing the discrepancy between individual samples and their centers.

Once inside the loop, the distance between each sample and all dynamic prototypes P is calculated, replacing traditional probability-based computation with a distance-based approach.

During prototype loss computation, we incorporate the sample-to-prototype distance $D_e(x, P)$, encouraging samples to move closer to their respective prototypes. Additionally, we compute the distance between samples and designated reciprocal points $d_e(x, \mathcal{P})$, enforcing separation by penalizing deviations from a learnable margin R. Finally, the three loss components are weighted and summed to the overall loss $\mathcal{L}$, updating the parameters of $\mathcal{F}_s, P, \mathcal{P}$ and R.

3. Results and discussions

3.1. Implementation Details

In our framework, the multimodal pretraining phase leverages the MIMIC-ECG dataset [7], which comprises 800,035 pairs of signal-text data. Each signal is recorded at a 500 Hz sampling rate for a duration of 10 seconds. Due to computational constraints, a subset of 100,000 high-quality samples was selected for pretraining in our experiments. For the ECG signal encoder, we employed both ResNet1D [15] and Vision Transformer (ViT) [8] models with varying hyperparameters to explore the impact of architectural choices on performance. For the text encoder, we utilized the pretrained MedCPT model [13], which is specifically tailored for medical text processing. The pretraining process was conducted over 50 epochs using the AdamW optimizer [14]. All experiments were executed on an NVIDIA 4070 GPU.

Dataset: During the fine-tuning phase for identity authentication, the datasetss were sourced from the following three repositories:

ECGID [19] contains 310 ECG recordings, obtained from 90 persons, digitized at 500 Hz with 12-bit resolution over a nominal ±10 mV range. The records were obtained from volunteers (44 men and 46 women aged from 13 to 75 years who were students, colleagues, and friends of the author).

MIT-BIH Arrhythmia [20] Database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory. The recordings were digitized at 360 Hz.

Autonomic Aging [26] which collects the high-resolution biological signals to describe the effect of healthy aging on cardiovascular regulation. The ECG data in Autonomic are recorded from 1121 healthy volunteers, which contains two different collection modes. The sampling rate of the ECG signal is 1000Hz and the length is longer than 8 minutes.

Metrics: In the experiments, two categories of evaluation metrics were employed: **Closed-set** and **Open-set** metrics. **Closed-set metrics** assess the authentication accuracy for enrolled users, focusing on scenarios where all test samples belong to known classes. We utilized a comprehensive set of metrics, including Accuracy (ACC), F1 Score, Precision, Recall, and Area Under the Curve (AUC).

Open-set metrics were designed to evaluate performance in more realistic scenarios. The test dataset contains a large number of signals from unregistered users. The Open Set Classification Rate (OSCR) is utilized to evaluate the model's balanced discrimination capability between known and unknown

Table 1: A comparative study of ECG signal encoders with different backbone architectures. The ResNet-based encoders include ResNet18, ResNet34, and ResNet50, while the ViT-based encoders consist of ViT-Tiny, ViT-Small, and ViT-Base.

dataset	backbone	Close-set evaluation					Open-set evaluation		
		ACC[%]	f1 score[%]	Precision[%]	Recall[%]	AUC[%]	OSCR[%]	FAR[%]	TNR[%]
ECGID	ResNet18	96.11	98.71	96.80	96.11	99.96	89.24	5.39	49.03
	ResNet34	96.67	98.52	97.17	96.67	99.96	89.79	5.35	54.67
	ResNet50	95.00	97.89	95.80	95.00	99.95	89.03	5.37	52.67
	ViT tiny	68.89	76.38	70.81	68.89	95.33	53.45	9.62	38.45
	ViT small	62.78	71.24	65.06	62.78	92.31	49.43	9.93	37.03
	ViT base	41.7	46.16	42.03	41.67	91.50	26.40	12.98	42.96
MITBIH	ResNet18	99.60	99.79	99.61	99.60	99.99	97.60	7.53	53.70
	ResNet34	99.43	99.79	99.45	99.43	99.99	95.16	8.09	55.69
	ResNet50	99.50	99.75	99.51	99.50	99.97	94.20	8.32	53.88
	ViT tiny	90.63	92.44	90.98	90.63	99.20	80.04	10.69	42.92
	ViT small	86.80	88.84	87.02	86.80	98.71	72.47	11.98	41.79
	ViT base	66.23	70.31	68.99	66.23	95.75	54.62	14.08	39.96
Autonomic	ResNet18	98.40	98.79	97.93	98.40	99.83	95.84	6.21	52.31
	ResNet34	98.52	98.84	98.04	95.52	99.90	94.40	6.56	29.55
	ResNet50	96.52	97.75	96.13	96.52	99.81	91.45	7.11	21.32
	ViT tiny	94.52	96.83	94.73	94.52	99.74	87.01	7.86	46.33
	ViT small	93.20	95.57	93.43	93.20	99.61	83.56	8.09	30.64
	ViT base	74.00	75.97	75.11	74.00	99.23	58.64	11.67	35.17

identity categories [6]. A threshold δ is employed to determine whether the unknown identity samples belong to the registered categories. The calculation of OSCR involves two key metrics. The Correct Classification Rate (CCR) measures the proportion of samples that are correctly classified:

$$CCR(\delta) = \frac{|\{x \in D_T^k \wedge \arg\max_k Prob(k|x) = k \wedge Prob(\hat{k}|x) \geq \delta\}|}{|D_T^k|}.$$
(10)

The False Positive Rate (FPR) assesses the proportion of unknown samples that are incorrectly classified as registered users:

$$FPR(\delta) = \frac{|\{x|x \in D_U \wedge \max_k Prob(k|x) \geq \delta\}|}{|D_U|}.$$
(11)

The OSCR is defined as the area under the curves of the CCR and the FPR across varying δ.

In addition, we chose the False Accept Rate (FAR) and True Negative Rate (TNR) as supplementary evaluation metrics. FAR measures the proportion of unregistered samples that are incorrectly authenticated as valid users, highlighting the vulnerability to false acceptance. TNR evaluates the proportion of unregistered samples correctly identified as unknown, providing insight into the robustness in distinguishing genuine from unauthorized inputs.

$$FAR(\delta) = \frac{|\{x|x \in D_U \wedge \max_k Prob(k|x) \geq \delta\}|}{|D_L| + |D_U|}.$$
(12)

3.2. Experiment of backbone of ECG Encoder

Dataset. In the backbone comparative experiments, three distinct datasets were applied, each differing in the number of

identity classes and sample sizes. Consequently, different data partitioning strategies were employed for each dataset. **ECGID** dataset, 41 available identity classes were selected, with 30 identity classes designated for enrolled users and 11 classes for the open set. Each sample was preprocessed by extracting 500 sampling points before and after the R-peak. **MIT-BIH** dataset, with 30 identity classes for enrolled users and 18 identity classes for the open set, maintaining the same preprocessing method as ECGID. **Autonomic** dataset, 100 identity classes were allocated for enrolled users and 50 identity classes for the open set, with 1000 sampling points extracted per sample.

Encoder Structure. In the experiments, different network architectures were used as the encoder for ECG signals, with representative models selected from ResNet and Vision Transformer (ViT). The ResNet-based encoders were configured with three different depths: [18, 34, 50], all employing 1D convolution (conv1D) as the primary convolutional operation. For ViT, conv1D was used to extract embedding features, and the full model consisted of 12 transformer blocks. The ViT-Tiny variant featured 3 attention heads, a multilayer perceptron (MLP) hidden dimension of 768, and an embedding dimension of 192. The ViT-Small variant had 6 attention heads, an MLP hidden dimension of 1536, and an embedding dimension of 384. The ViT-Base variant was designed with 12 attention heads, an MLP hidden dimension of 3072, and an embedding dimension of 768.

Table 1 presents the comparative results of different backbone architectures for ECG signal encoder. Overall, models utilizing the ResNet backbone outperform those based on ViT, particularly in open-set evaluation metrics. In terms of network complexity, both backbone types exhibit performance degra-

(a) Experiment with ECGID dataset   (b) Experiment with MIT-BIH dataset   (c) Experiment with Autonomic dataset
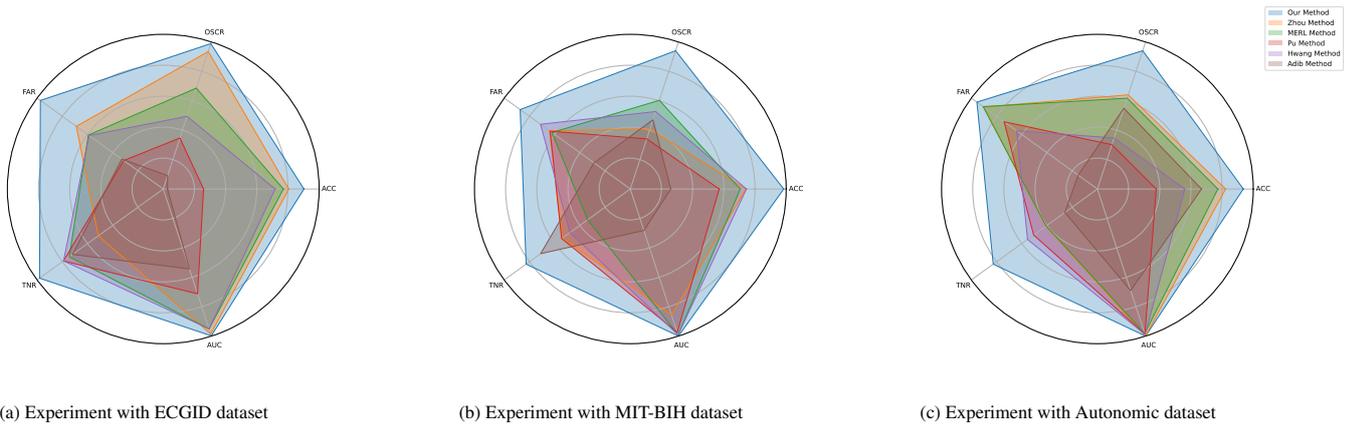
Figure 2: The experimental results comparing various baseline methods are presented using ACC, OSCR, FAR, TNR, and AUC.

dation as model complexity increases, with ViT being more significantly affected. Notably, on the ECGID dataset, the ViT-Base backbone achieves an ACC of only 41.7%. For the ResNet backbone, ResNet18 consistently achieves the best performance across most metrics on all datasets, indicating that convolutional networks are well-suited for capturing the waveform characteristics of ECG signals. In all subsequent experiments, the ResNet-based model is employed as the ECG signal encoder.

3.3. Comparative Experiment

Comparative Method. To ensure a fair comparison, the experimental settings for all baseline methods were aligned with those of the proposed approach. All selected baseline methods were evaluated on ECG classification. The Adib method [1] employs a generative adversarial network to address class imbalance, and we adopted its classification model for comparison. The Pu method [24] utilizes a highly generalizable binary neural network for classification. The Zhou method [34] leverages hard negative samples and multi-hypersphere learning to improve the capability of ECG signal encoder. The MERL method [18] integrates multimodal contrastive learning between clinical text and ECG signals, as well as unimodal contrastive learning within the ECG modality, to enhance feature extraction capabilities. The Hwang method [10] employs a ResNet-DenseNet architecture for multi-label classification tasks.

Figure 2 presents a visualization of key performance metrics across different comparative methods under three datasets. Across evaluations on three benchmark datasets, our proposed method consistently outperforms all baseline approaches across all key performance metrics. Regardless of dataset size, our method consistently achieves stable ACC. On the ECGID dataset, our method attains a TNR of 49.03% and a FAR as low as 5.39%, indicating its strong capability to filter out most unregistered samples. While the Zhou method and MERL demonstrate higher ACC compared to other comparative methods due to their well-designed classifier tailored for identity authentication. However, these two methods exhibit nearly a 20% gap in

TNR compared to our method. This suggests that these methods struggle to accurately distinguish between registered and unregistered users. Our method not only maintains a high ACC for registered users but also effectively excludes the majority of unregistered users. By leveraging clear decision boundaries, our method ensures a strong balance between accurate identity verification and open-set sample rejection.

3.4. Longer Proportion of Open-set Data

Dataset. Due to the limited number of samples in the ECGID and MIT-BIH datasets, the available open-set data is relatively scarce. To ensure a more comprehensive evaluation, we selected the Autonomic dataset for this experiment, as it provides a sufficient amount of data. Within this dataset, we selected 30 identity labels as the close-set data, while the open-set data was constructed by selecting [30, 60, 90, 120, 150, 180, 210, 240, 270, 300] identity categories as comparisons. These selections correspond to open-set to closed-set ratios ranging from 1:1 to 1:10, enabling a systematic analysis of the model's performance across varying levels of open-set complexity.

To simulate the diversity of open-set data in real-world scenarios, we constructed open-set datasets with varying proportions to evaluate the model's ability to distinguish closed-set data in the presence of large-scale open-set samples. Figure 3 presents the results obtained under the aforementioned experimental settings. The model achieves an ACC of 99.83% on the closed-set data, indicating its ability to correctly classify the vast majority of samples. Furthermore, across different open-set data proportions, the OSCR remains consistently above 95%. This demonstrates the model's robustness in maintaining high recognition accuracy even in the presence of external data interference, highlighting its effectiveness in distinguishing between known and unknown classes. As the proportion of open-set data increases, the FAR exhibits a sharp rise before stabilizing around 40%, while the TNR shows a declining trend, eventually settling at approximately 45%. This indicates that when the open-set dataset remains within a reasonable range, our method effectively identifies registered users with high accuracy. However, if the open-set data surpasses a

Table 2: Results of comparative experiments under varying proportions of open-set data. The performance is assessed using the OSCR and the FAR to reflect the model's effectiveness in distinguishing between registered and unregistered identities.

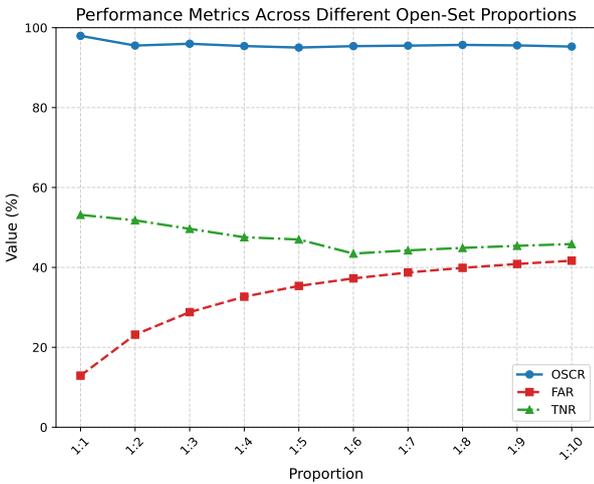| Method	Proportion									
	1:1		1:2		1:3		1:5		1:10	
	OSCR[%]	FAR[%]	OSCR[%]	FAR[%]	OSCR[%]	FAR[%]	OSCR[%]	FAR[%]	OSCR[%]	FAR[%]
Zhou	87.19	15.00	83.37	25.57	83.61	30.85	82.54	36.93	82.54	42.66
MERL	88.78	15.24	88.72	24.67	90.36	29.87	89.13	36.19	89.07	42.20
Pu	74.31	18.61	75.60	27.42	75.88	32.41	75.79	38.00	74.43	43.38
Hwang	73.87	18.36	69.37	28.66	70.15	33.36	68.16	38.91	67.69	43.87
Adib	11.57	23.81	11.10	32.89	10.87	37.39	10.95	41.58	10.80	45.50
Ours	97.96	12.92	95.53	23.17	95.98	28.80	95.03	35.36	95.27	41.68

Figure 3: Line chart illustrating the variations in OSCR, FAR, and TNR as the ratio of open-set data to close-set data changes.

certain threshold, the model may become less effective at distinguishing unregistered users, potentially leading to an increased acceptance of unauthorized samples.

Table 2 presents the experimental results comparing our method with multiple baseline approaches under varying open-set proportion settings. The evaluation metrics include OSCR and FAR, which are used to assess the model's ability to accurately recognize registered users in an open-set environment and to quantify the proportion of unregistered users mistakenly accepted. In the closed-set dataset, Zhou's method achieved an ACC of 94.56%, MERL attained 97.00%, Pu reached 92.18%, Hwang obtained 94.14%, and Adib achieved 24.00%. Our proposed method achieved an ACC of 99.83%, outperforming all other approaches and demonstrating superior recognition capability in a closed-set scenario. As the proportion of open-set data increases, all methods exhibit a rise in the FAR and a decline in the OSCR. However, across all experimental settings, our method consistently achieves the highest and most stable OSCR compared to all comparative methods. This indicates that our method effectively identifies registered users even in the presence of open-set data. Furthermore, in scenarios where

open-set data is less prevalent, our method demonstrates the lowest FAR, successfully rejecting the majority of unseen samples. These results highlight the robustness and efficiency of our method in balancing open-set recognition and false acceptance mitigation, making it particularly suitable for real-world applications where reliable user authentication is critical.

3.5. Ablation Study

Table 3 presents the results of several ablation studies. The multimodal pretraining significantly enhances the model's recognition accuracy. Additionally, the irrelevant sample repulsion learning module and the self-constraint center learning module effectively reduce the probability of open-set samples being misclassified as registered users.

In the ablation studies of each component, the softmax of classification loss in our method was computed using probabilities derived from the distance between samples and their corresponding prototypes. In contrast, in the baseline experiments without dynamic prototype learning, the standard cross-entropy loss with softmax was used to replace the distance.

As shown in Table 3, the model achieved nearly 18% higher ACC in closed-set recognition after multi-modal pre-training compared to the model without pre-training. It also contributed to worse performance in open-set metrics. The results demonstrate that the signal encoder, after multimodal pretraining, is able to capture identity-related information more effectively, thereby providing a stronger foundation for downstream identity authentication tasks.

Self-constraint center learning is a crucial module that effectively brings intra-class samples closer to the center of the corresponding identity label while reducing the dispersion of sample distributions. As a result, models without self-constraint center learning exhibit higher FAR, leading to the misclassification of some open-set samples as registered users. The effect of dynamic prototypes is similar to that of self-constraint center learning; however, the best results are achieved when both are applied simultaneously. The presence of irrelevant sample repulsion learning helps to push the registered samples further away from the dispersed open-set samples, and without irrelevant sample repulsion learning, FAR tends to increase slightly. The model achieves optimal performance only when all three sample distribution constraint methods are employed together.

(a) Close-set data in omplete model   (b) Close-set data in model only with B.1   (c) Close-set data in model only with B.2   (d) Close-set data in model only with B.3

(e) Open-set data in complete model   (f) Open-set data in model only with B.1   (g) Open-set data in model only with B.2   (h) Open-set data in model only with B.3
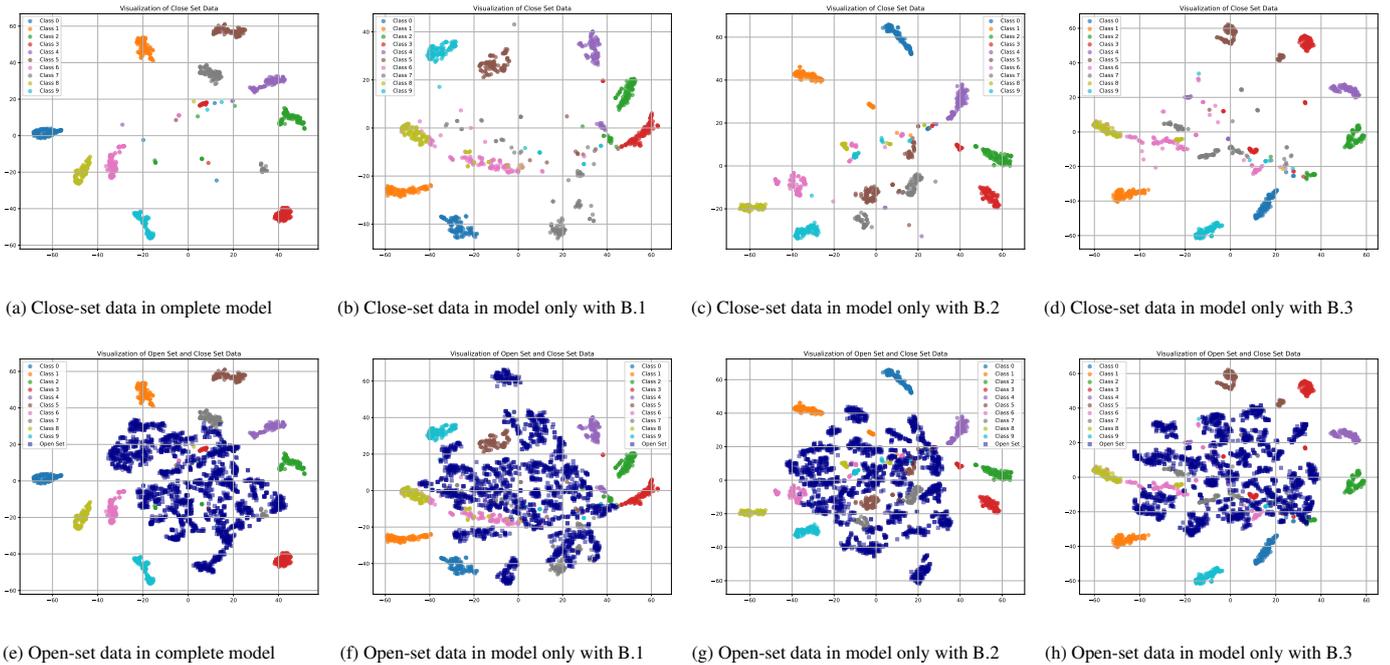
Figure 4: T-SNE visualizations of sample feature distributions under different ablation settings. Notably, (b) and (f) represent the results obtained using only the irrelevant sample repulsion learning module (only with B.1). (c) and (g) represent the results obtained using only the self-constraint center learning module (only with B.2). (d) and (h) represent the results obtained using only the dynamic prototype learning learning module (only with B.3).

Figure 4 illustrates the dimensionality-reduced feature distributions in the ablation study. To facilitate observation of the sample distributions, subfigures (4a)–(4d) depict the distributions of samples within the closed set, while subfigures (4e)–(4h) show the distributions of all samples when open-set data is included.

Subfigure (4a) and (4e) present the distributions using the complete proposed method, where the open-set data is clearly confined to a smaller area, with other close-set samples remaining relatively distant from the open-set sample distribution. In contrast, subfigures (4b) and (4f) show the outcome when only irrelevant sample repulsion learning and cross-entropy loss are used to guide model convergence. This approach also achieves a high ACC, successfully extracting the feature distribution of unknown labels, but results in a more scattered open-set sample distribution. A similar effect is observed when only the dynamic prototype method is applied, as shown in subfigures (4d) and (4h). However, when only the self-constraint method and cross-entropy loss are used, as depicted in subfigures (4c) and (4g), the model performs poorly in classification and produces a highly dispersed open-set samples distribution.

4. Summary and conclusions

This paper presents a novel ECG identity authentication method designed for open-set scenarios. The proposed method has been rigorously evaluated under varying proportions of open-set data. A key innovation of our method is the incorporation of contrastive learning with multi-modal data during pretraining, where ECG signals and text reports based on the

Table 3: Ablation study of the method. A denotes whether multi-modal pretraining is applied. B.1 denotes whether the reciprocal point is applied. B.2 denotes whether dynamic prototype is applied. B.3 denotes whether self-constraint center learning is applied.

A	B.1	B.2	B.3	ACC[%]	OSCR[%]	FAR[%]
×	✓	✓	✓	80.10	64.02	19.44
✓	✓	✓	✓	99.60	97.60	**7.53**
✓	✓	×	×	99.53	97.16	15.58
✓	×	✓	×	99.63	97.57	7.56
✓	×	×	✓	99.57	96.56	7.78
✓	✓	✓	×	99.53	97.12	15.11
✓	✓	×	✓	99.70	97.51	15.48
✓	×	✓	✓	99.13	93.68	8.40

fiducial feature are integrated to enhance the signal encoder's ability to represent ECG features comprehensively.

During the fine-tuning phase for the downstream identity authentication task, we introduce Self-constraint Center Learning, which further compacts the feature representations into a more discriminative subspace, leading to an identity recognition accuracy of 99.83%, surpassing comparative ECG classification methods. Additionally, we propose Irrelevant Sample Repulsion Learning, which effectively restricts the distribution of unseen open-set samples to a more constrained space, enabling the model to efficiently filter out unregistered identities, achieving a FAR as low as 5.39%.

Extensive experimental results demonstrate that our method maintains highly effective identity authentication performance even in the presence of large-scale open-set data, establish-

ing a new benchmark for ECG-based authentication in open-world settings. The ablation studies confirm the effectiveness of the proposed modules in enhancing identity recognition under open-set conditions. Incorporating all modules leads to a significant reduction in the FAR and a notable improvement in the OSCR.

However, current research still exhibits certain limitations. Specifically, when a large volume of open-set data is present, existing models struggle to effectively reject the majority of unregistered users. Consequently, future research efforts will focus on developing strategies to further reduce the FAR.

References

[1] Adib, E., Afghah, F., Prevost, J.J., 2022. Arrhythmia classification using cgan-augmented ecg signals, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE. pp. 1865–1872.

[2] Aslan, H.İ., Choi, C., 2024. Visgin: Visibility graph neural network on one-dimensional data for biometric authentication. Expert Systems with Applications 237, 121323.

[3] Boumbarov, O., Velchev, Y., Sokolov, S., 2009. Ecg personal identification in subspaces using radial basis neural networks, in: 2009 IEEE international workshop on intelligent data acquisition and advanced computing systems: technology and applications, IEEE. pp. 446–451.

[4] Chan, A.D., Hamdy, M.M., Badre, A., Badee, V., 2008. Wavelet distance measure for person identification using electrocardiograms. IEEE transactions on instrumentation and measurement 57, 248–253.

[5] Chen, G., Peng, P., Wang, X., Tian, Y., 2021. Adversarial reciprocal points learning for open set recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 8065–8081.

[6] Dhamija, A.R., Günther, M., Boult, T., 2018. Reducing network agnostophobia. Advances in Neural Information Processing Systems 31.

[7] Gow, B., Pollard, T., Nathanson, L.A., Johnson, A., Moody, B., Fernandes, C., Greenbaum, N., Waks, J.W., Eslami, P., Carbonati, T., et al., 2023. Mimic-iv-ecg: Diagnostic electrocardiogram matched subset. Type: dataset 6, 13–14.

[8] Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al., 2022. A survey on vision transformer. IEEE transactions on pattern analysis and machine intelligence 45, 87–110.

[9] Hoekema, R., Uijen, G.J., Van Oosterom, A., 2001. Geometrical aspects of the interindividual variability of multilead ecg recordings. IEEE Transactions on Biomedical Engineering 48, 551–559.

[10] Hwang, S., Cha, J., Heo, J., Cho, S., Park, Y., 2023. Multi-label ecg abnormality classification using a combined resnet-densenet architecture with resu blocks, in: 2023 IEEE EMBS Special Topic Conference on Data Science and Engineering in Healthcare, Medicine and Biology, IEEE. pp. 111–112.

[11] Irvine, J.M., Israel, S.A., Scruggs, W.T., Worek, W.J., 2008. eigenpulse: Robust human identification from cardiovascular function. Pattern Recognition 41, 3427–3435.

[12] Jin, J., Wang, H., Li, H., Li, J., Pan, J., Hong, S., 2025. Reading your heart: Learning ecg words and sentences via pre-training ecg language model. arXiv preprint arXiv:2502.10707 .

[13] Jin, Q., Kim, W., Chen, Q., Comeau, D.C., Yeganova, L., Wilbur, W.J., Lu, Z., 2023. Medcpt: Contrastive pre-trained transformers with large-scale pubmed search logs for zero-shot biomedical information retrieval. Bioinformatics 39, btad651.

[14] Kinga, D., Adam, J.B., et al., 2015. A method for stochastic optimization, in: International conference on learning representations (ICLR), San Diego, California;.

[15] Koonce, B., 2021. Resnet 50, in: Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization. Springer, pp. 63–72.

[16] Krishnamoorthy, L., Raju, A.S., 2024. Deep ensemble of vgg, resnet and inception for multimodal authentication system, in: 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), IEEE. pp. 1–6.

[17] Lee, S., Jeong, Y., Park, D., Yun, B.J., Park, K.H., 2018. Efficient fiducial point detection of ecg qrs complex based on polygonal approximation. Sensors 18, 4502.

[18] Liu, C., Wan, Z., Ouyang, C., Shah, A., Bai, W., Arcucci, R., 2024. Zero-shot ecg classification with multimodal learning and test-time clinical knowledge enhancement. arXiv preprint arXiv:2403.06659 .

[19] Lugovaya, T.S., 2005. Biometric human identification based on electrocardiogram. Master's thesis, Faculty of Computing Technologies and Informatics, Electrotechnical University 'LETI', Saint-Petersburg, Russian Federation .

[20] Moody, G.B., Mark, R.G., 2001. The impact of the mit-bih arrhythmia database. IEEE engineering in medicine and biology magazine 20, 45–50.

[21] Pereira, T.M., Conceição, R.C., Sencadas, V., Sebastião, R., 2023. Biometric recognition: A systematic review on electrocardiogram data acquisition methods. Sensors 23, 1507.

[22] Plataniotis, K.N., Hatzinakos, D., Lee, J.K., 2006. Ecg biometric recognition without fiducial detection, in: 2006 Biometrics symposium: Special session on research at the biometric consortium conference, IEEE. pp. 1–6.

[23] Porée, F., Kervio, G., Carrault, G., 2016. Ecg biometric analysis in different physiological recording conditions. Signal, image and video processing 10, 267–276.

[24] Pu, N., Wu, Z., Wang, A., Sun, H., Liu, Z., Liu, H., 2023. Arrhythmia classifier based on ultra-lightweight binary neural network, in: 2023 15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), IEEE. pp. 1–7.

[25] Qiang, Y., Dong, X., Liu, X., Yang, Y., Hu, F., Wang, R., 2024. Ecgmamba: Towards ecg classification with state space models, in: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE. pp. 6498–6505.

[26] Schumann, A., Bär, K.J., 2022. Autonomic aging–a dataset to quantify changes of cardiovascular autonomic function during healthy aging. Scientific Data 9, 95.

[27] Sumalatha, U., Prakasha, K.K., Prabhu, S., Nayak, V.C., 2024. Deep learning applications in ecg analysis and disease detection: An investigation study of recent advances. IEEE Access .

[28] Uwaechia, A.N., Ramli, D.A., 2021. A comprehensive survey on ecg signals as new biometric modality for human authentication: Recent advances and future challenges. IEEE Access 9, 97760–97802.

[29] Wang, G., Shanker, S., Nag, A., Lian, Y., John, D., 2024. Ecg biometric authentication using self-supervised learning for iot edge sensors. IEEE Journal of Biomedical and Health Informatics .

[30] Wang, Y., Agrafioti, F., Hatzinakos, D., Plataniotis, K.N., 2007. Analysis of human electrocardiogram for biometric recognition. EURASIP journal on Advances in Signal Processing 2008, 1–11.

[31] Wu, S.C., Hung, P.L., Swindlehurst, A.L., 2020. Ecg biometric recognition: unlinkability, irreversibility, and security. IEEE Internet of Things Journal 8, 487–500.

[32] Wu, S.C., Wei, S.Y., Chang, C.S., Swindlehurst, A.L., Chiu, J.K., 2021. A scalable open-set ecg identification system based on compressed cnns. IEEE Transactions on Neural Networks and Learning Systems 34, 4966–4980.

[33] Yang, H.M., Zhang, X.Y., Yin, F., Liu, C.L., 2018. Robust classification with convolutional prototype learning, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3474–3482.

[34] Zhou, S., Huang, X., Liu, N., Zhang, W., Zhang, Y.T., Chung, F.L., 2024. Open-world electrocardiogram classification via domain knowledge-driven contrastive learning. Neural Networks 179, 106551.