IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON CYBER-PHYSICAL-SOCIAL COMPUTING AND NETWORKING

Received December 24, 2017, accepted February 17, 2018, date of publication March 1, 2018, date of current version April 4, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2810882

Towards Brain Big Data Classification: Epileptic EEG Identification With a Lightweight VGGNet on Global MIC

HENGJIN KE[1], DAN CHEN[1], (Member, IEEE), XIAOLI LI[2], YUNBO TANG[1], TEJAL SHAH[3], AND RAJIV RANJAN[3,4]

[1]Computer School, Wuhan University, Wuhan 430072, China
[2]National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China
[3]Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.
[4]School of Computer Science, China University of Geosciences, Wuhan 430074, China

Corresponding author: Dan Chen (dan.chen@ieee.org) and Rajiv Ranjan (raj.ranjan@ncl.ac.uk)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772380 and in part by the Foundation for Innovative Research Groups of Hubei Province under Grant 2017CFA007.

ABSTRACT Brain big data empowered by intelligent analysis provide an unrivalled opportunity to probe the dynamics of the brain in disorder. A typical example is to identify evolving synchronization patterns from multivariate electroencephalography (EEG) routinely superimposed with intensive noise in epilepsy research and practice. Under the circumstance of insufficient *a priori* knowledge of subject dependency on domain problem, it becomes even more important to adaptively classify the synchronization dynamics to accurately characterize the intrinsic nature of seizure activities represented by the EEG. This paper first measures the global maximal information coefficient (MIC) of all EEG data channels to form a time sequence of correlation matrices. A lightweight VGGNet (Visual Geometry Group) is designed to adapt to the need to prune massive EEG datasets. The *VGGNet* characterizes the synchronization dynamics captured in the correlation matrices and then automatically identifies the seizure states of the EEG. Experiments are performed over the Children's Hospital Boston-Massachusetts Institute of Technology (CHB-MIT) scalp EEG dataset to evaluate the proposed approach. Seizure states can be identified with an accuracy, sensitivity, and specificity of [98.13% $\pm$ 0.24%], [98.85% $\pm$ 0.51%], and [97.47% $\pm$ 0.36%], respectively; the resulting performance is superior to those of most existing methods over the same dataset. The approach directly applies to raw EEG analysis, which holds great potential for handling brain big data.

INDEX TERMS Brain big data, pattern classification, VGGNet, synchronization measurement, EEG, epilepsy.

I. INTRODUCTION

Neuroscience research and practice have embraced the big data era. Brain big data maintain long term neural recordings of a large number of subjects under various conditions, which hold great potential to reveal the hidden mechanisms that drive brain activities. The recent boom in computational intelligence provides an unprecedented opportunity to probe brain dynamics based on brain big data. Synchronization measurement has long been a hotspot in neuroscience research in terms of both brain functions and malfunctions [1], e.g. diagnosis of brain diseases. The ability to find synchronization patterns in multivariate electroencephalography (EEG) possibly superimposed by intensive noise is increasingly important in feature extraction [2], complex oscillator networks, neural computing [3], and brain disorder detection [4]. Synchronization measurement of EEG manifests an effective means to characterize the underlying brain dynamics, e.g. identification and prediction of brain states. A typical example is to identify evolving synchronization patterns from multivariate EEG routinely superimposed with intensive noises in epilepsy research and clinical practice. The huge diversity of EEG belonging to different patients makes this task even more challenging.

Bivariate synchronization analyses have been extensively investigated in the neuroscience community. Among the classic bivariate methods, mutual information (MI) is salient for discrimination and robustness to noise [5] with its information theoretic interdependence measures [3].

Maximal Information Coefficient (MIC) has then emerged as the best bivariate synchronization measurement for analyses [6] in terms of nonlinearity and robustness to noise. Multivariate synchronous analysis methods have been developed, such as phase synchronization cluster analysis (PSCA), S-estimator [7], and correlation matrix analysis (CMA) [8]. Those cannot adapt to the difficulties of (1) uncertain levels of detail of synchronization measurement, (2) intensive embedded noises, and (3) limited computing capabilities at the same time.

Numerous methods have also been developed to classify EEG synchronization patterns, including linear (e.g. Kappa statistics [9] and K-means [10]) and non-linear classifiers (e.g. Support Vector Machine (SVM) [11]). EEG data are routinely non-linear and non-stationary in nature, and synchronization patterns (if any) embedded in EEG are inevitably highly nonlinear. This always results in poor performance for linear classifiers [9], [10]. In particular, Kappa is incapable of revealing synchronization patterns in detail, and K-means is often trapped at a local optimum due to its high sensitivity to noises and outliers. SVM applies to non-linear problems, while it cannot foster a general solution to EEG synchronization classification: (1) selection of the kernel function is problem-specific and (2) the space information among synchronization patterns is discarded.

To tackle these challenges, an appropriate solution should be able to (1) adaptively characterize the non-linear and non-stationary synchronization patterns of EEG with brain disorder belonging to different subjects, (2) capture the synchronization dynamics in detail under the circumstance of intensive noises, and (3) enable a general and cost-effective solution. The approach proposed in this study is designed as follows:

- It first organizes the Maximal Information coefficients (*MIC*) of all EEG data channels to form a time sequence of correlation matrices (CMMICs) to record the global synchronization dynamics in great detail. The *CMMIC* sequence can be easily transformed to observe synchronizations between clusters of channels, e.g. those in different brain regions. In other words, the spatial level-of-detail can be flexible per request. Variation in time windows can also result in change in temporal resolution. The *CMMIC* sequence forms the basis for EEG identification in this study, which has the merit of resistance to noises determined by the *MIC* theory.
- As for classification of EEG synchronization patterns, this study utilizes Convolutional Neural Networks (CNN) as it excels in adaptive selection of features. With a convolution operation capable of extracting distortion-invariant patterns, CNN gained great successes in video recognition, especially for the recently emerging VGGNet (Visual Geometry Group) [12]. A *CMMIC* sequence is inherently similar to a video in terms of both (1) the non-linearity of data elements in each matrix (frame) and (2) the dynamic evolution of matrices (frames). A lightweight *VGGNet* is then designed (Section IV-A) considering the need for efficiency and the much smaller scale of *CMMIC*s.

The proposed approach extracts the global synchronization features without a priori knowledge of EEG. The *VGGNet* model is trained in an off-line manner then applies to other subjects for on-line prediction of the states of epileptic EEG. Experiments are performed to evaluate the proposed approach over the Children's Hospital Boston-Massachusetts Institute of Technology (CHB-MIT) scalp EEG dataset (see `http://physionet.org/physiobank/database/chbmit` [13]). Experimental results indicate that this approach can classify seizure states with high accuracy, sensitivity, and specificity achieved. The overall performance is superior to those of most existing methods. The classifier holds great potential in minimizing false alarm of epilepsy seizure onset incurred by significant noise and interference in sophisticated scientific and engineering applications. It is less error prone as only one hyper-parameter of time window size needs to be set manually. The main contributions of this study include:

1) A lightweight classifier has been designed to identify epileptic EEG without the need for a priori knowledge on the EEG data. It exhibits excellent performance in seizure onset detection and can be generalized to analysis of other types of EEGs.
2) A complete solution has been developed to automatically characterize the synchronization dynamics of multivariate epileptic EEG superimposed by a high level of noise and interference. The risk of missing structural information of EEG incurred by excessive denoising is minimized.

The remainder of this paper is organized as follows: Section II presents related work and the objectives of this study. Section III introduces the proposed correlation matrix based on *MIC* (*CMMIC*). Section IV outlines the classifier using a lightweight VGGNet. Section V presents the performance evaluation of the proposed approach and gives a comparison with the state-of-the-art. Section VI concludes the paper with a summary.

II. RELATED WORK

Detection and classification of the patterns hidden in multivariate EEG has long been an interesting research issue in probing brain diseases such as epilepsy. Traditional methods focus on time frequency analysis and synchronization measurement. Recently, machine learning methods have boomed. The most salient works pursuing this direction are introduced as the follows:

Myers et al. proposed a seizure prediction and detection algorithm by calculating the Phase/Amplitude Lock Values (PLV/ALV). The algorithm achieved a sensitivity of 0.77, a precision of 0.88 and 0.17 false positive per hour over the CHB-MIT scalp EEG dataset [14].

In order to find the EEG segments with seizures and their onset and offset points, Lorena et al. developed a patient

non-specific strategy for seizure detection based on Stationary Wavelet Transform of EEG signals and achieved specificity of 99.9%, sensitivity of 87.5% and a false positive rate per hour of 0.9 over the CHB-MIT scalp EEG dataset [15].

Piotr et al. proposed a method to classify patient-specific synchronization patterns to predict seizure onset over a Freiburg dataset [2]. EEG synchronization was measured via cross-correlation, non-linear interdependence, dynamic entrainment or wavelet synchrony. Spatio-temporal patterns were then extracted to support seizure onset predication, which achieved a sensitivity of 71% and zero false positives.

Fergus et al. proposed a new method for generalizing seizure detection across different subjects without a priori knowledge about the focal point of seizures over the CHB-MIT scalp EEG dataset [16]. Classification was enabled by the k-NN algorithm and achieved a sensitivity of 88% and a specificity of 88%.

Morteza et al. proposed a density-based real-time seizure prediction algorithm based on a trained offline seizure detection model. The method achieved an accuracy rate of 86.56%, a precision rate of 86.53%, a recall rate of seizure prediction of 97.27%. The false prediction rate was 0.00215 per hour with their online signal prediction algorithm on the CHB-MIT dataset [17].

In contrast to the existing work, this study aims to find a solution with the capability of (1) detection of synchronization with robustness to the intensive noise embedded in the EEG with the evolving synchronization dynamics considered, (2) adaptive classification of the non-stationary synchronization patterns to capture the intrinsic nature of seizure activities represented by the EEG, and (3) high efficiency in classification to cater to the needs of potential big data applications.

III. CORRELATION MATRIX BASED ON MAXIMAL INFORMATION COEFFICIENT

This section first presents the operation process of the proposed approach. Synchronization measurement is performed to form the Correlation Matrix based on Maximal Information Coefficient.

A. OVERALL DESIGN

Considering the need for efficiency of analysis, this study attempts to minimize the efforts of conventional EEG preprocessing (basically denoising) that normally manifests as an onerous task. Another concern is that existing methods largely demand sufficient a priori knowledge and excessive hyper-parameter settings. Fig. 1 illustrates the overall design of the proposed approach, which operates in two phases: (1) feature extraction of synchronization dynamics, and (2) pattern classification upon the lightweight VGGNet. The unlabelled raw EEG data are segmented with the same window size (8 seconds in the experiments). All *MIC* measurements of all channel pairs in each time window are calculated and organized as a *CMMIC*. The *CMMIC* time sequences are then processed and classified by the lightweight VGGNet.

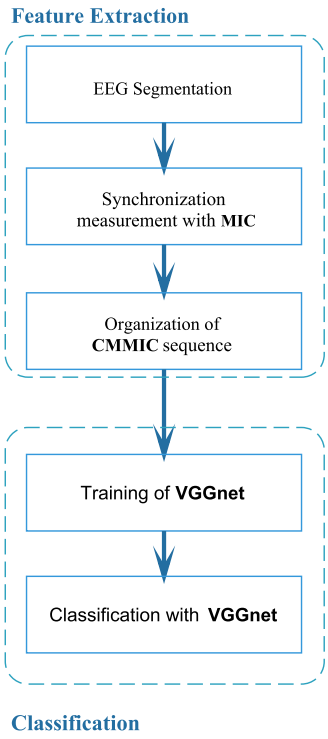

FIGURE 1. Overview of the proposed approach and its operation process.

B. MAXIMAL INFORMATION COEFFICIENT

MIC is intended to measure the linear or non-linear synchronization relationship between two random variables, e.g. bivariate EEG segments, which is part of a larger family of maximal information-based nonparametric exploration (MINE) statistics [6]. *MIC* is an informative measure to identify a subset of the strongest relations in a possibly very large data set.

Given two random variables, e.g. two time series, the data elements of each variable are rearranged in a descending/ascending order to get an ordered pair. For a finite set D of the ordered pair, the x-values and y-values of D are partitioned into x bins and y bins respectively (empty bins allowed). A pair of such partitions is named as an x-by-y grid. The maximum mutual information under each grid division is assigned to I^* by equation Eq.1 [6]:

$$I^*(D, G(b_1, b_2, \ldots, b_m)) = maxI(D/G) \tag{1}$$

where the maximum is identified across the whole G with x columns and y rows, and $I(D|G)$ denotes the mutual information of $D|G$.

The characteristic matrix of D is an infinite matrix with entries [6]:

$$M(D)_{x,y} = \frac{I^*(D, x, y)}{logmin\{x, y\}} \tag{2}$$

The *MIC* of a the original bi-variate data (sample size n and grid size less than $B(n)$) is given by [6]:

$$MIC(D) = \max_{xy<B(n)} \{M(D)_{x,y}\} \tag{3}$$

where $\omega(1) < B(n) \leq O\left(n^{1-\varepsilon}\right)$ for some $0 < \varepsilon < 1$. In this paper we use $B(n) = n^{0.6}$.

MIC is a positive real value with the following properties [5]:

1) Boundness, all entries of the characteristic matrix fall between 0 and 1;
2) Symmetry, the characteristic matrix remains the same when the x- and y-values of D are interchanged;
3) Invariant, the characteristic matrix is invariant under order-preserving transformations of the x- and y-values of D since the distribution $D|G$ depends only on the rank-order of the data.

The *MIC* measure can only indicate the synchronization strength of bivariate data. For an EEG dataset consisting of M channels, apparently $\frac{M\times(M-1)}{2}$ *MIC* measures should be calculated corresponding to all channel pairs.

C. CORRELATION MATRIX BASED ON MAXIMAL INFORMATION COEFFICIENT

This study extends the *MIC* measure to quantify the global synchronization of multivariate EEG, which combines *MIC* with a correlation matrix, i.e. Correlation matrix based on *MIC* (CMMIC). *CMMIC* can be formulated as Eq. 4.

$$CMMIC = \begin{bmatrix} MIC_{11} & MIC_{12} & \cdots & MIC_{1n} \\ MIC_{21} & MIC_{22} & \cdots & MIC_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ MIC_{n1} & MIC_{n2} & \cdots & MIC_{nn} \end{bmatrix} \quad (4)$$

where $MIC_{ij}(i, j = 1, \ldots, n)$ denotes the synchronization strength between channels i and j. As determined by the properties of *MIC*, *CMMIC* is a positive definite matrix: $MIC_{ij} \geq 0 \&\& MIC_{ii} = 1$. The trace value of *CMMIC* is equal to the number of data channels. An identify matrix will result IFF all channels are totally independent of each other, which is obviously very rare.

1) $\lambda \geq 0$
2) $p = \sum_{i=1}^{N} \lambda_i = tr(CMMIC) = \sum_{i=1}^{N} MIC_{ii} = \#Channels$

Each *CMMIC* is an instance of a synchronization pattern at a time point (or over a time slot) of the EEG. A *CMMIC* can be illustrated as a $N \times N$ symmetric image as shown in Fig. 2. A sequence of CMMICs in time order represents the evolving synchronization patterns (Section V-B).

IV. LIGHTWEIGHT *VGGNet* FOR EEG CLASSIFICATION

The *CMMIC* sequence is then processed and classified by the lightweight VGGNet. As EEG normally has a low spatial resolution, an excessively deep convolutional network does not apply to classify CMMICs. This section first describes the architecture of the *VGGNet* network and then details the parameter settings.

A. ARCHITECTURE OF THE LIGHTWEIGHT VGGNet

Fig. 3 illustrates the architecture of the lightweight VGGNet, which attempts to exploit as few layers as possible while gaining a high accuracy of classification. The *VGGNet* begins with a standalone *dropout* layer, followed by five *convolutional* layers with the same configurations and three *fully − connected* (FC) layers. ReLU activation function is adopted in all weighted layers (except the *dropout*). The final ReLU activation function of the *VGGNet* classifies the synchronization patterns [18], and outputs the final results of identification of the particular EEG segment. The details of the *dropout* technique and the pooling layer are as follows:

- "Dropout" aims to solve the overfitting problem by randomly dropping units from the neural network during training. Dropping out 20% of the input units and 50% of the hidden units was suggested in [19]. This study sets the dropout ratio as an empirical value 0.1 through a large number of experiments to avoid overfitting (Section IV-C).
- A pooling layer represents an area ($s \times s$) around a given location as an element (e.g. maximum of all elements in the area) and is useful in reduction of model parameters in image/video analysis. However, it will cause significant information loss when dealing with CMMICs as the latter have a low spatial resolution. **Unlike images or videos, the data elements in the feature matrix exhibit little continuity. The values of neighbor date elements can be significantly different as illustrated in Fig. 2.**

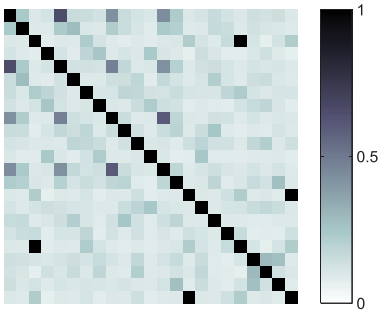

FIGURE 2. A *CMMIC* illustrated as a gray image. The value of each pixel represents the *MIC* measure of a pair of EEG channels.

B. BASIC PARAMETER SETTINGS OF THE VGGNet

Parameter Settings of the *VGGNet* are described in Table 1 with the number of parameters reported in the rightmost column. The overall parameter set (50,168) in *VGGnet* is in general much smaller than existing deep CNN models (see Section V for performance evaluation). The lightweight *VGGNet* differs from other VGG variants in: (1) five convolutional layers with the same configuration (*Convo2D*(2, (3, 3), 2 is the number of filters and (3,3) is the size of receptive fields) and (2) removal of pool layers. The receptive field is set small enough: 3×3 [20], which aims to convolve each *MIC* with the nearest neighbours only.

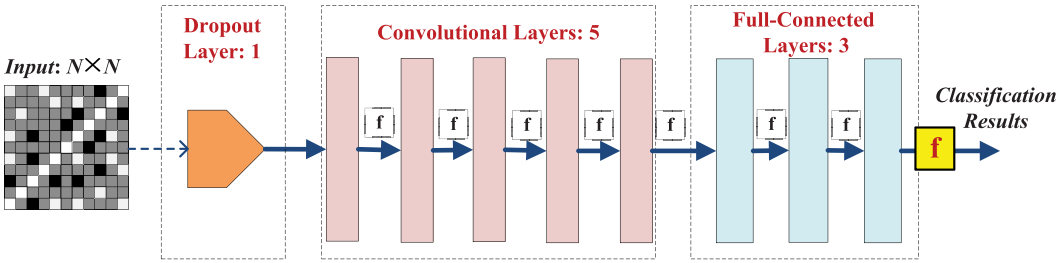

FIGURE 3. The architecture of the lightweight VGGNet. The *f* in the figure denotes the ReLU activation function.

TABLE 1. Parameter Settings of the VGGNet. The convolutional layer and activation output parameters are denoted as "[samples (2D-shape size) filters]". The ReLU activation function is omitted in the table for simplification.

Layer (type)	Output Shape	Parameter
dropout-1 (Dropout)	[None (23 23) 1]	0
conv2d-1 (Conv2D)	[None (21 21) 2]	20
conv2d-2 (Conv2D)	[None (19 19) 2]	38
conv2d-3 (Conv2D)	[None (17 17) 2]	38
conv2d-4 (Conv2D)	[None (15 15) 2]	38
conv2d-5 (Conv2D)	[None (13 13) 2]	38
flatten-1 (Flatten)	[None 338]	0
dense-1 (Dense)	[None 125]	42375
dense-2 (Dense)	[None 60)	7560
dense-3 (Dense)	[None 1)	61

C. CLASSIFICATION

This subsection first details the training of the *VGGNet* model and the testing over the trained model. The strategy for avoidance of the overfitting problem is then covered.

1) MODEL TRAINING AND TESTING

The lightweight *VGGNet* is trained using SGD. This study applies a very small weight decay to keep the model's training error low [21]. Weight initialization is performed conforming to that proposed in [22] and batch normalization is applied to the network [20]. The *objective* is to minimize the mean squared error in the *VGGNet*. The *VGGNet* processes the CMMICs (see Section III-C) as the initial inputs in model training.

After shuffling the whole sample space, CMMICs are divided into training sets, validation sets and test sets. A 5-fold cross validation algorithm is employed to evaluate the training performance of the classifier with training and validation sets. The performance of classification is reported with the test sets. During model training, for each layer:

- The forward propagation algorithm uses the outputs, weights and bias of the previous layer as the independent variables of the current activation function;
- The mean squared error is calculated based on the current outputs;
- The weights and bias of the previous layers are updated through a back propagation algorithm.

The above steps repeat until a steady state is reached, and the final training performance of classification can be evaluated.

The training is carried out by SGD optimizer using mini-batch (size: 50) gradient descent based on back-propagation with momentum (0.9) [21]. The training is regularized by weight decay (1e-4) and dropout (0.1). The update rule for weight following Eq. 5 [21]:

$$\begin{aligned} v_{i+1} &\leftarrow 0.9 \cdot v_i - 0.0001 \cdot \epsilon \cdot \omega_i - \epsilon \langle \frac{\partial L}{\partial \omega} |_{\omega_i} \rangle_{D_i}, \\ \omega_{i+1} &\leftarrow \omega_i + v_{i+1} \end{aligned} \tag{5}$$

where i is the iteration index, v is the momentum variable, ϵ is the learning rate, and $\langle \frac{\partial L}{\partial \omega} |_{\omega_i} \rangle_{D_i}$ is the average over the ith batch D_i of the derivative of the objective with respect to ω, evaluated at ω_i.

After the *VGGNet* model is trained, testing can be performed on the test sets (or new EEGs from other subjects). Here, the same parameter settings apply without the need for parameter update. After the input goes through the *dropout* layer and the five *convolutional* layers, the intermediate matrix will be flattened to a vector (with size of 338 at the *flatten* layer, see Table 1). The vector is passed through to the last three dense (FC) layers with outputs with sizes 125, 60 and 1 respectively. Finally, the state of each EEG segment (one *CMMIC*) is associated with can be identified (Seizure or Non-Seizure).

2) AVOIDANCE OF OVERFITTING

Two strategies are adopted to reduce overfitting of the *VGGNet* model: early stopping and dropout. In this study, the validation accuracy is monitored continuously until it stops ascending (patience: 10). The iteration of training will then stop on completion of the current epoch. Taking our experiments for example, the number of epochs was initially set to 300 while the iteration stopped at the 67th epoch (Section V).

The other strategy is "dropout", which temporarily drops units together with their connections at random from the neural networks during training. The central idea of dropout is to take a large model that overfits easily and repeatedly sample and train smaller sub-models from it. This prevents units from co-adapting too much on training. At the test stage, it can approximate the effect of averaging the predictions of all these sub-models by simply using a single unthinned model that has smaller weights, thus overfitting can be prevented in a simple manner at the cost of double training time [19].

V. EXPERIMENTS AND RESULTS

Experiments have been carried out to evaluate the performance of the proposed method. Experimental results are reported in terms of both synchronization dynamics and pattern classification. The testbed is a desktop with Intel i7 CPU (3.33GHz) and 24GB memory running 64bit Windows 7. The experiments concern both off-line training and on-line classification.

Off-Line Training: This procedure includes (1) calculation of all CMMICs and (2) training the neural network models. The bottleneck of step one is with *MIC* calculation, but it can be computed in a massively parallel manner to minimize the overhead [23]. As for the dataset (Section V-A) in this study, the model can be trained in a couple of minutes on the completion of step two.

On-Line Classification: This procedure includes (1) calculation of one *CMMIC* for evaluation and (2) state prediction based on the model from the last procedure, which takes less than 0.01 second.

A. DATA DESCRIPTION

The CHB-MIT scalp EEG dataset is used for this study (publicly authorized for open access). The dataset consists of EEG recordings from 22 patients (5 males, ages 3 - 22; 17 females, ages 1.5 - 19) with severe epilepsy caused by organic lesions, which were recorded simultaneously through 23 difference channels (FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FZ-CZ, CZ-PZ, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, P7-T7, T7-FT9, FT9-FT10, FT10-T8, and T8-P8) in 256Hz with 19 electrodes and a ground attached to the surface of the scalp. Most recordings contain multiple seizure occurrences.

This study investigates the EEG recordings with the same number of channels (from 18 patients). To avoid the problems of imbalanced samples, MCMC [24] sampling was used to balance the seizure states and non-seizure state samples:

- For each Epileptic seizure stage with size $S(seizure)$, denote *CMMIC* counts for seizure as $count(seizure) = \lfloor S(seizure)/S(window) \rfloor$, where $S(window)$ is the size of the time window.
- Denote *CMMIC* counts for non-seizure stage prior to epileptic seizure stage as $count(prior) = \lfloor \frac{1}{2} \times S(seizure)/S(window) \rfloor$.
- Denote *CMMIC* counts for non-seizure stage posterior to epileptic seizure stage as $count(posterior) = count(seizure) - count(prior)$.

B. EXPERIMENTS ON SYNCHRONIZATION DYNAMICS

Measurement of the evolution of relations among synchronization patterns (CMMICs) is an effective means to understand the roles of different data channels (i.e. brain regions). This study analyzes the change of synchronization strength in different channel pairs during seizure using the Apriori algorithm. The support degree tries to find distinct variation on synchronization measures between the seizure states and non-seizure states. The confidence degree tries to answer which interaction among synchronization features leads to epileptic seizure.

1) TOP-5 AND SUPPORT DEGREE

In the context of the Apriori algorithm, *support degree* ($support(A \rightarrow B) = P(A \cap B)$) denotes the probability of A and B simultaneously. The more frequently A and B appear simultaneously, the greater the association between A and B is. Synchronization dynamics of epileptic EEG is non-stationary in nature. **Support degree** is computed to probe the relations amongst the **MIC time series** obtained in the previous step to better understand the synchronization dynamics in connection with seizures.

The experimental results indicate that about 30% of synchronization between channels shows a decrease, while the others show an increase($about 58\%$) or invariant($about 12\%$) on seizures.

For the top five of all channel pairs, (i.e. $[< C4 - P4, FP2-F8 >, 82.4\%]$, $[< FZ-CZ, FP2-F4 >, 81.3\%]$, $[< FP2 - F4, T8 - P8 >, 80.2\%]$, $[< FP2 - F8, FT9 - FT10 >, 78\%]$, $[< P4 - O2, F8 - T8 >, 78\%]$), their *MIC*s increase on seizures with *support degree* of 47.3%.

In contrast, for other the top five pairs, (i.e., $[< T7 - P7, T8 - P8 >, 57.1\%]$, $[< C3 - P3, C4 - P4 >, 56\%]$, $[< P7 - T7, FT9 - FT10 >, 50.5\%]$, $[< P3 - O1, FT9 - FT10 >, 49.5\%]$, $[< T7 - P7, C3 - P3 >, 49.5\%]$), their *MIC*s decrease on seizures with *support degree* of 12.1%.

The results indicate that on seizures the probability of increase of synchronization strength is much higher than that of decrease cases.

2) TOP-5 AND CONFIDENCE DEGREE

Confidence Degree ($Confidence(A \rightarrow B) = P(B|A)$) is the probability of B in condition with A. If the confidence degree is 100%, then A and B can be bundled with the strongest association; Otherwise, a small value means that there is no obvious association between A and B. The confidence degrees between the top five increased channel-pairs are shown in Table 2, and the top five decreased channel-pairs are shown in Table 3.

The results of confidence degree show that (1) the top five channel pairs with *MIC* increase on seizures are likely to evolve in a similar manner in terms of synchronization; (2) those with *MIC* decrease do not exhibit this feature.

3) GLOBAL SYNCHRONIZATION STRENGTH

Fig. 4 presents the synchronization of Top-5 distinct variation on synchronization measures between the seizure states and non-seizure states. The values in seizure states are greater than those in non-seizure states.The synchronization property changes significantly from non-seizure states to seizure states and vice versa. The global synchronization matrices of average seizure features, normal features and their subtraction features are shown in Fig. 5. The negative values will be displayed in **white** in the subtraction features matrix.

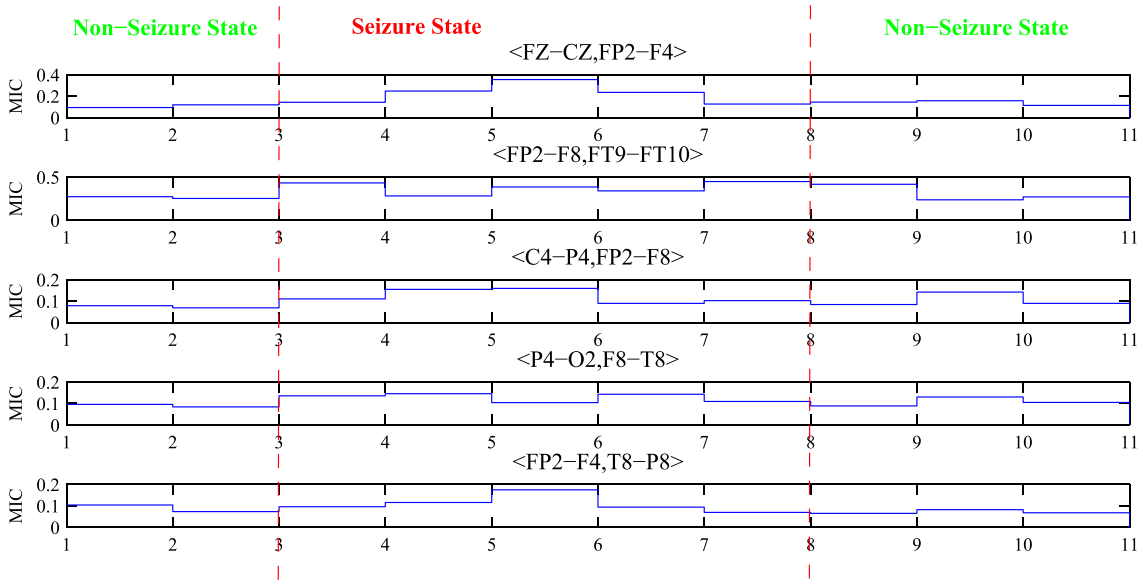

FIGURE 4. Synchronization of Top-5 distinct variation on synchronization measures between seizure states and non-seizure states for one instance of seizure onset. The X-axis denotes the time (8 seconds per unit), Y-axis denotes the synchronization value, each curve is marked with the index of channel pair. A non-seizure state is separated from seizure states by red dashed lines.

TABLE 2. Confidence degrees (last column) between TOP-5 ascent channel-pairs.

A	B	Conf $(B \rightarrow A)\%$
<FP2-F8, FT9-FT10 >	<FP2-F4, T8-P8 >	80.83
<P4-O2, F8-T8 >	<FZ-CZ, FP2-F4 >	81.09
<FP2-F4, T8-P8 >	<C4-P4, FP2-F8 >	81.34
<FZ-CZ, FP2-F4 >	<C4-P4, FP2-F8 >	81.34
<FP2-F8, FT9-FT10 >	<FZ-CZ, FP2-F4 >	82.44
<C4-P4, FP2-F8 >	<FZ-CZ, FP2-F4 >	82.44
<P4-O2, F8-T8 >	<FP2-F8, FT9-FT10 >	83.1
<FP2-F8, FT9-FT10 >	<P4-O2, F8-T8 >	83.1
<FP2-F4, T8-P8 >	<FP2-F8, FT9-FT10 >	83.1
<C4-P4, FP2-F8 >	<FP2-F4, T8-P8 >	83.57
<FP2-F4, T8-P8 >	<FZ-CZ, FP2-F4 >	83.79
<P4-O2, F8-T8 >	<C4-P4, FP2-F8 >	84.0
<FZ-CZ, FP2-F4 >	<P4-O2, F8-T8 >	84.51
<FZ-CZ, FP2-F4 >	<FP2-F4, T8-P8 >	84.94
<FZ-CZ, FP2-F4 >	<FP2-F8, FT9-FT10 >	85.92
<P4-O2, F8-T8 >	<FP2-F4, T8-P8 >	86.31
<FP2-F8, FT9-FT10 >	<C4-P4, FP2-F8 >	86.67
<FP2-F4, T8-P8 >	<P4-O2, F8-T8 >	88.74
<C4-P4, FP2-F8 >	<P4-O2, F8-T8 >	88.74
<C4-P4, FP2-F8 >	<FP2-F8, FT9-FT10 >	91.55

TABLE 3. Confidence degrees (last column) between TOP-5 descent channel-pairs.

A	B	Conf$(B \rightarrow A)\%$
<C3-P3, C4-P4 >	<T7-P7, C3-P3 >	47.83
<T7-P7, T8-P8 >	<P3-O1, FT9-FT10 >	47.83
<C3-P3, C4-P4 >	<P3-O1, FT9-FT10 >	50.0
<T7-P7, T8-P8 >	<T7-P7, C3-P3 >	50.0
<C3-P3, C4-P4 >	<P7-T7, FT9-FT10 >	51.12
<T7-P7, T8-P8 >	<P7-T7, FT9-FT10 >	53.34
<T7-P7, C3-P3 >	<C3-P3, C4-P4 >	55.01
<P3-O1, FT9-FT10 >	<T7-P7, T8-P8 >	56.42
<P7-T7, FT9-FT10 >	<T7-P7, C3-P3 >	56.53
<P3-O1, FT9-FT10 >	<C3-P3, C4-P4 >	57.5
<P7-T7, FT9-FT10 >	<C3-P3, C4-P4 >	57.5
<T7-P7, C3-P3 >	<P7-T7, FT9-FT10 >	57.78
<T7-P7, C3-P3 >	<T7-P7, T8-P8 >	58.98
<T7-P7, C3-P3 >	<P3-O1, FT9-FT10 >	60.87
<P3-O1, FT9-FT10 >	<T7-P7, C3-P3 >	60.87
<P7-T7, FT9-FT10 >	<T7-P7, T8-P8 >	61.54
<T7-P7, T8-P8 >	<C3-P3, C4-P4 >	62.51
<C3-P3, C4-P4 >	<T7-P7, T8-P8 >	64.11
<P7-T7, FT9-FT10 >	<P3-O1, FT9-FT10 >	71.74
<P3-O1, FT9-FT10 >	<P7-T7, FT9-FT10 >	73.34

The **darker** color denotes the higher synchronization measurement. On average, the global synchronization of seizure state is greater than that of non-seizure state.

The results from the above experiments indicate that characterization of synchronization dynamics can provide useful information to differentiate seizure states from the rest.

C. EVALUATION OF CLASSIFICATION PERFORMANCE

The lightweight *VGGNet* is trained using SGD for 300 epochs on CHB-MIT with mini-batch size of 50. The learning rate is set to 0.01. This study applies a weight decay of 1e-4, momentum of 0.9 and Nesterov momentum [21]. Weight initialization is performed conforming to that proposed in [22] and batch normalization is applied to the network [20]. Dropout rate is set as 0.1.

After being shuffled with random seed of 7, the data are divided into training sets, validation sets and test sets, which occupy 64%, 16% and 20% respectively. In the training phase, a 5-fold cross validation algorithm is employed to evaluate the training performance of lightweight *VGGNet* with training sets and validation sets. That is, all *CMMIC*s are divided into 5 fold by shuffling with 5 iterations performed. In each iteration, 4 fold are trained, and the remaining fold is used for validation. The final result is the average

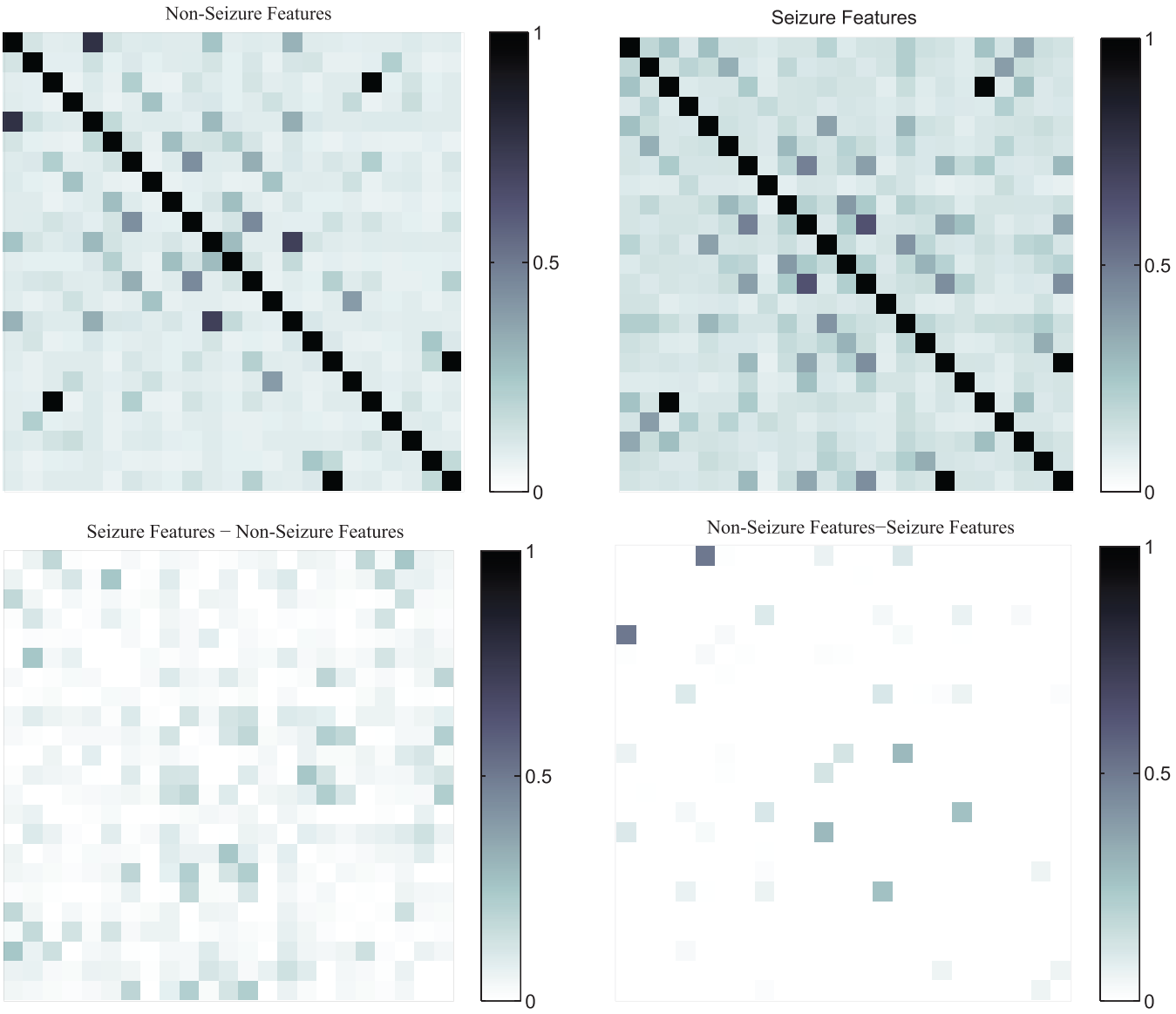

FIGURE 5. The global synchronization matrices of average seizure features, non-seizure features, and their subtraction features between a pair of EEG data.

of the outputs of 5 iterations. The results are reported in terms of sensitivity(*SEN*), specificity(*SPE*), accuracy(*ACC*), Precision and Recall. *SEN* and *SPE* describe the rate of correctly detecting seizure states and non-seizure states, respectively. *ACC* denotes the average performance of the classifier. *Precision* calculates the proportion of all correctly detected seizure onsets from all that were actually classified. *Recall* calculates the proportion of all correctly detected seizures from all correctly detected seizures and negative normals. After the classifier was trained, the performance was reported according to the testing sets.

Setting of time windows can affect the performance of the classifier. Fig. 6 shows a box chart of the classification performance with respect to segmentation. As the size of the time window increases (starting from 512), *SEN*, *SEP*, and *ACC* increase almost linearly with point of inflexion as size 768 and 1000 and then increase after that. The box height indicates the amount of variance, which shows the stationarity of the classification performance. With a window size of 2048 (8 seconds), the accuracy, sensitivity, and specificity reach the peak: [98.13% $\pm$ 0.24%], [98.85% $\pm$ 0.51%], and [97.47% $\pm$ 0.36%], respectively. This setting is then applied to all other experiments. The variance of most results is small, which indicates that the performance of lightweight *VGGNet* is relatively stable.

Figure 7 shows the accuracy and loss metrics for the training and validation processes. Here, *acc* and *loss* indicate the accuracy and error in training, respectively; *val_acc* and *val_loss* indicate the accuracy and error in validation, respectively. Obviously, overfitting does not occur in training stage as (1) *acc* and *val_acc* are high at the same time, and (2) no significant difference exists between *acc* and *val_acc* in all

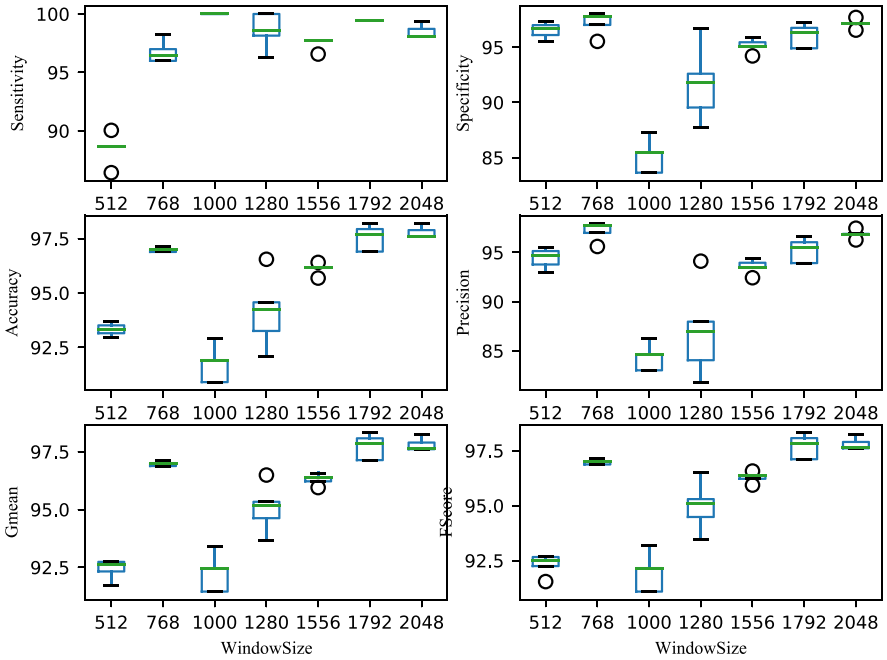

FIGURE 6. Relationship between performance (accuracy, sensitivity, specificity, precision, GMean and FScore) and window size. The Y-labels in each sub figure illustrate the related performance index and X-labels show the window size.

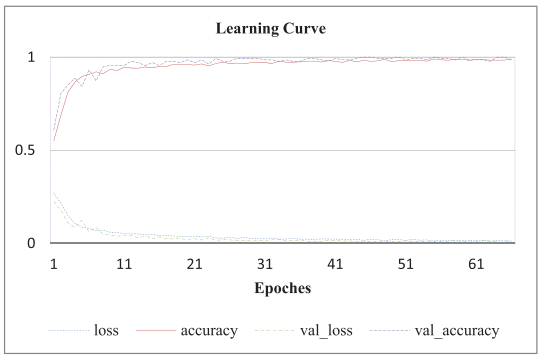

FIGURE 7. Accuracy and Loss rates in the training and validating processes.

TABLE 4. Performance Comparison. *SEN* and *SPE* describe the rate of correctly detecting seizure states and non-seizure states, respectively. *ACC* denotes the accuracy of classification. PK (A Priori Knowledge) shows the dependence of the approach on a priori knowledge.

Author Year	Classifier	Sens	Spec	Acc	PK
Fergus 2016[16]	k-NN	88	88	93	Y
Nasehi 2013[26]	IPSONN	98	-	-	Y
Morteza 2016[17]	MLP, Bayesian	86.53	97.27	86.56	Y
Lorena 2016[15]	LDA,NN	97.5	99.9	-	Y
Our Approach	Lightweight VGGNet	98.85	97.47	98.13	N

iterations. Results also indicate that overfitting does not occur in this case.

The area under the Receiver Operating Characteristic (ROC) curve, denoted as AUC, measures the performance of supervised classification rules. A satisfactory classification rule is reflected by an ROC curve which lies in the upper left triangle of the square. That is, it is above the counter-diagonal (the luck line in the left of Fig. 8) [25]. The ranking performance is promising when the AUC value is high. *Precision* and *recall* rates are of mutual influence, both of which will certainly be high in the ideal situation. However, in general, when the *accuracy* is high, the *recall* rate will be low, and vice versa. It is desirable that the Precision-Recall Curve is above the principal diagonal (the luck line on the right of Fig. 8). Fig. 8 shows the ROC Curve (left) and the related P-R curve (right) to evaluate the performance of the proposed *VGGNet* model. The figure illustrates the ranking performance on the k-fold cross validation (5-fold in this paper). The convex ROC/PR curve and the high AUC (both are 0.99) exhibit the excellent classification performance of the *VGGNet*.

A comparison between the proposed approach and the state-of-the-art especially including those with intelligent algorithms is presented in Table 4. The proposed approach achieves the highest **sensitivity** and **accuracy** over the same dataset CHB-MIT. Its performance is always superior except the *SPE* in [15]. Nevertheless, *SEN* is a much more critical indicator as it denotes whether seizures can be correctly detected.

SEN reflects the capability of the classifier to correctly identify an epileptic seizure (*SPE* for non-seizures). High *SEN* and *SPE* values are both desired. The box chart of classification performance with respect to segment size is shown in Fig. 6. Besides the above, indices including *GMEAN* and $F_1 - Score$ are measured to evaluate the capability of

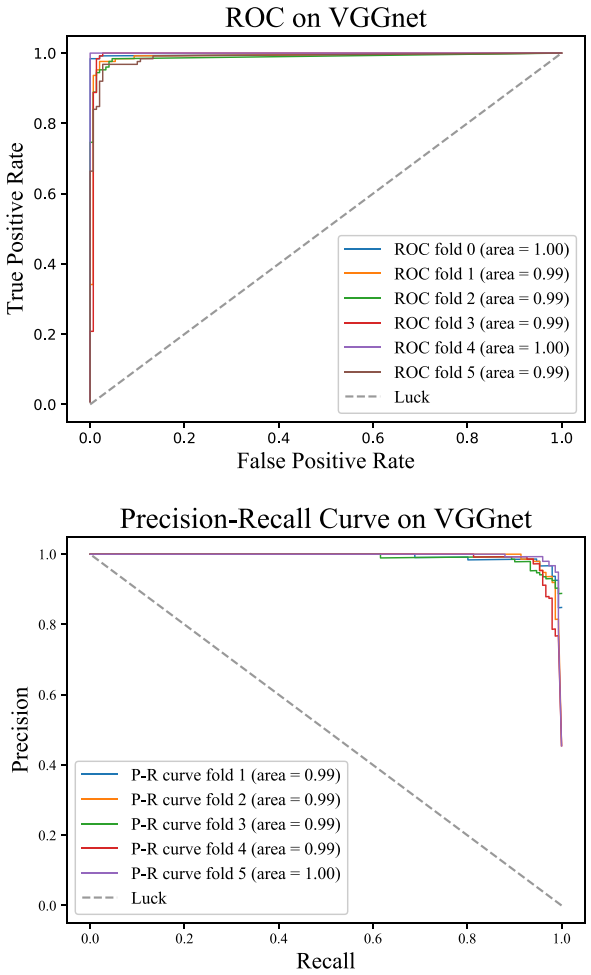

FIGURE 8. **ROC Curve (Top) & P-R curve (Button) on the *VGGNet*. Each curve denotes one fold of the 5-fold cross validation.**

the approach to detect both seizure and non-seizure states regardless of the percentage each state may exist in the whole dataset [17], thus the false alarm rate can be limited low:

$$Gmean = \sqrt{(Sensitivity \times Specificity)} \tag{6}$$

$$F_1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

Latency can be calculated to show the delay between the time point where the classifier detects a seizure activity and that marked by the expert:

$$\begin{aligned} Latency &= E\left(Distance_{onset}\right) \\ &= \sum_{i=1}^{N} P(i) * Distance^{i}_{onset} \\ &= \frac{1}{2} S(window) \end{aligned} \tag{8}$$

where $Distance_{onset}$ is the distance between start point of the time window and seizure onset marked by the expert; $S(window)$ is the size of time window. For the time window setting in this study (8 seconds), the latencies span from 1 to 4 seconds.

D. DISCUSSIONS

1) ADVANTAGES OF DEEP NEURAL NETWORKS

The latest neural networks (*NN*) are highly suited for EEG classification as they afford (1) Non-linearity: A *NN* consists of interacting neurons (linear or non-linear) and exhibits intensive non-linearity, (2) Adaptivity: A *NN* has the inherent ability to adjust the synaptic weights to adapt to the dynamics of the external environment such as arbitrary pattern change, (3) Fault Tolerance: When a part of a *NN* encounters a problem, the rest of the network will still function with the problem well contained, e.g. handling a segment contaminated with intensive interferences, and (4) convolutional neural networks (*CNN*) can adaptively select features [27].

2) NUMBER OF LAYERS

Although deep *NN* is widely adopted, an EEG classifier's performance does not necessarily rely on the number of layers. As spotted in [28], a smaller *CNN* architecture (e.g. SqueezeNet) can achieve the same accuracy as an extreme deep *NN* does, and it has merits in: (1) more efficient distributed training, (2) less overhead, and (3) easy deployability on embedded platforms with limited resources. Moreover, an extremely deep *NN* may suffer vanishing/exploding gradients and degradation problems.

3) GENERALITY

Most existing works on seizure detection and prediction have focused on patient-specific predictors with strong dependence on a priori knowledge of the patient [17], which demands either the sample should be trained and tested for the same patient or use manually set feature extraction rules on each specific patient relying on experts. As a contrast, this study uses samples from all patients under investigation, based on which a general EEG classification model is fostered to accurately detect seizure states of different subjects.

4) INDEPENDENCE OF A PRIORI KNOWLEDGE OF FREQUENCY FEATURES

Conventional classification approaches rely on time, frequency and spatial analysis of EEG [2]. The frequency bands should be customized for a particular patient, and identification of a set of suitable frequency bands itself is already a research challenge and makes it very difficult for a classification model to be generalized to different patients. Some approaches have been developed to address this problem, such as the ones upon Bayesian framework [29]. However, suitable frequency bands may only be achieved when a very large amount of EEG epochs are processed with complicated algorithms. Another problem is the size of time windows has to be long enough to avoid the risk of losing useful frequency information. For example, Piotr et al. had to form 1 or 5 min-long patterns of 12 or 60 frames to get frequency field information while bivariate synchronization was computed using 5s time windows [2]. To the best of our knowledge, existing classifiers need sufficient a priori knowledge as stated in the above. The proposed approach requires no a priori knowledge

at all. Furthermore, existing methods largely demand preprocessing of epileptic EEG to remove intensive noise/artifacts while it is not necessary for the proposed approach.

5) POTENTIAL IN BRAIN BIG DATA APPLICATIONS

The time complexity of the neural network is proportional to the product of the number of hidden neurons (N) and layers (L). As a small CNN framework, the lightweight *VGGNet* can achieve a time complexity as low as $O(L)$ given that the computation on each layer is propelled by cutting-edge GPUs and/or FPGAs. With the increase of patient sample size, incremental training samples can be used to update the model parameters via incremental online analysis and a more accurate model can be obtained [30], [31]. The overhead of processing the new CMMICs with the trained *VGGNet* model can be almost ignored, which makes it particularly suited for massive EEG identification while this persists as an onerous task for conventional counterpart approaches. The parallel *MIC* $++$ approach can reduce the time complexity of *MIC* synchronization measurement to $O(log_2(N))$ [23]. Using the latest high performance cyberinfrastructure [32]–[34], the model can be trained can be performed in a near-real-time manner. The proposed classification approach on the *CMMIC* sequence is naturally suitable for distributed classification using a model parallel to each machine (mapping onto GPUs) [32], [35] and/or parallel data processing over many compute nodes [36].

VI. CONCLUSIONS

It is an important issue to find synchronization patterns in multivariate EEG superimposed with intensive noise and accurately to classify them on this basis under the circumstance of insufficient a priori knowledge. Such capability can significantly benefit brain dysfunction research and practices, e.g. epilepsy.

This study extended the *MIC* method to measure global synchronization of multivariate EEG. The global *MIC* measures (*CMMIC*s) have been organized in time sequence to represent the evolving synchronization patterns. *CMMIC*s maintain abundant useful information to differentiate seizure states from the rest. A lightweight *VGGNet* is then designed to adaptively characterize the non-stationary patterns related to seizures and then classify them. The design alleviates the vanishing gradient problem and strengthens feature propagation, which leads to a substantial reduction of parameters.

Experiments have been performed to evaluate the proposed approach over the CHB-MIT scalp EEG dataset. The results show an improvement relative to existing methods, with accuracy, sensitivity, and specificity of [98.13% $\pm$ 0.24%], [98.85% $\pm$ 0.51%], and [97.47% $\pm$ 0.36%], respectively. The variance of most results is small, which indicates that the performance of the *VGGNet* is relatively stable.

The proposed approach achieves this performance without the need for denoising the EEG. Furthermore, the approach requires only one hyperparameter, which avoids the potential errors caused by excessive parameter settings. The overall work enables a general and cost-effective solution to classification of EEG and holds great potential for handling brain big data.

REFERENCES

[1] E. Gysels, "Phase synchronization for classification of spontaneous EEG signals in brain-computer interfaces," M.S. thesis, EPFL, Lausanne, Switzerland, Jan. 2005, doi: 10.5075/epfl-thesis-3397.
[2] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky, "Classification of patterns of EEG synchronization for seizure prediction," *Clin. Neurophysiol.*, vol. 120, no. 1, pp. 1927–1940, 2009.
[3] D. Cui *et al.*, "A new EEG synchronization strength analysis method: S-estimator based normalized weighted-permutation mutual information," *Neural Netw.*, vol. 82, pp. 30–38, Oct. 2016.
[4] D. Chen, X. Li, D. Cui, L. Wang, and D. Lu, "Global synchronization measurement of multivariate neural signals with massively parallel nonlinear interdependence analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 1, pp. 33–43, Jan. 2014.
[5] J. D. Bonita *et al.*, "Time domain measures of inter-channel EEG correlations: A comparison of linear, nonparametric and nonlinear measures," *Cognit. Neurodyn.*, vol. 8, no. 1, pp. 1–15, 2014.
[6] D. N. Reshef *et al.*, "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, 2011.
[7] C. Carmeli, M. G. Knyazeva, G. M. Innocenti, and O. D. Feo, "Assessment of EEG synchronization based on state-space analysis," *NeuroImage*, vol. 25, no. 2, pp. 339–354, 2005.
[8] D. Cui, X. Liu, Y. Wan, and X. Li, "Estimation of genuine and random synchronization in multivariate neural series," *Neural Netw.*, vol. 23, no. 6, pp. 698–704, 2010.
[9] A. J. C. Slooter *et al.*, "Seizure detection in adult ICU patients based on changes in EEG synchronization likelihood," *Neurocrit. Care*, vol. 5, no. 3, pp. 186–192, Dec. 2006.
[10] M. Le van Quyen *et al.*, "Preictal state identification by synchronization changes in long-term intracranial EEG recordings," *Clin. Neurophysiol.*, vol. 116, no. 3, pp. 559–568, 2005.
[11] E. Gysels, P. Renevey, and P. Celka, "SVM-based recursive feature elimination to compare phase synchronization computed from broadband and narrowband EEG signals in brain–computer interfaces," *Signal Process.*, vol. 85, no. 11, pp. 2178–2189, 2005.
[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representat.*, San Diego, CA, USA, 2015, pp. 1–14.
[13] A. L. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2010.
[14] M. H. Myers, A. Padmanabha, G. Hossain, A. L. de Jongh Curry, and C. D. Blaha, "Seizure prediction and detection via phase and amplitude lock values," *Front. Hum. Neurosci.*, vol. 10, p. 80, Mar. 2016.
[15] L. Orosco, A. G. Correa, P. Diez, and E. Laciar, "Patient non-specific algorithm for seizures detection in scalp EEG," *Comput. Biol. Med.*, vol. 71, pp. 128–134, Apr. 2016.
[16] P. Fergus, A. Hussain, D. Hignett, D. Al-Jumeily, K. Abdel-Aziz, and H. Hamdan, "A machine learning system for automated whole-brain seizure detection," *Appl. Comput. Inform.*, vol. 12, pp. 70–89, Jan. 2016.
[17] M. Behnam and H. Pourghassem, "Real-time seizure prediction using RLS filtering and interpolated histogram feature based on hybrid optimization algorithm of Bayesian classifier and hunting search," *Comput. Methods Programs Biomed.*, vol. 136, pp. 115–136, Aug. 2016.
[18] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8609–8613.
[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, vol. 37. Jul. 2015, pp. 448–456.
[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, 2012.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1502. Feb. 2015, pp. 1026–1034.
[23] C. Wang, X. Li, A. Wang, and X. Zhou, "Brief announcement: MIC++: Accelerating maximal information coefficient calculation with GPUs and FPGAs," in *Proc. 28th ACM Symp. Parallelism Algorithms Archit.*, New York, NY, USA, 2016, pp. 287–288.
[24] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer, 2004.
[25] D. J. Hand and R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," *Mach. Learn.*, vol. 45, no. 2, pp. 171–186, 2001.
[26] S. Nasehi and H. Pourghassem, "Patient-specific epileptic seizure onset detection algorithm based on spectral features and IPSONN classifier," in *Proc. Int. Conf. Commun. Syst. Netw. Technol.*, Gwalior, India, Apr. 2013, pp. 186–190.
[27] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 512–519.
[28] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer. (2016). "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size." [Online]. Available: https://arxiv.org/abs/1602.07360
[29] H.-I. Suk and S.-W. Lee, "A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 286–299, Feb. 2013.
[30] L. Wang, Y. Ma, J. Yan, V. Chang, and A. Y. Zomaya, "pipsCloud: High performance cloud computing for remote sensing big data management and processing," *Future Generat. Comput. Syst.*, vol. 78, no. 1, pp. 353–368, 2018.
[31] X. Chen *et al.*, "Design automation for interwell connectivity estimation in petroleum cyber-physical systems," *IEEE Trans. Comput.-Aided Design Integr.*, vol. 36, no. 2, pp. 255–264, Feb. 2017.
[32] W. Song, Z. Deng, L. Wang, B. Du, P. Liu, and K. Lu, "G-IK-SVD: Parallel IK-SVD on GPUs for sparse representation of spatial big data," *J. Supercomput.*, vol. 73, no. 8, pp. 3433–3450, Aug. 2017.
[33] L. Wang, J. Zhang, P. Liu, K.-K. R. Choo, and F. Huang, "Spectral–spatial multi-feature-based deep learning for hyperspectral remote sensing image classification," *Soft Comput.*, vol. 21, no. 1, pp. 213–221, Jan. 2017.
[34] L. Wang, Y. Ma, A. Y. Zomaya, R. Ranjan, and D. Chen, "A parallel file system with application-aware data layout policies for massive remote sensing image processing in digital earth," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 6, pp. 1497–1508, Jun. 2014.
[35] W. Xue *et al.*, "Ultra-scalable CPU-MIC acceleration of mesoscale atmospheric modeling on Tianhe-2," *IEEE Trans. Comput.*, vol. 64, no. 8, pp. 2382–2393, Aug. 2015.
[36] L. Wang *et al.*, "Particle swarm optimization based dictionary learning for remote sensing big data," *Knowl.-Based Syst.*, vol. 79, pp. 43–50, May 2015.

HENGJIN KE is currently pursuing the Ph.D. degree with the computer school, Wuhan University. His main research interests include machine learning and bioinformatics.

DAN CHEN is currently a Professor with the Computer School, Wuhan University, Wuhan, China. He was an HEFCE Research Fellow with the University of Birmingham, U.K. His research interests include data science and engineering, neuroinformatics, and complex systems.

XIAOLI LI received the B.Eng. and M.Eng. degrees from the Kunming University of Science and Technology, and the Ph.D. degree from the Harbin Institute of Technology, China, all in mechanical engineering. He is currently a Professor and a Vice Director with the National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, China. His research interests include neural engineering, computational intelligence, signal processing and data analysis, monitoring system, and manufacturing system.

YUNBO TANG is currently pursuing the M.D. degree with the computer school, Wuhan University. His main research interests include machine learning and bioinformatics.

TEJAL SHAH received the Ph.D. degree from the School of Computer Science and Engineering, University of New South Wales, Australia. She is currently a Post-Doctoral Researcher with Newcastle University. The focus of her research is on the development and application of semantic web technologies for analyzing big data across various disciplines such as healthcare, remote sensing, and smart homes.

RAJIV RANJAN received the Ph.D. degree in computer science and software engineering from the University of Melbourne in 2009. He is currently a Reader with the School of Computing Science, Newcastle University, U.K., also a Chair Professor with the School of Computer, China University of Geosciences, Wuhan, China, and also a Visiting Scientist at Data61, CSIRO, Australia. His research interests include grid computing, peer-to-peer networks, cloud computing, Internet of Things, and big data analytics.