2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Investigating the Impacts of Epilepsy on EEG-based Person Identification Systems

Dinh Phung, Dat Tran, Wanli Ma, Phuoc Nguyen and Tien Pham

***Abstract*— Person identification using electroencephalogram (EEG) as biometric has been widely used since it is capable of achieving high identification rate. Epilepsy is one of the brain disorders that involves in the EEG signal and hence it may have impact on EEG-based person identification systems. However, this issue has not been investigated. In this paper, we perform person identification on two groups of subjects, normal and epileptic to investigate the impact of epilepsy on the identification rate. Autoregressive model (AR) and Approximate entropy (ApEn) are employed to extract features from these two groups. Experimental results show that epilepsy actually have impacts depending on feature extraction method used in the system.**

I. Introduction

Person identification is the process of recognizing a person from a group. It recognizes the identity of a given person out of a closed pool of N [1]. The applications of person identification are found in video surveillance (public places, restricted areas) and information retrieval (police databases) [1]. In general, people can be identified by their biometrics such as voice, face, iris, retina, and fingerprint. It has been shown that electroencephalogram (EEG) can also be used as biometric for person identification.

A. Electroencephalogram

EEG is a measurement of the brain signals containing information generated by brain activities [2]. EEG signal is captured by using multiple electrodes either from inside the brain (invasive methods), over the cortex under the skull, or certain locations over the scalp (non-invasive methods) [2].

EEG signal carries genetic information; that is, there is a connection between genetic information and EEG of an individual [3]. Moreover, EEG features are universal as all living and functional persons have recordable EEG signal [4]. Therefore, EEG data can be suitably used for person identification [1], [3], [5], [6]. The use of brain wave patterns obtained from EEG data as a new modality for person identification has several advantages: 1) It is unique as individual's brain wave patterns are unique [7], [1]; and 2) It is universal as all living and functional persons have recordable EEG signals [4].

EEG signals are divided into five major bands, delta ($0.5 - 3$ Hz), theta ($4 - 7$ Hz), alpha ($8 - 13$ Hz), beta ($14 - 30$ Hz), and gamma (> 30 Hz) [2]. Delta waves are mainly associated with deep sleep and may also be observed in a waking state while theta waves are associated with creative inspiration and deep meditation. Alpha waves are the most popular in the brain activities. They appear in both relaxed awareness without attention and with concentration. Beta waves are the usual waking rhythms in the brain associated with active thinking, active attention, or solving problems. Gamma waves usually have low amplitudes, rare occurrence, and relate to left index finger, right toes, and tongue movement [2].

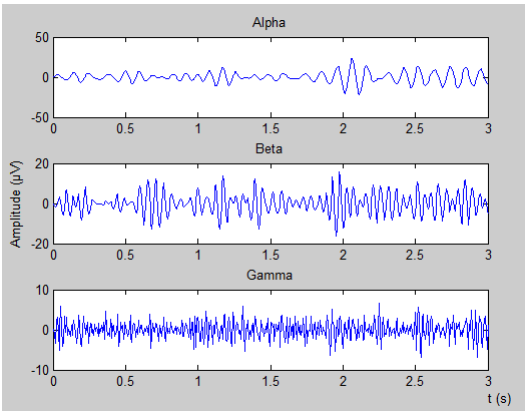

Fig. 1. Wavebands example

It is unlikely that the entire EEG could be more representative of brain dynamics than the individual frequency subbands. In fact, the subbands may yield more accurate information about constituent neuronal activities underlying the EEG and consequently, certain changes in the EEGs that are not evident in the original full-spectrum EEG may be amplified when each subband is analyzed separately [8].

B. Epilepsy

Epilepsy is chronic neurological disorder which is generally characterized the sudden and the recurrent seizures [9]. Epileptic seizures are manifestations of epilepsy, which are caused by the sudden development of synchronous neuronal firing in the cerebral cortex and are recorded using the EEG. They may be partial seizures, which occur only in a few channels of the EEG recording, or generalized seizures the whole brain, which involve in every channel of the EEG recording [9]. The shape of wave may contain useful information relating to the different psychological states of the brain [6]. Therefore, EEG signal parameters, extracted and analyzed using computers, are useful in diagnosing and assessing brain state, especially epilepsy [6]. So far, most

Dinh Phung, Dat Tran, Wanli Ma, Phuoc Nguyen and Tien Pham are of the Faculty of Education, Science, Technology & Mathematics, Canberra, Australia (email: {dinh.phung, dat.tran, wanli.ma, phuoc.nguyen, tien.pham}@canberra.edu.au).

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

studies have focused on detecting epilepsy such as in [6], [9], [10], [11], [12], [13], and there have been no reports on the impacts of epilepsy on the performance of person identification.

Accuracy is one of the crucial requirements of any person identification system, including EEG-based. Factors which may affect the accuracy of an EEG-based identification system can be signal noises, feature extraction methods, and/or classification algorithms. Mental disorder (such as epilepsy) may also have some effects on the performance as it has been shown that, brain state changes from less to more ordered state, from more to less chaotic, or from more to less complexity during epileptic seizures. We investigate this issue by employing feature extraction methods on EEG signals to obtain different features for EEG-based person identification systems. Two typical feature extraction methods chosen are Autoregressive model (AR) and Approximate entropy (ApEn). AR is to find a set of model parameters that best describe the signal generation system, therefore the AR features may not contain epileptic information. ApEn is an entropy-based feature extraction method that has shown its ability in reflecting the changes in the chaotic level and degree of complexity in time series as well [2], [9], [10]. Therefore ApEn features contain epileptic information. If epilepsy has impacts on EEG-based person identification systems, the identification rate of an AR-based system should be different from that of an ApEn-based system. We also chose two different EEG datasets. The first dataset consists of EEG data collected from normal subjects, and EEG data from epileptic subjects are included in the second dataset. We expect the change of AR-based person identification rates should be different from that of ApEn-based person identification rates measured on the same epileptic dataset since the epileptic information is included in ApEn features only.

II. Autoregressive Model

Autoregressive model (AR) is a predicting method to find a set of model parameters that best describe the signal generation system. The AR model, with the order p, is defined to be linearly related with respect to a number of its previous samples [2], i.e.

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + y(n) \quad (1)$$

where $x(n)$ is the data of the signal at the sampled point n, $a_k, k = 1, 2, ..., p$ are the AR coefficients, and $y(n)$ is the noise input.

AR modeling is an alternate for EEG spectral estimation. However, the AR model is only applicable to stationary signals. In contrast, EEG signal is complex, non-linear, non-stationary, and random in nature [2], [3], [6], [9], [10], [14], [15], [16], and they are considered stationary only within short intervals, i.e. "quasi-stationary" [17], [18]. Therefore, EEG signal is normally segmented into short intervals prior to AR modeling process.

AR model has a broad spectrum of applications ranging from identification, prediction and control of dynamical systems [19]. The AR modeling has been a popular feature extraction for EEG-based person identification as seen in [3], [20], [21].

III. Approximate entropy

Entropy is a measure of uncertainty. In brain-computer interface systems, entropy can be used to measure the level of chaos of the system [2]. It is a non-linear measure quantifying the degree of complexity in a time series [9]. The advantage of using entropy methods for EEG feature extraction is that EEG signals are complex, non-linear, non-stationary, and random in nature [2], [3], [6], [9], [10], [14], [15], [16]. Entropy is one of several approaches for non-linear analysis has been proposed for EEG feature extraction as randomness of non-linear time series data is well embodied by calculating entropies of the time series data [22]. Entropy reflects how well one can predict the behavior of each respective part of the trajectory from the other. Basically, higher entropy indicates more complex or chaotic systems, thus, less predictability [6].

Approximate entropy (ApEn) is a statistic measure that can estimate the complexity of signals from the discrete-time sequences, especially for real-time applications [9]. This measure can quantify the complexity or irregularity of the system. ApEn is less sensitive to noise and can be used for short-length data. In addition, it is resistant to short strong transient interferences such as spikes [9]. ApEn gives a robust entropy estimate from short and noisy data sets and increasing values correspond to more irregularity [6].

So far, ApEn has been popularly used for automatic epilepsy detection [6], [9], [11], [12], [23] as it has robust characteristics in the characterization of the epileptic patterns and low computational burden [12], [23]. In particular, it has been shown that there are significant differences between epileptic and normal EEG, and the degree of complexity for epileptic EEG signal is lower than that of normal EEG signal [9]. Epileptic EEG is more regular and less complex than the normal, and entropies of epileptic activity are less as compared to that of non-epileptic activity [6]. With these findings and the ApEn's robustness in epileptic pattern characterization, we assume that feature extraction by ApEn would be able to show the effects of epilepsy on the performance of person identification system.

Let $X = [x(1), x(2), ..., x(N)]$ be the data sequence containing N data points and let $x(i)$ be the subsequences of X such that $x(i) = [x(i), x(i+1), ..., x(i+m-1)]$, for $1 \leq i \leq N - m$ where m is the number of samples used for prediction.

The distance between any two of the above vectors, $x(i)$ and $x(j)$ is defined as: $d[x(i), x(j)] = max_{k=0,...,m-1}|x(i+k) - x(j+k)|$

Consider $M^m(i)$ is the number of $x(i)$ that satisfies $d[x(i), x(j)] < r$, where $r \in R+$ and is a filtering level.

$$C_r^m(i) = \frac{M^m(i)}{N-m+1}; i = 1, ..., N-m+1 \quad (2)$$

Define:

$$\phi_r^m = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} log(C_i^m(r)) \quad (3)$$

Approximate entropy is defined as:

$$ApEn(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \quad (4)$$

ApEn indicates that the system is regular (not complex) if two sequences similar for m points will remain similar at next m+1 points, and vice versa [9].

IV. Datasets

Our experiment was conducted on the Australian EEG (AEEG) dataset [24]. The dataset was collected in the John Hunter Hospital, New South Wales, Australia, over a period of 11 years. The recordings were made by using 23 electrodes (23 channels) placed on the scalp of a subject with the sampling rate of 167 Hz for about 20 minutes. The subset of the data used for our experiments consists of the EEG data of 80 subjects, being divided into two groups, namely normal and epileptic. The dataset is summarized in Table I. The normal group is chosen as a baseline so that we can compare the differences in the identification rates between the two groups.

Group	Number of subjects	Number of channels	Number of trials	Number of sssions	Trial length (seconds)
Normal	40	23	1	1	1200
Epileptic	40	23	1	1	1200

TABLE I
AEEG DATASET DESCRIPTIONS

V. Feature extraction

We extract EEG features based on AR and ApEn methods. The feature extraction is described in Fig. 2 and Fig. 3, in which EEG signal is firstly filtered into three wave bands, i.e, alpha (8-13 Hz), beta (14-26 Hz), gamma (30-45 Hz). Features are then extracted and classified separately for each sub-band. In particular, the filtered signal (about 1200 second length) is then fragmented into one-second [9] segments. Next, AR and ApEn features from every segment of each channel are extracted. The optimal order of AR was 16 (AR16). The selected parameters for ApEn in the Eq. 4 are $m = 4$, $r = 0.5 * std$, and $N = 167$ equaling to one-second data. Consequently, 16 features (by AR16) or one feature (by ApEn) from each sub-trial are extracted for one channel. After that, all the features from 23 channels are joined together to form a feature vector for each sub-trial [17]. In brief, there are about 1200 vectors of 368 or 23 features regarding to the feature extraction method, AR16 or ApEn, respectively.

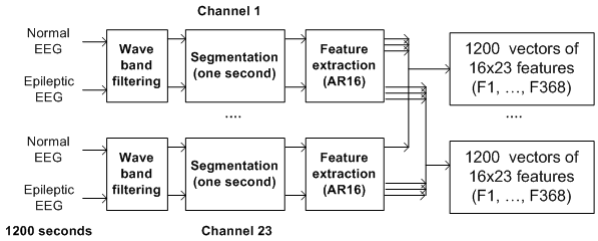

Fig. 2. AR16 Feature extraction

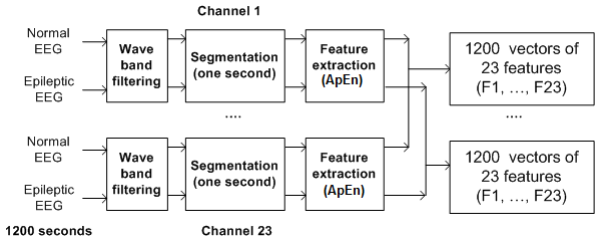

Fig. 3. ApEn Feature extraction

VI. Classification

The extracted features are used to train Linear Support Vector Machine (SVM) classifiers for person identification as described in [17], [18]. One advantage of SVM is the hyperplane selection that maximizes the margins, which is known to increase the generalization capabilities in classification. SVM also uses a regularization parameter C that enables accommodation to outliers and allows errors on the training set [17]. Originally, SVM was designed for binary classification; therefore, it cannot deal with multi-class classification directly [25]. The Multi-class SVM is described in [18]. The binary SVM classifiers can be combined to handle the multi-class case. In test phase, the voting strategy was used as follow: each binary classification was considered to be a voting where votes could be cast for all data points x. The final result was the class with maximum number of votes.

During the classification, two-third of the datasets are used for cross-validation training, and one-third are used for testing. Linear SVM classifiers are trained in 3-folds with parameter C ranging from 1 to 1000 in 5 steps [18]. The selection of parameter C is conducted by using a Weka's meta-classifier named CVParameterSelection. After finding the best parameters of C, the meta-classifier then uses them to build SVM's models on the training data. In the test phase, the testing data will be evaluated against the trained models for person identification.

VII. Experimental Results

Experimental results are presented in Tables II and III, and in two Figures 4 and 5.

Our task is to compare the change of identification rates between the AR and ApEn feature extraction methods from the normal dataset to the epileptic one. We are not aiming

to compare the identification rates between the two datasets on the same feature extraction method, however we can have a remark on why the epileptic group has slightly better accuracy than the normal one for the AR16 features as seen in Table II and Fig. 4. The maximum difference is 3.2% being observed in the gamma band in the test phase. Although it is not significant to see whether random difference between the two datasets or epilepsy has contributed to the performance of the classification results, the later case would be appropriate as [2], [9] state that epileptic EEG signals are less chaotic and complex than normal EEG signals. As a result, this may make epileptic EEG signals more suitable with linear methods such as Autoregressive model, thus, it probably helps to increase the accuracy of EEG-based person identification.

Band	AR16	
	NORMAL	EPILEPTIC
Gamma	91.9	95.1
Beta	91.6	93.2
Alpha	89.1	91.9

TABLE II

PERSON IDENTIFICATION RATES OF AR16

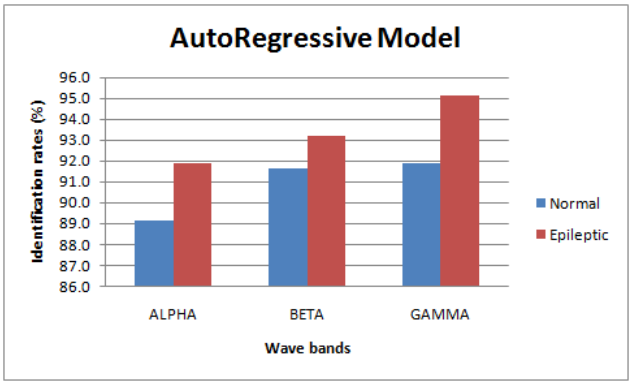

Fig. 4. Person identification rates of AR16

In contrast, epileptic group tends to have lower accuracy than the normal group for ApEn features. As demonstrated in the Table III and Fig. 5, the accuracy of epileptic group has a maximum 8.5%, recorded in ApEn's beta band, lower than the normal group. This would support the hypothesis that epilepsy has impact on the performance of EEG-based person identification systems. As stated above, epileptic EEG signals are less chaotic and complex than normal EEG signals [2], [9]. In addition, entropy estimation is based on chaotic levels, i.e, the higher entropy value reflects the more chaotic signals, and vice versa [2]. As a result, the less chaotic property of the epileptic EEG signals makes the lower entropy values, and this has been proved in [9]. Therefore, it causes the lower inter-class variation which results in the poorer identification rates of epileptic group.

Overall we can see a change on the person identification rates for the epileptic dataset. The AR-based person identification rates are higher than the ApEn-based ones.

Band	ApEn	
	NORMAL	EPILEPTIC
Gamma	64.2	62.5
Beta	70.8	62.3
Alpha	48.4	45.4

TABLE III

PERSON IDENTIFICATION RATES OF APEN

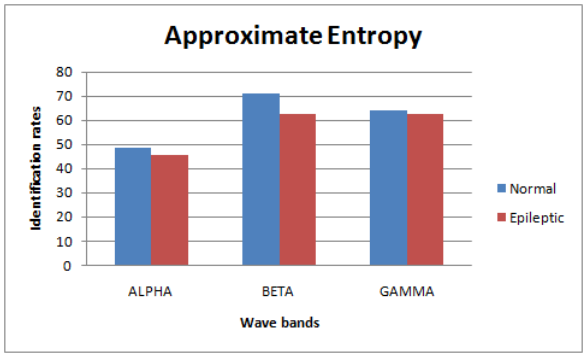

Fig. 5. Person identification rates of ApEn

VIII. CONCLUSIONS

We have found that epilepsy does have impacts on the accuracy of EEG-based person identification systems. The identification rates decrease for the epileptic dataset if the EEG features extracted from this dataset contains epileptic information. extraction method. This implies that a person identification system will not provide high performance on the users that have mental disorder such as epilepsy. For further investigation, we will conduct experiments on other feature extraction methods as well as on a larger scale of datasets for confirming the influences of epilepsy on person identification.

REFERENCES

[1] S. Marcel and J.l R. Millán. Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation. Idiap-RR Idiap-RR-81-2005, IDIAP, 0 2005.

[2] S. Sanei and J. A. Chambers. *EEG Signal Processing*. Wiley-Interscience, 2007.

[3] G. Mohammadi, P. Shoushtari, B. M. Ardekani, and M. B. Shamsollahi. Person identification by using ar model for eeg signals. *World Academy of Science, Engineering and Technology*, 1(11):918 – 923, 2007.

[4] H.A. Shedeed. A new method for person identification in a biometric security system based on brain eeg signal processing. In *WICT*, pages 1205–1210, 2011.

[5] M. K. Abdullah, K. S. Subari, J. L. C. Loong, and N. N. Ahmad. Analysis of the eeg signal for a practical biometric system. *Biomedical Engineering World Academy of Science, Engineering and Technology*, 4(8):944 – 949, 2010.

[6] N. Kannathal, M. L. Choo, U. R. Acharya, and P. K. Sadasivan. Entropies for detection of epilepsy in eeg. *Comput. Methods Prog. Biomed.*, 80(3):187–194, 2005.

[7] T. Pham, W. Ma, D. Tran, P. Nguyen, and D. Phung. A study on the feasibility of using eeg signals for authentication purpose. In *ICONIP*, pages 562–569, 2013.

[8] Hojjat Adeli, S. Ghosh-Dastidar, and N. Dadmehr. A wavelet-chaos methodology for analysis of eegs and eeg subbands to detect seizure and epilepsy. *Biomedical Engineering, IEEE Transactions on*, 54(2):205–211, 2007.

[9] Y. Kumar and M. L. Dewal. Complexity measures for normal and epileptic eeg signals using apen, sampen and sen. *IJCCT*, 2(7):6–12, 2011.

[10] U. R. Acharya; F. Molinari; S. V. Sree; S. Chattopadhyay; Kwan-Hoong Ng; J. S. Suri. Automated diagnosis of epileptic eeg using entropies. *Biomedical Signal Processing and Control*, 7:401–408, 2012.

[11] V. Srinivasan, C. Eswaran, and N. Sriraam. Approximate entropy-based epileptic eeg detection using artificial neural networks. *Information Technology in Biomedicine, IEEE Transactions on*, 11(3):288–295, 2007.

[12] S. Vukkadala, S. Vijayalakshmi, and S. Vijayapriya. Automated detection of epileptic eeg using approximate entropy in elman networks. *International Journal of Recent Trends in Engineering*, 1(1):307–312, 2009.

[13] A.S. Zandi, M. Javidan, G.A. Dumont, and R. Tafreshi. Automated real-time epileptic seizure detection in scalp eeg recordings using an algorithm based on wavelet packet transform. *Biomedical Engineering, IEEE Transactions on*, 57(7):1639–1651, 2010.

[14] W. Chaovalitwongse, O. Prokopyev, and P. Pardalos. Electroencephalogram (eeg) time series classification: Applications in epilepsy. *Annals OR*, 148(1):227–250, 2006.

[15] A. Hope and R. Rosipal. Measuring Depth of Anesthesia using Electroencephalogram Entropy Rates. Technical report, Department of Theoretical Methods, Slovak Academy of Sciences, Slovak Republic, 2001.

[16] G. Litscher. Electroencephalogram-entropy and acupuncture. *Anesthesia and Analgesia*, 102(6):1745–1751, 2006.

[17] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based BCIs. *Journal of neural engineering*, 4(2), 2007.

[18] P. Nguyen, D. Tran, X. Huang, and D. Sharma. A proposed feature extraction method for eeg-based person identification. *Int. Conf. on Artificial Intelligence*, 2012.

[19] R. Palaniappan, P. Raveendran, S. Nishida, and N. Saiwaki. Autoregressive spectral analysis and model order selection criteria for eeg signals. In *TENCON 2000. Proceedings*, volume 2, pages 126–129 vol.2, 2000.

[20] R.B. Paranjape, J. Mahovsky, L. Benedicenti, and Z. Koles'. The electroencephalogram as a biometric. In *Electrical and Computer Engineering, 2001. Canadian Conference on*, volume 2, pages 1363–1366 vol.2, 2001.

[21] N. Alex M. Poulos, M. Rangoussi and A. Evangelou. Person identification from the eeg using nonlinear signal classification. In *Methods of Info. in Medicine*, pages 41–64, 2001.

[22] Y. Song and P. Liò. A new approach for epileptic seizure detection: sample entropy based feature extraction and extreme learning machine. *Biomedical Science and Engineering*, 3:556–567, 2010.

[23] Satyanarayana Vollala and Karnakar Gulla. Automatic detection of epilepsy eeg using neural networks. *International Journal of Internet Computing (IJIC)*, 1(3):68–72, 2012.

[24] M. Hunter, R. L. L. Smith, W. Hyslop, O. A. Rosso, R. Gerlach, J. A. P. Rostas, D. B. Williams, and F. Henskens. The Australian EEG Database. *Clinical EEG and Neuroscience*, 36(2):76–81, April 2005.

[25] Z. Y. He and L. W. Jin. Activity recognition from acceleration data using AR model representation and SVM. *Machine Learning and Cybernetics*, January 2008.