34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Channel Selection for Epilepsy Seizure Prediction Method Based on Machine Learning

Nai-Fu Chang, Tung-Chien Chen, Cheng-Yi Chiang and Liang-Gee Chen

Abstract— **The studies on seizure prediction problem have shown great improvement these years. Machine learning based seizure prediction method shows great performance by doing pattern recognition on high-dimensional bivariate synchronization features. However, the computation loading of the machine learning based method may be too high to meet wearable or implantable devices with the power and area constraints. In this work, channel selection is proposed to reduce the channel number from 22 to less than 6 channels and therefore more than 93.73% of the computation loading is saved through the method. The best result shows successful rate of 60.6% in 3-channel cases of ECoG database and successful rate of 70% in 3-channel cases of EEG database.**

I. INTRODUCTION

Epilepsy is the world's second most common brain disorder with over 40 million people worldwide suffering from it [1]. For medically intractable Epileptics, it is the sudden, unforeseen way in which seizures occur that represents one of the most disabling aspects of the disease. Apart from the risk of serious injury, there is often an intense feeling of helplessness that has a strong impact on the everyday life of a patient [2]. A method capable of predicting the occurrence of seizures could significantly improve the therapeutic possibilities and thereby the quality of life for epilepsy patients [3].

A seizure prediction method based on machine learning is proposed by Mirowski and et al. [4] to predict seizures by doing pattern recognition on high-dimensional bivariate synchronization features and achieve outstanding sensitivity and low false alarm rates. The breakthrough of this method is that machine learning enables classification of high-dimensional feature vectors which aggregate into patterns. In contrast, the traditional method restricts feature to a low-dimensional vector or a scalar value. However, large dimension of the feature patterns may cause large time complexity owing to many filtering operation in continuous wavelet transform (CWT) and complicated arithmetic operations in coherence computation. For example, the computation capability of a laptop with Intel i7 core can only afford 14-channel feature extraction solely in real time. Besides, large-dimension feature patterns increase the computation loading of feature classification. It may not be feasible for the wearable or implantable devices which require low power consumption for real application.

Thanks Prof. Horng-Huei Liou, Prof. Sheng-Hong Tseng and Dr. Chih-Chuan Chen for annotating the EEG recording of patient in National Taiwan University Hospital. Thanks Dr. Ming-Kai Pan for annotating the MIT-CHB Scalp EEG database.

N.-F. Chang, C.-Y. Chiang, T.-C. Chen, and L.-G Chen are with the DSP/IC Design Lab, Graduate Institute of Electronics Engineering , National Taiwan University, Taipei, Taiwan. (e-mail: knife@video.ee.ntu.edu.tw).

TABLE I
PATIENT CHARACTERISTICS, TRAINING & TESTING SETS IN EEG DATA

Patient	1	3	6	7	10	22	NTUH	Total
Gender	F	F	F	F	M	F	M	
Seizure type	TLE	TLE	TLE	TLE	TLE	FLE	TLE	
Off-line training	2	4	3	2	1	1	3	16
Testing	5	3	4	1	1	1	5	20
Total	7	7	7	3	2	2	8	36

Gender: male (M), female (F). Seizure type: temporal lobe epilepsy (TLE), frontal lobe epilepsy (FLE).

In this paper, channel selection is proposed to reduce the feature pattern size and thus reduce the computation loading of feature extraction and feature classification. The remainder of this paper is organized as follows. Section II introduces the Electrocorticogram (ECoG) and Electroencephalogram (EEG) database, the machine learning based seizure prediction algorithm and channel selection method. The simulation results and computation analysis are described in Section III. Finally, Section IV concludes this work.

II. METHODS

A. ECoG and EEG Database

The study involves 21 Freiburg intracranial EEG recordings and 7 continuous long-term EEG recordings from CHB-MIT Scalp EEG database and National Taiwan University Hospital (NTUH). The Freiburg database [5] is a publicly available intracranial EEG database provided by the Epilepsy Center of the University Hospital of Freiburg, Germany. The database contains 6-channel ECoG recordings of 21 patients three focal and three extrafocal whose sampling rate is 256Hz. The CHB-MIT scalp EEG database [6], sampled at 256Hz, is composed of long-term scalp EEG recordings of 24 patients, and most of them have more than 22 channels. We select 6 out of 24 patients since some of the recordings are not suitable for seizure prediction problem which requires a long interictal between two seizure onsets and recording continuity without long disruption. There's also one continuous long-term 18-channel EEG recording of a patient at NTUH with the sampling rate of 200Hz. The patient characteristics, training set and testing set of these patients of EEG database are shown in Table I.

B. Machine Learning Based Seizure Prediction Algorithm

1) Feature Extraction: Sliding, non-overlapping 5 s windows for a given channel pair are used to compute wavelet coherence values. The frequency-specific phase of EEG signal of each channel is extracted by CWT at each band with gaussian mother wavelet [7]. Then, wavelet coherence [8]

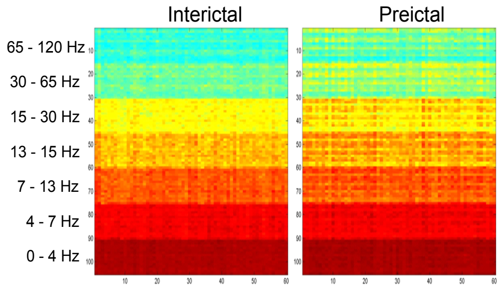

Fig. 1. The example of an interictal pattern and a preictal pattern of patient 4 in Freiburg database. The number of rows in the pattern is equal to the number of pairs times the number of frequency bands while the number of columns in the pattern corresponds to the duration of one pattern divided by 5 s. There is merely small difference between two patterns in the lower two bands. The similar phenomenon can be observed on other patterns.

for all channel pairs and frequency bands are computed to measure the phase synchronization.

2) Feature Aggregation: These synchrony features are further aggregated into patterns, matrixes with certain size. The number of rows in the pattern is equal to the number of pairs times the number of frequency bands. The number of columns in the pattern corresponds to the duration of one pattern divided by 5 s. For example, if 7 bands, 5 min and 6 channels are set, each pattern is a matrix with dimension 105 (7 bands times 15 channel pairs) by 60 (5 min / 5 s), a 6300-dimension vector in vector form.

By inspection we found that generally the feature values of two lower bands remain relatively constant, as shown in Fig. 1. Seeing that, we select only five higher bands: alpha (7-13 Hz), low beta (13-15 Hz), high beta (15-30 Hz), low gamma (30-65 Hz), and high gamma (65-120 Hz) instead of seven bands, therefore reduce the feature size by 28.5%.

3) Feature Classification: After feature extraction, the Support Vector Machine (SVM) provided by libsvm [9] is used to classify the patterns into preictal and interictal states. For classification on high-dimensional feature vectors, the performance of Linear SVM is comparable with SVM using kernel of higher order. Therefore, we choose Linear SVM with lower time complexity.

4) Post-processing: A two-in-a-row post-processing technique is adopted to reduce false alarms. An alarm would be generated only if there are two consecutive patterns classified as preictal.

Fig. 2 shows the overall block diagram of the machine learning based seizure prediction algorithm. We also adopt an on-line retraining method, which is different from the traditional off-line method to serve a fixed portion of the data as training set and the rest as testing set. In on-line retraining method, the size of training set gradually increases as more and more data become available. For the detailed description of on-line retraining method, please refer to [10].

C. Channel Selection

The comparison between feature pattern size and number of channels is shown in Fig. 3. As stated in Section II-B.2, the pattern size is proportional to the number of pairs, which relates to the number of channels of input data. Therefore, channel selection could be used to reduce the feature pattern

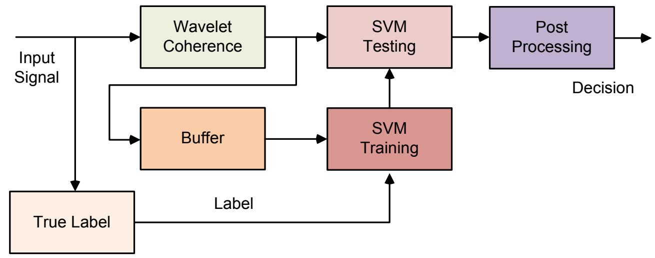

Fig. 2. The block diagram of the seizure prediction algorithm with on-line retraining method. The size of training set gradually increases as more and more data become available.

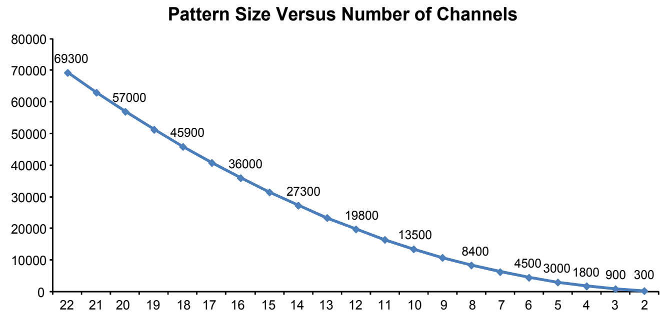

Fig. 3. The comparison between feature pattern size and number of channels is shown. The pattern size is proportional to the number of pairs, which relates to the number of channels of input data.

size and lower the computation loading of feature extraction and feature classification. Besides, bi- and multi-variate measures have a better performance on the seizure prediction problem, but the observed pre-ictal changes were found to be locally restricted to specific channels rather than occurring as a global phenomenon [11]. Therefore, channel selection may further improve sensitivity and decrease occurrences of false alarms of the seizure prediction method.

In ECoG database, performances of all channel pair combinations with number of channels from 2 to 6 are investigated. In EEG database, we choose seventy-five combinations of fixed channel pairs. Three to six channels around focal channels and extrafocal channels are selected for investigation. We also calculate the results of adaptive channel pairs, which aggregates the best results of seven patient for different numbers of channels. Adaptive channel pairs are investigated to test the effects of sensing many channels while only few qualified channels are used as the input data of the seizure prediction algorithm.

III. RESULTS AND DISCUSSION

A. Experiment Setup and Performance Evaluation Criterions

The ECoG and EEG input data are filtered by a band-pass filter of 1-100Hz, and then a notch filter is applied to offset power-line frequency (50 Hz or 60 Hz).

We set the prediction horizon as two hours. Only the patients whose results satisfy the following two conditions would be counted as successful patients. First, the false positive rate of the patient must be less than 0.2 per hour. Second, that at least one alarm is generated in the preictal period would be regarded as a successful prediction. Successful rate is equal to successful predictions divided by total number of testing seizures.

TABLE IV

RESULT OF CHANNEL SELECTION ON EEG TESTING SET

	N_p	S_p	S_r		N_p	S_p	S_r		N_p	S_p	S_r		N_p	S_p	S_r
All 22 ch	12	5	60%												
Fixed 6 ch				Fixed 5 ch				Fixed 4 ch				Fixed 3 ch			
1,2,3,4,6,7	11	4	55%	2,3,14,15,12	12	4	60%	2, 14, 8, 12	11	4	55%	2, 12, 14	10	3	50%
13,14,15,16,10,11	9	3	45%	2,3,14,15,8	10	3	50%	2, 14, 7, 11	12	6	60%	2, 3, 14	9	2	45%
2,3,6,7,14,15	10	3	50%	2,3,6,14,15	8	3	40%	2, 14, 6, 10	11	4	55%	2, 14, 15	8	3	40%
10,11,14,15,2,3	14	5	70%	2,3,10,14,15	10	3	50%	2, 14, 5, 9	9	3	45%	2, 3, 12	10	4	50%
2,3,6,7,12,16	13	4	65%	2,3,6,7,14	11	4	55%	2, 3, 14, 15	8	3	40%	8, 14, 15	9	3	45%
2,3,6,7,9,13	7	3	35%	2,3,6,7,12	14	4	70%	2, 3, 14, 12	11	4	55%	2, 3, 7	10	4	50%
1,2,3,4,14,15	10	4	50%	2,3,6,7,9	11	5	55%	2, 3, 13, 16	9	3	45%	11, 14, 15	9	2	45%
1,2,3,4,12,16	13	4	65%	10,11,14,15,2	12	5	60%	2, 3, 12, 16	11	3	55%	8, 11, 12	10	3	50%
1,2,3,4,9,13	8	4	40%	10,11,14,15,8	11	4	55%	2, 3, 9, 13	10	4	50%	5, 9, 10	10	2	50%
13,14,15,16,2,3	9	3	45%	10,11,14,15,5	8	2	40%	14, 15, 2, 8	8	3	40%	2, 8, 14	12	5	60%
13,14,15,16,4,8	5	2	25%	2,3,6,14,12	9	2	45%	14, 15, 1, 4	6	2	30%	3, 12, 15	8	3	40%
13,14,15,16,1,5	9	3	45%	2,3,6,12,16	12	3	60%	14, 15, 4, 8	8	2	40%	3, 8, 15	10	4	50%
10,11,14,15,4,8	13	5	65%	2,3,6,9,13	12	5	60%	14, 15, 1, 5	10	4	50%				
10,11,14,15,1,5	7	3	35%	10,14,15,2,8	11	4	55%	2, 3, 6, 7	11	5	55%				
2,3,6,10,14,15	9	4	45%	10,14,15,4,8	8	3	40%	10, 11, 14, 15	12	4	60%				
2,3,7,11,14,15	10	5	50%	10,14,15,1,5	8	2	40%	2, 3, 6, 14	9	3	45%				
2,3,6,14,15,12	12	3	60%	1,2,3,4,6	10	3	50%	2, 3, 6, 12	9	3	45%				
11,14,15,2,3,8	8	3	40%	13,14,15,16,9	9	3	45%	2, 3, 6, 9	8	4	40%				
2,3,14,15,8,12	13	5	65%	2,3,6,7,4	10	4	50%	10, 14, 15, 2	12	4	60%				
2,3,14,15,5,9	8	3	40%	9,10,14,15,16	11	4	55%	10, 14, 15, 8	11	4	55%				
2,3,4,14,15,16	11	3	55%	6,7,10,11,12	11	3	55%	10, 14, 15, 5	9	4	45%				
Adaptive 6ch	17	6	85%	Adaptive 5ch	17	6	85%	Adaptive 4ch	17	6	85%	Adaptive 3ch	14	5	70%

N_p: number of successful predictions, S_p: successful patients, S_r: successful rate, channel(ch). Channel position: 1 (FP_1-F_7), 2: F_7-T_7, 3: T_7-P_7, 4: P_7-O_1, 5: FP_1-F_3, 6: F_3-C_3, 7: C_3-P_3, 8: P_3-O_1, 9: FP_2-F_4, 10: F_4-C_4, 11: C_4-P_4, 12: P_4-O_2, 13: FP_2-F_8, 14: F_8-T_8, 15: T_8-P_8, 16: P_8-O_2.

TABLE II

RESULT OF CHANNEL SELECTION ON ECOG TESTING SET

	# of Successful Patients			Successful Rate		
Channel	Mean	Max	Min	Mean	Max	Min
All 6ch	16.0/21			69.7%		
All 5ch	14.8/21	16/21	14/21	62.4%	65.2%	57.6%
3f2r	15.0/21	16/21	14/21	62.1%	65.2%	57.6%
2f3r	14.7/21	16/21	14/21	62.6%	63.6%	62.1%
All 4ch	14.3/21	16/21	13/21	60.0%	66.7%	53.0%
3f1r	13.7/21	14/21	13/21	56.1%	59.1%	53.0%
2f2r	14.7/21	16/21	13/21	61.4%	66.7%	54.5%
1f3r	13.7/21	14/21	13/21	59.6%	62.1%	56.1%
All 3ch	12.2/21	15/21	10/21	50.0%	60.6%	40.9%
3f	13.0/21	13/21	13/21	50.0%	50.0%	50.0%
3r	13.0/21	13/21	13/21	48.5%	48.5%	48.5%
2f1r	12.6/21	15/21	10/21	53.5%	60.6%	45.5%
1f2r	11.6/21	13/21	10/21	46.6%	54.5%	40.9%
All 2ch	5.3/21	8/21	2/21	20.7%	34.8%	7.6%
2f	5.0/21	6/21	4/21	19.2%	25.8%	13.6%
1f1r	5.3/21	8/21	3/21	20.7%	28.8%	9.1%
2r	5.3/21	8/21	2/21	22.2%	34.8%	7.6%

channel (ch), focal channel (f), referenced or extrafocal channel (r).

TABLE III

RESULT OF DIFFERENT NUMBERS OF CHANNELS AND PATIENT SETS ON ECOG TESTING SET

	All patients			Original successful patients		
# of patients	21			16		
Total seizures for testing	66			53		
# of channels used	S_p	N_p	S_r	S_p	N_p	S_r
6	16.0	46.0	69.7%	16.0	46.0	86.8%
5	14.8	41.2	62.4%	14.5	40.8	77.0%
4	14.3	39.6	60.0%	14.1	39.4	74.3%
3	12.2	33.0	50.0%	12.0	32.8	61.9%
2	5.3	13.7	20.7%	4.7	13.1	24.8%

S_p: successful patients, N_p: number of successful predictions, S_r: successful rate.

B. Channel Selection Result on ECoG Database

The result on channel selection on ECoG database is shown in Table II. There are several combinations of different number of channels. To see the impact of channel positions, the channels are classified into focal and extrafocal channels and the average, maximum and minimum results of different channel pairs are investigated. The six-channel case has the best result, and the number of successful patients and the successful rate are sixteen and 69.7%, respectively. Generally, number of successful patients and successful rate both decrease as number of channels decreases. Still, some cases with few channels perform well. For example, the best result out of 9 combinations in 4-channel cases with two focal and two extrafocal shows sixteen successful patients and successful rate of 66.7% while the average result also shows 14.7 successful patients and successful rate of 61.4%. The best result out of 9 combinations in 3-channel cases with two focal and one extrafocal shows fifteen successful patients and successful rate of 60.6% while the average result also shows 12.6 successful patients and successful rate of 53.5%.

Over ten patients could be successfully predicted in all combinations of 3-, 4- and 5-channel cases, while the average results of 3-, 4-, 5-channel cases show 12.2, 14.3 and 14.8 successful patients. About 5.3 patients could be successfully predicted when only 2-channel data are used.

The comparison between sensitivity and number of channels used on ECoG testing set is shown in Table III. We classified the results into two group, all patients and originally successful patients to further test the impact of number of channels. Some original unsuccessful patients could become successful with less channels since some unqualified channels may be omitted. Therefore, the average numbers of successful predictions and successful patients in

TABLE V

COMPUTATION ESTIMATION OF DIFFERENT NUMBERS OF CHANNELS

	Computation					Equivalent Workin Frequency (MHz)					Ratio (%)				
	22ch	6ch	5ch	4ch	3ch	22ch	6ch	5ch	4ch	3ch	22ch	6ch	5ch	4ch	3ch
Filtering	1.8M /5 s	0.49M /5 s	0.408M /5 s	0.327M /5 s	0.245M 5 s	0.360	0.098	0.082	0.065	0.049	0.1%	0.4%	0.5%	0.6%	0.9%
Wavelet Transform	68.43M /5 s	18.66M /5 s	15.55M /5 s	12.44M /5 s	9.33M /5 s	13.69	3.73	3.11	2.49	1.87	3.6%	15.8%	19.2%	24.6%	34.6%
Coherence	292.7M /5 s	19M /5 s	12.67M /5 s	7.60M /5 s	3.8M /5 s	58.54	3.8	2.53	1.52	0.76	15.5%	16.0%	15.7%	15.0%	14.1%
SVM Training	549G /30 min	28.9G /30 min	18.8G /30 min	10.9G /30 min	4.9G /30 min	305	16.06	10.44	6.06	2.72	80.8%	67.8%	64.6%	59.8%	50.4%
SVM Testing	3.24M /300 s	0.212M /300 s	0.141M /300 s	0.085M /300 s	0.042M /300 s	1.1 E-02	7.1 E-04	4.7 E-04	2.8 E-04	1.4 E-04	0.0%	0.0%	0.0%	0.0%	0.0%
	Total Working Frequency					377.6	23.69	16.17	10.13	5.40					

the case of all patients are larger than the case of originally successful patients. Still, the performance of the seizure prediction method drops with less channels used.

C. Channel Selection Result on EEG Database

Table IV shows the result on channel selection of EEG database. The number of successful patients and the successful rate in 22-channel case are five and 60%, respectively. Seventy-five combinations of fixed three to six channel positions are selected as the input data to test the predictability of the seizure prediction method. The result is encouraging because the method shows certain predictability even with less number of channels. The case of two channels around left and right temporal lobes, T_7 and T_8, and an extrafocal channel, O_1 remains the same result as 22-channel case. Moreover, some cases exhibit better results with 70% successful rate. The method fails to predict patient 6 of CHB-MIT database in all cases.

From the result of adaptive channel pairs, the seizure onsets of six patients could be successfully predicted with number of channel from four to six. The successful rate can be further increased to 85% by adaptive channel selection. Five (patient 1, 7, 10 22 from CHB-MIT database and the patient from NTUH) patients exhibit good prediction result when some selected three to six channels are used, while patient 3 from CHB-MIT database only could be predicted with number of channels larger than three. Therefore, the system could be further improved by initially sensing many channels and finding the best channel pairs, and then only few qualified channels are used in the patients' normal lives.

D. Computation Analysis on Channel Selection

We calculate the amount of computation in each step of the algorithm flow and estimate the corresponding working frequency with different numbers of channels. The estimation is shown in the Table V. Although band selection already reduces 28.5% pattern size, the pattern size of 22-channel case is so large such that the computation of wavelet coherence and SVM training requires over 360M and 549G instructions, respectively. Although SVM training is only executed every thirty minutes, the computation is over 305M and the highest loading in 22-channel case. Even a laptop with advanced i7 processor can not meet the throughput requirements of feature extraction and feature classification in 22-channel case. With less number of channels, the time complexity of feature extraction and the pattern size for feature classification are largely decreased. In 3-, 4-, 5- and 6-channel cases, the overall algorithm requires much less computation costs with only 6.27%, 4.28%, 2.68% and 1.43% instructions of 22-channel case. This comparison greatly demonstrates the importance of channel selection to reduce the computation loading of machine learning based seizure prediction method.

IV. CONCLUSIONS

In this paper, channel selection is applied to reduce the feature pattern size and therefore the computation loading of both feature extraction and feature classification. Channel selection maintains certain or better predictability even with few channels used while the feature pattern size of the machine learning based seizure prediction method are largely reduced. The computation costs are reduced to less than 6.27% instructions of 22-channel case.

REFERENCES

[1] N. Sivasankari and et al., "An extensive review of significant researches on epileptic seizure detection and prediction using electroencephalographic signals," *Advances in Biomedical Research*, 2010.

[2] R. G. Andrzejak and et al., "Seizure prediction: Any better than chance?", *Clinical Neurophysiology*, vol. 120, pp. 1465-1478, 2009.

[3] Elger CE, "Future trends in epileptology," *Curr Opin Neurol*, vol. 14, pp. 185-186, 2001.

[4] P. Mirowski and et al., "Classification of patterns of EEG synchronization for seizure prediction," *Clinical Neurophysiology*, vol. 120, no. 11, pp. 1927-1940, Nov. 2009.

[5] Freiburg invasive EEG database, Epilepsy Center of the University Hospital of Freiburg, available at https://epi-lepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database/.

[6] A. L. Goldberger and et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, pp. e215-e220, Jun. 2000.

[7] M. Le Van Quyen and et al., "Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony," *J. Neurosci Method*, vol. 11, pp. 83-98, 2001.

[8] A. Grinsted, J. C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," *Nonlinear Processes in Geophysics*, vol. 11, pp. 561-566, 2004.

[9] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001, Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[10] C.-Y. Chiang and et al., "Seizure Prediction Based on Classification of EEG Synchronization Patterns with On-line Retraining and Post-Processing Scheme," *EMBC*, 2011, pp. 7564-7569.

[11] F. Mormann and et al., "Seizure prediction: the long and winding road," *Brain*, vol. 130, pp. 314-333, 2007.