ECG Biometric Authentication on Edge Devices Using Self-Supervised Contrastive Learning

Rashi Niyas P

***Abstract*—Wearable Internet of Things (IoT) devices are increasingly used for continuous physiological monitoring, enabling passive biometric authentication. This report investigates an electrocardiogram (ECG)-based user authentication system employing a Convolutional Neural Network (CNN) for feature extraction and self-supervised contrastive learning for training. We implemented and compared different ECG segmentation methods and contrastive learning frameworks. We demonstrate that R-peak to R-peak (R2R) segmentation combined with a Siamese network and Pearson Correlation Coefficient (PCC) similarity achieves promising results. Through systematic hyperparameter tuning of the Siamese loss margin (Λ), we identified an optimal training configuration yielding an estimated Equal Error Rate (EER) of approximately 13.35% on the ECG-ID dataset. While showing significant improvement over initial methods, the results highlight the inherent challenges of within-person ECG variability and the need for further optimization to meet stringent authentication requirements. The methodology is designed with consideration for potential deployment on resource-constrained IoT edge devices.**

***Index Terms*—Electrocardiogram (ECG), Biometric Authentication, Self-Supervised Learning, Contrastive Learning, Siamese Network, R2R Segmentation, Hyperparameter Tuning, Equal Error Rate (EER), IoT**

I. Introduction

The proliferation of wearable Internet of Things (IoT) devices has opened new avenues for continuous physiological data acquisition. Beyond health monitoring, this data, particularly the electrocardiogram (ECG), offers a promising modality for passive and continuous user authentication. ECG signals are considered unique to individuals, universally available, and difficult to replicate, making them suitable candidates for biometric applications.

Traditional authentication methods, such as passwords or fingerprint scans, are typically performed only at the start of a session. Continuous authentication, enabled by wearable devices, can enhance security by verifying user identity throughout an activity, preventing unauthorized access in dynamic environments. However, implementing continuous authentication without disrupting the user experience presents challenges.

Early approaches to ECG-based authentication often relied on extracting hand-crafted fiducial (landmark-based) or non-fiducial features from the signal, followed by traditional machine learning classifiers. While demonstrating feasibility, these methods can be sensitive to noise and signal variations, may lack generalizability to unseen data or users, and often require significant manual effort in feature engineering.

More recent work has explored using Convolutional Neural Networks (CNNs) for automated feature extraction from raw ECG signals. This approach mitigates the need for manual feature definition and can potentially learn more robust representations. However, many existing CNN-based methods employ supervised training, which requires large, labeled datasets and typically necessitates retraining when new users are added to the system, limiting scalability. Furthermore, evaluation is often limited to small, specific datasets, raising questions about real-world performance and generalizability across different recording conditions. Finally, the computational cost of complex neural networks can be a barrier for deployment on resource-constrained IoT edge devices.

To address these limitations, this project investigates an ECG-based biometric authentication system leveraging a CNN for feature extraction trained with **self-supervised contrastive learning**. Contrastive learning allows the model to learn discriminative features by comparing different parts of the data (similar vs. dissimilar pairs or triplets) without requiring explicit identity labels for every training sample. This approach enhances potential scalability and generalizability. We explore and compare different ECG preprocessing strategies and contrastive learning frameworks suitable for this task. Furthermore, we systematically tune key training hyperparameters to optimize the learned feature space for authentication performance, measured primarily by the Equal Error Rate (EER). The methodology is developed with an eye towards efficiency for potential deployment on IoT edge devices.

The main contributions of this work are: - Implementation and comparison of different ECG segmentation strategies (NPD, P2T, R2R) for fixed-length input generation for CNNs. - Implementation of a CNN encoder trained using a Siamese self-supervised contrastive learning framework based on Pearson Correlation Coefficient (PCC) similarity. - Systematic hyperparameter tuning of the Siamese contrastive loss margin (Λ) to optimize authentication performance, evaluated using comprehensive FAR/FRR analysis and EER calculation. - Analysis of the performance trade-offs and limitations of the developed system on a public ECG dataset (ECG-ID) with consideration for IoT deployment constraints.

The remainder of this report is structured as follows: Section II briefly reviews relevant prior work. Section III details the proposed system methodology, including data preprocessing, the CNN encoder, the Siamese contrastive learning framework, and the authentication process. Section IV presents and discusses the experimental results, outlining the project's progress through different phases. Section V concludes the report and outlines potential future work.

II. Related Work

ECG-based biometric authentication methods can be broadly categorized by their approach to feature extraction: fiducial, non-fiducial, and CNN-based learning.

Fiducial methods rely on precisely identifying specific points in the ECG waveform (P, Q, R, S, T waves) and computing features based on their amplitudes, durations, or intervals (e.g., RR interval, QT interval). Examples include methods using wavelet decomposition of RR intervals [4] or fiducial point analysis for IoT devices [5]. While interpretable, these methods are highly sensitive to noise and variations that affect peak detection accuracy.

Non-fiducial methods extract features from broader segments of the ECG signal without explicit landmark detection, often using signal processing techniques. Autocorrelation, Discrete Wavelet Transform (DWT), and other frequency/time-frequency analyses fall into this category [6], [7]. These methods can be more robust to noise affecting specific points but may require complex preprocessing and can sometimes lose subtle discriminative information.

Recently, CNN-based approaches have gained prominence for their ability to automatically learn hierarchical feature representations directly from raw or minimally processed ECG signals. CNNs have been applied for feature extraction in supervised identification tasks [8]–[10], achieving high accuracies on specific datasets. Some works combine CNN features with traditional classifiers or distance metrics for authentication [11], [12]. However, many existing CNN-based ECG biometric studies suffer from limitations such as evaluation on small datasets, lack of demonstrated generalizability to unseen data, and being primarily framed as identification systems (unable to handle truly unknown users unless retrained). Furthermore, the computational cost of these models can be high for deployment on edge devices.

Self-supervised learning, particularly contrastive learning, has shown significant promise in learning rich, generalizable representations from large amounts of unlabeled data in various domains. Recent work has applied contrastive learning to ECG for tasks like arrhythmia classification or representation learning [13], [14]. However, applying self-supervised contrastive learning specifically for ECG *biometric authentication* to learn features that discriminate individuals for subsequent verification against a database, especially considering IoT edge deployment, remains an active area of exploration. This work focuses on this specific application and investigates the effectiveness of Siamese contrastive learning for robust and generalizable ECG feature extraction for authentication.

III. Methodology

The proposed ECG biometric authentication system follows the pipeline presented in the baseline paper [1] and consists of two main phases: a training phase utilizing self-supervised

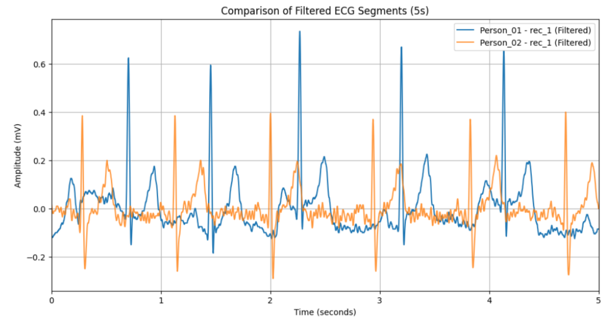

Fig. 1: Visual Comparison of Filtered ECG Segments (Person 01 vs Person 02, 5s), highlighting inter-person variability.

contrastive learning to train a CNN encoder, and an authentication phase where the trained encoder is used for user verification. The overall architecture is adapted for 1D ECG data.

Dataset We used the **ECG-ID Database** [3] for development and evaluation. This dataset is suitable for biometric tasks as it contains recordings from multiple persons and multiple recordings per person, collected over time. It comprises 310 ECG recordings from 90 unique individuals. Each recording is a 20-second Lead I ECG signal, digitized at 500 Hz with 12-bit resolution ($\pm$10mV range). Both raw and filtered signals are provided. The number of records per person varies from 2 to 20, collected over a period of up to 6 months, introducing natural within-person variability. The subjects are volunteers (44 male, 46 female) aged 13 to 75 years. The raw data contains both high and low frequency noise components.

Data Preprocessing and Segmentation Data preprocessing aims to clean the raw ECG signal and segment it into fixed-length inputs suitable for the CNN encoder. The steps are: 1. **Bandpass Filtering:** A Butterworth bandpass filter (0.5 Hz to 40 Hz cutoff) is applied to the raw signal to remove baseline wander, muscle noise, and powerline interference. 2. **Resampling:** The signal is resampled from its original 500 Hz to a target frequency of 200 Hz. This standardizes the sampling rate and reduces computational load. 3. **Segmentation:** We explored different methods for segmenting the continuous ECG signal into fixed-length inputs (1000 samples, corresponding to 5 seconds at 200 Hz). - *No Peak Detection (NPD):* Randomly extracts 1000-sample windows. Simple but segments contain a variable number of heartbeats. - *P-peak to T-peak (P2T):* Attempts to extract segments corresponding to a full cardiac cycle (P-wave onset to T-wave offset). Requires accurate P and T wave delineation. We found this method unreliable on the ECG-ID dataset, resulting in very few valid segments due to difficulty in consistent P/T detection. - *R-peak to R-peak (R2R):* Detects R-peaks using the Pan-Tompkins algorithm [2]. Extracts the signal between consecutive R-peaks (R-R intervals). Filters out excessively long intervals (indicating missed peaks, e.g., ¿ 400 samples at 200 Hz). Each valid R-R interval is resampled to a fixed size (200 samples). Finally, 5 consecutive resampled R-R

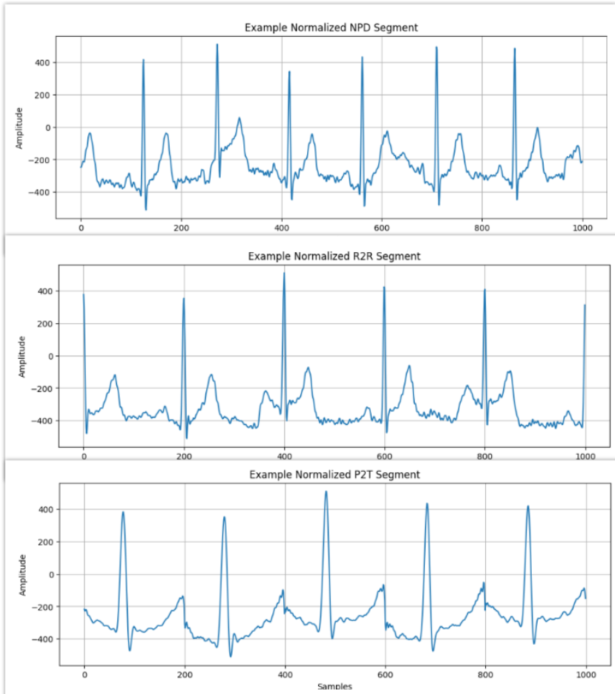

Fig. 2: Illustration of different ECG segmentation methods: (a) No Peak Detection (NPD), (b) R-peak to R-peak (R2R), (c) P-peak to T-peak (P2T). Figure based on [1].

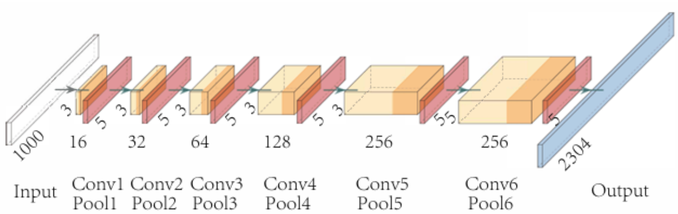

Fig. 3: Architecture of the 1D CNN Encoder.

pieces are concatenated to form a 1000-sample segment. This method aligns segments to heartbeats while being more robust than P2T. We adopted R2R segmentation for its balance of biological relevance and robustness on the dataset. Our R2R processing yielded 6372 segments from the 90 individuals.

4. **Normalization:** Each 1000-sample segment is normalized to the range [-512, 512] as in the baseline paper [1].

CNN Encoder Architecture A 1D Convolutional Neural Network (CNN) is used as the feature encoder to extract discriminative features from the preprocessed 1000-sample ECG segments. The architecture is based on Figure 4 of the baseline paper [1], consisting of six 1D convolutional layers with ReLU activation and Batch Normalization, interleaved with MaxPooling layers for downsampling. A final flatten layer is followed by a dense layer that outputs a 2034-dimensional feature vector. The model has approximately 8.3 million parameters.

Contrastive Learning Frameworks We investigate two self-supervised contrastive learning frameworks for training the CNN encoder: Triplet and Siamese networks. Both learn feature representations by enforcing that similar samples are closer than dissimilar samples in the embedding space. The distance/similarity metric used between feature vectors in both frameworks is the Pearson Correlation Coefficient (PCC).

Triplet Framework (Phase 2 Initial Implementation) In the Triplet framework, the training objective is based on triplets of samples: an Anchor, a Positive (similar to Anchor), and a Negative (dissimilar to Anchor). The goal is to train the encoder such that the distance between the Anchor and Positive embeddings is smaller than the distance between the Anchor and Negative embeddings by a margin Λ ($D(A,P) < D(A,N) - \Lambda$). We implemented a custom generator creating dynamic triplets and a Triplet Loss function using PCC and the margin Λ. This framework was used in our initial implementation phase (Phase 2) with NPD segmentation.

Siamese Framework (Phase 3 Implementation) In the Siamese framework, the training objective is based on pairs of samples: a Positive Pair (similar) or a Negative Pair (dissimilar). Two identical CNN encoders share weights and process the pair. The loss function encourages high similarity for positive pairs and low similarity for negative pairs, typically using a margin Λ. We implemented a generator for Siamese pairs and a custom Siamese contrastive loss function based on PCC similarity and Λ. This framework was adopted in Phase 3, combined with R2R segmentation.

Training Setup The R2R processed data (6372 segments) was split into reproducible training ($\sim$90%, 5734 segments) and validation ($\sim$10%, 638 segments) sets at the segment level using a fixed random state (42) for reproducibility. The CNN encoder was trained using the Adam optimizer (learning rate 0.0005) by fitting the contrastive learning network (Triplet or Siamese) for 100 or 200 epochs, monitoring performance on the validation set. Training loss convergence was observed for both Triplet and Siamese frameworks.

Authentication Process Once the CNN encoder is trained, it is used for authentication: 1. **Registration:** For each registered user, a set of their ECG segments is processed through the trained CNN encoder to obtain corresponding feature vectors. These vectors are stored in a database. Our evaluation setup used a reproducible split of 60 registered persons. Database templates were formed using 15 segments per registered user (900 DB segments total). 2. **Authentication Request:** A new ECG segment from a user attempting to authenticate is processed through the *same* trained CNN encoder to get its feature vector. 3. **Similarity Calculation:** The Pearson Correlation Coefficient (PCC) is calculated between the test segment's feature vector and *each* stored feature vector in the database. The maximum PCC score obtained against any registered user's template is taken. 4. **Decision:** If the maximum similarity score is greater than or equal to a predefined authentication threshold (Λ_{auth}), the user is authenticated. Otherwise, they are rejected.

Evaluation Methodology To evaluate the authentication performance, we split the 90 unique persons in the full dataset into a reproducible set of 60 registered persons and 30 unregistered persons using the same fixed random state (42). This person split was consistent across all evaluations. The database was built using the first 15 R2R segments from each of the 60 registered persons. The test set consisted of the remaining R2R segments from the 60 registered persons (genuine attempts) and all R2R segments from the 30 unregistered persons (impostor attempts). Total test segments were 5472 (2849 genuine, 2623 impostor).

Performance was assessed using standard biometric metrics: Accuracy, False Acceptance Rate (FAR), False Rejection Rate (FRR), Precision, Recall (TPR), and F1 Score. Crucially, we analyzed the trade-off between FAR and FRR across a range of authentication thresholds (Λ_{auth}) by plotting the FAR and FRR curves. The Equal Error Rate (EER), where FAR $\approx$ FRR, was estimated from the intersection of these curves.

IV. Results and Discussion

Phase 1: Initial Exploration Results Our initial exploration using simple template matching based on cross-correlation confirmed that while ECG exhibits inter-person differences (fig. 1), within-person variability due to noise, activity, or posture makes simple static templates insufficient for reliable authentication across different segments of the same user. A simple cross-correlation between a segment from Person 01 and Person 02 resulted in a low score (≈ 0.54), but similar low scores could occur for segments from the same individual under different conditions, leading to high False Rejection Rates. This highlighted the necessity of learning robust, invariant feature representations, motivating the use of advanced learning approaches.

Phase 2: NPD Segmentation and Triplet Framework Evaluation In Phase 2, we focused on implementing the baseline paper's approach using NPD segmentation and the Triplet contrastive learning framework (section III). Training with NPD segments showed initial promise in loss reduction. However, initial authentication evaluations revealed high error rates.

Using an evaluation setup with 50 registered / 20 impostor subjects and 5 averaged templates per user, at $\Lambda_{\text{auth}} = 0.95$: Accuracy was 0.6319, FAR was 0.2122, and FRR was 0.4488. These high error rates (21% FAR, 45% FRR) indicated significant overlap in score distributions.

A further evaluation using the refined 60 registered / 30 unregistered person split (10 DB segments/user, 5772 test segments) at $\Lambda_{\text{auth}} = 0.75$ resulted in: Accuracy=0.7032, FAR=0.5100, FRR=0.2021. The extremely high FAR (51%) at this lower threshold confirmed that NPD segmentation with the initial Triplet parameters did not yield sufficient separation of genuine and impostor score distributions. This poor performance, particularly the high FAR, highlighted the need to explore alternative preprocessing (segmentation) and contrastive learning frameworks.

Phase 3: R2R Segmentation and Siamese Framework Implementation and Tuning In Phase 3, motivated by the high error rates in Phase 2, we transitioned to using **R2R Segmentation** (section III) and the **Siamese Contrastive Learning** framework (section III). R2R segmentation proved more robust on the ECG-ID dataset than P2T and was hypothesized to better align segments to individual heartbeat morphology than NPD.

TABLE I: EER Results for Different Siamese Training Margins (Λ), R2R Segmentation, 100 Epochs

Training Λ	EER Threshold (Λ_{auth})	EER Rate
0.50	$\approx$ 0.9042	0.1365
0.60	$\approx$ 0.9147	0.1545
0.40	$\approx$ 0.9061	0.1566
0.20	$\approx$ 0.9268	0.1593
0.30	$\approx$ 0.9244	0.1629
0.80	$\approx$ 0.9510	0.1738
0.70	$\approx$ 0.9437	0.1773
0.95	$\approx$ 0.9848	0.1817
0.90	$\approx$ 0.9715	0.1818
0.10	$\approx$ 0.9773	0.2413
0.05	$\approx$ 0.9954	0.3127
0.01	$\approx$ 0.9993	0.3298

TABLE II: Metrics for Selected Models at Estimated EER Thresholds

Model Config	Auth Threshold	Accuracy	FAR	FRR
Siamese R2R, $\Lambda = 0.50$	$\approx$ 0.9049	0.8573	0.1441	0.1415
Siamese R2R, $\Lambda = 0.40$	$\approx$ 0.9061	0.8637	0.0987	0.1709

We implemented the Siamese framework with the custom PCC-based loss, which allows for a tunable margin Λ. Initial training of the Siamese R2R model for 100 epochs with a loss margin $\Lambda = 0.7$ showed promising convergence. Evaluation of this model at an authentication threshold $\Lambda_{\text{auth}} = 0.93$ (using the consistent 60 Reg / 30 Unreg, 15 DB segments/user split) yielded: Accuracy=0.8401, FAR=0.1475, FRR=0.1713. This demonstrated improved performance compared to the NPD+Triplet results, but the balanced error rate (EER $\approx 16\%$) was still relatively high.

To optimize this balanced performance, we systematically tuned the training margin Λ in the Siamese loss function. We trained separate Siamese R2R models (100 epochs each) for various Λ values ranging from 0.01 to 0.95 and calculated the FAR/FRR curve and EER for each.

The tuning results confirm that $\Lambda = 0.50$ yields the lowest EER rate (approximately 13.65%) among the tested margins. This signifies that training with a margin of 0.50 resulted in the best balance between FAR and FRR for this setup after 100 epochs.

Evaluating the specific performance metrics for models trained with the best performing margins at their respective EER thresholds:

The results in Table II confirm the EERs and show the specific trade-offs. The $\Lambda = 0.50$ model provides a balanced

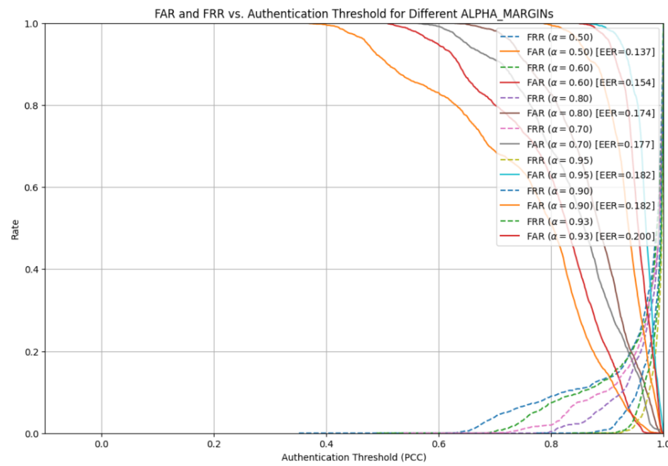

Fig. 4: Comparison of FAR and FRR Curves vs. Authentication Threshold for different Siamese training margins (Λ). The optimal $\Lambda = 0.50$ shows the lowest EER.

TABLE III: Comparison with State-of-the-Art

Method	Features	Framework	Dataset	Metric
Tantawi et al. [4]	Wavelet (Fiducial)	RBF NN	PTBDB (290 subj)	Acc: 97.7%
Hammad et al. [8]	CNN	QG-MSVM	PTBDB (290 subj)	Acc: 98.66%
Sepahvand et al. [11]	CNN	Distance	PTBDB (290 subj)	Acc: 99%
Hazratifard et al. [12]	CNN	Distance	PTBDB, ECGIDDB	Acc: 96.8%, 93.6%
Wang et al. [1]	CNN	Triplet CL + PCC	ECGIDDB (90 subj)	Acc: 98.77%
Our Work (Siamese R2R)	CNN (Learned)	Siamese CL + PCC	ECGIDDB (90 subj)	EER: 13.65 & 0.8637%

error rate around 14%, while the $\Lambda = 0.40$ model offers a lower FAR ($\approx 9.9\%$) but higher FRR ($\approx 17.1\%$) at its EER point.

Comparison with Existing Work Our implementation achieves an EER of approximately 13.65% on the ECG-ID dataset using R2R segmentation and Siamese contrastive learning tuned to $\Lambda = 0.50$. Below is a comparison with select recent works:

Limitations The achieved EER of $\approx 13.65\%$ is still relatively high for many real-world authentication systems, which often target EERs below 5% or even 1%. This indicates that further performance improvements are necessary.

V. Conclusion

This report presented an investigation into ECG biometric authentication using a CNN encoder trained with self-supervised Siamese contrastive learning. The resulting model, trained with $\Lambda = 0.50$ for 100 epochs, achieves an EER of approximately 13.65% on the ECG-ID dataset. Future work will focus on improving performance by training for longer durations and exploring data augmentation techniques.

References

[1] G. Wang *et al.*, "ECG Biometric Authentication Using Self-Supervised Learning for IoT Edge Sensors," *IEEE Trans. Biomed. Circuits Syst.*, vol. XX, no. Y, pp. ZZ-AA, 2024.

[2] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[3] T. S. Lugovaya, "Biometric human identification based on electrocardiogram," Master's thesis, Electrotechnical University 'LETI', Saint-Petersburg, Russia, 2005.

[4] M. M. Tantawi *et al.*, "A wavelet feature extraction method for electrocardiogram (ECG)-based biometric recognition," *Signal, Image Video Process.*, vol. 9, no. 6, pp. 1271–1280, Nov. 2015.

[5] G. Yang *et al.*, "Privacy-Preserving ECG Based Active Authentication for Wearable Devices," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1234–1246, 2021.

[6] M. Hejazi *et al.*, "ECG biometric authentication using non-fiducial features," *Signal Process.*, vol. 124, pp. 135–144, Jul. 2016.

[7] F. Agrafioti and D. Hatzinakos, "Signal validation for cardiac biometrics," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2010, pp. 1734–1737.

[8] M. Hammad *et al.*, "Multimodal Biometric Authentication Systems Using Convolution Neural Network," *IEEE Access*, vol. 7, pp. 26527–26542, 2019.

[9] J. Thentu *et al.*, "ECG-Based Biometric Authentication Using Convolutional Neural Networks," *J. Med. Syst.*, vol. 45, no. 4, Apr. 2021.

[10] A. J. Prakash *et al.*, "BAED: A secured biometric authentication system using ECG signals," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022.

[11] M. Sepahvand and F. Abdali-Mohammadi, "A novel multi-lead ECG personal recognition," *Biomed. Signal Process. Control*, vol. 68, p. 102766, 2021.

[12] M. Hazratifard *et al.*, "Ensemble Siamese Network (ESN) using ECG Signals," *Sensors*, vol. 23, no. 10, p. 4727, May 2023.

[13] Z. Chen *et al.*, "CL-ECG: Contrastive Learning for ECG Representation," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2021, pp. 123–130.

[14] H. Wei *et al.*, "Contrastive Heartbeats: Contrastive Learning for Self-Supervised ECG Representation," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 3, pp. 987–996, Mar. 2022.