
HIGH ACCURACY EEG BIOMETRICS IDENTIFICATION USING ICA AND AR MODEL

Chesada Kaewwit, Chidchanok Lursinsap & Peraphon Sophatsathit

*Advanced Virtual and Intelligent Computing Center
Faculty of Science, Chulalongkorn University, Thailand*

ckaewwit@gmail.com; lchidcha@chula.ac.th; peraphon.s@chula.ac.th

ABSTRACT

Modern biometric identification methods combine interdisciplinary approaches to enhance person identification and classification accuracy. One popular technique for this purpose is Brain-Computer Interface (BCI). The signal so obtained from BCI will be further processed by the Autoregressive (AR) Model for feature extraction. Many researches in the area find that for more accurate results, the signal must be cleaned before extracting any useful feature information. This study proposes Independent Component Analysis (ICA), k-NN classifier, and AR as the combined techniques for electroencephalogram (EEG) biometrics to achieve the highest personal identification and classification accuracy. However, there is a classification gap between using the combined ICA with the AR model and AR model alone. Therefore, this study takes one step further by modifying the feature extraction of AR and comparing the outcome with the proposed approaches in lieu of prior researches. The experiment based on four relevant locations shows that the combined ICA and AR can achieve higher accuracy than the modified AR. More combinations of channels and subjects are required in future research to explore the significance of channel effects and to enhance the identification accuracy.

Keywords: electroencephalogram (EEG), Autoregressive (AR), Independent Component Analysis (ICA), Biometrics, feature extraction, person classification.

Received: 16 January 2017

Accepted: 3 July 2017

INTRODUCTION

Biometrics is a person authentication and identification technique using several real organs. Biometrics function can be divided into two steps: identification and authentication. The first step performs verification and validation of individuals in a database or a group of persons. The latter step performs acceptance of individuals. Biometrics is likely to be feigned due to medical advance and information technology development such as plastic surgery, high resolution devices, or advanced digital tools. An electroencephalography (EEG) is considered as one biometric trait because it is generated from the human brain by recording the electrical activity of millions of neurons from the same position (Teplan, 2002). Nevertheless, owing to noninvasive, inexpensive, and portable EEG device, many researches focus on using EEG as biometrics (Maiorana et al., 2016; Mu, et al., 2016; Rocca, et al., 2012; Tangkraingij, et al., 2013).

Generally, EEG can be used to diagnose the function of the brain. The functional area in the human brain is divided into four areas, namely frontal, parietal, temporal and occipital lobes. The frontal lobe is liable for conscious thought. The parietal lobe combines sensory information from various senses. The temporal lobe performs auditory sense. Sense of sight is processed by the occipital lobe. EEG can be grouped into five different rhythms based on their frequencies: Delta rhythm (1-4 Hz) is seen during deep sleep in adults and in infants as an unusual activity; Theta rhythm (4-8 Hz) occurs in the drowsiness of adults and in the waking up of children; Alpha rhythm (8-12 Hz) is seen normally while relaxing with eyes closed; Beta rhythm (12-30 Hz) is associated with anxious thinking and active concentration; and Gamma rhythm (30-100 Hz) is associated with specific cognitive states. Many researches have focused on person authentication based on these EEG rhythms (Campisi et al., 2011; Rocca et al., 2012).

Furthermore, EEG biometrics can be used in Event-Related Potential (ERP). ERP measures brain waves while each subject does some activities such as looking at a picture, listening to audio, fisting, and so on. Many researches on ERP using EEG biometrics obtain satisfactor high accuracy (Kumari & Vaish, 2016; Ruiz-Blonded et al., 2016).

In neuroscience, Independent Component Analysis (ICA) is an important technique used to remove undesirable artifacts in EEG signals to obtain purer brain waves from each location of the brain (Albera et al., 2012). However,

there are only a few studies concentrating on ICA in EEG signal analysis (He & Wang, 2009; Tangkraingkiy et al., 2013), but none uses ICA to clean EEG signals. As a consequence, we set out to explore these combined techniques in this study to attain higher person identification accuracy.

The organization of this paper is as follows. Some related work will be recounted in the next section. Then, background theories are discussed, followed by the methodology. Next, the experimental results are described, and some final thoughts are given in the conclusion and future work section.

RELATED WORK

In Brain Computer Interface (BCI), it was noted that ICA could be used to extract the EEG signals so as to obtain sources of brain signals (Albera et al., 2012; Delorme et al., 2012). In EEG biometrics study, Tangkraingkiy's researches (Tangkraingkiy et al., 2010; Tangkraingkiy et al., 2013) used ICA to clean EEG signals collected from 20 subjects in the resting state. They grouped the data into varied lengths (500, 1000, 1500, and 3000 data points) and used artificial neuron network for classification. Their results showed that 4-channels of EEG at 1,000 data points per channel achieved high average identification accuracy at 98.51%. The limitations of their work were too many data points were used for classification, namely 4,000 data points for 4 channels. Unfortunately, such a highly accurate result could only be obtained in offline processing. He and Wang (2009) used the AR model and 5 dominating Independent Components from 5 brain areas with 7 subjects in motion tasks. The identification accuracy was not as high as anticipated.

For this reason, more sophisticated feature extraction techniques were required in the identification process such as spectral coherence (Del Pozo-Banos et al., 2014), power spectrum (Harshit et al., 2016), fuzzy entropy (Mu et al., 2016), AR model (Campisi et al., 2011; Maiorana et al., 2016; Rocca et al., 2012). These researches were made to apply the AR model for EEG biometrics. For example, Poulos (Poulos et al., 1999) used the AR model to achieve 80 to 95 percent accuracy with 4 subjects. Paranjape (Paranjape et al., 2001) used AR and variance/covariance matrices with statistical tools to model the EEG signal from a single channel. The experiment employed 40 subjects in the resting state with eyes opened and eyes closed. The accuracy percentage was only 80 percent. Yazdani (Yazdani et al., 2008) employed the AR model, the peak of power spectral density (PSD) for feature extraction,

and LDA for dimension reduction to improve higher accuracy. The resulting accuracy percentage reached 100 with 20 subjects. Riera (Riera et al., 2008) introduced a new feature extraction approach using combinations of five models, namely AR, Fourier Transform, Mutual Information, Coherence, and Cross Correlation. They applied Fisher's Discriminant Analysis as classifier in this study. There were 51 subjects and 36 intruders. The accuracy percentage ranged from 87.5 to 98.1 percent. Campi (Campisi et al., 2011) used reflection coefficients of the AR model to yield 96.08% accuracy with 48 subjects. The research of Rocca (Rocca et al., 2012) used three electrodes with sub-bands of EEG signals to classify 45 persons in the resting state with closed eyes. The AR stochastic and polynomial approaches were used for feature extraction in this study. Recognition rate was about 98.73 percent with 45 persons in the resting state. Maiorana (Maiorana et al., 2016) experimented and reported that using reflection coefficient of AR method gave a discriminating capability higher than using the PSD and COH as feature extraction methods. The results yielded higher identification accuracy when comparing to other features. Thus, the coefficient was considered as the main feature in this work.

BACKGROUND THEORIES

Independent Component Analysis (ICA)

Independent Component Analysis (ICA) is a technique of blind source separation (BSS) which divides multivariate signals into individual source signals or components that are statistically independent. ICA can be expressed by Eq. (1)

$$\mathbf{o} = \mathbf{A}\mathbf{b} \quad (1)$$

where \mathbf{o} denotes the mixture vector of mixing between the source signals \mathbf{b} and unknown mixing matrix \mathbf{A} . This technique can be set up in EEG as follows: let \mathbf{o} be the EEG recorded from the n -electrodes on the scalp, i.e. $\mathbf{o} = [o_1 \ o_2 \ o_3 \ \dots \ o_n]^T$, let $\mathbf{b} = [b_1 \ b_2 \ b_3 \ \dots \ b_n]^T$ be the source EEG signals. Both constitute an EEG vector of n signals at time t . Since the value of each b_i and mixing matrix \mathbf{A} are unknown, the known value o_i is multiplied by the inverse of \mathbf{A} , denoted by \mathbf{D} , to obtain an estimated source EEG signals \mathbf{y} of \mathbf{b} . The derived EEG signal becomes

$$\mathbf{y} = \mathbf{D}\mathbf{o} \quad (2)$$

where $\mathbf{y} = [y_1, y_2, y_3 \dots y_n]^T$. There are many algorithms which provide accurate estimated value of matrix \mathbf{D} such as JADE (Cardoso & Souloumiac, 1993), SOBI (Belouchrani et al., 1997), SOBIRO (Belouchrani & Cichocki, 2000), FastICA (Hyvärinen & Oja, 2000), ERICA (Cruces et al., 2002), and so on. In this work, the SOBIRO algorithm was selected as the main ICA process.

Autoregressive Model (AR)

Autoregressive model (AR) is a representation of a signal using its own previous values. It specifies that the present output is a linear combination of the previous z outputs combining with white noise. In general, AR(z) denotes an autoregressive model of order z , which is defined by Eq. (3).

$$x_t = \sum_{i=1}^z \varphi_i x_{(t-i)} + n_t \quad (3)$$

where x_t is the output signal of the sample point t , θ_i is the coefficient of AR, z is the order less than t , and n_t is a white noise input. The coefficient θ can be determined by Yule-Walker and solved by the Levinson method to obtain the reflection coefficient of the AR model (Maiorana et al., 2016). Anyhow, AR with the Burg method (Bos et al., 2002) is a popular approach for evaluating the AR coefficient and reflection coefficient since it yields a high resolution for small data points. The Burg method can evaluate the coefficient by first estimating the reflection coefficient defined as the last autoregressive parameter estimate for each model of z . Details of the Burg method can be found in de Hoon et al. (1996). Since Burg method can compute the reflection coefficients directly from signal x_t , this study adopted this method to compute reflection coefficients of each channel as the main feature for the proposed biometric identification.

k-Nearest Neighbor (k-NN)

k-Nearest Neighbor is a simple classifier with respect to calculating the distance to all trained data sets (\mathbf{V}_T^u) and test data set. Each test data (\mathbf{V}) is predicted as belonging to a class or a subject which is in the trained set having the shortest Euclidean distance. The decision is taken in accordance with the majority voting rule. The identity of a class depends on the number of occurrences from the votes (k). The Euclidean distance can be expressed by Eq. (4).

$$\mathbf{d}_{\text{Euclid}}^u(\mathbf{V}) = \min \sqrt{\sum_{i=1}^n (\mathbf{V}[i] - \mathbf{V}_T^u[i])^2} \quad (4)$$

where d_{Euclid} denotes the minimum distance between the test data set and the trained data set, n is the number of features, \mathbf{V} is the vectors in the test data set or the specific data set, u is the number of classes, T is the number of samples in the trained data set. The correct recognition rate (CRR) is used to measure the accuracy of classification. In essence, it represents the true positive numbers of each class. CRR is defined by Eq. (5) and Eq. (6) as follows:

$$CRR_f = \left(\frac{1}{u} \sum_{i=1}^u p_{(i)} \right) \times 100 \quad (5)$$

where f denotes the number of folds for cross validation, u is the number of subjects, $P_{(i)}$ is the number of correct recognitions of each class or subject, that is, the number of true positives of each class. Eventually, each CRR_f is averaged to be CCR, that is,

$$CRR = \text{Average}(CRR_f) \quad (6)$$

METHODOLOGY

Data Collection

Data of the EEG signals were obtained from the prior research data set of Tangkraingij. These EEG signals were recorded by the Chulalongkorn Comprehensive Epilepsy Program (CCEP) of King Chulalongkorn Memorial Hospital. There were 20 normal subjects, eight men and twelve women, tested in the resting state, motionless, and performed no task during recording. The ages of all the subjects ranged from 12 to 40 years. Recording was based on the 10-20 international system defining the location of scalp electrodes. Sixteen electrode placements were attached to the scalp of each subject at the following locations: Fp1, F7, T3, T5, Fp2, F8, T4, T6, F3, C3, P3, O1, F4, C4, P4, and O2 as shown in Figure 1(a). To simplify the reference to these experiments, they were labeled as ch1, ch2,..., ch16 in Figure 1(b), respectively.

During recording, the Mastoid area A1 and A2 were electrically linked and used as reference with the mono-polar montage. The EEG amplifier was the Grass model 8 plus. The sampling rate was 200 Hertz. The EEG signals were digitized and notch filtered at 50 Hertz by the BMSI board using the Stellate Harmony EEG software. The digitized EEG data were exported as EDF

(European Data Format). Electromyography (EMG) and Electrocardiogram (ECG) signals were subsequently removed from EEG signals. For each channel of each subject, 3,000 data points was collected in 15 seconds. This study selected four channels, namely F4, C4, P4, and O2 as depicted in Figure 1 based on prior work (Tangkraingij et al., 2013) to establish the relevant locations for classification having the highest average accuracy of 98.51%. Moreover, the selected channels were positions of the brain having the biological functions that were essential for person identification.

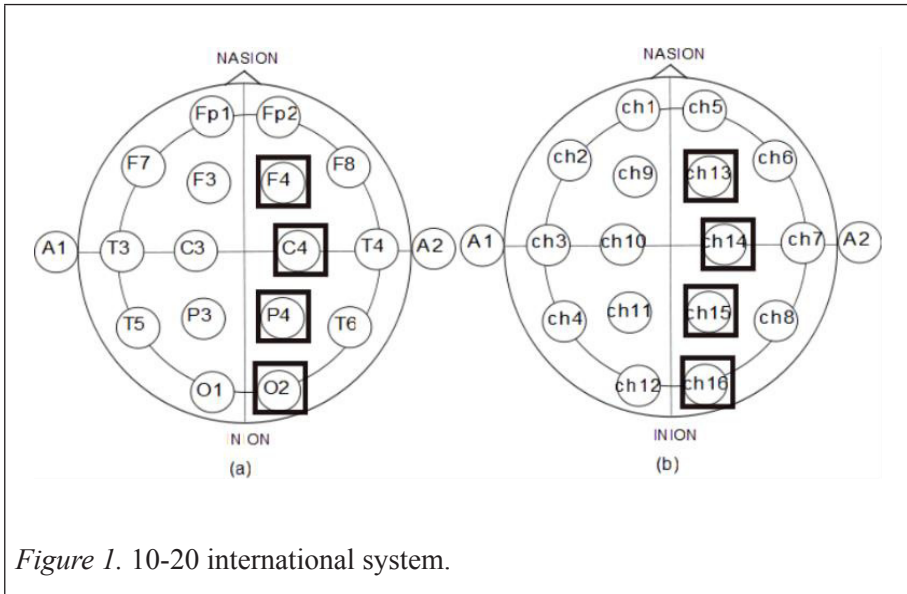


Figure 1. 10-20 international system.

Proposed Method and Experiment Setup

This study conducted 2 approaches. The first approach was called the modified AR method, consisting of 4 processes as shown in Figure 2. The first process obtained four selected channels EEG data filtering in the frequency range of 0.5-40 Hz. Subsequent processes were data segmentation, feature extraction with the AR model, and classification. The second approach was the proposed approach called ICA with the AR method, consisting of 5 processes as shown in Figure 3. The first process applied ICA using the SOBIRO algorithm to the EEG data set. The remaining four processes were the same as those of the modified AR method. Both approaches were experimented and compared with those of Tangkraingij et al. (2013).

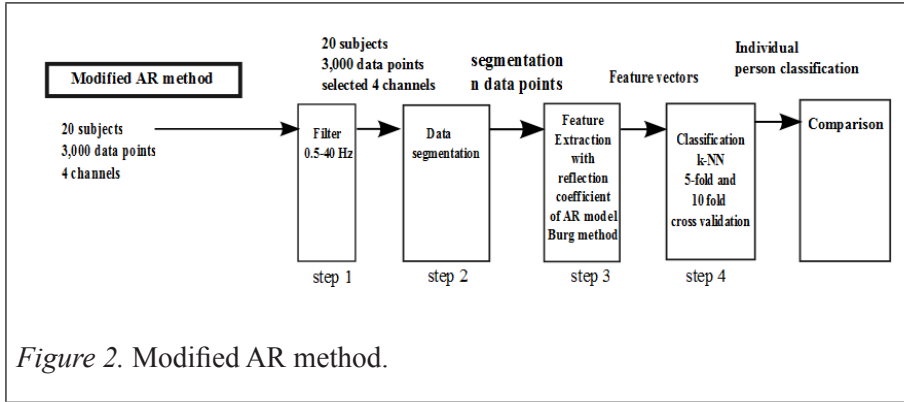


Figure 2. Modified AR method.

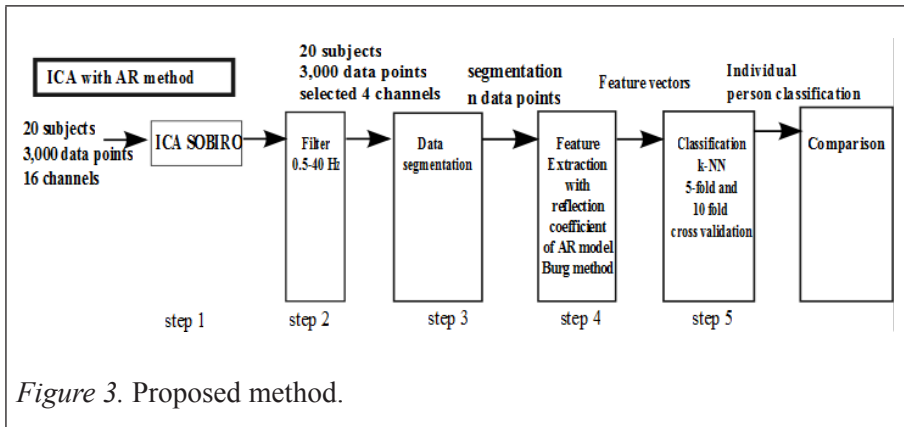


Figure 3. Proposed method.

Detailed implementation of the proposed approach is described in the algorithm below.

Algorithm ICA with the AR method

ICAwithAR: Compute CRR

Input: S_e^u : EEG signal matrix, where u is number subjects and e is number of channels [ch1, ch2,..., ch16]

Output: CRR_f^u : Average accuracy identification of each subject, where f denotes fold cross validation

1. idx5 ← determine indexing of trained and test data set for 5-fold cross validation
2. idx10 ← determine indexing of trained and test data set for 10-fold cross validation

3. $k \leftarrow$ number of majority votes measured by minimum Euclidean distance
4. //Step 1: ICA process
5. **for** $u = 1$ to 20 **do**
6. $\mathbf{S}_e^u \leftarrow$ ICA (\mathbf{S}_e^u)
7. **end for**
8. //Step 2: Filter process
9. **for** $u = 1$ to 20 **do**
10. $\mathbf{S}_e^u \leftarrow$ BandpassFilter (\mathbf{S}_e^u , 0.5, 40)
11. **end for**
12. $e = [\text{ch13}, \text{ch14}, \text{ch15}, \text{ch16}]$
13. //Step 3: Generate matrix \mathbf{V}^u for segmentation
14. $\mathbf{V}_{h_o}^u = [\mathbf{S}_1(1 + (h-1)n : hn), \mathbf{S}_2(1 + (h-1)n : hn), \dots]$
15. $\mathbf{V}_{j_o}^u = [\mathbf{S}_1(1 + (j-1)n/2 : (j-1)n/2), \mathbf{S}_2(1 + (j-1)n/2 : (j-1)n/2), \dots]$
where $\mathbf{V}_{h_o}^u$ is odd number [1,3,..], $\mathbf{V}_{j_o}^u$ is even number [2,4,..], n is the number of data points per segment.
16. //Step 4: Generate order z of AR using Burg method
17. **for** $z = 1$ to 60 **do**
18. **for each** segment of **do**
19. ${}^{(z)}\mathbf{V}_m^u \leftarrow$ ARBurg (\mathbf{V}_m^u, z)
20. **end for**
21. **end for**
22. //Step 5: Generate trained data set, ${}^{(z)}\mathbf{V}_{Tf}^u$ and test data set ${}^{(z)}\mathbf{V}_{Xf}^u$ with ${}^{(z)}\mathbf{V}_m^u$
idx5 and idx10 from
23. ${}^{(z)}\mathbf{V}_T^u = {}^{(z)}\mathbf{V}_m^u(\text{idx5}); \quad {}^{(z)}\mathbf{V}_X^u = {}^{(z)}\mathbf{V}_m^u(\text{idx5});$
24. ${}^{(z)}\mathbf{V}_T^u = {}^{(z)}\mathbf{V}_m^u(\text{idx10}); \quad {}^{(z)}\mathbf{V}_X^u = {}^{(z)}\mathbf{V}_m^u(\text{idx10});$
25. //k-NN classification
26. **for** $z = 1$ to 60 **do**
27. **for** $u = 1$ to 20 **do**
28. $\text{CRR}_f^u = \text{Average}\left(\text{CRR}_f\left(\text{k-NN}\left({}^{(z)}\mathbf{V}_T^u, {}^{(z)}\mathbf{V}_X^u, k\right)\right)\right)$
29. **end for**
30. **end for**

The following set ups were established to accommodate the experiment. Each process is as follows.

ICA

The objective of ICA is to clean and separate individual EEG signals from those obtained at other locations of the brain. All channels of each subject were processed with this method. This process is shown in line 4.

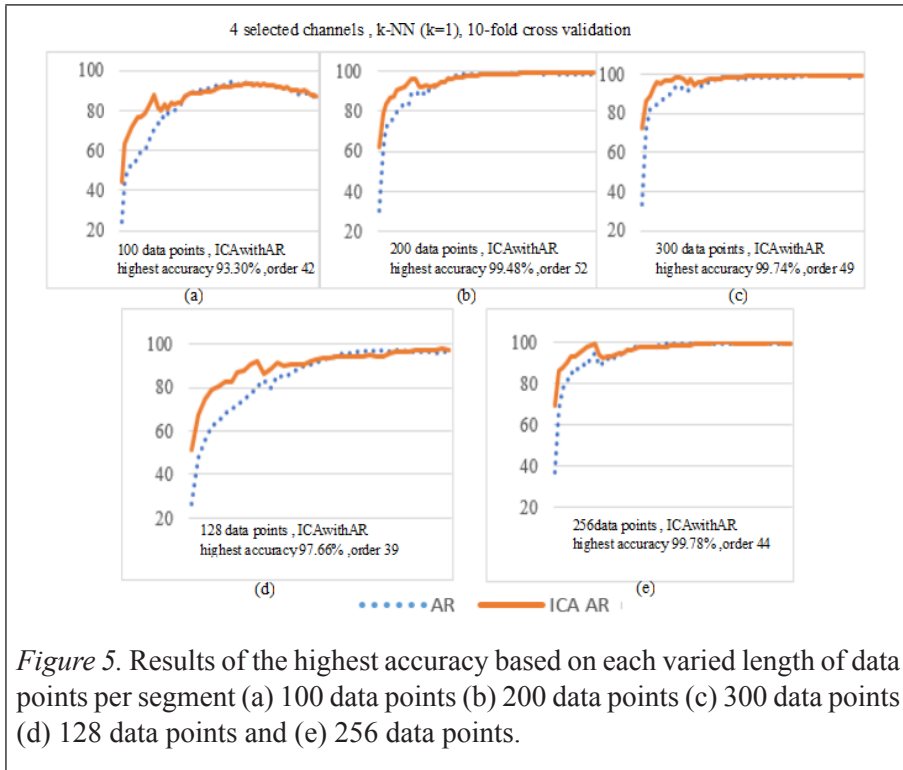


Figure 5. Results of the highest accuracy based on each varied length of data points per segment (a) 100 data points (b) 200 data points (c) 300 data points (d) 128 data points and (e) 256 data points.

In Figure 5, we compared the highest accuracy from varied lengths of each segment. Based on the limited number of data points per segment, 128 and 256 data points were used instead of 100, 200, and 300 data points so that the number of experiments would be reduced, yet still maintained relatively equivalent identification coverage. Moreover, with 128 and 256 data points, the highest accuracy with the lower order of the AR model was obtained as shown in Tables 1 and 4.

Table 1

Classification Accuracy of 20 Subjects with 128 Data Point Length Performed on the ICA with the AR Method for 5-fold and 10-fold Cross Validation with k-NN having k=1, 3 and 5

Order	5-fold cross validation			10-fold cross validation		
	k=1 %(SD)	k=3 %(SD)	k=5 %(SD)	k=1 %(SD)	k=3 %(SD)	k=5 %(SD)
39	97.22(0.39)	96.55(1.06)	95.77(2.13)	97.66(1.43)	96.66(2.02)	96.11(1.75)
40	97.00(0.84)	95.77(1.27)	95.66(2.05)	97.44(1.56)	96.54(2.46)	96.23(2.15)

1, 3, and 5 according to the experimental parameter setting. This ensured reliable generalization of the independent data set. All results from each fold are averaged to obtain the final results, which in turn was compared with Tangkraingij’s results (2013).

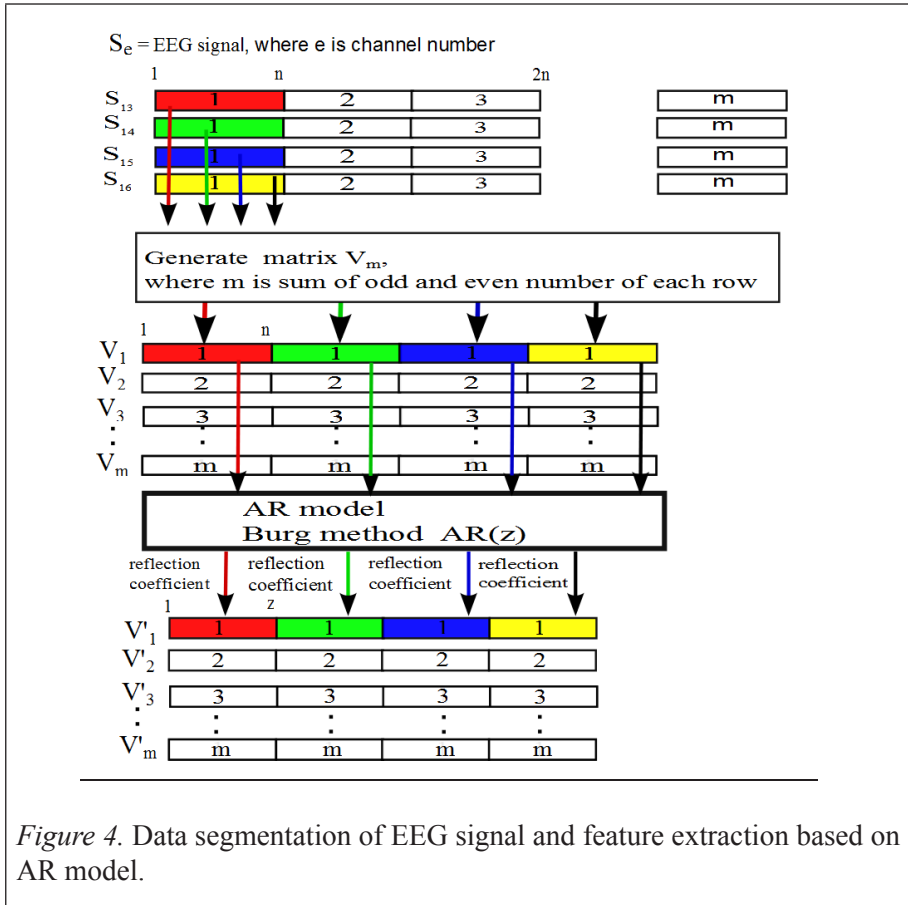


Figure 4. Data segmentation of EEG signal and feature extraction based on AR model.

EXPERIMENTAL RESULTS

The experiments for the authentication process for this work comprised of the following. Four channels were used for classifying 20 subjects. The k-NN classifier was set at 1, 3 and 5. Validation of the classifier used in this experiment was 5-fold and 10-fold cross validation. The number of data points per segment (100, 200, 300, 128, and 256) was evaluated according to the modified AR method and ICA with the AR method. The results are shown in Figure 5.

Table 1 shows the accuracy of ICA with the AR method using 5-fold and 10-fold cross validation. The highest accuracy of 5-fold and 10-fold appears in the 39th order of the AR model which were 97.22 and 97.66 percent, respectively, where k was equal to 1. It is apparent that the 10-fold cross validation reaches higher accuracy than that of the 5-fold.

Table 2

Classification Accuracy of 20 Subjects with 128 Data Points Performed on the AR Method for 5-fold and 10-fold Cross Validation with k-NN having k=1, 3 and 5

Order	5-fold cross validation			10-fold cross validation		
	k=1 %(SD)	k=3 %(SD)	k= 5 %(SD)	k=1 %(SD)	k=3 %(SD)	k= 5 %(SD)
30	95.44(1.26)	95.11(1.54)	94.77(1.39)	96.79(2.27)	96.12(2.19)	95.55(2.30)
31	95.44(1.63)	95.33(1.64)	94.88(1.38)	96.68(2.52)	96.33(1.79)	95.88(2.03)
32	95.77(1.60)	95.11(1.73)	95.11(1.49)	96.79(2.14)	96.55(2.41)	95.66(2.36)
33	95.66(1.73)	95.22(1.50)	95.22(0.92)	96.79(2.27)	96.22(2.41)	95.54(2.47)
.
.
.
39	95.44(1.94)	93.55(1.27)	95.66(0.91)	96.12(1.87)	96.12(2.06)	95.90(2.41)
40	95.77 1.60)	94.33(1.94)	95.66(1.43)	96.45(1.69)	96.01(2.43)	96.00(2.25)

Table 2 shows the accuracy of the AR method with 5-fold and 10-fold cross validation. The highest accuracy of 5-fold and 10-fold were 95.77 percent at the 32nd order and 96.79 percent at the 30th order. It can be seen that the 10-fold cross validation with k=1 still reaches higher accuracy than other values.

In Figure 6, the results of both approaches are different at the lower order. Apparently, ICA and AR seem to filter out unwanted noise better than the AR without ICA. The experimental results of Tangkraingkiy showed that four channels, F4, C4, P4, and O2 with 1,000 data points could identify 20 subjects with 98.50 percent accuracy. The comparisons are shown in Table 3.

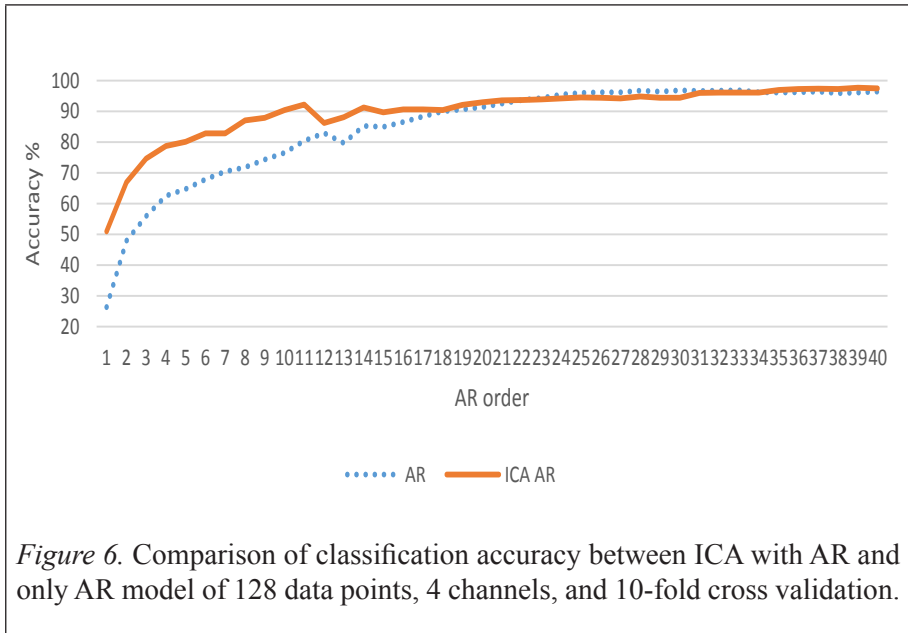


Figure 6. Comparison of classification accuracy between ICA with AR and only AR model of 128 data points, 4 channels, and 10-fold cross validation.

Table 3

Comparison of the Highest Accuracy for 4 Channels at 128 Data Points with Tangkraingkij's Results

ICA with AR (128 data points) %(SD)(th order)	AR (128 data points) %(SD)(th order)	Tangkraingkij's results (1,000 data points) %
97.66 (1.43)(39)	96.79 (2.14)(32)	98.50

The results of both proposed methods in Table 3 are slightly lower than those of Tangkraingkij. While Tangkraingkij used 1000 data points, we used only 128 data points which yielded as good as those of Tangkraingkij. For 256 data points, the results are shown in Table 4.

Table 4 shows the accuracy comparison of ICA with the AR method and only the AR method for 10-fold cross validation. The highest accuracy of each method was 99.78 and 99.29 percent, respectively, and k was equal to 1. It is apparent that the ICA with the AR method yields higher accuracy than only the AR method. As the order of AR increased, the two methods tended to become consistently accurate. The higher the order, the higher the accuracy became. The overall results are depicted in Figure 7.

Filter

Brain waves were divided into five bands of frequency resulting in different brain wave activities. Therefore, all five groups were merged and filtered in the frequency range of 0.5 to 40 Hz. This process is shown in line 8.

Data Segmentation

The aim of this process is to find a faster personal identification process by using short EEG data length. Each EEG signal from every channel of the 20 subjects is segmented into n data points, namely 100, 200, 300, 128 and 256 data points (or window size). To improve identification accuracy, each window was placed to overlap one another by 50% without differentiating patterns. Such a provision turned out to benefit less number of data segments for the experiment. In other words, the data length of 128 and 256 data points provided similar coverage of what the combined 100, 200, and 300 data points could do. Furthermore, the overlapping data length provision enhanced the number of feature vectors of classifications. This process is shown between lines 14 and 15. When $n = 128$ data points with 50 percent overlapping of the length, there were be 45 segments per channel. After processing with the AR model, the number of data points were be ordered and reduced based on the AR method. All segments with the same order from the 4 channels were concatenated resulting in 45 samples per subject. If $n = 256$ data points with the same 50 percent overlapping of the length, there would be 29 segments per channel which would result in 29 samples per subject.

Feature Extraction

This process extracts the features using the reflection coefficient of the AR model with the Burg method (Maiorana et al., 2016). The process was conducted by increasing the order of the AR model from 1 to 60 for each segment with n data points. In the AR model, the number of reflection coefficients was dependent on the order of the AR model. This process is shown between lines 17 and 20. For example, AR(10) yielded ten reflection coefficients from each segment. These reflection coefficients from the segments with the same order from 4-selected channels were concatenated to be a new sample as shown in Figure 4.

Classification

In this study, k-NN was chosen as the main classifier along with the 5-fold and 10-fold cross validation for evaluating the classification. k was set to

Table 4

Classification Accuracy of 20 Subjects with 256 Data Point Length Performed on the ICA with the AR Method and Only the AR Method for 10-fold Cross Validation with k-NN having k=1

Order	k_NN, k=1, 10-fold cross validation	
	ICA with AR %(SD)	AR % (SD)
9	97.49(1.27)	88.60(4.60)
10	98.63(1.93)	92.71(4.04)
11	<u>99.32(1.08)</u>	94.52(3.14)
12	94.08(3.92)	88.52(4.05)
.	.	.
.	.	.
.	.	.
41	99.55(0.93)	99.08(1.18)
42	99.57(0.90)	<u>99.29(1.13)</u>
43	99.57(0.90)	99.05(1.21)
44	<u>99.78(0.67)</u>	99.05(1.21)
45	99.78(0.67)	99.05(1.21)

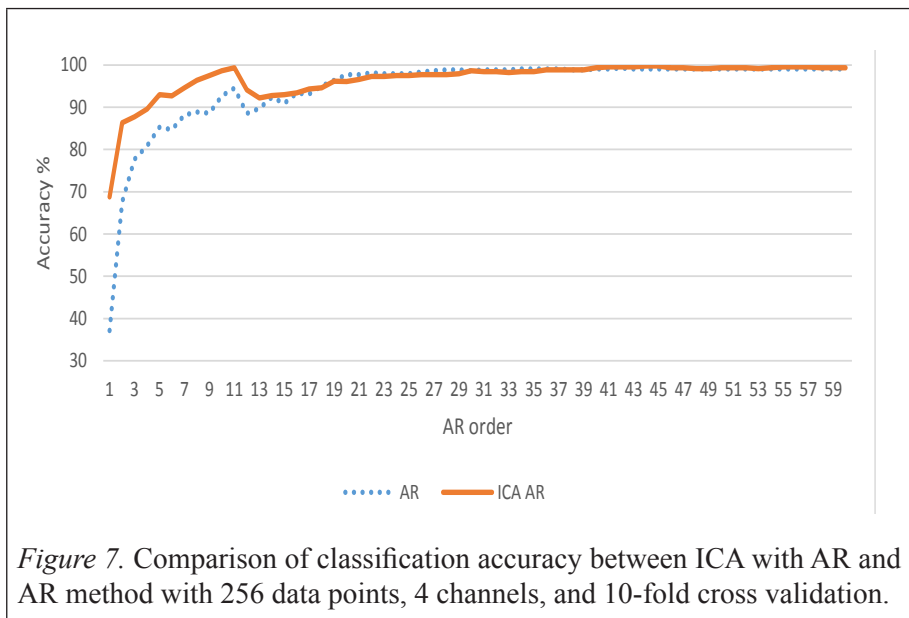


Figure 7. Comparison of classification accuracy between ICA with AR and AR method with 256 data points, 4 channels, and 10-fold cross validation.

The AR model order was in respect to the number of features used in the classification. The higher the order is, the more features were used. Having more features increased the complexity of the system but also tended to produce more accurate results. All in all, a low order system that still maintained high accuracy was ideal. It can be seen in Figure 7 that ICA with the AR method yielded excellent classification accuracy at order 11 and 44 with 99.32 and 99.78 percent, respectively. The first high accuracy can be considered as a practical implementation because the AR order of each channel can be reduced to 11. The experiment revealed that there were artifacts in EEG signals which perturbed the classification accuracy. By using the ICA method, we can removed the artifacts embedded in the EG signals, thereby resulting in higher accuracy. Table 5 represents the comparison of the highest accuracy of the three comparable methods.

Table 5

Comparison of the Highest Accuracy for 4 Channels at 256 Data Points with Tangkraingkij’s Results

ICA with AR (256 data points) %(SD)(th order)	AR (256 data points) %(SD)(th order)	Tangkraingkij’s results (1,000 data points) %
99.32(1.08)(11)	94.52(3.14)(11)	98.50
99.78(0.67)(44)	99.29(1.21)(42)	98.50

The highest accuracy of Tangkraingkij’s results was 98.50 based on ICA utilizing 1,000 data points in each of the four selected channels. Our AR only method outperformed the prior method by using only 256 data points with AR order 42, reducing from 256 to 42 features per each of the four selected channels. Furthermore, the proposed ICA with the AR method based on AR order 11, reducing from 256 to 11 features per each of the four selected channels yielded the best accuracy at 99.32 percent for AR order 11. Higher accuracy can be reached for AR order 44 at 99.78 percent. The proposed ICA with the AR method has a lower standard deviation than other methods at the same order. The lower the standard deviation is, the more reliable accuracy is obtained.

CONCLUSION AND FUTURE WORK

The study focused on applying two approaches to four channels of EEG, F4, C4, P4, and O2, to identify 20 subjects in the resting state in order to improve

higher person identification accuracy, namely ICA with the AR method and the modified AR method. The experimental results showed that ICA with the AR method yielded higher accuracy and lower standard deviation than using the modified AR method only with 5-fold and 10-fold cross validation. The experiment also revealed that the highest accuracy of 256 data points on both methods, utilized less data points than those of Tangkraingkij. Moreover, the number of features was reduced from 256 to 11 data points at 99.32% accuracy, while the highest accuracy (99.78%) used 44 features that were still less than those of Tangkraingkij's results. It was found that the accuracy of the ICA and the AR method was higher than modified AR method at low order. The results confirmed that by using ICA preprocessing and applying the AR model could clean EEG signals and enhance classification accuracy. Future improvement can explore 16- channel combinations to discover the significance of channel effects on subject classification accuracy.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Preecha Tangkraingkij at Sripatum University and Tayard Desudchit M.D. at The Chulalongkorn Comprehensive Epilepsy Program (CCEP), Faculty of Medicine, Chulalongkorn University for the EEG data used in this research.

REFERENCES

- Albera, L., Kachenoura, A., Comon, P., Karfoul, A., Wendling, F., Senhadji, L., & Merlet, I. (2012). ICA-based EEG denoising: A comparative analysis of fifteen methods. *Bull Pol Ac: Tech*, 60(3), 407-418. doi:10.2478/v10175-012-0052-3
- Belouchrani, A., Abed-Meraim, K., Cardoso, J. F., & Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2), 434-444. doi:10.1109/78.554307
- Belouchrani, A., & Cichocki, A. (2000). Robust whitening procedure in blind source separation context. *Electronics Letters*, 36(24), 2050-2051.
- Bos, R., De Waele, S., & Broersen, P. M. T. (2002). Autoregressive spectral estimation by application of the Burg algorithm to irregularly sampled data. *IEEE Transactions on Instrumentation and Measurement*, 51(6), 1289-1294. doi:10.1109/TIM.2002.808031

- Campisi, P., Scarano, G., Babiloni, F., DeVico Fallani, F., Colonnese, S., Maiorana, E., & Forastiere, L. (2011). Brain waves based user recognition using the “eyes closed resting conditions” protocol. *2011 IEEE International Workshop on Information Forensics and Security, WIFS 2011*, 00(c), 16-19. doi:10.1109/WIFS.2011.6123138
- Cardoso, J. F., & Soudoumiac, A. (1993). Blind beamforming for non-gaussian signals (Vol. 140, pp. 362-362).
- Cruces, S., Castedo, L., & Cichocki, A. (2002). Robust blind source separation algorithms using cumulants. *Neurocomputing*, 49, 87-118. doi:10.1016/S0925-2312(02)00525-8
- de Hoon, M. J. L., van der Hagen, T. H. J. J., Schoonewelle, H., & van Dam, H. (1996). Why Yule-Walker should not be used for autoregressive modelling. *Annals of Nuclear Energy*, 23(15), 1219-1228. doi:http://dx.doi.org/10.1016/0306-4549(95)00126-3
- Del Pozo-Banos, M., Alonso, J. B., Ticay-Rivas, J. R., & Travieso, C. M. (2014). Electroencephalogram subject identification: A review. *Expert Systems with Applications*, 41(15), 6537-6554. doi:10.1016/j.eswa.2014.05.013
- Harshit, R. S., Thomas, K. P., Smitha, K. G., & Vinod, A. P. (2016). Online Electroencephalogram (EEG) based biometric authentication using visual and audio stimuli. *The 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*.
- Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS ONE*, 7(2). doi:10.1371/journal.pone.0030135
- He, C., & Wang, Z. J. (2009). An independent component analysis (ICA) based approach for EEG person authentication. *3rd International Conference on Bioinformatics and Biomedical Engineering, iCBBE 2009(Dic)*, 1-4. doi:10.1109/ICBBE.2009.5162328
- Hyvärinen, a., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural networks : The Official Journal of the International Neural Network Society*, 13(4-5), 411-430. doi:10.1016/S0893-6080(00)00026-5

- Kumari, P., & Vaish, A. (2016). Feature-level fusion of mental task's brain signal for an efficient identification system. *Neural Computing and Applications*, 27(3), 659-669. doi:10.1007/s00521-015-1885-0
- Maiorana, E., Rocca, D. L., Member, S., Campisi, P., & Member, S. (2016). On the permanence of EEG signals for biometric recognition. *II*(1), 163-175. doi:10.1109/TIFS.2015.2481870
- Mu, Z., Hu, J., & Min, J. (2016). EEG-based person authentication using a fuzzy entropy-related approach with two electrodes. *Entropy*, 18(12). doi:10.3390/e18120432
- Paranjape, R. B., Mahovsky, J., Benedicenti, L., & Koles, Z. (2001). The electroencephalogram as a biometric. *Canadian Conference on Electrical and Computer Engineering 2001. Conference Proceedings (Cat. No.01TH8555)*, 2, 1363-1366. doi:10.1109/CCECE.2001.933649
- Poulos, M., Rangoussi, M., Chrissikopoulos, V., & Evangelou, A. (1999, 5-8 Sep 1999). *Parametric person identification from the EEG using computational geometry*. Proceedings of ICECS '99. The 6th IEEE International Conference on Electronics, Circuits and System.
- Riera, A., Soria-Frisch, A., Caparrini, M., Grau, C., & Ruffini, G. (2008). Unobtrusive biometric system based on electroencephalogram analysis. *Eurasip Journal on Advances in Signal Processing*, 2008. doi:10.1155/2008/143728
- Rocca, D. L., Campisi, P., & Scarano, G. (2012). EEG biometrics for individual recognition in resting state with closed eyes. *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 1-12. doi:10.1007/978-3-662-44485-6_22
- Ruiz-Blondet, M. V., Jin, Z., & Laszlo, S. (2016). CEREBRE: A novel method for very accuracy event-related potential biometric identification> *IEEE Transactions on Information Forensics and Security*, 11 (7), 1618-1629. doi: 10.1109/TIFS.2016.2543524
- Tangkraingkij, P., Lursinsap, C., Sanguansintukul, S., & Desudchit, T. (2010). Personal identification by EEG using ICA and neural network. *Computational Science and Its Applications–ICCSA 2010*, 419-430.

- Tangkraingkiij, P., Lursinsap, C., Sanguansintukul, S., & Desudchit, T. (2013). Insider and outsider person authentication with minimum number of brain wave signals by neural and homogeneous identity filtering. *Neural Computing and Applications*, 22(SUPPL.1), 463-476. doi:10.1007/s00521-012-0974-6
- Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*, 2, 1-11. doi:10.1021/pr0703501
- Yazdani, a., Roodaki, a., Rezatofghi, S. H., Misaghian, K., & Setarehdan, S. K. (2008). Fisher linear discriminant based person identification using visual evoked potentials. *2008 9th International Conference on Signal Processing*, 1677-1680. doi:10.1109/ICOSP.2008.4697459