

Epilepsy Analytic System with Cloud Computing

Chia-Ping Shen, *Student Member, IEEE*, Weizhi Zhou, Feng-Seng Lin, Hsiao-Ya Sung, Yan-Yu Lam, Wei Chen, Jeng-Wei Lin, Ming-Kai Pan, Ming-Jang Chiu, *Senior Member, IEEE*, and Feipei Lai, *Senior Member, IEEE*

Abstract - Biomedical data analytic system has played an important role in doing the clinical diagnosis for several decades. Today, it is an emerging research area of analyzing these big data to make decision support for physicians. This paper presents a parallelized web-based tool with cloud computing service architecture to analyze the epilepsy. There are many modern analytic functions which are wavelet transform, genetic algorithm (GA), and support vector machine (SVM) cascaded in the system. To demonstrate the effectiveness of the system, it has been verified by two kinds of electroencephalography (EEG) data, which are short term EEG and long term EEG. The results reveal that our approach achieves the total classification accuracy higher than 90%. In addition, the entire training time accelerate about 4.66 times and prediction time is also meet requirements in real time.

I. INTRODUCTION

ELECTROCARDIOGRAM (ECG), Phonocardiogram (PCG), and Electroencephalography (EEG) always supply important health information of the patients for physicians. The analysis of these biomedical data [1-3] is very important for clinical research field with signal processing and data mining. Techniques such as adaptive filtering, spectrum estimation, compression, time series processing, feature selection and pattern classification are used to analyze the signals. As a result of improving the quality of care and reducing the cost, physicians decided to use these intelligent algorithms to provide a decision making system. Therefore, biomedical data analytic system recently is developed as an intelligent system to capture, transmit, calculate, and distribute results to physicians and patients. In other words, the system has been designed as a solution to make diagnosis and to ensure the goal of automatic systems can be operated in both a real-time and an interactive environment.

C. P. Shen, W. Zhou, Y. Y. Lam, W. Chen, and F. Lai are with the Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan, R. O. C.

F. S. Lin and H. Y. Sung is with the Department of Computer Science and Information Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan, R. O. C.

M. J. Chiu and M. K. Pan is with the Department of Neurology, National Taiwan University Hospital, No.1, Changde St., Zhongzheng Dist., Taipei City 10048, Taiwan, R. O. C.

J. W. Lin is with the Department of Information Management, Tunghai University, No.181, Sec. 3, Taichung Port Rd., Xitun Dist., Taichung City 40704, Taiwan, R. O. C.

*Correspondence to: Mr. C. P. Shen, phone: 886-952707038; e-mail: b89505018@ntu.edu.tw

Electroencephalogram (EEG) signals record cerebral electric activities and detect events of epileptic seizures in patients with epilepsy, which afflicts approximately 1% of the population [4]. In general, seizures happen when clusters of brain neurons signal abnormally, which may temporarily cause anomaly in consciousness, behavior, movements, and actions of the patients. One way to analyze epilepsy is through recognizing EEG waveform of patients. EEG devices measure voltage changes from ionic current flows within the neurons of the brain [5], which shows temporal and spatial information of the brain, and is useful in the diagnosis of epilepsy. In addition, study [6] has also shown that EEG signals provide high sensitivity and specificity for the diagnosis of epilepsy. Recently, several studies [7-9] have proposed all kinds of algorithms for epileptic seizures using EEG signals. If successful, it would improve the quality of life and safety for patients with epilepsy. Thus, real-time EEG forewarning systems are emerging and necessary for seizure detection.

The web-based epilepsy analytic system (EAS) architecture we proposed contains three major portions: the client site, the server site and the database as depicted in Fig. 1. The client site, which is accessed by physicians or healthcare practitioners, provides a friendly graphical user interface to interact with the server site and the database. Users from the client site must get authentication from the server site via session services to validate the security information stored in database. At the server site, it is embedded all the functionalities under the web services, such as data preprocessing feature selection, feature extraction, and classifier. The database stores medical data collected from hospitals the National Taiwan University Hospital (NTUH) in our approach. All components of the system use the extensible markup language (XML) format for exchanging messages, and the communication mechanism is based on a simple object access protocol (SOAP) over HTTP handled internally by the .NET environment [10-11]. In conclusion, users can easily get the results from EAS.

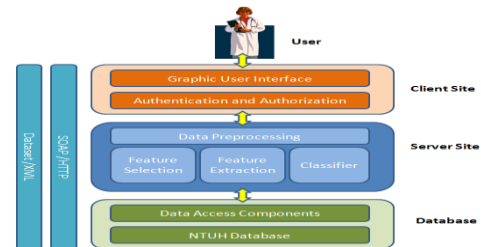


Fig. 1 The system architecture of EAS

In the following sections of the paper, we first elaborate the design of the overall methods of biomedical data analytic system. Detailed descriptions of system components, data preprocessing, features selection, and classifiers are illustrated. In Section III, the system implementation is described. The complete biomedical data analytic results are provided in Section IV. Finally, the paper concludes in Section V.

II. METHOD

A. Data Preprocessing

Infinite impulse response (IIR) is a property of signal processing systems [12]. IIR systems have an impulse response function that is non-zero over an infinite length of time. This is in contrast to finite impulse response (FIR) filters, which have fixed-duration impulse responses. The simplest analog IIR filter is an RC filter made up of a single resistor feeding into a node shared with a single capacitor. This filter has an exponential impulse response characterized by an RC time constant. Because the exponential function is asymptotic to a limit, and thus never settles at a fixed value, the response is considered infinite.

B. Wavelet Transform

Among all the available signal decomposition tools, the wavelet transform (WT), which simultaneously extracts the time and frequency characteristics of a signal, is most suitable for extracting the spike-and-wave features [13]. The multi resolution property of the WT allows the decomposition of a signal into a number of scales, each scale representing a particular compositional part of the signals. A WT reduces the original signals into a few parameters, while maintaining the major characteristics for differentiating the type of EEG records. In addition, the inherent properties of WT, which was proven as an efficient tool for biomedical signal processing [14], include good time and frequency location and across-sub-band similarity. With these properties, WT is suitable for inconsistent frequency characteristics in different time frames.

C. Feature Selection

Genetic algorithm (GA) is one of the artificial intelligence methods, which designed to simulate the nature process of biological evolution to solve problems. GA, presented by Prof. John Holland [15], evolved through the processes of natural selection, recombination and mutation. GA simulates the nature cell, its main component elements are genes, chromosomes, group, and fitness function. It evolves through the best genetic chromosome found in the current generation of chromosomes; the next generation of chromosomes could be more adapted to the environment.

To start from a population of randomly generated individuals represented in binary as strings of 0s and 1s. In each generation, the fitness of each individual in the population is evaluated and selected into the best one. In addition, the current populations can mutate, recombine and mate with each other to generate the next genetic generation. Then, the algorithm will check that if it achieves the termination condition or not. If this generation does not achieve the termination condition, otherwise, it will re-select the best individual which is used in the next iteration, as shown in Fig. 2.

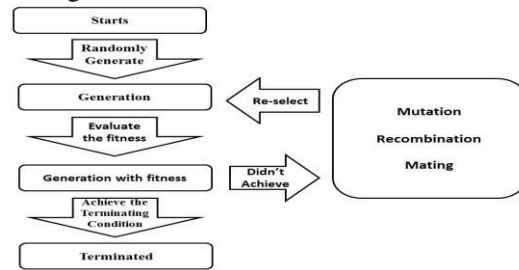


Fig. 2 Genetic Algorithm Flow Diagram

D. Support Vector Machines

The support vector machines (SVM) [16] map input feature vectors into a high dimensional space to realize a linear classification system. By feeding the algorithm with a set of training data, SVM can determine an optimal hyper-plane that minimizes the risks [17]. Note that it may not be useful to achieve high training accuracy. Therefore, a common way is to separate training data by mapping instances into high dimensional domain to build models. After data are mapped into a higher dimensional space, the number of variables becomes very large or even infinite. A typical approach to handle this difficulty is solving the dual problems shown in (1).

$$\begin{aligned}
 \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) - \sum_{i=1}^l \alpha_i \\
 \text{subject to} \quad & 0 \leq \alpha_i \leq c, \quad i = 1, \dots, l \\
 & \sum_{i=1}^l y_i \alpha_i = 0
 \end{aligned}
 \tag{1}$$

E. System Processing

As described, a multi-class SVM classification with GA feature selection requests many standard binary SVM invocations for different parameters [18]. These SVM invocations combined with genetic algorithm selection are independent, and we can execute them in parallel. Hence, the system explores the parallelism of these SVM invocations to speed up the whole task. With the concept of cloud computing architecture, several types of modules are implemented, each of which is a standalone service to complete a designed subtask. A backend server can execute a mixture of different modules on demand. To simplify the design of this

architecture, currently, communications among modules are going through different task lists on the database. The central control unit would activate these modules which will query the job list from the database. These modules would execute the job in order from the list and return the results until the job list is empty. The database will store the results and the system server would do the statistic of the results when the job is finished.

The following is the system architecture and the descriptions of major modules in Fig. 3.

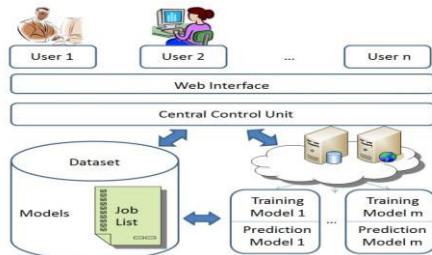


Fig. 3 The System Processing of EAS

III. SYSTEM IMPLANTATION

A. Data Acquisition

We collected two sets of clinical data from subjects receiving routine short term EEG examinations and long-term EEG monitoring in the Department of Neurology, NTUH. The first EEG data set was acquired from short-term EEG recordings of routine EEG examinations. The subjects were divided into two groups. In short term EEG, we included EEG data from 13 participants (5 women and 8 men), whose ages ranged from 20 to 89 years. The second EEG data set was acquired from five of those patients of epileptic. In long term EEG, we included EEG data from 5 participants (2 women and 3 men), whose ages ranged from 34 to 81 years. Thus, in this paper, we used the long term EEG as a validation set and tested only the recordings from the five epileptic patients using the model of short term EEG.

These EEG records are transformed from analog signals to digital signals by instrument manufacture. All the EEG data were preprocessed by eliminating artifacts, and the sample rate is about 200Hz, record channels are 16 channels. In addition, the EEG signals were digitally filtered by IIR, and extracted 1700 features [1] after 4 levels wavelet transform.

B. System Implementation

1) Home page

Home page is a major part of EAS that it is designed to allow users to train EEG data and predict the results. Users can adjust the required parameters and methods of SVM and GA on this page. After prediction, the system will show the results of the accuracy. In this page, we also provide the document of guide to introduce how to use the system.

2) EEG Monitoring Page

Fig. 4 shows the interface of EEG monitoring page which acquires EEG data by AJAX from database, and illustrates figures by javascript. Users can review the classification of EAS and annotation to determine if the classification is correct or not. Once the review is finished, the accumulation of feedback will retrain the new EEG classification model. The functions are explained as below:

EAS can display both unipolar and bipolar EEG signals. Channel names are shown in area (1). Area (2) plots the selected EEG record, and the pink area displays the annotated location of (4). Area (3) is the list of patients' records that has been analyzed in EAS. While users confirm the selected event results had been classified to wrong classes, users can click the option buttons to give a new annotation in the area (5). The modified annotations will be included in the next model training. Finally, users can navigate 10 seconds backward and forward of EEG charts by clicking the buttons in area (6). In summary, the EAS is an automatic EEG reading system that helps doctors to make clinical decision in real time.

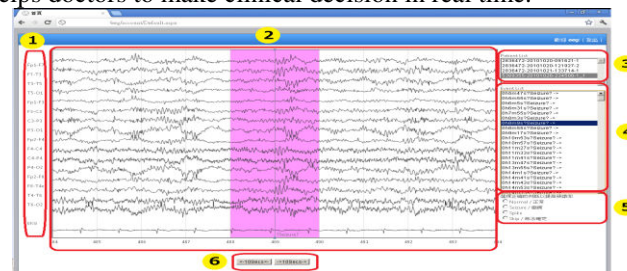


Fig. 4 Interface of EEG Monitoring Page

IV. RESULTS AND DISCUSSION

The EAS is written in C# language running on .NET platform. The functions provided from the LIBSVM 2.6 C# edition are modified as the classification web services. In addition, other various optional functions are also offered, such as genetic algorithm. In order to fulfill the flexibility, the system is implemented with modular characteristic which allows easy incorporation of new methods. There are two types of EEG data tested in the following statement.

1) Short Term EEG Data

The annotated short term EEG records we obtained includes 1939 2-second epochs of normal activity, 436 2-second epochs of spike activity, and 444 2-second epochs of seizure activity. Spikes and seizure do not occur frequently out of the norm. However, these two waveforms are of upmost detection priority and therefore more samples are needed for a strong prediction model. Half of the 2-second epochs are taken for training, and the other half are taken for prediction. In total, our system currently can output 1700 features for each 2-second epoch. We use these 1700 features together to obtain the results from the EAS (Table 1).

2) Long Term EEG Data

In order to validate that our EEG classification framework

can be generalized to data from different acquisition machines, we trained the classifiers using the entire short term EEG and tested the trained classifiers using long term EEG as a hold-out data set. The accuracies of short term EEG data were 83.02% (inter-ictal detection rate) and 100% (seizure detection rate). The testing results of using the hold-out long term EEG data are also given in Table I. The inter-ictal detection rate is 91.18% and the seizure detection rate is 99.22% which are close to the results of short term EEG data. It means that the features are useful and stable in classifying epileptic and nonepileptic patients. In addition, Table I also presents the comparisons with other literatures that both the accuracies of inter-ictal and seizure are higher than the others.

TABLE I RECOGNITION RATE OF EEG DATA

	Normal	Inter-ictal	Seizure
Short Term EEG Data	93.7%	83.0%	100%
Long Term EEG Data	92.8%	91.1%	99.2%
[19]	N/A	76 %	N/A
[20]	N/A	N/A	92.2%

3) System Performance

The EAS was built up on HiCloud 15 servers, which specifications are as follows: CPU: 1.0 GHz, RAM: 8GB, HDD: 100GB. Although GA-based feature selection is effective, however, it required 18.55 hours on one server. The time was reduced dramatically to 3.96 hours when fifteen generic servers were used as backend servers in the proposed framework. Fig. 5 shows the required time when different numbers of backend servers were used. As the number of backend servers increased, the overhead on the database and mutual communication between modules also increased. When more than 5 backend servers were used, the speedup was slow down. It means that the time consuming of overhead overtake the saving time of paralleling computing piece by piece. In addition, we also tested that the entire 10-second EEG perditions of data preprocessing, feature extraction to classifier needs 0.68 second to meet requirements in real time.

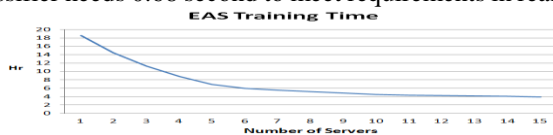


Fig. 5 The Training Time of EAS with Different Number of Servers

V. CONCLUSION

Our approach proposed a cloud computing system (EAS) for neurologists to read EEG records and help them to make clinical decisions. The EAS was tested using real data from NTUH and obtained good preliminary results. It reveals that the overall classification accuracy of the EAS is higher than 90%. In addition, the speedup of the entire training time is about 4.66 times, and prediction time also meets the real time requirements.

REFERENCES

- [1] C. P. Shen, W. H. Chen, J. M. Chen, K. P. Hsu, J. W. Lin, M. J. Chiu, and F. Lai, "Bio-signal Analysis System Design with Support Vector Machines based on Cloud Computing Service Architecture" in *Proc. IEEE Engineering in Medicine and Biology Society (EMBC)*, Argentina, August 2010, pp. 1421-1424.
- [2] W. C. Kao, W. H. Chen, C. K. Yu, C. M. Hong, and S. Y. Lin, "Portable real-time homecare system design with digital camera platform," *IEEE Trans. Consumer Electronics*, vol. 51, no. 4, pp. 1035-1041, 2005.
- [3] C. P. Shen, W. C. Kao, Y. Y. Yang, M. C. Hsu, Y. T. Wu, and F. Lai, "Detection of cardiac arrhythmia in electrocardiograms using adaptive feature extraction and modified support vector machines," *Expert Systems With Applications*, vol. 39, no. 9, pp. 556-561, July 2012.
- [4] H. M. d. Boer, M. Mula, and J. W. Sander, "The global burden and stigma of epilepsy," *Epilepsy & Behavior*, vol. 12, no. 4, pp. 540-546, February 2008.
- [5] D. E. Cragar, D. T. R. Berry, T. A. Fakhoury, J. E. Cibula, and F. A. Schmitt, "A review of diagnostic techniques in the differential diagnosis of epileptic and nonepileptic seizures," *Neuropsychology Review*, vol. 12, no. 1, pp. 31-64, March 2002.
- [6] W. Weng and K. Khorasani, "An adaptive structure neural network with application to EEG automatic seizure detection," *Neural Network*, vol. 9, pp. 1223-1240, August 1996.
- [7] E.D. Übeyli and I. Güler, "Features extracted by eigenvector methods for detecting variability of EEG signals," *Pattern Recognition Letter*, vol. 28, no.5, pp. 592-603, November 2007.
- [8] S. Ghosh-Dastidar and H. Adeli, "Principle component analysis enhanced cosine radial basis function neural network for robust epilepsy and seizure detection," *IEEE Trans. Biomedical Engineering*, vol. 55 no. 2, pp. 512-518, February 2008.
- [9] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in EEGs using time frequency analysis," *IEEE Trans. Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703-710, September 2009.
- [10] S. H. Hsieh, S. L. Hsieh, C. P. Shen, W. H. Chen, P. H. Cheng, K. P. Hsu, C. H. Chen, and F. Lai, "Web Services Based Bio-signal System Leveraging Support Vector Machines," *9th IEEE International Conference on Bioinformatics and BioEngineering (BIBE)*, Taichung, June 2009, pp. 254-259.
- [11] S. H. Hsieh, P. H. Cheng, C. H. Chen, K. H. Huang, and, P. H. Chen, "A newborn screening system based on service-oriented architecture embedded support vector machine," *J Med Syst*, vol. 34, no.4, pp. 727-733, August 2010.
- [12] http://en.wikipedia.org/wiki/Infinite_impulse_response
- [13] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Information Theory*, vol. 36, no. 5, pp. 961-1005, 1990.
- [14] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet presentation," *IEEE Trans. Intelligence*, vol. 11 no. 7, pp. 674-693, 1989.
- [15] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. 1992: MIT Press.
- [16] Cortes, C. and V. Vapnik, *Support-vector network*, 1995.
- [17] C. C. Chang and C. J. Lin, LIBSVM: a library for support vector machines, 2001. Software is available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [18] C. P. Shen, C. H. Liu, F. S. Lin, H. Lin, C. Y. F. Huang, C. Y. Kao, F. Lai, J. W. Lin, "A Multiclass Classification Tool Using Cloud Computing Architecture," *International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics*, Turkey, June 2012.
- [19] M. Lucia, J. Fritschy, P. Dayan, and D. Holder, "A novel method for automated classification of epileptiform activity in the human electroencephalogram-based on independent component analysis," *Med. Biol. Eng. Comput.*, vol. 46, pp. 263-272, 2008.
- [20] R. Yadav, M. Swamy, and R. Agarwal, "Model-Based Seizure Detection for Intracranial EEG Recordings," *IEEE Trans. Biomed. Eng.*, vol. 59, pp. 1419-1428, 2012.