Optimized Deep Learning for EEG Big Data and Seizure Prediction BCI via Internet of Things

Mohammad-Parsa Hosseini, *Senior Member, IEEE*, Dario Pompili, *Senior Member, IEEE*, Kost Elisevich, and Hamid Soltanian-Zadeh, *Senior Member, IEEE*

Abstract—A brain-computer interface (BCI) for seizure prediction provides a means of controlling epilepsy in medically refractory patients whose site of epileptogenicity cannot be resected but yet can be defined sufficiently to be selectively influenced by strategically implanted electrodes. Challenges remain in offering real-time solutions with such technology because of the immediacy of electrographic ictal behavior. The nonstationary nature of electroencephalographic (EEG) and electrocorticographic (ECoG) signals results in wide variation of both normal and ictal patterns among patients. The use of manually extracted features in a prediction task is impractical and the large amount of data generated even among a limited set of electrode contacts will create significant processing delays. Big data in such circumstances not only must allow for safe storage but provide high computational resources for recognition, capture and real-time processing of the preictal period in order to execute the timely abrogation of the ictal event. By leveraging the potential of cloud computing and deep learning, we develop and deploy BCI seizure prediction and localization from scalp EEG and ECoG big data. First, a new method for epileptic seizure prediction and localization of the seizure focus is presented. Second, an extended optimization approach on existing deep-learning structures, Stacked Auto-encoder and Convolutional Neural Network (CNN), is proposed based on principle component analysis (PCA), independent component analysis (ICA), and Differential Search Algorithm (DSA). Third, a cloud-computing solution (i.e., Internet of Things (IoT)), is developed to define the proposed structures for real-time processing, automatic computing and storage of big data. The ECoG clinical datasets on 11 patients illustrate the superiority of the proposed patient-specific BCI as an alternative to current methodology to offer support for patients with intractable focal epilepsy.

Index Terms—BCI, EEG big data, cloud computing, deep learning, epilepsy, seizure prediction and localization

1 Introduction

One percent of the world's population suffers from epilepsy, a chronic disorder characterized by the occurrence of spontaneous seizures. About 30 percent of patients remain medically intractable and may undergo surgical intervention; despite the latter, some may still fail to attain a seizure-free outcome [2]. The recent introduction of a closed loop system of localized electroencephalographic (EEG) recording and stimulus delivery (i.e., RNS; Neuropace) has provided greater opportunity to achieve control of this entity although further solutions are required to better actuate the system for optimal efficacy and to bring about an improved quality of life for these patients.

Motivation. The use of computers to help physicians in the acquisition, management, storage, and reporting of EEG signals is well established. To this end, there are computer-aided detection applications that use Brain Computer Interface (BCI). In order for a BCI system to work effectively, computational algorithms must reliably identify periods of increased probability of an impending ictal occurrence in order to abort its development. Such preictal periods may be of variable duration and may not afford suitable latency to provide current methodologies with sufficient time for signal deployment to achieve control in all circumstances. The development of an automated method that delivers on such short notice would optimize seizure control and bring about an improved quality of life.

Vision. Technological innovation with BCI for control of epilepsy must acknowledge the immediacy of seizure occurrence and the time constraints imposed upon effective delivery. Generally, there are three main steps in an automatic BCI system. These include data collection, data processing by computer and electronics to apply the desired action. The EEG represents the brain's spontaneous electrical activity which is recorded using multiple electrodes spatially distributed over the scalp [3], [4]. It is necessary to confirm, electrographically, the presence of epileptogenicity and thus a diagnosis of epilepsy.

The temporal dynamics of brain activity can be categorized into four states. The interictal or baseline state presents

- *M.P. Hosseini and D. Pompili are with the Department of Electrical and Computer Engineering, Rutgers University, New Brunswick, NJ 08854. E-mail: {parsa, pompili}@cac.rutgers.edu.*
- *K. Elisevich is with the Department of Clinical Neurosciences, Spectrum Health, Grand Rapids, MI 49503 and with the Division of Neurosurgery, College of Human Medicine, Michigan State University, Grand Rapids, MI 49503. E-mail: Kost.Elisevich@spectrumhealth.org.*
- *H. Soltanian-Zadeh is with the Medical Image Analysis Laboratory, Departments of Radiology and Research Administration, Henry Ford Health System, Detroit, MI 48202 and Control and Intelligent Processing Center of Excellence (CIPCE), School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran 14399, Iran. E-mail: hsoltan1@hfhs.org, hszadeh@ut.ac.ir.*

Manuscript received 7 Mar. 2017; revised 7 Oct. 2017; accepted 17 Oct. 2017. Date of publication 2 Nov. 2017; date of current version 7 Dec. 2017.
(Corresponding author: Mohammad-Parsa Hosseini.)
Recommended for acceptance by J. Zhu, A.-A. Liu, M. Chen, T. Tasdizen, and H. Su.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TBDATA.2017.2769670

between seizures. The preictal state precedes the clinical seizure or ictal activity. The ictal state identifies the interval during which activity manifests as a seizure. Finally the postictal state occurs following the ictus. In general, the ictal or seizure state occurs when the brain assumes a synchronized pattern of neuronoglial activity. Clinically, these may manifest in a number of ways ranging from partial seizures with or without loss of consciousness and a local electrographic expression to generalized seizures that have a widespread expression within both cerebral hemispheres [5]. Seizure prediction methodologies must identify the preictal state sufficiently well to differentiate it from other states and with sufficient timing in order to launch an appropriate signal that interrupts the evolution of the ictus.

Challenges. Several challenges exist in creating a seizure-prediction system in real-time. The differentiation of the preictal from the interictal state of an individual is problematic of itself in that artefactual features during interictal recording may mimic preictal patterns. Second, the EEG does not generate a stationary signal and electrographic ictal patterns will vary across different patients. As a consequence, a standard set of manually-extracted features may not scale well in a population of epilepsy patients. Hence, supervised feature extraction will be insufficient for learning algorithms. The third challenge relates to the spatio-temporal dynamics of the electrographic pattern. The approach currently used clinically in seizure detection involves the strategic placement of a limited number of cerebrocortical surface and/or depth electrodes in the vicinity of the epileptogenic site. The orientation of these electrodes naturally varies from patient-to-patient. Although high spatio-temporal resolution and discrete electro-optical mapping of neuronoglial activity are made possible [6], [7], large amounts of spatially oriented data are generated over relatively brief durations leading to a big data problem. This situation calls for safe storage of a large archive and for high computational resources to process the data in real-time.

Our Approach. The requirements of a practical BCI system include methods for signal processing, machine learning, and electrographic brain-state prediction in large data sets collected from user populations in real-time. Accordingly, next generation BCI systems must be connected to high-performance computing servers in order to be able to adopt predictive models and execute computations in real-time for large incoming datasets. Cloud computing suits this purpose by providing a simple way to access databases and computational resources through the global Internet. The key benefit of the proposed BCI centers upon the analysis and learning allowed from large amounts of unsupervised data, making it a practical method for developing a real-time patient-based seizure prediction and localization system.

Our Contributions. To address the challenges of predicting ictal events, a cloud-based BCI system solution for the big data problem in epilepsy is introduced. A system model is proposed for seizure prediction and localization of the seizure focus from EEG. We have developed and extended existing deep learning structures with an optimization module based on Principal Component Analysis (PCA), an extended-ICA model (I-ICA), and Differential Search Algorithm (DSA). Specifically, these contributions include the development of the following novel techniques:

- Developing a cloud-based BCI system as Internet of Things (IoT) to provide high computational resources and safe storage for the big data problem generated by implanted electrodes.
- An extended Convolutional Neural Network (CNN) structure with an optimization module for unsupervised feature extraction from big data.
- An extended Stacked Auto-encoder structure with an optimization module for unsupervised feature extraction and classification of big data.
- A system model for epileptic seizure prediction and seizure focus localization from big EEG data.

The proposed system has the ability for pervasive data collection and analysis, which is useful in real-life support for epilepsy patients. To study accuracy and performance, the system is evaluated and compared to other methods on a benchmark epilepsy dataset. In sum, our original contributions in this study are in four fold. First, we have developed a cloud-based platform as an IoT framework for BCI. Second, we performed a survey and presented state-of-the-art theories for big data analysis via deep learning structures. Third, we extended deep leaning structures using new machine learning techniques. Forth, a novel application scenario related to biomedical big data analytics called epileptic seizure prediction is developed by benchmark datasets.

Outline. The remainder of this paper is organized as follows. In Section 2, we provide a literature review. In Section 3, we present a cloud computing framework as IoT for medical big data processing. In Section 4, we introduce our solution for seizure prediction and seizure focus localization from big EEG data followed by an extended approach for feature extraction and classification. Then, in Section 5, we discuss the proof-of-concept prototype of the proposed BCI seizure predictor and show the results. Finally, in Section 6 we draw conclusions for the paper.

2 Literature Review

In this section, an overview of previous studies of seizure prediction systems and big data management of epilepsy is provided. Definition and identification of the preictal state from continuous ECoG in dogs with a naturally occurring epilepsy has been managed with a support vector machine (SVM) algorithm [8]. This was further elaborated subsequently by Zhang et al. (2016) using spectral power ratios of ECoG data and processing through a second-order Kalman filter with final input to a linear SVM classifier [9]. An elimination-based feature selection method was used by Wang et al. (2015) to improve efficacy through diminishing redundancy thereby improving upon processing time [10]. The latter was further refined through the use of artifact-free preictal and interictal EEG epochs characterized using global feature descriptors [11].

An integrated framework of an EEG-based BCI has been applied for more effective upper limb motor rehabilitation [12]. A faster learning algorithm for a self-organizing fuzzy neural network (SOFNN) by Coyle et al. (2009) [13] will expedite an EEG-based BCI system for neural-time-series prediction processing. Chaovalitwongse et al. (2007) has described a method for distinguishing abnormal (i.e., epileptogenic) EEG signals from normal brain activity through integration of chaos theory, k-nearest neighbor and statistical time series analysis [14]. To discriminate multi-class motor imagery EEG signals in a BCI system, a common Bayesian network is proposed by He et al. (2016) [15]. Two time-series classification techniques, to further distinguish EEG signals and identify epileptogenicity were developed by Chaovalitwongse et al. (2011) [16]. In general, existing works have focused on local processing and storage without considering multiple channels and big patient data. The current work is built upon preliminary findings using a multitier distributed computing structure based on the Mobile Device Cloud (MDC) and cloud computing for real-time seizure detection [17].

The deep-learning approach pertaining to BCI has been considered in very few works. Lu et al. (2016) manually extracted supervised frequency features from EEG records to train three Restricted Boltzmann Machines (RBM) [18]. These layers were stacked with a softmax regression to form a deep belief network (DBN) for motor imagery classification and adaptive EEG analysis. Deep Belief Networks can be trained on each EEG channel and the results combined by AdaBoost [19]. DBN is applied for EEG data correlation analysis and superiority of results to those of PCA is shown in [20]. In the current work, a deep-learning structure using cloud computing has been applied to address the big data analysis problem in epilepsy. In contrast to existing methods, the proposed method extracts unsupervised features from ECoG patterns to predict ictal activity.

3 Cloud Computing Framework

Cloud computing provides a limitless scale of computing power that can be made available on demand and, by way of the Internet, makes it ubiquitously available for an extensive global reach [21], [22]. There are many cloud platforms including Microsoft, Google and Amazon AWS. For the purposes of this study and because of prior proven efficacy with large scale processing, cloud usage was applied through Amazon Cloud, otherwise called Amazon Web Services (AWS). The cloud is generally broken into three layers based on the service provided: (1) Infrastructure as a Service (IaaS); (2) Platform as a Service (PaaS) and (3) Software as a Service (SaaS). These three layers all lend themselves to the infrastructural setup of the BCI as follows.

IaaS: This provides computing power, networking, storage and virtual orchestrators and operating systems. It is available at large scale and on demand with the ability to deliver High Performance Computing (HPC) which lends itself well to the processing required with rapid real-time epilepsy monitoring. An applicable BCI system dealing with large amounts of data from distributed electrodes requires storage capability and both rapid and timely event-related mining to produce intelligence in the forms of trends, predictions and recommendations. With a low cost of entry and ease of setup, the core engine of the BCI can be effectively deployed using the AWS HPC [23]. High Performance Computing processors allow the BCI system to function above a teraflop capacity or 1012 floating-point operations per second allowing for real-time results inspite of large data entry. The Health Insurance Portability and Accountability Act (HIPAA) and its Protected Health Information (PHI) provision also requires service providers to adhere to strict assurances regarding protection of personal data. A need for encryption and use of AWS HIPAA eligible [24], [25] services are required to host the BCI system.

PaaS: This uses an open source allowing developers from different constituencies to leverage the BCI to continue developing modules and customized features for their local environment in order to adapt the application to their practices and needs [26].

SaaS: This uses a cloud-based BCI application allowing a good deal of processing power to be made available and distributed globally with decreased reliance on local extensive computer infrastructure in order to complete predictions. Aside from the standard EEG recording units and other specialized detection tools; run analysis, simulations and other high-end processes can be initiated from relatively light client applications including smartphone apps.

The proposed method is implemented in AWS and developed with Internet of Things (IoT). Such a closed-loop automatic system can be implemented in two separate steps. The first consists of developing a BCI to predict seizure onset and the second concerns the application of a suitable neurostimulatory signal to abort the seizure [27]. In the first instance, an accurate forecasting mechanism of seizure onset is required. The proposed framework enables collection of EEG telemetry data for storage and analysis. The IoT framework is connected with the Amazon cloud computing services that include a number of elements. Simple Storage Service (S3) provides secure, durable, and highly-scalable cloud storage. Kinesis enables real-time processing of EEG streaming data on a large scale. Lambda enables the running of a deep learning process on virtual servers from Elastic resizable Compute Cloud (EC2) in response to events. Finally, Simple Notification Service (SNS) contains an option to send notifications to the patient, doctors or emergency rooms.

The proposed IoT (see Fig. 1) consists of the following components. *Authentication and registry*, to make the system secure in order for its value to be realized. *Message broker*, for sending and receiving messages by a publish and subscribe service; supports MQTT and HTTPS protocols. *Gateway*, to enable secure communication between devices and IoT. *Rules engine*, to process EEG data and trigger the execution of deep learning in AWS Lambda. *Thing shadows*, to publish the current state of EEG analysis for use by applications.

4 Proposed Work

A seizure prediction algorithm is developed for big data analysis of EEG recordings which is fit to be implemented as a real-time cloud-based service. The proposed solution is presented in four steps (see Fig. 2). In Section 4.1, the preprocessing and extracting time and frequency features are explained. In Section 4.2, the developed deep learning structures for high level feature extraction are presented. In Section 4.3, the proposed optimization module for deep learning structures are presented. Finally, in Section 4.4, the analysis step for extracted feature extraction is presented.

4.1 Preprocessing, Time & Frequency Features

For preprocessing the EEG data, a fourth-order Butterworth bandpass filter (0.5-150 Hz) is used for cutting frequencies. Then, to remove some unwanted frequencies, a notch filter set at 50 Hz is applied. In the next step, the phase distortion

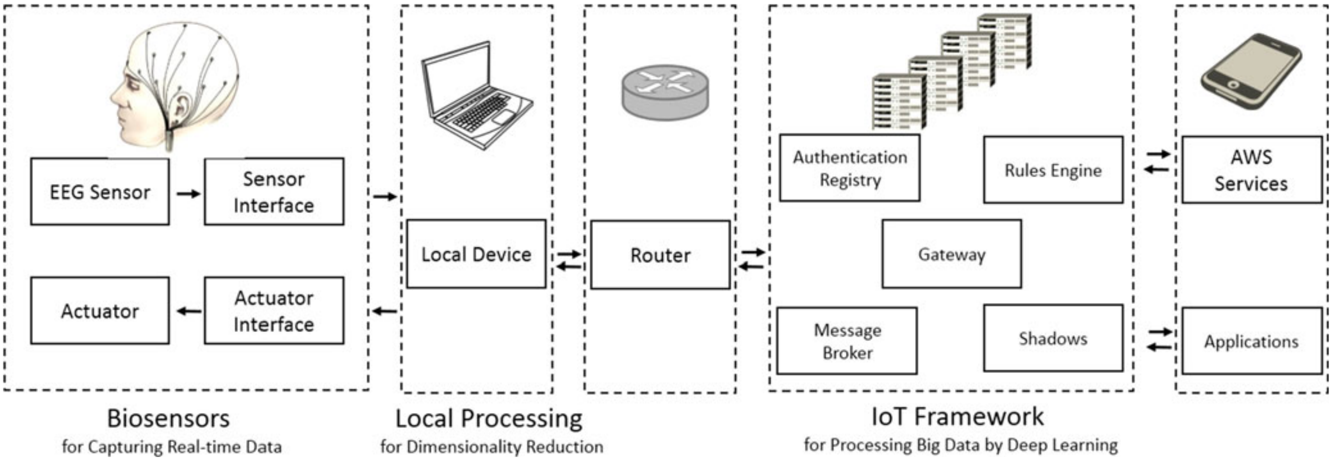

Fig. 1. The framework of a proposed IoT for seizure prediction consisting of authentication and registry, message brokering, gateway, rules engine and thing shadows. Such a closed-loop automatic system can be implemented in two separate steps. The first consists of developing a BCI to predict seizure onset and the second concerns the application of an suitable actuator signal to abort the seizure.

is canceled by using forward and backward filtering [28]. Because of the time-varying nature of EEG signals, wavelet transforms have been used to extract epileptic spikes and to capture the rhythmic nature of seizures [29], [30]. In addition, wavelet transforms are capable of capturing and localizing EEG transient features in both time and frequency domains. Therefore, the outputs of filtering, wavelet and ICA are used to extract preictal features. Several time and frequency-related features are considered including complexity, mobility, energy, entropy, correlation coefficients, Fast Fourier Transform (FFT), variance, skewness, kurtosis, mean, fractal dimension, frequency band power, peak amplitude, zero crossing, average spectral power, line length, maximal and minimal values, sum absolute value and some others. The appropriate wavelet and level of decomposition are chosen based on the input signal and application. Based on our evaluation of common wavelets, we use Daubechies 4 (db4) to find approximation and detail for EEG data. Since ictal activity with ECoG commonly occurs in the 3-25 Hz range, the detail coefficients have been investigated to find this frequency range. First, we consider a sampling frequency of the data (e.x. 500 Hz). By the Nyquist criteria, the maximum frequency of data is determined to be 250 Hz. Finally, by coefficient representation in each scale, the frequency range of 3-30 Hz is covered in scales of 4, 5 and 6.

4.2 Deep Learning for Feature Extraction

The extraction of meaningful features and patterns from large-scale EEG data for optimal data querying and analysis presents a significant challenge [31]. Deep structural learning has been recently advanced for computational methods addressing data processing and machine learning [32]. Deep learning structures use an hierarchical multilevel learning approach to extract meaningful abstract representations from raw data [33]. This property empowers deep networks with the capability for big data analysis. However, some weaknesses such as trap at local minima, lower performance and high computational time can occur with some applications such as EEG feature extraction. Therefore, new studies must seek appropriate optimization algorithms to obtain the best results from deep structures [34]. Constructing and training a new deep learning requires a considerable amount of labeled data as, otherwise, with a small amount of training data, the learning is ineffective. Therefore, in this study, instead of training a new network, fine-tuning of existing pretrained networks was undertaken for the seizure prediction task. This procedure is known as transfer learning which provides faster training and acceptable results with a small amount of labeled data. Two existing deep-learning structures, Stacked auto-encoder and CNN, for prediction of ictal onset by means of extraction of

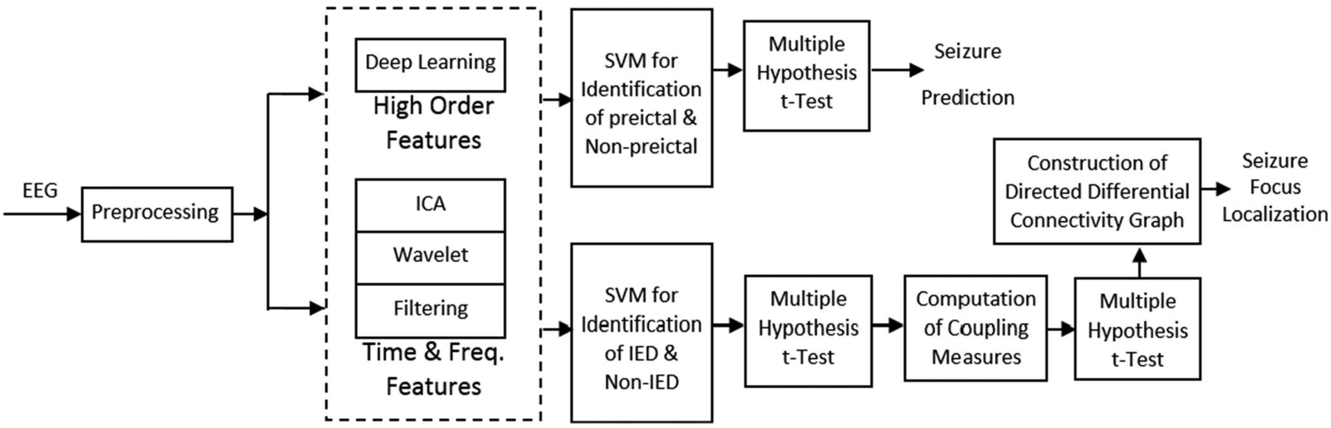

Fig. 2. Flowchart of the proposed system model for the analysis of EEG to predict the ongoing epileptic seizures and to localize the epileptogenic site.

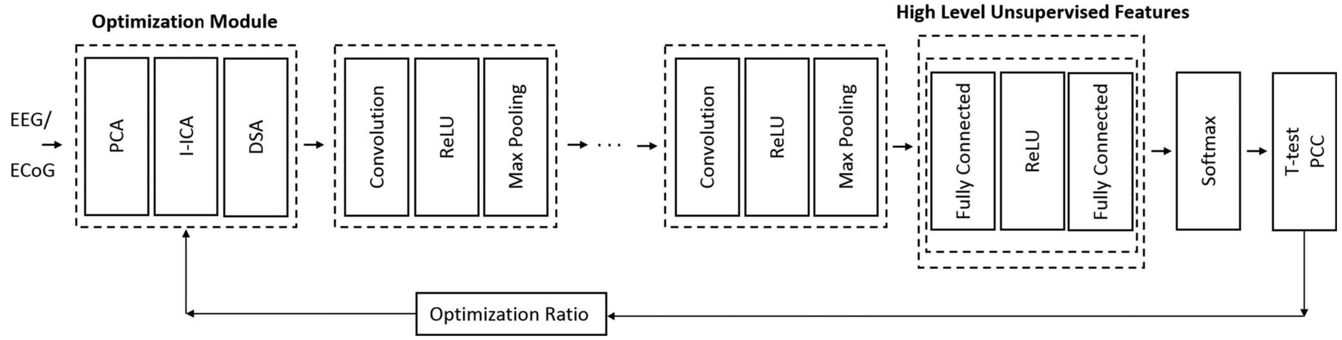

Fig. 3. The CNN consists of a multilayer structure. The input dimensions are defined in the first layer. The intermediate layers consist of a series of convolutional layers which are interspersed with rectified linear units (ReLU) and max-pooling layers. The last layer is designed for pattern classification via fully connected layers and softmax layer.

preictal feature patterns from unsupervised EEG data are presented here. Then, in the next section, we propose an optimization module to be deployed to optimize feature extraction by these deep structures.

Algorithm 1. CNN Forward and Backward propagation

Input: M-dimensional data, $\mathbf{x} = [x_1, \ldots, x_M]^T$
Output: Classification result as preictal (1) or nonpreictal (0) signal; $output \rightarrow (0, 1)$
begin
for $l := 1 \rightarrow \#HiddenLayers$ **do**
  for $i := 1 \rightarrow \#RowunitinLayerl$ **do**
    for $j := 1 \rightarrow \#ColumnunitinLayerl$ **do**
      Find the layer activations by,
      $y^l_{ij} = \varphi(x^l_{ij}) + b^l_{ij}$
      Compute next layer inputs by Eq. (12).
    end
  end
end
Keep the final output as y^l
Calculate error at the output layer.
begin
for $l := \#HiddenLayers \rightarrow 1$ **do**
  Find error partial derivation by Eq. (14).
  Find error at the previous laye by Eq. 15.
end
Calculatee the gradient of the error by Eq. (13).
END

Convolutional Neural Network. The developed CNN consists of a multilayer structure. The input dimensions are defined in the first layer. The intermediate layers consist of a series of convolutional layers which are interspersed with rectified linear units (ReLU) and max-pooling layers. Neurons are connected as rectangular grids in each convolutional layer where they have the same weights. In the pooling layer, small rectangular blocks from the convolutional layer are sub-sampled to find a single output [35], [36]. Finally, the last layer is designed for pattern classification via fully connected layers and the softmax layer (see Fig. 3).

The network is trained as a two-way classification problem—preictal state and nonpreictal state. Each of the layers respond to the input EEG signal but only a few layers are suitable for feature extraction. The first layer of CNN learning filters for basic features. Then, the primitive features are processed by deeper layers to develop higher level features. Features are extracted from the layer immediately before the classification layer since deeper layers combine all the primitive features into a more comprehensive signal representation.

Forward and backward propagation of algorithms are implemented in the CNN to find the output and to optimize the error, respectively. In order to formulate these steps, suppose a N × N square neuronal layer exists in the convolutional layer. By using a n × n filter, ω, the output is obtained by forward propagation,

$$x^l_{ij} = \psi\left(\sum_{a=0}^{n-1}\sum_{b=0}^{n-1}\omega_{ab}\,y^{l-1}_{(i+a)(j+b)}\right) \tag{1}$$

where ψ is the nonlinearity weight matrix. In this case, the size of output is (N − n + 1) × (N − n + 1). In the max-pooling layers, the size is reduced by sparseness. Then k × k regions are taken and the maximum in the regions is calculated to convert a single value output. Therefore, the size of output is reduced to $\frac{N-n+1}{k} \times \frac{N-n+1}{k}$.

For weight optimization, a back-propagation algorithm is applied to compute the derivative of the loss with respect to network parameters. Assuming error function, E, the gradient component for each weight can be found by applying the chain rule,

$$\frac{\partial E}{\partial \omega_{ab}} = \sum_{i=0}^{N-n}\sum_{i=0}^{N-n}\frac{\partial E}{\partial x^l_{ij}}\frac{\partial x^l_{ij}}{\partial \omega_{ab}} = \sum_{i=0}^{N-n}\sum_{i=0}^{N-n}\frac{\partial E}{\partial x^l_{ij}}\,y^{l-1}_{(i+a)(j+b)} \tag{2}$$

by which the gradient is computed,

$$\frac{\partial E}{\partial x^l_{ij}} = \frac{\partial E}{\partial y^l_{ij}}\frac{\partial y^l_{ij}}{\partial x^l_{ij}} = \frac{\partial E}{\partial y^l_{ij}}\frac{\partial}{\partial x^l_{ij}}(\psi(x^l_{ij})) \tag{3}$$

To find the weights of the convolutional layer, the error is back-propagated to the previous layer by the chain rule. Therefore, $\frac{\partial E}{\partial y^{l-1}_{ij}}$ is found by,

$$\sum_{a=0}^{n-1}\sum_{b=0}^{n-1}\frac{\partial E}{\partial x^l_{(i-a)(j-b)}}\frac{\partial x^l_{(i-a)(j-b)}}{\partial y^{l-1}_{ij}} = \sum_{a=0}^{n-1}\sum_{b=0}^{n-1}\frac{\partial E}{\partial x^l_{(i-a)(j-b)}}\omega_{ab} \tag{4}$$

The pseudocode of the forward and backward propagation is shown in Algorithm 1.

Stacked Autoencoder. This is a class of deep neural networks with multiencoders stacked together as hidden layers. The main property of a stacked autoencoder is that

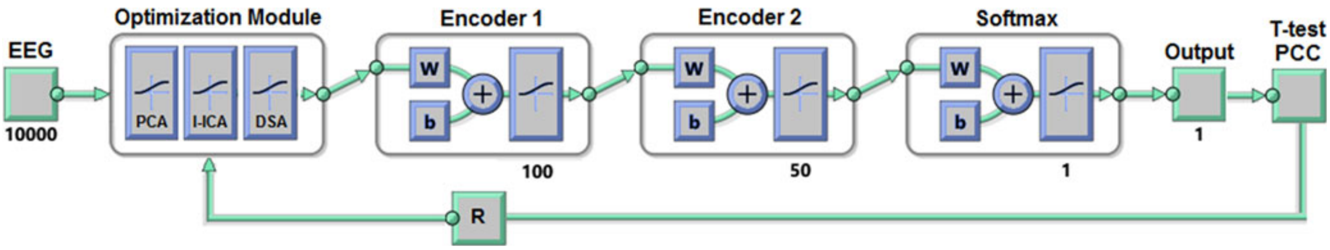

Fig. 4. The encoders from the autoencoders have been used to extract features. To form a deep network, the encoders from the autoencoders are stacked together and followed by a softmax layer.

of feature extraction from a large amount of unlabeled data, making it an applicable solution for the big data problem [37]. Here, this structure is developed with two separate encoders to capture the hierarchical grouping of the EEG input for a seizure prediction task. The encoder maps the input to a hidden representation. The size of the second hidden layer is designed with a lesser capacity than the first so the second encoder learns a smaller representation of the input data. Hidden layers are trained individually in an unsupervised method. The training data without labels are used to replicate the input from the output in the training step. To enforce a constraint on the sparsity of the output from the hidden layer, the impact of a sparsity regularizer is controlled. The first autoencoder tends to learn first-order features in the raw EEG input. The second-order features are extracted by inputting the primary features to the second hidden layer. A softmax layer is trained and the layers are joined to form a deep network as shown in Fig. 4. Finally, the network is trained a final time in a supervised manner. The pseudocode of the classification method is shown in Algorithm 2.

In the developed network, multiple nonlinear transformation layers are stacked together to represent a nonlinear function of the EEG data. A nonlinear transformation is applied to each layer's input and the input representation is provided in the output. Thus, there is no need to extract EEG features by manual engineering techniques for each patient. A gradient-log-normalizer of the categorical probability distribution as a softmax layer is used to classify the nonlinear function of the EEG as an interictal or preictal signal in the last layer. A softmax layer as a gradient-log-normalizer of the categorical probability distribution is used in the last layer. In order to predict an ictus, the network classifies the EEG segments as nonlinear inputs into preictal and nonpreictal. The softmax layer works as follows:

$$P(c_r|\mathbf{x}) = \frac{P(c_r)P(\mathbf{x}|c_r)}{\sum_{k=1}^{K} P(c_k)P(\mathbf{x}|c_k)} = \frac{exp(a_r)}{\sum_{k=1}^{K} exp(a_k)}, \qquad (5)$$

where $a_k = ln(P(c_k)P(\mathbf{x}|c_k))$, $P(c_k)$ is the class prior probability, and $P(\mathbf{x}|c_k)$ is the conditional probability of the sample given class k.

4.3 Optimization Module for Deep Learning

While deep learning offers a valuable and efficient approach to be used for solving a multiple problems, many limitations still exist in deep learning. The primary problem that often occurs in deep learning is the overfitting problem and it is always stuck at local minima. Once these problems occur, it results in lower performance and a higher computational time in deep learning. Therefore, the optimization algorithm can be taken into consideration to help make up for the limitations that occur in the deep learning method.

Algorithm 2. Deep Learning by Stacked Autoencoder

Input: M-dimensional data, $\mathbf{x} = [x_1, \ldots, x_M]^T$
Output: Classification result as preictal (1) or non-preictal (0) signal; $output \rightarrow (0, 1)$
begin
for $i := 1 \rightarrow \#HiddenLayers$ **do**
  Decrease the size of the ith hidden layer, P(i) < P(i-1)
  Train unsupervised the ith autoencoder
  Set explicitly the random number generator seed
  Control the impact of an L2 regularizer for weights
  Control the impact of a sparsity regularizer
  Control the output sparsity from the hidden layer
  Use the ith feature set for training in the next layer
end
Train supervised a softmax layer to classify ith features
Stack the encoders from the autoencoders with softmax
Compute the results on the test set $output \rightarrow (0, 1)$
Do fine tuning by retraining on the training data
END

To render greater efficiency and optimize the feature extraction of a complex large scale dataset in the setting of deep learning, optimization, using a method based on PCA, I-ICA, and DSA was undertaken. In this combined approach, PCA decorrelated the EEG data and the remaining higher-order dependencies were separated by I-ICA. Then, DSA was used to find the optimal solution by optimizing the search space in locating the region of global minimum.

Principal component analysis generates a diagonal covariance matrix from input data. Using a transformation, each dimension is normalized such that the covariance matrix is equal to the identity matrix. As a result, small trailing eigenvalues are discarded and computational complexity decreased by minimizing pairwise dependencies. Consider an EEG data matrix, $\mathbf{X}$, with a column-wise zero empirical mean, where each of the n rows identifies a different time and each of the p columns demonstrates a particular channel. The transformation

$$\mathbf{Y}_{np} = \mathbf{X}_{np}\mathbf{W}_{pp}, \qquad (6)$$

where $\mathbf{W}$ is a p × p matrix, maps a data matrix $\mathbf{X}$ to a new space that is uncorrelated over the dataset. The columns of $\mathbf{W}$ are the eigenvectors of $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}^T\mathbf{X}$ and are proportional to the empirical sample covariance matrix of X. For dimensionality reduction, it is unnecessary to retain all principal components and the first L largest principal components may be used. Therefore

$$\mathbf{Y}_{nl} = \mathbf{X}_{np}\mathbf{W}_{pl}, \qquad (7)$$

This truncated transform finds $\mathbf{Y}$ with n rows and only L columns. As a result, PCA learns a linear transformation to transform a set of correlated variables, $\mathbf{X}$, into a set of uncorrelated L features, $\mathbf{Y}$, called principal components. In the proposed structures for CNN and stacked autoencoder, the outputs are evaluated by t-test and Pearson Correlation Coefficient (PCC) [38]. If the p-value is not less than 0.05 with PCC at or near 0, using the optimization ratio as a feedback loop, the component of L, as largest principal component, is updated. This optimization ratio is used to find the best number of principal components to obtain an optimum output for deep learning with p-value less than 0.05 and PCC around 0.

The ICA is a computational method which separates observed data, $\mathbf{Y}$, in terms of an independent hidden source, $\mathbf{D}$

$$\mathbf{Y} = \mathbf{GD} + \mathbf{E}, \tag{8}$$

where $\mathbf{G}$ is the mixing matrix and $\mathbf{E}$ represents Gaussian noise. In the standard ICA, we assume $\mathbf{Y}$ and $\mathbf{D}$ have the same dimensions. However, this assumption may not be valid and requires a reversible-jump Markov chain Monte Carlo application to determine the dimension of $\mathbf{X}$. To solve this problem and to develop a method appropriate for dimensionality reduction of a big data input, I-ICA [39] followed upon PCA. This increases data reduction for systems with intermittently active sources. The ICA performance increases with preapplication of PCA. This removes small trailing eigenvalues before whitening and minimizes pairwise dependencies [40].

By defining a binary matrix, $\mathbf{Z}$, where its elements show activity of a kth hidden source for the ith data point, we have

$$\mathbf{Y} = \mathbf{G}[\mathbf{D} \odot \mathbf{Z}] + \mathbf{E}, \tag{9}$$

where $\odot$ denotes the element-wise multiplication. In this case, $\mathbf{Z}$ has infinitely many rows with finite nonzero elements so a potentially infinite number of hidden sources is available. For N data points and K hidden sources, the distribution of matrix $\mathbf{Z}$ is defined by

$$P(\mathbf{Z}|\pi_1, \ldots, \pi_K) = \prod_{k=1}^{K}\prod_{i=1}^{N} P(z_{ki}|\pi_k), \tag{10}$$

which is expanded to,

$$P(\mathbf{Z}|\pi_1, \ldots, \pi_K) = \prod_{k=1}^{K} \pi_k^{m_k}(1-\pi_k)^{N-m_k}, \tag{11}$$

where z_{ki} indicates activity of the kth source for the ith sample using a probability of π_k and $m_k = \sum_{i=1}^{N} z_{ki}$ indicates the total number of active sources. Finally, for inferring $\mathbf{D}$ hidden sources from $\mathbf{Y}$ observed data using the $\mathbf{G}$ mixing matrix, which has Z active sources, an inference is defined. Gibbs sampling is used to sample elements with $z_{ki} = 1$. This proceeds by sampling from the conditional distribution of one parameter given all others by Bayes rule [39]. Therefore, the result is a piecewise Gaussian distribution. For $d_{ki} > 0$, we have

$$P(d_{ki}|\mathbf{G}, d_{-ki}, y_i, z_i) = \mathcal{N}(x_{ki}; \mu_-, \sigma^2), \tag{12}$$

and for $d_{ki} < 0$ we have,

$$P(d_{ki}|\mathbf{G}, d_{-ki}, y_i, z_i) = \mathcal{N}(d_{ki}; \mu_+, \sigma^2), \tag{13}$$

where $\mu_\pm = \frac{g_k^T e_{ki}^\circ \pm \sigma_e^2}{g_k^T g_k}$, $\sigma^2 = \frac{\sigma_e^2}{g_k^T g_k}$, $e_{ki}^\circ = (e_{ki}|z_{ki} = 0)$, and $\mathbf{g}_k$ is the kth column of $\mathbf{G}$. Finally, by marginalizing over all possible values of $\mathbf{d}_{kt}$, the $P(y_i|\mathbf{G}, d_{-ki}, z_{-ki}, z_{ki} = 1)$ is found with,

$$\int [P(y_i|G, d_i, z_{-ki}, z_{ki} = 1)P(d_{ki})]d(d_{ki}) \tag{14}$$

In summary, the PCA expands EEG raw data into a set of orthogonal components which provides maximal decorrelation of the signals. This enables a separation of noise subspace from EEG data. The I-ICA, a scalable ICA for high dimensional data, reduces the dimension by extraction-independent and orthonormal methodology. The proposed combination of PCA with I-ICA therefore provides for dimensionality reduction in big data and generates a sparse representation of the raw data. This sparsity allows more efficient unsupervised feature extraction by deep-learning structures as demonstrated in the results.

DSA is an efficient novel evolutionary algorithm which is proposed to optimize solving of real-valued numerical problems. This algorithm is inspired from Brownian-like random walk movement [41]. It has been shown that DSA shows better performance in the solution of numerical optimization problems than other routine methods including JDE, JADE, SADE, ABC, EPSDE, GSA, CMA-ES, and PSO2011 [42]. We have used DSA to explore searching space to obtain it with a global minimum. As a result, the deep learning structures do not trap in the local minimums and the performance and the computational time is increased. The pseudocode of the proposed optimization method is described in detail in Algorithm 3 and pseudocode DSA part is in Algorithm 4. Results show that the proposed optimization model outperforms classic deep learning-based approaches in terms of training efficiency and seizure prediction accuracy.

4.4 Analysis Extracted Features

An increase in the functional connectivity of the brain during the interictal period has been reported [43], [44]. In this paper, we hypothesize that there are differences between the functional brain connectivity in the Interictal Epileptiform Discharges (IED) and non-IED periods. To find the differences, a differential connectivity graph (DCG) is constructed. Since the leading IED regions (sources) relate to the epileptogenic zone relative to the propagated IED regions (sinks) [45], [46], we estimate directed DCGs (dDCG) for different frequency bands and characterize them by an information emittance measure. Next, a multi-objective optimization method [47] is applied on the emittance values of all dDCG nodes in all frequency bands to identify the leading IED regions.

A nonlinear SVM with a GRBF kernel is used for classification of the extracted features [48], [49]. To improve the results, the Gaussian kernel parameters are optimized by maximizing a classical class separability criterion as the trace of the scatter ratio. Then, a quasi-Newton algorithm is used by exploiting a recently proposed criterion of decomposition of the objective. Fig. 5 shows the linear SVM classification for 2 subspace of

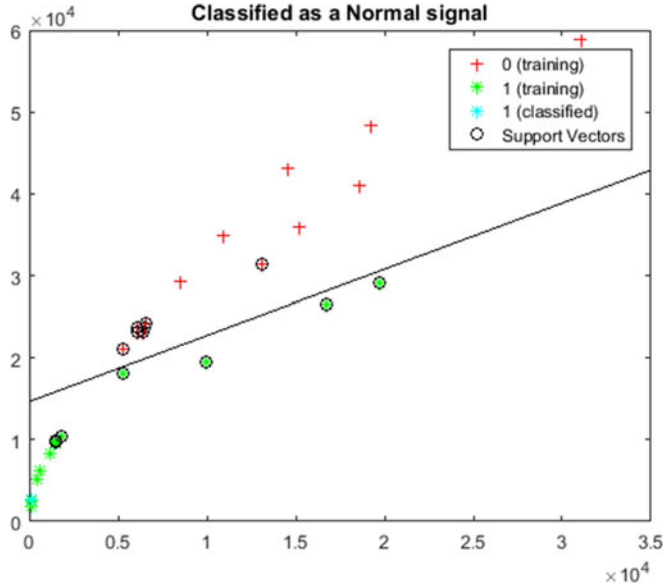

Fig. 5. Results of linear classification in one subset of features (P1, P2). Red signs show features extracted from the ictal group, green signs show features extracted from the interictal group, and the blue sign is a test signal, which is originally a normal brain activity that is correctly identified and classified as normal by the classifier.

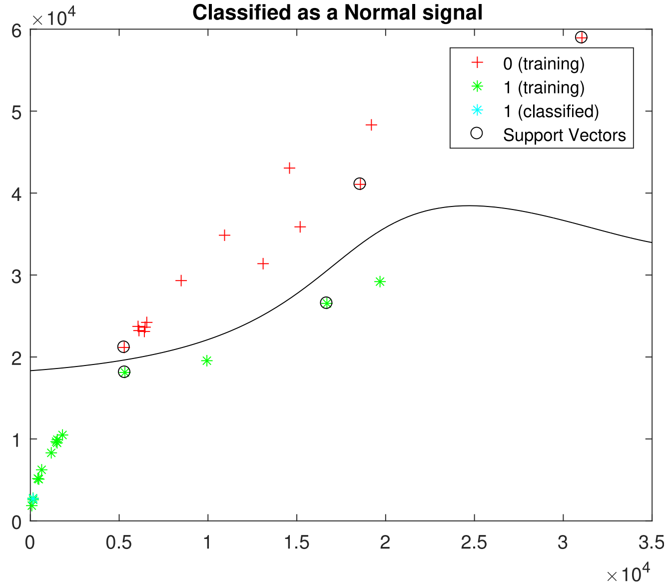

Fig. 6. Results of non-linear classification in one subset of features (P1, P2). Red signs show features extracted from the ictal group, green signs show features extracted from the interictal group, and the blue sign is a test signal, which is originally a normal brain activity that is correctly identified and classified as normal by the classifier.

extracted features (power and Sum of absolute elements) and Fig. 6 shows the result with non-linear SVM classifier.

Algorithm 3. Optimization Based on PCA + I-ICA + DSA

Input: D-dimensional EEG raw data $\mathbf{d} = [d_1, \ldots, d_D]^T$
Output: M-dimensional signal $\mathbf{x} = [x_1, \ldots, x_M]^T$,
$\odot$ denotes the element wise multiplication
BEGIN
Compute Covariance matrix of D
Choose P largest eigenvalues
Find Y by Eq. (2)
for $i := 1 \rightarrow N$ **do**
  $[Y, X, Z, E] = concatenation\{\mathbf{y}_i\ \mathbf{x}_i\ \mathbf{z}_i\ \mathbf{e}_i\}_{i=1}^N$
end
for $k := 1 \rightarrow K$ **do**
  Define z_{ki} as activity of kth source for ith sample
  Define $m_k = \sum_{i=1}^N z_{ki}$ as the active sources
  Calculate $p(\mathbf{Z}|\pi_1, \ldots, \pi_K)$ by Eq. (6)
  for $i := 1 \rightarrow N$ **do**
    Define $\mathbf{g}_k$ as the kth column of $\mathbf{G}$
    Find $\mu_\pm$ by $\frac{g_k^T e_{ki}^\circ \pm \sigma_e^2}{g_k^T g_k}$
    Find σ^2 by $\frac{\sigma_e^2}{g_k^T g_k}$
    Find e_{ki}° by $(e_{ki}|z_{ki} = 0)$
    Calculate $\mathbf{p}(x_{ki}|\mathbf{G}, x_{-ki}, y_i, z_i)$ by
    if $x_{ki} > 0$ **use** Eq. (7)
    if $x_{ki} < 0$ **use** Eq. (8)
  end
end
Find X by Eq. (4)
Extract $\mathbf{x} = [x_1, \ldots, x_M]^T$ from X
Apply DSA
END

A proposed approach searches for statistically significant connections among a large number of IED and non-IED time intervals. It selects the connections that change significantly between the IED and non-IED states. This approach decreases the effect of common information between the two states like background activity. Using a permutation-based multiple testing method [50], we estimate the distribution of a test statistic from different IED and non-IED time intervals under the null hypothesis and use the result to choose statistically significant connections. We start with the construction of a DCG which requires identification of IED and non-IED intervals and computation of a coupling measure. Identification of IED and non-IED time intervals is initially done manually by a collaborating epileptologist. Later, the manual results are used to develop an automatic method for this identification. An IED period may include one single IED or a burst of IEDs. A non-IED period is a time interval without any IED or abnormal event. For the coupling measure, we compare linear, nonlinear and directed coupling measures, including wavelet correlation coefficient, phase synchrony and transfer entropy and choose the most informative measure for the project. For preprocessing and feature extraction (i.e., to separate the signal (relevant information) from noise and background (irrelevant information)), we use the wavelet transform, which has been shown to be optimal for analyzing nonstationary EEG signals. A DCG for each of the frequency bands (i.e., wavelet scales) is constructed. Multiple hypothesis t-tests are applied to choose the wavelet bands that are most relevant and thus generate discriminating DCG between the IED and non-IED states. Then, for constructing a dDCG, the method of Amini et al. (2009) [51] is applied to estimate the drive-response relationship between the signals observed at the nodes of the constructed DCG. To identify the epileptogenic zone, we must classify the nodes of the constructed dDCG to the source and sink groups. To this end, we will use an index called Local Information (LI) [52] to measure the amount of information that passes through a node. This measure will depend on: (1) incoming

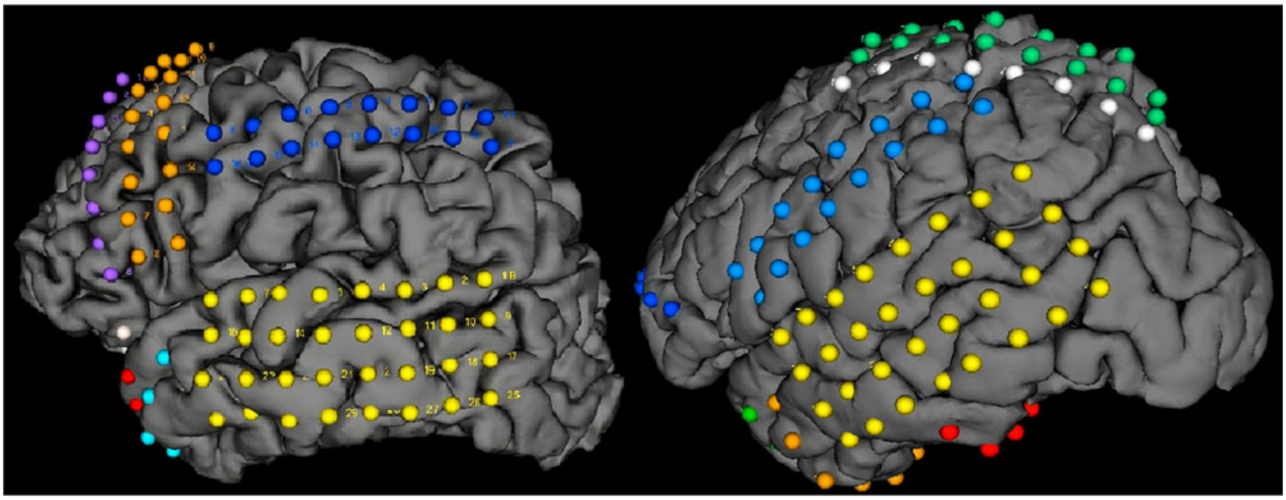

Fig. 7. Two different cases from our group at spectrum heath, grand rapids, MI that underwent wide placement of electrodes over the left hemisphere during a phase II investigation to determine the site of epileptogenicity. We were able to identify the sites of epileptic disturbance and implant a closed loop device with two electrode arrays placed in the locations that were found to be epileptogenic.

and outgoing connections; and (2) the amount of information carried by each connection, which will be calculated using the lagged mutual information (MI) between the signal pairs observed at the two ends of the connection.

Algorithm 4. DSA Algorithm

Input: Output of I-ICA
Output: Optimized problem
Do specify the DSA parameters
Do initialize the population
Out Iter=1
Evaluate superorganism & fitness
while stop criteria is not satisfy
Update stopover
Evaluate stopover fitness
if stopover-fitness > superorganism-fitness **then**
  Replace superorganism by stopover in new population
else discard the stopover in new population
if stop criteria is satisfy **then**
  Go to end
else Iter=Iter+1 **Repeat** while loop
END

5 Results

A proof-of-concept prototype of the proposed BCI seizure predictor was developed in the cloud and Autonomic Computing Center (CAC), Rutgers University. In this testbed, a benchmark dataset of EEG expression of epileptogenicity was employed along with an HP laptop and Intel i5 processor, an 8 GB RAM and battery capacity of 4,400 mAh and a super-cluster of computers hosted by AWS which offers cloud computing services. The proposed models were implemented in the Pytorch package with developing pre-trained networks on an EEG dataset [53], [54].

Data. ECoG data were collected at 1,000 Hz at Spectrum Health, Grand Rapids, MI from nine patients with focal epilepsy using in excess of 70 intracranial electrodes implanted over the cerebrocortical surface where each electrode corresponded to an individual channel on the ECoG recording (see Fig. 7). The data were then annotated by an epileptologist who noted the start and stop time for each preictal period along with the channels that exhibited the epileptic activity. A total 1,755 segments are used (585 interictal, 585 ictal, and 585 preictal). The ECoG datasets of two epilepsy patients with temporal and extratemporal epilepsy, jointly developed and released by the University of Pennsylvania and the Mayo Clinic and sponsored by the American Epilepsy Society, were used for this investigation [55], [56]. A total of 390 segments were used (130 interictal, 130 ictal segments, and 130 preictal). The dataset was recorded by 15 intracranial electrodes. Interictal and preictal data were segmented in 10 minute clips. A sampling rate of 5,000 Hz was applied and a scalp reference electrode used for referential recording. Preictal data segments covered a one hour duration prior to the ictus and the ictal horizon was defined as five minutes. The preictal horizon provided some guarantee that seizures could be forecast with sufficient warning to allow intervention in some circumstances. In Fig. 8, patterns of interictal and preictal segments are compared.

Outcome. The database contains a few independent cases each with a big data problem. These algorithms should be regulated against over-fitting. The proposed solution extracts the features in an unsupervised manner, decreasing the risk of over-fitting. Moreover, to evaluate the generality of the results, a leave-one-out approach is employed as an exhaustive cross validation technique on subjects. Using this technique, the model is fitted to subsets of patients and

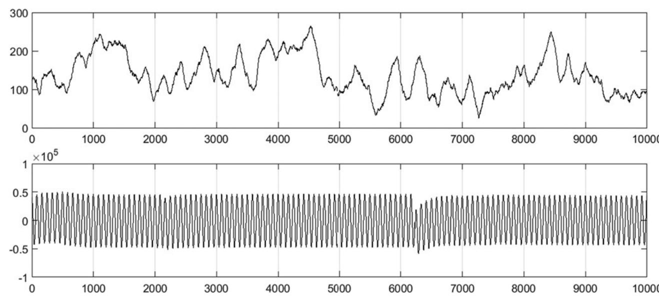

Fig. 8. A comparison of interictal (i.e., baseline) iEEG segment on top and preictal (i.e., before seizure) iEEG segment on bottom. The x-axis represents samples and the y-axis shows the signal in $\mu\mathrm{V}$.

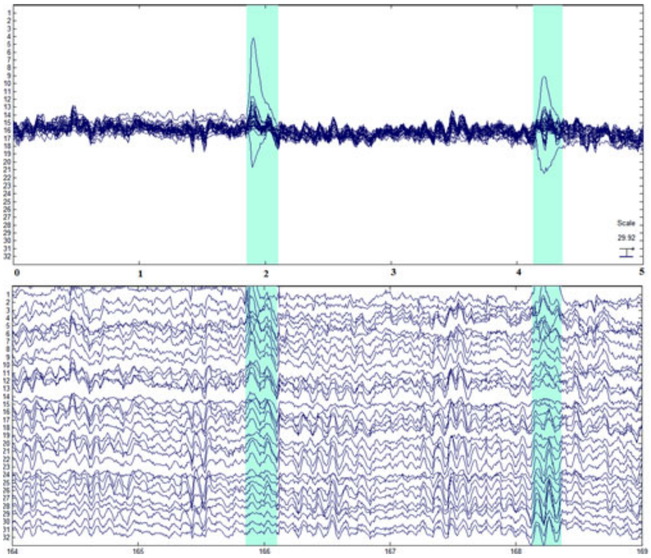

Fig. 9. EEG segments of 32 channels showing epileptiform normal variants in a few channels (top) and stacked EEG segments (bottom) in the highlighted areas. The horizontal line is Time [s] and the vertical line is Amplitude [μV].

the accuracy of the Stacked Auto-encoder and CNN models are found using the held-out sample [38].

Many EEG patterns that resemble epileptogenic abnormalities are not associated with epilepsy or any neurologic conditions. Small sharp spikes, wicket spikes, phantom waves and paroxysmal rhythmic discharges are some examples of these patterns [57]. Since such patterns have no clinical significance for seizure detection and prediction, they are termed epileptiform normal variants. These patterns are one of the major causes of false seizure detection in automatic methods Fig. 9 and their recognition is important for avoiding over-interpretation [58]. The deep networks are supervisor-trained to recognize them as normal patterns.

Mean square error using l2 and sparsity regularizers for unsupervised feature extraction in the first hidden layer of Stacked Auto-encoder is shown in Fig. 10 where the best training performance is 0.007 at epoch 100. The confusion matrix of the proposed method using Stacked Auto-encoder and CNN are shown in Table 1 and 2, respectively.

To evaluate the classification ability of the proposed unsupervised feature extraction and classification, the EEG feature sets are extracted for classification by other classification methods. The extracted features are based on fast

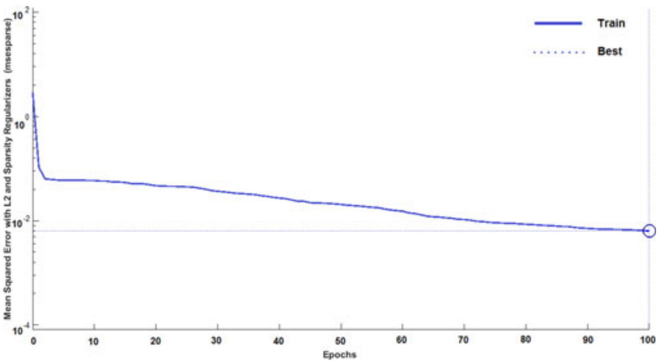

Fig. 10. The mean square error of unsupervised feature extraction by the first hidden layer with Stacked Autoencoder.

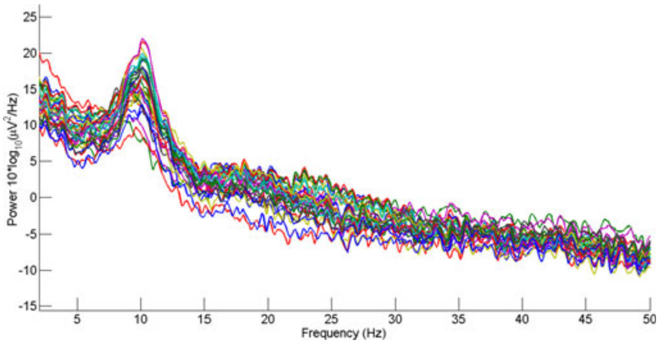

Fig. 11. The power log feature for interictal segments. The x-axis represents frequency in Hz and the y-axis shows the Power 10*log in $\mu\mathrm{V}^2/\mathrm{Hz}$.

Fourier transform, general energy average and energy STDV over time for each channel, power spectral density correlation coefficients, partial directed coherence of coefficients, power in band, low-gamma phase sync and log of energy in different frequency bands for each channel. Fig. 11 demonstrates the power log feature for interictal segments. To classify these manually extracted features, Random Forest, Liner SVM, Non-linear SVM and MLP Neural Network are used. Experimental results in Table 3 show that the proposed deep learning methods outperform previous methods for the EEG seizure prediction task.

Table 4 displays the results of a standard ANOVA analysis [1], while Fig. 12 shows some metrics on extracted features for interictal and preictal periods by quartile. The large F-statistic and small value of p in Table 1 correspond to a large difference in the center lines of the box plots in Fig. 12.

Message Queuing Telemetry Transport (QMTT) and RESTful Web Service protocols are used for sending data in the cloud [60]. The proposed framework collects telemetry EEG data from S3 after receiving a state in JSON format on MQTT topics. After a message is published on an MQTT topic, it is sent to MQTT message broker and subsequently to all subscribed clients. This communication is protected by X.509 certificates. To secure communication between sensors

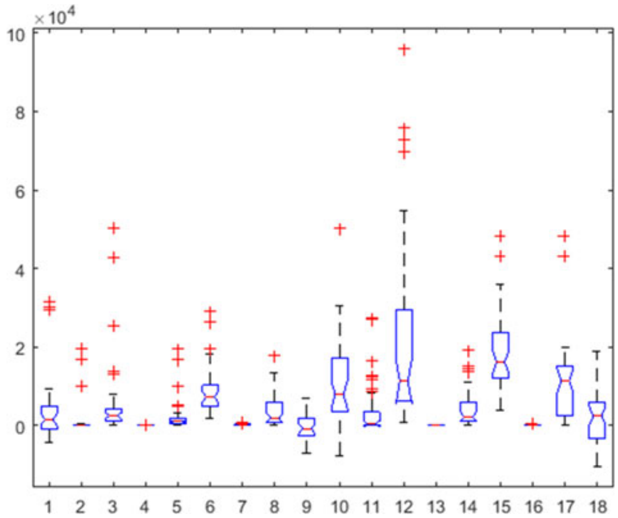

Fig. 12. Graphical depiction of some metrics on the extracted features for interictal group (#1 to #9). Amplitude maximum (multiplied by scaled factor 100), variance, energy (multiplied by scaled factor 0.01), skewness, power, sum of absolute value (multiplied by scaled factor 100), FFT, and mean value (multiplied by scaled factor 0.01) are shown, respectively. Outliers are plotted by plus signs. Same amount for preictal group (#10 to #18).

TABLE 1
Confusion Matrix for Stacked Auto-Encoder

	Output interictal	Output preictal	Total
Target interictal	610	40	650
Target preictal	32	618	650
Total	642	658	1300

The diagonal elements show the correct decisions. The results are shown as mean value of leave-one-patient-out.

TABLE 2
Confusion Matrix for CNN

	Output interictal	Output preictal	Total
Target interictal	619	31	650
Target preictal	22	628	650
Total	641	659	1300

The diagonal elements show the correct decisions. The results are shown as mean value of leave-one-patient-out.

TABLE 3
Accuracy, Precision, Sensitivity, Specificity, FPR, and FNR for Proposed Classification Compared with the Other Methods

Methods	Accuracy	Precision	Sensitivity	FPR	FNR
Opt. CNN	0.96	0.97	0.97	0.05	0.03
Opt. St. Autoencoder	0.94	0.95	0.95	0.06	0.05
Classic CNN	0.93	0.94	0.94	0.07	0.06
Classic St. Autoencoder	0.92	0.93	0.93	0.08	0.07
Random Forest	0.78	0.78	0.78	0.23	0.21
Non-linear SVM	0.76	0.77	0.77	0.25	0.23
Linear SVM	0.74	0.75	0.75	0.27	0.25
MLP Neural Network	0.69	0.71	0.71	0.32	0.29.

The results are shown as mean value of leave-one-patient-out.

TABLE 4
Source of Variability, Sum of Squares (SS) of Each Source, Degrees of Freedom (DF) of a Source, Mean Square (MS) as the Ratio SS/DF, Ratio of the Mean Squares as F-Statistic, and p-Value Derived from the Cumulative Distribution Function (CDF) of F Are Shown as the Standard ANOVA [59]

Group	Source	SS	df	MS	F-statistic	p-value
	Columns	2.89e+10	8	3.62e+08	11.45	1.85e-14
Interictal	Error	1.11e+10	351	3.16e+07		
	Total	1.40e+10	359			
	Columns	1.96e+10	8	2.45e+09	21.92	4.64e-27
Preictal	Error	3.92e+10	351	1.11e+08		
	Total	5.88e+10	359			

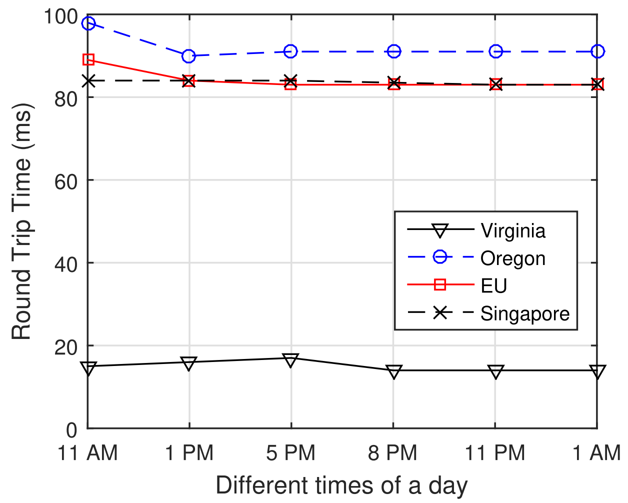

Fig. 13. Round Trip Time (RTT) between Amazon EC2 servers and a local machine for EEG segments. At any time, the lowest RTT is observed for the server located in Virginia, the closest to the study location.

and cloud, the device registry is used to store the certifications and information about the sensors. Using Rule Engine, the result of processing of the deep network in EC2 is extracted. Then, using Device Shadow, the state information is retrieved. An application which provides the neurostimulatory signal controls the sensors by requesting a change in its state.

The feasibility of using cloud computing is analyzed by the network latency offered by Amazon EC2 cloud servers. The Round Trip Time (RTT) for servers located at different geographical locations (i.e., Virginia, Oregon, Singapore, and Ireland) is calculated for 64B EEG segments at 10 days using the *ping* command. The shortest RTT is 15 ms for the Virginia server and the longest RTT is 97 ms for the Oregon server. The mean time is reported in Fig. 13 over different day times.

6 Conclusion and Future Work

There is great utility in the efficient handling and processing of big data for the management of complex medical conditions that may often require immediate intervention. This is exemplified in the context of medically intractable epilepsy and the use of implanted electrodes strategically targeting distinct sites of epileptogenicity within the brain. The requirements for safe storage and high computational resources for processing such data must also take into account the large variety of patterns characterized by fluctuations of signal amplitude and frequency that create significant challenges for reliable feature extraction. In order to address these issues, a novel cloud-based BCI providing real-time seizure prediction and seizure focal localization from EEG and ECoG data is proposed. An optimization method, as a layer in deep learning, decreases energy and computation time while retaining accurate classification of the event. The developed deep-learning methods provide unsupervised feature extraction as a suitable substitute to manual feature-extraction techniques of classification through a hierarchical learning process that extracts high-level and complex abstractions for data representations. The key benefit of the proposed method is summed up in the rapid analysis and learning provided for large amounts of unsupervised data. This translates into a means by which timely interjection in the epileptogenic process may prevent seizures from occurring bringing significant benefit to the patient who otherwise has no opportunity for such control.

A rapid coupling of reliable preictal detection and effective execution of an inhibitory signal delivery will determine the success of this therapeutic system. Future work will engage an implantable system currently in use (RNS, Neuropace) to apply these capabilities in a prospective manner to determine their utility. A further understanding of how an estimation of EEG variance can be used as feedback for responsive neurostimulation may further inform an automatic computing system is also worthy of further study. A decision-making process may also be implemented in order to select an appropriate preictal electrostimulatory signal in order to abort an evolving seizure as a second stage measure of the automatic computing system.

Acknowledgments

A preliminary shorter version of this work appeared in the *Proc. of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Greater Washington, D.C., USA, Dec. 2016 [1].

References

[1] M.-P. Hosseini, H. Soltanian-Zadeh, K. Elisevich, and D. Pompili, "Cloud-based deep learning of big EEG data for epileptic seizure prediction," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2016, pp. 1151–1155.

[2] M.-P. Hosseini, M. R. Nazem-Zadeh, D. Pompili, and H. Soltanian-Zadeh, "Statistical validation of automatic methods for hippocampus segmentation in mr images of epileptic patients," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, 2014, pp. 4707–4710.

[3] W. He, Y. Zhao, H. Tang, C. Sun, and W. Fu, "A wireless bci and BMI system for wearable robots," *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 46, no. 7, pp. 936–946, Jul. 2016.

[4] V. Gandhi, G. Prasad, D. Coyle, L. Behera, and T. M. McGinnity, "Eeg-based mobile robot control through an adaptive brain–robot interface," *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 44, no. 9, pp. 1278–1285, Sep. 2014.

[5] M.-R. Nazem-Zadeh, et al., "Lateralization of temporal lobe epilepsy using intrinsic property of water diffusion in fornix crus," *Anal.*, vol. 5, 2014, Art. no. 6.

[6] D. Kuzum, et al., "Transparent and flexible low noise graphene electrodes for simultaneous electrophysiology and neuroimaging," *Nature Commun.*, vol. 5, 2014, Art. no. 5259.

[7] M.-P. Hosseini, M.-R. Nazem-Zadeh, D. Pompili, K. Jafari-Khouzani, K. Elisevich, and H. Soltanian-Zadeh, "Automatic and manual segmentation of hippocampus in epileptic patients mri," arXiv:1610.07557, 2016.

[8] B. H. Brinkmann, et al., "Forecasting seizures using intracranial EEG measures and SVM in naturally occurring canine epilepsy," *PloS One*, vol. 10, no. 8, 2015, Art. no. e0133900.

[9] Z. Zhang and K. K. Parhi, "Low-complexity seizure prediction from iEEG/sEEG using spectral power and ratios of spectral power," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 3, pp. 693–706, Jun. 2016.

[10] N. Wang and M. R. Lyu, "Extracting and selecting distinctive EEG features for efficient epileptic seizure prediction," *IEEE J. Biomed. Health Inf.*, vol. 19, no. 5, pp. 1648–1659, Sep. 2015.

[11] L.-C. Lin, S. C.-J. Chen, C.-T. Chiang, H.-C. Wu, R.-C. Yang, and C.-S. Ouyang, "Classification preictal and interictal stages via integrating interchannel and time-domain analysis of EEG features," *Clinical EEG Neuroscience*, vol. 48, pp. 139–145, 2016.

[12] Z. XG, et al., "SSVEP-based brain-computer interface controlled functional electrical stimulation system for upper extremity rehabilitation," *IEEE Trans. Syst Man Cybern.: Syst.*, vol. 46, no. 7, pp. 977–988, Dec. 2016.

[13] D. Coyle, G. Prasad, and T. M. McGinnity, "Faster self-organizing fuzzy neural network training and a hyperparameter analysis for a brain–computer interface," *IEEE Trans. Syst. Man Cybern. Part B*, vol. 39, no. 6, pp. 1458–1471, Dec. 2009.

[14] W. A. Chaovalitwongse, Y.-J. Fan, and R. C. Sachdeo, "On the time series k-nearest neighbor classification of abnormal brain activity," *IEEE Trans. Syst. Man Cybern.-Part A: Syst. Humans*, vol. 37, no. 6, pp. 1005–1016, Nov. 2007.

[15] L. He, D. Hu, M. Wan, Y. Wen, K. M. von Deneen, and M. Zhou, "Common Bayesian network for classification of EEG-based multiclass motor imagery BCI," *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 46, no. 6, pp. 843–854, Jun. 2016.

[16] W. A. Chaovalitwongse, R. S. Pottenger, S. Wang, Y.-J. Fan, and L. D. Iasemidis, "Pattern-and network-based classification techniques for multichannel medical data signals to improve brain diagnosis," *IEEE Trans. Syst. Man Cybern.-Part A: Syst. Humans*, vol. 41, no. 5, pp. 977–988, Sep. 2011.

[17] M.-P. Hosseini, A. Hajisami, and D. Pompili, "Real-time epileptic seizure detection from EEG signals via random subspace ensemble learning," in *Proc. IEEE Int. Conf. Autonomic Comput.*, 2016, pp. 209–218.

[18] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted boltzmann machines," *IEEE Trans. Neural Syst. Rehabil. Eng.: A Publication Eng. Med. Biol. Soc.*, vol. 25, no. 6, pp. 566–576, Jun. 2017.

[19] X. An, D. Kuang, X. Guo, Y. Zhao, and L. He, "A deep learning method for classification of EEG data based on motor imagery," in *Proc. Int. Conf. Intell. Comput.*, 2014, pp. 203–210.

[20] Z. V. Freudenburg, N. F. Ramsey, M. Wronkiewicz, W. D. Smart, R. Pless, and E. C. Leuthardt, "Real-time naive learning of neural correlates in ECoG electrophysiology," *Int. J. Mach. Learn. Comput.*, vol. 1, no. 3, 2011, Art. no. 269.

[21] T. X. Tran, M.-P. Hosseini, and D. Pompili, "Mobile edge computing: Recent efforts and five key research directions," *IEEE COMSOC MMTC Commun. - Frontiers*, 2017.

[22] M.-P. Hosseini, "A cloud-based brain computer interface to analyze medical big data for epileptic seizure detection," in *Proc. 3rd Annu. New Jersey Big Data Alliance Symp.*, 2016.

[23] [Online]. Available: https://aws.amazon.com/hpc/

[24] [Online]. Available: http://buff.ly/1UcjCdi/

[25] S. Vahidian, S. Aïssa, and S. Hatamnia, "Relay selection for security-constrained cooperative communication in the presence of eavesdropper's overhearing and interference," *IEEE Wireless Commun. Lett.*, vol. 4, no. 6, pp. 577–580, Dec. 2015.

[26] M. P. Hosseini, "Developing a cloud base platform to improve health care in epileptic patients," *Annu. Res. Symp. Drexel IEEE Graduate Forum*, Philadelphia, USA, Mar. 2017.

[27] M. P. Hosseini, "Developing a Cloud Based Platform as a Service to Improve Public Health of Epileptic Patients in Urban Places," Reimagining Health in Cities: New Directions in Urban Health Research, Drexel University School of Public Health, Philadelphia, USA, 2015.

[28] M.-P. Hosseini, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, "Random subspace ensemble learning for epilepsy," *Artif. Intell. Med.*, 2018, to be published.

[29] S. Mercy, "Performance analysis of epileptic seizure detection using DWT & ICA with neural networks," *Int. J. Comput. Eng. Res.*, vol. 2, no. 4, pp. 1109–1113, 2012.

[30] M.-P. Hosseini, "Proposing a new artificial intelligent system for automatic detection of epileptic seizures," *J. Neurological Disorders*, vol. 3, no. 4, 2015.

[31] W. Zhang, et al., "Deep model based transfer and multi-task learning for biological image analysis," *IEEE Trans. Big Data*, 2016, doi: 10.1109/TBDATA.2016.2573280.

[32] M. Abavisani and V. M. Patel, "Multimodal sparse and low-rank subspace clustering," *Inf. Fusion*, vol. 39, pp. 168–177, 2017.

[33] M.-P. Hosseini, T. Tran, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, "Deep learning with edge computing for localization of epileptogenicity using multimodal rs-fMRI and EEG big data," in *Proc. IEEE Int. Conf. Autonomic Comput.*, 2017, pp. 83–92.

[34] S. Minaee, A. Abdolrashidi, and Y. Wang, "An experimental study of deep convolutional features for iris recognition," in *Proc. IEEE Signal Process. Med. Biol. Symp.*, 2016, pp. 1–6..

[35] M. Javanmardi, M. Sajjadi, T. Liu, and T. Tasdizen, "Unsupervised total variation loss for semi-supervised deep learning of semantic segmentation," arXiv:1605.01368, 2016.

[36] M.-P. Hosseini, T. Tran, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, "Multimodal rs-fmri and eeg for human brain-computer interface via deep learning," arXiv, 2017.

[37] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[38] M.-P. Hosseini, M.-R. Nazem-Zadeh, D. Pompili, K. Jafari-Khouzani, K. Elisevich, and H. Soltanian-Zadeh, "Comparative performance evaluation of automated segmentation methods of hippocampus from magnetic resonance images of temporal lobe epilepsy patients," *Med. Phys.*, vol. 43, no. 1, pp. 538–553, 2016.

[39] D. Knowles and Z. Ghahramani, "Infinite sparse factor analysis and infinite independent components analysis," in *Proc. Int. Conf. Ind. Compon. Anal. Signal Separation*, 2007, pp. 381–388.

[40] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Hoboken, NJ, USA: Wiley, 2004, vol. 46.

[41] B. Liu, "Composite differential search algorithm," *J. Appl. Math.*, vol. 2014, 2014, Art. no. 294703.

[42] P. Civicioglu, "Transforming geocentric cartesian coordinates to geodetic coordinates by using differential search algorithm," *Comput. Geosciences*, vol. 46, pp. 229–247, 2012.

[43] G. J. Ortega, R. G. Sola, and J. Pastor, "Complex network analysis of human ecog data," *Neuroscience Lett.*, vol. 447, no. 2, pp. 129–133, 2008.

[44] V. L. Towle, et al., "Identification of the sensory/motor area and pathologic regions using ECoG coherence," *Electroencephalography Clinical Neurophysiology*, vol. 106, no. 1, pp. 30–39, 1998.

[45] G. Alarcon, et al., "Origin and propagation of interictal discharges in the acute electrocorticogram. implications for pathophysiology and surgical treatment of temporal lobe epilepsy," *Brain*, vol. 120, no. 12, pp. 2259–2282, 1997.

[46] A. Hufnagel, M. Dümpelmann, J. Zentner, O. Schijns, and C. Elger, "Clinical relevance of quantified intracranial interictal spike activity in presurgical evaluation of epilepsy," *Epilepsia*, vol. 41, no. 4, pp. 467–478, 2000.

[47] K. Deb, "Multi-objective evolutionary algorithms: Introducing bias among pareto-optimal solutions," in *Advances Evolutionary Computing*. Berlin, Germany: Springer, 2003, pp. 263–292.

[48] M.-P. Hosseini, M. R. Nazem-Zadeh, F. Mahmoudi, H. Ying, and H. Soltanian-Zadeh, "Support vector machine with nonlinear-kernel optimization for lateralization of epileptogenic hippocampus in mr images," in *Proc. IEEE 36th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2014, pp. 1047–1050.

[49] K.-L. Hua, I. N. Sari, and M.-C. Yeh, "Human pose tracking using online latent structured support vector machine," in *Proc. Int. Conf. Multimedia Model.*, 2017, pp. 626–637.

[50] K. S. Pollard and M. J. van der Laan, "Resampling-based multiple testing: Asymptotic control of type i error and applications to gene expression data," 2003.

[51] L. Amini, et al., "Directed epileptic network from scalp and intracranial eeg of epileptic patients," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, 2009, pp. 1–6.

[52] G. Carter, "Time delay estimation for passive sonar signal processing," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 29, no. 3, pp. 463–470, Jun. 1981.

[53] R. T. Schirrmeister, et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, Aug 2017. [Online]. Available: http://dx.doi.org/10.1002/hbm.23730

[54] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," arXiv:1511.06448, 2015.

[55] M. Stead, et al., "Microseizures and the spatiotemporal scales of human partial epilepsy," *Brain*, vol. 133, pp. 2789–2797, 2010.

[56] B. H. Brinkmann, M. R. Bower, K. A. Stengel, G. A. Worrell, and M. Stead, "Large-scale electrophysiology: Acquisition, compression, encryption, and storage of big data," *J. Neuroscience Methods*, vol. 180, no. 1, pp. 185–192, 2009.

[57] T. A. Pedley, "Eeg patterns that mimic epileptiform discharges but have no association with seizures," in *Current Clinical Neurophysiology: Update on EEG and Evoked Potentials*. New York, NY, USA: Elsevier/North Holland, 1980.

[58] S. R. Benbadis and K. Lin, "Errors in eeg interpretation and misdiagnosis of epilepsy," *Eur. Neurology*, vol. 59, no. 5, pp. 267–271, 2008.

[59] M.-P. Hosseini, H. Soltanian-Zadeh, and S. Akhlaghpoor, "Computer-aided diagnosis system for the evaluation of chronic obstructive pulmonary disease on CT images," *Tehran Univ. Med. J. TUMS Publications*, vol. 68, no. 12, pp. 718–725, 2011.

[60] J. K. Zao, et al., "Pervasive brain monitoring and data sharing based on multi-tier distributed computing and linked data technology," *Frontiers Human Neuroscience*, vol. 8, 2014, Art. no. 370.

Mohammad-Parsa Hosseini received his BSc degree in electrical and electronic engineering in 2006. He studied MSc in biomedical engineering in 2007 and received a MSc degree in electrical and communication engineering in 2009. He was a graduate student in the ECE Dept. of Wayne State University, MI, and a graduate research assistant in Radiology and Research Administration Departments at Henry Ford Health System, MI in 2013. He is currently working towards his PhD degree in electrical and computer engineering and is a research and teaching assistant in the Dept. of ECE at Rutgers University, NJ, USA. He is collaborating with Medical Image Analysis Lab at Henry Ford Health System, MI, in medical imaging and data analysis. Also, he is collaborating with the Clinical Neurosciences Dept. Spectrum Health, MI, in computational neuroscience. His research focuses on deep/machine learning, signal/image processing, computer vision, big data, and cloud computing with future applications in the field of health care. He was a PhD research intern at Apple Inc., Silicon Valley, CA, during the summer of 2017. He has been teaching as an adjunct professor at several universities since 2009. He is a senior member of IEEE.

Dario Pompili received the Laurea (combined BSc and MSc) and the doctorate degrees in telecommunications and system engineering from the University of Rome "La Sapienza", Italy, and the PhD degree in ECE from the Georgia Institute of Technology, in 2001, 2004, and June 2007, respectively. He is an associate professor in the Department of ECE, Rutgers University. He is the director of the CPS Lab, which focuses on cloud computing, wireless communications and networking, acoustic communications, sensor networks, and data center management. He is a recipient of the NSF CAREER 2011, ONR Young Investigator Program 2012, and DARPA Young Faculty 2012 awards. He published more than a 100 refereed scholar publications: with about 6,500 citations, He has an h-index of 29 and a i10-index of 49 (Google Scholar, Feb'17). He is a senior member of both the IEEE Communications Society and the ACM.

Kost Elisevich, received the medical degree (MD) from the University of Western Ontario (UWO) in London, Canada, in 1978 and the PhD degree from the Departments of Anatomy and Physiology, University of Western Ontario (UWO) in London, Canada, in 1985, co-chair in the Department of Clinical Neurosciences for the Spectrum Health System and a fellow of the Royal College of Physicians and Surgeons of Canada and the American College of Surgeons. After receiving MD degree, he undertook his neurosurgical training in the Montreal Neurological Institute and other McGill University-affiliated hospitals. He was a member of the Department of Clinical Neurological Sciences (UWO) from 1987 to 1992 where he became associate professor and left for the Henry Ford Health System in Detroit, Michigan, where he served as system director of Epilepsy and Functional Neurosurgery, attaining a Clinical Professorship at Case Western University during this tenure. He joined with Spectrum Health and the College of Human Medicine of Michigan State University in Grand Rapids, Michigan in 2011 where he is professor of Neurosurgery and director of the Residency Program. His clinical interests include the epilepsy, brain tumors, pain, and headache disorders. He has more than 130 peer-reviewed publications in the basic and clinical neurosciences.

Hamid Soltanian-Zadeh received the BS and the MS degrees in electrical engineering: electronics from the University of Tehran, Iran, the MSc and the PhD degrees in electrical engineering: systems and bioelectrical sciences from the University of Michigan, MI, in 1986,1990, and 1992, respectively. Since 1988, he has been in the Department of Radiology, Henry Ford Health System, Detroit, MI, where he is currently a senior scientist and the director of the Medical Image Analysis Lab. Since 1994, he has also been in the Department of ECE, University of Tehran, where he is currently a full professor. He has active research collaboration with Wayne State University, Detroit, and the Institute for Research in Fundamental Sciences (IPM), Tehran. His research interests include medical imaging, signal and image processing and analysis, pattern recognition, and neural networks. He has co-authored more than 800 papers in journals and conference records or as book chapters in these areas. He has served on the scientific committees of several national and international conferences and review boards of more than 60 scientific journals. He is an associate member of the Iranian Academy of Sciences and has served on the study sections of the National Institutes of Health, National Science Foundation, American Institute of Biological Sciences, and international funding agencies. He is a senior member of IEEE.