# Permanence of ECG Biometric: Experiments Using Convolutional Neural Networks

Abhishek Ranjan
Nymi Inc.
Toronto, Canada
aranjan@nymi.com

## Abstract

*ECG biometric has emerged as an appealing biometric primarily because it is difficult to spoof. Because ECG is a continuous measure of an electrophysiological signal, it is difficult to mimic, but at the same time, its day-to-day variations impact its permanence. In this paper, we present a study of the permanence of ECG biometric using a Convolutional Neural Network based authentication system and a multi-session ECG dataset collected from 800 users. The authentication system achieved an equal error rate of 2% on ECG-ID database, improving the state-of-the-art. Using this system, we designed a series of rigorous experiments by varying the days elapsed between when enrollment and authentication are performed. The results show that, despite controlling for posture, equal error rate increases as days pass. Simply including more data to enrollment does improve the accuracy, but more recent data are significantly more advantageous.*

## 1. Introduction

Electrocardiogram (ECG) has been studied as a biometric for several decades [17]. Being a continuous measure of cardiac cycles, it has many advantages over some other conventional biometrics, such as fingerprint, face, and iris. It is more difficult to lift and circumvent to the extent that recently it has been utilized as a liveness indicator for fingerprints [13]. Furthermore, it is suitable for applications requiring continuous and unobtrusive biometric monitoring such as during driving or machine operation [19].

While typical medical grade ECG capture is a relatively cumbersome process requiring careful placement of 12-leads, recent advances in capture technology allow portable, unobtrusive, and yet high quality, capture of 1-lead ECG. Using commercially available integrated circuits, many consumer grade devices now utilize 1-lead ECG for biometric [3] or healthcare purposes [2, 1]. Considering this rising trend, the present study focuses on such 1-lead ECG, also termed as *off-the-person* ECG by da Silva et al [5].

ECG is a measure of heart's electrical depolarization and repolarization throughout the cardiac cycle. Factors that influence this cycle also cause variations in an individual's ECG [10, 25]. They include physical posture, activity, electrolytes saturation, medications, stress, and cardiac conditions. As more data are becoming available, new idiosyncrasies of ECG as a biometric are further being revealed [23]. This makes it pertinent to rigorously study its biometric characteristics.

Two of the fundamental characteristics of a biometric trait are uniqueness and permanence [12]. Because ECG is tightly related to an individual's cardiovascular system, its uniqueness has been studied in both cardiology [10] and biometrics [11]. Permanence, however, has remained understudied. Permanence of a biometric trait refers to the invariability of the trait over time [12]. A trait with significant variability over time may not be appropriate as a practical biometric.

In this paper, we report on a study of permanence conducted on a dataset with ECG records collected from 800 individuals over 10 consecutive days under multiple postures. To conduct this study, we developed a novel 2D representation of ECG and designed a deep Convolutional Neural Network (CNN [9]) based authentication system. We evaluated our system's accuracy on publicly available ECG-ID multi-session dataset [16]. Having verified the high accuracy of our system, we used the system to conduct three permanence experiments examining the impact of day-to-day ECG variations on biometric authentication. To the best of our knowledge, this is the first reported study of this kind. Some of the main contributions of the paper are:

1. A novel 2D representation of an ECG record that encodes temporal properties of multiple cardiac cycles.
2. Design of a CNN based multi-session authentication

system with $2\%$ equal error rate (EER) on ECG-ID dataset.

3. Comprehensive study of the permanence of ECG biometric by analyzing the impact of day-to-day ECG variations on authentication accuracy.

In section 2, we present background work in ECG biometric, focusing on multi-session authentication and application of deep learning to ECG recognition. In section 3, we present the details of the proposed 2D ECG representation and the design and evaluation of the CNN based system. In section 4, we describe two sets of experiments on permanence and discuss results. Finally, in section 5, we draw conclusions and discuss design implications for future ECG biometric authentication systems.

## 2. Background

Past couple of decades have seen a rise in ECG biometric research using off-the-person ECG. A majority of this research used single session data in which enrollment and authentication ECG signals were captured in a single continuous session [17]. Accuracy results for biometric systems based on single session evaluation have been encouraging, many claiming less than $1\%$ EER [17, 6]. These results, however, may not represent how well such a system would perform over long-term, regular usage.

### 2.1. Multi-session ECG

Multi-session ECG authentication, with enrollment and authentication signals captured across two or more different sessions, has lately become a more relevant problem primarily because of its similarity to real-world use cases. Error rates calculated using multi-session datasets have been reported to be significantly worse than those using single session. According to a summary by Luz et al. [6], EER of some of the best performing multi-session ECG biometric systems range from $5\%$ to $18\%$.

Two sessions in a dataset may differ in several different ways, e.g., signals may be captured under different postures or heart-rate, different hydration levels, or just simply be captured on different days. When users were enrolled and authenticated under two different postures, Wahabi et al. [23], reported their algorithm's EER across five different postures (69 users) ranged from 5.61% to 24.1%. After introducing a separate posture-recognition step, the system's accuracy on a smaller dataset of 52 users was comparable to that with same posture enrollment and authentication [24].

Even with the same posture, ECG has been known to have day-to-day variations [25]. On a smaller multi-session dataset (47 users), Wahabi et al. [23], observed higher error rates (ranging from 5% to 15%) when enrollment and testing sessions were separated in time. Even though these results are based on relatively small datasets, they hint at a large gap in the current understanding of multi-session ECG biometric. Our study reported in this paper attempts to narrow this gap by systematically analyzing how a state-of-the-art ECG biometric system's accuracy changes over days.

### 2.2. Deep learning for ECG biometric

A wide range of features and algorithms have been proposed to effectively capture the aforementioned variations in ECG signals [17]. Early research focused on fiducial features [11], followed by a prevalence of non-fiducial features such as autocorrelation and Wavelet transform coefficients [21, 7]. Many successful systems applied Linear Discriminant Analysis or Principal Component Analysis for dimension reduction and k-Nearest Neighbor for classification [17]. In general, most of these approaches required expert feature selection based on time and frequency domain properties of ECG.

Deep learning has recently achieved outstanding success in other similar recognition tasks such as face and speech recognition [15]. Some of the main advantages of deep learning over traditional approaches include: (1) data driven representation learning, obviating careful feature selection, (2) high learning capacity, and (3) ability to transfer learning from one dataset to another through learned features [15]. Not surprisingly, this has motivated its application to ECG recognition as well. In particular, CNN has recently been used for representation learning [14, 6] and Recursive Neural Network (RNN) for modeling temporal behavior [26, 22]. Encouraged by success of these systems, we designed our ECG authentication system based on a CNN architecture using a novel 2D representation of ECG.

## 3. System Design and Evaluation

In this section, first, we motivate and describe our novel 2D representation of ECG. Using this representation, we next propose the CNN architecture at the core of the system. We briefly describe the private dataset we used for pretraining the CNN and the pretraining process itself. We next give an overview of the biometric template creation from a user's ECG records. Finally, we provide details of the system evaluation process using ECG-ID dataset. It should be noted that all the various ECG records used in this paper were first resampled at 125Hz and passed through a $1-20$Hz Butterworth bandpass filter.

### 3.1. 2D representation

A single channel ECG is a non-periodic one-dimensional data, but it has a generally repetitive structure reflecting the underlying cardiac cycles (a.k.a., "PQRST wave" in electrocardiography [20]). We leverage this property to generate a 2D representation of ECG. In this paper, this repeti-

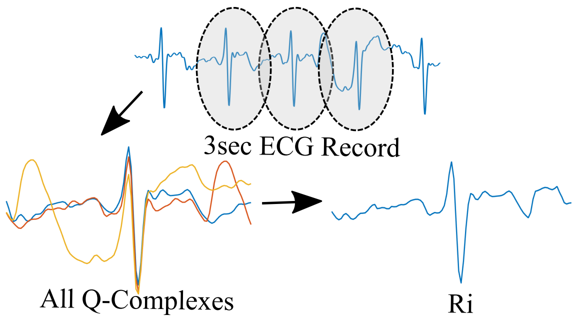


Figure 1: Generating $R_i$ from ECG record

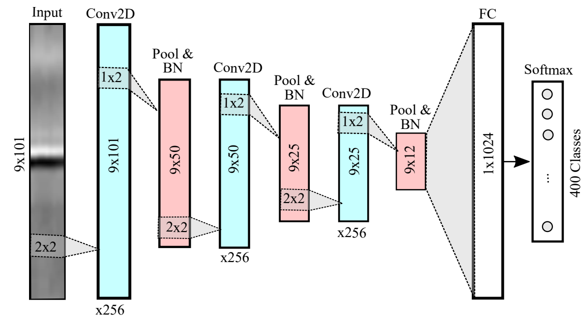


Figure 3: CNN architecture

tive structure is referred to as a *Q-complex* and is defined as a zero-mean normalized 800msec long ECG segment centered around an R-peak. It is extracted using the R-peak detection algorithm proposed by Pan et al [18]. Since our system processes records captured at 125Hz sampling frequency, a Q-complex has 101 samples. Note that a fixed-length Q-complex only approximately corresponds to an actual PQRST wave which has a variable length. This specific length was determined empirically for best authentication performance.

To create a 2D representation of an ECG record, first, it is divided into many overlapping 3 sec long windows. Secondly, for every window, a representative vector ($R_i$) is calculated as the median of all the Q-complexes in that window (see figure 1). Finally, a 2D representation ($I_i$) is generated by stacking all $R_i$s from the same record (see figure 2). For all the experiments in this paper, an $I_i$ is a 9x101 array and is generated by stacking 9 $R_i$s extracted from a 10sec long ECG record.

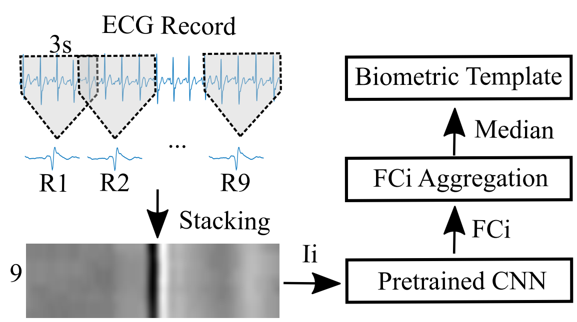


Figure 2: Authentication System

It should be noted that many $I_i$s can be generated from a single record by permuting the order in which $R_i$s are stacked. This property can be used to augment training data. Furthermore, because a single Q-complex represents temporal variations in one cardiac cycle, this 2D representation spatially encodes cardiac repetitions. Along a row (horizontal axis), two neighboring points in $I_i$ are generated by the same cardiac cycle, whereas, along a column, neighbors represent the same cardiac state across two different cardiac cycles (see figure 2).

Off-the-person ECG collected in the real world typically has myoelectric or motion noise which may distort an individual Q-complex (e.g., the ECG in figure 1). Because $R_i$ is a median of many Q-complexes, it is robust to such distortions. Recently, Wu et al. have proposed a 1D CNN based model to completely discard distorted Q-complexes from training and testing [26]. Our median based noise-reduction approach allows the model to incorporate a Q-complex even if it's partly distorted.

### 3.2. CNN architecture

A 2D representation of ECG gives us the opportunity to incorporate many features from the more mature field of computer vision to ECG biometrics (such as 2D convolution and max-pooling). To describe our CNN architecture, we follow the terminology and guidelines proposed by Goodfellow et al [9]. Figure 3 illustrates the architecture. The network has three 2D Convolutional layers (*Conv2D*). In each layer, there are 256 maps and each map uses 2x2 kernel and rectified linear activation function (*ReLu*) to introduce non-linearity. Each Conv2D layer is followed by *max-pooling* (1x2 kernel) and *batch normalization* layers (Pool & BN in figure 3). The final layer is a *fully-connected* layer with 1024 $tanh$ activation units and 20% *dropout*. The 1024-dimensional output vector is referred to as $FC$ in this paper. $FC$ is fed to a multi-class *softmax* layer for regression.

### 3.3. Our ECG dataset

Training a deep CNN requires a large dataset due to large number of model parameters [9]. Furthermore, a rigorous study of permanence of biometrics also demands a large multi-session dataset. In this context, we observed a significant gap in the current research. The median user size for all the research works surveyed by Odinaka et al. [17] is

22 (maximum 520) and that by Pinto et al. [20] is 97 (maximum 1019). Even for the largest dataset reported above only 46 users had 5 session data [23].

For a more comprehensive study, we used our own private dataset. This dataset consists of ECG records captured from 800 individuals from general population without any screening for cardiac conditions. These records were captured over a period of 10 consecutive days under two postures: comfortably seated on a chair and standing up. There are total $16,000$ records (= 800 users x 10 days x 2 postures), each record being 60sec long. A commercially available mobile device was used to capture ECG data at 125Hz across a person's right wrist and left index finger.

For the purposes of CNN training and experiments, the dataset was split into two non-overlapping subsets: 400 users for training the CNN (labeled *CNN400*) and 400 users for enrollment and authentication for the permanence experiments (labeled *EA400*).

### 3.4. Pretraining and template creation

We trained the CNN on *CNN400* dataset with 400-class *softmax* regression using categorical *cross-entropy* loss function [9]. In total $56,937$ $I_i$ were generated using all the 20 records from every user. By randomly permuting $R_i$s in each $I_i$, training data was further expanded to approximately 1.7 million $I_i$s. Using *Adam* optimizer [9] with learning rate 0.0001, the CNN was trained until validation accuracy (on 30% of samples) stopped improving. All the learned parameters of the network, excluding the softmax layer, were saved as *pretrained CNN*.

Through supervised learning, the pretrained CNN learned to represent an input $I_i$ as an $FC$, and, thus, our system first converted an ECG record to an $I_i$ and then to an $FC$. For biometric enrollment, every enrollment ECG record of a user was converted to an $FC$ and a biometric template of the user ($FC_T$) was defined as the median of all the $FC$s. During authentication, a query ECG record was converted to an $FC$ and classified as a match if $d_{cosine}(FC, FC_T) < Th$, where $Th$ is a predetermined threshold and $d_{cosine}()$ is *cosine distance* metric.

### 3.5. System evaluation

The authentication accuracy of our biometric system was estimated using the ECG-ID dataset [16, 8] using EER metric. We selected this dataset for two main reasons: (1) it is one of the largest publicly available multi-session dataset (89 users with 2 or more sessions) exclusively created for biometric authentication, and (2) it has been known to be particularly challenging as it consists of records collected under unrestrained movement [20, 22, 26, 7]. Of the two sessions per user, one was used for enrollment and the other was used for authentication.

Since the pretrained CNN was trained using one dataset and being evaluated on a different dataset, we introduced a *supervised fine-tuning* step to every user's enrollment (as recommended by Goodfellow et al [9]). First, the ECG-ID dataset was randomly divided into two non-overlapping subsets: 60% (53 users) for fine-tuning the pretrained CNN and 40% (36 users) for running authentication tests. During enrollment, a biometric profile was created for each of the 36 users using their first session ECG records.

To create a biometric profile for a user, a fine-tuning dataset was created with 54 classes (53 fine-tuning users + 1 enrollment user). A 54 class *softmax* layer was appended to the end of the pretrained CNN (i.e., replacing the 400-class layer in figure 3) and the entire network was fine-tuned for 10 epochs using the fine-tuning dataset. Using this fine-tuned CNN, $FC_T$ was created following the steps in subsection 3.4.

During authentication, a test dataset was created using the second session ECG records from all the 36 users. Since each session record is 20sec long, 2 test $I_i$s were created per user. Using this dataset, a receiver operator characteristic (ROC) curve was generated and EER was estimated. To reduce the impact of subject selection bias, the entire enrollment and authentication process was repeated 20 times, each time with a different group of 53 users for fine-tuning. The median EER was estimated to be 2.0%, with the best EER of all the runs being 0.9% and the worst being 3.5%.

To the best of our knowledge, this accuracy is on-par with or better than other published authentication results on ECG-ID dataset (see Pinto et al. [20] for a compilation). We present an EER comparison in table 1. Note that Wu et al. [26] proposed a system that achieved an EER of 0.52% by discarding invalid beats during authentication making it unclear how long an ECG record was needed for a single authentication attempt. Since our system was forced to process every 10sec record (regardless of the noise level), it is difficult to compare the two systems.

Table 1: Comparison of Authentication accuracy

| System | Authentication Accuracy |
|---|---|
| [4] | EER=2.4% |
| [27] | EER=15.0% |
| [22] | minimum EER=1.5% |
| Our system | EER=2.0% (minimum EER=0.9%) |

## 4. Permanence Experiments

Electrocardiography researchers have been aware of day-to-day variations in ECG for decades [25] and have incorporated these variations when defining a "normal" ECG. In biometrics research, however, this has largely remained understudied [23]. In this section, we report on a series of experiments we conducted to explore the impact of day-to-day ECG variation on authentication accuracy.

The primary instrument used in all the experiments was the authentication system described in the previous section. For all the experiments, we used *EA400* dataset with 10 days data from 400 users (see subsection 3.3). Two common steps were followed across all the experiments: (1) enroll all the users using ECG records captured on certain days, and (2) authenticate using ECG records captured on certain other days. By varying the enrollment and authentication days, we studied the variations in accuracy over time.

In what follows, an experiment labeled $E_{l,m,n,...} \rightarrow A_{x,y,z,...}$ refer to a setup in which enrollment and authentication data were generated using records from days $l, m, n, ...$ and $x, y, z, ...$, respectively. For example, an experiment labeled $E_{1,2} \rightarrow A_{10}$ used records captured on the 1st and the 2nd days for enrollment and the 10th day for authentication. Results of these experiments were recorded as EERs.

## 4.1. Impact of day-to-day variation

The purpose of the first set of experiments is to examine the impact of day-to-day ECG variations on authentication accuracy. Because the disparity between enrollment and authentication samples would grow larger over days, we hypothesized that for a fixed enrollment, authentication accuracy would gradually worsen over days. To test this hypothesis, a single day's data was used for enrollment and authentication was performed on the 2nd, 5th, and 8th day after enrollment. In particular, we conducted nine experiments labeled: $E_1 \rightarrow A_2$, $E_1 \rightarrow A_5$, $E_1 \rightarrow A_8$, $E_2 \rightarrow A_3$, $E_2 \rightarrow A_6$, $E_2 \rightarrow A_9$, $E_3 \rightarrow A_4$, $E_3 \rightarrow A_7$, $E_3 \rightarrow A_{10}$. Figure 4 shows that EERs gradually worsen as the number of days elapsed between enrollment and authentication increases.

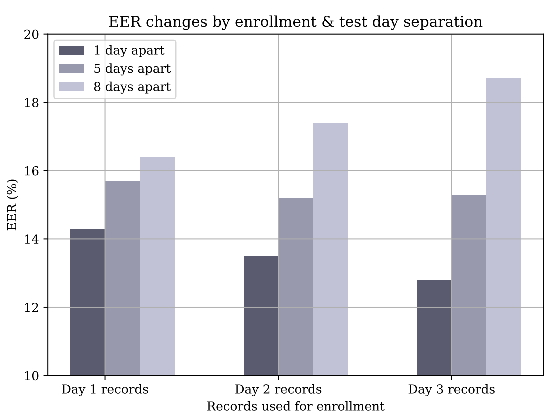


Figure 4: Impact of elapsed time on EER

## 4.2. Impact of larger training data

Because increasing training data size typically improves recognition accuracy, we hypothesized that any significant increase in the number of enrollment days would improve EER. Extending this hypothesis based on the results of the previous set of experiments, we further hypothesized that selecting enrollment days adjacent to the authentication day would be more beneficial than days further away.

To test these hypotheses, we ran 6 experiments. Picking day 3 as the authentication day, we ran $E_2 \rightarrow A_3$, $E_{7,8,9,10} \rightarrow A_3$, and $E_{1,2,4,5} \rightarrow A_3$; and picking day 8 as the authentication day, we ran $E_7 \rightarrow A_8$, $E_{1,2,3,4} \rightarrow T_8$, $E_{6,7,9,10} \rightarrow A_8$. Figure 5 shows accuracy improvements when 4 distant days were used for enrollment and further significant improvements when 4 adjacent days were used.

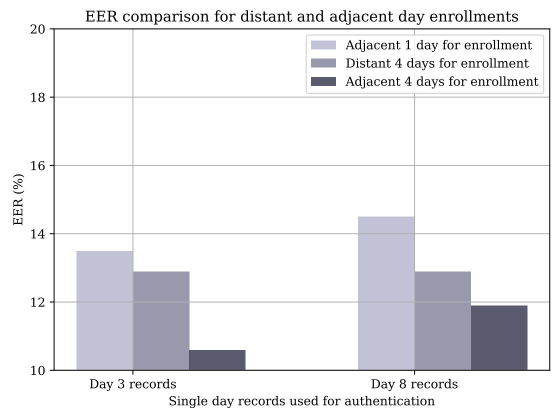


Figure 5: Impact of elapsed time on EER

# 5. Conclusions

Uniqueness and permanence are two fundamental characteristic of a biometric trait. While the uniqueness of ECG biometric has recently gained significant attention, permanence largely remains unexplored. In this paper, we reported on the design of an ECG authentication system and a study of the permanence of the biometric by analyzing the impact of day-to-day variations in ECG on system accuracy. Because our goal was to be able to generalize the results of the study to a wider range of systems, the accuracy of our system was crucial. The higher the accuracy, the wider the applicability of the study results.

To develop such a system, we proposed a novel 2D representation of ECG and designed a CNN based architecture. The system was trained on a large private dataset of $400$ users and rigorously evaluated on publicly available multi-session ECG-ID dataset. With an EER of $2\%$, our system's accuracy was on-par with or better than existing systems.

Having verified the accuracy, we used our system as an instrument to conduct a series of permanence experiments.

In the first set of experiments, the time elapsed between enrollment and authentication was considered as the independent variable. Next set of experiments also included the number of ECG records as an independent variable. System accuracy (in terms of EER) was the dependent variable. Results indicate that with a single session enrollment, authentication accuracy degrades as days pass. However, when more ECG records from different days are included, the accuracy significantly improves. More recent enrollment data can further improve accuracy.

Although accuracy degradation may partly be attributed to our system's limitations, the high accuracy on ECG-ID dataset indicates that the results would still be widely applicable to most existing systems. In particular, our results strongly suggest that real-world deployments would significantly benefit from incorporating recent ECG data into biometric templates.

## References


[1] Alivecor. https://www.alivecor.com.

[2] Apple watch series 4. https://www.apple.com/ca/apple-watch-series-4/health/.

[3] Nymi inc. https://nymi.com.

[4] S. Y. Chun. Single pulse ecg-based small scale user authentication using guided filtering. In *IEEE ICB*, pages 1–7, June 2016.

[5] H. P. da Silva, C. Carreiras, A. Lourenço, A. Fred, R. C. das Neves, and R. Ferreira. Off-the-person electrocardiography: performance assessment and clinical correlation. *Health and Technology*, 4(4):309–318, Apr 2015.

[6] 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 TIFS*, 13(5):1258–1270, May 2018.

[7] 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 *InfoSec*, pages 5–10, Nov 2015.

[8] 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. Peng, and H. E. S. HE. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, June 2000.

[9] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[10] L. S. Green, R. L. Lux, C. W. Haws, R. R. Williams, S. C. Hunt, and M. J. Burgess. Effects of age, sex, and body habitus on qrs and st-t potential maps of 1100 normal subjects. *Circulation*, 71:244–253, 1985.

[11] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold. Ecg to identify individuals. *Pattern Recogn.*, 38(1):133–142, Jan. 2005.

[12] A. K. Jain, A. A. Ross, and K. Nandakumar. *Introduction to Biometrics*. Springer, Boston, MA, 2011.

[13] M. Komeili, N. Armanfard, and D. Hatzinakos. Liveness detection and automatic template updating using fusion of ecg and fingerprint. *IEEE TIFS*, 13(7):1810–1822, July 2018.

[14] R. D. Labati, E. Muñoz, V. Piuri, R. Sassi, and F. Scotti. Deep-ecg: Convolutional neural networks for ecg biometric recognition. *Pattern Recognition Letters*, 2018.

[15] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(436), May 2015.

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

[17] I. Odinaka, P. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbaugh. Ecg biometric recognition: A comparative analysis. *IEEE TIFS*, 7(6):1812–1824, Dec 2012.

[18] J. Pan and W. J. Tompkins. A real-time qrs detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236, March 1985.

[19] J. R. Pinto, J. S. Cardoso, A. Lourenco, and C. Carreiras. Towards a continuous biometric system based on ecg signals acquired on the steering wheel. *Sensors*, 2017.

[20] J. R. Pinto, J. S. Cardoso, and A. Lourenço. Evolution, current challenges, and future possibilities in ecg biometrics. *IEEE Access*, 6:34746–34776, 2018.

[21] K. N. Plataniotis, D. Hatzinakos, and J. K. M. Lee. Ecg biometric recognition without fiducial detection. In *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, pages 1–6, Sept 2006.

[22] R. Salloum and C. . J. Kuo. Ecg-based biometrics using recurrent neural networks. In *IEEE ICASSP*, pages 2062–2066, March 2017.

[23] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos. On evaluating ecg biometric systems: Session-dependence and body posture. *IEEE TIFS*, 9(11):2002–2013, Nov 2014.

[24] S. Wahabi, S. Pouryayevali, and D. Hatzinakos. Posture-invariant ecg recognition with posture detection. In *IEEE ICASSP*, pages 1812–1816, April 2015.

[25] J. L. Willems and P. F. Poblete. Day-to-day variation of the normal orthogonal electrocardiogram and vectorcardiogram. *Circulation*, 16:1057–1064, 1972.

[26] B. Wu, G. Yang, L. Yang, and Y. Yin. Robust ecg biometrics using two-stage model. In *IEEE ICPR*, pages 1062–1067, Aug 2018.

[27] E. K. Zaghouani, A. Benzina, and R. Attia. Ecg based authentication for e-healthcare systems: Towards a secured ecg features transmission. In *IWCMC*, pages 1777–1783, June 2017.