IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 29, 2019, accepted May 24, 2019, date of publication June 19, 2019, date of current version July 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923856

Biometric Recognition Using Multimodal Physiological Signals

SIMONE BIANCO AND PAOLO NAPOLETANO
University of Milano-Bicocca, 20126 Milan, Italy

Corresponding author: Simone Bianco (simone.bianco@unimib.it)

This work was supported in part by the project TEcnologie INnovative per i VEicoli Intelligenti (TEINVEIN), Codice Unico Progetto (CUP): E96D17000110009 - Call "Accordi per la Ricerca e l'Innovazione," cofunded by the POR FESR 2014-2020 (Programma Operativo Regionale, Fondo Europeo di Sviluppo Regionale - Regional Operational Programme, European Regional Development Fund), and in part by the project The Home of Internet of Things (Home IoT), Codice Unico Progetto (CUP): E47H16001380009–Call "Linea R&S per Aggregazioni," cofunded by the POR FESR 2014-2020.

ABSTRACT In this paper, we address the problem of biometric recognition using the multimodal physiological signals. To this end, four different signals are considered: heart rate (HR), breathing rate (BR), palm electrodermal activity (P-EDA), and perinasal perspitation (PER-EDA). The proposed method consists of a convolutional neural network that exploits mono-dimensional convolutions (1D-CNN) and takes as input a window of the raw signals stacked along the channel dimension. The architecture and training hyperparameters of the proposed network are automatically optimized with the sequential model-based optimization. The experiments run on a publicly available dataset of multimodal signals acquired from 37 subjects in a controlled experiment on a driving simulator show that our method is able to reach a top-1 accuracy equal to 88.74% and a top-5 accuracy of 99.51% when a single model is used. The performance further increases to 90.54% and 99.69% for top-1 and top-5 accuracies, respectively, if an ensemble of models is used.

INDEX TERMS Biometric identification, multimodal physiological signals, machine learning, convolutional neural network, hyperparameters optimization.

I. INTRODUCTION

Automatic identification and authentication of people is a task of enormous interest for numerous application fields, including security, domotics, automotive etc [1]. These tasks can be performed in different ways [2], [3]: typing in login credentials, digital fingerprints recognition, speech recognition, face recognition, handwriting recognition, DNA recognition, recognition using biometric data, recognition using physiological signals, recognition using inertial signals, etc. Among all these ways, the recognition of physiological signals has highlighted very interesting features. Firstly, these signals capture unique characteristics among the subjects [3] and therefore can be very robust to attempts at fraud. Secondly, physiological signals can be used for aliveness detection [4]. Thirdly, such signals can be nowadays easily acquired thanks to the rapid development of non-invasive (mostly off-the-shelf) wearable sensors [5], [6]. Last, the use of this signals can be integrated with the use of traditional authentication and identification methods in order to improve the goodness and thus reducing the errors [7]–[10].

Examples of common physiological signals adopted in authentication and identification methods are: electrocardiogram (ECG) [3], [11], electroencephalogram (EEG) [12], heart-rate (HR), heart-rate variability (HRV), pupil-dilation, blood-pressure, respiration rate (BR), Galvanic Skin Response (GSR) [13], etc. These signals can be exploited individually or in combination. In the last case, the recognition is multimodal [14], [15].

A recent paper published a large amount of physiological data and telemetries acquired by people engaged in simulated driving tasks with different levels of stress, with the aim to enable research into driving behaviors under neatly abstracted distracting stressors [16]. This data includes physiological signals of n = 68 subjects that individually drove for about 70 minutes. In particular, the available signals are: perinasal perspiration (PER-EDA), palm Electrodermal activity (P-EDA), heart rate (HR) and breathing rate (BR). Although

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

the purpose of the database was different, this is particularly suitable for the study of people identification/recognition methods.

In this paper we present a novel biometric recognition approach based on multimodal physiological signals. The proposed method uses Convolutional Neural Networks (CNNs) to extract features that permit to perform closed-set identification. This study is one of the first that exploits multimodal physiological signals in combination with CNNs for biometric recognition.

The rest of the paper is organized as follows: section II reviews the state of the art; in section III we introduce the dataset used in this work; section IV describes the proposed method; the experimental results are reported in section V, and finally section VI concludes the paper.

II. RELATED WORK

Israel et al. stated in 2005 that ECG traces express cardiac features that are unique to an individual [3]. The electrical currents generated by the heart depends on subjective features that are position and size of the heart and physical conditions of the body [17]. This findings enabled the use of ECG signals for recognition of human identities [18]–[21]. Since then, a plenty of methods based on machine learning techniques have been presented [11], [22], [23]. Most of these methods have been tested on publicly available databases that have been created for different purposes, such as automatic ECG interpretation. Some examples of databases are the PTB diagnostic ECG Database [24], Telemetric and Holter ECG Warehouse [25], MIT-BIH [26], etc. The method proposed by Labati *et al.* [11] based on deep learning achieves 100% of accuracy considering about 50 human subjects.

Galvanic Skin response (GSR) is another electrical phenomena that measure the Electrodermal Activity (EDA) of the human body. As for ECG, GSR depends on physical conditions and thus is quite subjective. GSR has been successfully employed for human identification and authentication [13], [27], [28]. Cornelius et al. experimented the use of GSR, acquired by a band worn on the wrist, for the identification and verification of 8 subjects using Support Vector Machines (SVM) [27]. Results achieved are quite good but the number of subjects experimented is quite limited. Another physiological pattern that can be employed for biometric application is the breathing rate. User authentication has been obtained by analyzing acoustic characteristics of the breathing rate [29]. The dataset experimented is composed of three types of breathing gestures: sniff (two quick consecutive inhalations), normal breathing, and deep breathing performed by 10 volunteers. The results achieved are on average about 90%.

A multimodal human identification method that exploits ECG, GSR, and airflow biosignals has been presented by Camara *et al.* [15]. The method is based on an ensemble of classifier and has been tested on database (not public) that includes 6 subjects. Performance are of about 99%.

Blasco and Peris-Lopez [14] evaluated the feasibility of using low-cost wearable sensors, such as photoplethysmogram (PPG), electrocardiogram (ECG), accelerometer (ACC), and galvanic skin response (GSR), for biometric verification. They acquired a database of 25 subjects using an home made device. The subjects involved performed three actions for a total of about 13 minutes: resting state, walking, and seated (after a gentle stroll). The proposed method obtained 0.99 area under the curve and 0.02 equal error rate with only 60 s of training data.

III. MATERIALS AND METHODS

The pipeline of a biometric system based on machine learning techniques includes: acquisition of raw signals, signal normalization and segmentation, feature extraction and classification. In the case of convolutional neural network, the last two module are merged.

A. DESCRIPTION OF THE DATABASE

The database used in this paper has been released in 2017 by Taamneh *et al.* [16]. It contains multimodal signals acquired in a controlled experiment on a driving simulator that involves $n = 68$ volunteers that drove the same highway under four different conditions: no distraction, cognitive distraction, emotional distraction, and sensorimotor distraction. The database includes also a special driving session, where all subjects experienced a startle stimulus in the form of unintended acceleration-half of them under a mixed distraction, and the other half in the absence of a distraction. Fig. 1 shows the driving simulator set up and some samples extracted from physiological signals recorded during the driving session.

The signals included in the database are speed, acceleration, brake force, steering, and lane position signals, perinasal perspiration (PER-EDA), palm EDA (P-EDA), heart rate (HR), breathing rate (BR), and facial expression signals. Moreover, biographical and psychometric covariates as well as eye tracking data were also obtained. Perinasal perspiratory signals commensurate with electrodermal (EDA) in the palm, for this reason is called perinasal EDA (PER-EDA) [30].

The database includes two sessions of experiments:

- EXPERIMENT-I: focused on the effect of cognitive, emotional and sensorimotor stressors on driving behaviors under typical conditions.
- EXPERIMENT-II: focused on the effect of stress on reactivity to a startling event while driving; this startling event was unintended acceleration.

In this paper we focus on the EXPERIMENT-I and we consider four signals: PER-EDA, P-EDA, HR and BR. Perinasal perspiratory signals have been extracted from the thermal facial videos using the S-Interface software [16] applied to a region of interest. Palm EDA has been acquired with the Shimmer3 GSR sensor (Shimmer, Dublin, Ireland) that has a measurement range of 10-4700 kΩ. HR and BR signals have been acquired with the adrenergic sensor Zephyr BioHarness 3.0 (Zephyr Technology, Annapolis, MD). The sensor detects a heart rate range of 25-240 bpm and a breathing rate range of 4-70 bpm. The adrenergic sensor was connected to a chest

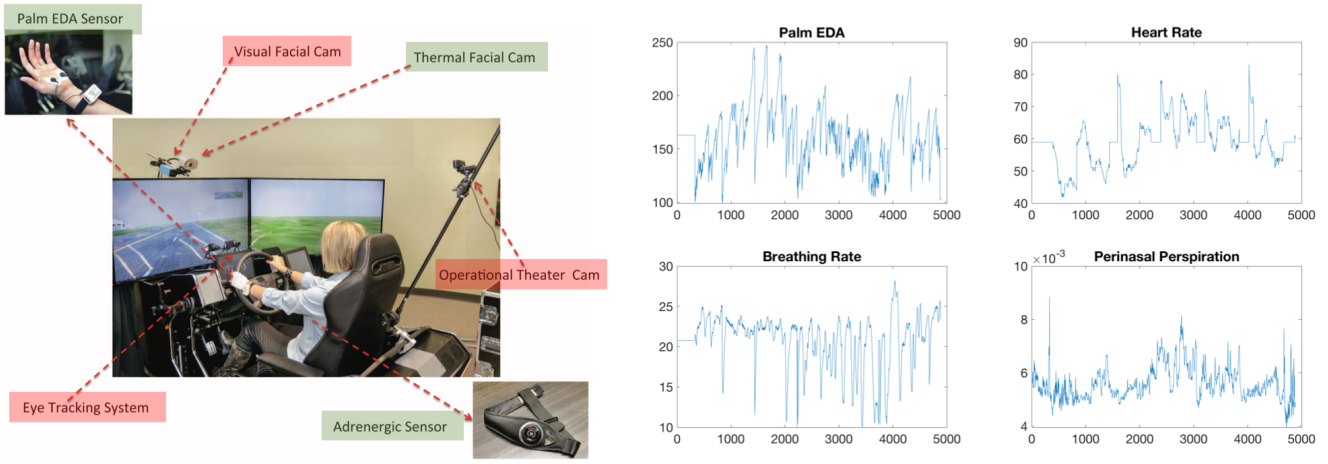

FIGURE 1. Left side of the figure: Image depicting the experimental setup adopted in [16] to record the physiological signals: Highlighted in green are the sensors used to acquire the signals actually used in this paper. Right side of the figure: Graphical representation of the physiological signals acquired during a driving session (of about 5000 seconds) of the experiments presented in [16] and used in this paper. Left side of this figure is courtesy of authors of the paper [16], and it is licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

strap that was worn underneath the subject's clothing. Fig. 1 shows the PER-EDA, P-EDA, HR and BR signal related to the Experiment-I and Experiment-II of the subject 1.

1) EXPERIMENT-I

This experiment is composed of 7 subsequent sessions:

- Baseline (B): subjects sat quietly in a dimly lit room, listening to soothing music;
- Practice drive (PD): the subjects familiarized themselves with the driving simulator;
- Relaxing drive (RD): the subjects had to drive following some instructions;
- Loaded Drive (LD): Driving with no secondary activity (no additional stressor);
- Cognitive Drive (CD): Driving under a cognitive stressor. The cognitive stressor was mathematical questions;
- Emotional drive (ED): Driving under an emotional stressor. The emotional stressor was emotionally stirring questions posed orally by the experimenter in two phases. There were two sets of questions: a set with less pointed questions and a set with more pointed questions
- Sensorimotor drive (MD): Driving under a sensorimotor stressor. The sensorimotor stressor was texting back words, sent one by one to the subject's smartphone; this texting exchange took place in two phases.

The last three phases were developed alternating phases with distraction and phases without distractions. We decide to discard the session B because the physiological signals related to this session are not available.

B. DATA CLEANING, NORMALIZATION AND SEGMENTATION

A cleaning procedure has been applied to the database with the aim of removing outliers signals from the dataset and removing those subjects with missing or not valid signal values. We removed from the database all the signals that had a peak of the HR, BR and P-EDA respectively outside the numeric ranges [40-120], [4-40], and [28-628] for at least 30% of the signal duration. In case of an amount of invalid values below than 30%, such invalid values have been substituted by the mean of the signal.

To the scope of this paper, we have explored the use of HR, BR, PER-EDA and P-EDA to biometric recognition. Since, multimodal recognition is possible only if, for each time instant, a valid value of each one of the four aforementioned physiological signals exists, we discard all the subjects with missing or not valid signals in any of the four signals. The initial number of subjects available is 68, after the cleaning and discarding procedures, the number of subjects is 37.

Before being processed, each signal is segmented by using a sliding window of size W and overlap L. An overlap of 50% means that two subsequent windows has 50% of their values in common. After having investigated several values for W ranging from 50 to 400 seconds and three values of L ranging from 25 to 75 % we decide to set $W = 60$ seconds and $L = 50\%$.

IV. PROPOSED METHOD

The proposed method is a Convolutional Neural Network that exploits mono-dimensional convolutions (1D-CNN). The proposed 1D-CNN takes as input the raw physiological signals in windows of $W = 60$ samples (as described in section III) and performs multimodal learning by concatenating the four monodimensional signals considered along the channel dimension, resulting in an input size equal to $1 \times 60 \times 4$.

Given the very limited amount of training data available, the architecture of the proposed 1D-CNN is rather shallow: the input layer is followed by a first convolutional layer conv-1 with ReLU activation and pooling layer MaxPool-1.

TABLE 1. Detailed architecture of the proposed 1D-CNN.

Layer	Size	Stride/Padding	Output size
INPUT	$1 \times 60 \times 4$		
CONV-1	n_1@$1 \times 5 \times 4$	1/0	$1 \times 56 \times n_1$
RELU			$1 \times 56 \times n_1$
MAXPOOL-1	1×2	2/0	$1 \times 28 \times n_1$
CONV-2	n_2@$1 \times 5 \times n_1$	1/0	$1 \times 28 \times n_2$
RELU			$1 \times 28 \times n_2$
MAXPOOL-2	1×2	2/0	$1 \times 14 \times n_2$
DROPOUT-1	prob.$= p_d$		$1 \times 14 \times n_2$
FC-1	$14 \cdot n_2 \times h_1$		h_1
RELU			h_1
DROPOUT-2	prob.$= p_d$		h_1
FC-2	$h_1 \times 37$		37
RELU			37
SOFTMAX			37

TABLE 2. Numerical ranges and data type of the hyperparameters optimized with SMBO.

Hyperparameter	Range	Type
n_1 (CONV-1)	[4, 32]	integer
n_2 (CONV-2)	[4, 32]	integer
p_d (DROPOUT-1,DROPOUT-2)	[0.3, 0.7]	continuous
h_1 (FC-1)	[10, 200]	integer
augmRounds	[0, 5]	integer
augmNoise	[1e-5, 1e-2]	continuous

This is followed by a second convolutional layer conv-2 with ReLU activation [31] and pooling layer MaxPool-2. Then, we have two fully connected layers fc-1 and fc-2, each with ReLU activation, followed by the softmax classification layer. In order to reduce overfitting, each fully connected layer is preceded by a dropout layer. To further deal with the limited training data available, we perform offline data augmentation by replicating the training set by adding Gaussian noise. The detailed architecture of the proposed 1D-CNN is reported in Table 1 with the size of each layer, its eventual stride and padding, and the size of its activation map. From the architecture it possible to see that it has some free hyperparameters. In the following subsection we describe how we optimize them.

A. HYPERPARAMETERS OPTIMIZATION

Using the architecture defined in the previous section as the backbone, we want to optimize the values of the following hyperparameters:

- n_1: the number of filters in conv-1;
- n_2: the number of filters in conv-2;
- p_d: the dropout probability of dropout-1 and dropout-2;
- h_1: the number of neurons in fc-1;
- *augmRounds*: the number of data augmentation rounds performed (i.e. number of augmented versions of the training set);
- *augmNoise*: standard deviation of the Gaussian noise used for data augmentation.

TABLE 3. Hand-crafted features used in the preliminary experiment and extracted from each of the four physiological considered (HR, BR, P-EDA, and PER-EDA) on windows with width $W = 60$ seconds.

Features	
Minimum	$min = \min_{j=1,\dots n}(x_j)$
Maximum	$max = \max_{j=1,\dots n}(x_j)$
Mean	$\mu = \frac{1}{n}\sum_{j=1}^{n} x_j$
Median	$Me = x_{0.5} : F(x_{0.5}) = 0.5$
Standard Deviation (SD)	$\sigma = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j - \mu)^2}$
Variance	$\sigma^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu)^2$
Fourth central moment	$m_4 = E(x - \mu)^4$
Fifth central moment	$m_5 = E(x - \mu)^5$
Skewness	$S = \frac{E(x-\mu)^3}{\sigma^3}$
Kurtosis	$K = \frac{E(x-\mu)^4}{\sigma^4}$
Root Mean Square (RMS)	$RMS = \sqrt{\frac{1}{n}\sum_{j=1}^{n} x_j^2}$
Interquartile Difference	$ID = x_{0.75} - x_{0.25}$
Total Sum	$TS = \sum_{j=1}^{n} x_j$
Range	$R = max - min$
Entropy	$H(x) = -\sum_{j=1}^{n} p(x_j)\log_2 p(x_j)$
SD of the intervals between two successive peaks	$SDNN = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(PP_j - \overline{PP})^2}$
RMS of the differences between two successive peaks	$RMSSD = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n-1}(PP_{j+1} - PP_j)^2}$
Number of pairs of successive peaks intervals that differ by more than 50 ms	$pNN50 = p(\lvert PP_{j+1} - PP_j\rvert > 50)$
Sum of the spectral power components	$SP = \frac{1}{n}\sum_{j=1}^{f} \lvert FFT_j\rvert^2$
Mean of the spectral components	$\mu_f = \frac{1}{n}\sum_{j=1}^{n} FFT_j$
Median of the spectral components	$Me_f = FFT_{0.5} : F(f_{0.5}) = 0.5$

To this end, Sequential Model-Based Global Optimization (SMBO) [32]–[34] is used. SMBO has been used in many applications where the evaluation of the fitness function $f : X \rightarrow \mathbb{R}$ is expensive to be evaluated. SMBO works by approximating f with a surrogate that is cheaper to evaluate and optimize. New locations within the hyperparameters domain X on which the original function f should be evaluated are sequentially selected by optimizing an acquisition function S, that defines a balance between exploration and exploitation, i.e. between exploring new areas in the objective space and exploiting areas that are already known to provide good values. In this work we use Gaussian process (GP) [35] as the surrogate to model the objective function, expected improvement [36] as the acquisition function, and top-1 accuracy as the fitness function. The maximum number of function evaluation is set to 70. The search ranges for the hyperparameters considered and their types are reported in Table 2. All the CNN configurations explored by SMBO are trained for a total of 4000 epochs, with a batch size of 512, using the Stochastic Gradient Descent with Momentum (SGDM) as optimizer [37], with initial Learning Rate equal to 0.1 and a piece-wise decay policy with the reduction by a 0.5 factor every 800 epochs.

V. EXPERIMENTAL RESULTS

In this section we report the performance of biometric identification from physiological signals in terms of top-1, top-3 and top-5 accuracy, i.e. we measure the number of times that the right identity is the prediction given by

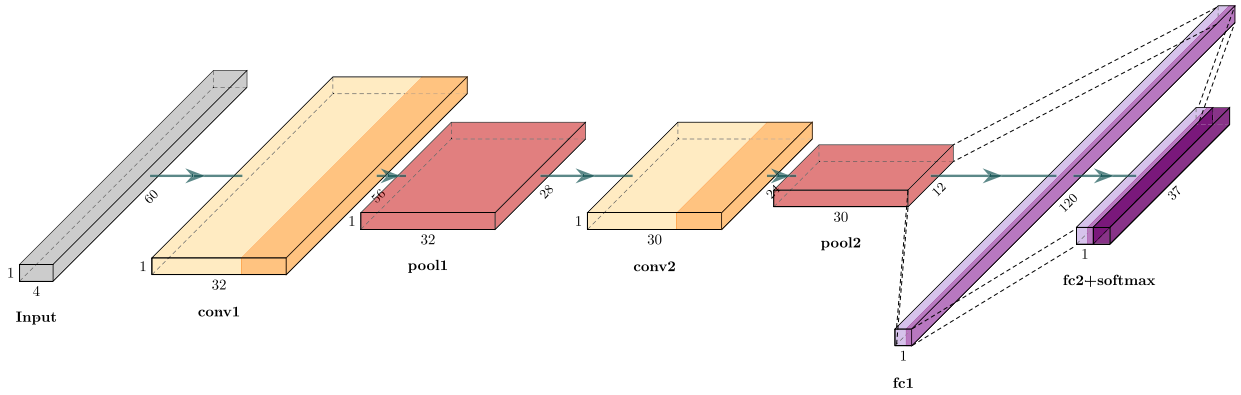

FIGURE 2. Activation maps size of each layer of the proposed 1D-CNN with the hyperparameters optimized by SMBO.

TABLE 4. Results of the preliminary experiment for biometric identification from HR, BR, P-EDA, and PER-EDA physiological signals in terms of top-1 (a), top-3 (b), and top-5 (c) micro accuracy. For each table the global best results is reported in bold, while the best result for each column is underlined. The results are reported by performing a five-fold cross-validation.

	HR	BR	P-EDA	PER-EDA	ALL
KNN	44.76 (± 2.19)	12.03 (± 1.35)	29.51 (± 0.70)	16.83 (± 1.22)	78.97 (± 1.94)
SVM	49.69 (± 2.07)	12.57 (± 1.85)	36.32 (± 1.77)	19.77 (± 1.26)	82.74 (± 1.40)
ANN	51.43 (± 2.28)	14.94(± 1.15)	41.37 (± 1.94)	22.63 (± 1.33)	82.90 (± 1.11)
STACK	51.46 (± 2.16)	13.54 (± 1.60)	38.34 (± 1.10)	20.77 (± 1.28)	**85.50** (± 1.24)

a)

	HR	BR	P-EDA	PER-EDA	ALL
KNN	68.22 (± 1.13)	25.73 (± 1.37)	51.13 (± 0.75)	34.52 (± 1.46)	91.11 (± 0.94)
SVM	72.14 (± 2.15)	26.03 (± 2.12)	57.29 (± 0.53)	36.80 (± 1.65)	94.94 (± 0.57)
ANN	82.91 (± 1.58)	33.35 (± 2.29)	69.83 (± 2.65)	44.73 (± 1.43)	95.52 (± 0.61)
STACK	75.38 (± 1.64)	25.98 (± 1.79)	59.41 (± 0.57)	36.62 (± 1.77)	**96.37** (± 0.55)

b)

	HR	BR	P-EDA	PER-EDA	ALL
KNN	77.58 (± 0.98)	34.15 (± 1.30)	62.84 (± 1.94)	44.64 (± 2.16)	94.80 (± 0.86)
SVM	75.44 (± 1.29)	33.39 (± 1.56)	62.80 (± 0.68)	44.60 (± 1.62)	97.02 (± 0.19)
ANN	92.56 (± 1.01)	45.54 (± 1.99)	82.20 (± 2.28)	59.04 (± 1.63)	**98.24** (± 0.50)
STACK	83.06 (± 1.09)	35.81 (± 1.58)	70.04 (± 1.22)	47.89 (± 1.53)	98.17 (± 0.39)

c)

TABLE 5. Best hyperparameter configuration found by SMBO for the proposed 1D-CNN architecture.

Hyperparameter	Value
-Number of filters in CONV-1	$n_1 = 32$
-Number of filters in CONV-2	$n_2 = 30$
-Dropout probability	$p_d = 0.6213$
-Number of neurons in FC-1	$h_1 = 120$
-Number of data augmentation rounds	$augmRounds = 4$
-Standard deviation of Gaussian noise for data augmentation	$\sigma = 3.8701e-4$

the model with the highest probability, or it is respectively within the three or five predictions with the highest probabilities. The performance measures are computed on the whole dataset by performing a five-fold cross-validation.

First we run a preliminary experiment by extracting commonly used features [38] on each of the four physiological signals considered (i.e. HR, BR, P-EDA, and PER-EDA). Whatever is the signal, 21 different features are extracted from each segment of length W. The extracted features, reported in Table 3, describe different properties of the signal in both the time and frequency domains. For each fold, the training features are normalized in a way that each feature component ranges between 0 and 1. The test features are then normalized using the values adopted to normalize the training features. Three different classifiers are trained to classify both the features extracted from the individual signals, and the concatenation of the features extracted from all the signals: k-NN, Support Vector Machines (SVM), and Artificial Neural Networks (ANN).

The k-NN uses Euclidean distance and $k = 1$, thus resulting in the 1-nearest neighbor classifier.

The SVMs used have a Radial Basis Function (RBF) kernel. The optimal values of C and γ are found by grid search.

Due to the small quantity of data available the ANNs used have a shallow architecture. The hidden layers are unsupervisedly trained as Autoencoders, stacked together

TABLE 6. Results for the biometric identification with multimodal physiological signals, i.e. HR, BR, P-EDA, and PER-EDA, in terms of top-1 (a), top-3 (b), and top-5 (c) micro accuracy. For each performance metric the best result is reported in bold. The results are reported by performing a five-fold cross-validation.

Method	Top-1 acc.	Top-3 acc.	Top-5 acc.
KNN	78.97 (± 1.94)	91.11 (± 0.94)	94.80 (± 0.86)
SVM	82.74 (± 1.40)	94.94 (± 0.57)	97.02 (± 0.19)
ANN	82.90 (± 1.11)	95.52 (± 0.61)	98.24 (± 0.50)
STACK	85.50 (± 1.24)	96.37 (± 0.55)	98.17 (± 0.39)
proposed 1D-CNN	88.74 (± 0.52)	98.59 (± 0.35)	99.51 (± 0.17)
proposed 1D-CNN(SE)	**90.54** (± 0.92)	**98.81** (± 0.42)	**99.69** (± 0.05)

and are followed by a softmax layer. The number of hidden layers and the number of neurons in each of them, i.e. H_1 and H_2, are selected by grid search in the range [0, 5, 10, 15, 20, 30, 40, 50] for individual features and [0, 5, 10, 15, 20, 40, 60, 80, 100, 120, 140, 160] for concatenated features. The maximum number of hidden layers is fixed to two. Given the possible ranges for H_1 and H_2, if $H_1 = 0$ or $H_2 = 0$ the resulting ANN has just one hidden layer. If $H_1 = H_2 = 0$, the resulting ANN has no hidden layers and the input features are directly fed to the softmax layer.

As an additional classifier we consider the stacked classifier, that performs the classification on the basis of the predictions of the k-NN, SVM and KNN through majority vote.

The results in terms of top-1, top-3, and top-5 micro accuracy are reported in Table 4. From the results it is possible to see that the overall best results are obtained by using the concatenation of the features extracted from all the four signals considered and using a stacked classifier, reaching a top-1 accuracy of 85.50%. The physiological signals that permit to obtain a higher recognition rate are HR and P-EDA reaching a top-1 accuracy of 51.46% and 41.37% respectively. Considering top-3 and top-5 accuracy, the best results are again obtained using the stacked classifier on the concatenation of all the features, reaching respectively a recognition rate of 96.37% and 98.24%.

In the second experiment we compare the best solutions identified in the preliminary experiment with the proposed method. The best hyperparameters found by SMBO for the proposed 1D-CNN are reported in Table 5, while a graphical representation if its activation maps is reported in Figure 2. The resulting 1D-CNN has a total of 60280 parameters to be trained. The total time required by SMBO using a single NVIDIA Titan V GPU is about 20.5 hours (approx. 17.5 minutes for evaluation), with the minimum time for an evaluation equal to approx. 4 minutes and the maximum equal to approx. 24 minutes.

The recognition rate in terms of top-1, top-3, and top-5 micro accuracy are reported in Table 6. From the results it is possible to see that the proposed method is able to improve in terms of all the three metrics considered with respect to both the individual classifiers and the stacked one fed with hand-crafted features [38]. As a reference the improvement in top-1 accuracy is 5.84% with respect to the best individual

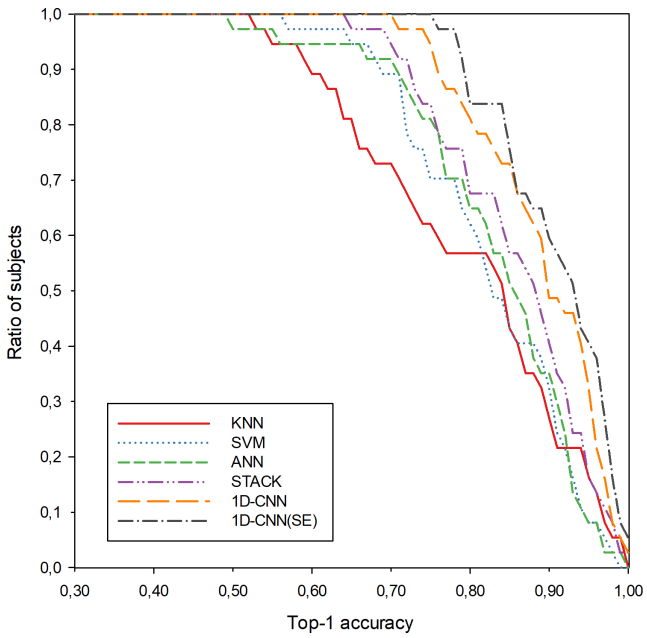

FIGURE 3. Curves representing the ratio of subjects having a top-1 accuracy above a given threshold in the range [0.3, 1.0] for the methods compared in Table 4.

classifier (i.e. ANN) and 3.24% with respect to the stacked one. With a top-1 accuracy density [39], i.e. top-1 accuracy divided by the number of trainable parameters (in millions), of 1472.1 we can assert that the proposed 1D-CNN uses its parameters very efficiently.

Since the best results in the preliminary experiment are obtained by using a stacked classifier, we investigate if an ensemble of 1D-CNNs is able to improve the results of the single 1D-CNN proposed. The ensemble is build taking inspiration from the Snapshot Ensemble (SE) technique [40]. In SE, the different members of the ensemble are obtained at different epochs of the training of the same model. In this work instead, we create the ensemble by taking the members from the list of the models evaluated by SMBO for the hyperparameter optimization. The considered ensemble is composed by three elements: the best model identified by SMBO and the two before it. From the results reported in Table 6 it is possible to see that this ensemble is able to improve the top-1 accuracy of the proposed 1D-CNN by 1.8%, reaching a recognition rate of 90.54%. Furthermore we can observe that for both the single-model 1D-CNN and

the ensemble 1D-CNN(SE) the top-3 and top-5 accuracy are higher than 98.5% and 99.5% respectively.

As a further comparison we report in Figure 3 a plot with the ratio of subjects having a top-1 accuracy above a given threshold in the range [0.3, 1.0]. From the plot is possible to notice how the curves of the proposed methods are highest ones for all the thresholds. In particular, 1D-CNN and 1D-CNN(SE) are able to achieve a top-1 accuracy higher than 99% for the 5.4% and 8.1% of the users respectively.

VI. CONCLUSION

In this paper we addressed the problem of biometric recognition using multimodal physiological signals. Four different physiological signals have been considered: of heart rate (HR), breathing rate (BR), palm Electrodermal Activity (P-EDA), and perinasal perspitation (PER-EDA).

The proposed solution exploited a Convolutional Neural Network with mono-dimensional convolutions (1D-CNN), taking as input a window of the raw signals stacked along the channel dimension. Both the architecture and training hyperparameters of the proposed solution were automatically optimized with Sequential Model-Based Optimization (SMBO).

Experiments run on a publicly available dataset of multimodal signals acquired from 37 subjects in a controlled experiment on a driving simulator show that our method is able to reach a top-1 accuracy equal to 88.74% and a top-5 accuracy of 99.51% when a single model is used, and 90.54% and 99.69% for top-1 and top-5 accuracy respectively if an ensemble of three models is used.

As a future work we plan to increase the dataset both in terms of subjects involved and in terms of physiological signals considered, in order to be able to exploit deeper architectures inspired by famous 2D-CNNs. We also plan to study the combination of physiological signal with audio and image/video signals for more robust biometric recognition.

ACKNOWLEDGMENTS

The authors would like to thank the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research.

REFERENCES

[1] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*. New York, NY, USA: Springer, 2007.

[2] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[3] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern Recognit.*, vol. 38, no. 1, pp. 133–142, Jan. 2005.

[4] Y. Wang, K. N. Plataniotis, and D. Hatzinakos, "Integrating analytic and appearance attributes for human identification from ECG signals," in *Proc. Biometrics Symp., Special Session Res. Biometric Consortium Conf.*, Aug. 2006, pp. 1–6.

[5] M. M. Rodgers, V. M. Pai, and R. S. Conroy, "Recent advances in wearable sensors for health monitoring," *IEEE Sensors J.*, vol. 15, no. 6, pp. 3119–3126, Jun. 2015.

[6] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 1, pp. 1–12, Jan. 2010.

[7] S. Kun, Y. Gongping, W. Bo, Y. Lu, L. Dunfeng, S. Peng, and Y. Yilong, "Human identification using finger vein and ECG signals," *Neurocomputing*, vol. 332, pp. 111–118, Mar. 2019.

[8] M. Hammad, Y. Liu, and K. Wang, "Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint," *IEEE Access*, vol. 7, pp. 26527–26542, 2018.

[9] M. Komeili, N. Armanfard, and D. Hatzinakos, "Liveness detection and automatic template updating using fusion of ECG and fingerprint," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 7, pp. 1810–1822, Jul. 2018.

[10] S. A. Israel, W. T. Scruggs, W. J. Worek, and J. M. Irvine, "Fusing face and ECG for personal identification," in *Proc. 32nd Appl. Imag. Pattern Recognit. Workshop*, Oct. 2003, pp. 226–231.

[11] R. D. Labati, E. Muñoz, V. Piuri, R. Sassi, and F. Scotti, "Deep-ECG: Convolutional neural networks for ECG biometric recognition," *Pattern Recognit. Lett.*, to be published. doi: 10.1016/j.patrec.2018.03.028.

[12] S. Yang, F. Deravi, and S. Hoque, "Task sensitivity in EEG biometric recognition," *Pattern Anal. Appl.*, vol. 21, no. 1, pp. 105–117, 2016.

[13] I. Martinovic, K. B. Rasmussen, M. Roeschlin, and G. Tsudik, "Pulse-response: Exploring human body impedance for biometric recognition," *ACM Trans. Privacy Secur.*, vol. 20, no. 2, p. 6, 2017.

[14] J. Blasco and P. Peris-Lopez, "On the feasibility of low-cost wearable sensors for multi-modal biometric verification," *Sensors*, vol. 18, no. 9, p. 2782, 2018.

[15] C. Camara, P. Peris-Lopez, J. E. Tapiador, and G. Suarez-Tangil, "Non-invasive multi-modal human identification system combining ECG, GSR, and airflow biosignals," *J. Med. Biol. Eng.*, vol. 35, no. 6, pp. 735–748, Nov. 2015.

[16] S. Taamneh, P. Tsiamyrtzis, M. Dcosta, P. Buddharaju, A. Khatri, M. Manser, T. Ferris, R. Wunderlich, and I. Pavlidis, "A multimodal dataset for various forms of distracted driving," *Sci. Data*, vol. 4, Aug. 2017, Art. no. 170110.

[17] B. P. Simon and C. Eswaran, "An ECG classifier designed using modified decision based neural networks," *Comput. Biomed. Res.*, vol. 30, no. 4, pp. 257–272, Aug. 1997.

[18] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis, "Analysis of human electrocardiogram for biometric recognition," *EURASIP J. Adv. Signal Process.*, vol. 2008, 2007, Art. no. 148658.

[19] K. Revett, F. Deravi, and K. Sirlantzis, "Biosignals for user authentication-towards cognitive biometrics?" in *Proc. Int. Conf. Emerg. Secur. Technol.*, Sep. 2010, pp. 71–76.

[20] J. Blasco, T. M. Chen, J. Tapiador, and P. Peris-Lopez, "A survey of wearable biometric recognition systems," *ACM Comput. Surv.*, vol. 49, no. 3, p. 43, Dec. 2016.

[21] J. R. Pinto, J. S. Cardoso, and A. Lourenço, "Evolution, current challenges, and future possibilities in ECG biometrics," *IEEE Access*, vol. 6, pp. 34746–34776, 2018.

[22] N. Venkatesh and S. Jayaraman, "Human electrocardiogram for biometrics using DTW and FLDA," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3838–3841.

[23] 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.

[24] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, and R. Mark, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220. 2000.

[25] J.-P. Couderc, "The Telemetric and Holter ECG Warehouse (THEW): The first three years of development and research," *J. Electrocardiol.*, vol. 45, no. 6, pp. 677–683, 2012.

[26] G. B. Moody and R. G. Mark, "The MIT-BIH arrhythmia database on CD-ROM and software for use with it," in *Proc. Comput. Cardiol.*, Sep. 1990, pp. 185–188.

[27] C. Cornelius, R. Peterson, J. Skinner, R. Halter, and D. Kotz, "A wearable system that knows who wears it," in *Proc. 12th Annu. Int. Conf. Mobile Syst., Appl. Services*, 2014, pp. 55–67.

[28] A. Bianchi and I. Oakley, "Wearable authentication: Trends and opportunities," *Inf. Technol.*, vol. 58, no. 5, pp. 255–262, Oct. 2016.

[29] J. Chauhan, Y. Hu, S. Seneviratne, A. Misra, A. Seneviratne, and Y. Lee, "BreathPrint: Breathing acoustics-based user authentication," in *Proc. 15th Annu. Int. Conf. Mobile Syst. Appl. Services*, 2017, pp. 278–291.

[30] D. Shastri, M. Papadakis, P. Tsiamyrtzis, B. Bass, and I. Pavlidis, "Perinasal Imaging of physiological stress and its affective potential," *IEEE Trans. Affect. Comput.*, vol. 3, no. 3, pp. 366–378, Jul. 2012.
[31] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
[32] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 2546–2554.
[33] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Proc. Int. Conf. Learn. Intell. Optim.* Springer, 2011, pp. 507–523.
[34] J. Bergstra, D. Yamins, and D. D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *Proc. 30th Int. Conf. Int. Conf. Mach. Learn.*, 2013, pp. I-115–I-123
[35] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, vol. 2. Cambridge, MA, USA: MIT Press, no. 3, 2006,
[36] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, no. 4, pp. 455–492, 1998.
[37] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*. [Online]. Available:https://arxiv.org/abs/1609.04747
[38] S. Bianco, P. Napoletano, and R. Schettini, "Multimodal car driver stress recognition," in *Proc. 13th EAI Int. Conf. Pervasive Comput. Technol. Healthcare*, 2019, pp. 302–307.
[39] S. Bianco, R. Cadene, L. Celona, and P. Napoletano, "Benchmark analysis of representative deep neural network architectures," *IEEE Access*, vol. 6, pp. 64270–64277, 2018.
[40] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft, and K. Q. Weinberger, "Snapshot ensembles: Train 1, get M for free," 2017, *arXiv:1704.00109*. [Online]. Available: https://arxiv.org/abs/1704.00109

SIMONE BIANCO received the B.Sc. and M.Sc. degrees in mathematics and the Ph.D. degree in computer science from the University of Milano-Bicocca, Italy, in 2003, 2006, and 2010, respectively, where he is currently an Assistant Professor of computer science with the Department of Informatics, Systems and Communication. His current research interests include computer vision, machine learning, optimization algorithms, and color imaging.

PAOLO NAPOLETANO received the master's degree in telecommunications engineering from the University of Naples Federico II, in 2003, and the Ph.D. degree in information engineering from the University of Salerno. He is currently an Assistant Professor of computer science with the Department of Informatics, Systems and Communication, University of Milano-Bicocca. His current research interest includes machine learning for multi-modal data classification and understanding.