Seizure Prediction on EEG Using Common Spatial Pattern and Convolutional Neural Network

Yuan Zhang, *Senior Member, IEEE,* Yao Guo, *Student Member, IEEE,* Po Yang, *Senior Member, IEEE,* Wei Chen, *Senior Member, IEEE,* Benny Lo, *Senior Member, IEEE,* and Wenyao Xu, *Senior Member, IEEE*

***Abstract*—Seizure prediction can warn patients to take effective measurement in time before the oncoming onset, which is critical to protecting their lives. Compared to seizure detection that identifies the inter-ictal state and the ictal state, there are far fewer researches on seizure prediction because high similarity makes it challenging to distinguish between the pre-ictal state and the inter-ictal state. In this paper, a novel algorithm to predict seizure is proposed using common spatial patterns (CSP) and convolutional neural network (CNN). Firstly, we generate additional artificial EEG signals by combining segmented pre-ictal signals to solve the trial imbalance problem between the two states. Secondly, an extractor we designed for feature extraction in both time domain and frequency domain employing wavelet packet decomposition and CSP to decrease training time while increasing overall accuracy. Finally, a shallow CNN uses the extracted feature matrix as input to facilitate discriminating between the pre-ictal state and the inter-ictal state. Our proposed algorithm is evaluated on 23 patients from Boston Children's Hospital-MIT scalp EEG dataset employing a leave-one-out cross-validation approach and achieves a sensitivity of 92.2% and false prediction rate (FPR) of 0.11/h. Experimental results verify the particularly good performance compared to other state-of-the-art methods.**

***Index Terms*—seizure prediction, EEG, common spatial patterns, convolutional neural network**

I. Introduction

EPILEPSY is a common chronic brain disease with approximately 50 million patients worldwide, the premature death rate of whom is 2 to 3 times that of disease-free individuals, and it poses a heavy burden on the patients, their families and society [1], [2]. Accordingly, the study of seizure prediction has always played an important role in the field of biomedicine to offer hope of orthobiosis and proper functioning to those patients who cannot be cured by surgery [3]. However, accurate and generalized seizure prediction algorithm is critically difficult to develop due to two important factors [4], [5]. Firstly, EEG signal itself is highly complex and varies irregularly over time. Secondly, the pre-ictal and the inter-ictal EEG states across individuals vary entirely. For these reasons, an automated patient-specific seizure prediction approach which can reduce the serious consequences of seizure by obtaining the oncoming onset alarm is both highly challenging and desired.

In the early stages, classification between the ictal state and the inter-ictal state for seizure detection was most popular [6]–[9]. Automatic seizure detection can quickly and accurately estimate an approximate period of seizure, greatly reducing the time for doctors to view EEG records. In one of our previous works, we utilize extended correlation-based feature selection and logistic model trees to classify the ictal state, the inter-ictal state and the normal state, with the highest accuracy of 97.6% [10]. Although it can detect the seizures occurred, it is powerless to obtain the information on upcoming epilepsy which does not make sense for clinical emergency treatment. Unfortunately, only limited attention has been paid to seizure prediction due to lack of the unified assessment criteria, until Maiwald unequivocally defined the term for prediction methods and predict the oncoming onset using dynamical similarity index and threshold crossing [11]. Thereafter, more research on seizure prediction has emerged on different dataset [12]–[14].

Nowadays, machine learning is the most advanced technique to predict seizure. Among them, dynamical similarity index, mean phase coherence, phase locking value, zero-crossings are effective algorithms to extract features, and Gaussian mixture models, Adaboost, SVM and convolutional neural network (CNN) are utilized widely to obtain the output of prediction. However these methods cannot achieve high sensitivity and low false prediction rate (FPR) simultaneously. In addition, most methods ignore the data imbalance problem that the pre-ictal signals are far less than the inter-ictal signals. Therefore, we adopt data augmentation to balance data and utilize common spatial patterns (CSP) together with CNN to forecast the oncoming seizure. As an effective spatial filtering algorithm CSP can search the component signal which best transduces the cerebral activity in seizure prediction. CNN has already been successfully used including but not limited to face recognition, natural language processing and emotion recognition, with few development in EEG data analysis. Moreover, to the best of our knowledge, the combined model of CSP and CNN is for the first time applied to EEG data classification.

The main contributions of our work are as follows: (1) We design a novel algorithm comprising of two consecutive components. The first part is an extractor to extract feature vectors obtained by CSP from raw EEG signals and their

Yuan Zhang and Yao Guo are with Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan, 250022, China.(e-mail: yzhang@ujn.edu.cn)

Po Yang is with the Faculty of Engineering and Technology, Liverpool John Moores University, L69 7ZX, UK.

Wei Chen is with the Center of Intelligent Medical Electronics at School of Information Science and Technology, Fudan University, Shanghai, 200433, China.

Benny Lo is with the Hamlyn Centre, Imperial College London, London,SW7 2AZ, UK.

Wenyao Xu is with Department of Computer Science and Engineering, State University of New York at Buffalo, New York, 14260-1660, USA.

several frequency band. Then, a shallow CNN with 2 blocks is constructed to predict the oncoming seizure. The designed extractor can extract temporal, spatial and frequency characteristics of each class from multi-channel EEG signals, while the CNN can be trained for seizure prediction. (2) We implement signal segmentation and recombination in the time domain to augment the amount of pre-ictal EEG signals, to prevent the model from suffering from the trial imbalance problem. (3) The balanced data are applied to our architecture. The average prediction accuracy reaches 92.2% meanwhile the average FPR is 0.11/h, outperforming most state-of-the-art seizure prediction methods in recent literature.

The remainder of the paper is organized as follows. Section II introduces typical works using machine learning. Section III provides the details of our proposed method. In Section IV, the results of this method are presented. Section V presents a discussion on the results, and comparisons with related work. Finally, the paper is concluded in Section VI.

II. Related work

Studies on seizure prediction can be divided into two categories according to the type of classifier employed. In the first category, one or two kinetic indicators and threshold crossing are used to forecast an impending seizure onset. The seizure is expected to come when the indicator is apparently above or below the calculated threshold [15], [16]. Once the increasing or decreasing tendency in the value over time appears, an alarm is triggered to warn an approaching onset. Among them, Iasemidis studies the T-index of the largest exponent [17], found that the largest Lyapunov curve for the pivotal channels in temporal lobe present a dynamical change before the onset. In the second category, the entire EEG signals from patients are divided into segments (several seconds in general as a trial), and then labeled as inter-ictal, pre-ictal, and ictal [18]. Among them, ictal signals have no contribution to seizure prediction and are discarded prior to classification. In this approach, training a binary machine learning classifier to effectively distinguish between the two states is the most important component for success. An unsolvable disadvantage of the first approach over the second one, is that no single or two features simultaneously have remarkable changes for all-patients when the seizure come. The feature extraction and classification is the most important procedure for the second approach, and suitable choice and design can produce superior performance [19].

Zandi proposed a novel method that applied zero-crossing interval histogram and variational Bayesian Gaussian mixture model to predict the oncoming onset of 20 patients from the Vancouver General Hospital database [20]. Turky used CSP to extract a feature set which was fed into LDA classifier to distinguish between the pre-ictal EEG segments and the inter-ictal EEG segments [21]. Mayer used phase/amplitude lock values (PLV/ALV) to calculate the phase and amplitude difference between EEG electrodes local and remote to the epileptic event [22]. Dongrae used EEG signals of 21 patients from CHB-MIT and applied phase locking value to the gamma frequency band decomposed by EMD, NEMD and NA-MEMD algorithms [23]. Recent research efforts have focused on developing a method to extract features that can effectively predict seizures for patients.

With the rapid advancement of deep learning, CNN becomes the most headline-grabbing method for seizure prediction. Truong utilized the short-time fourier transform (STFT) on 30s EEG windows without overlap to extract time-frequency information as an input of classifier [24]. After standardization, a CNN structure with 3 convolution layers, each layer including a batch normalization unit, a convolution unit and a max pooling unit, is trained to separate the pre-ictal trials from the inter-ictal trials. They used 13 patients from the same database to test the proposed methodology. The average seizure prediction sensitivity reached 81.2% with a FPR being 0.16/h. Khan designed a CNN architecture with 6 convolutional layers to extract features which have the ability to differentiate pre-ictal from inter-ictal EEG segments [25]. The detail coefficients obtained by the wavelet transform of each EEG channel at assorted scale was fed as input to CNN. They used 15 patients from the CHB-MIT dataset to test the proposed methodology and achieved an average FPR of 0.142/h.

We note that all previous researches ignore that the length of the pre-ictal signals is much less than the inter-ictal signals to prediction seizure. We balance the data firstly by generating extra pre-ictal states and discard some inter-ictal states. Moreover, CSP as a feature extraction method with superior performance in the brain computer interface filed lacks of attention in seizure prediction. Only [21] used CSP as an extractor for binary classification, nevertheless the results for binary classification were unsatisfactory. We observe that extracting features in different frequency and temporal bands simultaneously is an ideal method for improving the classification precision. In addition, due to EEG data with multi-channel signals, as a special two-dimensional signal, whose number of channel and samples extreme dissimilarity. Raw signals after simple pre-processing such as cutting or flipping is still unfeasible for direct use as inputs. EEG signals after time-frequency transform can be fed into deep CNN, it takes too much time in training which is undesirable in clinic. The CNN with multilayer is easy to overfitting in terms of small dataset, which is unfortunately aligned with light seizure of only several hundred EEG trials. To solve these challenges, we design a lightweight CNN to identify the pre-ictal state and the inter-ictal state.

III. Material and methodology

A. EEG DATA

The EEG data used in this paper is acquired from the CHB-MIT EEG dataset including scalp EEG (sEEG) recordings of 23 patients suffering from medically intractable focal epilepsy. In order to assess whether patients can be alleviated by surgical intervention, the EEG data is collected by the Neurofile NT digital video EEG system. Each case refers to EEG signals of a patient, containing between 9 to 24 continuous EDF files (EDF is a proprietary format for storing EEG signals and EEG signals in an EDF are referred to as a sample in this paper.) and

TABLE I: The detailed description of the CHB-MIT EEG dataset. Gender: Female (F) and Male (M). Seizure type: Simple partial seizure (SP), Complex partial seizure (CP) and Generalized tonic-clonic seizure (GTC). Brain Location: Frontal, Temporal, Occipital and Parietal. No. of seizures: The number of seizures.

Patient ID	Gender	Age	Seizure type	Brain location	No. of seizures
01	F	11	SP, CP	Frontal	7
02	M	11	SP, CP, GTC	Temporal	3
03	F	14	SP, CP	Frontal	7
04	M	22	SP, CP, GTC	Temporal	4
05	F	7	CP, GTC	Frontal	5
06	F	1.5	CP, GTC	Temporal/Occipital	10
07	F	14.5	SP, CP, GTC	Temporal	3
08	M	3.5	SP, CP, GTC	Frontal	5
09	F	10	CP, GTC	Temporal/Occipital	4
10	M	3	SP, CP, GTC	Temporal	7
11	F	12	SP, CP, GTC	Parietal	3
12	F	2	SP, CP, GTC	Temporal	40
13	F	3	SP, CP, GTC	Temporal/Occipital	12
14	F	9	CP, GTC	Frontal/Temporal	8
15	M	16	SP, CP, GTC	Temporal	20
16	F	7	SP, CP, GTC	Temporal	10
17	F	12	SP, CP, GTC	Temporal	3
18	F	18	SP, CP	Frontal	6
19	F	19	SP, CP, GTC	Frontal	3
20	F	6	SP, CP, GTC	Temporal/Parietal	8
21	F	13	SP, CP	Temporal/Parietal	4
22	F	9	-	Temporal	3
23	F	6	-	Temporal	7

an annotation document clearly states the electrode utilized and the time of seizure start and seizure end in each EDF file. Each recording of case 10 lasts for 2 hours, the recordings of case 4, case 6, case 7, case 9 and case 23 last for 4 hours, and each recording of other cases lasts for 1 hour.

The valid EEG recordings sum up to 664 samples and approximately 983 hours. The start and end time of seizure is recorded in the annotation by clinical experts after visual inspection. Each recording is named as Chb_n, where i denotes patient ID and n indicates the nth sample for patient i. All the detailed information of the 23 cases is listed in Table I.

B. Pre-processing

Since most of the EEG recordings were contaminated by the power interference at 50 Hz while abnormal discharge of seizure mainly occurs in the frequency ranging from 5Hz to 50Hz, a fifth-order Butterworth filter was applied to truncate the 5-50Hz frequency band. Consequently, the filtered data mentioned hereafter all refer to the EEG signals in the frequency ranges of 5-50Hz. Because the electrode utilized for each patient in multiple experiments is somehow different, it is difficult to analyze without selecting their common channels. Therefore, we have picked 18 channels that all patients have, including FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ and CZ-PZ. Epilepsy seizure EEG recordings contain inter-ictal, pre-ictal and ictal durations (as shown in Fig. 1). The seizure prediction horizon (SPH) needs to be defined before algorithm implementation. However, the pre-ictal horizon prescribed is still controversial. In our work, we follow the SPH defined by Maiwald. They believe that 30 minutes as the pre-ictal horizon is an appropriate range which can effectively remind the patient without causing over term tension [11]. As explained above the epilepsy prediction only performs recognition between the pre-ictal state and the ictal state, hence the ictal data are discarded and the rest of data are wielded as the inter-ictal period.

We encounter the following issues on truncating the pre-ictal period. Firstly, seizures for most samples often start before 30 minutes, however, the SPHs do not satisfy the 30-minute required. In such circumstance, we have to fill the part less than 30 minutes with the latter part of a previous consecutive sample to get close 30 minutes. Secondly, some

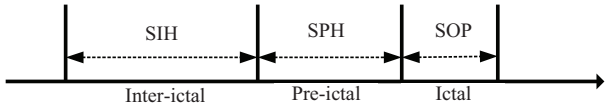


Fig. 1: Definition of the seizure inter-ictal period (SIH), seizure pre-ictal period (SPH) and seizure occurrence period (SOP).

EDF recordings are lost, resulting in interruption between two adjacent samples. For example, chb13_62 starts record at 04:20:55 and seizure at the 851th s (about 14 minutes), however, the previous record chb13_60 ends at 03:20:41. Accordingly the duration of only 14 minutes is segmented as a pre-ictal state of this seizure. Each truncated recording includes pre-ictal and inter-ictal state and is divided into 5s EEG signals as trials for binary classification. The data segmentation method is illustrated in Fig. 2. Fig. 3 shows two sample trial segments of the inter-ictal state and the pre-ictal state, which clearly indicates the difference between these two types of signals.

C. Data augmentation

The trial imbalance problem is not unique but could be knottier for seizure prediction when machine learning is concerned. As for the CHB-MIT dataset, the percentage of pre-ictal trials to the inter-ictal trials can be less than 1:15 in most cases. In general, undersampling and oversampling methods are used together to generate a balanced dataset from imbalance dataset. To balance the two types of data, undersampling reduces the size of the majority class, and oversampling generates extra artificial data, respectively. Compared to undersampling, oversampling is more difficult to achieve due to too many sampling points in an EEG signal. Traditional data augumentation methods such as Bootstrapping [26] and SMOTE [27], cannot generate an authentic artificial sample that is difficult to extricate from real samples. In this step we explore two schemes through extensive experiments in generating the pre-ictal trials. The first one is to recombine EEG signals by means of multi-segment cutting and splicing [28] and the second one is to generate EEG data by generative adversarial networks (GAN) [29]. Both the algorithm complexity and the training time needs to be taken into consideration. In addition, the artificial EEG data generated by GAN lack channel correlation because the EEG signals are generated from separate single channel. For these reasons, the former scheme is finally adopted to generate extra pre-ictal EEG signals.

The idea of our pre-ictal trial augmentation is to first split each training EEG trial into 3 segments, and then generate new artificial trials as a concatenation of segments coming from diverse and randomly selected training trials of the pre-ictal state (as shown in Fig. 4). We randomly discard some of the inter-ictal signals from a training set to make the ratio of the inter-ictal trial to the pre-ictal trial reach 2. Meanwhile, we generate additional pre-ictal signals and put them in the training set to make the ratio of the inter-ictal trial to the pre-ictal trial reach 3/2. This scheme facilitates generating a large number of new trials, which are different from the original ones but are closely relevant and may be similar to other trial, since they are part of the real trials and have the same temporal structure. By adding these new data to the original training set to enrich the feature space in a correlative way can ease the training of the subsequent machine learning algorithms.

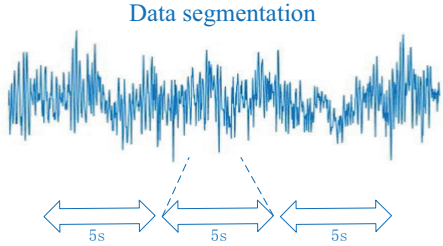
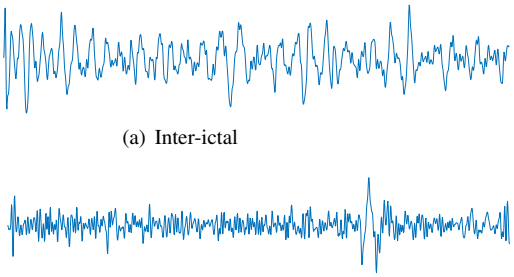


Fig. 2: Data segemention without overlapped.



(a) Inter-ictal

(b) Pre-ictal

Fig. 3: Two example trials segemented from the inter-ictal state and the pre-ictal state.

D. Feature extraction

1) Common spatial pattern: As an extension to PCA, CSP could find a projection matrix composing of several pairs of space filtering vector. The multi-channel EEG signals are projected into a new space through the projection matrix [30] so that the variance of one class is maximized and the other is minimized by the following function:

$$J(\omega) = \frac{\omega^T C_1 \omega}{\omega^T C_2 \omega} \tag{1}$$

The detailed process of CSP is as follows.

The covariance of each trial of the two kinds of EEG signals is calculated by equation (7),

$$C = \frac{E_{N*T} E_{N*T}^T}{trace(E_{N*T} E_{N*T}^T)} \tag{2}$$

where E_{N*T} denotes the trail's original EEG signal, N is the number of channels, T is the number of points, and $trace(X)$ is the sum of diagonal elements of matrix X. The sum of covariance matrices for both classes' EEG signals is calculated as follows:

$$C = C_1 + C_2 \tag{3}$$

The covariance matrix C is decomposed by eigenvalue decomposition as:

$$C_c = U_c A_c U_c^T \tag{4}$$

where U is the $N * N$ diagonal matrix of eigenvector, and A_c is the $N * N$ matrix of eigenvalue. To remove channel-to-channel correlation, the whitening matrix P is calculated by equation (5).

$$P = A_c^{-1/2} U_c^T \tag{5}$$

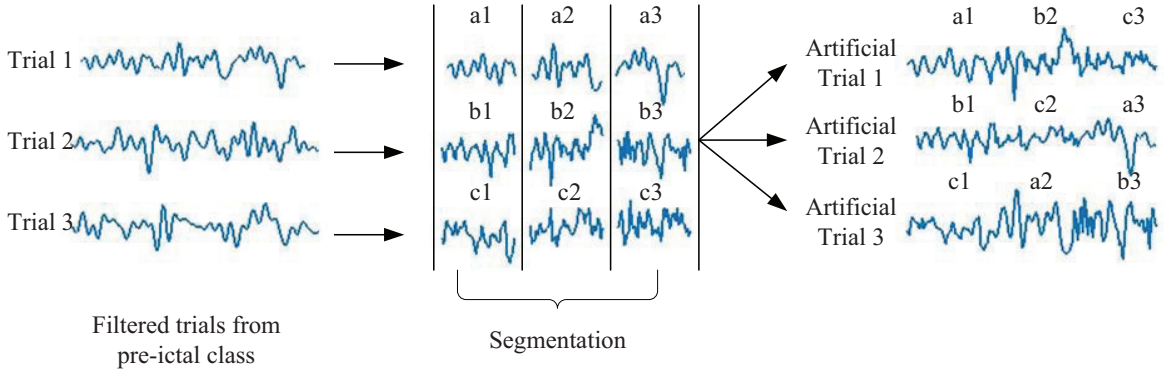


Fig. 4: The flow diagram of data augumentation.

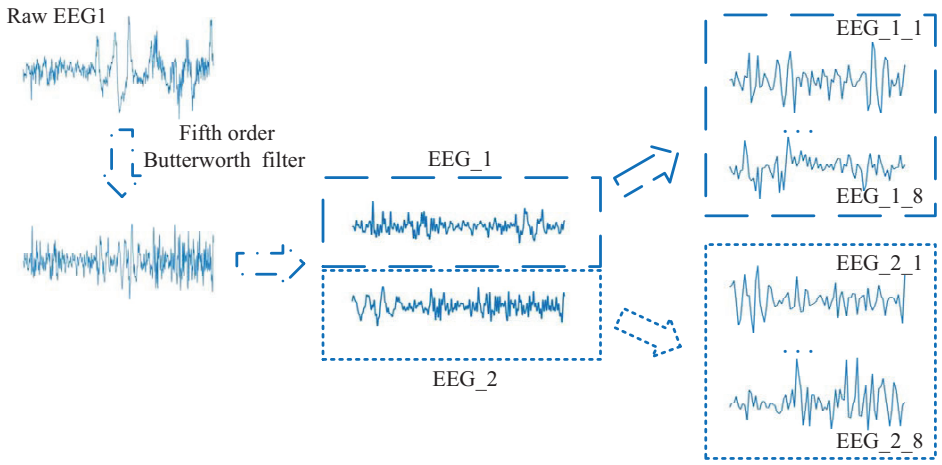


Fig. 5: The flow diagram of the extractor.

Using the characteristics of the same eigenvector after whitening, S_c can be obtained as:

$$S_c = PC_cP^T \tag{6}$$

$$S_c = BA_cB^T \tag{7}$$

The spatial filter W is obtained by equation (8).

$$W = \left(B^T P\right)^T \tag{8}$$

Z can be calculated by spatial filtering using equation (9).

$$Z_{N*T} = W_{N*N}E_{N*T} \tag{9}$$

Features can be obtained by extracting the first m and the last m line of Z by equation (9).

2) Feature matrix: CSP can effectively extract discriminative patterns from EEG signals. In [21] the authors utilize CSP to extract features from raw EEG signals. Nevertheless, the performance of this spatial filter depends on the operational frequency band of the EEG and there exists difference in the operational frequency between individuals. It is unfeasible to manually select a specific frequency range for each subject. Therefore, different from [21], we extracted the features from 9 frequency bands. The wavelet packet decomposition, as a superior time-frequency analysis tool, divides EEG signals into 8 sub-bands with the same frequency span. The classifier used in Section III.E assigns different weight to each frequency band to automatically select the suitable ones for all the patients. We obtain the 9*18 feature matrix from 8 sub-bands and 1 original data including EEG_1, EEG_1_1,..., and EEG_1_8 as shown in Fig. 5. In addition, to make full use of the temporal correlation in the feature matrix, we divide the EEG trial into 2 segments of 2.5s (such as EEG_1 and EEG_2 as shown in Fig. 5), and extract features to finally obtain a feature matrix with size of $18 * 18$. The features after normlization is fed into classifier. The overall implenmentation process of the extractor is depicted in Fig. 5.

E. Classification and evaluation

CNN is a classical deep neural network most commonly applied in computer vision and natural language processing. In our work, the feature matrix after normlization is fed into a shallow CNN as input with 2 blocks (as shown in Fig. 3). Each block consists of a convolution unit (C1, C2) with a leaky rectified linear unit (Leaky Relu) activation function, a max pooling unit (S1, S2) and a dropout unit with rate of 0.5 (the dropout layer of the first block is not shown in Fig. 3).

3*3 kernel with the stride of 1*1 is applied to C1 and C2. Leaky Relu activation is applied to the convolution results

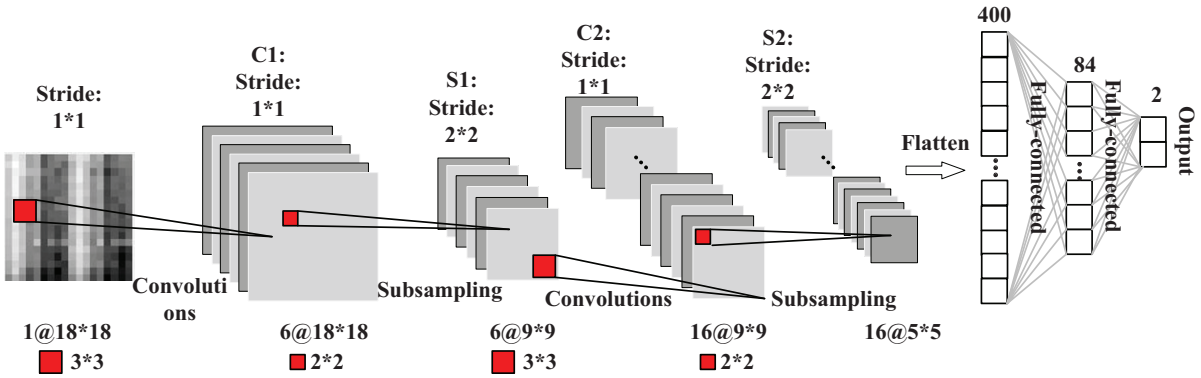


Fig. 6: 2-Layer CNN architecture for seizure prediction.

before the max pooling layer. The first block and the second block have 6 and 16 convolution kernels, respectively. Features extracted further by the two blocks are flattened and connected to two fully connected layers with output sizes of 84 and 2, respectively. The former fully connected layer and the latter use a sigmoid activation function and a soft-max activation function, respectively. Because of the limited available datasets, we design a shallow CNN to prevent overfitting. In addition, we randomly pick one of the seizure samples from the training set as a validation set to further overcome overfitting. After each training epoch, the accuracy is calculated with respect to the validation set to check if the network starts to overfit.

Although the output of the classifier represents the state of a 5s EEG trial, each trial is analyzed independently without considering the association between the previous and latter states. A good predictor depends on several trials to forecast the oncoming state, however, a prediction result relying on only one trial will result in high false alarm rate. In order to eliminate the possibility of such case, Kalman filtering is used to reduce mispredictions. The Kalman filtering equation is described as follows.

$$f_{out}\left[n\right] = \frac{\sum_{k=n-T}^{n} O\left[k\right]}{T} \tag{10}$$

where $O\left[k\right]$ denotes the output of classifier. The $O\left[k\right]$ equals 0 when output is inter-ictal state and the $O\left[k\right]$ equals 1 when the output is pre-ictal state. T denotes the necessary time for continuous monitoring and is set to 30 in our experiment. $f_{out}\left[n\right]$ is the output after filtering. When it reaches 1, the alarm is triggered.

Cross-validation is a technique used to evaluate whether the results of a statistical analysis can be generalized to a separate data set. In order to obtain reliable output, we choose the leave-one-out cross-validation to verify the model in real scenarios. Suppose there are N seizures data for a certain patient. Each seizure data is adopted as a single test set, and the remaining $N-1$ samples are used as training set to obtain N prediction results for each seizure. The average of the classification accuracy of the N results is defined as the performance of the classifier. Comparing with k-fold cross validation, leave-one-out cross-validation is deterministic in that there could be no random factors happening and the whole process is repeatable.

Two evaluation metrics to measure algorithm performance are applied in our work: sensitivity and FPR. Sensitivity is defined as the percentage of seizure correctly predicted in the total number of seizures, which can measure the ability of correct seizure identification. FPR represents the ratio per hour of inter-ictal trials which are currently misclassified as pre-ictal trials to all the pre-ictal trials. It is the index to calculate the possibility of misdiagnosis.

IV. Results

In this work, we evaluate the proposed algorithm on CHB-MIT sEEG dataset using leave-one-out cross-validation. Table I presents the results of the proposed method on 23 patients. An average sensitivity of 92.2% and an average FPR of 0.11/h is achieved. The proposed predictor reaches 100% sensitivity except for patient 2, patient 6, patient 9, patient 14 and patient17. Among them, the patients with low sensitivity and higher FPR, such as patient 3, 13 and 17, have common characteristics that the number of seizures detected is too less so that the pre-ictal trials in the trainning set are much less comparing with other patients.

The visualization of the 18*18 feature matrix of the pre-ictal state and inter-ictal state for patient 1 is presented in Fig. 7. The larger the feature value, the darker the corresponding point will appear in Fig. 7. On the contrary, lighter color indicates corresponding smaller features. It is evident that the two types of features have clear distinction. The maximum of feature values in each column always lie in the end row for the inter-ictal state in Fig.7 (a), however for the pre-ictal state in Fig.7 (b), it is on the contrary that their maximum value lies in the first row. Thus, it fully verifies that our extractor has strong ability to learn their differences.

In order to evaluate the performance of our designed CNN, we feed the features of each round obtained from patient 1 into linear discriminant analysis (LDA) and support vector machine (SVM). Patient 1 has a total of 7 seizures, therefore 7 rounds experiment need to be implemented if we utilize leave-one-out cross-validation. Fig. 8 shows the accuracy of SVM, LDA and CNN for 7 rounds on patient 1. The poly function and 12 are selected as kernel function and the degree of kernel

function, respectively. The two selected parameter produce good accuracy in most rounds. It can be observed from Fig. 8 that CNN offers the best performance in all rounds. The superior performance in terms of accuracy clearly indicates that our CNN is powerful in seizure prediction.

TABLE II: The performance of the proposed algorithm on 23 patients.

Case ID	Number of Seizure	Number of channels used	SEN (%)	FPR (/h)
01	7	18	100	0.001
02	3	18	67.7	0.2
03	7	18	100	0.05
04	4	18	100	0.005
05	5	18	100	0.091
06	10	18	80	0.04
07	3	18	100	0.07
08	5	18	100	0.05
09	4	18	100	0.115
10	7	18	100	0.14
11	2(3*)	18	100	0.01
12	21(40*)	18	100	0.11
13	11(12*)	18	67.7	0.17
14	8	18	75	0.19
15	17(20*)	18	100	0.13
16	9(10*)	18	100	0.16
17	3	18	67.7	0.14
18	6	18	100	0
19	3	18	100	0.1
20	8	18	100	0.06
21	4	18	100	0.14
22	3	18	67.7	0.45
23	7	18	100	0.012
Total	157	-	92.2	0.11

* Two seizures are combined when the second one is in the postseizure interval of the first one.

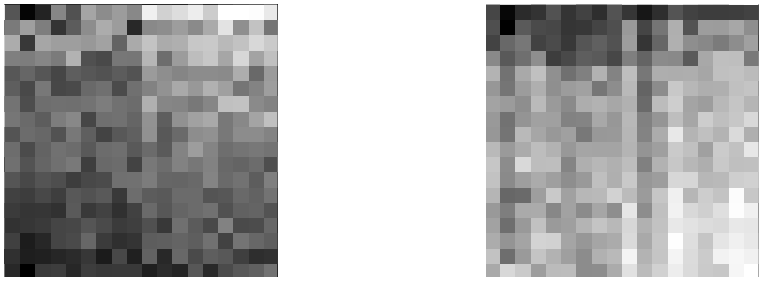
V. Discussion

Threshold crossing and machine learning classifier are two popular directions for seizure prediction, some of which have high sensitivity or low FPR. Our results with highly sensitivity and low FPR are compared with the-state-of-art methods using the same CHB-MIT dataset and Freiburg Hospital EEG dataset (FH dataset). Table III summarizes some key information of these works in chronological order.

The two datasets used in Table III are CHB-MIT dataset and FH dataset which belong to sEEG and intracranial EEG (iEEG), respectively. The iEEG signals have higher signal-to-noise ratio and spatial resolution compared to sEEG signals. Because the sEEG is less proximity to neural electrical activity, hence, it's highly susceptible to power frequency interference, baseline drift and other noises from the external environment compared with the iEEG. This leads the fact that the same algorithm has lower sensitivity and higher FPR rate for sEEG. Nevertheless, iEEG data collection is prone to infection and may introduce other complications during craniotomy. Consequently, seizure prediction based on sEEG is more suitable to promote in a real-life scenario. This trend can also be recognized from Table III that CHB-MIT, a popular EEG dataset collected from the scalp is increasingly acknowledged by researchers in this field.

The research endeavors on epilepsy prediction, from thresholding crossing to conventional machine learning and then to deep learning, are roughly outlined in Table III. [11], [31], [32], [33] and [34] adopt threshold crossing to predict seizure, but has been unable to realize the satisfactory performance. Hence, people gradually give up this scheme. Some conventional machine learning classifiers did a good job [2], [22], [35]. Among them, [2] used univariate spectral power and SVM classifer to achieve a high sensitivity of 98.3% and FPR of 0.29/h, and [35] achieved a sensitivity of 95.4% and FPR of 0.36/h using Phase-match error, deviation and LS-SVM classifer. Both of them were are tested on iEEG dataset. Another interesting point lies in that, although [35] is not so good as [2] in terms of sensitivity and FPR, it were tested on more patient cases for better potential clinical usage. Comparing with them, we use sEEG with more patient cases in the experiments, realizing lower FPR.

Compared to those methods tested on the CHB-MIT dataset, the highest sensitivity is reached by our method. Comparing with [36], they used CNN with 6 blocks whose layers are apparently more than ours, however their result is even slightly weaker. These facts discussed above demonstrate that good



(a) Inter-ictal (b) Pre-ictal

Fig. 7: Two features extracted from the inter-ictal state and the pre-ictal state.

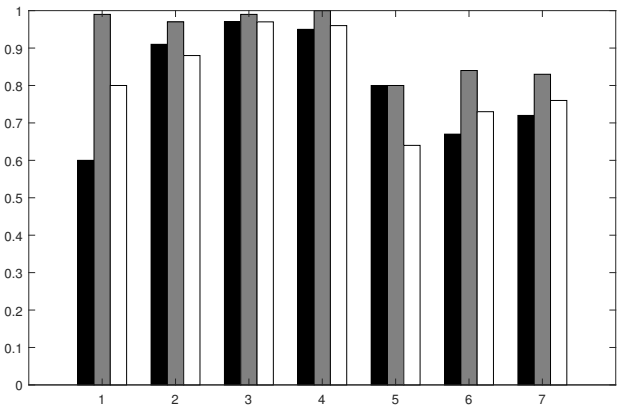


Fig. 8: The comparison between using different classifiers. Among them, the black color, gray color and white color represent LDA, CNN and SVM as the classifier, respectively.

performance can be obtained by the proposed method for seizure prediction.

VI. Conclusion

In this paper, we propose a novel approach for seizure prediction based on data augumentation, CSP and CNN to perform data equalization, feature extraction and classification on 23 patients from CHB-MIT dataset. In order to solve the trial imbalance problem, we generate the pre-ictal trials and discard some inter-ictal trials. According to the temporal-frequency characteristics of EEG, an extractor using CSP in 9 frequency bands and 2 temporal bands is designed. A 2-layer shallow CNN as a classifier is designed for seizure prediction. Extensive experimental results demonstrate that the proposed algorithm outperforms other state-of-the-art methods in terms of sensitivity and FPR. The present study on seizure prediction provides an effective option based on sEEG signals in clinical diagnosis.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China under Grant 61572231, in part by the National Key R&D Program of China under Grant 2017YFE0112000, in part by the Shanghai Municipal Science and Technology Major Project under Grant 2017SHZDZX01, and in part by the Shandong Provincial Key Research & Development Project under Grant 2017GGX10141.

References

[1] L. D. Iasemidis, S. Deng-Shan, C. Wanpracha, S. J Chris, P. M. Pardalos, J. C. Principe, P. R. Carney, P. Awadhesh, V. Balaji, and T. Konstantinos, "Adaptive epileptic seizure prediction system," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 616–627, 2003.

[2] P. Yun, L. Lan, K. K. Parhi, and N. Theoden, "Seizure prediction with spectral power of eeg using cost-sensitive support vector machines," *Epilepsia*, vol. 52, no. 10, pp. 1761–1770, 2011.

[3] K. Gadhoumi, J. M. Lina, F. Mormann, and J. Gotman, "Seizure prediction for therapeutic devices: A review," *Journal of Neuroscience Methods*, vol. 260, pp. 270–282, 2016.

[4] D. R. Freestone, P. J. Karoly, and M. J. Cook, "A forward-looking review of seizure prediction," *Current Opinion in Neurology*, vol. 30, no. 2, p. 1, 2017.

[5] W. Jia, N. Kong, F. Li, X. Gao, S. Gao, G. Zhang, Y. Wang, and F. Yang, "An epileptic seizure prediction algorithm based on second-order complexity measure," *Physiological Measurement*, vol. 26, no. 5, pp. 609–625, 2005.

[6] D. P. E. Mcsharry, T. He, L. A. Smith, and L. Tarassenko, "Linear and non-linear methods for automatic seizure detection in scalp electro-encephalogram recordings," *Medical & Biological Engineering & Computing*, vol. 40, no. 4, pp. 447–461, 2002.

[7] L. Logesparan, E. Rodriguez-Villegas, and A. J. Casson, "The impact of signal normalization on seizure detection using line length features," *Medical & Biological Engineering & Computing*, vol. 53, no. 10, p. 929, 2015.

[8] L. Logesparan, A. J. Casson, and E. Rodriguez-Villegas, "Erratum to: Optimal features for online seizure detection," *Medical & Biological Engineering & Computing*, vol. 54, no. 8, pp. 1–1, 2016.

[9] Y. Li, X. D. Wang, M. L. Luo, K. Li, X. F. Yang, and Q. Guo, "Epileptic seizure classification of eegs using time–frequency analysis based multiscale radial basis functions," *IEEE Journal of Biomedical and Health Informatics*, 2018.

[10] Y. Guo, Y. Zhang, M. Mursalin, W. Xu, and B. Lo, "Automated epileptic seizure detection by analyzing wearable eeg signals using extended correlation-based feature selection," in *IEEE International Conference on Wearable and Implantable Body Sensor Networks*, 2018.

[11] T. Maiwald, "Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic," *Physica D-nonlinear Phenomena*, vol. 194, no. 3, pp. 357–368, 2004.

[12] O. Kocadagli and R. Langari, "Classification of eeg signals for epileptic seizures using hybrid artificial neural networks based wavelet transforms and fuzzy relations," *Expert Systems with Applications*, vol. 88, 2017.

[13] X. Cao, S. Wang, L. Iasemidis, S. Wong, and W. A. Chaovalitwongse, "An adaptive pattern learning framework to personalize online seizure prediction," *IEEE Transactions on Big Data*, vol. PP, no. 99, pp. 1–1, 2017.

[14] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, "Epileptic seizure prediction using relative spectral power features," *Clinical Neurophysiology Official Journal of the International Federation of Clinical Neurophysiology*, vol. 126, no. 2, pp. 237–248, 2015.

[15] S. Blanco, A. Garay, and D. Coulombie, "Comparison of frequency bands using spectral entropy for epileptic seizure prediction." *Isrn Neurology*, vol. 2013, no. 1, p. 287327, 2013.

[16] Z. Zhen, C. Ziyi, Z. Yi, D. Shouhong, Z. Yang, M. Tian, and T. Xianghua, "Construction of rules for seizure prediction based on approximate entropy," *Clinical Neurophysiology*, vol. 125, no. 10, pp. 1959–1966, 2014.

[17] L. D. Iasemidis, S. D-S, P. M. Pardalos, . Chaovalitwongse, W., . Narayanan, K., . Prasad, A., . Tsakalis, K., P. R. Carney, and J. C. Sackellares, "Long-term prospective on-line real-time seizure prediction," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 532–544, 2005.

[18] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using eeg signals," *Computers in Biology & Medicine*, vol. 99, p. S001048251830132X, 2018.

[19] M. Piotr, M. Deepak, L. Yann, and K. Ruben, "Classification of patterns of eeg synchronization for seizure prediction." *Clinical Neurophysiology*, vol. 120, no. 11, pp. 1927–1940, 2009.

[20] S. Z. Ali, T. Reza, J. Manouchehr, and G. A. Dumont, "Predicting epileptic seizures in scalp eeg based on a variational bayesian gaussian mixture model of zero-crossing intervals," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 5, pp. 1401–1413, 2013.

[21] T. N. Alotaiby, S. A. Alshebeili, F. M. Alotaibi, and S. R. Alrshoud, "Epileptic seizure prediction using csp and lda for scalp eeg signals," *Computational Intelligence and Neuroscience*, vol. 2017, no. 6, pp. 1–11, 2017.

[22] M. H. Myers, A. Padmanabha, G. Hossain, A. L. D. J. Curry, and C. D. Blaha, "Seizure prediction and detection via phase and amplitude lock values," *Frontiers in Human Neuroscience*, vol. 10, 2016.

[23] D. Cho, B. Min, J. Kim, and B. Lee, "Eeg-based prediction of epileptic seizures using phase synchronization elicited from noise-assisted multivariate empirical mode decomposition," *IEEE Trans Neural Syst Rehabil Eng*, vol. PP, no. 99, pp. 1–1, 2016.

[24] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, S. Ippolito, and O. Kavehei, "Convolutionalneuralnetworks for

TABLE III: Comparison to prior works.

Method	EEG data source	Number of used cases	Number of used seizures	features	classfier	SEN	FPR(/h)	SPH
Maiwald et al., 2004 [11]	FH	21	88	Dynamical similarity index	Threshold crossing	42	<0.15	30
Winterhalder et al., 2006 [31]	FH	21	88	Phase coherence, lag synchronization	Threshold crossing	60	0.15	30
Park et al., 2011 [2]	FH	18	80	Univariate spectral power	SVM	98.3	0.29	30
Li et al., 2013 [32]	FH	21	87	Spike rate	Threshold crossing	72.7	0.11	50
Zheng et al.,2014 [33]	FH	10	50	Mean phase coherence	Threshold crossing	>70	<0.15	30
Eftekhar et al., 2014 [34]	FH	21	87	Multiresolution N-gram	Threshold crossing	90.95	0.06	20
Aarabi & He, 2014 [37]	FH	21	87	Bayesian inversion of power spectral density	Rule-based decision	87.07	0.2	30
Parvez & Paul, 2017 [35]	FH	21	87	Phase-match error, deviation	LS-SVM	95.4	0.36	30
Aarabi et al., 2017 [38]	FH	10	28	Univariate and bivariate features	SVM	86.7	0.126	30
Zandi et al., 2013 [20]	CHB-MIT	3	18	Zero crossings, similarity/ dissimilarity index	-	83.81	0.165	40
Myers et al., 2016 [22]	CHB-MIT	10	31	Phase/Amplitude locking value	-	77	0.17	60
Khan et al., 2017 [25]	CHB-MIT	-	131	Wavelet transform	CNN	87.8	0.14	10
Cho et al., 2017 [23]	CHB-MIT	21	65	Phase locking value	SVM	82.44	-	5
Chu et al., 2017 [39]	CHB-MIT	13	125	Fourier Transform coefficients, PSD	-	83.81	0.165	86
Truong et al., 2017 [36]	CHB-MIT	13	64	STFT spectral images	CNN	81.2	0.16	5
Alotaiby et al., 2017 [21]	CHB-MIT	24	170	Common spatial pattern statistics	LDA	81	0.47	60
The proposed method	CHB-MIT	23	157	Combination of Common spatial pattern statistics	CNN	92.2	0.11	30

FH: Freiburg Hospital intracranial EEG dataset; SEN: sensitivity; FPR: false prediction rate; SPH: seizure pre-ictal horizon; CNN: convolutional neural network; SVM: support vector machine; LS-SVM: least squares-support vector machine.

seizureprediction using intracranial and scalp electroencephalogram," *Neural Networks*, vol. 105, p. S0893608018301485, 2018.

[25] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal onset seizure prediction using convolutional networks," *IEEE Transactions on Biomedical Engineering*, vol. PP, no. 99, pp. 1–1, 2017.

[26] C. Yin, Z. Feng, Y. Lin, and S. Belongie, "Fine-grained categorization and dataset bootstrapping using deep metric learning with humans in the loop," in *Computer Vision & Pattern Recognition*, 2016.

[27] J. Mathew, M. Luo, C. K. Pang, and H. L. Chan, "Kernel-based smote for svm classification of imbalanced datasets," in *Conference of the IEEE Industrial Electronics Society*, 2016.

[28] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 871–890, 2015.

[29] S. Palazzo, C. Spampinato, I. Kavasidis, D. Giordano, and M. Shah, "Generative adversarial networks conditioned by brain signals," in *IEEE International Conference on Computer Vision*, 2017.

[30] S. Kumar, K. Mamun, and A. Sharma, "Csp-tsm: Optimizing the performance of riemannian tangent space mapping using common spatial pattern for mi-bci." *Computers in Biology & Medicine*, vol. 91, pp. 231–242, 2017.

[31] M. Winterhalder, B. Schelter, T. Maiwald, A. Brandt, A. Schad,

A. Schulzebonhage, and J. Timmer, "Spatio-temporal patient–individual assessment of synchronization changes for epileptic seizure prediction," *Clinical Neurophysiology*, vol. 117, no. 11, pp. 2399–2413, 2006.
[32] S. Li, W. Zhou, Q. Yuan, and Y. Liu, "Seizure prediction using spike rate of intracranial eeg." *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, vol. 21, no. 6, pp. 880–886, 2013.
[33] Y. Zheng, G. Wang, K. Li, G. Bao, and J. Wang, "Epileptic seizure prediction using phase synchronization based on bivariate empirical mode decomposition," *Clinical Neurophysiology*, vol. 125, no. 6, pp. 1104–1111, 2014.
[34] E. Amir, J. Walid, E. I. Jamil, T. G. Constandinou, and T. Christofer, "Ngram-derived pattern recognition for the detection and prediction of epileptic seizures," *Plos One*, vol. 9, no. 6, p. e96235, 2014.
[35] M. Z. Parvez and M. Paul, "Seizure prediction using undulated global and local features," *IEEE transactions on bio-medical engineering*, vol. 64, no. 1, pp. 208–217, 2016.
[36] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, and O. Kavehei, "A generalised seizure prediction with convolutional neural networks for intracranial and scalp electroencephalogram data analysis," 2017.
[37] A. Ardalan and H. Bin, "Seizure prediction in hippocampal and neocortical epilepsy using a model-based approach," *Clinical Neurophysiology*, vol. 125, no. 5, pp. 930–940, 2014.
[38] A. Aarabi and B. He, "Seizure prediction in patients with focal hippocampal epilepsy," *Clinical Neurophysiology*, vol. 128, no. 7, p. 1299, 2017.
[39] H. Chu, C. K. Chung, W. Jeong, and K. H. Cho, "Predicting epileptic seizures from scalp eeg based on attractor state analysis," *Computer Methods & Programs in Biomedicine*, vol. 143, no. C, pp. 75–87, 2017.

Zhang Yuan Yuan Zhang received his M.S. degree in Communication Systems and Ph.D. degree in Control Theory & Engineering (Biomedical Engineering) both from Shandong University, China, in 2003 and 2012respectively. He is currently an Associate Professor at University of Jinan,China. Dr. Zhang was a visiting professor at Computer Science Department,Georgia State University, USA, in 2014. As the first author or corresponding author he has published more than 50 peer reviewed papers in international journals and conference proceedings, 1 book chapters, and 6 patents in the areas of Smart Health and Biomedical Big Data Analytics. He has served as Leading Guest Editor for six special issues of IEEE, Elsevier, Springer and InderScience publications, including IEEE Internet of Things Journal special issue on Wearable Sensor Based Big Data Analysis for Smart Health and IEEE Journal of Biomedical and Health Informatics (JBHI)special issue on Pervasive Sensing and Machine Learning for Mental Health. He has served on the technical program committee for numerous international conferences. He is an associate editor for IEEE Access and Internet of Things (Elsevier). Dr. Zhang's research interests are Wearable Sensing for Smart Health, Machine Learning for Auxiliary Diagnosis, and Biomedical Big Data Analytics. His research has been extensively supported by the Natural Science Foundation of China, China Postdoctoral Science Foundation, and Natural Science Foundation of Shandong Province with total grant funding over 1.4million RMB. Dr. Zhang is a Senior Member of both IEEE and ACM. For more information, please refer to http://uslab.ujn.edu.cn/index.html.

Yao Guo Yao Guo received her BS in Communication Engineering in 2016 from University of Jinan, China. She is a master student majoring in computer science at Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan. She is a member of both ACM and IEEE. Her current research interests include biomedical signal/image processing machine learning and pattern recognition.

Po Yang is a Senior Lecturer in Computer Science at Liverpool John Moores University. He has generated over 60 international journal and conference papers (over 20 IEEE Trans/Journal) in the fields of Pervasive Healthcare, Image Processing, Parallel Computing and RFID related internet of things (IoT) applications. His research work has been supported by European FP6/FP7/H2020, UK EPSRC, Innovate UK, including 8 EU projects and 4 EPSRC/TSB funded projects. From 2013-2015, he worked as a technical manager in 3 EU projects (MHA: No 600929, CHIC: No 600841, CARRE: No 611140) and 1 EPSRC project (No EP/L023830/1), which are all within healthcare fields. He serves as the Associate Editor in IEEE Journal of Translational Engineering in Health and Medicine and IEEE Access, the Guest Editor for 7 International Journals, the Reviewers for 18 International Journals (10 Trans), and have chaired/co-chaired over 10 IEEE international conferences and workshops. He received 4 Best Paper Awards from IEEE PiCOM 2015, HPCSC 2013, IEEE ICAC 2008, IEEE ICAC 2007.

Wei Chen is Professor and Director of Center of Intelligent Medical Electronics at School of Information Science and Technology, Fudan University. She received her B. Eng. degree in 1999 and M. Eng. degree in 2002 on telecommunication systems and smart sensor systems, from School of Electrics and Information Engineering, Xian Jiaotong University, China. She obtained her Ph.D. degree in 2007 on performance monitoring and impairment mitigation for optical communication systems, from the Department of Electrical & Electronics Engineering, The University of Melbourne, Australia. She worked at Bell Laboratories Germany, Alcatel-Lucent, Stuttgart, Germany as an intern in 2005 and she was a research assistant in 2007 at the Department of Electrical & Electronics Engineering, The University of Melbourne, Australia. From 2007 to 2015, she was an Assistant Professor at Eindhoven University of Technology, the Netherlands. From 2009 to 2013, she served as Chair of the Theme Health Care at the Department of Industrial Design, Eindhoven University of Technology, the Netherlands. Since Oct. 2015, she has been a full professor and director of Center of Intelligent Medical Electronics at Department of Electronic Engineering, School of Information Science and Technology, Fudan University. Her research interests include medical monitoring system design using wearable sensors, patient health monitoring, sleep monitoring, brain activity monitoring, wireless body area networks, ambient intelligence, personalized and smart environment, smart sensor systems, and signal processing. She has been the supervisor for more than 10 PhD candidates and graduated 5 PhD students, among them 3 with Cum Laude. She has published two books as the main editor and she has more than one hundred publications in prestigious journals and international conferences. As PI and co-PI Prof. Chen has led eight projects including EU, STW NL, and China Minister of Science and Technology (MOST) projects. Currently, as PI she is the consortium leader of one MOST major project, one Shanghai Municipal of Science and Technology project and as co-PI she is conducting one Shanghai Municipal of Science and Technology Major project. She is a Senior Member of IEEE and associate editor of IEEE Journal on Biomedical Health Informatics (J-BHI).

Benny Lo is a Senior Lecturer of the Hamlyn Centre and the Department of Surgery and Cancer, Imperial College London. He obtained his BASc in Electrical Engineering from the University of British Columbia, Canada, MSc (Distinction) in Electronics from King's College London and PhD in Computing from Imperial College London. Dr Lo's research mainly focuses on Body Sensor Networks (BSN) and Wearable Technologies for healthcare and wellbeing. As one of the pioneers in BSN, Dr Lo has introduced numerous new platform technologies, and novel approaches for wellbeing, and healthcare applications. Much of his research has translated for clinical applications, and he has also led a few pioneering clinical trials of novel sensing technologies, such as the clinical validation of a new spectroscopic sensor for blood perfusion monitoring for micro-vascular surgeries. His innovative work has been recognized with numerous innovation awards, such as the IMAGINE challenge. The impact of his research is apparent from his many highly cited publications. He has published over 150 peer reviewed publications in the field of BSN. His work has led to numerous awards, such as the Medical Futures Award. He is a Senior Member of IEEE, an Associate Editor of the IEEE Journal of Biomedical Health Informatics, Chair of the IEEE EMBS Wearable Biomedical Sensors and Systems Technical Committee, and a member of the IEEE EMBS Standards Committee. His expertise has been well regarded by his peers and he has been invited as expert reviewers and expert panel members, such as the panel for the Research Assessment Exercise of Tampere University of Technology and the International Advisory Board Member for IOP Physiological measurement.

Wenyao Xu received the Ph.D. degree from the University of California at Los Angeles, Los Angeles, USA, and both the Master and Bachelor degree from Zhejiang University, China. Wenyao Xu is an Associate Professor with tenure of the Computer Science and Engineering Department, University at Buffalo (SUNY). His group has focused on exploring novel sensing and computing technologies to build up innovative Internet-of-Things (IoT) systems for high-impact human-technology applications in the fields of Smart Health and Cyber-Security. Results have been published in peer-reviewed top research venues across multiple disciplines, including Computer Science conferences (e.g., ACM MobiCom, SenSys, MobiSys, UbiComp, ASPLOS, ISCA, HPCA and CCS), Biomedical Engineering journals (e.g., IEEE TBME, TBioCAS, and JBHI), and Medicine journals (e.g., LANCET). To date, his group has published over 150 papers, won six best paper awards, two best paper nominations and three international best design awards. His inventions have been filed within U.S. and internationally as patents, and have been licensed to industrial players. His research has been reported in high-impact media outlets, including the Discovery Channel, CNN, NPR and the Wall Street Journal. Dr. Xu is an Associate Editor of IEEE Transactions on Biomedical Circuits and Systems (TBCAS), serves the technical program committee of numerous conferences in the field of Smart Health and Internet of Things, and has been a TPC co-chair of IEEE Body Sensor Networks in 2018.