Minimal Preprocessing of ECG Signals for Deep Learning-Based Biometric Systems

Zofia Mizgalewicz
Faculty of Physics
University of Warsaw
Warsaw, Poland
z.mizgalewicz@student.uw.edu.pl

Christian R. Cuenca
Dipartimento di Informatica
Università degli Studi di Milano
Milan, Italy
christian.cuenca@studenti.unimi.it

Massimo W. Rivolta
Dipartimento di Informatica
Università degli Studi di Milano
Milan, Italy
massimo.rivolta@unimi.it

Ruggero Donida Labati
Dipartimento di Informatica
Università degli Studi di Milano
Milan, Italy
ruggero.donida@unimi.it

Fabio Scotti
Dipartimento di Informatica
Università degli Studi di Milano
Milan, Italy
fabio.scotti@unimi.it

Vincenzo Piuri
Dipartimento di Informatica
Università degli Studi di Milano
Milan, Italy
vincenzo.piuri@unimi.it

Roberto Sassi
Dipartimento di Informatica
Università degli Studi di Milano
Milan, Italy
roberto.sassi@unimi.it

***Abstract*—In recent years, subject identification through electrocardiograms (ECGs) broaden the possibilities of existing biometric systems. In this study, we proposed a novel ECG-based biometric identification method designed to be computationally inexpensive, while being sufficiently accurate for a broad variety of application contexts. Specifically, we adapted an established deep learning model known as Deep-ECG to process raw ECG data with minimal preprocessing. We examined the robustness of the model by investigating the identification accuracy across three experiments, obtaining results comparable to more complex state-of-the-art methods. For all experiments, we utilized the SHAREE dataset, containing 24h Holter recordings form 139 subjects, collected in uncontrolled conditions and trained the network by randomly selecting ECG segments during daytime. In the first experiment, we quantified the performance by varying the number of subjects to identify and the number of ECG leads concurrently fed in input. In the second experiment, we varied the number of training samples per individual and the duration of the ECG segments. In the third experiment, we reimplemented the original pipeline of the Deep-ECG model to compare the performance with the new approach with minimal preprocessing. We obtained that the new approach achieved similar performance to the original Deep-ECG model. Also, the new approach obtained accuracies $> 80\%$ for individual leads and $> 90\%$ for multiple leads when using ECG segments of 2 seconds. Using this ECG duration, the minimal number of training samples per individual to achieve an accuracy $> 80\%$ was 100. Our study showed that the computational cost of the Deep-ECG model could significantly be improved by changing the pipeline previously proposed with another one with minimal preprocessing. The source code replicating the results of this study is available on GitHub.**

***Index Terms*—ECG, Biometrics, Deep Learning**

I. Introduction

Biometric recognition certainly brings advantages over the traditional methods of authentication (token-based or password-based) in physical systems (airport gates or public transportation turnstiles) or digital services [1]. Widely used methods are fingerprints, face recognition, iris recognition, but recent studies claim the possibility of using electrical signals from the beating heart as a method of recognition, with the electrocardiogram (ECG) used as a proxy to such electrical trait. The advantages of this approach are manifold: I) "Liveness detection": an ECG signal can only be detected by individuals who are alive; II) "High security": ECG-based biometric is particularly secure since it is difficult to artificially reproduce a heartbeat; and III) "Combined information": an ECG can give information about the identity as well as subject's health [1].

Unlike other recognition methods, fingerprint recognition for example, ECG-based biometric systems have reached no consensus for the standardization of the acquisition phase. The literature proposed multiple acquisition modalities by varying the ECG device (on/off-person) [2], [3], duration of the ECG signals [4], number of ECG leads [5], posture of the subject [6], *etc*. Moreover, several are the attempts to put together databases to fairly assess the performance of these biometric systems under different acquisition settings. For example, Pouryayevali *et al.* proposed a private database with a large number of subjects, acquired in different sessions and with different postures [6].

As in other fields, deep learning (DL) has emerged as a valuable ally to enable ECG-based biometric systems to reach significant performance. For example, the very first convolutional neural networks introduced in this field achieved remarkable identification rates [3], [7]. Both methodologies preprocessed the ECG prior subject identification by means of DL models. Specifically, the method proposed by Chu *et al.* [7] first detected the timings of each heartbeat in a given ECG segment, and then concatenated two randomly picked beats. While methods based on DL showed remarkable accuracy for a wide set of applications, most studies in the literature are based on very deep neural networks, which are computationally expensive and difficult to be integrated into wearable and edge-devices. In our previous study instead [3], we hypothesized that most of the personal traits would lie within the QRS complex, and the model we proposed, *i.e.*, named Deep-ECG, processed 8 concatenated QRS complexes. This model consisted in a relatively efficient Convolutional Neural Network (CNN). The QRS complex is known to be the most stable ECG wave over time (with very little dependency on heart rate too), and considering 8 of them would make the identification more robust. However, processing only QRS complexes completely removes other subject's characteristics such as depolarization of the atria (P-wave) and repolarization of the ventricles (T-wave). In addition, both methods (and others more recent, *e.g.*, [8]) did not leverage the potential identification traits due to heart rate variability, which can potentially boost the performance [9]. Another limitation of this method consisted in the computationally expensive algorithm used to compose the set of QRS complexes to provide as input to the CNN.

In this study, we hypothesize that Deep-ECG could be revised by removing entirely the preprocessing phase that detects and concatenates the QRS complexes. The approach proposed is designed to be computationally efficient and portable in edge devices. Specifically, this approach processes the raw ECG with minimal preprocessing, thus potentially leveraging both morphological characteristics of the heartbeat as well as the heart rate variability. Another advantage of the proposed method with respect to the original Deep-ECG consists in the capability of performing an accurate biometric identification based on signals acquired for relevantly shorter time (2 s instead of 20 s), thus potentially increasing the usability of the biometric system and the possible number of application scenarios. In addition, since the cardiac electrical activity can be described by means of a 3-dimensional space (the so-called vectorcardiogram), we also hypothesize that Deep-ECG could leverage multiple leads directly in inputs (as in [8], or typically done in vectorcardiography), and then information carried by the different leads would be fused at feature representation level. Specifically, the goal of this study is the quantification of the identification accuracy that the proposed Deep-ECG with minimal preprocessing could achieve while varying i) the number of subjects to identify; ii) the number of leads concurrently put in input to the network; and iii) the time duration of the ECG segments. We finally compare the obtained results with a re-implementation of the original Deep-ECG model. Our method can accurately perform biometric identifications, thus enhancing the security [10] of medical and consumer applications.

II. Materials and Methods

A. Dataset

In the study, we utilized the SHAREE Database [11] from Physionet [12], which comprised 24 h Holter ECG recordings from 139 hypertensive patients, including 90 males and 49 females, all aged $>$ 55 years. Individual recordings had varying duration in between 16 and 24 hours and contained 3-lead ECG signals, sampled at $f_s = 128$ Hz with 8 bit precision. The ECG leads were III, V3 and V5. In these recordings, we considered the first 15 hours after ensuring that Holter devices were mostly activated at the same time ($> 90\%$ of the recordings started before 12:00 pm). In addition, we discarded the first 5 minutes since they contained artifacts which were not related to the cardiac activity. Each ECG signal was minimally preprocessed using the filters described in sec. II-B.

The dataset was used to train the DL model and test its performance. Both sets were generated by segmenting the signals. Specifically, within each signal we established a list of points equally spaced of T seconds apart. From this list, we randomly selected $2N$ points as the starting positions of the T-second-long segments. The approach ensured that selected segments did not overlap and retained all characteristics of the ECG signal. The parameter N represented the number of training and testing samples per individual.

B. Signal preprocessing

Since ECG signals are often corrupted by noise, which may hamper the performance of automated systems, a preprocessing pipeline is usually recommended to reduce the effect of such noise. In addition, the preprocessing allows to remove non-cardiac-related components from the signals, which could otherwise bias the system in leveraging ECG characteristics not relevant for the identification.

In this work, we implemented a minimal standard ECG preprocessing pipeline. It comprised the application of a third order band-pass Butterworth filter with cut-off frequencies 0.5 Hz and 40 Hz to reduce baseline wander and high-frequency noise. Additionally, we applied an IIR notch filter to eliminate power line interference at 50 Hz. Finally, some of the recordings contained not-a-number values. Here, we performed linear interpolation to impute these missing values.

Figure 1 shows an example of a 3-s segment extracted from a 3-lead ECG signal. Segments with one or multiple leads served as a direct input for the DL model.

C. Network

The Deep-ECG model consisted of six convolutional layers using Rectified Linear Units, three max-pooling layers for subsampling, three Local Response Normalization layers to mimic the natural inhibition of a biological neural network, a dropout layer, a fully connected layer, and a softmax layer.

Fig. 1: Example of preprocessed 3-lead ECG segment of 3 s. The ECG is visualized with the standard printing format of 25 mm/s. With this format, small horizontal segments represent 40 ms whereas small vertical segments indicates 0.1 mV.

For a detailed description of its architecture, please refer to [3].

The original Deep-ECG architecture was modified by adjusting the input and output layers depending on the experiments performed. Specifically, the input layer was designed to take in input multiple-lead configurations. The proposed approach was based on a feature-level fusion strategy. The input dimension was $L \times (T f_s)$ where L is the number of leads concurrently put in input to the CNN. Regarding the output, it consisted of a k-class softmax layer, where k corresponds to the number of individuals enrolled in the biometric system. The discrete numerical output of the CNN corresponded to the identification code of a single individual. We experimentally evaluated different values of L and T to detect the best configuration of the proposed method.

D. Experiments and evaluation of the performance

We performed three different experiments to quantify the performance of the network by varying several variables, including number of subjects, number and length of training samples per individual, and number of leads in input to the network.

In the first experiment (Experiment 1), we varied the number of subjects from 20 up to the full dimension of the dataset, *i.e.* 139, while keeping fixed both N and T. In addition, we repeated this experiment by varying the number of leads L. Specifically, we considered lead III, V3 and V5 individually, the pair of leads III+V3, and all three leads combined.

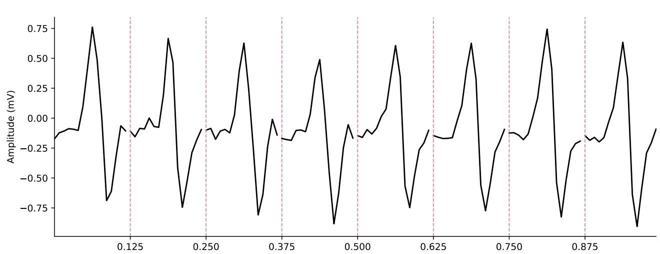


Fig. 2: Example of feature vector in input to the original implementation of the Deep-ECG model, which comprised 8 QRS complexes.

In the second experiment (Experiment 2), with all 139 subjects included, and considering the pair of leads III+V3 ($L = 2$), we varied the number N and duration T of ECG segments derived from each subjects.

Finally, in the third experiment (Experiment 3), we compared the new approach with the original strategy used to train the Deep-ECG model in [3]. For the original preprocessing, briefly, we determined the positions of R peaks in the filtered signals using a re-implementation of Pan-Tompkins algorithm [13]. Similarly to the other experiments, for each subject's ECG, we extracted 15 hours of continuous ECG recording, which we divided into 10 second segments. Then, we randomly selected $2N$ segments. From each segment, we extracted a vector of all QRS complexes by taking a time window of 0.125 s around each R point. To select the eight QRS complexes least corrupted by noise, we first estimated the correlation of each QRS complex with the average QRS pattern of the specific segment. We selected the eight QRS complexes with the highest correlation values for our feature vector. If a segment contained fewer than eight QRS complexes, we completed the feature vector by replicating the QRS complex with the highest correlation value until eight complexes were included. Segments without any R points were discarded. Figure 2 shows an example of one of the feature vectors.

The rank-1 accuracy metric was used to quantify the performance obtained in each experiment [14]. The source code to reproduce the results of this paper is available on GitHub[1].

E. Training strategy

After the dataset was divided into train and test sets using the stratification technique to maintain the distribution of sample classes proportionally between the training and test sets, the datasets were further divided into batches of 16. The train set was randomly shuffled before training the network.

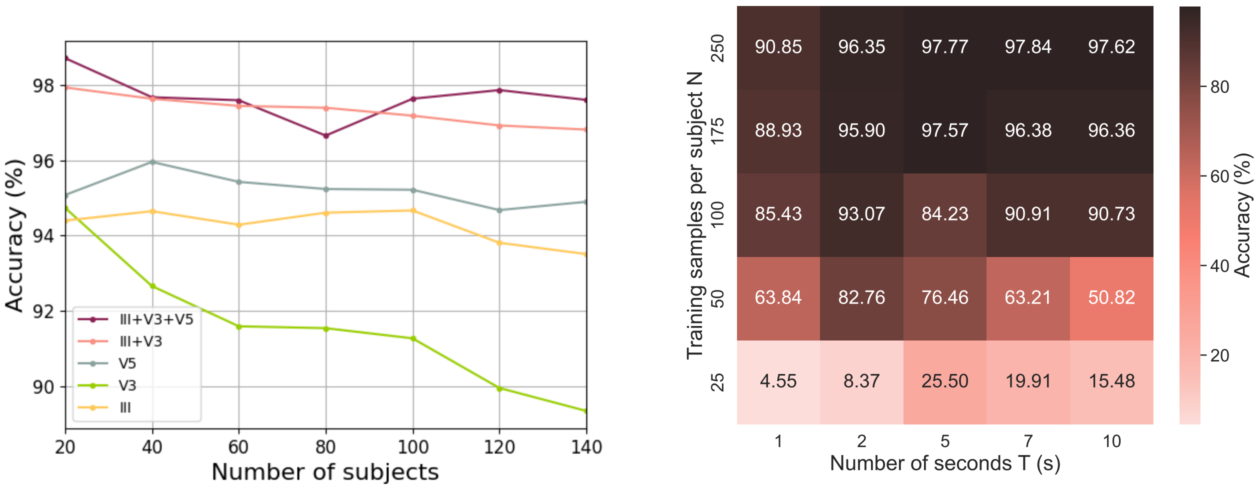
In all experiments performed, the neural network was trained for 200 epochs minimizing the cross-entropy loss by means of stochastic gradient descent with learning rate of 0.01.

III. Results

A. Experiment 1: Varying number of subjects

Experiment 1 investigated the identification accuracy of the Deep-ECG network with varying number of subjects and a fixed number and length of training samples. For this experiment, we chose $N = 250$ training samples per individual and $T = 2$ s for both training and testing. Figure 3a shows the achieved accuracy for different ECG lead configurations. The results indicated a decrease in accuracy as the number of subjects increased, particularly for configurations using only one ECG lead. Despite this trend, the accuracy remained

[1] https://github.com/mizgii/DeepECG



(a) (b)

Fig. 3: Experiment 1 (a): Accuracy obtained by the model while varying the number of patients to identify from 20 to 139 and considering different combinations of leads (V3, III, V5, V3+III, V3+III+V5), $N = 250$ training samples per individual, $T = 2$ second each. Experiment 2 (b): varying number of training samples per individual N from 25 to 250 and ECG segment length T from 1 to 10 seconds, while considering the pair of leads III+V3.

relatively high across all configurations. The use of multiple leads generally resulted in better accuracy, highlighting the benefit of multi-dimensional cardiac activity representation.

B. Experiment 2: Varying segment parameters

Experiment 2 investigated the changes in accuracy across different ECG duration T from 1 to 10 seconds and number of training samples N from 25 to 250 per patient. Figure 3b shows the identification accuracy achieved for different pairs of parameters N and T. The overall trend showed an improvement in accuracy with longer ECG duration and more training samples. However, the increase in accuracy was not linear, and there were instances where longer ECGs did not necessarily result in significantly higher performance, suggesting a more complex relationship between sample duration, number of training samples, and identification capability.

C. Experiment 3: Comparison with original Deep-ECG

In Experiment 3, we trained the Deep-ECG model using the original preprocessing phase proposed in [3] for each lead individually. We used $L = 1$ and $N = 250$. The accuracies obtained were 95.20%, 95.56% and 96.56% for lead V3, III and V5, respectively. The results showed that Deep-ECG, as proposed in this paper with minimal preprocessing, could achieve similar performance to the original model.

D. Computational time

We executed the experiments using a PC with Intel (R) Core Xeon (R) W-2135 CPU @ 3.70GHz with 32 GB RAM, and an NVIDIA - TITAN Xp. We implemented all methods using Python and Pytorch.

The longest training time was approximately 1600 s, which was obtained in Experiment 2 when using 10 s ECGs with two leads. For this case, the inference time was approximately 3 s for 250 ECG segments.

IV. Discussion and Conclusion

In this study, we extended the original Deep-ECG model by changing the input layer to deal with ECGs minimally preprocessed. The minimal preprocessing phase reduced drastically the computational cost since neither heartbeat detection nor QRS complex selection were necessary. The performance achieved with the three experiments in place indicated that removing the preprocessing was feasible.

The results of Experiment 1 showed a trend already known in the literature [1]: when the number of subjects to be identified increased, the performance dropped (Fig. 3a). This trend was observed with larger dataset as well. For example, Carreiras *et al.* investigated this phenomenon on a dataset including 618 subjects, while achieving a good identification rate ($\approx 85\%$) [15]. Similarly, the drop obtained in our experiment was not severe and the new model managed to identify subjects accurately.

Another trend confirmed with Experiment 1 was the dependency of the performance to the number of leads included in the model [1], [8]. The highest recognition rates were obtained when using all three leads concurrently (Fig. 3a). It is worth mentioning that the lead setting used to collect the ECGs in this database approximated the three orthogonal leads commonly employed in vectorcardiography. Therefore, the information carried by the three individual leads corresponded to (approximately) independent views of the cardiac electrical activity, which could potentially boost the identification performance. Our results corroborated this hypothesis, but with minimal improvement with respect to using two leads only. Finally, regarding the optimal individual leads to use, Dong *et al.* [16] achieved the highest performance on the frontal plane with III and V1, while the less effective was V6. The results of Jekova *et al.* [17] were instead more similar to ours with the sagital plane more relevant (*i.e.*, V5) and V3 less effective. We believe that the optimal lead configuration still need to be identified.

In general, the more training samples per individual are provided during training, the higher is the identification accuracy. In Experiment 2, we investigated the relationship between the number of training samples per individual, the ECG duration and the performance, concluding that about 100 one-second windows were necessary for a sufficient recognition rate for such uncontrolled setting (Holter recordings). This is equivalent to about a minute and a half of recording but collected random during daylight. Moreover, when using about three to five minutes of registration, we achieved significantly better performance. Whether by increasing the duration of segments or increasing to the number of segments (Fig. 3b), there was an improvement in accuracy. This result was in line with the study of Ramos *et al.*, who found that collecting longer ECG segments brought better results only up to a certain time duration [4].

The training strategy employed in this study presented both advantages and disadvantages. On the one hand, the random selection of ECG segments during the day would make the Deep-ECG model more robust to noise (*e.g.*, subject's movements) and could be used as a strategy in edge devices. On the other hand, the enrollment phase of the subject would last significantly longer. With respect to this aspect, the major limitation of the study was the impossibility to test the temporal stability of the method. Indeed, the database contained only a single recording per subject, thus preventing the re-assessment of the performance in a second moment.

In conclusion, the obtained results suggested that the proposed method can effectively be applied in real time applications and could be efficiently implemented in wearable and edge devices

ACKNOWLEDGMENT

This work was supported in part by project MUSA – Multilayered Urban Sustainability Action - project, funded by the European Union - NextGenerationEU, under the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment Line 1.5: Strengthening of research structures and creation of R&D "innovation ecosystems", set up of "territorial leaders in R&D". This work was also supported in part by project SERICS (PE00000014) under the MUR NRRP funded by the EU - NextGenerationEU. This work was also supported in part by Key Digital Technologies Joint Undertaking (KDT JU) in EdgeAI "Edge AI Technologies for Optimised Performance Embedded Processing" project, grant agreement No 101097300. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the Italian MUR. Neither the European Union nor Italian MUR can be held responsible for them.

REFERENCES

[1] T. M. 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, 2023.

[2] M. S. Islam and N. Alajlan, "Biometric template extraction from a heartbeat signal captured from fingers," *Multimedia Tools and Applications*, vol. 76, no. 10, p. 12709–12733, 2016.

[3] R. Donida Labati, E. Muñoz, V. Piuri, R. Sassi, and F. Scotti, "Deep-ECG: Convolutional neural networks for ECG biometric recognition," *Pattern Recognition Letters*, vol. 126, pp. 78–85, 2019.

[4] 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, 2021.

[5] F. Porée, G. Kervio, and G. Carrault, "ECG biometric analysis in different physiological recording conditions," *SIViP*, vol. 10, no. 2, p. 267–276, 2014.

[6] S. Pouryayevali, S. Wahabi, S. Hari, and D. Hatzinakos, "On establishing evaluation standards for ECG biometrics," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014.

[7] 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, p. 51598–51607, 2019.

[8] P. Melzi, R. Tolosana, and R. Vera-Rodriguez, "ECG biometric recognition: Review, system proposal, and benchmark evaluation," *IEEE Access*, 2023.

[9] N. Akhter, H. Gite, G. Rabbani, and K. Kale, *Heart Rate Variability for Biometric Authentication Using Time-Domain Features*. Springer International Publishing, 2015, p. 168–175.

[10] S. De Capitani di Vimercati, S. Foresti, and P. Samarati, "Managing and accessing data in the cloud: Privacy risks and approaches," in *Proc. of the 7th International Conference on Risks and Security of Internet and Systems (CRiSIS)*, 2012, pp. 1–9.

[11] P. Melillo, R. Izzo, A. Orrico, P. Scala, M. Attanasio, M. Mirra, N. De Luca, and L. Pecchia, "Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis," *Plos One*, vol. 10, no. 3, p. e0118504, March 2015.

[12] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

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

[14] A. Jain, P. Flynn, and A. Ross, *Handbook of Biometrics*. Springer US, 2007. [Online]. Available: https://books.google.it/books?id=WfCowMOvpioC

[15] C. Carreiras, A. Lourenço, H. Silva, A. Fred, and R. Ferreira, "Evaluating template uniqueness in ecg biometrics," in *Informatics in Control, Automation and Robotics: 11th International Conference, ICINCO 2014 Vienna, Austria, September 2-4, 2014 Revised Selected Papers*. Springer, 2016, pp. 111–123.

[16] X. Dong, W. Si, and W. Yu, "Identity recognition based on the qrs complex dynamics of electrocardiogram," *IEEE Access*, vol. 8, pp. 134 373–134 385, 2020.

[17] I. Jekova, V. Krasteva, and R. Schmid, "Human identification by cross-correlation and pattern matching of personalized heartbeat: Influence of ECG leads and reference database size," *Sensors*, vol. 18, no. 2, p. 372, 2018.