

Learning Deep Off-the-Person Heart Biometrics Representations

Eduardo José da Silva Luz, Gladston J. P. Moreira, Luiz S. Oliveira,
William Robson Schwartz, *Member, IEEE*, and David Menotti[✉], *Member, IEEE*

Abstract—Since the beginning of the new millennium, the electrocardiogram (ECG) has been studied as a biometric trait for security systems and other applications. Recently, with devices such as smartphones and tablets, the acquisition of ECG signal in the off-the-person category has made this biometric signal suitable for real scenarios. In this paper, we introduce the usage of deep learning techniques, specifically convolutional networks, for extracting useful representation for heart biometrics recognition. Particularly, we investigate the learning of feature representations for heart biometrics through two sources: on the raw heartbeat signal and on the heartbeat spectrogram. We also introduce heartbeat data augmentation techniques, which are very important to generalization in the context of deep learning approaches. Using the same experimental setup for six methods in the literature, we show that our proposal achieves state-of-the-art results in the two off-the-person publicly available databases.

Index Terms—Electrocardiogram, biometric systems, deep learning, off-the-person category.

I. INTRODUCTION

ELECTROCARDIOGRAM (ECG) as a biometric is a recent research topic and it is a trend in the literature [1]–[4]. The ECG is a robust biometric from a security point of view as it is difficult to fake or steal. However, it is also difficult to use it on day-to-day biometric applications since it often demands uncomfortable interventions for signal acquisition process. Nonetheless this reality can change due to new forms of capturing the ECG signal. According to Silva *et al.* [5] there are three types of acquisition categories been investigated nowadays: *in-the-person*, *on-the-person*, and *off-the-person*.

Manuscript received April 3, 2017; revised October 20, 2017 and November 20, 2017; accepted December 4, 2017. Date of publication December 18, 2017; date of current version January 30, 2018. This work was supported in part by UFOP, UFPR, UFMG, FAPEMIG (APQ-02825-14, PPM-00540-17, APQ-00567-14), in part by CAPES, and in part by CNPq under Grant 307010/2014-7, Grant 311053/2016-5, and Grant 428333/2016-8. The work of E. J. da Silva Luz was supported by an IBM Ph.D. fellowship. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Adams W. K. Kong. (*Corresponding author: David Menotti.*)

E. J. da Silva Luz and G. J. P. Moreira are with the Computing Department, Federal University of Ouro Preto, Ouro Preto 35400-000, Brazil (e-mail: eduluz@iceb.ufop.br; gladston@iceb.ufop.br).

L. S. Oliveira and D. Menotti are with the Department of Informatics, Federal University of Paraná, Curitiba 81.531-980, Brazil (e-mail: lesoliveira@inf.ufpr.br; menotti@inf.ufpr.br).

W. R. Schwartz is with the Department of Computer Science, Federal University of Minas Gerais, Belo Horizonte 31270-010, Brazil (e-mail: william@dcc.ufmg.br).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIFS.2017.2784362

Majority of devices used for ECG measurements are in *on-the-person* category. Devices in this category usually require the usage of some electrodes attached to the skin surface greased with conductive gel. Nowadays, the standard devices used for medical heartbeat analysis come from this category and some few databases acquired with those devices are used to evaluate methods aiming ECG as biometry.

Acquisition on *in-the-person* category are even more invasive than the last one. In this category, there are equipments designed to be used inside the human body, such as surgically implanted ones, subdermal applications or even ingested in the form of pills. To the best of our knowledge, in the literature, there is no available database build upon those equipments.

Contrasting with the *in-the-person* category, there is the *off-the-person* category. Devices in this category are designed to measure ECG without skin contact or with minimal skin contact (e.g., through fingers). According to [5], this category is aligned with future trends of medical application where pervasive computer systems are a reality. Furthermore, research on such equipments could bring Heart biometrics to mass production equipments such as smartphones and tablets and thus leveraging the usage of ECG as biometric on daily basis. In this direction, an equipment has been released in 2014 and the acquisition of ECG signal using mobile devices become practical (see Fig. 1), but the application on focus is the stroke detection, not biometrics. To the best of our knowledge, there is no databases acquired with this equipment. Therefore, our work considers publicly databases acquired with prototype equipments idealizing this new one, as we discuss further.

A systematic review on databases for evaluating Heart biometric systems is presented in [2]. According to them, most of the ECG databases are in *on-the-person* category and they are not designed for biometrics. Two reasonably recent ECG databases were specifically designed for biometrics, i.e., The Check Your Biosignals Here initiative (CYBHi) [6] and The University of Toronto Database (UofTDB) [3]. They belong to the *off-the-person* category and are publicly available, which makes them most suitable for benchmarking Heart biometric systems nowadays. Based on results reported to date on those databases, biometrics in *off-the-person* category is an open problem especially when one considers multiple acquisition sessions in the analysis. The ECG signal can morph due to cardiac frequency variations and this effect is more apparent when multiple sessions are employed. Design ECG representation or extract good features to overcome such problems is a challenging task.



Fig. 1. Example of commercial off-the-person ECG mobile equipment not yet used for Biometrics. Source: <https://www.alivecor.com/en/>.

Techniques to automate the creation of features have been extensively investigated in the literature, especially those based on deep learning. Since 2012, several outstanding results have been achieved in different tasks such as natural language processing, voice and object recognition, among others due to the employment of representations obtained by deep learning [7]–[10], particularly convolutional networks (CN).

Roughly speaking, the success of deep learning lies on two main aspects. First, given a large amount of data, the network is able to learn deep (several layers) and discriminative representations directly from the data themselves. Second, nowadays, we have the computational power to process such a large amount of data running parallel algorithms producing complex and powerful representations.

Inspired by those outstanding results achieved with deep learning on other tasks, in this work we introduce the usage of deep learning techniques, more specifically convolutional networks (CN), for producing discriminative representations for off-the-person Heart biometrics.

Our main contribution is a biometric system that can robustly represent a single ECG heartbeat from two different views: raw signal and spectrogram. Furthermore, we show that the fusion of the two approaches significantly reduces the system error. We also present the following secondary contributions:

- 1) This study is the first one to employ deep learning techniques to develop person verification using ECG. Experimental results demonstrate that the Equal Error Rate (EER) obtained by our proposal is significantly reduced compared to six baseline methods on two public off-the-person databases (CYBHi and UofTDB).
- 2) For the first time, the spectrogram of a heartbeat signal is represented using deep learning techniques.

- 3) As the availability of a large amount of data is an issue for deep learning approaches, we also investigate how data augmentation techniques can help to mitigate heartbeat morphing due to cardiac frequency variations.

Our experiments demonstrated, in four different evaluation setups, that the proposed method far exceeds the six methods considered as baseline, reducing the error by more than half (i.e., 50%).

The remainder of this paper is organized as follows. Related works of biometrics on ECG databases and methods are presented in Section II. Section III presents the proposed methodology for learning deep Heart biometrics representations. The benchmark used here, composed of databases and baseline methods used in an experimental comparison, is shown in Section IV. In Section V, we show our experimental setup and perform experiments by empirically comparing our method with the four methods considered as baselines. Finally, conclusions are presented in Section VI.

II. RELATED WORKS

Majority of works published in the literature are carried out with private *on-the-person* databases or public databases designed for medical investigation. To the best of our knowledge, the first work using the ECG as a biometry was proposed by Biel *et al.* [11] in 2001, in which the authors proposed methods to identify 22 individuals using ECG fiducial points, principal component analysis (PCA) and a generative model classifier (GMC). After that seminal work, other methods have been proposed [1], [12]–[16].

Irvine *et al.* [12] use 15 fiducial points chosen by a particular feature selection method (the Wilks Lambda approach) and a linear discriminant analysis classifier to identify 104 subjects. Using an approach based on the wavelet coefficients extracted from ECG raw wave, Chiu *et al.* [13] achieved 100% of verification rate for 35 healthy individuals and they found that adding heartbeats of arrhythmic individuals significantly degraded the verification rate.

Irvine *et al.* [14], focused their method exclusively on PCA and showed that PCA is a great tool for removing noise from ECG signal without loss of relevant information. However, before the PCA, all heartbeats are subsampled by a factor of four and then filtered to eliminate baseline wander.

Odinaka *et al.* [17] first pre-process each heartbeat by a bandpass Butterworth filter, with cutoff frequencies at 0.5 and 400 Hz. The spectrogram of each heartbeat is then estimated using a 64 ms length window with 54 ms overlap. Mean and standard deviation of each spectrogram bin was then used to compose feature vector. The authors used a distance metric based on the log-likelihood ratio for identification. Odínaka *et al.* [15], investigated the impact of intra-session evaluation against inter-session evaluation. They built a large and realistic database and made acquisitions spaced in days or even months apart from 269 subjects. They proposed a method for Heart biometric recognition based on *spectrogram* and achieved outstanding results.

Luz *et al.* [1] showed that the ECG is viable as a biometry even at very low frequencies, such as 30 Hz and 20 Hz, and that even ECG signals captured by low-cost devices

can provide good biometrics. However, according to Odinaka *et al.* [15], for a more realistic evaluation scenario, multiple acquisition sessions should be considered and due to the absence of public databases containing multiple sessions, few authors evaluated their methods in a more realistic way.

A QRS detection-free Heart biometric (i.e., no fiducial point detection) using a coarse grained-structure based on phase space trajectory of ECG signal is proposed by Fanga and Chanb [16]. They performed tests using 100 subjects and showed that it is possible to improve significantly results using more than one ECG lead. Also following a QRS detection-free approach, Agrafioti and Hatzinakos [18], used first-order statistics in conjunction with linear discriminant analysis to extract features from 5 consecutive seconds of ECG signal. Their method proved to be robust and able to eliminate the need for signal segmentation and consequently mitigate the error associated with this step.

Regarding *off-the-person* category, there are four well-known datasets in the literature [2], [21], two private and two publicly available. One private dataset was proposed in [19] where 90 seconds of the ECG signal was collected with two *Ag-AgCl* electrodes held between thumb and index finger. The data were collected from 50 subjects at 1 KHz and 12-bit ADC resolution. Each subject participated in three sessions. The proposed dataset is evaluated in [19] and for that, each heartbeat is firstly filtered by a 4th order bandpass Butterworth filter, with cutoff frequencies at 1 and 100 Hz. Thus, for every subject contained in the database (enrollment step), a template is created by averaging all of his/her heartbeats. The feature representation is obtained by computing the fifth level of the discrete wavelet transform for each heartbeat of the test set as well as the templates using Daubechies scalar wavelet (Db3) as the mother function. Also, a novel distance metric is proposed, i.e., the wavelet distance measure (WDIST).

The other private database was proposed in [21], in which the ECG was collected from 112 individuals of different ages, gender, and nationalities. For data acquisition, a commercial ECG device was used, and the signals were collected in rest condition, with the individual in sitting state and electrodes on both thumbs. The data was recorded in two sessions, spaced out for approximately 45 days, and each session lasted 15 seconds at a sampling rate of 250 Hz. This database was then used by authors to evaluate several ECG biometric methods and impressive results were reported in [20] with a fusion of methods.

The publicly available databases are the CYBHi [6] and UofTDB [3]. They were designed exclusively for biometrics purposes and therefore have multiple sessions spaced in days or even months apart, as described in Section IV-A. Some previous works have been employed to off-the-person category [14], [17]–[19], surveyed in Section IV-B which are used as baselines for the experiments. However, to the best of our knowledge, this is the first work to report results in both databases.

Table I summarizes outstanding results for on-the-person and off-the-person non-fiducial based works. It is difficult to compare methods evaluated in on-the-person databases with the ones evaluated in off-the-person databases since

TABLE I
OUTSTANDING RESULTS FOR ON-THE-PERSON AND OFF-THE-PERSON WORKS. *AUTHORS RE-IMPLEMENTED THE METHOD IN [18]

Works	subjects (#)	DB Category	sessions (#)	Performance
[15]	269	on-the-person	multiple sessions	5.58% (EER)
[18]	27	on-the-person	single session	<1% (EER)
[19]	50	off-the-person	multiple sessions	95% (Accuracy)
[12]	43	on-the-person	single session	100% (Accuracy)
[1]	193	on-the-person	single session	>99% (Accuracy)
[16]	100	on-the-person	single session	99% (Accuracy)
[13]	35	on-the-person	single session	0.83% (FAR) 0.86% (FRR)
[3]*	46	off-the-person	multiple sessions	18% (EER mean)
[20]	164	off-the-person	multiple session	10.81% (EER)

on-the-person databases are more controlled and often single session. According to Table I, it can be seen that for on-the-person and single session databases there is little room for improvement. However, for off-the-person bases and considering multiple sessions the problem is still open.

III. PROPOSED APPROACH

In this work, we deal with off-the-person category databases and this category of the database has low signal-to-noise ratio, in other words, excessive noise in the signal. Although it is more difficult to find fiducial points in off-the-person ECG due to noise and some studies in the literature discourage the use of fiducial point-based methods [3], [14], in [21], the authors have shown that methods based on fiducial points can obtain good results and contribute to improving off-the-person ECG biometric recognition. Thus, in this work, we investigate four non-fiducial methods and two methods based on fiducial points. The four non-fiducial methods considered as baseline are: wavelet-based method [19], spectrogram based method [17], eigenPulse method [14] and the AutoCorrelation (AC) / Linear Discriminant Analysis (LDA) method [18]. The non-fiducial methodologies are HBS: heartbeat shape template (HBS) method [22] and Fiducial feature based template (FFT) [20]. Those methods are considered state-of-the-art methods in the literature [3], [20].

We propose a new non-fiducial method based on convolutional networks (CN) for learning feature representations directly from the ECG using two approaches: on raw heartbeat signal and on heartbeat spectrogram. We propose a fusion of those approaches on score level, i.e., late fusion. An overview of the proposed method is shown in Fig. 2, detailed in the following sections.

A. CN Based Feature Representation

We propose deep architectures for each approach, based on [23]–[25], since CNs with such architectures

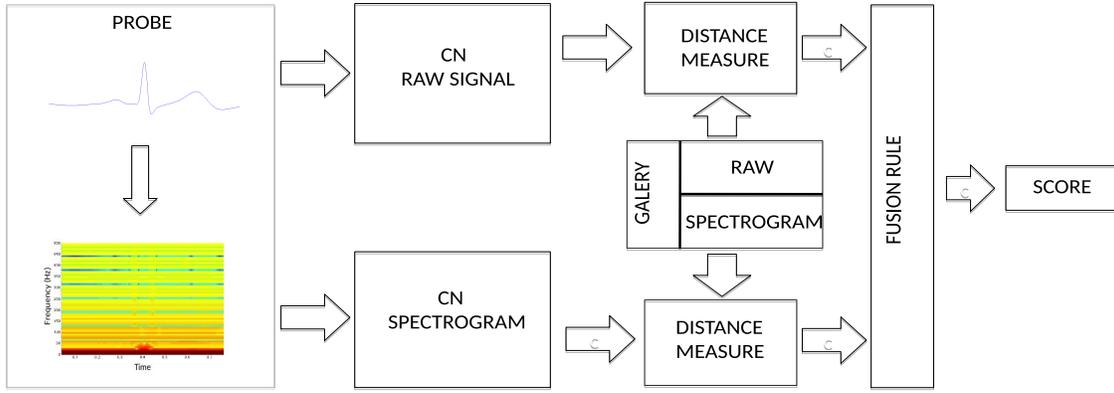


Fig. 2. Overview of proposed method.

have achieved state-of-the-art performance in several tasks [7], [26], [27].

Initially some concepts about CN are described and then our methodology is presented, aiming both 1D (signal) and 2D (spectrogram) inputs.

The network architecture is composed of the f operations of convolution, activation (ReLU), pooling, normalization and finally fully connected layers.

The convolutional operation can be defined as:

$$s(t) = \int_{-\infty}^{\infty} x(a)w(t-a) da \quad (1)$$

where x is referred to as the input and w as the kernel or filter and both have the same dimensions. For this work, the output s is referred as feature map or feature vector.

Considering the discrete nature of digitalized signals, the computational implementation can be defined as:

$$s(t) = \sum_1^n x(n)w(t-n) \quad (2)$$

where n is often smaller than the length of input signal. There is also one hidden parameter associated with equation 2: the stride. It defines the sliding step in which each signal is convolved with the other (or with a kernel). It is useful to control the downsampling rate of the output.

The convolutional operation can promote sparse connectivity, parameter sharing and equivariant representations [28]. These properties assure less computational cost, in time and space, regarding the convolutional networks with unshared weights. However, the convolution operation is not invariant to other transformation such as changes in scale or object rotation inside images and signals. Thus, other operations must be used along the network such as activation and pooling.

An *activation* function is applied after the convolution to adjust the output values to a given range. The Rectified Linear Unit (ReLU) is the one adopted here:

$$activation(s) = \begin{cases} 1 & \text{if } s < 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

This activation process aims at mimicking the function of a biological neuron allowing activation only when a certain amount of energy is reached.

Pooling operation highlights the output of convolution step aiming to keep relevant information and discard unnecessary details [29]. Similar to a convolution, the pooling operation is performed by passing a sliding window of size $n \times m$ on the signal. Within the window, a mathematical operation is performed, such as sum, average, or maximum. For instance, in maximum operator, called max-pooling, the largest value contained in the window is preserved and used for the construction of the new post-pooling signal. It is useful to achieve invariance to signal transformations and robustness against noise. It is also used to represent the feature maps in a more compact shape when the stride parameter are larger than one. For example, a pooling operation with stride size two results in an output signal that is half the size of the input.

The *normalization* operator is bioinspired from the competitive interaction observed in natural neural systems such as the gain control contrast mechanism in the cortical area [30], [31].

Fully-connected layers are useful for controlling the final size of the feature vector and also for classification [28].

1) *1D CN on Raw ECG*: Convolution operation of a raw ECG signal is illustrated in Fig. 3. In our specific case, the ECG signal has length of 800 ms and kernels are several orders of magnitude smaller, four in the illustration in Fig. 3.

For raw signal, we designed nine CN architectures, five architectures with small receptive fields in first convolutional layers inspired by [23] and others with large receptive field on the first convolutional layers inspired by [24] and [25], with the exception that the filters are one dimensional (See Table III) and proportionally adapted to heartbeat sample size. To construct the architectures, we also vary the width and depth of the networks.

It is important to note that for the UofTDB the CN architectures for the raw data had to be adapted since the sampling rate of UofTDB is five times smaller than that of the base CYBHi.

2) *2D CN on Spectrogram*: Spectrogram is the visual representation of the energy of a signal expressed as a function of frequency and time. In this work, a single ECG heartbeat is

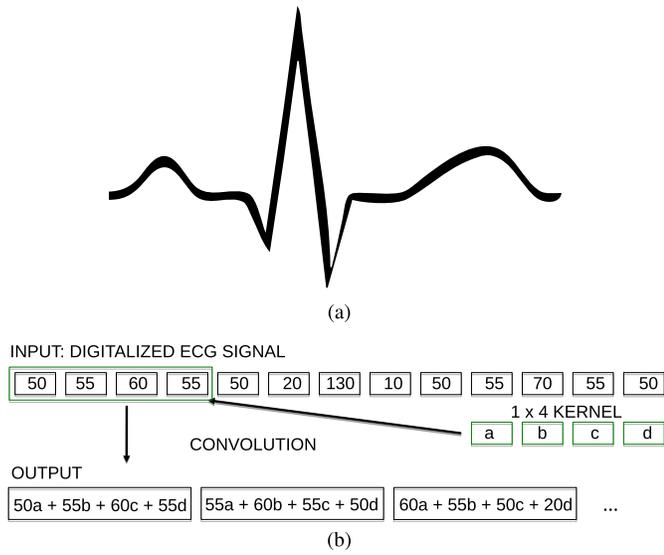


Fig. 3. Convolutional process of a raw ECG. (a) Raw signal. (b) Convolutional process.

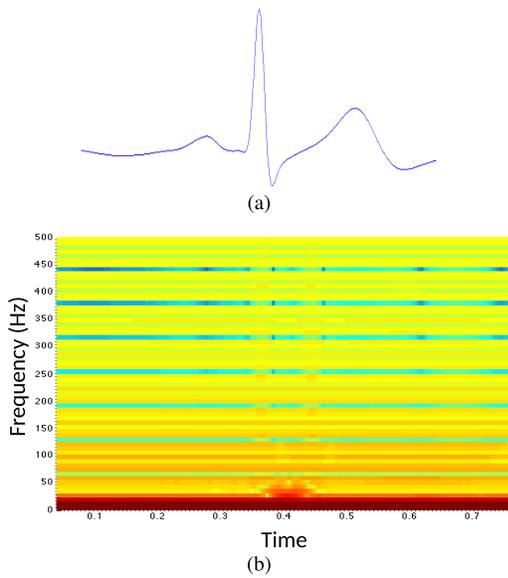


Fig. 4. Spectrogram process of a raw ECG. (a) Mean signal of CYBHi T1 session. (b) Spectrogram of the signal in (a) constructed with a 80 ms length window and overlapping of 76 ms.

segmented in several windows, which may be with or without overlap. The short-time Fourier transform [32] of each window is then calculated in order to find square of the magnitude of dominating frequency.

Fig. 4 illustrates the spectrogram for the mean heartbeat of an entire session (named T1) from CYBHi. This spectrogram is constructed with window size of 80 ms and overlap of 76 ms. For more details, see Section V.

We have also design nine architectures for the spectrogram approach, five architectures with small receptive fields in the first convolutional layers [23] and the others with large receptive field in the first convolutional layers [24], [25]. To construct the architectures, we vary width and depth of the networks.

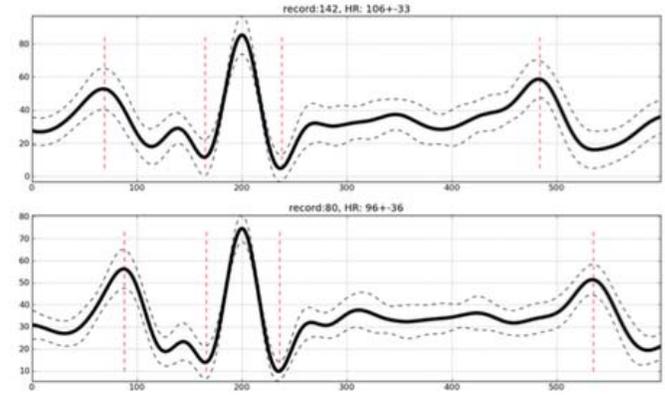


Fig. 5. Heartbeats morphing due to different heart rates between sessions. Source [4].

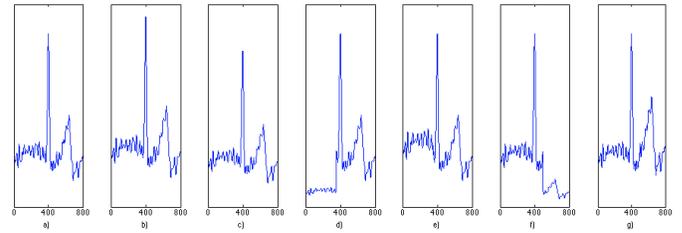


Fig. 6. Example of data augmentation on one heartbeat from record 1. a) Normal heartbeat. b) Heartbeat in scale 10% bigger. c) Heartbeat in scale 10% smaller. d) P-wave attenuated. e) Gain on P-wave. f) T-wave attenuated. g) Gain on T-wave.

3) *Data Augmentation and Dropout*: Silva *et al.* [4], make a detailed analysis on the differences between heartbeats of the same person on different sessions. Authors pointed out that in some situations, the heartbeats of the same individual presented significant changes, as can be seen in Fig. 5.

Aiming to build a more robust CN to this signal morphing, we use data augmentation. Every heartbeat of training dataset was replicated in six more heartbeats. Two new versions of the heartbeat are created by changing the scale by 10%. Other two versions are created by changing only the P-wave portion of the heartbeat, with 30% gain and attenuation, and the same procedure is applied to the T-wave as well. The criteria for data augmentation can be seen in detail in Fig. 6 and the rationale is to force the network to recognize the heartbeat even when some portion are compromised (See Fig. 9(b)).

Overfitting can be reduced by using the dropout operator [33]. Dropout is a regularizer layer which randomly sets a certain percentage of the activations of fully connected layers to zero during training. Data augmentation and dropout are used for both CN approaches proposed here.

B. CN Fusion

To investigate the impact of fusion both CN approaches, we considered three simple fusion strategies at score level: sum rule, mean rule and multiplication rule. However, we stress that more complex fusions, such as on feature level, could outperform the fusion strategies employed here.

IV. BENCHMARKS

In this section, we first describe the two publicly off-the-person ECG databases used in this work. Then, we describe baseline methods used in the comparison with our proposed approach, also the ones used in [3] and [20]. Four baseline methods [14], [17]–[19] have been selected due to their non-fiducial nature and other two are fiducial based [20], [22]. The rationale for this selection is to allow the investigation of the impact/viability of fiducial points based methods on off-the-person ECG signals.

A. Databases

Off-the-person ECG databases are the new trend for biometrics with ECG signals. In this category, the acquisition of the signal is feasible (easy and comfortable), without the inconvenience of placing electrodes on the chest or the need of conductive gel usage. Although there are some issues in these new trends such as excessive noise, off-the-person acquisition can boost the usage of ECG as day-to-day biometrics. Nowadays, it is possible to find commercial off-the-shelf ECG monitoring equipment integrated to mobile phones, as already mentioned and exemplified in Fig. 1.¹

The off-the-person databases considered in this work are the CYBHi [6] and UofTDB [3] and their prototypes used for acquisition are shown in Fig. 7. We present a brief description of the databases in the following sections.

1) *CYBHi*: *Check Your Biosignals Here initiative Database* is a public off-the-person database. Data acquisition was performed using two differential lead electrodes at hand palms and fingers and a virtual ground [34]. The acquisition was made at 1K Hz sample frequency and 12 bits resolution. The acquisition device used is the biosignalsPLUX research.² The dataset is divided into two types of experimental protocols: short-term and long-term.

The *short-term* is composed of ECG data collected from 65 people. Sessions were taken at intervals of two days and demographics showed 65 healthier subjects, in which 49 are males and 16 are females, with an average age of 31.1 ± 9.46 years old. In the proposed experimental protocol [6], participants were stimulated by specific audio and video content.

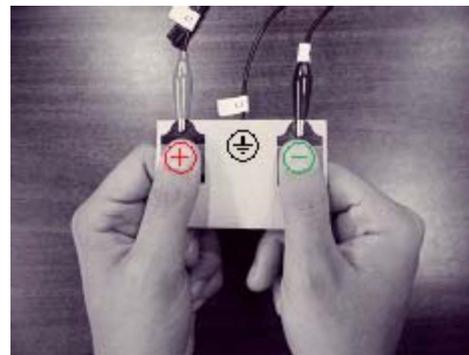
The *long-term* acquisition has data acquired from 63 health subjects in two distinct two-minute sessions, collected 3 months apart. Among the subjects, there are 14 males and 49 females, with an average age of 20.68 ± 2.83 years old. The ECG signal of both sessions were collected on the fingers (one finger of the right hand and the other of the left hand) and in the sitting position.

The *long-term* session is more challenging from the biometric point of view and consequently, it is chosen for the experiments of the present work. We call the first session as T1 and the second session, made three months later, as T2.

2) *UofTDB*: The *University of Toronto Database* is the largest off-the-person database publicly available in the literature, containing data collected from 1020 individuals.



(a)



(b)

Fig. 7. Database acquisition hardware prototypes. (a) CYBHi prototype. Source [6]. (b) UofTDB prototype. Source [3].

The data acquisition was collected with dry electrodes Ag/AgCl placed on thumbs of both hands at a sample rate of 200 Hz and 12 bits resolution. Note that the sample rate used in this database is five times smaller than the one used in CYBHi, making the heartbeat representation of this database smaller. The Vernier EKG Sensor and the Vernier Go!Link interface were used as acquisition device.³ Length of the acquisition recordings varies between 2 to 5 minutes.

This database was proposed to allow investigation of three aspects that influence the quality of the ECG signal as biometrics:

- 1) *Large number of individuals*. The database contains data from 1020 subjects with age ranging from 18 to 52 years old.
- 2) *Posture*. Five postures/conditions were used during the acquisition - sitting, standing, supine, tripod and exercise.
- 3) *Multiple sessions*. Few databases in the literature perform acquisition in multiple spaced sessions. To address this issue, the authors of the database collected the data in six sessions (S1, S2, S3, S4, S5 and S6) spanned over a period of six months.

Although the database has data of 1020 subjects, not all subjects have data in all sessions. The first session is the largest, with data from 1012 subjects and only 100 subjects were selected to participate in the other 5 sessions (See Table II). Among the 100 selected, only 46 participated in

¹<http://www.alivetec.com/store/alivetecor-heart-monitor-for-ios-and-android>

²<http://www.biosignalsplux.com>

³<http://www.vernier.com>

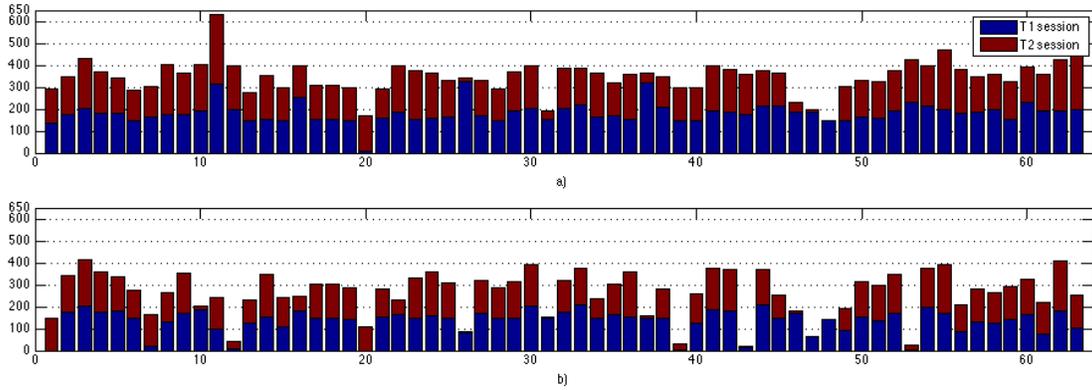


Fig. 8. Histogram of heartbeat distribution before (a) and after (b) the outlier removal algorithm for CYBHi DB.

TABLE II
UofTDB SUBJECT DISTRIBUTION BY ACQUISITION
CONDITION AND SESSION

Session	Condition				
	sit	stand	exercise	supine	tripod
1	1012	0	0	0	0
2	72	72	0	0	0
3	76	5	71	0	0
4	63	0	0	0	0
5	0	0	0	63	63
6	65	65	0	0	0

all subsequent sessions. As suggested by the creators of the database in [3], in order to perform an inter-session evaluation, we only considered individuals in which the heartbeats were acquired in seated condition and who participated in all five sessions (S1, S2, S3, S4 and S6), adding up 46 subjects.

B. Baseline Methods

Following is a brief description of baseline methods and a highlight on the modifications made exclusively for this work. Modifications in the methods were made in order to better adapt them to the CYBHi and UofT databases. To maintain compatibility with the work published in [3], all six baseline methods were implemented using a window of 700 ms for segmentation, 200 ms before the R peak and 500 ms after. Also, all methods are analyzed by the same distance metric during the verification, i.e., the Euclidean distance. For sake of reproducibility and fair comparison our implementation is available in *our CSILab website*.⁴

1) *Wavelet Based Method*: In the original work published in [19], to perform verification, the authors proposed a new distance metric based on wavelet coefficients (WDIST). However, in present work we analyze the features under the same distance metric to make a direct comparison between them. Thus, the distance proposed in [19] is not used.

2) *Spectrogram Based Method*: Odnaka *et al.* [17] used a distance metric based on log-likelihood ratio, which was also

discarded by the present work since our analysis focus on verification mode.

3) *EigenPulse Method*: Based in [14], the covariance matrix for all heartbeats of the training set is then estimated and all heartbeats (train and test sets) are projected onto eigenvectors. The principal components corresponding to 99.9% of the signal were used as feature vectors.

4) *AC/LDA Method*: Unlike the other methods, the window used for segmentation is 6 seconds length, being 200 ms before the R wave and 5800 ms after. The 6 seconds window length yields better results for UofTDB as pointed by [3] and we have also validated this parameter for CYBHi. Here we chose to use the information of R-peak location to guarantee the same number of pairs comparisons during the verification process. However, it is worth mentioning that the original method proposed in [18] works regardless of the detection of fiducial points. To extract the features, the normalized autocorrelation coefficients of each window are estimated following the same approach of [3]. The autocorrelation lag is fixed in 900 ms and the feature vector is reduced with PCA before being projected onto a new space with Linear Discriminant Analysis. No filters were used for this method.

5) *FFT Method*: The FFT method proposed in [20] consists of fifteen features extracted from ECG segments, amplitudes and ratios. The features are formed by a junction of five fiducial points (P, Q, R, S, and T): length in time of PQ, QR, RS, ST, PR, QS, and RT segments; amplitude in millivolts of PR, QR, SR, and TR; ratio of amplitudes of QP and QR, QR and SR; ratio of an SR and ST segments. The fifteen-dimensional feature vector is then used to represent a unique ECG heartbeat. The fiducial point extractor used here is an implementation of us based on the work in [35].

6) *HBS Method*: For the heartbeat shape (HBS) template [22], the signal is resampled to 1200 Hz and then normalized on the interval [0,1]. The resampled signal is filtered by a second derivative Gaussian kernel and divided into several bins. A window of five heartbeats is used to compute the average heartbeat bins to form the HBS feature vector. It is worth noting that to maintain the same number of intra-class and inter-class pairs from other methods, the five-heartbeat window is shifted by one heartbeat at a time.

⁴www.csilab.ufop.br

TABLE III
ARCHITECTURES FOR CN AT RAW ECG SIGNAL

Name	Type	Input Size	# of Filters	Filter Size /Stride/Pad	Relu	Norm
Arch A						
Conv1	conv	1x800	96	1x181/1/0	yes	yes
Pool1	max pooling	1x620	N/A	1x2/2/0	no	no
Conv2	conv	1x310	128	1x51/1/0	yes	yes
Pool2	max pooling	1x260	N/A	1x2/2/0	no	no
Conv3	conv	1x130	512	1x17/1/0	yes	yes
Pool3	max pooling	1x114	N/A	1x2/2/0	no	no
FC4	full. conn.	1x29	4096	1x29/1/0	yes	no
FC5	full. conn.	1x1	4096	1x1/1/0	yes	no
FC6	full. conn.	1x1	63	1x1/1/0	no	no
Drop7	Dropout	1x1	63	N/A	no	no
Cost8	Softmax	N/A	N/A	N/A	N/A	N/A
Arch B						
Conv1	conv	1x800	96	1x81/1/0	yes	yes
Pool1	max pooling	1x720	N/A	1x2/2/0	no	no
Conv2	conv	1x360	128	1x17/1/0	yes	yes
Pool2	max pooling	1x344	N/A	1x2/2/0	no	no
Conv3	conv	1x172	256	1x7/1/0	yes	yes
Pool3	max pooling	1x166	N/A	1x2/2/0	no	no
Conv4	conv	1x83	512	1x7/1/0	yes	yes
Pool4	max pooling	1x77	N/A	1x2/2/0	no	no
FC5	full. conn.	1x38	4096	1x38/1/0	yes	no
FC6	full. conn.	1x1	4096	1x1/1/0	yes	no
FC7	full. conn.	1x1	4096	1x1/1/0	no	yes
FC8	full. conn.	1x1	63	1x1/1/0	no	no
Drop9	Dropout	-	N/A	N/A	no	no
Cost10	Softmax	N/A	N/A	N/A	N/A	N/A
Arch C						
Conv1	conv	1x800	96	1x81/1/0	yes	yes
Pool1	max pooling	1x720	N/A	1x2/2/0	no	no
Conv2	conv	1x360	128	1x17/1/0	yes	yes
Pool2	max pooling	1x344	N/A	1x2/2/0	no	no
Conv3	conv	1x172	256	1x7/1/0	yes	yes
Pool3	max pooling	1x166	N/A	1x2/2/0	no	no
Conv4	conv	1x83	256	1x7/1/0	no	no
Conv5	conv	1x77	256	1x7/1/0	yes	yes
Pool5	max pooling	1x71	N/A	1x2/2/0	no	no
Conv6	conv	1x35	256	1x7/1/0	no	no
Conv7	conv	1x29	256	1x7/1/0	yes	yes
Pool7	max pooling	1x23	N/A	1x2/2/0	no	no
FC8	full. conn.	1x11	4096	1x11/1/0	yes	no
FC9	full. conn.	1x1	4096	1x1/1/0	yes	no
FC10	full. conn.	1x1	4096	1x1/1/0	yes	no
FC11	full. conn.	1x1	4096	1x1/1/0	no	yes
FC12	full. conn.	1x1	63	1x1/1/0	no	no
Drop13	Dropout	1x1	63	N/A	no	no
Cost14	Softmax	N/A	N/A	N/A	N/A	N/A

C. Fusion of Baseline Methods

Recently, Islam *et al.* [20] have shown that it is possible to leverage ECG biometric results by fusing several state-of-the-art ECG representations. Islam *et al.* [20], used a heuristic to select the best fusion combination at score level. For the present work, we have following the same methodology presented in [20] to select the best combination of the six baseline methods presented here. The individual result of each method is used to compute the weights for the weighted sum fusion rule.

V. EXPERIMENTS

Experiments were conducted on an Intel (R) Core i7-5820K CPU @ 3.30GHz 12-core machine, 64GB of DDR4 RAM and one GeForce GTX TITAN X GPU. In this work, we use the MatConvNet library for the convolutional networks [36] linked to NVIDIA CuDNN.

Next, experimental results for the selected architectures are presented as well as for the baseline methods. For comparison

TABLE IV
ARCHITECTURES FOR CN AT SPECTROGRAM

Name	Type	Input Size	# of Filters	Filter Size /Stride/Pad	Relu	Norm
Arch D						
Conv1	conv	128x128	64	11x11/1/0	yes	yes
Pool1	max pooling	118x118	N/A	2x2/2/0	no	no
Conv2	conv	59x59	128	7x7/1/0	yes	yes
Pool2	max pooling	53x53	N/A	2x2/2/0	no	no
Conv3	conv	26x26	256	3x3/1/1	yes	yes
Pool3	max pooling	26x26	N/A	2x2/2/0	no	no
Conv4	conv	13x13	512	3x3/1/1	yes	yes
Pool4	max pooling	13x13	N/A	2x2/2/0	no	no
FC5	full. conn.	6x6	4096	6x6/1/0	yes	no
FC6	full. conn.	1x1	4096	1x1/1/0	no	yes
FC7	full. conn.	1x1	4096	1x1/1/0	no	no
FC8	full. conn.	1x1	63	1x1/1/0	no	no
Drop9	Dropout	1x1	63	N/A	no	no
Cost10	Softmax	N/A	N/A	N/A	N/A	N/A
Arch E						
Conv1	conv	128x128	32	7x7/1/0	yes	yes
Conv2	conv	122x122	64	5x5/1/0	yes	yes
Pool2	max pooling	118x118	N/A	2x2/2/0	no	no
Conv3	conv	59x59	128	3x3/1/0	yes	yes
Conv4	conv	57x57	192	3x3/1/0	yes	yes
Pool4	max pooling	55x55	N/A	2x2/2/0	no	no
Conv5	conv	27x27	256	3x3/1/0	yes	yes
Pool5	max pooling	25x25	N/A	2x2/2/0	no	no
Conv6	conv	12x12	256	3x3/1/0	yes	yes
FC4	full. conn.	10x10	4096	10x10/1/0	yes	no
FC5	full. conn.	1x1	4096	1x1/1/0	no	yes
FC6	full. conn.	1x1	4096	1x1/1/0	no	no
FC6	full. conn.	1x1	63	1x1/1/0	no	no
Drop7	Dropout	1x1	63	N/A	no	no
Cost8	Softmax	N/A	N/A	N/A	N/A	N/A

purposes, all six baseline methods, see description in the Section IV-B, were implemented here and evaluated with the same protocol.

A. Preprocessing

1) *Heartbeat Segmentation*: To prepare for heartbeat segmentation, raw ECG signal is firstly filtered with a 4th order, zero-phase Butterworth bandpass filter with cutoff frequencies at 0.5 Hz and 40 Hz. The well-known Pan & Tompkins [37] algorithm is used to locate the QRS wave. Once the QRS wave is located, the heartbeat is delimited by an 800 ms window centered at R peak. Also, each heartbeat is zero mean and unit variance normalized, i.e., by subtracting its subject mean heartbeat and dividing by its standard deviation. Although QRS detection is done with a filtered version of the signal, the effective segmentation is done on raw signal. The rationale for this is to preserve the signal so that each method performs its own preprocessing and subsequently feature extraction.

Once the heartbeat segmentation is performed, the total number of heartbeats correctly segmented for CYBHi and UofTDB databases are 21, 886 and 61, 312, respectively.

Last preprocessing step subtracts mean value, computed on the training set, from all heartbeats.

2) *Outlier Removal*: To exclude highly corrupted heartbeats or segmentation defects, an outlier removal algorithm is employed. As it is difficult to reproduce the outlier removal algorithms of previous works [3], we adopt a simple and reproducible algorithm. The algorithm consists in calculating the mean heartbeat of each subject per session and then

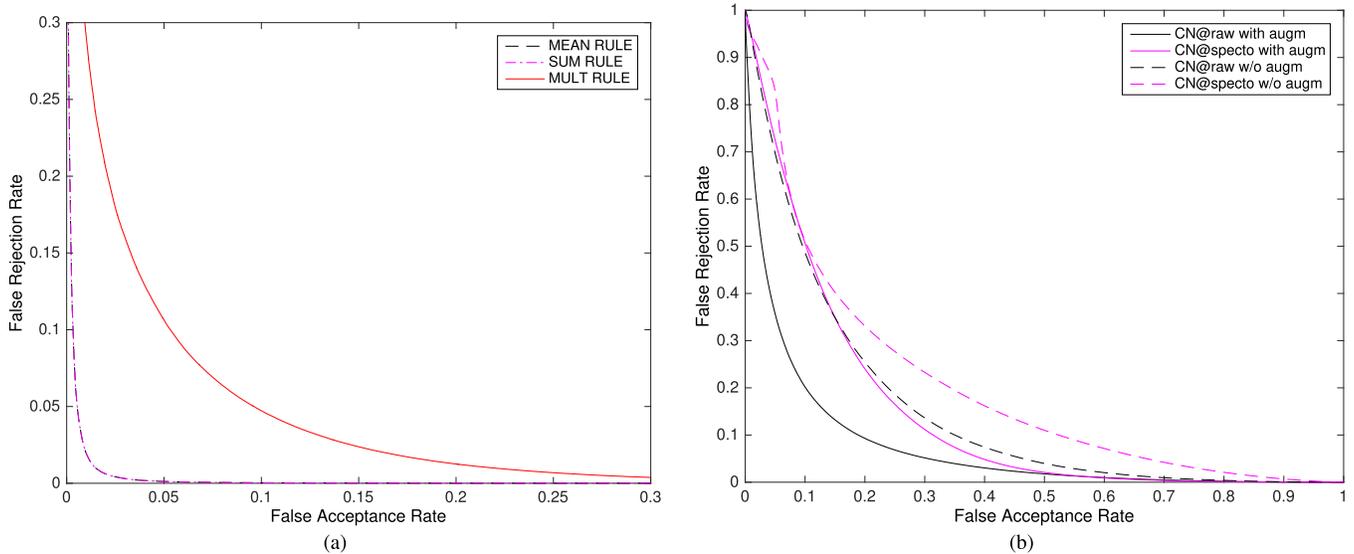


Fig. 9. Impact of fusion rules and data augmentation. (a) DET curve for fusion rules on CYBHi training set (T1-T1). (b) DET curve for data augmentation on CYBHi T1-T2 evaluation.

TABLE V

EER VALUES OBTAINED FOR THE BASELINE METHODS AND OUR APPROACH EACH SCENARIO. *WE FOLLOW THE FUSION METHODOLOGY PROPOSED IN [20]. FOR CYBHi DATABASE, THE FUSION ALGORITHM SELECTED ALL METHODS AND FOR UoFTDB THE FUSION ALGORITHM SELECTED: ODINAKA et al.(2010), AGRAFIOTI et al.(2008), ISLAM et al.(2014) AND ISLAM et al.(2017) (FFT)

Database	Scenario		Baseline Methods						Our Approach			
	Enroll	Probe	Odinaka et al. (2010)	Agrafioti et al. (2008)	Irvine et al. (2008)	Chan et al. (2008)	Islam et al.(FFT) (2017)	Islam et al. (2012)	Baseline* (Fusion)	CN@ raw	CN@ spectro	CN@ fusion
CYBHi	T1	T1	23.81	22.14	32.5	27.39	47.49	42.12	11.12	1.17	19.57	1.33
CYBHi	T1	T2	30.36	33.37	36.76	49.37	48.56	49.60	15.98	14.13	26.38	12.78
CYBHi	T2	T1	29.25	40.88	37.54	29.25	48.56	49.59	24.51	15.60	20.48	13.93
UoFTDB	S1	S2+S3+S4+S6	36.44	32.29	36.08	36.44	43.95	42.96	30.89	16.92	19.37	14.27

compute the Euclidean distance of each heartbeat to the mean heartbeat. We discard 15% of the events detected by the QRS detector algorithm, which corresponds to the mean and one and a half standard deviations for the CYBHi database, while for the UoFTB database corresponds to the mean and two and a half standard deviations. In addition, heartbeats whose maximum value is not the point referring to R-wave are discarded, due to bad segmentation. The number of heartbeats after outlier removal is 17, 226 for CYBHi database and 50, 401 for UoFTDB. After outlier removal, subjects who did not have at least one single valid heartbeat in one of the sessions were excluded from the analysis. Fig. 8 shows a histogram of the number of heartbeats per individual for CYBHi database.

B. Representation Learning

The heartbeat, represented by a 800 ms length signal or a 128×128 spectrogram image, is feed-forwarded through network layers. Each layer represents one or more CN operations: convolutional filter, pooling, stride, rectification (RELU), normalization (L2 Norm). Convolutional stride is set to one and the padding is sometimes used to preserve the original size of the signal. Pooling layer performs max-pooling operation

and when there is downsampling ($stride > 1$), it happens in conjunction with pooling. The stack of layers is followed by *Fully-Connected* (FC) layers and the last FC layer is for classification. Thus, final number of channels is equal to number of classes and final layer is a soft-max loss one. The FC layer with 1×1 filter size is used for dimension reduction and rectified linear activation. After the learning process, the last layer (the classifier layer) is removed and the new network output can be used as feature vector for identity verification problem, such as done in [7] and [8]. In this way, the CN becomes a feature extractor.

For training the CN, stochastic gradient descent with mini-batches of size 100 and momentum coefficient of 0.9 are employed. Three different learning rates are considered during 80 training epochs, i.e., 0.01 for the first 40 epochs, then 0.001 for more 20 epochs, and finally 0.0001 for the last 20 epochs. The stopping criterion used in the training is the number of epochs, i.e., 80. With higher values for training epochs, no improvements have been noticed. The weights of filters are randomly initialized with zero mean Gaussian distribution and deviation of 10^{-2} . During the training phase, 10% of data is reserved for validation and dropout operation is placed before the last layer with 10%.

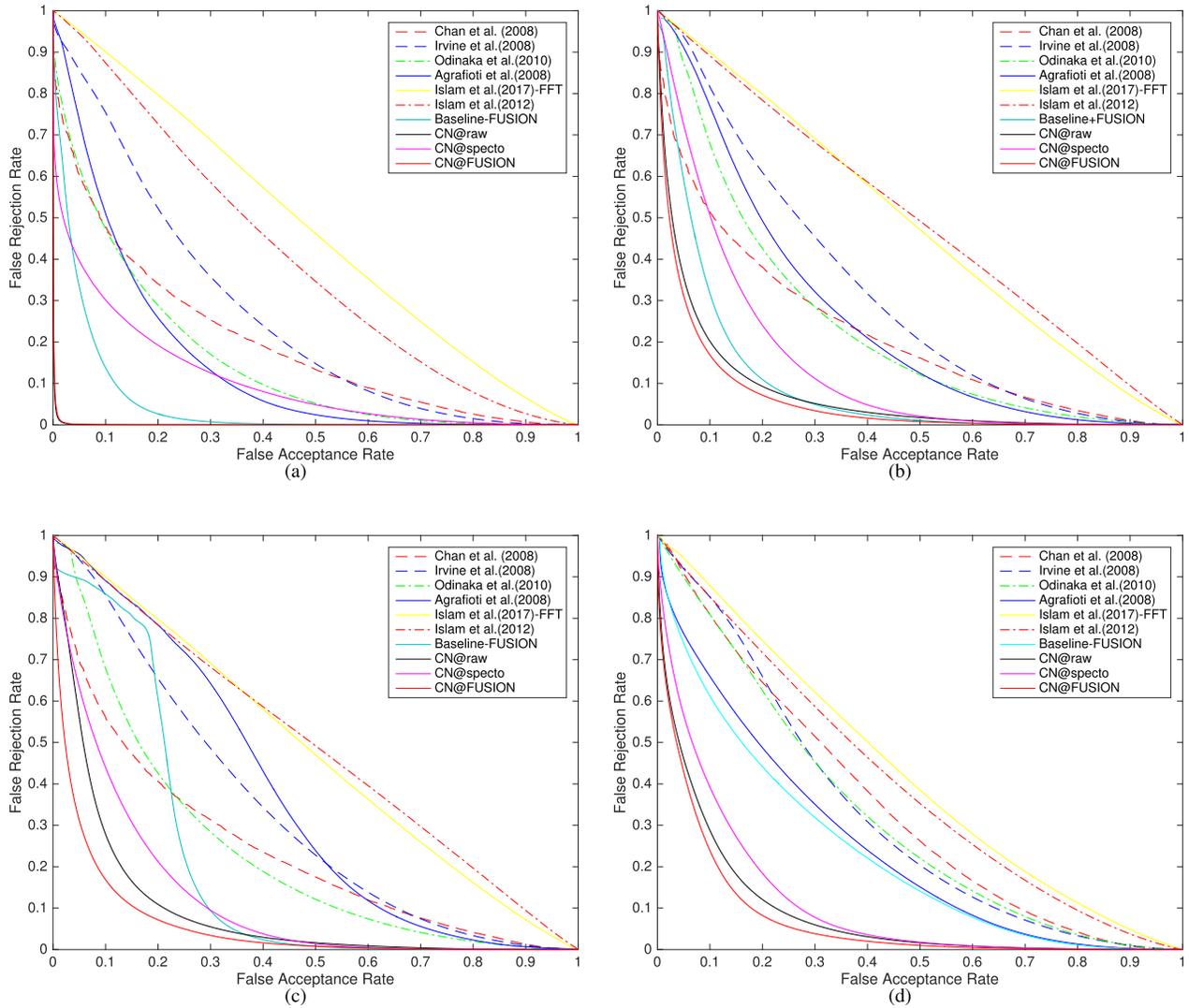


Fig. 10. Final results on CYBHi and UofTDB. [Best viewed in color]. (a) INTRA SESSION evaluation for CYBHi. Training and Testing on T1. (b) INTER.1 evaluation for CYBHi. Training and enrollment on T1 and T2 as probe. (c) INTER.2 evaluation for CYBHi. Training and enrollment on T2 and T1 as probe. (d) INTER.3 evaluation for UofTDB. Training with S1 and enrollment on S2, S3, S4, and S6.

A grid search is employed to achieve best architectures and the ones with lower error on the validation set are shown on and Table IV and Table III. Parameters evaluated on grid search are the size of receptive fields, number of filters per layer and number of layers. Before that, a grid search was also performed in order to achieve optimum figures for spectrogram image and architecture D was chosen for that task due to its lowest computational cost. The selected figures for spectrogram are window width of 80 ms and overlapping of 95%.

C. Experimental Setup

Evaluation is carried out in two scenarios: intra-session (same-session) and inter-session (across-session). For intra-session evaluation, the model and parameters are estimated based on half the data of first session (T1) of CYBHiDB and half the data of first session (S1) of UofTDB. The evaluation is carried on the other halves of the same sessions. Regarding inter-session evaluation, we evaluate the proposed method

according to evaluation setup used in the literature for the CYBHi DB [4] and in order to investigate if the proposed methods is robust to other databases, we performed an inter-class evaluation also in the UofTDB. Thus, we have the following scenarios:

INTER.CYBHiDB.T1-T2: Training on all the instances of CYBHi T1 data and evaluate on CYBHi T2. In this scenario, the instances of T1 would be considered enrollment and T2 the probe. This setting results in 1, 214, 920 intra-class pairs and 67, 316, 504 inter-class pairs after outlier removal.

INTER.CYBHiDB.T2-T1: Training on all the instances of CYBHi T2 data and evaluate on CYBHi T1. In this scenario, the instances of T2 would be considered enrollment and T2 the probe. This setting results in 1, 265, 443 intra-class pairs and 69, 050, 481 inter-class pairs after outlier removal.

INTER.UofTDB.S1-S2+S3+S4+S6: Training with all the instances of UofTDB S1 data and evaluate on the S2, S3, S4, and S6. Session 5 (S5) was discarded because our analysis considers only controlled acquisition, that is, the acquisition

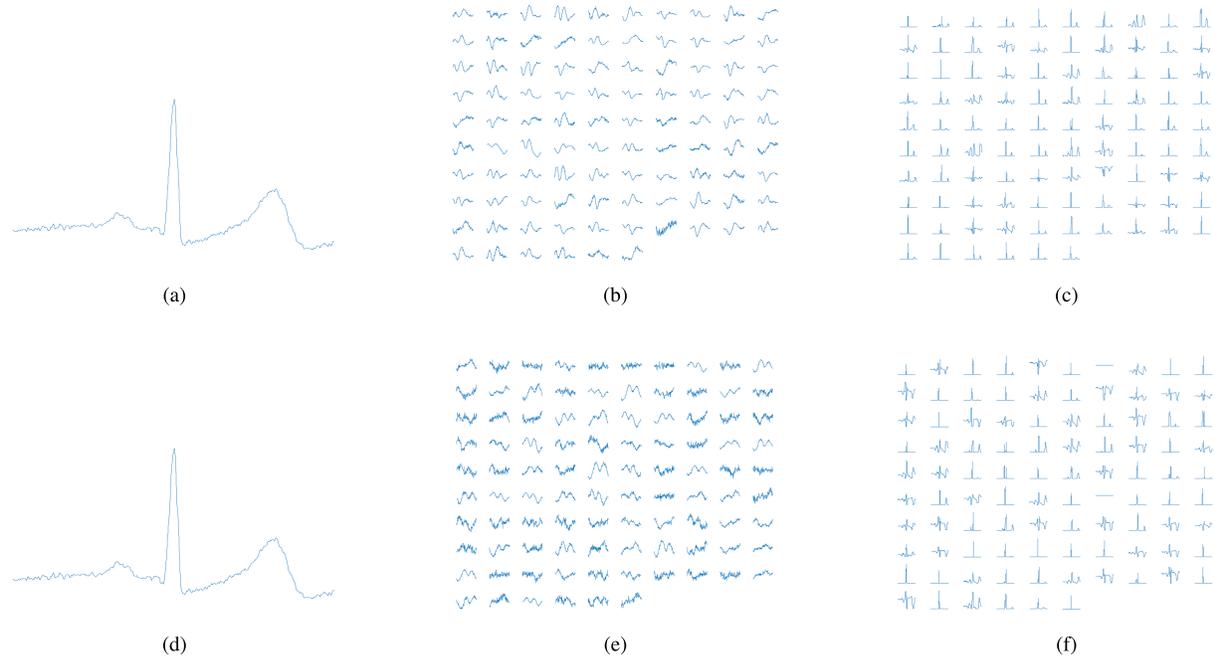


Fig. 11. Outputs for Arch B CN training with and without data augmentation. (a) Input raw ECG signal. (b) First layer filters for CN trained with data augmentation. (c) Output after fourth layer for CN trained with data augmentation. (d) Input ECG signal. (e) First layer filters for CN trained without data augmentation. (f) Output after fourth layer for CN trained without data augmentation.

made with the subject in a seated condition. Session 5 does not contemplate this condition. This setting results in 3, 510, 175 intra-class pairs and 136, 131, 670 inter-class pairs after outlier removal.

D. Results and Discussion

For both datasets, the best results were obtained with architectures B and E on data augmentation. Furthermore, mean and sum rules achieved the best results (See Fig. 9(a)) during fusion. Therefore, the sum rule was chosen for the final results presented here.

Fig. 10 and Table V show the results of proposed experimentation. Results in Fig. 10(a, b, c) are related to the CYBHi database. CYBHi database was chosen for model construction while the UofT database was reserved to evaluate the generalization of the proposed method (See Fig. 10(d)). Thus only the more realistic evaluation (inter-class) was performed for the UofT database. Results for intra-class approach can be seen in Fig. 10(a) and results for inter-class approach can be seen in Fig. 10(b, c, d).

From Fig. 10(a) it can be observed that there is a specialization of the model over data of the first session, especially the model trained with raw ECG data. However, the method still achieve good generalization in other sessions (See Fig. 10(b-d)), outperforming the six baseline methods.

The intermediate architecture B, designed for raw signal, achieved better results than architectures A and C. Our hypothesis is that there is a compromise between the amount of data available for training and depth of the network.

Fusion between the two approaches proposed here, CN trained with raw ECG data and CN trained with the image of the spectrogram, offers a significant improvement,

suggesting that the information of the two models complement each other (See Fig. 10). The fusion applied to the baseline methods also resulted in a significant improvement although not enough to overcome the results achieved by CN at raw ECG data.

Data augmentation technique proposed in this work is fundamental for CN performance as can be seen in Fig. 9(b). In Fig. 11(b, e), we have the first layer filters generated from two networks: one trained with data augmentation and another without data augmentation. It is notable that the kernel functions of the first layer of the model trained with data augmentation are sharper and well defined. This may be due to more data and consequently more training time. The output of filters in Fig. 11(c, f) suggests that the P and T waves of the ECG wave have less influence on the activation outputs of the fourth layer on data augmented model.

VI. CONCLUSIONS

In this work we investigated the use of CN to represent a single heartbeat aiming biometrics approaching the ECG raw signal and its spectrogram representation. The features generated by the CNs provided lower EER for person recognition (verification) when compared to the six baseline methods considered here, under the same evaluation protocol and in two benchmark databases for Heart biometrics.

Despite results shown here, there is still room for improvements, especially if one considers using off-the-person ECG equipment for high-security applications, such as online payment systems or bank transactions. However, the usage of multiple and consecutive ECG heartbeats (i.e., a sequence) to authenticate a subject, as proposed in [1], may further leverage the results presented here and should be investigated in

future work. Another promising research direction to improve Heart biometric systems with CN would be to build very large and representative databases.

ACKNOWLEDGMENTS

The authors thank NVIDIA Corporation for the donation of two Titan Black, one Titan X, and one Titan X Pascal GPUs.

REFERENCES

- [1] E. J. D. S. Luz, D. Menotti, and W. R. Schwartz, "Evaluating the use of ECG signal in low frequencies as a biometry," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2309–2315, 2014.
- [2] 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.
- [3] S. Wahabi, S. Pouryayeyali, 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.
- [4] 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/Oct. 2013, pp. 1–8.
- [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 Technol.*, vol. 4, no. 4, pp. 309–318, 2015.
- [6] 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, 2014.
- [7] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf.*, vol. 1, no. 3, 2015, p. 6.
- [8] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 815–823.
- [9] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2014.
- [10] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, Nov. 2011.
- [11] 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.
- [12] J. M. Irvine, B. K. Wiederhold, L. W. Gavshon, and M. D. Wiederhold, "Heart rate variability: A new biometric for human identification," in *Proc. Int. Conf. Artif. Intell.*, 2001, pp. 1106–1111.
- [13] C.-C. Chiu, C.-M. Chuang, and C.-Y. Hsu, "A novel personal identity verification approach using a discrete wavelet transform of the ECG signal," in *Proc. IEEE Int. Conf. Multimedia Ubiquitous Eng. (MUE)*, Apr. 2008, pp. 201–206.
- [14] J. M. Irvine, S. A. Israel, W. T. Scruggs, and W. J. Worek, "eigenPulse: Robust human identification from cardiovascular function," *Pattern Recognit.*, vol. 41, no. 11, pp. 3427–3435, 2008.
- [15] 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.
- [16] S.-C. Fanga and H.-L. Chanb, "QRS detection-free electrocardiogram biometrics in the reconstructed phase space," *Pattern Recognit. Lett.*, vol. 34, no. 5, pp. 595–602, 2013.
- [17] I. Odinaka *et al.*, "ECG biometrics: A robust short-time frequency analysis," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.*, Dec. 2010, pp. 1–6.
- [18] F. Agrafioti and D. Hatzinakos, "ECG based recognition using second order statistics," in *Proc. IEEE 6th Annu. Commun. Netw. Services Res. Conf. (CNSR)*, May 2008, pp. 82–87.
- [19] A. D. C. Chan, M. M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 2, pp. 248–253, Feb. 2008.
- [20] S. Islam, N. Ammour, N. Alajlan, and M. Abdullah-Al-Wadud, "Selection of heart-biometric templates for fusion," *IEEE Access*, vol. 5, pp. 1753–1761, Feb. 2017.
- [21] M. S. Islam and N. Alajlan, "Biometric template extraction from a heartbeat signal captured from fingers," *Multimedia Tools Appl.*, vol. 76, no. 10, pp. 12709–12733, 2017.
- [22] M. S. Islam, N. Alajlan, Y. Bazi, and H. S. Hichri, "HBS: A novel biometric feature based on heartbeat morphology," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 445–453, May 2012.
- [23] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [25] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2014, pp. 818–833.
- [26] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet. (2013). "Multi-digit number recognition from street view imagery using deep convolutional neural networks." [Online]. Available: <https://arxiv.org/abs/1312.6082>
- [27] C. Szegedy *et al.*, "Going deeper with convoluti," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [28] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>
- [29] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 111–118.
- [30] W. S. Geisler and D. G. Albrecht, "Cortical neurons: Isolation of contrast gain control," *Vis. Res.*, vol. 32, no. 8, pp. 1409–1410, 1992.
- [31] E. Rolls and G. Deco, *Computational Neuroscience of Vision*. Oxford, U.K.: Oxford Univ. Press, 2002.
- [32] J. B. Allen and L. Rabiner, "A unified approach to short-time Fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. 11, pp. 1558–1564, Nov. 1977.
- [33] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. (2012). "Improving neural networks by preventing co-adaptation of feature detectors." [Online]. Available: <https://arxiv.org/abs/1207.0580>
- [34] H. Silva, A. Lourenço, R. Lourenço, P. Leite, D. Coutinho, and A. Fred, "Study and evaluation of a single differential sensor design based on electro-textile electrodes for ECG biometrics applications," in *Proc. IEEE Sensors*, Oct. 2011, pp. 1764–1767.
- [35] M. S. Islam and N. Alajlan, "Augmented-Hilbert transform for detecting peaks of a finger-ecg signal," in *Proc. IEEE Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2014, pp. 864–867.
- [36] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. Int. Conf. Multimedia*, 2015, pp. 689–692.
- [37] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BEM-32, no. 3, pp. 230–236, Mar. 1985.



Eduardo José da Silva Luz received the Electrical Engineering degree from the Federal University of Minas Gerais, Belo Horizonte, Brazil, in 2005, and the master's degree in computer science from the Federal University of Ouro Preto (UFOP), Ouro Preto, Brazil, in 2012, where he is currently pursuing the Ph.D. degree. In 2017, he was nominated for an IBM Ph.D. Fellowship. He has been with the Computing Department, UFOP, as an Assistant Professor, since 2013.

His research interests include biomedical signal processing, embedded systems, pattern recognition, and machine learning.



Gladston J. P. Moreira received the master's degree in mathematics from the Federal University of Minas Gerais, Belo Horizonte, Brazil, in 2003, and the Ph.D. degree in electrical engineering, in 2011. He is currently an Associate Professor with the Computing Department, Federal University of Ouro Preto, Ouro Preto, Brazil, where he is a Collaborator Professor with the Post-Graduate Program in Computer Science.

His research interests include multi-objective optimization, pattern recognition, and spatial statistics.



Luiz S. Oliveira received the Ph.D. degree in engineering from École de Technologie Supérieure, Université du Québec, Montreal, QC, Canada, in 2003. From 2004 to 2009, he was a Professor with the Computer Science Department, Pontifical Catholic University of Parana, Curitiba, Brazil. In 2009, he joined the Federal University of Paraná, Curitiba, where he is currently a Professor with the Department of Informatics and the Head of the Graduate Program in Computer Science.

His current interests include pattern recognition and machine learning.



William Robson Schwartz received the B.Sc. and M.Sc. degrees from the Federal University of Paraná, Curitiba, Brazil, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, all in computer science. He is currently a Professor with the Department of Computer Science, Federal University of Minas Gerais, Brazil.

His research interests include computer vision, computer forensics, smart surveillance, biometrics, and image processing.



David Menotti received the Computer Engineering and Informatics Applied master's degrees from the Pontifical Catholic University of Paraná, Curitiba, Brazil, in 2001 and 2003, respectively. In 2008, he received the co-tutelage Ph.D. degree in computer science from the Federal University of Minas Gerais, Belo Horizonte, Brazil, and the Université Paris-Est/Groupe ESIEE, Paris, France. From 2008 to 2015, he was with the Computing Department, Federal University of Ouro Preto, Ouro Preto, Brazil, as an Assistant/Adjoint Professor. From

2013 to 2014, he was on his sabbatical year with the Institute of Computing, University of Campinas, Campinas, Brazil, as a post-doc/collaborator researcher. He has been an Associate Professor with the Department of Informatics, Federal University of Paraná, Curitiba, since 2015.

His research interests include pattern recognition, computer vision, and machine learning.