

Received 7 November 2023, accepted 23 November 2023, date of publication 30 November 2023,
date of current version 18 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3338191

RESEARCH ARTICLE

Advancing ECG Biometrics Through Vision Transformers: A Confidence-Driven Approach

ONORATO D'ANGELIS¹, LUCA BACCO¹, LUCA VOLLERO, (Member, IEEE),
AND MARIO MERONE¹, (Member, IEEE)

Research Unit of Computer Systems and Bioinformatics, Department of Engineering, Università Campus Bio-Medico di Roma, 00128 Rome, Italy

Corresponding author: Mario Merone (m.merone@unicampus.it)

ABSTRACT Over the past two decades, Electrocardiography (ECG) has gained significant momentum in the field of biometrics, offering a compelling alternative for person identity recognition based on physical/biological traits. Its inherent difficulty to be circumvented and its ability to enable liveness detection make it particularly appealing compared to other popular identifiers such as face, fingerprint, and iris. As a result, ECG has garnered attention from the computer vision community working on biometrics applications. We present a novel biometric method for personal recognition that leverages I-lead signals acquired off-the-person. Through fine-tuning a pre-trained Vision Transformer (ViT) model, we achieve remarkable results in recognizing individuals based on a single 2D image of their ECG recording obtained from as little as three heartbeats. Extensive evaluation on the CYBHi database, with enrollment and testing phases separated by a three-month time window, simulating a real-world, long-term identification scenario, demonstrates the robustness of our multiclass approach. Specifically, our system achieves a remarkable single sample-based identification accuracy of over 70% with a pool of 63 individuals, along with an equal error rate of only 0.48% in the 1-vs-1 authentication task. Additionally, we evaluated our approach on the very recent Heartprint database to assess the robustness of our approach with more subjects, larger separation time windows, and continuous training settings, obtaining again remarkable performance with respect to the state-of-the-art. While the promising capabilities of ECG-based biometrics are evident, given various security challenges, using such methods as standalone authentication could raise caution among users. To address this concern and enhance the system's dependability, we introduce a confidence-based rejection rule. Integrating this mechanism improves both identification and authentication performances, while it could also enable the system to detect out-of-database individuals.

INDEX TERMS Biometric recognition, identification, authentication, electrocardiogram, imposter detection, vision transformer.

I. INTRODUCTION

Electrocardiography is a well-established medical technology, extensively used for diagnose and monitor cardiovascular diseases. Since the early 2000s, electrocardiograms (ECGs), which depict the electrical activity of an individual's heart, have been recognized as suitable candidates for computer-based biometrics systems (CBBSs) even in single-lead configurations [1], [2], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenhua Guo¹.

Although physiological and behavioral traits such as fingerprints, iris, facial features, gait, and handwriting have gained popularity, "hidden" and dynamic biometrics like ECG offer significant advantages for biometric system implementation. ECGs meet the universality, uniqueness, and acceptability criteria for users, as outlined in various studies [4], [5], and provide additional unique benefits. ECG tracings enable detection of liveness and are difficult to circumvent, ensuring robustness against presentation attacks greater than other biometric features [6], [7], [8], [9], [10], [11]. Furthermore, ECGs may be less computationally

intensive to process due to their one-dimensional nature [12], and more straightforward to acquire (measurability/collectability).

Nonetheless, electrocardiography is not without limitations as a source of biometric information. For instance, collectability is better achieved using one-lead off-the-person approaches, which employ dry electrodes to collect signals from users' fingertips instead of wet electrodes attached to the skin. This method is less invasive and costly [13], which makes it more suitable for real-world applications [14], [15], [16]. However, off-the-person recordings are more prone to noise and artifacts [13], [17], necessitating a more rigorous pre-processing stage for filtering and signal cleaning.

Another challenge in electrocardiography is its intra-subject variability, which may be influenced by an individual's physical and mental state [18]. This is particularly evident when considering separate training and testing sessions [19], [20] and the time elapsed between these sessions [3], with more significant degradation observed over longer periods [21], [22], [23]. However, ECGs captured from the fingers are less affected by this issue [24], increasing the demand for off-the-person frameworks that collect finger-based signals.

In this study, we used off-the-person ECG recordings from the CYBHi and Heartprint datasets [25], [26] to introduce a novel system for both **authentication** (also known as verification) of individuals (1-vs-1) and **identification** across a group of subjects (1-vs-N) while evaluating its resilience to temporal variability. Refer to Section IV-A for an overview of the dataset and Section V for the experiments handled in this work.

Given the nature of the more common identifiers, the Computer Vision (CV) community has traditionally been at the forefront of developing and advancing biometrics tasks. Furthermore, the landscape is rapidly evolving, and recent breakthroughs in the field have introduced innovative approaches, including the revolutionary *Vision Transformer* (ViT) architecture [27], which we employ as our core model. The ViT model has previously been utilized in various tasks associated with ECG like the detection of cardiac arrhythmias [28], [29], atrial fibrillation [30], [31], and congestive heart failures [32].

In our approach, we fine-tuned a pre-trained ViT model (see Section IV-C for more details) to recognize individuals based on a single 2D image of their ECG recording, obtained by averaging only three consecutive heartbeats.

Despite achieving remarkable performances in all the various settings (see Section VI), as the other physical/biological sources, there is a general concern about the use of biometric systems as standalone authentication methods, granting access to resources to mistakenly recognized users. To address this aspect, we integrate a novel confidence-based rule into our system to allow the system to reject doubt samples and improve the reliability of the entire pipeline.

To the best of our knowledge, this study is the first to analyze model outputs to reduce false acceptance rates at the expense of rejecting more samples. Furthermore, we investigate the use of such a method in identifying individuals outside the database.

Our research thus presents the following notable practical contributions to the field of biometrics using electrocardiography and ViTs:

- This is the first study to investigate the application of ViTs for biometric systems based on electrocardiography;
- Our system surpasses state-of-the-art results for authentication on the long-term CYBHi dataset, demonstrating high reliability even after three months from the enrollment phase;
- We propose a novel approach for rejecting difficult samples by analyzing the variance of predictions, resulting in a reduced false acceptance rate;
- We investigate the use of the confidence-based rejection method as an *imposter detector*, enabling the identification of out-of-database individuals, a concept not previously explored.

II. RELATED WORKS

Over the past two decades, numerous studies have been conducted to investigate and improve the feasibility of using electrocardiograms for biometric purposes, specifically for authentication and identification. For authentication (also referred to as verification), a typical approach involves comparing the similarity of incoming sample patterns to individuals' templates or latent representations using a (dis)similarity metric, such as Euclidean or Mahalanobis distances [33], [34]. Another employed technique is the Dynamic Time Warping [35], [36], which can be applied to unsynchronized sequences of varying lengths. However, this approach necessitates storing user templates, posing potential security and privacy risks. In contrast, classifier-based approaches commonly used in identification tasks use methods such as Support Vector Machines [37], k-Nearest Neighbors [38], and Random Forests [39].

Biometric approaches can be broadly categorized into two types based on the type of features they employ: fiducial and non-fiducial. Fiducial features are based on the morphology of the signal and involve detecting specific points, i.e., the peaks of P, Q, R, S, and T waves. However, fiducial features require extensive feature engineering [40], [41], [42]. Examples of such features include amplitudes, peak ratios, time intervals between the peaks, and distances. Furthermore, they are less robust against noisy signals [12], [43]. On the other hand, non-fiducial features are obtained from signal transformations, such as discrete cosine and wavelet-based transforms [20], [44], [45], or from statistical features, such as autocorrelation [46]. Non-fiducial features offer greater robustness as they do not rely on detecting characteristic points, but instead make extensive use of higher SNR

signal projections. Moreover, with the rise of deep learning [12], [41], [47], [48], a popular approach is to use raw signals as input and allow the neural network to learn meaningful representations across its layers.

In evaluating the effectiveness of our approach using the CYBHi dataset, most previous studies have focused on either intra-session experiments or term inter-session experiments. Intra-session experiments involve obtaining training and test samples from a single session [49], [50], [51], while term inter-session experiments involve enrollment and testing on data separated by a brief period [52]. Being a completely different setup from ours, we did not include works using term setups for the comparison analysis reported in Section VI. These limitations hinder the evaluation of models in real-world scenarios, where enrollment and testing sessions may be separated by a significant amount of time. Only a few studies have investigated the long-term robustness of their systems [22], [53], [54], [55], [56], similar to our approach. However, not all of them employed all the 63 subjects in the dataset. For instance, Jyotishi and Dandapat [55] excluded two individuals from their inter-sessions tests and Lourenço et al. [57], which employed record from only 32 subjects. We excluded these works from the comparison analysis as well. In addition, some works have exploited the CYBHi dataset for assessing the detection of fiducial points and outliers [58], [59].

Concerning Heartprint, this dataset addressed the challenges of authentication and identification across both intra- and inter-session experiments, encompassing setups for both short-term and long-term durations [26]. In a subsequent work, Ammour et al. [62], as we did, converted signals into 2D images. However, they employed spectrogram images to train their system via deep contrastive learning paradigm. They evaluated their system for the identification in short-term intra- and inter-session experiments. The dataset was employed under multimodal scenarios, too, fusing ECG recordings with fingerprints signals [63] to robustness to spoofing attacks: integrating ECG features with others derived by other kinds of signal is a recent trend in ECG biometrics [64].

Regarding the core model of our system, Vision Transformers (ViTs) [27] have gained interest after the Transformers architecture was introduced in the field of Natural Language Processing [65] and later adapted to vision problems. ViTs have outperformed traditional Convolutional Neural Networks (CNNs) by utilizing self-attention mechanisms, achieving state-of-the-art performance on various visual tasks. In the field of ECG data, ViTs have been shown to improve the classification of tetanus severity [32] and congestive heart failure [66] when combined with other types of networks. However, their application to ECG-based biometric tasks has not been extensively explored. Recently, a particular version of ViTs, the Data Efficient Image Transformer (DEIT), has been employed to extract features from ECGs and fingerprints for biometrics [63].

III. MOTIVATION

In recent years, there has been a growing interest in the field of ECG biometrics, driven by the quest for more robust and efficient authentication and identification systems. ECG biometrics offers significant advantages, especially when compared to other common modalities such as fingerprints, which are more vulnerable to circumvention techniques. However, harnessing the potential of ECGs for biometric applications presents its own set of challenges. Current ECG-based approaches often require extensive feature engineering and struggle with noise and variability over time. We saw the potential of ViTs to enhance ECG-based biometrics, by directly processing the raw signals as images. ViTs excel at capturing intricate patterns and relationships in ECG signals, offering an alternative to grid-like CNNs. While ViTs have shown promise in ECG recognition tasks, their application to biometrics remains relatively unexplored. Thus, our motivation was to explore ViTs' applicability in biometrics, aiming to improve system accuracy and reliability.

Additionally, we introduce a novel confidence-based rejection rule to bolster system security and resilience against malicious or erroneous access attempts.

IV. MATERIALS AND METHODS

We propose a biometric system, illustrated in Figure 1. Due to the low signal-to-noise ratio of the dataset, we opted for a non-fiducial approach, using raw segments converted into images. However, our pipeline relies on a segmentation algorithm based on the detection of fiducial R peaks, categorizing our system as a hybrid approach. Indeed, R peaks are easily detectable even under very low signal-to-noise ratios. Specifically, the pre-processing step involves signal filtering, segmentation, and conversion of three consecutive heartbeats into images to be used as input for the model.

For the core model, we employed the Vision Transformers architecture [27] and integrated a fully-connected classifier layer comprising 63 neurons (one for each class or subject's identity) with softmax activation to normalize its outputs into a probability distribution. Additionally, we propose a novel post-processing algorithm based on the variance of the model's output to identify and, eventually, reject input heartbeats for which the model shows a low confidence, resulting in a more reliable system.

A. DATASETS

1) CYBHi

In this work, we utilized the CYBHi (Check Your Biosignals Here initiative) dataset [25] to evaluate the effectiveness of the proposed system for ECG biometrics. This dataset is publicly available¹ and was specifically collected for ECG biometrics research. The recordings were taken at a sampling frequency of 1 kHz with a resolution of 12 bit using dry Ag/AgCl electrodes. None of the participants reported any

¹<https://zenodo.org/record/2381823#.ZBMiyHbMKUI>

TABLE 1. Summary table of the literature related works. Following, the employed acronyms: kNN (k-Nearest Neighborhood), SVM (Support Vector Machine), CNN (Convolutional Neural Network), LSTM (Long Short-Term Memory), RNN (Recurrent Neural Network), CWT (Continuous Wavelet Transform), WDM (Wavelet Distance Measure), STFT (Short-Time Fourier Transform), DEIT (Data Efficient Image Transformer).

Data	Work	Task	Session setup	Features	Model
CYBHi	Lourenço et al. (2012) [57]	Authentication / Identification	Short-term	fiducial features	kNN & SVM
	Lourenço et al. (2012) [58]	Fiducial points detection	Short-term	1D raw signal	multiple algorithms combination
	Lourenço et al. (2013) [59]	Outliers detection	Short-term	1D raw signal	DMEAN
	Santos et al. (2012) [60]	Authentication	Short-term	PCA coefficients	kNN
	da Silva et al. (2013) [61]	Authentication	Inter-	1D raw signal	SVM
	da Silva et al. (2017) [53]	Authentication	Inter-	1D raw signal + 2D spectrogram	1D & 2D CNNs + Fusion rule
	Hammad et al. (2019) [52]	Authentication	Short-term	Hashed fiducial features	Feed-Forward NN
	Hammad et al. (2019) [49]	Authentication	Intra-	Fiducial features + 2 signal image	2D CNN
	Jyotishi et al. (2020) [50]	Identification	Intra-	sequences of 1D raw signal	LSTM
	Belo et al. (2020) [54]	Authentication / Identification	Intra- & Inter-	1D raw signal	Temporal CNN / RNN
	Ibtehaz et al. (2021) [22]	Identification	Inter-	1D raw signal	EDITH
Jyotishi et al. (2021) [55]	Authentication	Intra-	sequences of 1D raw signal	Attention-based hierarchical LSTM	
Zhu et al. (2022) [51]	Authentication	Intra-	1D raw signal + 2D spectrogram	1D & 2D SEResNets	
Heartprint	Islam et al. [26]	Authentication / Identification	All	1D raw signal / CWT	WDM / VGGNet CNN
	Ammour et al. [62]	Identification	Intra- & Inter- short-term	2D STFT image	EfficientNetB5
	Ammour et al. [63]	Identification	Short-term	2D STFT image + fingerprints	DEIT + CNN

health issues, allowing for the exclusion of pathological signals.

We chose to work with the CYBHi dataset, as widely considered the most suitable for testing biometric recognition systems [13]. This choice is primarily due to the specific acquisition hardware and protocols. A crucial aspect of this dataset is its incorporation of two acquisition sessions spaced over time for each subject. This unique feature enables us to thoroughly evaluate the proposed system’s capabilities in handling challenges highlighted in the literature, such as the low signal-to-noise ratio [67] resulting from the off-the-person framework and its overall reliability over time.

In particular, we employed the long-term subset, which consists of two-minute-long ECG signals recorded from 63 participants (14 males and 49 females) aged between 18 and 24 years in two sessions three months apart, namely S1 and S2.

2) HEARTPRINT

The Heartprint database [26] is a recently published repository of ECG recordings featuring data from 199 individuals (comprising 130 males and 69 females) spanning ages 18 to 68, with a predominant representation from South-Asian and Arabian ethnic groups. These ECG signals were meticulously captured at the fingertip level using the dry electrodes of the ReadMyHeart device.² Each recording (at least 2 for each session and subject) was gathered for 15 seconds with a sampling frequency of 250 Hz.

Like the CYBHi dataset, the Heartprint dataset is publicly accessible³ and encompasses multiple sessions that are spaced across various time intervals. In detail, this comprehensive dataset is composed of four distinct sessions: sessions S1 and S2, comprising data from all 199 subjects, are separated by a relatively short average interval of

²https://dailycare.en.ec21.com/ReadMyHeart_-_Handheld_ECG_Recording-976239_976240.html

³https://figshare.com/articles/dataset/Heartprint_A_Multisession_ECG_Dataset_for_Biometric_Recognition/20105354/3

47.5 days; session S3R, involving 109 subjects, was recorded during activities involving reading, with an average temporal separation of 1054.7 days from Session S1; and session S3L, featuring 78 subjects, exhibits a considerable time gap of 1572.2 days when compared to Session S1. In this work, we employed all the four sessions for our experiments.

B. PRE-PROCESSING

The pre-processing phase is a crucial step in ensuring the robustness of the biometric system, especially considering the low signal-to-noise ratio of the targeted ECG signals.

The first step in the pre-processing phase was to filter the signals using a zero-phase Butterworth bandpass filter of order 4, with cutoff frequencies at 0.5 Hz and 50 Hz. This filtering approach helped to remove baseline wander artifacts, power-line noise, and electromyographic interference that typically occur at high frequencies [41], [68], [69].

Next, the Hamilton algorithm [70] was used to identify the R peaks in the ECG signals, and, for each peak, a time window of [−200 ms, +400 ms] was considered to segment the single heartbeats. We also employed the DMEAN outlier removal algorithm [59] to detect and remove noisy R peaks.

To increase the signal-to-noise ratio (SNR), we considered three consecutive peaks with an overlap of 1 (to increase the number of samples in the training and the validation sets), we summed the peaks and divided them by the mean. Finally, we extracted 2D images from the resulting samples.

C. VISION TRANSFORMER (ViT)

Since their advent in 2017 [65], Transformers have revolutionized the field of Natural Language Processing (NLP). More recently, their architecture has been proposed in CV problems [27] to handle 2D images that are presented to the ViT model as a sequence of patches. Such patches are embedded subdivision of fixed-size of the input image. For each patch, its absolute position within the image is added to its representation.

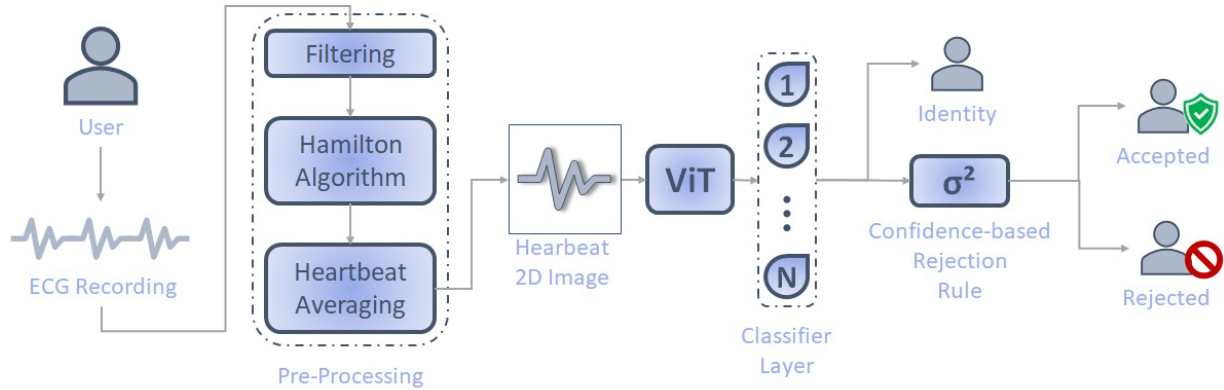


FIGURE 1. Schematic overview of the proposed biometric system.

These models are usually pre-trained to learn inner representations of the input data (i.e., texts in the NLP scope and images in the CV one). The pre-trained model, integrated with a new classifier layer, can be trained on a target task: this step is generally called fine-tuning. The advantage of this two-steps process is that, thanks to pre-training, the models can be fine-tuned on way smaller datasets achieving larger performance than training the models from scratch, often surpassing previous state-of-the-art results.

We thus exploited the ViT model presented by Dosovitskiy et al.⁴ [71], which they first pre-trained with images at resolution 224×224 from *ImageNet-21k*⁵ and then fine-tuned on a multi-class classification task with images at resolution 384×384 from *ImageNet 2012*⁶ [72]. To exploit this model in particular, we thus employed patches of resolution 32×32 .

Following the *pre-training and then fine-tuning* paradigm, we first replaced the old classifier layer to a newly initialized one. Then, together with a newly initialized fully-connected layer with softmax activation, we fine-tuned the ViT model with a cross-entropy loss to the target task, i.e., the classification of the identity of the individual presenting in input to the model their ECG-based 2D image. We trained the model for 1000 epochs, incorporating an *early stopping* mechanism monitoring the loss on the validation set, causing training to halt if the loss fails to decrease for 10 consecutive epochs. Furthermore, we incorporated a *learning rate* scheduler during training that gradually decreased the learning rate using a cosine function. More precisely, the learning rate decayed over half a period, with a maximum number of iterations set to 10. The initial learning rate was fixed at $2e-5$ and the *Adam* optimizer [73] is used.

D. CONFIDENCE-BASED REJECTION RULE

During inference, before returning the identity classification, the entire system employs a threshold mechanism to act as a rejection rule to reject samples for which the model shows low confidence. This module retrieves the probability

distribution given in output by the classifier layer (i.e., the values given by the softmax activation), and computes its variance.

The variance within the softmax values computed for the i -th test sample is thus computed with the following equation

$$\sigma_i^2 = \frac{\sum_{j=1}^N |a_j - \mu_a|^2}{N} \quad (1)$$

where a_j is the softmax activation value for j -th neuron of the classification layer, N is the number of neurons of the classification layer (i.e., the number of subjects), and μ_a is the mean value of the softmax activations from the classifier layer.

As shown in Figure 2, when the model is particularly confident in its decision (i.e., it shows a high value for one subject), the Gaussian bell results to be relatively large. On the other hand, when the model is not that confident about its decision (i.e., it shows a number of mid-to-low values for several subjects), it shows a narrower bell around the mean value, thus, a lower variance. Therefore, if the variance surpasses a predetermined threshold, the system considers the response reliable and thus returns the identity response. Otherwise, the model is considered not confident in providing a classification, which is then rejected. To determine the threshold value, the variance of each softmax distribution derived from the correctly classified samples in the *validation set*. The threshold could thus be selected from the quantiles, representing the amounts of validation samples rejected under a certain threshold. The threshold value is thus a trade-off between the performances and the percentage of rejected samples.

V. EXPERIMENTS

In our study, we conducted a series of experiments to evaluate our system from several perspectives.

A. CONFIGURATION ANALYSIS

1) FILTERING CONFIGURATION

Since different pre-processing pipelines can lead to substantially different results, we assessed our systems in two other

⁴<https://huggingface.co/google/vit-base-patch32-384>

⁵<https://image-net.org/>

⁶<https://image-net.org/challenges/LSVRC/2012/>

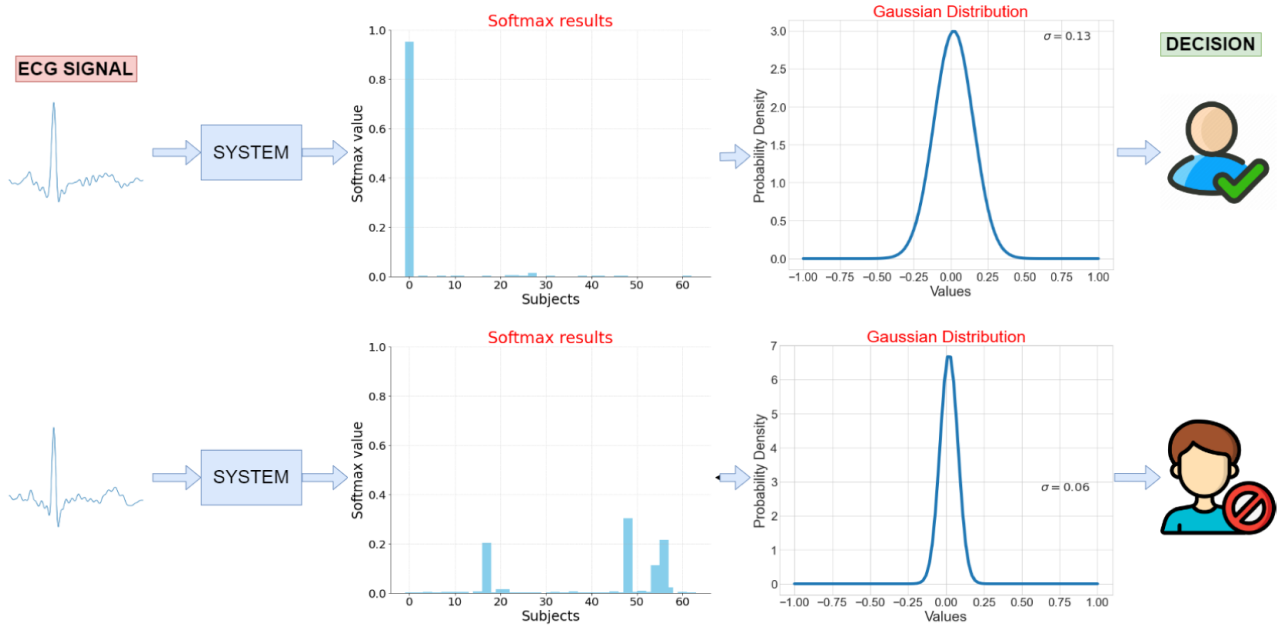


FIGURE 2. Example of the confidence-based rejection mechanism: by elaborating the ECG image, the system outputs the softmax values for each class (i.e., subject), represented by the bar plot on the center of the figure. Then, the mean and the standard deviation (and the variance) are computed. The second graph reports the Gaussian bell derived from the softmax distribution. Based on the value of the variance (i.e., the width of the bell), the system decides to either accept the decision (top) or reject it (bottom).

different filtering configurations: (i) in an attempt to remove more high-frequency noise from the electrocardiograms, we lowered the high cutoff frequency to 30 Hz from the band-pass filter described in Section IV-B; (ii) following the empirical considerations extracted from the heuristic analysis of the signal [74], we employed a 300-order band pass Finite Impulse Response (FIR) filter with Hamming window and cutoff frequencies at 5 Hz and 20 Hz.

In particular, we compared these configurations following the most realistic scenario, i.e., the inter-session setup. We employed S1 (enrollment) as the training set and S2 (test) of the CYBHi dataset to assess the performance.

2) MODEL CONFIGURATION

We conducted a comparative analysis between two distinct model architectures: the Multi-Class and the Multi-Expert Vision Transformer (MC-ViT and ME-ViT) models. The former involves a single ViT model with a sole on-top classifier layer, featuring N neurons, where N corresponds to the number of subjects in the dataset. This model is designed for identifying the individual to whom an ECG sample belongs among the N subjects in the database. The identity is determined by the subject-associated neuron n_i with the highest confidence, calculated as follows:

$$\text{Identity} = \underset{i \in \{1, \dots, N\}}{\text{argmax}} (S_i), \quad S_i = \frac{e^{n_i}}{\sum_{j=1}^N e^{n_j}} \quad (2)$$

where S_i represents the softmax value of the i -th neuron.

The ME-ViT model, on the other hand, offers an alternative approach, utilizing multiple ViTs, each specializing in recognizing ECG samples from a specific subject, framed as a binary classification task. In essence, each individual ViT serves as an expert for a single subject, trained to classify whether an ECG sample belongs to that particular subject or not. Consequently, the multi-expert system aggregates the outputs of these individual models, providing the identity associated with the expert that exhibits the highest confidence in positively classifying its respective subject as follows:

$$\text{Identity} = \underset{i \in \{1, \dots, N\}}{\text{argmax}} (S_i), \quad S_i = \frac{e^{n_1}}{e^{n_1} + e^{n_2}}, \quad n_1 > n_2 \quad (3)$$

where S_i represents the softmax value of the i -th expert model, computed for the n_1 neuron associated with the positive classification of the designated subject, where its probability is higher than the negative neuron n_2 .

One notable advantage of the ME-ViT model is its practical applicability. In scenarios necessitating the addition of a new subject to the system, the ME-ViT model only requires training a new expert dedicated to that specific subject. This stands in contrast to the MC-ViT model, which would mandate a complete retraining process for the entire system.

Both model configurations were trained on the CYBHi S2 dataset and subsequently assessed using the S1 dataset. Within the ME-ViT framework, the training data for each expert comprises not only the images of the chosen subject but also a selection of randomly sampled images from other individuals.

B. AUTHENTICATION AND IDENTIFICATION

To assess the performance of our system in a more realistic scenario, we conducted *inter-sessions* experiments. Here, data from one session (e.g., S1) was used in the enrollment phase to train the models, while data from the other session (e.g., S2) was used for testing. This allowed us to evaluate the performance of our system when a time gap of three months existed between the two sessions.

However, to provide a comprehensive comparison with previous literature, we also performed *intra-session* experiments. These experiments involved randomly splitting data from the same single session (e.g., for CYBHi, either S1 or S2) into two subsets for training/validation and testing purposes. We employed an 80-10-10 fixed training-validation-test split ratio for each subject.

In the authentication setup, we aimed to determine if a new input sample (i.e., a new image) belongs to a specific subject by testing the ability of the model to discriminate whether the model recognizes the individual or confuses them with someone else. To be consistent with previous literature, we evaluated the models using the Equal Error Rate (EER), a commonly used metric in biometric applications, particularly for verification. The EER is determined as the point on the Receiver Operating Characteristic (ROC) curve at which the False Acceptance and Rejection Rates (FAR and FRR, respectively) are equal. The FAR represents the rate at which the system incorrectly accepts an imposter, while the FRR indicates the rate at which the system incorrectly rejects an authorized user. The EER serves as an overall performance metric for biometric systems, indicating when the system is equally likely to accept an imposter as it is to reject a genuine user. A lower EER indicates better performance.

In the identification task, we evaluated the ability of our system to classify the identity of a new input sample among all 63 individuals in the database. To evaluate the system's performance, we used accuracy, a popular metric in the literature measuring the percentage of correctly classified samples out of the entire test set.

For both tasks, to have a fair comparison with the state of the art, we assessed the performance of our system both with and without the proposed rejection rule.

C. IMPOSTER DETECTION

In this setting, we assessed the overall system's capability, which encompasses both the model and the rejection rule presented in Section IV-D, to discern individuals who are not part of the database. To accomplish this, we conducted the following procedure: we randomly selected 13 subjects and excluded them from the training set. Subsequently, we trained the model using the remaining 50 subjects and assessed its performance using data from the 13 excluded subjects. To mitigate any potential bias introduced by the random selection process, we repeated this procedure five times and reported the average performance. Performance was measured in terms of the detection rate, representing the

percentage of data samples correctly identified by the machine learning model as belonging to the imposter class.

Within this framework, the imposter detection mechanism operates based on the confidence exhibited by the model's output. Specifically, if the variance falls below a predefined threshold, the system categorizes the sample as not belonging to any of the subjects in the database. Conversely, if the variance surpasses the threshold, the sample is considered part of the database. For establishing the threshold value, we computed the variance for each softmax distribution derived from correctly classified samples in the *validation set*. We then utilized this list of variances to determine the threshold by calculating the quantile, which represents the proportion of validation samples rejected under that threshold. Our analysis also involved assessing the system's behavior under multiple thresholds.

VI. RESULTS AND DISCUSSION

A. CONFIGURATIONS ANALYSIS

1) FILTERING CONFIGURATION

Regarding the filtering configuration, Table 2 reports the performance in terms of the identification accuracy. As can be noted, the system employing the filtering configuration with the frequency band of 0.5-50 Hz resulted in the highest accuracy performance. It thus seems to indicate that considering frequency information above 30 Hz may contain relevant information for identifying individuals through ECG signals.

TABLE 2. Results, in terms of identification accuracy (%), obtained by our model in the three different filtering configurations. The experiments were conducted in an inter-session setup, training on S1 (enrollment) and evaluating the performance on S2 (test) of the CYBHi dataset.

Filtering setting	Accuracy (%)
Butter-worth (0.5 - 50 Hz)	68.00
Butter-worth (0.5 - 30 Hz)	60.00
Finite Impulse Response (5 - 20 Hz)	58.00

2) MODEL CONFIGURATION

Regarding the model configuration, the multi-class model outperformed the multi-expert one, both for the authentication task (0.48% vs 0.58% of EER) and the identification one (70% vs 64% of accuracy).

Furthermore, for the authentication task, in the case of the multi-expert model, since the identity is known, there is also the possibility to have the sample analyzed by the corresponding expert only and then to evaluate only the response of the single expert. In this additional configuration, we obtained an average EER of 6.65% with a standard deviation of 8.97%. It implies that analyzing the combination of all the experts is more advantageous than using a single expert. In particular, we attribute the relatively high EER to the fact that the single experts tend more toward non-authentication of test samples, resulting in being more accurate in recognizing when an image does not belong to

their related subject (precision of $99.8\% \pm 27.8\%$) as opposed to the vice versa ($36.5\% \pm 24.6\%$).

In general, these results seem to suggest that exploiting the knowledge acquired from all the subjects helps the ViT model.

B. AUTHENTICATION AND IDENTIFICATION

1) RESULTS WITHOUT THE REJECTION RULE

To perform a fair comparison with the literature, we analyzed the results of only the MC-ViT model, i.e., without the rejection rule. Several works in the literature have presented their performance results using different datasets like the PTB dataset [75] and the FANTASIA dataset [76]. However, a key aspect worth noting is that these datasets typically consist of only one recording session per subject. Recent works, including ours, show high performance for intra-session recognition tasks, i.e., tasks where the data used for training and testing come from the same recording. In addition, to provide a comprehensive analysis, we include intra-session results in our study, but we primarily direct the reader's attention to inter-session testing, as this is the main focus of our work. By evaluating the performance of our model across recording sessions, we intend to assess its ability to generalize to new data belonging to recording sessions other than the one used in training, which is critical for real-world applications.

Table 3 provides an overview of the authentication results achieved by our model on the CYBHi dataset in comparison to prior research efforts. It is important to note that not all studies conducted experiments across all four combinations of intra- and inter-session scenarios. Specifically, while Zhu et al. [51] demonstrated commendable performance in the S1 vs S1 setup, they did not furnish results for the inter-session settings, preventing us from directly comparing their system in those scenarios. However, for the inter-session setup, our model yielded remarkable results, boasting an EER of 0.51% for S1 vs S2 and 0.48% for S2 vs S1, a substantial improvement over previous results, notably surpassing the outcomes reported by da Silva et al. [53] and Jyotishi and Dandapat [55]. Concerning intra-session performance, our model achieved results on par with the previous state-of-the-art work by Belo et al. [54], exhibiting minimal or no errors. However, it is crucial to highlight a significant advantage of our system: it operates with just three heartbeats, whereas their approach requires over ten seconds of data acquisition. This streamlined experimental setup aligns more closely with practical scenarios, where extended acquisition times may not be desirable and could potentially hinder user acceptance of the biometric device.

Table 4 provides the outcomes in the identification task on the CYBHi dataset. It is essential to acknowledge that not all research works have reported results for all four experimental setup combinations. Specifically, Jyotishi et al. [50] and Zhu et al. [51] limited their reporting to the S1 vs S1 scenario, while Ibtehaz et al. [22] exclusively presented inter-session results. Within the intra-session settings, our model

TABLE 3. Results of our approach and comparison with the literature for the authentication task on the CYBHi dataset. The scores are reported in terms of equal error rate in two different experimental setups: intra-session (i.e., S1 vs S1 and S2 vs S2) and inter-sessions (i.e., S1 vs S2 and S2 vs S1), in which the first term refers to the training session and the second one refer to the testing session. We also report the number of heartbeats (hb) or seconds (s) employed.

Work	Equal Error Rate (%)				Hb / s
	S1 vs S1	S2 vs S2	S1 vs S2	S2 vs S1	
(2013) [61]	-	-	9.10	9.37	5 hb
(2017) [53]	1.33	-	12.78	13.93	1 hb
(2019) [49]	4.47 (90:10)		-	-	1 hb
(2020) [54]	0.2	0.0	2.2	4.1	>10 s
(2021) [55]	3.98	-	13.43	12.05	2 s
(2022) [51]	0.86	-	-	-	1.4 s
MC-ViT	0.0	0.0	0.51	0.48	3 hb

TABLE 4. Results of our approach and comparison with the literature for the identification task on the CYBHi dataset. The scores are reported in terms of identification accuracy in two different experimental setup: intra-session (i.e., S1 vs S1 and S2 vs S2) and inter-sessions (i.e., S1 vs S2 and S2 vs S1), in which the first term refer to the training session and the second one refer to the testing session. We also report the number of heartbeats (hb) or seconds (s) employed.

Work	Identification Accuracy (%)				Hb / s
	S1 vs S1	S2 vs S2	S1 vs S2	S2 vs S1	
(2020) [50]	79.37		-	-	2 s
(2020) [54]	100 (50:50)	100 (50:50)	60.30	61.00	>10 s
(2021) [22]	-	-	71.60	73.93	1 hb
(2022) [51]	98.5 (90:10)	-	-	-	1.4 s
MC-ViT	99.00 (90:10)	99.00 (90:10)	68.00	70.00	3 hb

achieved 99% of accuracy, outperforming, in general, previous research endeavors. Only Belo et al. [54] reported a higher accuracy. However, it is important to highlight that their system necessitates over ten seconds for the identification process. For the inter-session tests, our system attained an accuracy of 68% in the S2 vs S1 scenario and 70% in the S2 vs S1 scenario. While these results are marginally below those reported by Ibtehaz et al. [22], comprehensive comparison between our systems cannot be drawn as they did not furnish performance metrics in terms of EER.

Finally, in Table 5, we present the results of our MC-ViT approach on the Heartprint dataset across three distinct experimental setups. As for the CYBHi dataset, the first two sets of settings pertain to intra- and inter-sessions experiments, encompassing all four available sessions. The third setup involved training the model using two sessions that were relatively close in time (S1 + S2) and testing it on a third session that was considerably separated in time (either S3R or S3L). In conducting these experiments, we encountered the challenge of dealing with underrepresented subjects, for whom the database provided significantly fewer samples compared to CYBHi. Given the well-known data-hungry nature of Vision Transformer (ViT) models, lacking translation equivariance and locality of the CNNs [27], we excluded subjects with ten or fewer samples (after removing outliers), resulting in 103, 104, 102, and 74 subjects for S1, S2, S3R, and S3L, respectively. Although our results may not be

TABLE 5. Results of our approach for both the identification and authentication tasks on the *Heartprint* dataset. The scores are reported in terms of identification accuracy in three different experimental setups: intra-session (top block), inter-sessions (middle block), and inter-sessions with continuous training (bottom block), in which the first term(s) refer to the training session(s) and the second one refer to the testing session.

Sessions	Id Accuracy (%)	EER (%)
S1 vs S1	98.01	0.00
S2 vs S2	99.35	0.00
S3R vs S3R	94.59	7.62
S3L vs S3L	95.61	4.83
S1 vs S2	49.02	9.03
S2 vs S1	52.68	8.47
S1 vs S3R	53.45	8.93
S2 vs S3R	67.94	7.62
S1 vs S3L	48.71	11.66
S2 vs S3L	56.90	12.17
S1 + S2 vs S3R	61.42	5.75
S1 + S2 vs S3L	51.60	8.18

directly comparable to those presented in previous literature, they shed light on several facets of our approach and the dataset.

Consistent with the results obtained on CYBHi, the intra-session experiments demonstrated performances on par with or even surpassing the state-of-the-art [26], [62], achieving approximately 98% identification accuracy and an EER of around 5%. In the case of inter-sessions experiments, our system reported results that were less favorable than those on CYBHi, but it exhibited significantly improved EER compared to the work of Islam et al. [26]. The EER values transitioned from approximately 16% when testing on S1 and S2 to around 50% when testing on S3R and S3L, down to approximately 9% and 12%, respectively.

It is noteworthy that settings involving training on S2 generally outperformed those involving training on S1. This can be attributed to the fact that S3R and S3L are closer in time to S2 than to S1. However, it is essential to clarify that this assumption may have been influenced by the selection of different subjects for S1 and S2. Yet, the hypothesis still applies to the third set of results, in which testing on S3R (closer in time to S1 and S2) yielded better outcomes than testing on S3L. Interestingly, combining S1 and S2 for training resulted in improved performance compared to using only S1 but lagged behind using only S2. This seems to point more to the direction of possible problems with the signals of the first session. To confirm that, training and testing on either S1 or S2 produced worse outcomes than testing on either S3R or S3L.

2) RESULTS WITH THE REJECTION RULE

Figure 3 illustrates the impact that changing the threshold (based on the quantiles) for the rejection option has on our whole system: it is evident that increasing the threshold enhances the performance of the model at the cost of reducing the number of classified samples. For the CYBHi dataset (S2 vs S1, Figure 3a), the maximum accuracy,

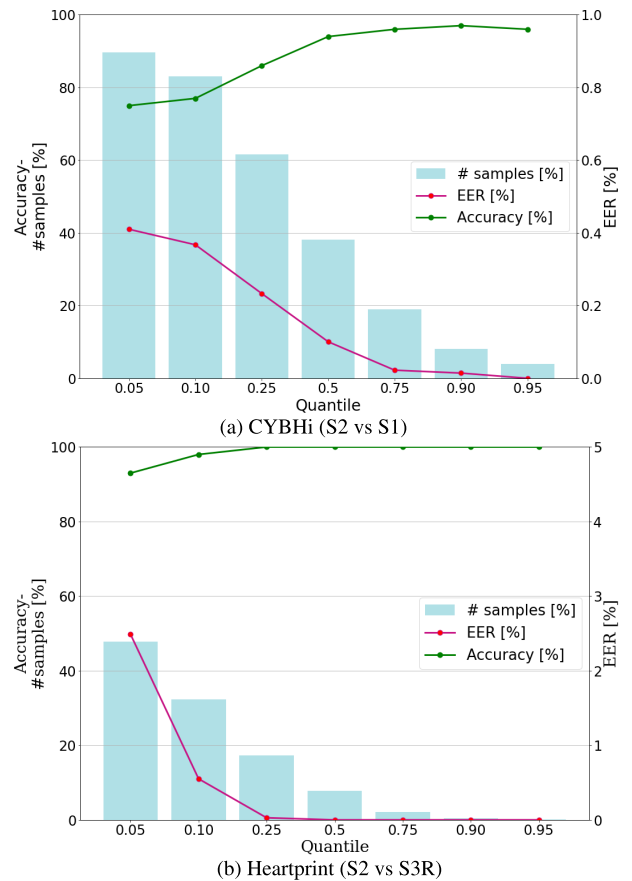


FIGURE 3. The plot displays the results obtained at varying rejection rule tolerance. The X-axis denotes the threshold quantile, the Y-axis on the left represents the percentage of the Accuracy (green line) for the identification task, and the percentage of the accepted samples (light blue bars), while the Y-axis on the right represents the percentage of the EER metric for the authentication task (red line).

at 97%, is achieved with a threshold of 0.90, but only 8% of the samples are classified. On the other hand, the lowest threshold at 0.05 accepts 89.7% of the samples with an accuracy of 75%. Therefore, selecting the appropriate threshold depends on the performance tolerance intended for the biometric system as well as the tolerated rejection rate. In the case of the Heartprint dataset (Figure 3b), we observed that the maximum accuracy, a perfect 100%, is achieved when using a relatively high threshold of 0.25. However, it is important to note that this threshold results in the classification of only 17% of the samples. Additionally, the lowest threshold at 0.05 accepts 48% of the samples with an accuracy of 93%.

This behavior can be attributed to the unique characteristics of the dataset. Specifically, due to the limited sample size in the second session, the validation set comprises only a few images for each subject. Consequently, the confidence-based algorithm tends to overfit on this small number of samples. In such scenarios, where additional samples from the recordings are not readily available, we recommend considering the use of tolerance at lower quantiles as a potential solution.

C. IMPOSTER DETECTION PERFORMANCE

Figure 4 reports the average performance obtained in the 5-fold sessions for the imposter detection presented in Section V-C, and the related standard deviation, for each quantile threshold. For the CYBHi dataset (S2 vs S1), the plot shows that a favorable trade-off is achieved when the threshold is set to 0.5, which corresponds to the median value: in this case, the imposter detection rate achieves 91.4%. By choosing a higher threshold, the system achieves nearly 100% in terms of imposter detection rate.

Regarding the Heartprint dataset, we employed the discarded subjects (the ones with ≤ 10 samples in the training set) as imposters. For all the chosen quantile thresholds, the system achieved a detection rate of 100%. Again, the algorithm seems to overfit on the small-sized validation set. Clearly, the system is more restrictive and, therefore, more secure, but, at the same time, it increases the false rejection rate. Once more, the choice of the threshold should be based on a trade-off between being more restrictive or permissive.

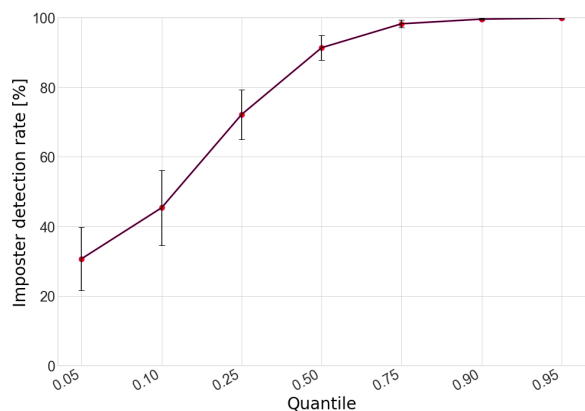


FIGURE 4. The plot displays the results: the X-axis denotes the threshold quantile, and the Y-axis represents the imposter detection rate over the quantile thresholds. For each quantile threshold, the error bar is also reported.

VII. CONCLUSION

To our knowledge, this is the first study investigating the application of ViTs for biometric systems based on electrocardiography.

Our system outperforms state-of-the-art results for authentication on the long-term CYBHi dataset, demonstrating high reliability even after three months from the enrollment phase.

We propose a confidence-based rejection method to reject difficult samples by analyzing the variance of predictions, resulting in a reduced false acceptance rate. Finally investigate the use of the confidence-based rejection method as an imposter detector, enabling the identification of out-of-database individuals, a concept not previously explored.

ACKNOWLEDGMENT

(Onorato d'Angelis and Luca Bacco contributed equally to this work.)

AUTHORS CONTRIBUTIONS

Onorato d'Angelis: Methodology, Software, Data Curation, Investigation, Formal analysis, and Writing—Original Draft; **Luca Bacco**: Methodology, Software, Formal Analysis, Visualization, Investigation, Validation, and Writing—Original Draft; **Luca Vollero**: Methodology, Supervision, Validation, and Writing—Review and Editing; **Mario Merone**: Conceptualization, Methodology, Formal Analysis, Writing—Original Draft, Supervision, and Project Administration.

CONFLICT OF INTEREST/COMPETING INTERESTS

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

DATA AVAILABILITY

In this work, we utilized the Check Your Biosignals Here initiative (CYBHi) and the Heartprint datasets [25], [26], publicly available on <https://zenodo.org/record/2381823#.ZBMiyHbMKUI> and https://figshare.com/articles/dataset/Heartprint_A_Multisession_ECG_Dataset_for_Biometric_Recognition/20105354/3, respectively.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] 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.
- [2] T. W. Shen, W. J. Tompkins, and Y. H. Hu, "One-lead ECG for identity verification," in *Proc. 2nd Joint 24th Annu. Conf. Annu. Fall Meeting Biomed. Eng. Soc.* [Eng. Med. Biol.], vol. 1, 2002, pp. 62–63.
- [3] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbaugh, "ECG biometric recognition: A comparative analysis," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1812–1824, Dec. 2012.
- [4] E. L. van den Broek, "Beyond biometrics," *Proc. Comput. Sci.*, vol. 1, no. 1, pp. 2511–2519, May 2010.
- [5] A. S. Rathore, Z. Li, W. Zhu, Z. Jin, and W. Xu, "A survey on heart biometrics," *ACM Comput. Surv.*, vol. 53, no. 6, pp. 1–38, Nov. 2021.
- [6] R. Tolosana, M. Gomez-Barrero, C. Busch, and J. Ortega-Garcia, "Biometric presentation attack detection: Beyond the visible spectrum," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1261–1275, 2020.
- [7] A. Czajka and K. W. Bowyer, "Presentation attack detection for iris recognition: An assessment of the state-of-the-art," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 1–35, 2018.
- [8] A. B. Singh and R. Rani, "Iris biometric presentation attack: Types and detection techniques—A review," in *Soft Computing: Theories and Applications*. Singapore: Springer, 2022, pp. 415–426.
- [9] S. Volkova, "Attacks on facial biometrics systems: An overview," in *Proc. Comput. Appl. Manag. Sustain. Develop. Prod. Ind. (CMSD)*, vol. 12251, 2022, pp. 20–25.
- [10] C. Shen, S. Yu, J. Wang, G. Q. Huang, and L. Wang, "A comprehensive survey on deep gait recognition: Algorithms, datasets and challenges," 2022, *arXiv:2206.13732*.
- [11] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Presentation attacks in signature biometrics: Types and introduction to attack detection," in *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection*. Cham, Switzerland: Springer, 2019, pp. 439–453.

- [12] J. R. Pinto, J. S. Cardoso, and A. Lourenço, "Evolution, current challenges, and future possibilities in ECG biometrics," *IEEE Access*, vol. 6, pp. 34746–34776, 2018.
- [13] M. Merone, P. Soda, M. Sansone, and C. Sansone, "ECG databases for biometric systems: A systematic review," *Expert Syst. Appl.*, vol. 67, pp. 189–202, Jan. 2017.
- [14] A. Bansal and R. Joshi, "Portable out-of-hospital electrocardiography: A review of current technologies," *J. Arrhythmia*, vol. 34, no. 2, pp. 129–138, Apr. 2018.
- [15] H. T. Haverkamp, S. O. Fosse, and P. Schuster, "Accuracy and usability of single-lead ECG from smartphones—A clinical study," *Indian Pacing Electrophysiol. J.*, vol. 19, no. 4, pp. 145–149, Jul. 2019.
- [16] E. Merdjanovska and A. Rashkovska, "Comprehensive survey of computational ECG analysis: Databases, methods and applications," *Expert Syst. Appl.*, vol. 203, Oct. 2022, Art. no. 117206.
- [17] M. Ingale, R. Cordeiro, S. Thentu, Y. Park, and N. Karimian, "ECG biometric authentication: A comparative analysis," *IEEE Access*, vol. 8, pp. 117853–117866, 2020.
- [18] S. Pouryayevali, S. Wahabi, S. Hari, and D. Hatzinakos, "On establishing evaluation standards for ECG biometrics," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 3774–3778.
- [19] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos, "On evaluating ECG biometric systems: Session-dependence and body posture," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 11, pp. 2002–2013, Nov. 2014.
- [20] R. Sorvillo, L. Bacco, M. Merone, A. Zompanti, M. Santonico, G. Pennazza, and G. Iannello, "Single beat ECG-based identification system: Development and robustness test in different working conditions," in *Proc. IEEE Int. Workshop Metrol. Ind. IoT (MetroInd4.0&IoT)*, Jun. 2021, pp. 538–543.
- [21] F. Porée, G. Kervio, and G. Carrault, "ECG biometric analysis in different physiological recording conditions," *Signal, Image Video Process.*, vol. 10, no. 2, pp. 267–276, Feb. 2016.
- [22] N. Ibtihaz, 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.
- [23] 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, p. 3446, Apr. 2022.
- [24] M. S. Ramos, J. M. Carvalho, A. J. Pinho, and S. Brás, "On the impact of the data acquisition protocol on ECG biometric identification," *Sensors*, vol. 21, no. 14, p. 4645, Jul. 2021.
- [25] 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.
- [26] 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, p. 141, Oct. 2022.
- [27] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [28] Y. Dong, M. Zhang, L. Qiu, L. Wang, and Y. Yu, "An arrhythmia classification model based on vision transformer with deformable attention," *Micromachines*, vol. 14, no. 6, p. 1155, May 2023.
- [29] P. Bing, Y. Liu, W. Liu, J. Zhou, and L. Zhu, "Electrocardiogram classification using TSSST-based spectrogram and ConViT," *Frontiers Cardiovascular Med.*, vol. 9, Oct. 2022, Art. no. 983543.
- [30] S. Pratiher, A. Srivastava, Y. B. Priyatha, N. Ghosh, and A. Patra, "A dilated residual vision transformer for atrial fibrillation detection from stacked time-frequency ECG representations," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 1121–1125.
- [31] H. Manda, S. Dash, and R. K. Tripathy, "Time-frequency domain modified vision transformer model for detection of atrial fibrillation using multi-lead ECG signals," in *Proc. Nat. Conf. Commun. (NCC)*, Feb. 2023, pp. 1–5.
- [32] T. Liu, Y. Si, W. Yang, J. Huang, Y. Yu, G. Zhang, and R. Zhou, "Inter-patient congestive heart failure detection using ECG-convolution-vision transformer network," *Sensors*, vol. 22, no. 9, p. 3283, Apr. 2022.
- [33] Y. N. Singh and S. K. Singh, "Evaluation of electrocardiogram for biometric authentication," *J. Inf. Secur.*, vol. 3, no. 1, pp. 39–48, 2012.
- [34] J. R. Pinto and J. S. Cardoso, "An end-to-end convolutional neural network for ECG-based biometric authentication," in *Proc. IEEE 10th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Sep. 2019, pp. 1–8.
- [35] V. N. and S. Jayaraman, "Human electrocardiogram for biometrics using DTW and FLDA," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3838–3841.
- [36] Y. Chu, H. Shen, and K. Huang, "ECG authentication method based on parallel multi-scale one-dimensional residual network with center and margin loss," *IEEE Access*, vol. 7, pp. 51598–51607, 2019.
- [37] H.-S. Choi, B. Lee, and S. Yoon, "Biometric authentication using noisy electrocardiograms acquired by mobile sensors," *IEEE Access*, vol. 4, pp. 1266–1273, 2016.
- [38] M. N. Dar, M. U. Akram, A. Usman, and S. A. Khan, "ECG biometric identification for general population using multiresolution analysis of DWT based features," in *Proc. 2nd Int. Conf. Inf. Secur. Cyber Forensics (InfoSec)*, Nov. 2015, pp. 5–10.
- [39] N. Belgacem, A. Nait-Ali, R. Fournier, and F. Berekci-Reguig, "ECG based human authentication using wavelets and random forests," *Int. J. Cryptogr. Inf. Secur.*, vol. 2, no. 2, pp. 1–11, 2012.
- [40] Q. Zhang and D. Zhou, "Deep arm/ear-ECG image learning for highly wearable biometric human identification," *Ann. Biomed. Eng.*, vol. 46, no. 1, pp. 122–134, Jan. 2018.
- [41] A. N. Uwaechia and D. A. Ramli, "A comprehensive survey on ECG signals as new biometric modality for human authentication: Recent advances and future challenges," *IEEE Access*, vol. 9, pp. 97760–97802, 2021.
- [42] T. M. C. Pereira, R. C. Conceição, V. Sencadas, and R. Sebastião, "Biometric recognition: A systematic review on electrocardiogram data acquisition methods," *Sensors*, vol. 23, no. 3, p. 1507, Jan. 2023.
- [43] J. M. Irvine, S. A. Israel, W. Todd Scruggs, and W. J. Worek, "EigenPulse: Robust human identification from cardiovascular function," *Pattern Recognit.*, vol. 41, no. 11, pp. 3427–3435, Nov. 2008.
- [44] J. Pinto, J. Cardoso, A. Lourenço, and C. Carreiras, "Towards a continuous biometric system based on ECG signals acquired on the steering wheel," *Sensors*, vol. 17, no. 10, p. 2228, Sep. 2017.
- [45] E. Kalai Zaghoulani, A. Benzina, and R. Attia, "ECG based authentication for e-healthcare systems: Towards a secured ECG features transmission," in *Proc. 13th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2017, pp. 1777–1783.
- [46] M. Komeili, W. Louis, N. Armanfard, and D. Hatzinakos, "Feature selection for nonstationary data: Application to human recognition using medical biometrics," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1446–1459, May 2018.
- [47] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," *Comput. Biol. Med.*, vol. 122, Jul. 2020, Art. no. 103801.
- [48] N. Musa, A. Y. Gital, N. Aljojo, H. Chiroma, K. S. Adewole, H. A. Mojeed, N. Faruk, A. Abdulkarim, I. Emmanuel, Y. Y. Folawiyi, J. A. Ogunmodede, A. A. Oloyede, L. A. Olawoyin, I. A. Sikiru, and I. Katb, "A systematic review and meta-data analysis on the applications of deep learning in electrocardiogram," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 7, pp. 9677–9750, Jul. 2023.
- [49] M. Hammad, S. Zhang, and K. Wang, "A novel two-dimensional ECG feature extraction and classification algorithm based on convolution neural network for human authentication," *Future Gener. Comput. Syst.*, vol. 101, pp. 180–196, Dec. 2019.
- [50] 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.
- [51] G. Zhu, M. Ma, Y. Huang, K. Wang, and G. Yang, "Dual-domain low-rank fusion deep metric learning for off-the-person ECG biometrics," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 2914–2918.
- [52] M. Hammad, G. Luo, and K. Wang, "Cancelable biometric authentication system based on ECG," *Multimedia Tools Appl.*, vol. 78, no. 2, pp. 1857–1887, Jan. 2019.
- [53] E. J. da Silva Luz, G. J. P. Moreira, L. S. Oliveira, W. R. Schwartz, and D. Menotti, "Learning deep off-the-person heart biometrics representations," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1258–1270, May 2018.

- [54] 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, p. 4078, Jul. 2020.
- [55] D. Jyotishi and S. Dandapat, "An attention based hierarchical LSTM architecture for ecg biometric system," *TechRxiv*, Jun. 2021.
- [56] P. Melzi, R. Tolosana, and R. Vera-Rodriguez, "ECG biometric recognition: Review, system proposal, and benchmark evaluation," *IEEE Access*, vol. 11, pp. 15555–15566, 2023.
- [57] A. Lourenço, H. Silva, and A. Fred, "ECG-based biometrics: A real time classification approach," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Sep. 2012, pp. 1–6.
- [58] A. Lourenço, H. Silva, P. Leite, R. Lourenço, and A. L. N. Fred, "Real time electrocardiogram segmentation for finger based ECG biometrics," in *Proc. Biosignals*, 2012, pp. 49–54.
- [59] A. Lourenço, H. Silva, C. Carreiras, and A. Fred, "Outlier detection in non-intrusive ECG biometric system," in *Proc. 10th Int. Conf. Image Anal. Recognit. (ICIAR)*, Póvoa do Varzim, Portugal. Berlin, Germany: Springer, Jun. 2013, pp. 43–52.
- [60] M. S. Santos, A. L. N. Fred, H. Silva, and A. Lourenço, "Eigen heartbeats for user identification," in *Proc. BIOSIGNALS*, 2013, pp. 351–355.
- [61] H. P. da Silva, A. Fred, A. Lourenço, and A. K. Jain, "Finger ECG signal for user authentication: Usability and performance," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2013, pp. 1–8.
- [62] N. Ammour, R. M. Jomaa, M. S. Islam, Y. Bazi, H. Alhichri, and N. Alajlan, "Deep contrastive learning-based model for ECG biometrics," *Appl. Sci.*, vol. 13, no. 5, p. 3070, Feb. 2023.
- [63] N. Ammour, Y. Bazi, and N. Alajlan, "Multimodal approach for enhancing biometric authentication," *J. Imag.*, vol. 9, no. 9, p. 168, Aug. 2023.
- [64] E. Maiorana, C. Romano, E. Schena, and C. Massaroni, "BIOWISH: Biometric recognition using wearable inertial sensors detecting heart activity," *IEEE Trans. Dependable Secure Comput.*, early access, Apr. 19, 2023, doi: [10.1109/TDSC.2023.3268360](https://doi.org/10.1109/TDSC.2023.3268360).
- [65] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2017, vol. 30.
- [66] P. Lu, C. Wang, J. Hagenah, S. Ghiasi, V. consortium, T. Zhu, L. Thwaites, and D. A. Clifton, "Improving classification of tetanus severity for patients in low-middle income countries wearing ECG sensors by using a CNN-transformer network," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 4, pp. 1340–1350, Apr. 2023.
- [67] P. Melzi, R. Tolosana, and R. Vera-Rodriguez, "ECG biometric recognition: Review, system proposal, and benchmark evaluation," *IEEE Access*, vol. 11, pp. 15555–15566, 2023.
- [68] S. Z. Fatemian and D. Hatzinakos, "A new ECG feature extractor for biometric recognition," in *Proc. 16th Int. Conf. Digit. Signal Process.*, Jul. 2009, pp. 1–6.
- [69] B. Singh, P. Singh, and S. Budhiraja, "Various approaches to minimise noises in ECG signal: A survey," in *Proc. 5th Int. Conf. Adv. Comput. Commun. Technol.*, 2015, pp. 131–137.
- [70] P. Hamilton, "Open source ECG analysis," in *Proc. Comput. Cardiol.*, 2002, pp. 101–104.
- [71] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [72] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [73] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [74] F. Canento, A. Lourenço, H. Silva, and A. L. N. Fred, "On real time ECG segmentation algorithms for biometric applications," in *Proc. BIOSIGNALS*, 2013, pp. 228–235.
- [75] R. Bousselet, D. Kreiseler, and A. Schnabel, "Nutzung der ekgsignal-datenbank cardiodat der ptb über das internet," *Biomedizinische Technik*, vol. 40, no. S1, 1995.
- [76] N. Iyengar, C. K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *Amer. J. Physiol.-Regulatory, Integrative Comparative Physiol.*, vol. 271, no. 4, pp. R1078–R1084, Oct. 1996.



ONORATO D'ANGELIS received the degree in biomedical engineering from Università Campus Bio-Medico di Roma, in 2021, where he is currently pursuing the Ph.D. degree in information technology for biomedicine. His current research interests include machine learning, deep learning, time-series analysis, computer vision, and the application of artificial intelligence to medicine.



LUCA BACCO received the Ph.D. degree in information technology for biomedicine from Università Campus Bio-Medico di Roma, in 2023. He is currently a Biomedical Engineer. His current research interests include artificial intelligence, with a focus on natural language processing and machine/deep learning for applications in healthcare.



LUCA VOLLERO (Member, IEEE) received the Ph.D. degree. He is currently an Associate Professor of computer science with Università Campus Bio-Medico di Roma. He is also the Co-Founder of the innovative startup Heremos, operating in the field of remote monitoring of frail persons and patients. His current research interests include signal processing and digital imaging, AI, distributed systems, and embedded systems optimization. He is a member of the IEEE Computer Society, the IEEE Communication Society, ACM, and SIAM.



MARIO MERONE (Member, IEEE) received the Ph.D. degree in bioengineering and bioscience, in 2017. He is currently an Assistant Professor with Università Campus Bio-Medico di Roma. He is also the Founding Partner of an innovative start-up called BPCOMEDIA S.R.L., focused on technology transfer of research results. He has extensive experience in the field of artificial intelligence (machine learning and deep learning for time-series analysis). His research has led to 40 publications, all focused on AI for healthcare. He is the scientific leader of several projects involving the application of artificial intelligence in the medical field. He is a member of the IEEE Computer Society, the IEEE Computational Life Sciences, and the IEEE Sensors Council.