Deep Learning for EEG-based Biometric Recognition

Emanuele Maiorana

Section of Applied Electronics, Department of Engineering, Roma Tre University, Via V. Volterra 62, 00146 Roma, Italy; e-mail: emanuele.maiorana@uniroma3.it, phone: +39.06.57337365

Abstract

The exploitation of brain signals for biometric recognition purposes has received significant attention from the scientific community in the last decade, with most of the efforts so far devoted to the quest for discriminative information within electroencephalography (EEG) recordings. Yet, currently-achievable recognition rates are still not comparable with those granted by more-commonly-used biometric characteristics, posing an issue for the practical deployment of EEG-based recognition in real-life applications. Within this regard, the present study investigates the effectiveness of deep learning techniques in extracting distinctive features from EEG signals. Both convolutional and recurrent neural networks, as well as their combinations, are employed as strategies to derive personal identifiers from the collected EEG data. In order to assess the robustness of the considered techniques, an extensive set of experimental tests is conducted under very challenging conditions, trying to determine whether it is feasible to identify subjects through their brain signals regardless the performed mental task, and comparing acquisitions collected at a temporal distance greater than one year. The obtained results suggest that the proposed networks are actually able to exploit the dynamic temporal behavior of EEG signals to achieve high-level accuracy for brain-based biometric recognition.

Keywords: Biometrics, Electroencephalography, Deep Learning, Convolutional Neural Networks, Recurrent Neural Networks.

1. Introduction

Biometric recognition systems are nowadays deployed in many real-life applications, ranging from border control to verification on a mobile device. Although several biometric traits have already achieved the plateau of their productivity, such as face, fingerprint, and iris, prominent researches are still carried out to design innovative recognition systems, relying on other characteristics, which may possess useful properties not available in mainstream solutions. In this regard, in addition to the usage of physical or behavioral traits, cognitive biometrics [1], involving the acquisition of biosignals generated by the nervous system while performing a specific task, has been recently proposed for high-security applications. Notable examples belonging to this category include the heart activity [2], the skin electrodermal activity [3], and also the brain activity [4]. Exploiting such characteristics as biometric identifiers allows to inherently guarantee liveness detection. Moreover, recording the aforementioned activities at a distance is impossible, thus making it therefore very hard to implement presentation attacks on systems based on them.

As far as brain activity is concerned, among the possible solutions available to record the associated signals, such as functional magnetic resonance imaging (fMRI), near-infrared spectroscopy (NIRS), positron emission tomography (PET), magnetoencephalography (MEG), and electroencephalography (EEG), this latter has been mostly considered in the context of biometric recognition, due to the relative inexpensiveness of the associated acquisition devices, and their ease of use [5]. However, despite the fact that the first proposals of using EEG signals for biometric recognition purposes date back to the 80s [6], several issues still have to be properly addressed before EEG-based biometric systems could be realistically deployed in practical applications [7]. From a hardware point of view, it would be in fact desirable to design innovative acquisition devices, which could allow to collect brain signals with good signal-to-noise ratio (SNR) without being too cumbersome and uncomfortable to be worn. Currently, wet electrodes are commonly preferred to measure the electric potential difference between distinct positions on the scalp surface, yet they require to put electrolyte gel on the scalp, thus causing subject inconvenience and a non-negligible setup time. Alternatives based on dry electrodes exists, yet the quality of the acquired signals is still not comparable with the one guaranteed by wet electrodes, and the typically-used electrode shape may cause ache for prolonged acquisitions [8].

More importantly, novel approaches are required to extract distinctive features from the collected EEG signals, in order to evaluate whether recognition rates comparable with those achieved through other well-established biometric traits can be guaranteed. Although several studies have mentioned the possibility of reaching perfect recognition performance when using EEG-based identifiers [9, 10], their reliability is highly questionable, since data collected from few users, or during a single acquisition session, are often employed in such evaluations [11]. As deeply discussed in [12] and [13], a proper analysis of the discriminative capabilities of EEG signals has to be conducted using data collected during multiple acquisition sessions, in order to avoid that the estimated performance depends on session-specific exogenous conditions, instead of only on the personal characteristics of the involved subjects. When this precaution is taken into account, the attainable recognition rates significantly worsen, testifying the need for further research on the extraction of discriminative features from EEG signals [14].

Taking the aforementioned aspects into account, the effectiveness of employing deep learning techniques to process EEG signals for biometric recognition purposes is investigated in this paper, in order to assess whether their usage could allow achieving high-level recognition performance, even under challenging realistic scenarios. Specifically, the major contributions of the reported study include:

- the proposal of several novel neural network architectures, exploiting convolutional, recurrent, as well as combinations of such kinds of network, to be applied on different representations of the considered EEG data. The comparison between the performance achievable through the proposed networks, and through those already presented in literature, sheds some light on which network characteristics should be considered to attain high recognition accuracy in EEG-based recognition systems;

- the exploitation of hand-crafted features, extracted from the recorded brain activity, as input to neural networks for recognition purposes. As outlined in Section

2, state-of-the-art biometric approaches applying deep learning to EEG signals have so far exploited only raw temporal data as input to the proposed networks. The tests here described have been instead conducted to investigate whether, and under which conditions, it could be useful to feed feature representations of EEG data to the considered networks, instead of applying these latter directly to raw EEG signals;

- the usage of the largest multi-session dataset ever employed to evaluate an EEG-based biometric recognition systems, in terms of enrolled subjects, employed elicitation protocols, number of recording sessions, and temporal distance between enrolment and recognition stages, to test the proposed deep learning approaches. As outlined in Section 2, the applications of deep learning for EEG-based biometric recognition has been so far investigated considering datasets comprising signals collected during at most two sessions, from a number of users in the order of a dozen, while performing a single task. The tests here reported have been instead conducted on a database comprising EEG recordings taken from 45 users during five sessions, for each of which six distinct acquisition protocols have been considered. The proposed networks have been therefore designed in order to be suitable for recognition purposes under different mental conditions;

- the first proper evaluation of the feasibility of recognizing subjects using their brain signals, regardless the performed mental task. In addition to a task-dependent modality, in which EEG signals acquired according to the same protocol are compared, a task-independent modality has been also taken into account in the performed tests, investigating whether the proposed networks can extract, from the EEG signals of the considered users, distinctive information which could remain stable across both time and performed brain activity.

The paper is organized as follows. Section 2 illustrates the current state of the art on the use of deep learning techniques for EEG signal processing, for both biometric and non-biometric purposes. The architectures investigated in this paper, as well as the representations of EEG data used as their input, are outlined in Section 3, where also the preprocessing applied to the considered brain signals is described. The performed tests and the obtained results are then discussed in Section 4, while conclusions are eventually drawn in Section 5.

2. Deep Learning of EEG Signals

In order to be employed for a given application, EEG signals are typically processed to generate feature-based representations, which should possess properties specifically suited for the sought purpose. For instance, in case certain brain states have to be recognized, yet without the need for discriminating between different subjects, the desired characteristics should be invariant to both inter- and intra-subject differences. This kind of features would be needed for applications implementing brain computer interfaces (BCIs), whose aim is to understand the intentions of a user through his/her brain signals, or for medical purposes, where abnormal brain activities should be identified as markers of possible diseases. Conversely, EEG representations with limited intra-subject variability, yet able to retain inter-subject differences, are required when processing brain signals for people recognition purposes in biometric systems. Thanks to their ability in automatically learning the features resulting in a (local) optimum of the considered performance metrics, deep learning techniques have been recently adopted to cope with both the aforementioned scenarios.

Convolutional neural networks (CNNs) have been in fact recently exploited to design several BCIs [15]. For instance, brain signals captured from subjects elicited by motor imagery (MI) stimuli are interpreted in [16], while a system able to detect braking intention under different car driving conditions is proposed in [17]. Medical applications have been also developed as in [18], where automatically-learned brain features allow to predict and localize the focus of a seizure in medically refractory patients. Emotion recognition systems, able to identify human's emotive state by analyzing and processing brain signals, have been proposed in [19] and [20], while systems for monitoring subjects' fatigue and drowsiness have been considered in [21]. Deep learning techniques have been also proposed to preprocess EEG signals by detecting artifacts in [22].

Although time-domain inputs are fed to the considered CNNs in most of the aforementioned applications, such as in [17], [19], and [22], the processing applied to the collected EEG signals often include a feature extraction step, as in [16] and [21] where representations in the frequency domain have been extracted before applying deep learning, in [18] where time-frequency features expressed through wavelets have been exploited, and in [20] where connectivity characteristics have been instead used. The networks employed in all the aforementioned approaches comprise a limited number of convolutional layers, i.e., between one and three, followed by a fully-connected layer and a softmax.

CNNs have been also employed for EEG-based biometric recognition purposes as in [23], where brain signals captured in both eyes-closed (EC) and eyes-open (EO) scenarios have been processed by a network with two convolutional layers to classify ten subjects. A shallow network with a single convolutional layer has been proposed in [24] to extract features from EEG signals acquired from 30 subjects during the presentation of visual stimuli. Visually-evoked potentials (VEPs) deriving from visual stimuli shown to 40 subjects have been also processed by a four-convolutional-layer CNN architecture in [25], while signals acquired from 100 subjects performing a driving task in a virtual-reality interactive driving platform have been used as inputs of a network comprising three convolutional layers in [26]. The same number of convolutional layers has been employed in [27] to deal with EEG signals collected from 50 subjects while performing MI tasks, and also in [28] to classify the signals taken in EC and EO conditions from the 109 subjects available in the Physionet EEG Database. A steady state visual evoked potential (SSVEP) protocol has been considered in [29], where signals captured from ten subjects have been processed through six convolutional layers, as well as in [30], where recordings from eight subjects have been used as input to a network with two convolutional layers. A rapid serial visual presentation (RSVP) protocol has been used in [31], where a network with two convolutional layers has been trained on the event-related potentials (ERPs) extracted from the EEG signals collected from 15 genuine users. The same protocol has been also considered in [32], where an adversarial CNN with four layers has been used to process inputs from ten subjects.

In addition to CNNs, deep learning techniques involving recurrent neural networks (RNNs) have been also applied to EEG data processing. Actually, being the acquired signals an expression of the dynamic behavior of brain activity, exploiting networks able to keep an internal memory state for managing the temporal characteristics of sequences of inputs seems a natural choice. Long short-term memory (LSTM) networks, the most-widely employed kind of RNN architecture, have been in fact used to investigate brain disorders as in [33], implement BCI applications for MI or meaningful event detection as in [34], [35], and [36], recognize emotions as in [37], and also estimate cognitive workload as in [38] and [39]. As for applications based on CNNs, the inputs of the adopted recurrent networks can be given by either the raw EEG signals as in [36] and [37], by features extracted from them during preprocessing, using for instance PSD [35], [38] or wavelet characteristics [34], or even by representations obtained through CNNs [33].

Few recent works have also tested the effectiveness of RNNs in processing EEG signals for biometric recognition systems, as in [40] where recordings from eight subjects in relaxed EC conditions have been fed to LSTM networks, or in [41] where data taken from 109 subjects performing MI tasks have been considered.

It has however to be remarked that most of the studies proposed in literature regarding EEG-based biometric recognition systems using deep learning approaches have been carried out considering improper experimental conditions, with the reliability of the reported results severely compromised. In fact, single-session EEG data have been considered in several works as in [41], while acquisitions from two sessions have been used either in both enrolment and recognition stages as in [40], or to estimate only verification performance in terms of false rejection rate (FRR) on the same subjects employed to train the considered network for identification [31]. Proper experimental tests have been performed only in a handful of works, summarized in Table 1. Yet, as can be seen from the reported summary, a significant number of subjects have been taken into account only in [25] and [27], with a single mental task considered in both cases, and recognition performed over EEG signals lasting minutes, due to the need for extracting ERPs. The recognition accuracy achieved in such works, estimated comparing signals taken from distinct sessions, is typically low, even when considering datasets with few users as in [32]. It is also worth mentioning that, differently from what has been done for BCI and medical applications, approaches applying deep learn-

Table 1: State-of-the-art on deep learning techniques for EEG-based biometric identification.

Paper	**Sessions**	**Period**	**Subjects**	**Channels**	**Protocol**	**EEG Duration**	**Network**	**Accuracy**
[25]	2	1 week	40	19	VEP	5 min	CNN, 3 conv layer	98.8%
[27]	2	1 week	50	19	MI	4 min	CNN, 3 conv layer	81.2%
[30]	2	1 day	8	9	SSVEP	10 s	CNN, 2 conv layer	96.7%
[32]	3	8 days	10	16	RSVP	0.5 s	Adv. CNN, 4 conv layer	71.6%

ing techniques to EEG signals for biometric recognition purposes have used only raw temporal data as inputs of their architectures.

The tests reported in this study have been designed while taking into account the limitations of all previous attempts. Novel neural network architectures, exploiting CNNs and RNNs both separately and jointly, have been therefore defined and fed with both raw EEG data and hand-crafted features extracted from them, as described in Section 3. The proposed solutions have been tested over a large dataset comprising EEG data collected from 45 subjects performing six mental tasks during five distinct recording sessions, spanning a period larger than one year, as outlined in Section 4.

3. EEG-based Biometric Recognition with Deep Learning Techniques

The general framework of the proposed EEG-based identification systems is depicted in Figure 1. During enrolment, the EEG recordings of U users are collected, and employed to train a neural network. At the identification stage, a novel EEG acquisition is preprocessed and segmented into frames, which are fed as inputs to the trained network to estimate the identity of the presented subject, first for each frame, and then on the basis of the whole recorded EEG signal. The preprocessing applied to the collected EEG signals, common to all the considered systems, is presented in Section 3.1. The representations derived from preprocessed data to be used as inputs of the proposed networks are outlined in Section 3.2. The descriptions of the networks exploited to perform EEG-based biometric recognition are then given in Sections 3.3 and 3.4.

3.1. Preprocessing

It is assumed that EEG recordings are acquired from a user u employing C electrodes placed on the user's scalp, generating the raw signals $\mathbf{s}^{(u,c)}$, $c = 1, \ldots, C$.

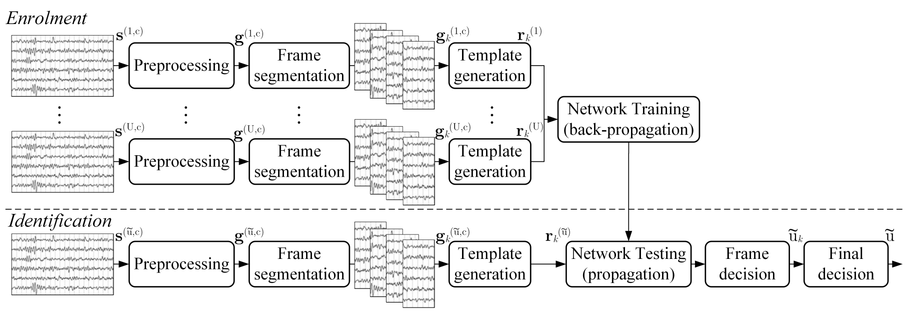

Figure 1: Graphical outline of the proposed EEG-based identification systems.

A spatial filtering is first applied to the collected data, in order to improve their signal-to-noise ratio (SNR), using a common average referencing (CAR) filter [42]. The mean voltage sensed over the entire scalp is subtracted from the raw EEG signals of each channel, obtaining $\mathbf{v}^{(u,c)} = \mathbf{s}^{(u,c)} - \frac{1}{C}\sum_{c=1}^{C}\mathbf{s}^{(u,c)}$. A frequency band-pass filtering is then performed, limiting the considered signals within the $[\alpha, \beta] = 8 \div 30$Hz subband, which previous studies have shown to contain the most discriminative and permanent EEG content [12], [13]. Whatever the acquisition frequency employed for data collection, a downsampling to $D = 64$Hz is then performed to reduce the computational complexity of the subsequent processing, without discarding significant information given the considered subband, thus obtaining the preprocessed signals $\mathbf{g}^{(u,c)}$.

A segmentation process is eventually carried out, generating from each EEG acquisition $\mathbf{s}^{(u,c)}$ a set of K overlapping frames $\mathbf{g}_k^{(u,c)}$, $k = 1, \dots, K$, each lasting $L = 5$s, with an 80% overlap factor between consecutive frames. The so-generated frames are treated as EEG samples in the considered systems, and therefore employed as training data upon which the proposed networks are learned, as well as query probes whose owner should be identified during recognition.

3.2. Data Representations

In order to be fed to the considered networks, the frames obtained from preprocessed EEG signals can be represented in different formats. In fact, CNNs require inputs in the form of matrices or tensors, whereas RNNs works with sequences of data. As outlined in Section 2, both kinds of network have been employed in literature using as inputs either raw time-dependent EEG signals, or features extracted from them.

In this latter case, hand-crafted characteristics are first generated from the treated raw EEG data, before using them to automatically learn some alternative descriptors during network training.

In order to perform a comprehensive analysis of the deep learning architectures suitable for EEG-based biometric recognition, the present work reports in Section 4 an extensive set of experimental tests, carried out while taking into account several distinct representations of an EEG frame.

3.2.1. Frame-based Representations

An EEG frame generated according to the preprocessing described in Section 3.1, comprising the C signals $\mathbf{g}_k^{(u,c)}$, $c = 1, \ldots, C$, can be represented in a format $\mathbf{r}_k^{(u)}$ consisting of:

- a matrix $\mathbf{M}_T$ with size $C \times (L \cdot D)$, whose rows represent the temporal behavior of the EEG signals acquired through the considered channels. Such data could be used as input to a CNN, as proposed in [25], [27], [30], and [32];
- a sequence $\mathbf{S}_T$ of $L \cdot D$ vectors having size $C \times 1$. Such data could be used as input for an RNN, as proposed in [40];
- a set $\mathbf{M}_\psi$ of C sequences, each comprising the $L \cdot D$ elements of a channel. This representation can be exploited by training a joint RNN + CNN network, where each sequence is modeled with an RNN, and the last internal states of each RNN, expressed through H coefficients, are organized into a $C \times H$ matrix used as input to a CNN.

The aforementioned approaches directly apply deep learning techniques to raw EEG data in order to extract discriminative features, and use them to classify the considered traits. State-of-the-art deep learning approaches to EEG-based biometric recognition have actually resorted to $\mathbf{M}_T$ or $\mathbf{S}_T$ representations of EEG data.

Yet, hand-crafted features could be alternatively derived from each EEG channel, and used as inputs for standard classifiers or deep neural networks, as already done for BCI and medical applications [21]. A frame can be therefore represented also as:

- a matrix $\mathbf{M}_F$ with C rows, each containing W features associated to a channel of the considered frame, to be used as input to a CNN;
- a vector $\mathbf{m}_F$ comprising $C \cdot W$ elements, obtained concatenating the rows of $\mathbf{M}_F$. This approach is actually the one commonly adopted in most state-of-the-art EEG-based biometric recognition systems, exploiting standard classifiers such as nearest neighbor (NN) or support vector machines (SVMs) to perform classification [13].

With respect to the exploitation of the temporal behavior of EEG signals, resorting to hand-crafted features makes the employed representations independent of the adopted sampling rate D and frame duration L. The adopted features should be obviously selected in order to retain the discriminative information contained in EEG signals. Given the high reliability shown in several studies [13], auto-regressive (AR) models and mel-frequency cepstrum coefficients (MFCC) are here jointly used to provide a highly-discriminative manually-defined EEG representation.

Specifically, AR modeling has proven to be one of the most effective representations for EEG signals employed in biometric recognition systems [43]. Each c-th channel $\mathbf{g}_k^{(u,c)}$ of a given k-th frame can be modeled as a realization of an AR process of order O, obeying the linear difference equation $\mathbf{g}_k^{(u,c)}[t] = -\sum_{o=1}^{O} a_{k(O,o)}^{(u,c)} \cdot \mathbf{g}_k^{(u,c)}[t-o] + \mathbf{w}[t]$, where $\mathbf{w}[t]$ is a realization of a white noise process with standard deviation $\sigma_{k(O)}^{(u,c)}$, and $a_{k(O,o)}^{(u,c)}$, $o = 1, \ldots, O$, are the AR coefficients. These latter can be estimated through the Yule-Walker equations [44], which can be solved iteratively employing the Levinson algorithm. Such approach introduces the reflection coefficients $P_{k,(O)}^{(u,c)}$ of order O, employed in the proposed systems to represent each EEG channel. Their estimation can be obtained with the Burg's method [44], directly operating on the observed data $\mathbf{g}_k^{(u,c)}$, rather than passing through their autocorrelation function, with $O = 12$ in the performed tests.

As for the employed MFCC modeling, it has been traditionally used in applications involving speech processing, and recently exploited also for the biometric analysis of brain data [45], showing its ability in achieving superior performance with respect to other descriptors, such as those based on power spectral density (PSD) or spectral coherence (COH) [13]. In the adopted implementation, a filter bank with 18 mel-scaled triangular band-pass filters is first applied to the EEG spectrum $\mathbf{G}_k^{(u,c)}$ of each c-th channel. The natural logarithm of the resulting cepstral bins is then evaluated to filter the underlying neural activation signals from the effects of their propagation through the skull [46]. A discrete cosine transform (DCT) is eventually performed on the resulting values, with the desired MFCC representation obtained by selecting the first $O = 12$ DCT coefficients, excluding the DC component.

In the performed tests, described in Section 4, a total of $W = 24$ features are therefore associated to each channel when generating the representations $\mathbf{M}_F$ and $\mathbf{m}_F$.

3.2.2. Sub-frame-based Representations

Alternative representations can be generated by splitting each frame into overlapping sub-frames, then describing each available sample as a sequence of elements. Specifically, systems performing the division of each frame into $Q = 17$ sub-frames lasting $J = 1$s, using a 75% overlap factor, have been considered in the performed tests. When resorting to such approach, a frame can be interpreted as different kinds of sequences, that is, as:

- a sequence $\mathbf{S}_\tau$ of Q matrices, each with size $C \times (J \cdot D)$. A joint CNN + RNN network can be trained over these data, with a recurrent network receiving as input the feature representations generated by a CNN trained over raw data;
- a sequence $\mathbf{S}_\zeta$ of Q matrices, each with size $C \times W$, being W the number of hand-crafted features associated to each channel in a sub-frame. A joint CNN + RNN network can be trained over these data, with a recurrent network receiving as input the feature representations generated by a CNN trained over the adopted hand-crafted characteristics;
- a sequence $\mathbf{S}_F$ of Q vectors, each comprising $C \cdot W$ elements defined by manually extracting features from the channels of a sub-frame. Such data could be used as input for an RNN.

Eventually, a frame can be represented as a set $\mathbf{M}_\phi$ of C sequences, each made of Q vectors with size $1 \times W$, containing the features extracted from each sub-frame. This

Table 2: Employed representations of an EEG frame, used as inputs for the considered classifiers.

$\mathbf{r}_k^{(u)}$	**Source**	**Type**	**Elements**	**Structure**	**Input to**
$\mathbf{M}_T$	Frame	Matrix	Raw data	$C \times (L \cdot D)$	CNN
$\mathbf{S}_T$	Frame	Sequence	Raw data	$L \cdot D$ vectors with C coefficients	RNN
$\mathbf{M}_\psi$	Frame	Sequences	Raw data	C sequences, each with $L \cdot D$ coeff.	RNN + CNN
$\mathbf{M}_F$	Frame	Matrix	Features	$C \times W$	CNN
$\mathbf{m}_F$	Frame	Vector	Features	$C \cdot W$	Standard classifiers
$\mathbf{S}_\tau$	Sub-frame	Sequence	Raw data	Q matrices with size $C \times (J \cdot D)$	CNN + RNN
$\mathbf{S}_\zeta$	Sub-frame	Sequence	Features	Q matrices with size $C \times W$	CNN + RNN
$\mathbf{S}_F$	Sub-frame	Sequence	Features	Q vectors with $C \cdot W$ coefficients	RNN
$\mathbf{M}_\phi$	Sub-frame	Sequences	Features	C sequences, each with Q $1 \times W$ vectors	RNN + CNN

representation can be exploited by training a joint RNN + CNN network, where each sequence is modeled with an RNN, and the last internal states of each RNN, expressed through H coefficients, are organized into a $C \times H$ matrix used as input to a CNN.

A resume of the tested representations, with their main characteristics and the network to which they are supposed to be fed, is reported in Table 2. A diagram depicting the generation of the employed EEG representations is also reported in Figure 2. The networks considered in the performed tests are discussed in the following sections.

It is just worth mentioning that, in addition to the aforementioned frame representations, two additional tensorial structures could have been considered when dividing a frame into overlapping sub-frames, namely:

- a $C \times (J \cdot D) \times Q$ tensor $\mathbf{M}_\tau$, containing EEG raw data to be directly provided as input to a CNN;
- a $C \times W \times Q$ tensor $\mathbf{M}_\upsilon$, containing feature extracted from each channel of a sub-frame, to be provided as input to a CNN.

Nonetheless, such representations are not considered in the following discussion, since they are conceptually similar to $\mathbf{M}_T$ and $\mathbf{M}_F$, respectively, and yet they achieve worse recognition performance while being characterized by an higher computational complexity.

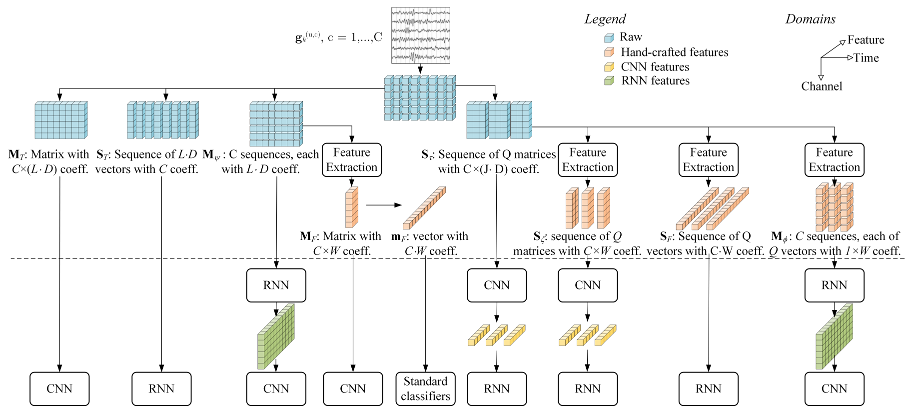

Figure 2: Graphical outline of the proposed representations of an EEG frame.

3.3. CNN Architectures

CNNs represent the class of deep neural networks commonly employed to process data with a grid-like structure, like matrices or tensors. The layers of a CNN perform a convolution (actually, a cross-correlation) between their inputs and a set of kernel filters, whose coefficients have to be learned during the training stage, thus producing outputs fed to the next layer.

Most of the CNNs employed in the EEG-based biometric recognition systems mentioned in Section 2 employ square kernels having size $[5 \times 5]$ as in [25], [27], and [31], with $[3\times3]$ and $[4\times4]$ filters also used in [29]. Such choice results in architectures with a limited number of convolutional layers, typically in the order of two or three, since pooling layers reducing the size of the treated data, and providing translation-invariant properties, are typically employed after them, thus making the number of employed electrodes C a limiting factor for the depth of the network. In fact, C is usually in the order of some dozens, and should be possibly kept to numbers as low as possible to minimize the discomfort of users when wearing the EEG acquisition devices.

Conversely, uni-dimensional kernels are here employed to allow the processing of each channel to be performed independently one from the others. Specifically, $[1 \times L]$ filters, with $L \in [3 \div 6]$, are employed to process the considered representations in the time- and feature-based domains, while a single $[C \times 1]$ filter is employed to compute the connectivity among the available channels. The overall number of layers in the proposed architectures is therefore independent on the number of used channels C, since this latter only affects the size of the spatial filter employed in a single layer.

The CNN architectures applied to each of the considered EEG representations are detailed in Table 3. As input, they can receive either raw EEG temporal data when using $\mathbf{M}_T$ or $\mathbf{S}_\tau$ as frame representations, hand-crafted features when exploiting $\mathbf{M}_F$ or $\mathbf{S}_\zeta$, or also features generated by an RNN when adopting $\mathbf{M}_\psi$ or $\mathbf{M}_\phi$ representations. In all the considered cases, the employed convolutional layers produce 1024 features, which can be either provided to a fully-connected (FC) layer whose size depends on the number U of considered users, followed by a softmax layer based on a cross-entropy loss function, or used as input to another stage exploiting an RNN, as when considering $\mathbf{S}_\tau$ and $\mathbf{S}_\zeta$ representations.

In the adopted implementations, each convolutional layer is followed by a batch normalization layer. Rectified linear unit (ReLU), max pooling (MP), and dropout (DO) at 50% layers are also used. Network training is performed on the representations $\mathbf{r}_k^{(u)}$, $k = 1, \ldots, K$, $u = 1, \ldots, U$, derived from the EEG signals recorded from U subjects during enrolment, using Adam [47] as gradient-based optimization algorithm.

3.4. RNN Architectures

RNNs represent the class of deep neural networks commonly employed to process data characterized by a temporal dynamic behavior. This kind of network receive inputs, update its hidden states depending on previous computations, thus maintaining a memory of the information processed so far, and produce outputs making predictions associated to either each element of the sequence or the whole sequence altogether.

LSTM units have been here exploited to model the considered sequences of inputs. Networks based on LSTM units are capable of learning long-term dependencies handling a memory cell ($\mathbf{l}$) using an input gate ($\mathbf{i}$), an output gate ($\mathbf{o}$), and a forget gate ($\mathbf{f}$), which regulate the flow of information into and out each unit. Letting $\varsigma(x) = (1 + \mathrm{e}^{-x})^{-1}$ be the *sigmoid* function, and $\eta(x) = 2\sigma(2x) - 1$ be the *hyper-*

Table 3: Employed CNNs architectures, depending on the chosen EEG representation.

(a) input: $\mathbf{M}_T, \mathbf{M}_\psi, \mathbf{M}_\phi$

#	Layer	Filters
L_1	Conv	(1×5×1)×32
L_2	Relu	-
L_3	MP	1×2
L_4	Conv	(1×5×32)×64
L_5	Relu	-
L_6	MP	1×2
L_7	Conv	(1×5×64)×128
L_8	Relu	-
L_9	MP	1×2
L_{10}	Conv	(1×5×128)×256
L_{11}	Relu	-
L_{12}	MP	1×2
L_{13}	Conv	(1×5×256)×512
L_{14}	MP	1×2
L_{15}	Conv	(C×1×512)×512
L_{16}	Relu	-
L_{17}	DO	-
L_{18}	Conv	(1×6×512)×1024
L_{19}	Relu	-

(b) input: $\mathbf{S}_\tau$

#	Layer	Filters
L_1	Conv	(1×5×1)×128
L_2	Relu	-
L_3	MP	1×2
L_4	Conv	(1×5×128)×256
L_5	Relu	-
L_6	MP	1×2
L_7	Conv	(1×5×256)×512
L_8	Relu	-
L_9	MP	1×2
L_{10}	Conv	(C×1×512)×512
L_{11}	Relu	-
L_{12}	DO	-
L_{13}	Conv	(1×4×512)×1024
L_{14}	Relu	-

(c) input: $\mathbf{M}_F, \mathbf{S}_\varsigma$

#	Layer	Filters
L_1	Conv	(1×5×1)×256
L_2	Relu	-
L_3	MP	1×2
L_4	Conv	(1×5×256)×512
L_5	MP	1×2
L_6	Conv	(C×1×512)×512
L_7	Relu	-
L_8	DO	-
L_8	Conv	(1×3×512)×1024
L_9	Relu	-

bolic tangent function, the LSTM unit is updated at time-step t as:

$$\begin{cases} \mathbf{i}[t] = & \varsigma(\mathbf{W}_i\mathbf{x}[t] + \mathbf{U}_i\mathbf{h}[t-1] + \mathbf{b}_i) \\ \mathbf{f}[t] = & \varsigma(\mathbf{W}_f\mathbf{x}[t] + \mathbf{U}_f\mathbf{h}[t-1] + \mathbf{b}_f) \\ \mathbf{o}[t] = & \varsigma(\mathbf{W}_o\mathbf{x}[t] + \mathbf{U}_o\mathbf{h}[t-1] + \mathbf{b}_o) \\ \mathbf{l}[t] = & \mathbf{f}[t] \odot \mathbf{l}[t-1] + \mathbf{i}[t] \odot \eta(\mathbf{W}_c\mathbf{x}[t] + \mathbf{U}_c\mathbf{h}[t-1] + \mathbf{b}_c) \\ \mathbf{h}[t] = & \mathbf{o}[t] \odot \eta(\mathbf{l}[t]), \end{cases} \tag{1}$$

where $\mathbf{i}, \mathbf{f}, \mathbf{o} \in \mathbb{R}^Y$ are the activation vectors for input gate, forget gate, and output gate respectively. $\mathbf{h} \in \mathbb{R}^Y$ is the output vector of the LSTM unit, $\mathbf{l} \in \mathbb{R}^Y$ is the cell-state vector, $\mathbf{U} \in \mathbf{R}^{Y \times Y}$, $\mathbf{b} \in \mathbb{R}^Y$ are the weight matrices and bias vector for the gates to be learned during training, $\odot$ is the Hadamard element-wise product, $\mathbf{x} =\in \mathbb{R}^Z$ is the input vector to the LSTM unit, and $\mathbf{W} \in \mathbb{R}^{Y \times Z}$, with $Z = 1$ for $\mathbf{S}_T$, $Z = C$ for $\mathbf{M}_\psi$, $Z = 1024$ for $\mathbf{S}_\tau$ and $\mathbf{S}_\zeta$, and $Z = W$ for $\mathbf{M}_\phi$.

Bidirectional LSTM (BLSTM) networks, where distinct units are used to process information from past (backwards) and future (forward) states, thus generating two separate hidden layers with access to the past and future context of every element in the sequence, have been here employed. The performed tests have actually highlighted the superior performance achievable by BLSTM, with respect to LSTM, in terms of both recognition accuracy and processing time. The last outputs $\mathbf{h}$ of the employed LSTM units, obtained when the input sequences have been completely parsed, are provided as input either to a 50% dropout layer, followed by an FC layer producing U outputs and a softmax layer, or to a CNN as when $\mathbf{M}_\psi$ and $\mathbf{M}_\phi$ are considered. The hidden layer of the employed RNNs is composed by $Y = 1300$ units when processing inputs as $\mathbf{S}_T$, $\mathbf{S}_\tau$, $\mathbf{S}_\zeta$, and $\mathbf{S}_F$, therefore generating a representation with $H = 2600$ coefficients, while $Y = 180$ units are used for each channel when dealing with the EEG representations $\mathbf{M}_\psi$ and $\mathbf{M}_\phi$, for which therefore $H = 360$. Adam [47] is used for training LSTM networks too.

Tests with deep RNNs, stacking LSTM units on top of other LSTM units thus generating a temporal hierarchy, have also been performed, yet the associated results are not reported in the following being them typically worse than those achievable with networks having a single layer of (bidirectional) units.

4. Experimental Evaluation

An extensive set of experimental tests has been carried out in order to evaluate the effectiveness of using the proposed deep learning approaches for EEG-based biometric recognition. The considered database is described in Section 4.1, while the performed tests and the obtained results are outlined in Section 4.2.

4.1. Employed Database

In order to evaluate whether a user could be recognized using the brain activity recorded while performing a given activity, a huge database containing EEG acquisitions from 45 subjects, taken during six different mental tasks, has been taken into

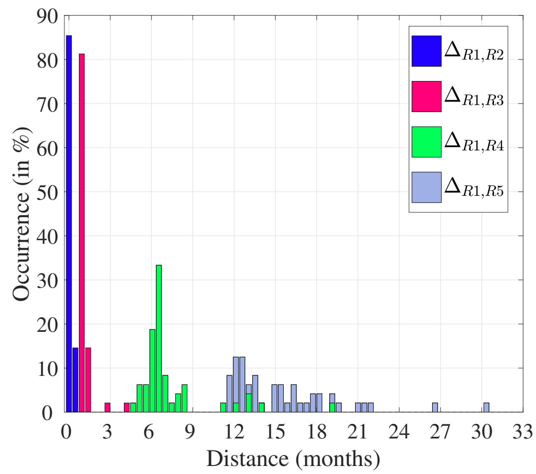

Figure 3: Acquisitions' sessions temporal distribution (adapted from [13]).

account. Specifically, the longitudinal database described in [13], together with extensions of the database employed in [25], [27], and [48], have been exploited to properly test the proposed neural networks.

Brain signals collected at an original sampling rate of 256 Hz, using a GALILEO BE Light amplifier equipped with $C = 19$ wet electrodes, are here considered. Five different recording sessions, indicated in the following as $R_1, R_2, \ldots, R_5$, have been arranged for each of 45 subjects, all in healthy conditions and with an age in the range between 21 to 34 years, with an average of 25 years, at the time of the first acquisition. The average distances between the first and the other four recordings are: $\bar{\Delta}_{R_1,R_2} = 1$ week, $\bar{\Delta}_{R_1,R_3} = 1$ month, $\bar{\Delta}_{R_1,R_4} = 7$ months, and $\bar{\Delta}_{R_1,R_5} = 16$ months, with the distributions of the time distances Δ_{R_1,R_n}, $n = 2, \ldots, 5$ elapsed between the first and the n-th EEG acquisition of each subject depicted in Figure 3.

EEG signals have been collected from each subject according to the following six elicitation protocols:

- resting state with eyes closed (EC), with subjects keeping their eyes closed while relaxing for four minutes. This represents the most-commonly-used protocol when acquiring EEG signals for biometric recognition purposes, and commonly guarantees the best possible recognition accuracy;

- resting state with eyes open (EO), with subjects asked to fix a light point on the screen for four minutes while keeping a relaxed state;

- motor imagery (MI), during which subjects are asked to perform an imaginary movement depending on which one out of four images is shown on the screen, contemplating right-arm, left-arm, right-leg and left-leg movements. As described in [27], each stimulus is randomly displayed for three seconds, after which an empty black screen is shown for 1.5 seconds. Twenty occurrences of each stimulus are here considered, with each session lasting six minutes;

- speech imagery (SI), where subjects are asked to mentally reproduce the sound of a vowel observed on the screen. Each of the 5 vowels is shown 20 times in random order for three seconds, with two seconds of blank screen between consecutive stimuli, for an overall recording period of about eight minutes;

- visual stimulation (VS), during which eight images, each of a different geometric shape, are displayed on a screen for $250\mathrm{ms}$, followed by a black screen lasting $450\mathrm{ms}$. Each geometric shape is presented in a random order for 60 times, therefore resulting in a total session period of about five minutes. As described in [48], the appearance of each image generates a visually-evoked potentials (VEP) in the observer, while a peculiar response is elicited when the target shape is shown;

- mathematical computation (MC), with subjects asked to perform cognitive tasks consisting of simple mathematical operations such as sums and differences of integers. During each session, 28 operations are shown on the screen, each for five seconds with a two-seconds interval between consecutive demands, for a total session duration of about three minutes.

The employed protocols stimulate different brain activation patterns, being the considered tasks related to distinct neural oscillatory activities. In general, it can be assumed that slower brain rhythms are dominant during relaxed states, while the faster ones are typical of information processing operations. EEG signals recorded in EC conditions show a predominant contribution of $\alpha = [8 \div 13]\mathrm{Hz}$ oscillations in the parieto-occipital region of the scalp [49]. A widespread activity reduction resulting in a neuronal α desynchronization, and corresponding to an increased arousal due to basic processing of visual information [50], is commonly observed turning to EO resting conditions, together with an increase in $\theta = [4 \div 8]$Hz band power. Attentional and semantic memory demands lead to a further suppression of α activity, with an involvement of central and frontal brain regions [51]. Performing SI and MI tasks in fact results in increased oscillatory activity in $\beta = [4 \div 8]$Hz band by the motor cortex in the centro-lateral side of the scalp [52]. The need for concentrating and successfully encoding new information in tasks based on VS and MC implies greater involvement of the frontal lobe, with an increase in $\delta = [0.5 \div 4]$Hz, β, and $\gamma = [30 \div 40]$Hz bands [53].

The considered EEG data are treated as continuous streams to which the processing described in Section 3 is applied, without any connection to the specific events which may have triggered them in the considered elicitation protocols. Differently from [25], [27], [48], and also [13], the temporal references to the shown stimuli are therefore discarded during the enrolment and recognition phases of the systems.

As already remarked in Section 1, the database used for the experimental tests here described is by far the largest multi-session dataset ever employed to evaluate an EEG-based biometric recognition systems, in terms of enrolled subjects and employed elicitation protocols.

4.2. Tests and Results

In order to properly test an EEG-based biometric recognition systems, it is required to compare data collected during distinct recording sessions. Taking this into account, very challenging conditions have been considered in the performed experimental tests, by comparing EEG signals captured at an average distance of at least 15 months for each considered subject.

In this scenario, the recognition performance achievable when employing standard classifiers has been first evaluated, as detailed in Section 4.2.1. Deep-learning-based approaches already proposed in literature for EEG processing have also been analyzed, as outlined in Section 4.2.2. The results obtained when employing the EEG representations and networks proposed in Section 3 are then discussed in Section 4.2.3. In all the

Table 4: Rank-1 IR (mean and confidence interval, in %) over a single EEG frame achievable with standard classifiers applied to the $\mathbf{m}_F$ EEG representation (best performance for each protocol in bold).

Classifier	**Protocol**					
	EC	**EO**	**MI**	**SI**	**VS**	**MC**
NN	66.8 $\pm$ 0.3	53.7 $\pm$ 0.2	54.8 $\pm$ 0.2	57.6 $\pm$ 0.2	54.1 $\pm$ 0.3	57.5 $\pm$ 0.2
LDA	**81.8 $\pm$ 0.4**	68.4 $\pm$ 0.3	73.3 $\pm$ 0.3	72.1 $\pm$ 0.2	70.1 $\pm$ 0.2	71.9 $\pm$ 0.2
SVM	81.2 $\pm$ 0.3	66.3 $\pm$ 0.2	69.7 $\pm$ 0.2	72.6 $\pm$ 0.3	69.6 $\pm$ 0.2	71.7 $\pm$ 0.3
LDA + SVM	81.6 $\pm$ 0.3	**69.0 $\pm$ 0.2**	**75.1 $\pm$ 0.3**	**73.2 $\pm$ 0.2**	**73.3 $\pm$ 0.3**	**72.6 $\pm$ 0.4**

considered cases, enrolment has been performed using the first three available sessions of each user, while data acquired during the fifth session has been used for estimating identification accuracy. Whenever required, the fourth session has been used to provide validation data, in order to determine the most-suitable number of epochs needed to properly train the exploited systems. A cross-validation evaluation modality has been carried out, repeating for five times the training and testing procedure over each of the considered protocols and classifiers. At each iteration, 150s of the data available in each of the first three session have been used for training purposes, 90s of the fourth recordings for validation, and 90s of the fifth recordings for testing purposes.

4.2.1. Performance of standard classifiers

As a baseline reference, the performance attainable with standard classifiers applied to the $\mathbf{m}_F$ representation of the considered EEG frames has been first evaluated. The considered classifiers include state-of-the-art supervised approaches such as NN with Manhattan distance [12], Fisher linear discriminant analysis (LDA) with Manhattan distance [54], SVM with Gaussian kernels and a one-versus-one comparison design [55], and also a novel approach considered for the first time in this work, exploiting SVM classifiers over LDA projections of the original representations. The obtained recognition accuracies are reported in Table 4 in terms of rank-1 identification rate (IR), for identification probes consisting of a single frame, that is, five seconds of EEG data. The obtained results confirm EC as the protocol guaranteeing the best possible recognition accuracy. The MI, SI, VS, and MC protocols all achieve similar performance levels, while the EO acquisition modality performs notably worse than the others. The highest accuracies are achieved using the LDA + SVM classifier here introduced. Nonetheless, even considering that the reported results are referred to a scenario where acquisitions taken at a time distance greater than one year are compared, further performance improvements would be highly desirable in order to allow EEG-based biometric systems be deployed for practical applications.

4.2.2. Performance of state-of-the-art neural networks

Neural networks already proposed in literature for EEG processing have been employed to perform people identification too. The considered networks include both general-purpose architectures proposed in field-leading works regarding EEG signals, as well as networks specifically used for EEG-based biometric recognition.

Tests have been in fact conducted using the *Deep ConvNet* and the *Shallow Net* architectures designed in [56] to perform oscillatory signal classification, as well as the *EEGNet* framework proposed in [57] to perform automatic EEG feature extraction in BCI applications. Furthermore, the neural networks specifically designed to cope with EEG-based biometric recognition in [25], [27], and [30], have also been considered. All the aforementioned approaches apply CNNs to EEG templates based on raw temporal data, analogous to the $\mathbf{M}_T$ representation defined in Section 3.2.

It has to be specified that the application of neural networks as originally described in the aforementioned papers has not guaranteed proper recognition rates. The following modifications have had to be made in order to improve the performance achievable with deep learning techniques already proposed for EEG processing:

- the employed filters have been adapted to be applied to EEG signals sampled at 64Hz. Lowering the employed sampling rate facilitates the process of parameter learning during network training, due to the possibility of using shorter filters, thus requiring less training time, and allowing to achieve better performance, with respect to the use of the original proposals at 128Hz or 256Hz in [56], [57], and [30];

- for all the considered approaches, the number of kernels employed at each convolutional layer has had to be notably increased, with respect to the original versions proposing the usage of few filters as in [57], improving the achievable recognition performance at the cost of an increased computational complexity.

The number of learnable parameters in each considered architecture is detailed in Table 5;

- the number of dropout layers in *Deep ConvNet* has been notably reduced, since the original version proposed in [56] could not have been trained to give acceptable performance;
- batch normalization layers have been employed in the networks proposed in [25] and [27], which otherwise would have reached far lower recognition rates;
- instead of exponential linear unit (ELU), square, or logarithm, rectified linear unit (ReLU) has been employed as activation function, due to the better accuracies obtained with it.

A batch size consisting of 64 samples has been employed during training in all the performed evaluations, carried out on an Nvidia GeForce GTX Titan X GPU, using Adam [47] as gradient-based optimization algorithm, with an initial learning rate at 0.0001. For each state-of-the-art network, the recognition rates achieved in the considered scenarios are reported in Table 5, together with the associated computational complexity, expressed in terms of number of learnable parameters, and time required on average to perform training over samples belonging to a single protocol. Interestingly, the obtained recognition rates are far better than those reported in [25], [27], and [48], where EEG signals captured using MI and VS protocols are treated while considering the temporal references of the employed stimuli, which are here instead discarded.

Analyzing the obtained results, it can be seen that using square kernels in the employed convolutional layers, as in [25] and [27], ensues in general worse performance than resorting to uni-dimensional kernels, with the exception of the EC scenario, and also requires considerable training time. The network in [30], derived from the *Shallow Net* in [56], outperforms this latter thanks to the usage of smaller pooling layers. *EEG-Net*, which employs more conv layers than [30], guarantees the best results among the considered state-of-the-art approaches. However, it can outperform a standard classifier using LDA and SVM only when considering the EO and VS protocols, while

Table 5: Rank-1 IR (mean and confidence interval, in %) over a single EEG frame achievable with state-of-the-art neural networks proposed for EEG processing (best performance for each protocol in bold). The computational complexity of the employed networks is also reported.

Network	# parameters (millions)	Training time (min)	Protocol					
			EC	EO	MI	SI	VS	MC
Deep ConvNet [56]	1.233	21	76.5 ± 0.2	62.7 ± 0.3	60.9 ± 0.2	54.2 ± 0.3	59.6 ± 0.3	48.5 ± 0.2
Shallow Net [56]	0.506	16	77.2 ± 0.4	64.4 ± 0.3	67.8 ± 0.3	52.1 ± 0.2	68.7 ± 0.2	45.4 ± 0.2
EEGNet [57]	1.002	47	$\mathbf{81.7 \pm 0.3}$	$\mathbf{74.6 \pm 0.2}$	$\mathbf{72.8 \pm 0.2}$	$\mathbf{70.5 \pm 0.3}$	$\mathbf{74.6 \pm 0.3}$	$\mathbf{72.5 \pm 0.3}$
[25]	26.887	45	66.8 ± 0.3	60.6 ± 0.2	59.4 ± 0.2	47.8 ± 0.2	56.3 ± 0.3	44.6 ± 0.2
[27]	44.371	70	81.4 ± 0.3	68.2 ± 0.2	65.3 ± 0.2	56.8 ± 0.2	63.3 ± 0.2	53.6 ± 0.2
[30]	0.534	17	79.6 ± 0.3	68.0 ± 0.2	68.5 ± 0.3	54.2 ± 0.2	69.3 ± 0.2	42.6 ± 0.4

worse results have been obtained in all the other scenarios. Furthermore, due to the large size of its kernels, the employed *EEGNet* requires more time than *Deep ConvNet* and *Shallow Net* to be trained. Even though *Deep ConvNet* comprises more layers than *EEGNet*, it achieves worse results due to the excessively small size of the employed kernels.

The networks presented in Section 3 have been designed taking into account the strengths and weaknesses observed in the considered state-of-the-art approaches for EEG processing: batch normalization is used after each conv layer, while limiting dropout. Networks comprising the largest possible number of layers, given the usage of uni-dimensional kernels with a length in between those used in *Deep ConvNet* and *EEGNet*, have been defined. The number of filters exploited in each conv layer is significantly larger than those employed in [56], [57], and [30]. A single conv layer with spatial filters, computing measures of connectivity between channels, is included. The same considerations have been taken into account when designing the networks to be applied on hand-crafted feature representations of EEG data.

4.2.3. Performance of the proposed neural networks

The performance achievable when applying the proposed deep learning approaches to the EEG frame representations detailed in Section 3 are reported in Table 6, where identification probes consisting of a single frame are taken into account.

The reported results show that the proposed networks notably outperform both standard classifiers and state-of-the-art network architectures proposed in literature for

Table 6: Rank-1 IR (mean and confidence interval, in %) over a single EEG frame achievable with the deep learning approaches proposed in Section 3, considering different acquisition protocols (best performance for each protocol in bold).

Represent.	**Network**	**Protocol**					
		EC	**EO**	**MI**	**SI**	**VS**	**MC**
$\mathbf{M}_T$	CNN	85.9 ± 0.3	75.1 ± 0.5	79.5 ± 0.4	77.1 ± 0.2	79.5 ± 0.7	80.8 ± 0.3
$\mathbf{S}_T$	RNN	60.9 ± 1.0	45.5 ± 0.3	43.9 ± 1.3	48.7 ± 3.4	44.9 ± 0.4	46.1 ± 1.5
$\mathbf{M}_\psi$	RNN + CNN	82.6 ± 0.3	72.4 ± 0.2	76.6 ± 0.2	76.1 ± 0.2	77.2 ± 0.2	77.8 ± 0.3
$\mathbf{M}_F$	CNN	90.6 ± 0.2	80.6 ± 0.2	85.4 ± 0.3	83.4 ± 0.2	$\mathbf{81.6 \pm 0.3}$	85.3 ± 0.2
$\mathbf{S}_\tau$	CNN + RNN	84.0 ± 0.2	73.2 ± 0.2	77.2 ± 0.2	76.2 ± 0.2	78.6 ± 0.4	78.5 ± 0.3
$\mathbf{S}_\zeta$	CNN + RNN	$\mathbf{91.4 \pm 0.2}$	$\mathbf{81.9 \pm 0.2}$	$\mathbf{86.2 \pm 0.3}$	$\mathbf{84.1 \pm 0.3}$	81.4 ± 0.3	$\mathbf{86.1 \pm 0.3}$
$\mathbf{S}_F$	RNN	89.7 ± 0.3	78.8 ± 0.2	79.9 ± 0.4	79.1 ± 0.3	77.2 ± 0.2	81.5 ± 0.3
$\mathbf{M}_\phi$	RNN + CNN	81.3 ± 0.2	73.8 ± 0.2	79.6 ± 0.3	76.7 ± 0.2	77.1 ± 0.2	80.8 ± 0.2

EEG processing, with improvements in recognition rates ranging from 7% to 14%, depending on the considered stimulation protocols. Again, the highest IRs are achieved when considering the EC protocol, with three networks attaining performance at about 90% of correct recognition using probe EEG signals lasting only five seconds, and taken after a time-lapse of about 15 months from enrolment.

The obtained results also suggest that it is not advisable to train RNNs directly over raw EEG data, as when considering the $\mathbf{S}_T$ and $\mathbf{M}_\psi$ representations, and performed in [40]. On the other hand, the end-to-end network here proposed, applied on the $\mathbf{M}_T$ temporal representations of EEG signals, outperforms both classical approaches and state-of-the-art deep-learning-based proposals, requiring a time comparable with the one of the employed *EEGNet* to be trained on a single protocol. More interestingly, the best recognition rates are achieved when either RNNs or CNNs are trained over hand-crafted features instead of raw EEG data, with the joint exploitation of both kinds of networks typically guaranteeing the best recognition performance for EEG probes made of a single EEG frame. Due to the usage of EEG representations defined over a much smaller space, such as $\mathbf{M}_F$, networks using hand-crafted features as inputs are also characterized by a limited computational complexity, comparable with the one of *Shallow Net*. It therefore seems that extracting discriminative information from the

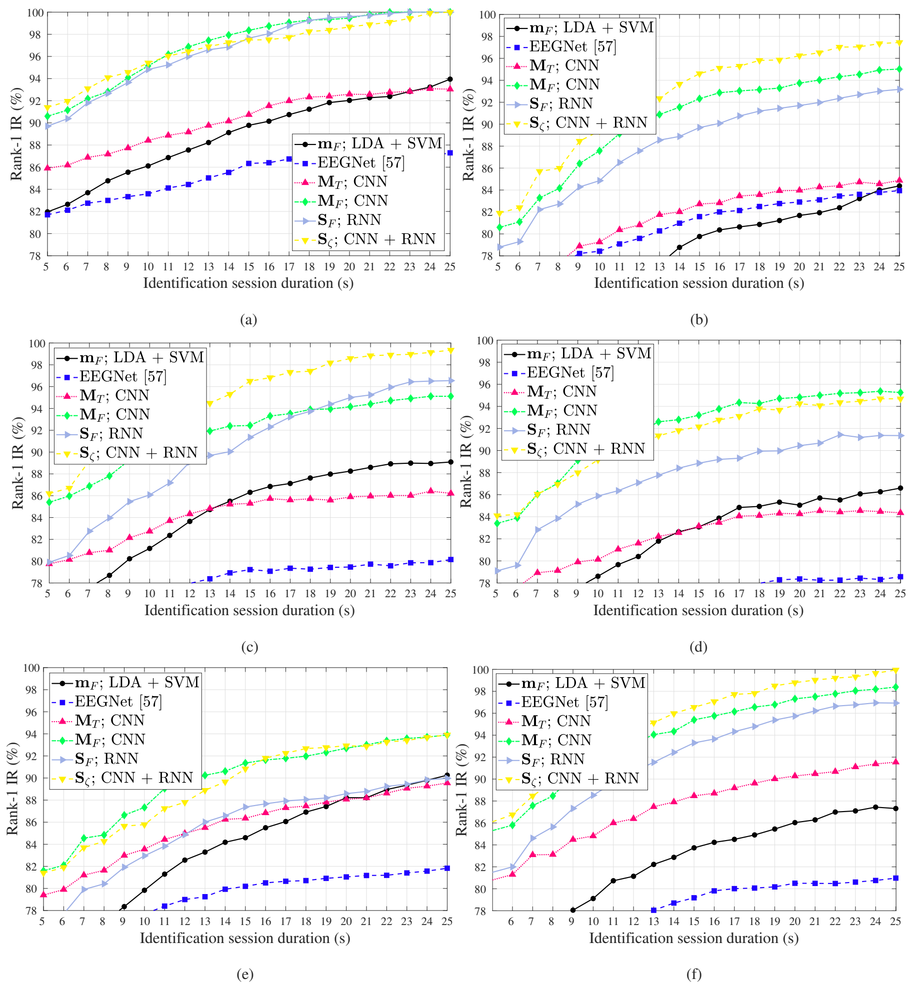

Figure 4: Rank-1 IR (mean, in %) versus identification time, considering different acquisition protocols. (a): EC; (b): EO; (c): MI; (d): SI; (e): VS; (f): MC.

originally acquired EEG signals, in order to reduce the amount of noise and the overall intra-class variability, turns out to be useful for effectively training neural networks, and thus achieving high recognition results. Similar behaviors have been observed when applying deep learning techniques to speech biometrics, with the effective use of frequency representations derived from the considered samples as input to neural networks [58].

When EEG probes longer than a single frame are available for identification, the

Table 7: Rank-1 IR (mean and confidence interval, in %) over a single EEG frame achievable with the best state-of-the-art approaches, and the neural networks here proposed, for EEG probes captured in any of the considered acquisition modalities.

$\mathbf{m}_F$; LDA + SVM	**Deep learning Approaches**				
	EEGNet [57]	Proposed			
		$\mathbf{M}_T$; CNN	$\mathbf{M}_F$; CNN	$\mathbf{S}_F$; RNN	$\mathbf{S}_\zeta$; CNN + RNN
77.9 $\pm$ 0.2	79.7 $\pm$ 0.3	84.4 $\pm$ 0.2	**86.9 $\pm$ 0.3**	82.8 $\pm$ 0.2	84.3 $\pm$ 0.3

outcomes from different frames can be combined according to a majority-voting rule to perform a decision, as depicted in Figure 1. Figure 4 shows the rank-1 IRs attainable, for increasing lengths of the employed identification samples, when considering the best representations and architectures among those presented in Section 3, that is, $\mathbf{M}_T$ with CNN, $\mathbf{M}_F$ with CNN, $\mathbf{S}_F$ with RNN, and $\mathbf{S}_\zeta$ with joint CNN + RNN.

It can be noticed that IRs greater than 90% can be achieved for all the considered stimulation protocols with identification sessions lasting at most 12 seconds, with perfect accuracy achievable for the EC scenario exploiting 22 seconds of EEG recordings for identification purposes. A significant improvement with respect to state-of-the-art approaches is therefore guaranteed by the proposed deep learning methods also when considering identification probes longer than a frame.

Further tests have also been carried out to verify whether neural networks can be exploited to effectively train a classifier for identifying a person regardless the specific mental task being performed. Toward this aim, the best architectures here proposed have been trained over sets of frames taken from 150s of consecutive recordings randomly-chosen from each of the first three sessions, while considering all the EEG data captured according to the six considered acquisition protocols as a whole. Frames taken from the fifth session of any of the considered protocols have been then used as identification probes, therefore performing the recognition phase without caring about the specific mental task performed by the subjects.

The achieved results are shown in Table 7, where the IRs attainable using a single frame as identification probe, exploiting either the proposed deep learning methods or the best state-of-the art approaches, are reported. The reported performance rates con-

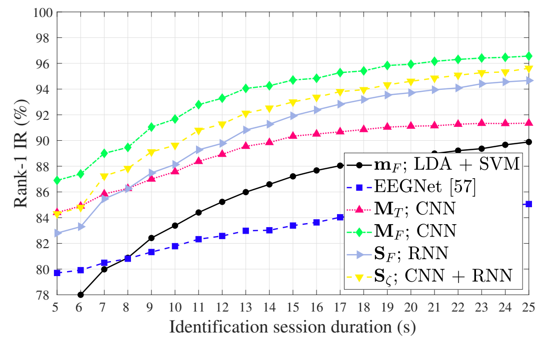

Figure 5: Rank-1 IR (mean, in %) versus identification time, for EEG probes captured in any of the considered acquisition modalities.

firm the large superiority of the deep learning approaches here proposed over methods known in literature to perform EEG-based identification. Interestingly, the obtained IRs are better than those achieved when performing training and testing over EEG data captured over a single protocol, with the exception of those attained with reference to EC conditions. Such behavior can be explained considering the greater amount of training data, i.e., six times larger than those exploited for each individual protocol, available when performing this kind of test. Having the possibility of exploiting larger amount of EEG data also notably assists the training of networks applied to raw EEG data represented as $\mathbf{M}_T$, with the performance gap with respect to the use of hand-crafted features notably reduced, compared to the task-dependent scenarios considered in Table 6. The amount of data available for training purposes is therefore highly-relevant for properly exploiting the designed architectures, especially when resorting to EEG representations in the form of raw temporal signals.

The behaviors reported in Figure 5, showing the rank-1 IRs obtained for increasing lengths of the considered identification session, testify that the networks here proposed achieve recognition performance greater than 96% with probes lasting 20s, performing EEG-based identification regardless the performed mental task, and considering a temporal distance of 15 months between enrolment and recognition. Conversely, approaches based on state-of-the-art networks are less effective than standard classifiers using LDA and SVM.

5. Conclusions

The present work has thoroughly evaluated the suitability and effectiveness of exploiting deep learning approaches to perform EEG-based biometric identification. An extensive set of experimental tests has been performed over a large longitudinal database comprising EEG data captured from 45 subjects during five recording sessions, with six acquisition modalities considered to evaluate the recognition performance achievable when the involved subjects perform different mental tasks.

The obtained results have highlighted the large superiority of the classifiers here proposed over standard methods exploiting LDA and SVM, and also over state-of-the-art methods employing neural networks to process EEG data. The proposed EEG representations and network architectures have allowed to achieve interesting IRs, larger than 96% for identification sessions lasting about 20s, and taken at a temporal distance greater than one year from the enrolment, when performing EEG-based biometric recognition irrespective of the performed mental task, therefore providing a solid foundation motivating the use of the considered biometric modality in real-life practical applications.

The performed tests have also shown the usefulness of training CNNs and RNNs over hand-crafted features extracted from the acquired EEG data, instead of directly on the raw data themselves. The present study therefore highlights the need for both investigating the definition of discriminative hand-crafted features to be used as inputs of neural networks, as well as designing novel end-to-end deep learning approaches which could be effectively trained, in order to fully exploit the discriminative capabilities of EEG biometric data and improve the currently achievable recognition performance. Toward such aim, approaches resorting to models such as generative adversarial networks (GANs) and autoencoders [59], or exploiting transfer learning from networks trained for other purposes on larger databases [60], could be for instance evaluated.

References

[1] K. Revett, Cognitive biometrics: a novel approach to person authentication, Int. Jour. Cognitive Biometrics 1 (1) (2012) 1–9 (2012).

[2] B. Chen, Y. Li, N. Zeng, Centralized wavelet multiresolution for exact translation invariant processing of ecg signals, IEEE Access 7 (2019) 42322–42330 (2019).

[3] S.Bianco, P. Napoletano, Biometric recognition using multimodal physiological signals, IEEE Access 7 (2019) 83581–83588 (2019).

[4] Q. Gui, M. Ruiz-Blondet, S. Laszlo, Z. Jin, A survey on brain biometrics, ACM Computing Surveys 51 (6) (2019) 1–38 (2019).

[5] P. Campisi, D. La Rocca, Brain waves for automatic biometric-based user recognition, IEEE Transactions on Information Forensics and Security 9 (5) (2014) 782–800 (2014).

[6] H. H. Stassen, Computerized recognition of persons by EEG spectral patterns, Electroencephalography and Clinical Neurophysiology 49 (1–2) (1980) 190–194 (1980).

[7] S. Yang, F. Deravi, On the usability of electroencephalographic signals for biometric recognition: A survey, IEEE Transactions on Human-Machine Systems 47 (6) (2017) 958–969 (2017).

[8] A. Pinegger, S. Wriessnegger, J. Faller, G. Muller-Putz, Evaluation of different EEG acquisition systems concerning their suitability for building a brain[U+0096]computer interface: Case studies, Frontiers in Neuroscience 10 (441) (2016) 1–11 (2016).

[9] Y. Chen, A. Atnafu, I. Schlattner, W. T. Weldtsadik, M.-C. Roh, H. Kim, S.-W. Lee, B. Blankertz, S. Fazli, A high-security EEG-based login system with RSVP stimuli and dry electrodes, IEEE Transactions on Information Forensics and Security 11 (12) (2016).

[10] M. Ruiz-Blondet, Z. Jin, S. Laszlo, CEREBRE: A novel method for very high accuracy event-related potential biometric identification, IEEE Transactions on Information Forensics and Security 11 (7) (2016) 1618–1629 (2016).

[11] J.-H. Kang, Y. C. Jo, S.-P. Kim, Electroencephalographic feature evaluation for improving personal authentication performance, Neurocomputing 287 (2018) 93–1019 (2018).

[12] E. Maiorana, D. La Rocca, P. Campisi, On the permanence of EEG signals for biometric recognition, IEEE Transactions on Information Forensics and Security 11 (1) (2016) 163–175 (2016).

[13] E. Maiorana, P. Campisi, Longitudinal evaluation of EEG-based biometric recognition, IEEE transactions on Information Forensics and Security 13 (5) (2018) 1123–1138 (2018).

[14] B. Armstrong, M. Ruiz-Blondet, N. Khalifian, K. Kurtz, Z. Jin, S. Laszlo, Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics, Neurocomputing 166 (2015) 59–67 (2015).

[15] L. Bozhkov, Overview of deep learning architectures for classifying brain signals, KSI Transactions on knowledge society 9 (1) (2016).

[16] Y. Tabar, U. Halici, A novel deep learning approach for classification of EEG motor imagery signals, Journal of Neural Engineering 14 (2017).

[17] L. Hernández, O. Mozos, J. Ferrandez, J. Antelis, EEG-based detection of braking intention under different car driving conditions, Frontiers in neuroinformatics 12 (29) (2018).

[18] M.-P. Hosseini, D. Pompili, K. Elisevich, H. Soltanian-Zadeh, Optimized deep learning for EEG big data and seizure prediction BCI via internet of things, IEEE Transactions on Big Data 3 (4) (2017).

[19] Z. Wen, R. Xu, J. Du, A novel convolutional neural networks for emotion recognition based on EEG signal, in: International conference on security, pattern analysis, and Cybernetics (SPAC), 2017 (2017).

[20] S.-E. Moon, S. Jang, J.-S. Lee, Convolutional neural network approach for EEG-based emotion recognition using brain connectivity and its spatial information, in: IEEE ICASSP, 2018 (2018).

[21] D. Gurve, S. Krishnan, Deep learning of EEG time-frequency representations for identifying eye states, Advances in data science and adaptive analysis 10 (2) (2018).

[22] P. Nejedly, et al., Intracerebral EEG artifact identification using convolutional neural networks, Neuroinformatics 17 (2) (2019) 225–234 (2019).

[23] L. Ma, J. Minett, T. Blu, W.-Y. Wang, Resting state EEG-based biometrics for individual identification using convolutional neural networks, in: IEEE EMBC, 2015 (2015).

[24] Q. Gui, W. Yang, Z. Jin, A residual feature-based replay attack detection approach for brainprint biometric systems, in: IEEE WIFS, 2016 (2016).

[25] R. Das, E. Maiorana, P. Campisi, Motor imagery for EEG biometrics using convolutional neural network, in: IEEE ICASSP, 2018 (2018).

[26] Z. Mao, W. Yao, Y. Huang, EEG-based biometric identification with deep learning, in: IEEE EMBC, 2017 (2017).

[27] R. Das, E. Maiorana, P. Campisi, Visually evoked potential for EEG biometrics using convolutional neural network, in: EUSIPCO, 2017 (2017).

[28] T. Schons, G. Moreira, P. Silva, V. Coelho, E. Luz, Convolutional network for EEG-based biometric, in: CIARP, 2017 (2017).

[29] H. El-Fiqi, et al., Convolution neural networks for personidentification and verification using steady statevisual evoked potential, in: IEEE International Conference on Systems, Man, and Cybernetics, 2018 (2018).

[30] T. Yu, C.-S. Wei, K.-J. Chiang, M. Nakanishi, T.-P. Jung, EEG-based user authentication using a convolutional neural network, in: IEEE EMBS International Conference on Neural Engineering, 2019 (2019).

[31] Q. Wu, Y. Zeng, C. Zhang, L. Tong, B. Yan, An eeg-based person authentication system with open-set capability combining eye blinking signals, Sensors 18 (335) (2018) 1–18 (2018).

[32] O. Ozdenizci, Y. Wang, T. Koike-Akino, D. Erdogmus, Adversarial deep learning in EEG biometrics, Vol. 26, 2019, pp. 710–714 (2019).

[33] S. Roy, I. Kiral-Kornek, S. Harrer, Deep learning enabled automatic abnormal EEG identification, in: IEEE EMBC, 2018 (2018).

[34] M. Li, W. Zhu, M. Zhang, Y. Sun, Z. Wang, The novel recognition method with optimal wavelet packet and lstm based recurrent neural network, in: IEEE International Conference on Mechatronics and Automation, 2017 (2017).

[35] C.-L. Z. T.-J. Luo, F. Chao, Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network, BMC Bioinformatics 19 (1) (2018).

[36] R. K. Maddula, et al., Deep recurrent convolutional neural networks for classifying P300 BCI signals, in: Graz Brain-Computer Interface Conference, 2017 (2017).

[37] S. Alhagry, A. Fahmy, R. El-Khoribi, Emotion recognition based on EEG using lstm recurrent neural network, International Journal of Advanced Computer Science and Applications 8 (10) (2017).

[38] R. G. Hefron, et al., Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation, Pattern Recognition Letters 94 (15) (2017).

[39] S. Kuanar, et al., Cognitive analysis of working memory load from EEG, by a deep recurrent neural network, in: IEEE ICASSP, 201 (201).

[40] X. Zhang, L. Yao, S. Kanhere, Y.Liu, T. Gu, K. Chen, Mindid: Person identification from brain waves through attention-based recurrent neural network, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2 (3) (2018) 1–23 (2018).

[41] Y. Sun, F.-W. Lo, B. Lo, Eeg-based user identification system using 1d-convolutional long short-term memory neural networks, Expert Systems With Applications 125 (2019) 59—-26 (2019).

[42] D. McFarland, L. McCane, S. David, J. Wolpaw, Spatial filter selection for EEG-based communication, Electroencephalography and Clinical Neurophysiology 103 (3) (1997) 386–394 (1997).

[43] P. Campisi, et al., Brain waves based user recognition using the [U+0093]eyes closed resting conditions[U+0094] protocol, in: IEEE WIFS, 2011 (2011).

[44] S. Kay, Modern Spectral Estimation. Theory and Applications, Prentice-Hall, 1988 (1988).

[45] P. Nguyen, D. Tran, X. Huang, D. Sharma, A proposed feature extraction method for EEG-based person identification, in: International Conference on Artificial Intelligence (ICAI), 2012 (2012).

[46] P. Davis, C. Creusere, J. Kroger, Subject identification based on EEG responses to video stimuli, in: IEEE International Conference on Image Processing (ICIP), 2015 (2015).

[47] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: ICLR, 2015 (2015).

[48] R. Das, E. Maiorana, P. Campisi, EEG biometrics using visual stimuli: A longitudinal study, IEEE Signal Processing Letters 23 (3) (2016) 341–345 (2016).

[49] E. Niedermeyer, F. Da Silva, Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, Doody's all reviewed collection, Lippincott Williams & Wilkins, 2005 (2005).

[50] R. Barry, A. Clarke, S. Johnstone, C. Magee, J. Rushby, EEG differences between eyes-closed and eyes-open resting conditions, Clinical Neurophysiology 118 (2007) 2765–2773 (2007).

[51] W. Klimesch, M. Doppelmayr, T. Pachinger, H. Russegger, Event-related desynchronization in the alpha band and the processing of semantic information, Cognitive Brain Research 6 (2) (1997) 83–94 (1997).

[52] D. J. McFarland, L. A. Miner, T. M. Vaughan, J. R. Wolpaw, Mu and beta rhythm topographies during motor imagery and actual movements, Brain topography 12 (3) (2000) 177–186 (2000).

[53] T. Harmony, T. Fernandez, J. Silva, J. Bernal, L. Diaz Comas, A. Reyes, E. Marosi, M. Rodriguez, M. Rodriguez, EEG delta activity: an indicator of attention to internal processing during performance of mental tasks, International Journal of Psychophysiology 24 (2) (1996) 161–171 (1996).

[54] E. Maiorana, D. La Rocca, P. Campisi, Eigenbrains and eigentensorbrains: Parsimonious bases for EEG biometrics, Neurocomputing 171 (2016) 638[U+0096]–648 (2016).

[55] K. Brigham, B. V. Kumar, Subject identification from electroencephalogram (EEG) signals during imagined speech, in: IEEE BTAS, 2010 (2010).

[56] R. Schirrmeister, J. Springenberg, L. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for eeg decoding and visualization, Human Brain Mapping 38 (11) (2017) 5391—5420 (2017).

[57] V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung, B. J. Lance, EEG-Net: A compact convolutional neural network for EEG-based brain-computer interfaces, Journal of Neural Engineering 15 (5) (2017) 5391—5420 (2017).

[58] A. Boles, P. Rad, Voice biometrics: Deep learning-based voiceprint authentication system, in: IEEE System of Systems Engineering Conference (SoSE), 2017 (2017).

[59] A. Ditthapron, N. Banluesombatkul, S. Ketrat, E. Chuangsuwanich, T. Wilaiprasitporn, Universal Joint Feature Extraction for P300 EEG Classification Using Multi-Task Autoencoder, IEEE Access 7 (2019) 68415–68428 (2019).

[60] C. Tan, F. Sun, W. Zhang, Deep Transfer Learning for EEG-Based Brain Computer Interface, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018 (2018).