A deep descriptor for cross-tasking EEG-based recognition

Mariana RL Mota[1]**, Pedro HL Silva**[1]**, Eduardo JS Luz**[1]**, Gladston JP Moreira**[1]**, Thiago Schons**[1]**, Lauro AG Moraes**[1]**, and David Menotti**[2]

[1]**Department of Computing, Universidade Federal de Ouro Preto, Ouro Preto, Minas Gerais, Brazil.**

[2]**Department of Informatics, Universidade Federal do Paraná, Curitiba, Paraná, Brazil.**

Corresponding author:

Gladston JP Moreira[1]

Email address: gladston@ufop.edu.br

ABSTRACT

Due to the application of vital signs in expert systems, new approaches have emerged, and vital signals have been gaining space in biometrics. One of these signals is the electroencephalogram (EEG). The motor task in which a subject is doing, or even thinking, influences the pattern of brain waves and disturb the signal acquired. In this work, biometrics with the EEG signal from a cross-task perspective are explored. Based on deep convolutional networks (CNN) and Squeeze-and-Excitation Blocks, a novel method is developed to produce a deep EEG signal descriptor to assess the impact of the motor task in EEG signal on biometric verification. The Physionet EEG Motor Movement/Imagery Dataset is used here for method evaluation, which has 64 EEG channels from 109 subjects performing different tasks. Since the volume of data provided by the dataset is not large enough to effectively train a Deep CNN model, it is also proposed a data augmentation technique to achieve better performance. An evaluation protocol is proposed to assess the robustness regarding the number of EEG channels and also to enforce train and test sets without individual overlapping. A new state-of-the-art result is achieved for the cross-task scenario (EER of 0.1%) and the Squeeze-and-Excitation based networks overcome the simple CNN architecture in three out of four cross-individual scenarios.

INTRODUCTION

Nowadays, different biometric modalities are being explored. They have been gradually replacing the logging/password systems, once it represents the future path in terms of digital security. Among them, one of the categories which has gained attention is the vital signals based biometric modalities, such as the Electrocardiogram (ECG) (Luz et al., 2018; Silva et al., 2019; Garcia et al., 2017) and the Electroencephalogram (EEG) (Schons et al., 2017; Poulos et al., 1999; Carrión-Ojeda et al., 2020). The EEG, initially used for clinical purposes, have already been used by consumers and incorporated into portable devices for applications not directly related to medicine, as in the gaming industry, education systems aimed at monitoring engagement and in EEG-based Brain-Computer Interface (Sawangjai et al., 2019). Nonetheless, automatic disease detection (Islam et al., 2015) and an individual identification employing brain waves are still challenging tasks (Yang and Deravi, 2017).

The EEG signal is the measurement of the electrical activity of the brain captured by electrodes strategically positioned on the scalp (Boubakeur et al., 2017). The first use of the EEG signal as biometrics dates back to 1980 in Stassen's seminal work (Stassen, 1980), using a speech recognition technique to characterize the EEG spectrum pattern of a person. Since then, several approaches have investigated the EEG as a biometric modality (Das et al., 2018; Fraschini et al., 2015; Yang et al., 2018; Gui et al., 2019).

Researches on EEG as biometrics have shown that this signal satisfies three out four requirements to be considered as a biometric modality (Jain et al., 2004): 1) collectability: it can be measured and quantified; 2) unicity: it provides distinction among individuals; 3) universality: the element exists in all person. The fourth requirement, persistence over time, is still an open problem in the literature (Marcel and Millán, 2007; Kostílek and Šťastnỳ, 2012).

Figure 1. Example of an EEG headband with few electrodes and no need for conductor gel.

Many authors have addressed the EEG signal as a biometric modality with traditional machine learning techniques (Fraschini et al., 2015; Singh et al., 2015; Yang et al., 2018). Today, deep learning represents the state-of-the-art for several computer vision and pattern recognition problems and also there are works in the literature addressing deep-learning-based biometrics on EEG (Ma et al., 2015; Das et al., 2017, 2018; Mao et al., 2017; Wang et al., 2019; Wilaiprasitporn et al., 2019). Although, it is well known that methods based on deep learning need large amounts of data, especially, deep convolution-based architectures, a workaround for this issue is to artificially create new samples, such as the data augmentation techniques. In a preliminary work (Schons et al., 2017), a data augmentation technique was proposed to increase the training data, improving the network performance, and enabling the deep learning model to converge for the Physionet - EEG Motor Movement/Imagery Dataset.

It has been shown in previous works evidence that EEG has biometric potential under the performance of different tasks (Vinothkumar et al., 2018; DelPozo-Banos et al., 2018; Kong et al., 2018; Kumar et al., 2019; Fraschini et al., 2019). In this work, the explore EEG as a biometric modality under this challenging multi-task perspective. In a real situation, a person can be in motion, or even, imagining an action, and with that, generating interference that can be captured by the EEG signal. Thus, a deep feature descriptor approach is proposed in order to mitigate the impact of Movement/Imagery tasks on the EEG biometrics. A evaluation whether EEG biometrics benefit from any specific task, or from its nature (motor or imaginary) is conducted. The relevance of the number of channels is also investigated as well as a more robust model based on Squeeze-and-Excitation blocks (Hu et al., 2018). The Squeeze-and-Excitation blocks promote a channel-wise feature response and since the EEG signal can have up to 64 channels and each channel captures different nuances of brain movement, these blocks are a promising research path for this type of biometric modality.

This work extends the one presented in the 22nd CIARP (Iberoamerican Congress on Pattern Recognition) (Schons et al., 2017). Here, the following improvements are described:

- It presents results for a cross-task (one task is used to train and another to test) multi-task (two or more tasks are used to train and another to test), and cross-individual scenario (the individual used to train is not used to test) extending the approach showed in (Schons et al., 2017), considering motor task interference.

- It presents an extensive literature review.

- It explores the Squeeze-and-Excitation blocks instead of conventional CNN blocks.

- A better description of the methodology/experimentation process and detailed analysis regarding parameter tuning.

The contributions of this work are summarized as follow:

- The new state-of-the-art method for EEG-based biometric verification evaluated in a protocol aiming a cross-task scenario.

- A data augmentation technique based on the exploration of the overlap between signals used during the training phase.

- Competitive results with fewer EEG channels.

- An in-depth discussion about the implications of using multiple motors / imaginary tasks for biometrics with EEG.

The results reported in this work corroborates with the feasibility of the EEG as a biometric modality, once the evaluation carried with different tasks in different runs have reported promising results. Although our experiments reveal that physical motor tasks (not imaginary) hinder the task of identification, making the problem more challenging.

With more ubiquitous acquisition method (as illustrated in Figure 1) and the advancement of hardware accelerators for deep learning, there is a large potential for its usage in real-life applications.

The remaining of this work is organized as follows. Related works are presented in the second section. In the third section, the proposed methodology is presented, while the experiments and results are described in the fourth section. Finally, the conclusions are pointed out in the fifth section.

DATASET AND RELATED WORKS

In this section, a popular benchmark dataset is presented: EEG Motor Movement/Imagery dataset (Fraschini et al., 2015; Yang et al., 2018; Sun et al., 2019), available in Physionet Resource[1]. The EEG Motor Movement/Imagery dataset is one of the most comprehensive datasets of EEG signals and also contemplates motor tasks during the acquisition sessions.

The majority of the works presented in this section use the EEG Motor Movement/Imagery dataset (Goldberger et al., 2000; Schalk et al., 2004). However, other datasets were considered by authors in the literature; thus, the presented methods are categorized according to the signals type used in the evaluation: baseline task or single-task and multi-task.

EEG Motor Movement/Imagery Dataset

The EEG Motor Movement/Imagery dataset is a popular open-access dataset that consists of over 1,500 recordings of EEG signals obtained from 109 volunteers, with one-minute or two-minutes lengths. The EEG signal was obtained from 64 electrodes distributed over the scalp. The EEG subset contains a total of 109 subjects and 14 acquisition records. From the acquisition records, there are two motor and two imaginary task types and two baseline tasks.

The baseline tasks capturing occurred in a controlled scenario in which the individual follows the command to open and close their eyes during specifics moments. Except for the baseline, all other tasks have three acquisition sessions.

The records of three sessions of four different motor/imaginary tasks result in 12 sessions. Regarding the 12 sessions, there are four distinct motor or imaginary task types (T1-T4), each one with three runs (R1-R3). The activities assigned to the tasks are:

- Task 1 (T1): In a screen, a target appears on the left or right side, and a subject is requested to open and close the corresponding fist on the same side until the target disappears. After that, the subject may relax.

- Task 2 (T2): In a screen, a target appears on the left or right side, and the subject is requested to imagine opening and closing the corresponding fist on the same side until the target disappears. After that, the subject may relax.

- Task 3 (T3): In a screen, a target appears at the top or bottom of a screen. If the target is on the top, the subject is asked to open and close both fists. Otherwise, if the target is in on the bottom, the subject is requested to open and close both feet. This task runs until the target disappears. After that, the subject may relax.

- Task 4 (T4): On a screen, a target appears at the top or bottom of a screen. If the target is on the top, the subject is asked to imagine opening and closing both fists. Otherwise, if the target is in on the bottom, the subject is requested to imagine opening and closing both feet. This task runs until the target disappears. After that, the subject may relax.

---
[1] https://physionet.org/content/eegmmidb/1.0.0/

Related Works

It is possible to summarize the related works evaluated under the EEG Motor Movement/Imagery dataset into two groups: (1) the ones focused on the baseline task, i.e., eyes close (EC) and eyes open (EO) tasks, and (2) the ones focused on the multi-task recognition. Regarding the first group, the individuals are in a rest state with the eyes closed or opened (Del Pozo-Banos et al., 2014). In contrast, in the second group, the individuals were requested to do simple motor or imaginary tasks.

The work in (Gui et al., 2019) presents a survey on EEG signals biometry. The authors reviewed the state-of-the-art for the recognition task and concluded that the EEG may has some advantages over other biometrics markers. One of the significant advantages of the EEG to consider is that the user must agree and consent to record the EEG signal what difficulties the theft of this kind of signal. Gui et al. (2019) also explored all steps regarding the EEG signal, from the acquisition to the classification, and described several EEG datasets from the literature. Moreover, the authors also detected open problems related to brain biometrics, such as multi-modality and the persistence of the EEG. According to the authors, the EEG signal has a great potential as biometry from the theoretical point of view and due to results obtained from empirical experiments.

Singh et al. (2015) proposed an approach to find discriminative characteristics described by unique patterns from the relations generated among the cerebral regions. The authors preprocessed the data by applying a low-pass filter to reduce the noise and used a technique called Magnitude Squared Coherence, which relies on the phase constancy while a brain area is interacting with another, generate the signal descriptors. The classification process performed the K-Nearest Neighbors (KNN) algorithm. The proposed approach evaluated 64, 10, 6, and 5 channels for the EEG signal capturing and achieved a final accuracy score for the identification reported task of 100% with 64 electrodes. This experiment corroborates the hypothesis that the more channels/information employed, the better will be the result.

Fraschini et al. (2015) evaluated a biometric verification approach using all the 64 electrodes available in the dataset. The proposed approach can be divided into four steps: (1) filter the raw data to allow a study of the specific frequency data from the cerebral network; (2) estimate the statistical independency in pairs among EEG temporal series; (3) create a weighted graph where each edge represents a functional connection among the connected electrodes on the head; and (4) perform a characterization of the cerebral functional organization aiming a centrality measure to qualify the importance of each node inserted in the graph. The result is a 64-length-vector used to classify and to report the results segmented by the frequency. The authors reported an EER of 4.4% using the gamma frequency (30-50 Hz) and observed a difficulty enhancement for individual verification when the frequency band gets slower.

CNNs are also evaluated for the EEG problem. Ma et al. (2015) created a CNN architecture with two convolution layers, two pooling layers, and one fully connected layer. However, the authors did not use all 109 individuals. Instead, Ma et al. (2015) evaluated their approach with 10 individuals for testing. The authors divided the 60 seconds of EO and EC in 55 seconds in a one-second fragment for training and the remaining five seconds for testing. The identification rate reported was in the range of 64% to 86%.

Das et al. (2017) evaluated a CNN with four convolutional layers, two max-pooling, one rectified linear unit (ReLU), and a softmax-loss layer. The dataset used to perform the experiments were acquired from 40 subjects over two distinct sessions separated by a week in a visual stimuli environment. The correct recognition rate (CRR) reported was in the range between 80.65% and 98.8% in the best scenario.

Das et al. (2018) also performed EEG biometry with a CNN containing four convolutional layers, two max-pooling, a ReLU, and a softmax-loss layer. The dataset evaluated consists of 40 subjects performing imaginary arms and legs movements collected in two sessions with an interval of two weeks. The authors achieved the accuracy scores of 81.25% and 93% for rank-1 and rank-2, respectively.

Mao et al. (2017) proposed another approach using CNN. The authors used convolutional layers with ReLu and max-polling, followed by fully connected layers with softmax. The evaluation was made in a dataset containing data of 100 subjects using 64 channels during a driving experiment. The CRR achieved was 97%.

Maiorana and Campisi (2018) analyzed the discriminative characteristics of EEG in longitudinal behavior, aiming to verify the pertinence across time. The study utilized the signals captured from 45 users during approximately 36 weeks from the first EEG data recording in five to six sessions. The results showed that aging could damage the EEG traits, but yet the authors could achieve EER below 2% from one data collection to another distant in time.

Wang et al. (2019) evaluated the impact of different motor and imaginary tasks in the identification

scenario. The authors used all the 109 individuals of the EEG Motor Movement/Imagery Dataset and also created a dataset with 59 subjects performing attention and descriptive tasks. The proposed approach is a Graph Convolutional Neural Network-based to extract discriminating features from the EEG graphs created by the Phase Locking Value (PLV) algorithm. Using the same task to train and test, the authors reported an average CRR of 99.96% and 98.94% for the EEG Motor Movement/Imagery Dataset and the created datasets, respectively. A degradation of 86.21% was observed when trained on resting-state data and tested in diverse states in the EEG Motor Movement/Imagery Dataset. However, mixing signals in different states for training the GCNN resulted in an average CRR of 99.98% with the Physionet bank and 99.96% with the author's personal acquisition bank.

Yang et al. (2018) also evaluated the impact of motor/imaginary tasks during the recognition process in biometric identification and verification tasks. The authors used a discrete Wavelet transform in each electrode and created the feature vector as the concatenation of the standard deviations of all electrodes. The vector created is used as the input for a Linear Discriminant Analysis (LDA) classifier. A majority voting rule merges all classifier decisions of all time windows. In one of the experiments performed, the authors trained on tasks T1 + T2, tested on T1R2 using 9 electrodes and reached an identification accuracy of 100% and an EER of 2,63%. They also analyzed the impact of using only 3 channels and achieved a maximum accuracy of 89% when training with T1R1 + T1R3 and testing with T1R2. In the verification scenario, the EER is approximately three times greater using the reduced number of electrodes.

To find the best set of electrodes for an identification problem, Alyasseri et al. (2020) used the EEG Motor Movement/Imagery Dataset as a single matrix containing the signals from all tasks performed by the 109 individuals. The authors propose a hybrid optimization technique to find the most relevant channels to extract discriminant characteristics, using the Flower Pollination Algorithm (FPA) and the β-hill-climbing algorithm (β-hc). Besides, the proposed approach presented a study of the best domain to extract the characteristics of an EEG signal, comparing the use of characteristics in the time domain, frequency domain, and both simultaneously. The hybrid method FPAβ-hc obtained better results than isolated FPA in most cases. The accuracy reported of 96.05% was achieved using a Support Vector Machine (SVM) classifier with the RBF kernel and the characteristics in both the time and frequency domain simultaneously.

In the work proposed in (Sun et al., 2019), the authors applied a hybrid convolution and recurrent deep neural network with a Long-Short Time Memory (LSTM) for individual identification. The authors tested the use of 4, 16, 32, and 64 electrodes. Besides, instead of using 12-seconds length segments, a 1-second length segment was used and reached 0.41% of EER with 16 electrodes. The focus of their discussion is the trade-off between performance and the use of the electrodes/recording time. The authors outweighed the loss in EER to the reduction of the electrodes, and, according to the authors, 12-second segments are not feasible for real applications. Despite the outstanding results presented by the authors, it is worth highlighting that the protocol is different from the one showed in this work and in (Yang et al., 2018). The authors evaluated two different scenarios compared to the one presented in (Yang et al., 2018). The rationale for the first scenario is that the authors use 90% to train/validate and 10% to test, that is, the same individual signal is present in training and testing. This scenario is more straightforward than the one proposed by Yang et al. (2018). The reported results cannot be compared to the presented in (Yang et al., 2018), once it uses only a one-second signal, and the evaluation protocol is different.

Some works in the literature do not use the EEG Motor Movement/Imagery dataset but present relevant results and alternative techniques that inspired the development of this work.

The work of El-Fiqi et al. (2018) proposed a CNN with raw steady-state visual evoked potentials (SSVEPs) for individual identification and verification. In the SSVEPs acquisition protocols, the subject focuses its attention on a repetitive visual stimulus while recording the activity response signals. The study explores two SSVEP datasets with raw data of four and ten subjects that were seated and focused on three groups of flashing LED stimuli blinking at 13Hz, 17Hz, and 21Hz frequencies. The signals were recorded at 256Hz by eight electrodes placed on the parietal-occipital area. The performance of three classical methods (support vector machine, random forest, and shallow feed-forward network with one layer) with spectral features were compared against the proposed CNN method, which works directly with the raw signal. The deep learning-based approach achieved an averaged identification accuracy of 96.80% and an averaged verification accuracy of 98.34%. These results outperform those obtained by other classical methods. Besides, the proposed CNN needs no complex techniques for feature representation or extraction to achieve good results, making it feasible for real-time systems.

Table 1 summarizes these related works.

Table 1. Related works for EEG-based biometric.

Work	Database	Classes	Acquisition	Approach	Channels	Result
Singh et al. (2015)	Physionet	109	Motor/Imaginary tasks	Magnitude Squared Coherence	64	Acc = 100%
Fraschini et al. (2015)	Physionet	109	Rest	Eigenvector	64	EER = 4.4%
Ma et al. (2015)	Physionet	10	Rest	CNN	64	Acc = 88%
Das et al. (2017)	Own	40	Visual stimuli	CNN	17	CRR = 98.8%
Das et al. (2018)	Own	40	Imaginary arms/legs movement	CNN	17	CRR = 93%
Mao et al. (2017)	BCIT	100	Driving car	CNN	64	CRR = 97%
Maiorana and Campisi (2018)	Own	45	Sit	Hidden Markov models (HMMs)	9	EER < 2%
Wang et al. (2019)	Physionet	109	Motor/Imaginary tasks	GCNN + PLV	64	CRR = 99.98% FAR = 1.65%
Yang et al. (2018)	Physionet	108	Motor/Imaginary tasks	Discrete Wavelet Transform + LDA	9	Acc = 100% EER = 2.63%
Alyasseri et al. (2020)	Physionet	109	Motor/Imaginary tasks	FPA + β-hill	35	Acc = 96.05%
Sun et al. (2019)	Physionet	109	Motor/Imaginary tasks	1D-Conv. LSTM	16	EER = 0.41%
El-Fiqi et al. (2018)	SSVEP database	14	Visual stimulus	CNN	8	Verification Acc = 98.34%

PROPOSED APPROACH AND EVALUATION SCENARIOS

In this section, the proposed method is presented and outlined as the pipeline shown in Figure 2. The method is evaluated on the EEG Motor Movement/Imagery Dataset (Goldberger et al., 2000; Schalk et al., 2004), under the three scenarios evaluated in this work and presented in Section .

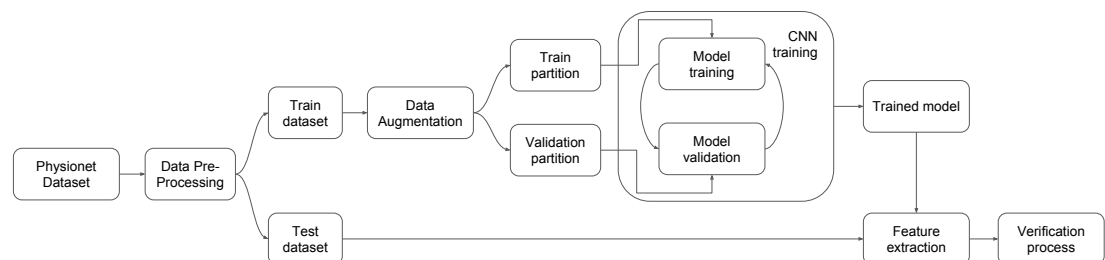



Figure 2. Proposed pipeline.

In order to execute the pipeline proposed in Figure 2, preprocessing methodology is developed along with the data augmentation protocol, the CNN model, and the respective data representation along with the evaluation process. The steps presented in Figure 2 are described in details in the next subsections.

Data preprocessing

Biometrics, disease detection, and brain death detection are two of many applications that could be employed based on the EEG signal. Regardless of the application, preprocessing is required since the signal is very susceptible to noise. Several approaches have been used in the literature for data processing, such as Common Spatial Pattern (CSP) (Yong et al., 2008), Wavelet transforms (Kumar et al., 2009; Ghandeharion and Ahmadi-Noubari, 2009), and Independent Component Analysis (ICA) (Delorme et al., 2007).

For biometrics purposes, the band-pass FIR/IIR filters are popular preprocessing techniques. Once each frequency is related to a specific cerebral activity (Boubakeur et al., 2017), filtering out unwanted frequency-bands aids in the detection process. In this work, a band-pass filter is applied aiming only

259 the intermediate frequency-bands of the signal spectrum. Since the data acquired by equipment used in
260 Physionet EEG Motor Movement/Imagery Dataset was sampled at 160Hz, it has well-defined components
261 up to 80Hz. Besides, it is not usual to employ the full spectrum, since some frequency bands are more
262 discriminant than others for biometry (Fraschini et al., 2015). According to Yang and Deravi (2017),
263 signals related to EEG biometrics have higher energy in the frequency spectrum below 50Hz. In that
264 manner, filters are built for the following bands: 01-50Hz, 10-30Hz, and the gamma band (30-50Hz),
265 following the ones proposed by Fraschini et al. (2015).

266   The data division in training and testing dataset is defined according to the evaluated scenario. For
267 the first one (baseline task evaluation), the EC is used for testing. For the second scenario (multi-task
268 evaluation), the data division proposed by Yang and Deravi (2017) is followed. In the third scenario
269 (cross-task evaluation), proposed in this work, the T1R2 is used as testing. Finally, the forth scenario
270 (cross-individual evaluation), also proposed in this work, the EC and EO are used for testing.

271 **Data augmentation**
272 The evaluation protocol in (Fraschini et al., 2015) proposes the division of the EEG data into 12-second
273 segments (1920 samples). Since the recording motor/imagery tasks have 120 seconds, each task has a
274 total of 10 segments of 12 seconds without overlap. From both 60-second baseline recording sessions
275 (EO and EC), five segments are created, disregarding the overlap. A total of only five or ten segments for
276 each individual is not enough for the CNN training to converge.

277   A workaround for this issue is a data augmentation technique for the training set. The strategy consists
278 of abundantly extract segments with overlapping from the EEG segments. In this work, the impact of the
279 data augmentation technique based on segments overlapping is investigated. The technique consists of
280 creating new segments by a sliding window that moves by a constant step called stride. Figure 3 shows an
281 example of the data augmentation technique used.

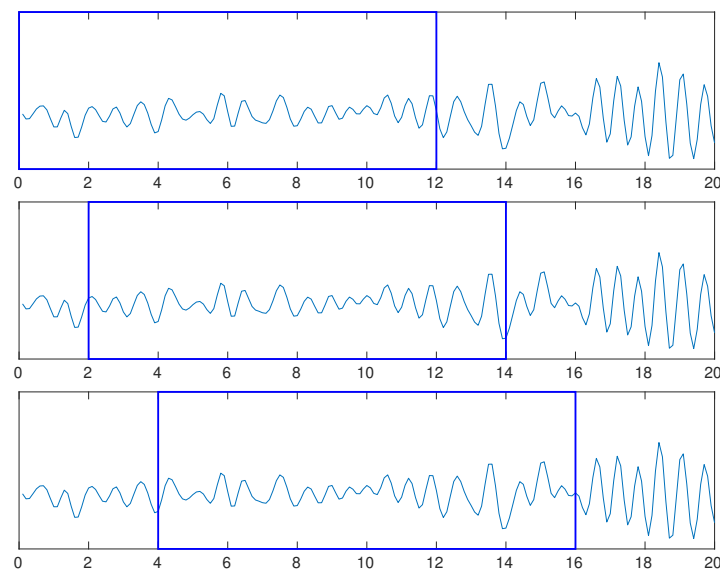



Figure 3. Example of the proposed data augmentation based on overlapping with a step of two-second size. Each slide of the window produces a new training instance.

282   From the whole signal of an individual record, new segments are created. The extraction of the first
283 segment starts from time zero and finishes within 12 seconds. The 12-second record represents 1920
284 samples at a sampling frequency of 160Hz. The next segment starts after a specific stride in $stride$ seconds
285 and finishes in $stride + 12$ seconds. The sliding window finishes at the end of the signal.

286 **CNN training and data representation**
287 After the data augmentation step, the data is split into two subsets: train and validation. The first 90%
288 signals resulted from the data augmentation are used for training, and the remaining 10% for validation
289 (model adjusting during training), as presented in Figure 4. It is worth highlighting that not necessarily

290 10% of the entire segment is used for validation, but 10% of the signals resulted from the data augmentation
291 step. For example, if the Data Augmentation process output is 500 segments, the first 450 segments (90%)
292 are for training and the last 50 segments (10%) for validation.

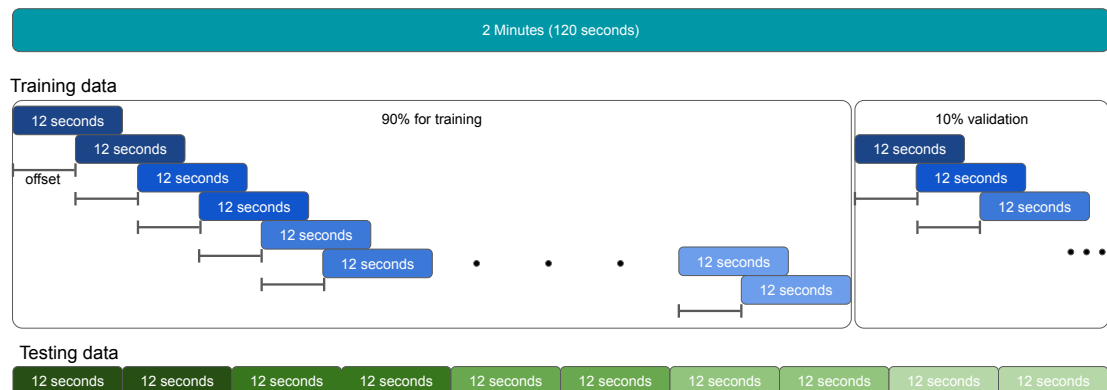



Figure 4. Data split proposed.

293   At first, the CNN architectures are evaluated with training and validation data. This evaluation is
294 handled on the baseline evaluation perspective, once it is a more constrained scenario in comparison to
295 the multi-task scenario perspective.

296   Two different one-dimensional CNN architectures with large receptive fields are investigated, once
297 those with small kernels did not converge.

298   The two architectures are the result of the empirical adjustments in the depth and width according to
299 the error measured in the validation data, during the training phase.

300   The first CNN architecture is composed of three convolution layers, each one followed by max-pooling,
301 and four fully-connected layers with a dropout and a softmax layer. The convolution stride is set to one
302 and no padding is used. The main among architectures investigated here is related to the number of
303 convolutional layers. The second architecture is composed of five convolution layers with larger receptive
304 fields in the initial layers. Both architectures are presented in Figure 5.

305   Both architectures are evaluated according to the protocol proposed by Fraschini et al. (2015) and
306 thus, the one that best fits the data can be determined.

307   Each architecture is trained for n epochs, in which n relies on the convergence of the training and
308 validation error. The training process resembles a simple classification problem for an identification task
309 (close-gallery) on a biometry context. Thus, a conventional supervised classification problem. Upon this
310 fact, each sample provided to a CNN model is feed-forwarded and results in an output. This output is a
311 probability that a sample belongs to one of the classes.

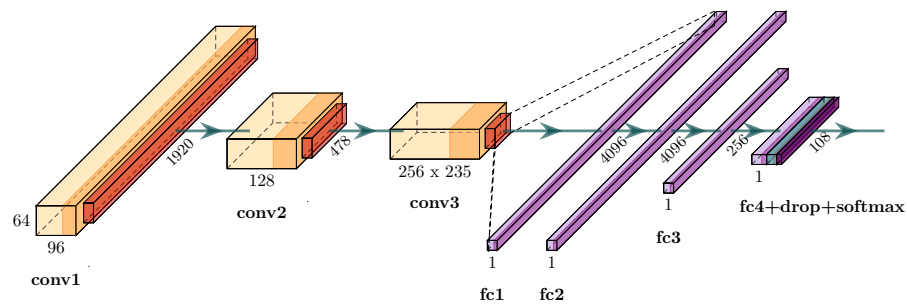
312   However, for the verification mode, one needs a feature vector as the output of a network instead of a
313 probability vector of classes. To address this requirement, layers responsible for handling the classification
314 (Soft-max, Dropout, and FC4. See in Figure 5) should be removed. The feature extraction uses the FC3
315 output which can now be used as a feature vector.

316   From this point, the trained model starts to operate as a deep feature descriptor. Once a 12-seconds
317 segment is provided to the network, the output is a feature vector of the same size as the FC3 output. For
318 architecture A this size corresponds to 256 dimensions, while for architecture B, 4096 dimensions.

319 **Biometric verification**

320 Different from the training data in which data augmentation techniques were applied, there is no overlap
321 for the testing data. That is, for the EC session, five segments of 12 seconds are extracted for each subject,
322 while in the motor movement/imaginary tasks, ten segments, as shown in Figure 4.

323   One of the most popular measures to compare methods for the verification mode in the literature is
324 the Equal Error Rate (EER). This measure is the point where the False Acceptance Rate (FAR) and the
325 False Rejection Rate (FRR) are equal in the Detection Error Trade-off (DET) curve. Comparing in an
326 all-against-all scheme, a distance score (such as euclidean distance) is computed for each pair of feature

A

B

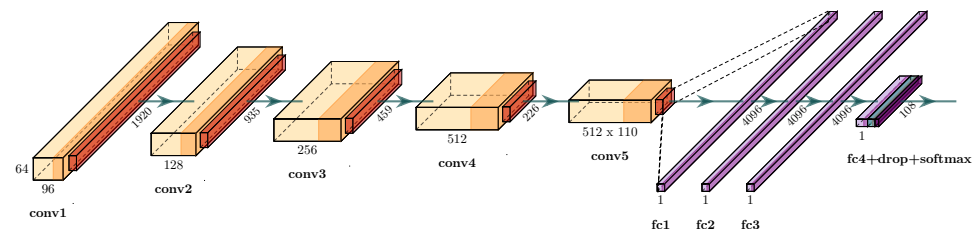



Figure 5. Architecture evaluated in this work. A) Architecture proposed in which *conv1*, *conv2* and *conv3* have filters size equal to 11x1, 9x1 and 9x1, respectively, and stride equal to 1. *Pool1* with stride equal to 4 and *pool2* and *pool3* with stride equal to 2 and filter size equal to 2 for all three. The padding is equal to zero for all convolutional and pooling layers. B) Architecture proposed in which *conv1*, *conv2*, *conv3*, *conv4* and *conv5* have filters size equal to 51x1, 17x1, 7x1, 7x1 and 7x1, respectively, and stride equal to 1 all five. *Pool1*, *pool2*, *pool3*, *pool4* and *pool5* are max pooling with filter size and stride equal to 2 for all pooling layers. The padding is equal to zero for all convolutional and pooling layers.

vectors. Ideally, one should have small scores between vectors from the same class (genuine) and large scores between vectors from different subjects (impostor).

In biometric verification mode, the number of impostor pairs is much higher than the genuine ones. A genuine pair occurs when two feature vectors are from the same individual. In contrast, an impostor pair corresponds to two feature vectors from two different individuals. As the number of impostor pairs is higher than the genuine ones, the biometric verification mode simulates a spoofing attack on a system. Thus, usually, the verification protocols is a challenging scenario.

Evaluation scenarios

The first scenario adopted in this work is the one proposed by Fraschini et al. (2015) referenced in this work as baseline task evaluation. The evaluation protocol related to first scenario ensures the acquisition of the training data on the open-eye task, while the closed-eye task composes the testing data from 109 individuals. It is worth to highlight that the acquisition of training data and testing data came from two different sessions.

Therefore, the first studied scenario disregards the impact of the motor/imaginary tasks for the biometric verification. Contrasting to that, Yang et al. (2018) carried out experiments to evaluate such conditions, i.e., the impact of motor activity on EGG signal for biometrics. Those experiments are the second scenario considered here and referenced as multi-task evaluation.

Yang et al. (2018) proposed three experimental protocols to evaluate the EEG biometric in the EEG Motor Movement/Imagery dataset (Goldberger et al., 2000; Schalk et al., 2004). The protocols are:

1. Protocol P1: aims to investigate the influence of some regions of the scalp for biometric purposes. Three regions are selected: the frontal lobe (F), with electrodes AF3, AFz, and AF4, the motor cortex (M), with the electrodes C1, Cz, and C2, and the occipital lobe (O), with O1, Oz, and O2

349 electrodes. For each training, the signals from R1 and R3 of one task were selected and R2 from
350 the same task was used for testing.

2. Protocol P2: explores the impact of training in the first and the third runs of the same mo-
tor/imaginary task and testing in the second run of all the remaining tasks and the baseline ones.
Using only a subset of nine electrodes, 24 experiments were evaluated.

3. Protocol P3: aims to explore the impact on the combination of different motor and imaginary tasks
for training. The authors reserve the second run of the first task (T1R2) to evaluate the performance
of the training tasks. They start training with T1R1 and T1R3 separately and finish with the fusion
of both with the other tasks and runs (all runs of T2, T3, T4, and T4).

358 The protocols P1 and P2 use the second run of all tasks as testing. In the protocol P3, only the task
359 T1R2 is used as testing. Furthermore, all protocols are evaluated from both biometric identification and
360 verification scenarios.

361 The use of several data acquisition sessions/runs may be unfeasible for real biometric systems once
362 it is onerous for the user, and, typically, the enroll session occurs only on the first time the user gets in
363 touch with the system. Considering that, a third evaluation scenario is proposed in which a new protocol
364 to validate the proposed approach is evaluated, using one task from a single run (session) for training
365 referenced as Cross-task evaluation. Inspired by the third protocol (P3) proposed by Yang et al. (2018),
366 the T1R2 is fixed as the testing data, and all the remaining tasks and runs are used for training. It is worth
367 stressing that different from Yang et al. (2018), the remaining tasks and their respective runs are used
368 separately for each training setup, instead of combining them. In that sense, train and testing in different
369 task types should be a more challenging scenario when compared with a scenario that uses the same task
370 in both training and test sets.

371 In all the three presented scenarios, samples from the same individual were present in both the training
372 and testing sets. It is well known that such evaluation scheme tend to favor the classifiers. To avoid this
373 issue and to assess the robustness of the proposed approach, a fourth protocol is proposed, forcing the
374 division between training sets and test oriented to individuals. That is, the individual used to train the
375 model is not used to test/evaluate the approach. This scenario was called cross-individual evaluation.
376 Therefore, the Deep Learning models are trained with signals from the first 55 individuals from the
377 Physionet database and tested with the last 54.

378 Due to the large number of possible combinations involving the 12 imaginary/motor tasks and the
379 objective of this study, only the Eyes Open and Closed conditions were used to evaluate this fourth
380 scenario.

381 Since the cross-individual evaluation is a more complex scenario and could benefit from a more robust
382 model, the impact of using a state-of-art technique called Squeeze-and-Excitation Network (Hu et al.,
383 2018) is evaluated. Section details the Squeeze-and-Excitation Networks.

Squeeze-and-Excitation Network

385 The Squeeze-and-Excitation (SE) block is an architecture unit first introduced by Hu et al. (2018) in order
386 to capture the dependency between the convolution channels of a network. It has two main steps: Squeeze
387 and Excitation.

388 The Squeeze step aggregates the global information of all the input channels and shrinking the spatial
389 dimensions in a feature map space U by using the global average pooling. The Excitation step transforms
390 each channel of the feature map U with weights obtained with the output of the previous step modulated
391 for each channel.

392 One can easily obtain an SE network by stacking several SE blocks, replacing some components, or
393 integrating SE blocks in some already known architecture, which is one of the great advantages of this
394 method. In addition, in Hu et al. (2018) the authors revealed state-of-the-art performance across multiple
395 datasets and applications using models based on SE blocks.

396 Toward this direction, it is investigated the use of these units in the best model obtained in the other
397 three scenarios evaluated. In order to determine the best SE network, the first step was to reproducing one
398 experiment of the baseline task evaluation scenario (eyes open for training and eyes closed for testing)
399 and trained a series of SE models using all 109 individuals for training. The performance of five models
400 obtained by merging SE blocks with the best architecture are investigated as follows:

401 • SE Model 1: all convolution layers were replaced by SE blocks;

402 • SE Model 2: add a single SE block after the input layer;

403 • SE Model 3: add a single SE block before the fully connected layers, that is, after all convolution
404   and pooling layers;

405 • SE Model 4: add a SE block after the input layer and before fully connected layers;

406 • SE Model 5: add a SE block after each convolution layer.

407   It is also evaluated the most promising reduction factor (hyper-parameter of SE blocks) and analyzed
408 six different values for all models.

RESULTS AND DISCUSSION

410 This section presents and discusses the protocols, the experimental setup and the obtained results.

Experimental setup

412 All the CNN operations are conducted using the TensorFlow library, along with the high level API Keras.
413 The specifications of the experimental computational environment are 64GB of DDR4 RAM, an Intel (R)
414 Core i7-5820K CPU 3.30GHz 12-core, and a GeForce GTX TITAN X GPU. The source is available on
415 Github.[2]

416   Two CNN architectures are evaluated with three band-pass filters for the baseline task: eyes open (EO)
417 for training/validation and eyes closed (EC) for testing. First, the performance of the CNN architectures
418 are evaluated on this simple task evaluation protocol by training for 20 epochs. Then, the effect of using
419 different strides on the CNN performance is investigated which also tends to change the data volume
420 created by the data augmentation technique and the time to train the CNN model. Thus, the best stride
421 parameter is used for the remaining experiments.

422   Model weights are randomly initialized, and the optimization method used here is the Stochastic
423 Gradient Descent with momentum coefficient of 0.9. A 10% dropout sub-sampling operation is placed
424 before the softmax layer in order to minimize the over-fitting. Using a learning rate sequence of
425 $lr = [0.01^2, 0.001^{18}]$, in which the number superscript represents the amount of epochs using that learning
426 rate. For the present experimentation, there is no improvement when training for more than 20 epochs.
427 After the training and the removal of the last layers of the CNN, the model resembles one embedding
428 model or a deep feature descriptor for a 12 seconds length EEG input signal.

429   The data augmentation technique applied to the training data created a total of 384 samples for each
430 individual from the EO data. For the motor (or imaginary) tasks (T1-T4), it was extracted from each
431 individual a total of 889 or 905 samples (signals with 123 seconds and 125 respectively). In the verification
432 mode, the data augmentation technique produces a total of 1,086 genuine pairs (intra-class) and 146,610
433 impostor pairs (inter-class) for the evaluation on the baseline tasks (EO-EC). A total of 4,809 intra-class
434 pairs and 571,392 inter-class pairs were generated for the multi-task (Motor-Imaginary) evaluation. In
435 that sense, the multi-task evaluation is a more complex and computationally expensive than the simple
436 tasks evaluation context.

Baseline tasks evaluation

438 The results regarding the combination of CNN architectures and the frequencies-bands are presented in
439 Table 2.

440   Architecture B, represented by Figure 5 (b), reached an EER of 0.65% in 30-50 frequency band.
441 Architecture A, depicted in Figure 5 (a), achieved the best overall figures (0.19% EER) for the gamma
442 frequency (30-50 Hz). The results in terms of decidability are also reported. The decidability is a measure
443 of how separable two distributions are (Ratha et al., 2001). In this case, it indicates how far the impostor
444 scores are from the genuine scores. As one may see in Table 2, architecture A performs better for all
445 frequency bands, specially for the gamma frequency.

446   Both architectures A and B perform well for the 30-50 Hz band and worse for 01-50 Hz, but
447 architecture A outperforms architecture B. One of the possible reasons is the filter sizes, which are smaller
448 in the first one and could lead to more details by extracting local features.

---

[2]https://github.com/ufopcsilab/EEG-Multitask

Table 2. EER and decidability reported of the two proposed architectures. EER presented in percentage.

Frequency Band	EER (%)	Decidability (%)	Architecture
10-30	5.06	3.22	
30-50	0.19	7.02	A
01-50	9.73	2.50	
10-30	6.85	2.84	
30-50	0.65	3.61	B
01-50	9.64	2.20	

In Table 3 is presented the EER from the proposed approach against the work proposed by Fraschini et al. (2015). As one can see, the proposed method with the use of convolution neural networks significantly reduced the EER.

Table 3. EER reported on EO-EC. EER presented in percentage. [#]Different evaluation protocol.

Reports	Approach	EER(%)
Fraschini et al. (2015)	Eigenvector Centrality	4.40
[#]Sun et al. (2019)	CNN + LSTM	0.41
Proposed Method	CNN	0.19

The experiments are under the same scenario: all electrodes (64), all individuals (109), training in EO (eyes open), and testing in EC (eyes closed). Furthermore, as the test protocol used is the same, a comparison becomes valid between both methods.

Despite the outstanding results reported by Sun et al. (2019), the evaluated scenario is different from the one proposed in (Fraschini et al., 2015) and used here.

Figure 6 shows the performance of the architecture A employing the DET curve over all three frequency bands evaluated. The curve related to the spectrum of 30-50 Hz surpasses all other experiments described in the literature. Those results are in line with the ones presented in (Fraschini et al., 2015) in which the 30-50 Hz range frequency overcomes the other frequencies for EEG biometrics-based. The reported results in this work differ from those reported by Fraschini et al. (2015) in the sense that there is a discrepancy in other frequency ranges, which are more significant here.

Stride evaluation

To evaluate the impact of different strides in data augmentation, the first step started using a stride of 20, which represents 125 milliseconds (ms) step in the time domain, to generate the samples aiming the baseline task (EO). Then, the size of the step was successively increased by 20 up to 200 and trained Architecture A for each one. In that sense, the impact of overlap from 125 ms to 1250 ms is analyzed. Baseline task EC, without overlapping, was reserved for evaluation (test) in all cases. Table 4 shows the EER for each stride.

Table 4. EER obtained for different strides. EER presented in percentage.

Stride	20	40	60	80	100	120	140	160	180	200
EER	0.09	0.09	0.37	0.76	0.65	0.65	0.74	0.83	0.37	1.11

One may see from Table 4 that the stride increase harms the proposed approach performance in the verification task. There is a trade-off on the stride size, the greater the stride, the smaller are the number of samples generated and the greater is the EER. Upon this fact, the smallest stride (20) resulted in 38,208

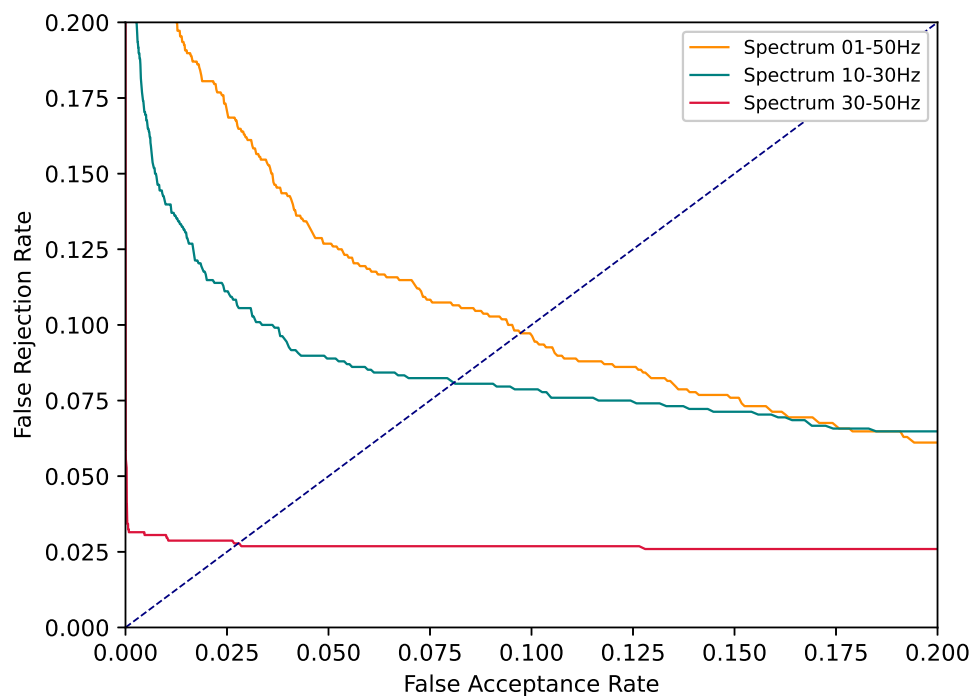



Figure 6. DET curves comparing all spectrum evaluated with Architecture A.

training samples while the larger one (200) generated only 3,885. The larger stride obtained the worst result, with an EER of 1,11%. However, the stride of 40 produces the same performance of stride 20 and with half the number of samples for training. Thus, the stride equal to 40 is the one chosen to be used in the remaining protocols.

Multi-task evaluation

The following experiments were based on the protocols proposed by Yang et al. (2018), which were presented in Section . It was used the best CNN architecture and stride obtained on the above experiments to perform the experiments on the verification scenario.

Protocol P1: region/task pairing

Figure 7 shows the location of the nine selected electrodes to perform this experiment.

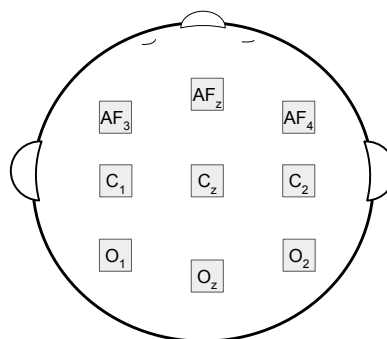



Figure 7. Selected electrodes on motor cortex, frontal and occipital lobule. Based on (Yang et al., 2018).

The proposed method is evaluated with signals from the three regions of the scalp (see Figure 7)

484 combined. The results have shown that the performance is better using all nine electrodes. The best result
485 is reported for T1R1 + T1R3 as training and T1R2 as test (EER of 0.12%). Among the experiments carried
486 with only three electrodes, the best result is obtained with task four (T4) and motor cortex electrodes
487 (EER of 0,85%). The worst results are related to the frontal lobe. A comparative analysis can be made
488 with the results obtained in (Yang et al., 2018), witch reached an EER of 7.83% when training and testing
489 with the same task T1 using only the data from the occipital lobe. The same experiment performed here
490 achieved 1.07% EER, showing that the proposed approach is robust even with a reduction in the number
491 of channels.

Protocol P2: Mismatch training/testing test

493 After applying the methodology described in section with protocol P2, results showed no improvement
494 when the train and the test come from the same task. The lowest EER, 0.08%, came from training
495 with T2R1+T2R3 and testing with the baseline task EO. Table 5 synthesizes the EER obtained for each
496 experiment.

Table 5. EER reported of mismatch training/testing. EER presented in percentage.

Train \ Test	T1R2	T2R2	T3R2	T4R2	EO	EC
T1R1+T1R3	0.12	0.29	0.42	0.42	0.10	0.37
T2R1+T2R3	0.19	0.29	0.56	0.69	0.08	0.56
T3R1+T3R3	0.21	0.19	0.29	0.19	0.18	0.36
T4R1+T4R3	0.12	0,13	0.27	0.27	0.20	0.36

497   The proposed approach outperformed in all scenarios the best figures reported by Yang et al. (2018)
498 when training with T1 and testing with T2.

499   It is worth emphasizing that the physical (non-imaginary) tasks T3 and T4 are the most difficult for
500 the biometric problem. Even when training and testing with tasks of the same nature (See Table 5), the
501 results are worse. Our hypothesis is that in these cases, noise from muscle movement can interfere with
502 signal acquisition.

Protocol P3: heterogeneous training

504 In this protocol, the samples of the individuals are organized in mini-batches of size 100 and shuffled. All
505 the nine electrodes from Figure 7 are used and T1R2 fixed as test set for all experiments. Table 6 shows
506 the tasks used for training and the EER obtained for each one.

507   As one may see on Table 6, stacking up different tasks hinder the proposed approach verification
508 performance. The worst-case occurred in P3.11, which has a larger training set (EER of 1.91%). The best
509 scenario occurs when the system is trained and evaluated with one run from the same task (EER of 0,1%).
510 In some cases, accumulate tasks do not significantly affect performance.

Comparative analysis

512 Protocol 1 led to the same conclusion then Yang et al. (2018): one could not determinate the most relevant
513 electrode placement among the different regions of the scalp, but using all of them (all nine electrodes)
514 enhances performance.

515   Protocol 2, both in this work and presented in (Yang et al., 2018), showed that using different tasks to
516 train and test does not affect performance. However, the proposed approach does not improve with the
517 accumulation of tasks for training, different from reported by Yang et al. (2018). Moreover, a lower EER
518 for all scenarios was reported, including Protocol 3.

519   One may see that the proposed CNN method has a significant improvement over the SOTA method
520 proposed by Yang et al. (2018), considering that, the experiments were conducted under the same scenario
521 with nine electrodes (See Table 7).

Cross-task evaluation

523 To evaluate the cross-task scenario, a new protocol based in (Yang et al., 2018) is proposed. Among the
524 12 sessions, it is selected the T1R2 for testing. Besides, the remaining tasks (and runs) are used as training

Table 6. EER obtained by accumulating tasks/runs. EER presented in percentage.

Protocol	Train	Test	EER (%)
P3.1	T1R1	T1R2	0.10
P3.2	T1R3	T1R2	0.44
P3.3	T1R1, T1R3	T1R2	0.22
P3.4	T2R1, T2R2, T2R3	T1R2	0.25
P3.5	T1R1, T1R3, T2R1	T1R2	1.25
P3.6	T1R1, T2R1, T2R2, T2R3	T1R2	1.14
P3.7	T1R1, T1R3, T2R1, T2R2	T1R2	1.29
P3.8	T1R1, T1R3, T2R1, T2R2, T2R3	T1R2	1.27
P3.9	T1R1, T1R3, T2R1, T2R2, T2R3, T3R1	T1R2	1.58
P3.10	T1R1, T1R3, T2R1, T2R2, T2R3, T3R1, T4R1	T1R2	1.47
P3.11	T1R1, T1R3, T2R1, T2R2, T2R3, T3R1, T4R1, T4R2	T1R2	1.91

Table 7. EER reported on T1R2. Both metrics are presented in percentages.

Reports	Train	Test	EER (%)
Yang et al. (2018)	T1 & T2	T1R2	2.63
Proposed Approach	T1 & T2	T1R2	0.27

and validation. In that manner, eleven models to evaluate the T1R2 set are created. It is conducted the training following the methodology described in Section . Also, the impact of using only nine electrodes (channels) instead of the 64 electrodes available on the Physionet Dataset is analyzed.

From the results presented in Table 8, one may see that CNN has converged for all training data, and the EER is lower than 0.39% for all of them. With 64 channels, the best scenario reaches 0,02% for T3R3 and an error, on average, of 0,16%. With nine channels, the lower EER achieved was 0.06% for T1R3 and the average error was 0,15%.

From these results, five questions arises:

What is the impact of using different tasks from the same run in biometric verification? Despite the parity of reported EER, the model trained with task T2 (same task as in T1, but in the imaginary state) presented a small and similar EER with both nine channels and 64, outperforming the models trained with tasks T3-T4. Despite task T3 showing the smallest EER with nine channels, it increased significantly with 64. In that sense, tasks from the same nature, even if it is imaginary, perform better than other ones. Although, the proposed CNN still performs well for the different tasks.

What is the impact of using the same task but different runs in biometric verification? From all experiments with 64 channels, the ones that use the same tasks from different runs had a good performance and a small variation. Upon this fact, one may affirm that EEG biometry benefits of data acquisition in which the same task is performed. Using nine electrodes, the results presented high variance, which may happen due to the reduction in the number of channels used to train the model.

What is the impact of using the same task but imaginary from different runs in biometric verification? Different from the experiments using the T1 task, the experiments engaging the T2 task have a noticeable variation both with nine and 64 electrodes. Despite this fact, imaginary tasks perform similar to the ones which are not imaginary but with more variance. Toward this direction, EEG biometry benefits from data acquisition in which the task performed are the same in training and testing, even if it is imaginary or not.

Table 8. EER reported on T1R2. EER presented in percentage.

Test	Train	EER (%) 9 channels	EER (%) 64 channels
	T1R1	0.21	0.17
	T2R1	0.25	0.15
	T3R1	0.10	0.19
	T4R1	0.09	0.19
	T2R2	0.15	0.17
T1R2	T3R2	0.12	0.25
	T4R2	0.39	0.26
	T1R3	0.06	0.17
	T2R3	0.12	0.04
	T3R3	0.08	0.02
	T4R3	0.27	0.06

What is the impact of using different tasks (imaginary or not) in biometric verification? Different tasks, motor or imaginary, for training and testing did not affect the performance of the system substantially. The biggest errors came from training with T4, a task completely different from the one used for test. On the other hand, training with a different motor task (T3) presented the same behavior from training with the same imagine task, aleatory, but satisfactory.

In that manner, EEG biometric may benefit from using tasks of the same nature, both motor tasks. In the same manner, it benefits from using the same task (imaginary or not). One possible hypothesis is that the activated areas from the brain are different from each task and the respective nature. The electrodes capture this behavior that directly influences the EEG curve. However, those changes are not substantial enough to affect very much the biometric verification process.

What is the impact of the number of electrodes used in biometric verification? Using fewer electrodes for training and testing did not affect considerably the performance of the system. Among all the performed experiments, the best result, 0.02% EER for training with T3R3, occurred with 64 channels, and the worst occurred with nine channels and an EER of 0.39%. This could indicate that EEG biometric benefits from using more channels, but the training accomplished with T3R1, T4R1, T2R2, T3R3, and T1R3, almost half of them, showed better performance with only nine.

One hypothesis to justify this result is that the use of 64 electrodes has redundancies and that using nine is enough to capture the most important signals emitted during the performance of a task, motor, or imaginary. Even in cases where the EER was higher using fewer channels, this increase was not significant. In that sense, one may conclude that using 64 channels adds unnecessary financial and computational cost since nine channels proved to be enough to achieve good results.

Cross-individual evaluation

Since the Cross-individual evaluation is more challenging, it has more room for improvement. Thus, in this scenario, we evaluate the SE Blocks. We also compared networks based on SE blocks against networks based on simple CNN blocks under the same conditions.

In total, 30 models were evaluated on training data by changing the hyper-parameter r six times for each of the five SE models proposed. The r given by $r = 2^x$ ranged from 1 to 32 ($x = 0, 1, 2, 3, 4, 5$). The r hyper-parameter is used to enhance the information of the channels **in the excitation operation** and it is related to the division of the original number of channels. If $r = 64$, the 64 original signals were divided 64 and the result will be only one channel in middle, which could mix up the original channels information.

The best results were achieved for SE Model 2 with $r = 2$, reaching an EER of 0.18% and even better

for SE Model 3 with $r = 32$, with an EER of 0.11%, both greater than the results reported without the SE blocks. These two models are then used to evaluate the proposed Cross-individual scenario.

To refer to the first half of the dataset (55 first individuals), EO_1 and EC_1 are used. EO_1 is used to refer to the first half of the individuals present on the eyes open task and EC_1, for the first half of the eyes close task. EO_2 and EC_2 are the second half of the eyes open and eyes closed task respectively.

Table 9 shows the obtained results. One can observe that all reported results are closer to the reported in the baseline task evaluation (EER = 0,19%). This endorses the robustness of the proposed approach. SE Model 2 reached the best EER, 0.18% when testing with a different set of individuals under the same condition (eyes open). When trained with EC_1 the best result occurred for test with $E0_1$ (EER = 0.36%).

Table 9. Results obtained with SE Model 2 with r=2 (SE2r2) and SE Model 3 with r=32 (SE3r32) both trained with signs of half of the individuals.

Train	Test	CNN Arch. A	SE2r2 EER(%)	SE3r32 EER(%)
EO_1	EO_2	0.55	0.18	0.92
	EC_2	0.93	0.51	0.39
EC_1	EO_2	0.40	0.41	0.55
	EC_2	0.97	0.36	1.11

SE Model 3 showed lower performance than SE Model 2 for most experiments. Thus, we believe SE-block based network is a promising approach as a feature extractor for EEG signals.

CONCLUSION

In the present work, it was evaluated the use of deep descriptors to extract features from the EEG signal for biometric purposes. The main focus was on one particular aspect: what is the impact of a motor (or even imaginary) activity in the EEG biometric modality? To investigate this question, the proposed approach with the Physionet EEG Motor Movement/Imagery Dataset was evaluated.

The proposed deep learning-based method achieved outstanding figures, overcoming the state-of-the-art methods even for the multi-task scenario. It is worth emphasizing that the proposed method is capable of performing well, even when trained on different tasks. For instance, it reaches an EER of 0.12% when trained with the task T4, corresponding to imagining moving the feet, and tested with the task T1, equivalent to performing the motor movement of closing the fists. One can conclude that the electrodes capture the motor (or imaginary) interference, however, those changes are not substantial enough to inhibit the biometric verification process. This work also evaluated the impact of the Squeeze-and-Excitation blocks for the EEG biometric problem. The Squeeze-and-Excitation blocks explore the interdependence between the channels of the signal, allowing better quality feature extraction and, therefore, better results.

Due to the nature of the proposed CNN-based approach, it allows simultaneous processing of all EEG channels, which is an advantage from a computational point of view. However, the use of multiple channels could turn the biometric modality impractical for real-world applications. Thus, the impact of using fewer channels (electrodes) is investigated, and a possible conclusion is that the use of all channels does not necessarily improve the result. The reported results showed that only 9 electrodes allowed to achieve competitive results, or even better, compared to the use of 64 electrodes in similar tasks.

ACKNOWLEDGMENTS

We appreciate the reviewers and the editor for their detailed examination of our manuscript, which helped us to improve important aspects of the presentation.

REFERENCES

Alyasseri, Z. A. A., Khader, A. T., Al-Betar, M. A., and Alomari, O. A. (2020). Person identification using eeg channel selection with hybrid flower pollination algorithm. *Pattern Recognition*, page 107393.

Boubakeur, M. R., Wang, G., Zhang, C., and Liu, K. (2017). Eeg-based person recognition analysis and criticism. In *2017 IEEE International Conference on Big Knowledge (ICBK)*, pages 155–160. IEEE.

Carrión-Ojeda, D., Fonseca-Delgado, R., and Pineda, I. (2020). Analysis of factors that influence the performance of biometric systems based on eeg signals. *Expert Systems with Applications*, page 113967.

Das, R., Maiorana, E., and Campisi, P. (2017). Visually evoked potential for eeg biometrics using convolutional neural network. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 951–955. IEEE.

Das, R., Maiorana, E., and Campisi, P. (2018). Motor imagery for eeg biometrics using convolutional neural network. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2062–2066. IEEE.

Del Pozo-Banos, M., Alonso, J. B., Ticay-Rivas, J. R., and Travieso, C. M. (2014). Electroencephalogram subject identification: A review. *Expert Systems with Applications*, 41(15):6537–6554.

Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis. *Neuroimage*, 34(4):1443–1449.

DelPozo-Banos, M., Travieso, C. M., Alonso, J. B., and John, A. (2018). Evidence of a task-independent neural signature in the spectral shape of the electroencephalogram. *International Journal of Neural Systems*, 28(01):1750035.

El-Fiqi, H., Wang, M., Salimi, N., Kasmarik, K., Barlow, M., and Abbass, H. (2018). Convolution neural networks for person identification and verification using steady state visual evoked potential. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1062–1069. IEEE.

Fraschini, M., Hillebrand, A., Demuru, M., Didaci, L., and Marcialis, G. L. (2015). An eeg-based biometric system using eigenvector centrality in resting state brain networks. *IEEE Signal Processing Letters*, 22(6):666–670.

Fraschini, M., Pani, S. M., Didaci, L., and Marcialis, G. L. (2019). Robustness of functional connectivity metrics for eeg-based personal identification over task-induced intra-class and inter-class variations. *Pattern Recognition Letters*, 125:49–54.

Garcia, G., Moreira, G., Menotti, D., and Luz, E. (2017). Inter-patient ecg heartbeat classification with temporal vcg optimized by pso. *Scientific Reports*, 7(1).

Ghandeharion, H. and Ahmadi-Noubari, H. (2009). Detection and removal of ocular artifacts using independent component analysis and wavelets. In *2009 4th International IEEE/EMBS Conference on Neural Engineering*, pages 653–656. IEEE.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220.

Gui, Q., Ruiz-Blondet, M. V., Laszlo, S., and Jin, Z. (2019). A survey on brain biometrics. *ACM Computing Surveys (CSUR)*, 51(6):112.

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141.

Islam, M. K., Rastegarnia, A., and Yang, Z. (2015). A wavelet-based artifact reduction from scalp eeg for epileptic seizure detection. *IEEE journal of biomedical and health informatics*, 20(5):1321–1332.

Jain, A. K., Ross, A., and Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1).

Kong, X., Kong, W., Fan, Q., Zhao, Q., and Cichocki, A. (2018). Task-independent eeg identification via low-rank matrix decomposition. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 412–419. IEEE.

Kostílek, M. and Šťastný, J. (2012). Eeg biometric identification: Repeatability and influence of movement-related eeg. In *2012 International Conference on Applied Electronics*, pages 147–150. IEEE.

Kumar, M. G., Saranya, M., Narayanan, S., Sur, M., and Murthy, H. A. (2019). Subspace techniques for task-independent eeg person identification. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4545–4548. IEEE.

Kumar, P. S., Arumuganathan, R., Sivakumar, K., and Vimal, C. (2009). An adaptive method to remove ocular artifacts from eeg signals using wavelet transform. *Journal of Applied Sciences Research*, 5(7):741–745.

Luz, E. J. S., Moreira, G. J. P., Oliveira, L. S., Schwartz, W. R., and Menotti, D. (2018). Learning deep

off-the-person heart biometrics representations. *IEEE Transactions on Information Forensics and Security*, 13(5):1258–1270.

Ma, L., Minett, J. W., Blu, T., and Wang, W. S. (2015). Resting state eeg-based biometrics for individual identification using convolutional neural networks. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2848–2851. IEEE.

Maiorana, E. and Campisi, P. (2018). Longitudinal evaluation of eeg-based biometric recognition. *IEEE Transactions on Information Forensics and Security*, 13(5):1123–1138.

Mao, Z., Yao, W. X., and Huang, Y. (2017). Eeg-based biometric identification with deep learning. In *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 609–612. IEEE.

Marcel, S. and Millán, J. d. R. (2007). Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 29(4):743–752.

Poulos, M., Rangoussi, M., Chrissikopoulos, V., and Evangelou, A. (1999). Person identification based on parametric processing of the eeg. In *ICECS'99. Proceedings of ICECS'99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No. 99EX357)*, volume 1, pages 283–286. IEEE.

Ratha, N. K., Senior, A., and Bolle, R. M. (2001). Automated biometrics. In *International Conference on Advances in Pattern Recognition*, pages 447–455. Springer.

Sawangjai, P., Hompoonsup, S., Leelaarporn, P., Kongwudhikunakorn, S., and Wilaiprasitporn, T. (2019). Consumer grade eeg measuring sensors as research tools: A review. *IEEE Sensors Journal*, 20(8):3996–4024.

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., and Wolpaw, J. R. (2004). Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043.

Schons, T., Moreira, G. J. P., Silva, P. H., Coelho, V. N., and Luz, E. J. S. (2017). Convolutional network for eeg-based biometric. In *Iberoamerican Congress on Pattern Recognition*, pages 601–608. Springer.

Silva, P. L., Luz, E., Moreira, G., Moraes, L., and Menotti, D. (2019). Chimericaldataset creation protocol based on doddington zoo: A biometric application with face, eye, and ecg. *Sensors*, 19(13):2968.

Singh, B., Mishra, S., and Tiwary, U. S. (2015). Eeg based biometric identification with reduced number of channels. In *2015 17th International Conference on Advanced Communication Technology (ICACT)*, pages 687–691. IEEE.

Stassen, H. (1980). Computerized recognition of persons by eeg spectral patterns. *Electroencephalography and Clinical Neurophysiology*, 49(1):190 – 194.

Sun, Y., Lo, F. P.-W., and Lo, B. (2019). Eeg-based user identification system using 1d-convolutional long short-term memory neural networks. *Expert Systems with Applications*.

Vinothkumar, D., Kumar, M. G., Kumar, A., Gupta, H., Saranya, M., Sur, M., and Murthy, H. A. (2018). Task-independent eeg based subject identification using auditory stimulus. In *Proc. Workshop on Speech, Music and Mind*, volume 2018, pages 26–30.

Wang, M., El-Fiqi, H., Hu, J., and Abbass, H. A. (2019). Convolutional neural networks using dynamic functional connectivity for eeg-based person identification in diverse human states. *IEEE Transactions on Information Forensics and Security*, 14(12):3259–3272.

Wilaiprasitporn, T., Ditthapron, A., Matchaparn, K., Tongbuasirilai, T., Banluesombatkul, N., and Chuangsuwanich, E. (2019). Affective eeg-based person identification using the deep learning approach. *IEEE Transactions on Cognitive and Developmental Systems*, 12(3):486–496.

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

Yang, S., Deravi, F., and Hoque, S. (2018). Task sensitivity in eeg biometric recognition. *Pattern Analysis and Applications*, 21(1):105–117.

Yong, X., Ward, R. K., and Birch, G. E. (2008). Robust common spatial patterns for eeg signal preprocessing. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2087–2090. IEEE.