Array 6 (2020) 100026

Contents lists available at ScienceDirect

Array

journal homepage: www.elsevier.com/journals/array/2590-0056/open-access-journal

Event related potential (ERP) as a reliable biometric indicator: A comparative approach

M. Sabeti [a,**], R. Boostani [b], E. Moradi [c,*]

[a] Department of Computer Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran
[b] CSE & IT Dept., Faculty of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran
[c] Department of Neurosurgery, Shahid Beheshti University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Keywords:
EEG
Subject-dependent features
Person identification
Person verification

ABSTRACT

There is growing tendency to elicit subject-dependent features from physiological data like electroencephalogram (EEG) signals, which is the most relative one to the subject's brain, though the behavior of this signal does not obey any grammatical pattern. In this regard, several attempts have been made toward eliciting subject-dependent features from EEGs to promisingly verify/identify subjects. In this paper, we have comprehensively assessed state-of-the-art EEG features to empirically show their pros and cons. Herein the resting-EEG and ERP-EEG of 20 healthy subjects are utilized and the conventional features are extracted from them. These features are then fed to three types of classifiers. Among the deployed features, spectral coherence and correlation features provide the best verification and identification results in terms of classification accuracy and equal error rate. Empirical results demonstrate that the ERP-EEG features are more discriminative than the resting-EEG features because they reveal the response of a specific circuit of the neural brain system.

1. Introduction

Biometric systems have been widely used in secure and financial organizations such as military, banks, airports, security agencies and famous companies. Biometric systems verify their queries according to their unique physiological data such as fingerprint [1, 2]], palm print [3], face, ear shape, iris textures, vocal tract system (speech signal) [1], fiducial points of electrocardiogram (ECG) [4], distribution of electroencephalogram (EEG) energy in different bands [5, 6]] and DNA structure [7]. The mentioned biometric systems suffer from various limitations and can be deceived by impostors. For instance, speech recognition system can be deceived by a recorded tape; iris biometric systems can be deceived by wearing a lens; an image or video tape is the bottleneck of face and ear-shape biometric systems. DNA verification system is the most secure biometric system but its verification process is a bit time consuming and expensive.

ECG and EEG biometric systems cannot be deceived since these stochastic signals should be directly acquired from the surface of the body. Among the mentioned biometric indicators, it is expected that decoding the EEG variations [8-13] provides exclusive features characterizing the brain activity of subjects. There are some evidences about differences among people's brain in terms of anatomical and functional traits [14, 15]]. Moreover, it has been proven that there are some genes, which have considerable effects over different brain areas and EEG frequencies [16-18]. Since EEG signals represent the reflection of ionic currents through exclusive neural connections of each subject, it carries specific information and imitating of this signal is impossible. It seems that EEG is the most appropriate technique [19, 20]] to monitor the brain activity because of high temporal resolution, portability and low price, though EEG suffers from some limitations such as volume conduction, common sources, potential leakage over the scalp and additive noise (e.g., muscle artifacts). As the research trend shows, EEG will be used as the most effective and popular brain monitor tool in the future [21, 22]]. Nevertheless, EEG as a verification biomarker has some bottlenecks such as noisy recording and also its content is influenced by stress and mood.

A few attempts have been made to extract exclusive brain signature for each subject by applying a suitable transform to his/her EEG signals and considering the elicited coefficients as his unique mental features. Fraschini et al. [23] investigated the possibility of person identification by measuring the brain functional connectivity. Nonetheless, this feature is sensitive to the phase difference between each pair of EEG channels. They estimate functional connectivity between electrodes using the

* Corresponding author.
** Corresponding author.
E-mail addresses: sabeti@shirazu.ac.ir (M. Sabeti), moradieh@sbmu.ac.ir (E. Moradi).

https://doi.org/10.1016/j.array.2020.100026
Received 13 June 2019; Received in revised form 30 March 2020; Accepted 12 April 2020
Available online 18 April 2020

phase lag index and then, generate the connectivity matrix. Afterward, the weighted network is estimated from the connectivity matrix and the nodal eigenvector centrality is computed. The highest recognition rate is obtained in the Gamma and Beta bands, while other frequency bands yield poor results. Their results show that using functional connectivity improves the performance compared to that of power spectrum features.

Rocca et al. [24] propose the spectral coherence-based connectivity between different brain regions which measures the extension of correlation function in the frequency domain. They execute this method over 108 subjects and obtain 97.5% and 96.3% verification accuracy during eye-closed and eye-open conditions, respectively. These convincing results are achieved in the Alpha and Delta bands. Gui et al. [25] evaluate two similarity features for EEG-based identification: Euclidean distance and dynamic time wrapping (DTW) methods. Their results on the channel Oz provide the best identification accuracy for both mentioned features. Using Euclidean distance yields 80% accuracy over 30 subjects, while the best accuracy of DTW yields 68% verification rate. Zhao et al. [26] implement a security system verifying the users (4 females and 6 males) directly from their EEG signals. After preprocessing stage to remove frequency interference and artifacts, they extracted features including coefficients of autoregressive (AR) model, power spectrum and center frequency (frequency which makes the power spectrum maximize) from EEG signals. Then, k-nearest neighbor (kNN) classifier is applied to classify the obtained features with the average accuracy of 97.63%. Huang et al. [27] extract equivalent root mean square values from scalp EEGs and apply them to a multi-layer perceptron (MLP) neural network. Their results over 45 subjects yield 100% verification rate while it provided 95.1% over 122 subjects.

It is noteworthy to say that the verification results over the occipital, temporal and parietal lobes are much higher than the other scalp parts in the eye-close condition [16], while the verification results are proper over the anterior part under the eye-open state [16]. Therefore, a few studies are conducted to verify the subjects by employing less number of EEG channels (e.g., only one on the occipital lobe). Riera et al. [28] propose an EEG-based authentication system by using just two dry electrodes (Fp1 and Fp2). They recorded EEG signals from the subjects within the interval of 34 ± 74 days. AR coefficients, power spectral density (PSD) and three statistical measures are extracted from their EEG signals. They have achieved the verification rate of 98.10% over 51 registered subjects and 36 intruders.

Another EEG based subject-dependent biomarker is the pattern of evoked related potential (ERP) which has been used for subject verification [29]. ERP shape reflects valuable information about high-level neural processing like attention and memory. Armstrong et al. [30] extract ERP template by synchronous averaging, and then subject-dependent features are elicited from the ERP of each subject. They finally apply their subjects' ERP features to different classifiers and achieve 82–97% verification rate. They have achieved the identification rate of 92% for cross-correlation, 89% for divergent auto-encoder (DIVA), 82% for Naive discriminant learning (NDL) and 83% by support vector machines (SVM), respectively. In addition, their findings imply the stability of the ERP features through six months recording.

Das et al. [31] extract spatio-temporal patterns from visual evoke potential (VEP). They show that the subject-dependent information of ERP component, over the occipital region, occurs from 120 to 200 ms after the visual stimulus. Palaniappan [32] have recorded EEG signals through a specific mental task from several subjects in order to verify them. In this regard, they have extracted AR coefficients, PSD differences over the inter-hemispheric channels, linear complexity and approximate entropy from the EEG signals. They propose a modified four fold cross validation procedure to demonstrate the robustness of this authentication system against impostors.

This study aims to investigate the performance of state-of-the-art signal processing techniques to characterize the EEG signals of 20 subjects, in both resting and cognitive state, for an EEG-based authentication system in both identification and verification modes. The candidate features in this study are local binary pattern (LBP) [33, 34]], AR coefficients, wavelet coefficients, eigen-vectors of principal component analysis (PCA), correlation value, spectral coherence and phase difference. The reminder of this paper is structured as follows. The description of the collected dataset is brought in Section 2 and the explanation of the extracting features are brought is Section 3. Classifiers are discussed in Section 4. Experimental results and discussions are illustrated in Section 5. The paper is finally concluded in Section 6 and a horizon for the future work is suggested.

2. Data collection

Twenty subjects (all male and healthy) have voluntarily enrolled in this study. The average of their age is 33.4 years old with standard deviation (Std) of 9.29. They have had no history of pain, neurologic and psychiatric disorder. During the EEG recording, each participant is seated on a comfortable chair with open eyes. To mitigate the effect of muscle artifact, the neck is firmly supported by the chair, and the feet are rested on a footstep. In this study, the signal recording process can be divided into two phases, where the resting-EEG and ERP-EEG signals are caught from the same subjects. In the first recording state, the participants do not perform any mental task and no stimulus is imposed. Each trial recording takes 2 min. In the second recording state, ERP-EEG signals are recorded through oddball stimulus paradigm (cognitive task), where audio stimuli are randomly applied to the subjects. The participants are asked to discriminate the infrequent target stimuli (high pitch beep) from the frequent standard stimuli (low pitch beep) by pressing a button. The cognitive ability of distinguishing between the low and high tones is confirmed before beginning the experiment. This cognitive experiment takes 20 min for each subject.

Electrophysiological data are recorded using a Neuro-Scan 24 channels Synamps system, with a signal gain equal to 75 K (150- at the headbox). Through the recording paradigm, EEG signals are caught from 20 electrodes (Fpz, Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, and O2) via an Electrocap molded according to the 10–20 standard system with reference to linked earlobes plus vertical electrooculogram (VEOG) and horizontal electrooculogram (HEOG). The sampling rate of the EEG setup is 200 Sa/s. The blink artifacts are corrected using an efficient technique, which is described in [35, 36]]. In the first preprocessing phase, the artifacts are eliminated by an experienced physician using visual inspection. In addition, the EEG signals are filtered by a band pass filter with cutoff frequencies of 0.5 and 45 Hz.

3. Methodology

In this Section, state-of-the-art features for the EEG-based authentication systems are briefly explained.

3.1. Local binary pattern (morphological feature)

Local binary pattern (LBP) [33, 34]] is an efficient descriptive method for texture analysis. For each sample in the signal, this feature is computed by comparing the current sample value to its left and right side neighbors and therefore, it measures the local variation for each sample. LBP feature is computed as:

$$LBP_{LHS}(x[n]) = \sum_{k=0}^{L-1} S(x[n+k-L] - x[n])2^{2L-1-k} \tag{1}$$

$$LBP_{RHS}(x[n]) = \sum_{k=L}^{2L-1} S(x[n+k+1-L] - x[n])2^{2L-1-k} \tag{2}$$

$$LBP(x[n]) = LBP_{LHS}(x[n]) + LBP_{RHS}(x[n]) \tag{3}$$

where $x[n]$ represents the nth sample of signal x, $L$is the number of samples considered on either side of the current sample. Here L is set to 8 and the function $S$is defined as follows:

$$S(p) = \begin{cases} 1 if p \geq 0 \\ 0 otherwise \end{cases} \tag{4}$$

3.2. Spectral coherence

Spectral coherence [24] quantifies the level of synchrony between two stationary signals at a specific frequency f. This feature for two EEG channels (i and j) is estimated by:

$$coh_{i,j}(f) = \frac{|S_{i,j}(f)|^2}{S_{j,j}(f).S_{i,i}(f)} \tag{5}$$

where $S_{i,j}(f)$ is the cross spectrum between the ith and jth channels, $S_{i,i}(f)$ and $S_{j,j}(f)$ are the respective auto-spectra. $coh_{i,j}(f) = 0$ is occurred when no synchrony exists at the frequency f, and $coh_{i,j}(f) = 1$ shows the maximum synchrony is happened at frequency f.

3.3. Phase difference

The phase lag index (PLI) [23] is estimated using pair-wise statistical interdependence between EEG time series in two different channels. In contrast to spectral coherence, PLI reveals the features related to volume conduction, common sources and active reference; therefore, it presents a more suitable estimation of functional interactions among the brain regions. The PLI varies between 0 (interaction with zero-phase lag) and 1 (maximum interaction), and it evaluates the asymmetry of the distribution of instantaneous phase differences between pairs of channels. The PLI is determined as:

$$PLI = |\langle sign(\sin(\Delta\phi(t_k)))\rangle| \tag{6}$$

where $\Delta\phi$ is the difference between instantaneous phases for two time series in the interval $[-\pi, \pi]$, t_k are discrete steps and $\langle . \rangle$denotes the average over the time. Finally, a functional connectivity matrix containing the PLI values between each pair of electrodes is obtained.

3.4. Wavelet coefficients

Wavelet transform [37] has been repeatedly used for analyzing EEG signals and representing its content in both time and frequency domains. In fact, wavelet transform is a bank of filters, which their bandwidths are set in a dyadic manner [38]. Wavelet coefficients *c(j,k)* of the mother wavelet $\Psi_{j,k}(n)$ (jth scale and shift k), is determined by the following relation:

$$c(j,k) = \sum_{n \in z} f(n)\Psi\left(2^j(n-k)\right) \tag{7}$$

where *f(n)* is the input discrete signal. After decomposing each EEG channel into m scales, five features are extracted from each decomposed signal within a time frame. These features are the energy, maximum, minimum, mean, and standard deviation of the decomposed signals. In this research, the number of extracted features from 20 EEG channels for the 5 scale became $5 \times 5 \times 20 = 500$.

3.5. Principal component analysis (PCA) features

PCA [39, 40]] is an orthogonal linear transform which projects a high dimensional input to a lower dimensional subset of features, where the new features are orthogonal and independent to each other. This feature reduction is carried out by multiplying the input vectors to the eigen-vectors, corresponding to the larger eigen-values. Given a threshold determines the informative eigen-vectors in which data is scattered along those directions (just features with a high variance are selected). First, the employed EEG signals are segmented into successive 1s windows. Next, the covariance matrix of EEG signals is estimated for all channels and 20 eigen-vectors are extracted. For each time-frame over 20 channels, $20 \times 20 = 400$ features are generated.

3.6. Correlation-based features

The correlation [41] between different EEG channels is determined by the following function:

$$R_x(\tau_t) = E\{x(t+\tau_t)x^T(t)\} \tag{8}$$

where τ_t is the time delay between a signal and its shifted version. The cross-correlation value is determined between two by two EEG channels and therefore, $20 \times 20 = 400$ features are produced for each time frame. In fact, these features contain spatio-temporal information of the 20 EEG channels. Fig. 1 depicts stages of the implemented correlation-based feature.

1. Estimate time delayed covariance matrix between two consecutive EEG frames as

$$R_x^i(\tau_t) = E\{x(t+\tau_t)x^T(t)\}$$

where the covariance matrix $R_x^i(\tau_t)$ includes spatio-temporal information of EEG signals in each time frame.

2. Takes an average over the time-lag covariance matrices of the two by two channels:

$$\tilde{R}_x(\tau_t) = \frac{\sum_{i=1}^{T-1} R_x^i(\tau_t)}{T-1}$$

where T shows the number of trials.

3. Arrange the elements of final matrix $\tilde{R}_x(\tau_t)$ into a vector.

3.7. AR coefficients

One of the most powerful tools for signal modeling is the autoregressive (AR) model [42], in which a sample is predicted according to the weighted average of its p former samples, where p determines the model order. A stationary signal *x[n]* is modeled by the following relation:

$$x[n] = \sum_{i=1}^{p} \hat{a}_i x[n-i] \tag{9}$$

where $\hat{a}_i$ denotes the AR model coefficients. In this paper, the Burg method is employed to estimate the AR coefficients based on the summation of both forward and backward prediction errors. In addition, the finite sample criterion [42] is used to select the best order of the AR model, considering the residual variance and the prediction error. In this study, the best order of AR model is set to 8; therefore, the number of elicited features for 20 channels becomes 160.

4. Classifiers

In this part, we introduce the k-nearest neighbor (kNN), support vector machine (SVM) and random forest. Although kNN [43] is not a new method, it is still used in wide range of applications, especially for those multi-class problem that the distribution of their classes is multimodal. kNN is famous due to its simplicity, interpretability and good performance. kNN is a lazy classifier in the recall phase while it is the only classifier that does not need any training phase. kNN is a local classifier and for each test sample x_0, its k nearest neighbors should be

1. Estimate time delayed covariance matrix between two consecutive EEG frames as

$$R_x^i(\tau_t) = E\{x(t+\tau_t)x^T(t)\}$$

where the covariance matrix $R_x^i(\tau_t)$ includes spatio-temporal information of EEG signals in each time frame.

2. Takes an average over the time-lag covariance matrices of the two by two channels:

$$\widetilde{R}_x(\tau_t) = \frac{\sum_{i=1}^{T-1} R_x^i(\tau_t)}{T-1}$$

where T shows the number of trials.

3. Arrange the elements of final matrix $\widetilde{R}_x(\tau_t)$ into a vector.

Fig. 1. The correlation-based feature.

found and the label of samples, with majority vote, is assigned to x_0. Through the cross validation phase, the value of k is set to 5.

Support vector machine (SVM) [44] is designed based on the statistical learning theory. SVM tries to maximize the margin width simultaneous with minimizing the empirical error in order to provide a good generalization property. Using kernel function maps the input samples to a new high dimensional space that probability of separating the projected samples by a hyper-plane is increased.

Random forest [45] classifier is operated by constructing an ensemble of decision trees in the training phase and predicts the label, which is the mode of the classes of the individual trees. Decision trees work by choosing the finest feature to divide the data and expanding the leaf nodes of the tree until the termination criterion is met. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. After the construction of the decision tree, a tree-pruning phase is applied to diminish the size of the decision tree. Pruning is supported by trimming the branches of the initial tree to simplify the interpretation capability of the decision tree.

5. Experimental results and discussion

In this section, the verification and identification results are separately illustrated in terms of classification accuracy and equal error rate (EER). The results are demonstrated for three authentication scenarios in terms of the resting-EEG, ERP-EEG and averaged ERP. In the identification system, we are confronted with a multi-class problem, where the outputs of each classifier could be 1 to N (where N is the number of participants). On the other hand, in verification system, we are faced with a two-class problem, where each EEG signal could be classified either as authorized user or impostor.

In the first stage, both resting-EEG and ERP-EEG signals are divided into successive 1 s windows (except for wavelet transform that the window length is set to 4 s). The successive windows have 50% overlap. The mentioned features are extracted from each windowed signal, then these features are applied to kNN, SVM and random forest classifiers. Results in Tables 1–3 demonstrate the identification and verification accuracy (in terms of mean and Std) over twenty healthy subjects. In this study, SVM uses linear kernel function, and sequential minimal optimization (SMO) for objective-function minimization. In addition, for optimizing the hyper-parameters of each SVM, an interval is considered for the parameter C and SVM is run for different values of C (in the train set) and the best value of C is selected. Since the SVM is binary classifier (only two classes), in identification task, K(K – 1)/2 binary SVM models using the one-versus-one is applied where K is the number of class labels.

As we see, the results of ERP-EEG using different features are significantly superior to that of the resting-EEG features. This is therefore ERP-EEG features reveals the high-level activity of neural systems (e.g., attention, memory and cognition) rather using the resting-EEG which reveals the integration of the whole activities of the brain system. Tables 1–3 show that both spectral coherence and correlation are the most accurate methods for both EEG-based identification and verification

Table 1
The mean and Std of accuracy of kNN classifier for resting-EEG and ERP-EEG signals.

	Identification		Verification	
Method	ERP-EEG	Resting-EEG	ERP-EEG	Resting-EEG
LBP	84.88 ± 0.00	63.48 ± 0.00	98.57 ± 1.65	97.05 ± 1.90
Spectral coherence	98.64 ± 0.00	60.67 ± 0.00	99.87 ± 0.13	96.74 ± 1.85
Phase difference	56.93 ± 0.00	35.39 ± 0.00	96.14 ± 289	94.86 ± 4.09
Wavelet coefficients	99.26 ± 0.00	71.35 ± 0.00	99.93 ± 0.00	97.44 ± 2.13
Correlation-based	99.06 ± 0.00	65.73 ± 0.00	99.91 ± 0.12	97.22 ± 2.03
PCA	91.26 ± 0.00	38.20 ± 0.00	99.13 ± 0.95	95.90 ± 1.18
AR	79.82 ± 0.00	76.40 ± 0.00	98.19 ± 1.14	98.26 ± 1.25

Table 2
The mean and Std of accuracy of SVM classifier for resting-EEG and ERP-EEG signals.

	Identification		Verification	
Method	ERP-EEG	Resting-EEG	ERP-EEG	Resting-EEG
LBP	91.18 ± 0.00	69.66 ± 0.00	98.58 ± 1.49	96.88 ± 2.33
Spectral coherence	99.06 ± 0.00	57.87 ± 0.00	99.56 ± 1.41	95.45 ± 2.11
Phase difference	89.22 ± 0.00	48.88 ± 0.00	97.48 ± 2.37	93.60 ± 6.52
Wavelet coefficients	23.76 ± 0.00	38.20 ± 0.00	95.00 ± 0.29	95.00 ± 0.25
Correlation-based	99.21 ± 0.00	67.42 ± 0.00	99.93 ± 0.11	96.85 ± 2.39
PCA	89.31 ± 0.00	39.89 ± 0.00	98.97 ± 0.77	95.65 ± 1.80
AR	94.25 ± 0.00	77.53 ± 0.00	99.38 ± 1.26	97.56 ± 1.53

Table 3
The mean and Std of accuracy of random forest classifier for resting-EEG and ERP-EEG signals.

Method	Identification		Verification	
	ERP-EEG	Resting-EEG	ERP-EEG	Resting-EEG
LBP	87.76 ± 0.22	73.31 ± 3.18	98.59 ± 1.48	96.94 ± 1.60
Spectral coherence	99.00 ± 0.00	66.66 ± 1.52	99.88 ± 0.14	97.08 ± 1.75
Phase difference	93.84 ± 0.21	67.67 ± 1.58	99.24 ± 0.81	97.25 ± 2.00
Wavelet coefficients	88.96 ± 4.08	55.70 ± 4.46	99.32 ± 0.99	93.65 ± 5.00
Correlation-based	99.02 ± 0.00	67.78 ± 1.82	99.88 ± 0.15	97.36 ± 2.01
PCA	97.76 ± 0.12	65.25 ± 2.30	99.54 ± 0.48	96.54 ± 1.66
AR	86.90 ± 0.29	77.25 ± 1.43	98.37 ± 1.32	97.53 ± 1.28

systems.

The significance of the ERP-EEG results implies the uniqueness of the cortical activities for each subject. On the other hand, the brain is an open system (cybernetic) and it receives several inputs from the five types of sensors along with cortical activities during the mental and emotional excitation. Therefore, unlike several attempts for distinguishing subjects according to their resting-EEG features, the identification result is not promising due to the wide range of factors influencing the EEG signals [14]. Our results on the ERP-EEG signals over three classifiers show that the spectral coherence and correlation have the best discriminative information for the person identification task. The Student's t-test is applied to determine the difference between the efficiency of the mentioned features is significant or not. The significance level of $p < 0.05$ is considered statistically significant. Our statistical evaluation over the three classifiers shows that these two features (spectral coherence and correlation) are significantly ($p < 0.05$) superior to the other compared features.

Both correlation and spectral coherence can measure the relationship between two signals. Coherence value is a correlation coefficient in the frequency domain, while the correlation feature shows the similarity of signals in the time domain. Spectral coherence measures the degree of linear dependency of similar frequency components between two signals. In other words, it can measure the degree of information flow between groups of neurons activities of the EEG signal. The excellent results given by the correlation and spectral coherence show the way which functional networks can cooperate with each other during neurocognitive processes. These are a high-resolution way to quantify the degree of dynamic connectivity between the diverse areas in the brain [46, 47]]. Additionally, our results show difference between the efficiency of wavelet coefficients in Tables 1–3. Since wavelet decomposition adaptively provides dyadic resolution over the frequency range, these frequency divisions are matched with the EEG standard bands. However, the accuracy of wavelet may be changed with several factors such as mother wavelet and number of decomposition levels. Efficiently setting of these factors will lead to high verification and identification accuracy.

In addition to the classification accuracy, false accept rate (FAR) and false reject rate (FRR) and EER are measured [48]. A False acceptance allows an impostor to get an access, and a false reject denies an access to the enrolled user. These criteria are simply defined below:

$$FAR = \frac{Number\ of\ accepted\ impostors}{Total\ number\ of\ impostors} \tag{10}$$

$$FRR = \frac{Number\ of\ rejected\ genuines}{Total\ number\ of\ genuines} \tag{11}$$

Table 4
The EER% of kNN classifier for resting-EEG and ERP-EEG signals.

Method	Person Identification		Person Verification	
	ERP-EEG	Resting-EEG	ERP-EEG	Resting-EEG
LBP	0.14	0.05	0.14	0.05
Spectral coherence	0.00	0.00	0.00	0.00
Phase difference	0.03	0.41	0.03	21.85
Wavelet coefficients	0.00	0.00	0.00	0.00
Correlation-based	0.00	0.00	0.00	0.00
PCA	0.04	4.92	0.01	12.34
AR	0.46	1.36	0.46	0.29

Table 5
The EER% of SVM classifier for resting-EEG and ERP-EEG signals.

Method	Person Identification		Person Verification	
	ERP-EEG	Resting-EEG	ERP-EEG	Resting-EEG
LBP	0.06	0.00	16.99	18.34
Spectral coherence	0.00	0.29	0.00	2.35
Phase difference	0.34	0.00	23.95	35.00
Wavelet coefficients	49.67	0.00	50.00	50.00
Correlation-based	0.00	0.00	0.00	0.20
PCA	2.49	5.56	17.45	37.14
AR	0.19	0.00	9.98	22.05

Table 6
The EER% of random forest classifier for resting-EEG and ERP-EEG signals.

Method	Person Identification		Person Verification	
	ERP-EEG	Resting-EEG	ERP-EEG	Resting-EEG
LBP	0.00	0.00	0.37	0.00
Spectral coherence	0.00	0.00	0.00	0.00
Phase difference	0.00	0.00	0.07	0.00
Wavelet coefficients	0.02	0.44	0.00	25.30
Correlation-based	0.00	0.00	0.00	0.00
PCA	0.00	0.00	0.00	19.45
AR	0.35	0.29	0.12	5.70

$$EER = \frac{FAR + FRR}{2} \quad when \quad FAR = FRR \tag{12}$$

Tables 4–6 show the EER (%) for the deployed resting-EEG and ERP-EEG signals in both verification and identification systems. Our results shows that the random forest and kNN obtain the better results in comparison with SVM.

It is noticeable to state that specialists visually analyze the ERP waveforms for diagnosis purposes by measuring their amplitude and latency. In this regard, the ERP signals are extracted from the EEGs and then applied to the classifiers to see how discriminative these waveforms are for authenticating the subjects. The ERP-EEG epochs following a given stimulus are averaged (over 80 stimuli) to remove the background EEG, for eliciting the ERP template. Fig. 2 shows the extracted ERP signal on Fz, Cz and Pz channels for a randomly selected subject. Since the most of classifiers require multiple examples from each participant in order to learn input–output mappings robustly; consequently, it is not sufficient to simply create one ERP trial from each participant for training of the classifiers. A bootstrapping procedure is applied to both train and test data, where 100 ERP trials are generated for each participant [30].

Finally, the samples of ERP waveform are considered as features to be applied to kNN, SVM and random forest classifiers. Table 7 shows the accuracy and EER for the ERP-based verification and identification systems. Unlike our expectation, the classification results show the ERP waveforms are not suitable for the identification task, while its verification results are plausible.

Despite the achieved excellent ERP-EEG based verification and identification results, there are certain aspects that may need further investigation. These important points are:

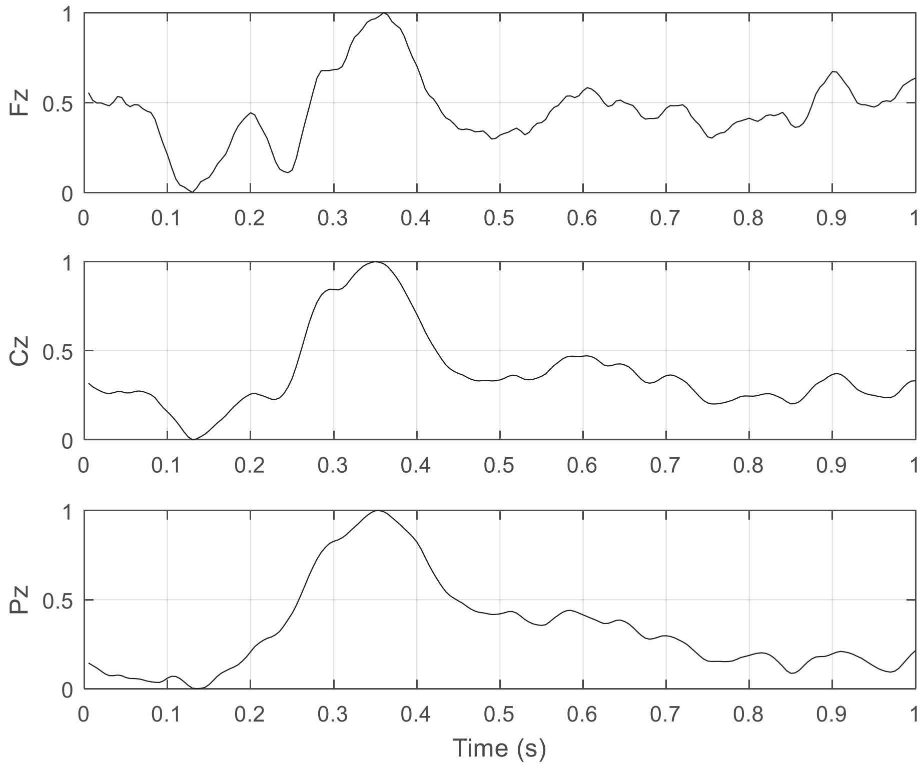

Fig. 2. The synchronous averaging over ERP-EEG signal on Fz, Cz, Pz channels for a randomly selected subject.

- Our results demonstrate that the ERP-EEG signals have high inter-subject variability and low inner-subject reliability; therefore, ERP-EEG-based individual identification is a good candidate to design a reliable authentication system (in confirmation with previous studies [30, 50]]). Nonetheless, to address the robustness, participants should return to the EEG lab after their first session to analyze the stability of our results over the time.
- EEG signals may be influenced by motion artifacts and body movement. In order to reduce the effect of these undesired components, it is essential to select better filtering algorithm to obtain signals with better quality. ERP is less sensitive to the background noise (the ongoing EEG) because the background EEG and additive noises behave like the white/pink Gaussian noise and therefore they are attenuated and faded through the averaging. Synchronous averaging assumes the ERP template is fairly constant in response to each stimulus and through the long time, brain fatigue leads to skew the ERP latency and amplitude.
- The quality of EEG depends on the quality EEG setup. An apparatus with more number of electrodes (e.g., 128 or 256) is usually more accurate and if the process is performed by a high quality apparatus with dense array, the results might be increased due to covering more spatial resolution and better quality of signals. Although the results are acceptable with 24 electrodes, the acquisition of EEG signals with more number of electrodes will definitely add more discriminative information.

Table 7
The obtained results of the three classifiers for ERP (P300) component.

Method	Person Identification		Person Verification	
	Identification rate	EER (%)	verification rate	EER (%)
kNN	65.10 ± 0.00	1.82	96.79 ± 1.51	2.11
SVM	38.10 ± 0.00	10.04	95.29 ± 1.10	1.50
Random forest	73.14 ± 0.00	3.92	97.11 ± 1.53	1.50

- The important limitation that must be addressed in future work is the sample size. The sample size is small in our study. To prove the usefulness of these tools for subject identification task, these measures should be applied to larger population. Since, a normal EEG varies with the age, these features must be applied to study of EEG signals during different intervals of age (it is shown that EEG in childhood generally shows slower frequency oscillation than an adult).
- EEG recording in the resting state needs no active involvement of subjects during the recording; therefore, the probable effect of mental fatigue is highly reduced. Moreover, our results show the existence of exclusive information in the resting-EEG of each subject, which is emerged from the exclusive connections in the brain system of each subject [24]. EEG signals in the time domain obey not grammatical behavior and they behave like a white/color noise, nonetheless it seems that the subject-dependent information is captured through this signal. It is shown the power spectrum of EEG signal, especially in the range of Alpha band contains subject information that can be used in person verification [49].

6. Conclusion and future work

In this paper, we investigate the role of state-of-the-art features in verification and identification of subjects via classifying their resting-EEG, ERP-EEG and the averaged ERP features. The universality, measurability, uniqueness and robustness against fraudulent attacks are favorite characteristic of use of resting-EEG and ERP-EEG signals for human recognition. As we have expected, the verification results provide more accuracy rather the identification results in terms of classification accuracy and equal error rate. This difference is emerged from this fact that the identification process is a multi-class problem (in this research 20 classes), while the verification process is a two-class problem. Moreover, the results of ERP-EEG features provide higher accuracy rather than the two other approaches, because ERP-EEG contains the information of both background EEG and ERP, which are richer than each of them. Although ERP has a grammatical pattern and reveals the response of a specific sub-system of the brain, the background EEG contains unique information for each subject and combination of them provide robust features. We suggest the ERP-EEG based authentication system as an efficient biometric tool due to its convincing results.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors of this paper present a special thanks to Dr. Gregory W. Price who has given us his database including resting-EEG and ERP-EEG signals of healthy participants.

References

[1] Jain A, Ross AA, Nandakumar K. Introduction to biometrics. Spriger US; 2011.
[2] Biomedical informatics and technology. In: Pham TD, Ichikawa K, Oyama-Higa M, Coomans D, Jiang X, editors. First international conference on communications in computer and information science. Japan: Springer; 2013.
[3] Ramalho M, Singh S, Correia PL, Soares LD. Biometric identification through palm and dorsal hand vein patterns. 2011. IEEE International Conference on Computer as a Tool, Portugal.
[4] Yazdani S, Vesin JM. Extraction of QRS fiducial points from the ECG using adaptive mathematical morphology. Digit Signal Process 2016;56:100–9.
[5] Britton JW, Frey LC, Hopp JL. Electroencephalography (EEG): an introductory text and atlas of normal and abnormal findings in adults, children, and infants. Chicago: American Epilepsy Society; 2016.
[6] Ghofrani N, Boostani R. Reliable features for an ECG-based biometric system. 17th Iranian Conference of Biomedical Engineering (ICBME); 2010.
[7] Jain A, Ross AA, Prabhakar S. An introduction to biometric recognition. IEEE Trans Circ Syst Video Technol 2004;14(1):4–20.
[8] Campisi P, Rocca DL. Brain waves for automatic biometrics-based user recognition. IEEE Trans Inf Forensics Secur 2014;9(5):782–800.
[9] Sabeti M, Katebi SD, Boostani R. Entropy and complexity measures for EEG signal classification of schizophrenic and control participants. Artif Intell Med 2009;47(3): 263–74.
[10] Parvinnia E, Sabeti M, Jahromi MZ, Boostani R. Classification of EEG Signals using adaptive weighted distance nearest neighbor algorithm. Journal of King Saud University-Computer and Information Sciences 2014;26(1):1–6.
[11] Nasihatkon B, Boostani R, Jahromi MZ. An efficient hybrid linear and kernel CSP approach for EEG feature extraction. Neurocomputing 2009;73(1–3):432–7.
[12] Fattahi D, Nasihatkon B, Boostani R. A general framework to estimate spatial and spatio-spectral filters for EEG signal classification. Neurocomputing 2013;119: 165–74.
[13] Sabeti M, Boostani R, Katebi SD, Price GW. Selection of relevant features for EEG signal classification of schizophrenic patients. Biomed Signal Process Contr 2007; 2(2):122–34.
[14] Sanei S, Chambers JA. EEG signal processing. England: John Wiley & Sons; 2007.
[15] Nunez PL, Srinivasan R. Electric fields of the brain: the neurophysics of EEG. New York: Oxford University Press; 2006.
[16] Pozo-Banos MD, Alonso JB, Ticay-Rivas JR, Travieso CM. Electroencephalogram subject identification: a review. Expert Syst Appl 2014;41:6537–54.
[17] Singh YN, Singh SK, Ray AK. Bioelectric signals as emerging biometrics: issues and challenges. ISRN Signal Processing 2012;712032:1–13.
[18] Hitchcock DC. Evaluation and combination of biometric authentication systems. Master Thesis; 2003.
[19] Boostani R, Moradi MH. A new approach in the BCI research based on fractal dimension as feature and Adaboost as classifier. J Neural Eng 2004;1(4):212–7.
[20] Alimardani F, Boostani R, Blankertz B. Weighted spatial based geometric scheme as an efficient algorithm for analyzing single-trial EEGs to improve cue-based BCI classification. Neural Network 2017;92:69–76.
[21] Pozo-Banos MD, Travieso CM, Weidemann CT, Alonso JB. EEG biometric identification: a through exploration of the time-frequency domain. J Neural Eng 2015;12:1–23.
[22] Pal A, Gautam AK, Singh YN. Evaluation of bioelectric signals for human recognition. International Conference on Intelligent Computing, Communication & Convergence 2015;48:746–52.
[23] Fraschini M, Hillebrand A, Demuru M, Didaci L, Marcials GL. An EEG-based biometric system using eigenvector centrality in resting state brain networks. IEEE Signal Process Lett 2015;22(6):666–70.
[24] Rocca DL, Campisi P, Vegso B, Cserti P, Kozmann G, Babiloni F, Fallani DV. Human Brain Distinctiveness based on EEG spectral coherence connectivity. IEEE (Inst Electr Electron Eng) Trans Biomed Eng 2014;61(9):2406–12.
[25] Gui Q, Jin Z, Blondet MVR, Laszlo S, Xu W. Towards EEG biometrics: similarity-based approaches for user identification. IEEE International Conference on Identity, Security and Behavior Analysis, Hong Kong 2015:1–6.
[26] Zhao Q, Peng H, Hu B, Liu Q, Liu L, Qi Y, Li L. Improving individual identification in security check with an EEG based biometric solution. LNAI 2010;6334:145–55.
[27] Huang X, Altahat S, Tran D, Shutao L. Human identification with electroencephalogram (EEG) for the future network security. LNCS 2013;7873: 575–81.
[28] Riera A, Soria-Frisch A, Caparrini M, Grau C, Ruffini G. Unobtrusive biometric system based on electroencephalogram analysis. EURASIP J Appl Signal Process 2008:1–8.
[29] Nazhvani AD, Boostani R, Afrasiabi S, Sadatnezhad K. Classification of ADHD and BMD patients using visual evoked potential. Clin Neurol Neurosurg 2013;115(11): 2329–35.
[30] Armstrong BC, Ruiz-Blondet MV, Khalifian N, Kurtz KJ, Jin Z, Laszlo S. Brainprint: assessing the uniqueness, collectability and permanence of a novel method for ERP biometrics. Neurocomputing 2015;166:59–67.
[31] Das K, Zhang S, Giesbrecht B, Eckstein MP. Using rapid visually evoked EEG activity for person identification. USA: IEEE International Conference on Engineering in Medicine and Biology Society; 2009.
[32] Palaniappan R. Two stage biometric authentication method using thought activity brain waves. Int J Neural Syst 2008;18(1):59–66.
[33] Kumar TS, Kanhangad V, Pachori RB. Classification of seizure and seizure-free EEG signals using local binary patterns. Biomed Signal Process Contr 2015;15:33–40.
[34] Kaya Y, Uyar M, Tekin R, Yildirim S. 1D-local binary pattern based feature extraction for classification of epileptic EEG signals. Appl Math Comput 2014;243: 209–19.
[35] Sabeti M, Katebi SD, Boostani Price GW. A new approach for EEG signal classification of schizophrenic and control participants. Expert Syst Appl 2011; 38(3):2063–71.
[36] Semlitch HV, Anderer P, Schuster P, Presslich O. A solution for reliable and valid reduction of ocular artifacts applied to the P300 ERP. Psychophysiology 1986;23: 696–703.
[37] Chen D, Wan S, Xiang J, Bao FS. A high-performance seizure detection algorithm based on discrete wavelet transform (DWT) and EEG. PloS One 2017:0173138.
[38] Burrus S. Introduction to wavelet and wavelet transforms. Prentice Hall; 1997.
[39] Subasi A, Gursoy MI. EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Syst Appl 2010;37(12):8659–66.
[40] Lekshmi SS, Selvam V, Rajasekaran MP. EEG signal classification using principal component analysis and wavelet transform with neural network. India: International Conference on Communications and Signal Processing; 2014.
[41] Sabeti M, Boostani R. Separation P300 event-related potential using time varying time-lag blind source separation algorithm. Comput Methods Progr Biomed 2017; 145:95–102.
[42] Stoica P, Moses RL. Introduction to spectral analysis. USA: Prentice Hall; 1997.
[43] Fix E, Hodges Jr JL. Discriminatory analysis, non-parametric discrimination, and consistency properties. Int Stat Rev 1989;57(3):238–47.
[44] Wang L. Support vector machine: theory and applications. Springer-Verlag; 2005.
[45] Breiman L. Random forests. Machine Learning Journal 2001;45(1):5.
[46] Seleznov I, Zyma I, Kiyono K, Tukaev S, Popov A, Chernykh M, Shpenkov O. Detrended fluctuation, coherence, and spectral power analysis of activation rearrangement in EEG dynamics during cognitive workload. Front Hum Neurosci 2019;13:270.
[47] Borges AFT. Spectral and coherence estimates on electroencephalogram recordings during arithmetical tasks. master thesis; 2009.
[48] Alimardani F, Boostani R. Improvement of the performance of fingerprint verification using a combinatorial approach. Biomed Eng: Applications, Basis and Communications 2018;30(3):1850019.
[49] Haegens S, Cousijn H, Wallis G, Harrison PJ, Nobre AC. Inter- and intra-individual variability in alpha peak frequency. Neuroimage 2014;92:46–55.
[50] Hu B, Liu Q, Zhao Q, Qi Y, Peng H. A real-time electroencephalography (EEG) based individual identification interface for mobile security in ubiquitous environment. IEEE Asia-Pacific Services Computing Conference 2011:436–41.