Efficient Epileptic Seizure Prediction Based on Deep Learning

Hisham Daoud and Magdy A. Bayoumi, *Life Fellow, IEEE*

***Abstract*—Epilepsy is one of the world's most common neurological diseases. Early prediction of the incoming seizures has a great influence on epileptic patients' life. In this paper, a novel patient-specific seizure prediction technique based on deep learning and applied to long-term scalp electroencephalogram (EEG) recordings is proposed. The goal is to accurately detect the preictal brain state and differentiate it from the prevailing interictal state as early as possible and make it suitable for real time. The features extraction and classification processes are combined into a single automated system. Raw EEG signal without any preprocessing is considered as the input to the system which further reduces the computations. Four deep learning models are proposed to extract the most discriminative features which enhance the classification accuracy and prediction time. The proposed approach takes advantage of the convolutional neural network in extracting the significant spatial features from different scalp positions and the recurrent neural network in expecting the incidence of seizures earlier than the current methods. A semi-supervised approach based on transfer learning technique is introduced to improve the optimization problem. A channel selection algorithm is proposed to select the most relevant EEG channels which makes the proposed system good candidate for real-time usage. An effective test method is utilized to ensure robustness. The achieved highest accuracy of 99.6% and lowest false alarm rate of 0.004 h^{-1} along with very early seizure prediction time of 1 h make the proposed method the most efficient among the state of the art.**

***Index Terms*—Classification, deep learning, epilepsy, EEG, interictal, preictal, seizure prediction.**

I. Introduction

EPILEPSY is defined according to the International League Against Epilepsy (ILAE) report [1], as a neurological brain disorder identified by the frequent occurrence of symptoms called epileptic seizure due to abnormal brain activities. Seizure's characteristics include loss of awareness or consciousness and disturbances of movement, sensation or other cognitive functions. The overall incidence of epilepsy is 23–100 per 100,000. People at extremes of age are the most affected age group while the disease crests among young individuals in ages between 10 to 20 years old [2].

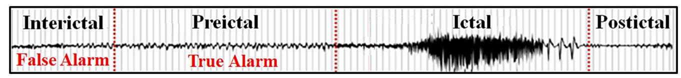

Fig. 1. Brain states in a typical epileptic EEG recording.

Epilepsy has a high disease burden where 50 million people worldwide have epilepsy and there are about two million new patients recorded every year. Up to 70% of the epileptic patients could be controlled by the Anti-Epileptic Drugs (AED) while the other 30% are uncontrollable [2].

Electroencephalogram (EEG) is the electrical recording of the brain activities and is considered the most powerful diagnostic and analytical tool of epilepsy. Physicians classify the brain activity of the epileptic patients according to the EEG recordings into four states: preictal state, which is defined by the time period just before the seizure, ictal state which is during the seizure occurrence, postictal state that is assigned to the period after the seizure took place and finally the interictal state which refers to the period between seizures other than the previously mentioned states [3], these four states are illustrated in Fig. 1.

Due to unexpected seizure times, epilepsy has a strong psychological and social effect in addition to it could be considered a life-threatening disease. Consequently, the prediction of epileptic seizures would greatly contribute to improving the quality of life of epileptic patients in many aspects, like raising an alarm before the occurrence of the seizure to provide enough time for taking proper action, developing new treatment methods and setting new strategies to better understand the nature of the disease. According to the above categorization of the epileptic patient's brain activities, the seizure prediction problem could be viewed as a classification task between the preictal and interictal brain states. An alarm is raised in case of detecting the preictal state among the predominant interictal states indicating a potential seizure is coming as shown in Fig. 1. The prediction time is the time before the seizure onset when the preictal state is detected.

In the literature, there are various methods proposed to address the seizure prediction problem trying to reach high classification accuracy with early prediction. Since EEG signals are different across patients due to the variations in seizure type and location [4], most seizure prediction methods are therefore patient-specific. In these methods, supervised learning techniques are used through two main stages which are feature extraction

Manuscript received March 1, 2019; revised May 4, 2019 and June 26, 2019; accepted June 28, 2019. Date of publication July 17, 2019; date of current version November 4, 2019. This paper was recommended by Associate Editor L. Najafizadeh. *(Corresponding author: Hisham Daoud.)*

H. Daoud is with the Center for Advanced Computer Studies, University of Louisiana at Lafayette, Lafayette, LA 70503 USA (e-mail: hgd8133@louisiana.edu).

M. A. Bayoumi is with the Department of Electrical and Computer Engineering, University of Louisiana at Lafayette, Lafayette, LA 70503 USA (e-mail: mab0778@louisiana.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBCAS.2019.2929053

and classification between preictal states and interictal states. In [5], the authors categorize the feature extraction schemes in terms of localization into univariate and bivariate and in terms of linearity into linear and nonlinear Multiple features are sometimes combined to capture the brain dynamics that ends up in dimensionality increase. The extracted features are used to train the classifier that could then be used for the analysis of new EEG recordings to predict the occurrence of the seizure by detecting the preictal state.

In the previous studies, the extracted features are categorized into three main groups: time domain, frequency domain and non-linear features. The authors in [6] used some statistical measures like variance, skewness and kurtosis as time domain features. In [7], the authors calculated the spectral power of the EEG signals for frequency domain analysis. Some nonlinear features that are derived from the dynamic systems' theory were investigated such as Lyapunov exponent [8] and dynamic similarity index [9]. Based on the selected features, a prediction scheme that detects the preictal brain state is implemented. Most of the previous work proposed machine learning based prediction schemes like Support Vector Machine (SVM). SVM classifier is used in numerous studies like [7], [10], [11] to predict the epileptic seizures. SVMs achieved outstanding results over other types of classifiers in terms of specificity and sensitivity [5].

Deep learning algorithms achieved great success in multiple classification problems for various applications like computer vision and speech recognition. Some previous work utilized deep learning in the classification stage for seizure prediction problem. In [12], the authors applied multi-layer perceptron to the extracted features. In [13] and [14], the authors used a convolutional neural network as a classifier that is applied on the extracted features from EEG data to predict seizures.

The main challenge of the previously proposed methods is to determine the most discriminative features that best represent each class. The computation time needed to extract these features depends on the process complexity and is considered another challenge especially in real-time application. Motivated by these challenges and due to the significance of the early and accurate seizure prediction, we developed deep learning based seizure prediction algorithms that combine the feature extraction and classification stages into a single automated framework.

In this paper, we aim at automatic extraction of the most important features by developing deep learning based algorithms without any preprocessing. Multi-Layer Perceptron is applied to the raw EEG recordings as a simple architecture of multiple trainable hidden layers, then Deep Convolutional Neural Network (DCNN) is used to learn the discriminative spatial features between interictal and preictal states. In another proposed model, Bidirectional Long Short-Term Memory (Bi-LSTM) Recurrent Neural Network is concatenated to the DCNN to do the classification task. An Autoencoder (AE) based semi-supervised model is proposed and pre-trained using transfer learning technique to enhance the model optimization and converge faster. For the system to be suitable for real-time usage, computation complexity should be considered, therefore we introduce a channel selection algorithm to select the best representing channels from the multi-channel EEG recording.

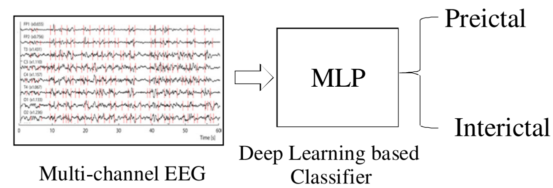

Fig. 2. Block Diagram of MLP based Seizure predictor.

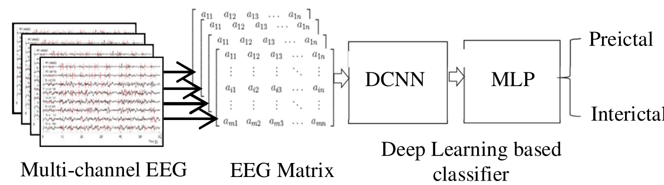

Fig. 3. Block Diagram of DCNN + MLP based Seizure predictor.

The used testing method proves the robustness of the proposed algorithms over different seizures.

II. Methodology

In this paper, we propose four deep learning based models for the purpose of early and accurate seizure prediction taking into account the real-time operation. The seizure prediction problem is formulated as a classification task between interictal and preictal brain states, in which a true alarm is considered when the preictal state is detected within the predetermined preictal period as shown in Fig. 1. In spite of the abundant research work done in seizure prediction, there is no standard duration for the preictal state. In our experiments, the preictal duration was chosen to be one hour before the seizure onset and interictal duration was chosen to be at least four hours before or after any seizure as in [15]. Raw EEG data without any preprocessing and without handcrafted features extracting is used as the input to all the models. The discriminative features are learned automatically using the deep learning algorithms in order to reduce the overhead and speed up the classification task. Due to the limited number of seizures for each patient, there is an imbalance between preictal and interictal samples. Obviously, the number of interictal samples is much larger than the number of preictal samples, and the classifiers tend to be more accurate toward the class with the larger number of training samples [16]. In our experiments, we selected the number of interictal samples to be equal to the number of preictal samples to make the data balanced. The EEG signals are divided to non-overlapping five seconds segments, each segment is considered as a training batch.

In our first model, Multi-layer Perceptron (MLP), a simple deep neural network, is trained on the selected patients to learn the network parameters that are able to do the classification task. The block diagram of the model is shown in Fig. 2. To enhance the classification accuracy, we propose the second model that relies on Deep Convolutional Neural Network (DCNN) which extracts the spatial features from different electrodes' locations and uses MLP for the classification task as illustrated in Fig. 3. In order to use DCNN, EEG data is represented by a matrix

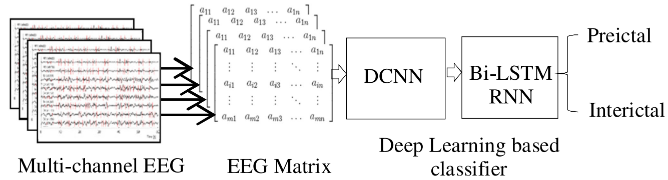

Fig. 4. Block Diagram of DCNN + Bi-LSTM based Seizure predictor.

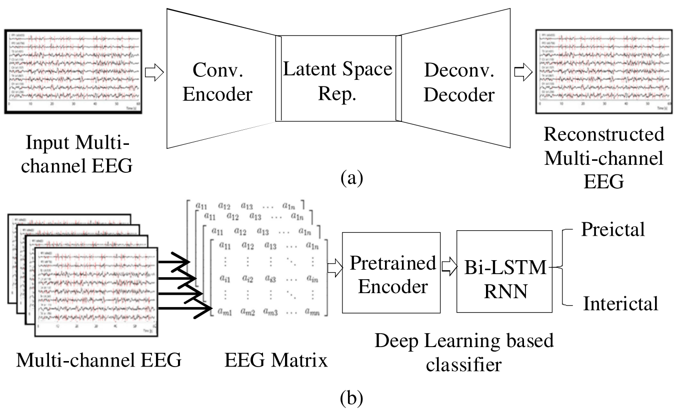

Fig. 5. Block Diagram of the semi-supervised DCAE + Bi-LSTM model, (a) pre-training phase of DCAE to generate the reconstructed EEG signals from the latent space representation through unsupervised learning and (b) pre-trained classifier that predicts seizures through supervised learning.

with one dimension is the number of channels and the other dimension is the time steps. In our third model, proposed in [17], DCNN is utilized and concatenated with a Bidirectional Long Short-Term Memory (Bi-LSTM) Network as the model back-end to do the classification as shown in Fig. 4. LSTM networks are known for their excellence in learning temporal features while maintaining long-time sequences dependencies which helps in early prediction. Prediction problems are handled better using Bi-LSTM as it uses information from both previous and next time instances. For the sake of training time reduction, we developed the fourth model that implements Deep convolutional Autoencoder (DCAE) architecture. In DCAE, we pre-trained the model front-end, DCNN, in an unsupervised manner. Then, the training process is launched with some initial values that will help the network to converge faster and enhance the network optimization which in turn reduce the training time and increase the accuracy. Transfer learning approach is used to train the DCAE to improve the generalization across different seizures for the same patient. After training the AE, the trained encoder is connected to Bi-LSTM network for classification. Fig. 5 illustrates the two parts of the DCAE model. We propose a channel selection algorithm to reduce the number of EEG channels which successively reduce the computation complexity and allocated memory making the system suitable for real-time application.

A. Dataset

In this paper, we trained the proposed models and evaluated their performance on the CHB-MIT EEG dataset recorded at

TABLE I
INFORMATION OF EEG DATA OF THE SELECTED PATIENTS

Case	ID-Gender-Age	Number of Seizures	Total Seizure Time (s)
1	1-F-11	7	442
2	3-F-14	7	402
3	7-F-14.5	3	325
4	9-F-10	4	276
5	10-M-3	7	447
6	20-F-6	8	294
7	21-F-13	4	199
8	22-F-9	3	204

Children's Hospital Boston [18], [19] which is publicly available [20]. The dataset composed of long-term scalp EEG data for 22 pediatric subjects with intractable seizures and one recording with missing data. The recordings were taken during several days after anti-seizure medication withdrawal to characterize their seizures and evaluate their candidacy for surgical intervention. Most cases have EEG recordings from surface electrodes of 23 channels in accordance with the International 10–20 system. The sampling rate of the acquired EEG signals is 256 samples per second with 16-bit resolution. There are some variations in many factors between all subjects such as interictal period, preictal period, number of channels, and recording continuity. Therefore, we chose eight subjects in this study such that the pre-determined interictal and preictal periods are satisfied, the recordings are not interrupted and the full channels' recordings are available. Table I summarizes the details about the EEG recordings used in our experiments.

B. Multi-Layer Perceptron

Multilayer Perceptron (MLP) is considered one of the most widely used artificial neural network (ANN). MLP consists usually of three successive layers, called: input layer, hidden layers, and output layer [21]. Deep ANNs are composed of multiple hidden layers that enable the network to learn the features better using the non-linear activation functions. The ANN idea is motivated by the structure of the human brain's neural system. A typical ANN is a buildup of connected units called neurons. These artificial neurons incorporate the received data and transmit it to the other associated neurons, much like the biological neurons in the brain. The output of a neuron in any ANN is computed by applying a linear or non-linear activation function to the weighted sum of the neurons' output in the preceding layer. When the ANN used as a classifier, the final output at the output layer indicates the appropriate predicted class of the corresponding input data.

In our first proposed seizure prediction model, Fig. 2, we apply the raw EEG after segmentation to MLP with four hidden layers as depicted in Fig. 6. The number of units in each layer is 300, 100, 50, 20 starting from the first hidden layer to the fourth one. The total number of trainable parameters is 8,870,291 which is considered high due to the fully connected architecture. The model is trained with backpropagation and optimized using RMSprop algorithm. The loss function used is the binary cross

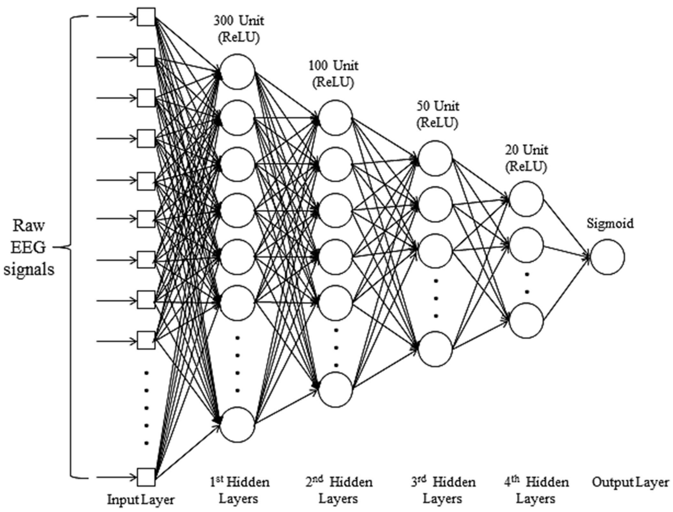

Fig. 6. The architecture of the proposed MLP based classifier.

entropy defined by (1).

$$l\left(y, \hat{y}\right) = -\left[y \log\left(\hat{y}\right) + \left(1 - y\right) \log\left(1 - \hat{y}\right)\right] \tag{1}$$

where $\hat{y}$ and y are the desired output and the calculated output respectively and $l(y, \hat{y})$ is the loss function.

Rectifier Linear Unit (ReLU) activation function [22], as defined by (2), is used across the hidden layers to add nonlinearity and to ensure robustness against noise in the input data.

$$f\left(x\right) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{2}$$

where x is the sum of the weighted input signals and $f(x)$ is the ReLU activation function.

Sigmoid activation function (3) is selected for the output layer. to predict the input data class.

$$p_i = \frac{1}{1 + e^{-x_i}} \tag{3}$$

where x_i is the sum of the weighted input signals and p_i is the probability of the input example being preictal.

C. Deep Convolutional Neural Network

Convolutional Neural networks (CNNs) have shown great success in different pattern recognition and computer vision applications [23]. This is due to the ability of CNN to automatically extract significant spatial features that best represents the data from its raw form without any preprocessing and without any human decision in selecting these features [24]. The sparse connectivity and parameter sharing of CNN give it high superiority regarding the memory footprint as it requires much less memory to store the sparse weights. The equivariant representation property of the CNN increases the detection accuracy of a pattern when it exists in a different location across the image [25]. A typical CNN formed of three types of layers: convolution layer, pooling layer and fully connected layer. The convolution layer is used to generate the feature map by applying filters with trainable weights to the input data. This feature map is then down-sampled by applying the pooling layer to reduce the features' dimension and therefore the computational complexity. Finally, the fully connected layer is applied to all the preceding layer's output to generate the one-dimensional feature vector. CNN is used as a feature extractor to replace the complex feature engineering used in previous work.

The proposed DCNN architecture model is shown in Fig. 7, in which the EEG segment is converted into a 2D matrix to be suitable for the DCNN. The architecture consists of four convolutional layers and three maximum pooling layers interchangeably. We chose the number of kernels in each convolution layer to be 32 with kernel size of 3 × 2 to cover the non-square matrix of EEG data. The maximum pooling layers have pool size of 2 × 2. RELU activation function is used across all the convolutional layers. Batch Normalization technique [26] is used to improve the training speed and reduce overfitting through adding some noise to each layer's activation.

The Batch Normalization Transform is defined as:

$$BN_{\gamma,\beta}\left(x_i\right) = \gamma \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \tag{4}$$

where x_i is the vector to be normalized in a mini-batch $\mathrm{B} = \{x_1, x_2, \ldots x_m\}$. μ_B and σ_B^2 are the mean and variance of the current mini-batch of x_i, respectively. ϵ is a constant added to the mini-batch variance for numerical stability. γ and β are learned parameters used to scale and shift the normalized value respectively [26].

The proposed DCNN architecture is used as the front-end feature extractor in our three proposed models in Figs. 3, 4, 5(b) which helps in spatial feature extraction from the different electrodes position on the scalp. The number of trainable parameters is drastically decreased when employing DCNN due to the weight sharing property. The number of trainable parameters in the second model, DCNN + MLP, is almost 520K, while in the third and fourth model, DCNN + Bi-LSTM and DCAE + Bi-LSTM, the number of trainable parameters is almost 28K.

D. Bidirectional-LSTM Recurrent Neural Network

Recurrent neural network (RNN) is a type of neural network that can maintain state along the sequential inputs. It can process a temporal sequence of data depending on the processing done on the previous sequences. This property of RNN makes it suitable for applications like prediction of time series data. The typical architecture of RNN is trained using backpropagation through time (BPTT) which has some drawbacks like exploding and vanishing gradients and information morphing.

Long Short Term Memory Networks (LSTMs) [27] are a type of RNN, implemented to overcome the problems of basic RNN. LSTMs are able to solve the problem of vanishing gradient by maintaining the gradient values during the training process and backpropagate it through layer and time, thus LSTM has the capability of learning long-term dependencies. LSTM cell, as shown in Fig. 8 consists of three controlling gates that could store or forget the previous state and use or discard the current state. Any LSTM cell computes two states at each time step: a cell state (c) that could be maintained for long time steps and a hidden state (h) that is the new output of the cell at each time

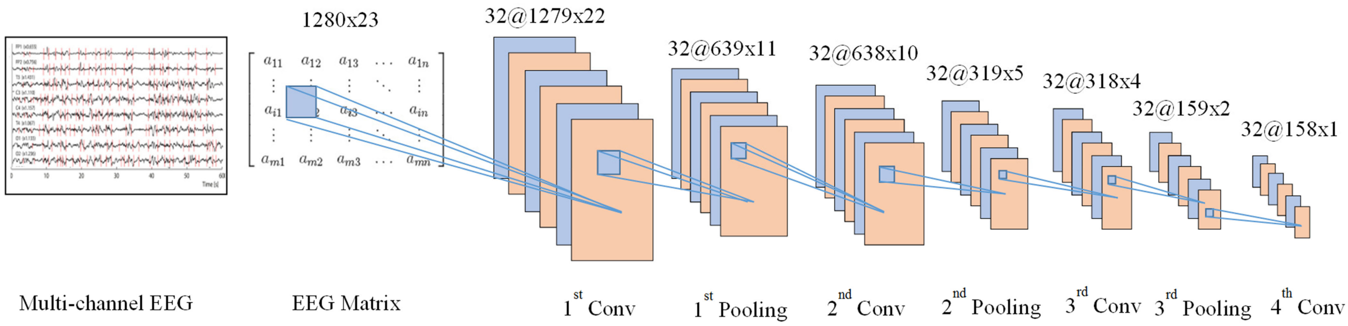

Fig. 7. The architecture of the proposed DCNN front-end in DCNN based models.

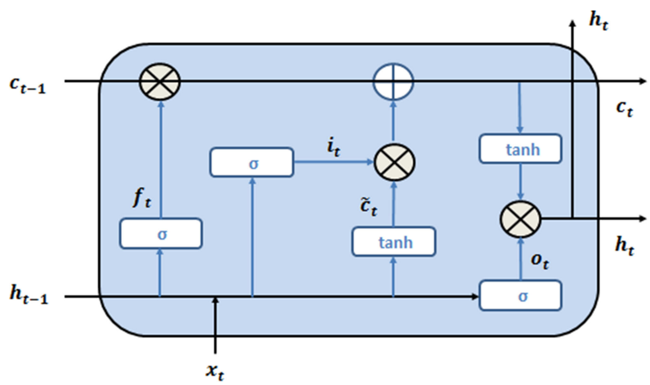

Fig. 8. Basic LSTM cell.

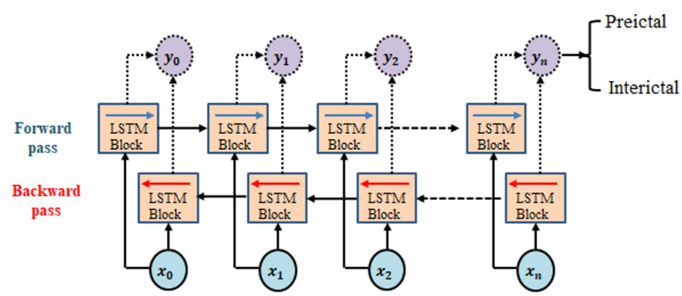

Fig. 9. The unrolled Bidirectional LSTM Network.

step. The mathematical expressions governing the cell gates' operation are defined as follows:

$$f_t = \sigma\left(W_{fh}h_{t-1} + W_{fx}x_t + b_f\right) \tag{5}$$

$$i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \tag{6}$$

$$o_t = \sigma\left(W_{oh}h_{t-1} + W_{ox}x_t + b_o\right) \tag{7}$$

$$\tilde{c}_t = tanh\left(W_{ch}h_{t-1} + W_{cx}x_t + b_c\right) \tag{8}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \tag{9}$$

$$h_t = o_t \circ \tanh\left(c_t\right) \tag{10}$$

where x_t is the input at time t, c_t and h_t are the cell state and the hidden state at time t respectively. W and b denote weights and biases parameters respectively. σ is the sigmoid function and $\circ$ is the Hadamard product operator. $\tilde{c}_t$ is a candidate for updating c_t through the input gate.

The input gate i_t decides whether to update the cell with a new cell state $\tilde{c}_t$, while the forget gate f_t decides what to keep or forget from the previous cell state and finally the output gate o_t decides how much information to be passed to the next cell.

Instead of using LSTM as the classifier, we used a Bidirectional-LSTM (Bi-LSTM) network [28] in which each LSTM block is replaced by two blocks that process temporal sequence simultaneously in two opposite directions as depicted in Fig. 9. In the forward pass block, the feature vector generated from the DCNN is processed starting from its first-time instance to the end, while the backward pass block processes the same segment in the reverse order. The network output at each time step is the combined outputs of the two blocks at this time step. In addition to the previous context processing in standard LSTM, Bi-LSTM processes the future context which enhances the prediction results. Using Bi-LSTM as a classifier enhances the prediction accuracy through extracting the important temporal features in addition to the spatial features extracted by the DCNN.

Bi-LSTM is used in two proposed models in Figs. 4, 5(b), as the back-end classifier that works on the feature vector generated by DCNN. The proposed network consists of a single bidirectional layer that predicts the class label at the last time instance after processing all the EEG segments as shown in Fig. 9. We chose the number of units, dimensionality of the output space, to be 20. Dropout regularization technique is utilized to avoid overfitting. The dropout is applied to the input and the recurrent state with factor of 10% and 50% respectively. The sigmoid activation function is used for prediction of the EEG segment's class and RMSprop is selected for optimization.

E. Deep Convolutional Autoencoder

Autoencoders (AEs) are unsupervised neural networks whose target is to find a lower dimensional representation of the input data. This technique has many applications like data compression [29], dimensionality reduction [30], visualizing high dimensional data [31] and removing noise from the input data. The AE network has two main parts namely, encoder and decoder. The encoder compresses the high dimensional input data into lower dimensional representation called latent space representation or bottleneck and the decoder is retrieving the data back to its original dimension. The simple AE uses fully

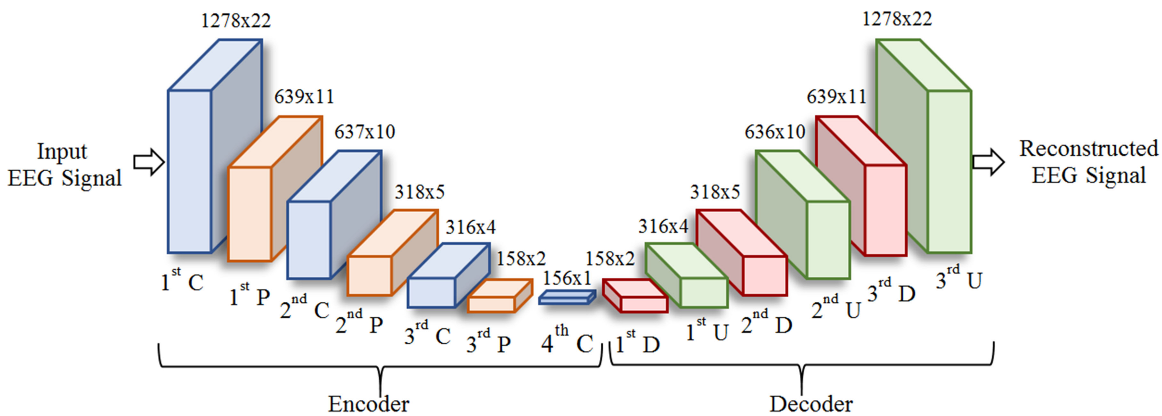

Fig. 10. The architecture of the proposed DCAE. C stands for convolution, P for pooling, D for deconvolution and U for upsampling layer.

connected layers for the encoder and decoder. The aim is to learn the parameters that minimize the cost function which expresses the difference between the original data and the retrieved one. Deep Convolutional Autoencoder (DCAE) replaces the fully connected layers in the simple AE with convolution layers.

Due to the limited EEG dataset for each patient, we decided to extend our work to develop an unsupervised training algorithm using DCAE as shown in Fig. 5(a). The proposed architecture of the DCAE model is depicted in Fig. 10. We used the same proposed DCNN model as an encoder and added the decoder network to build the DCAE. Unsupervised learning is deployed using transfer learning technique by training the DCAE on all the selected patients' data (not patient-specific). Transfer learning helps to obtain better generalization and enhance the optimization of our prediction model and therefore reducing the training time.

In the DCAE, Fig. 10, the encoder part consists of convolution and pooling layers interchangeably, while in the decoder part, the deconvolution and upsampling layers are used to reconstruct the original EEG segment. The encoder output is the latent space representation which is low dimensional features that best represent the EEG input segment. On the other hand, the decoder output is the reconstructed version of the original input. The learned encoder parameters are saved to be used later for training the prediction model in Fig. 5(b) allowing the training process to have a good start point instead of random initialization of the parameters which reduces the training time drastically.

Training of the DCAE is done using unlabeled EEG segments (balanced data of preictal and interictal segments) of all the selected patients. RELU activation function is used across all the convolutional layers. Batch Normalization technique is used to improve the training speed and to reduce overfitting. The DCAE is optimized using RMSprop optimizer. The mean square error is utilized as the cost function and is defined as

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(\overset{\prime(i)}{x} - x^{(i)} \right)^2 \tag{11}$$

where $x^{(i)}$ is the input EEG signal and $\overset{\prime(i)}{x}$ is the reconstructed EEG signal. m is the number of training examples and θ is the parameters being learned.

After DCAE training, the pre-trained encoder is used as a front-end of the fourth proposed model, DCAE + Bi-LSTM, as shown in Fig. 5(b) while the back-end is Bi-LSTM network. We used the same network architecture of the DCNN and Bi-LSTM that is used in the third model (DCNN + Bi-LSTM). Training of this model is done in a supervised manner to predict the patient-specific seizure onset. Since we used both unsupervised and supervised learning algorithms, this model is considered a semi-supervised learning model.

F. EEG Channel Selection

We introduce an EEG channel selection algorithm to select the most important and informative EEG channels related to our problem. Decreasing the number of channels helps with reducing the features' dimension, the computation load and the required memory for the model to be suitable for real-time application. The proposed channel selection algorithm is explained in Table II. We provide the algorithm with the EEG preictal segments for each patient and the measured prediction accuracy by running our fourth model, DCAE + Bi-LSTM using all channels. On the other hand, the algorithm will output the reduced channels that give the same accuracy by omitting redundant or irrelevant channels. We start by computing the statistical variance defined by (12) and the entropy defined by (13) for all the available channels (23 channels) of the preictal segments. Then, we select the channels with highest variance entropy product that provide the same given prediction accuracy. This is done through an iterative process by training our model on the reduced channels over each iteration. The variance is estimated as

$$\sigma^2(X_c) = \frac{1}{N} \sum_{i=1}^{N} (x_c(i) - \mu_c)^2 \tag{12}$$

where X_c, μ_c and N are the EEG data after normalization, mean and number of samples of channel c, respectively. The entropy of channel c is calculated as

$$H(X_c) = -\sum_{i=1}^{N} p(x_c(i)) \log_2 p(x_c(i)) \tag{13}$$

TABLE II
THE PROPOSED EEG CHANNEL SELECTION ALGORITHM

```
Algorithm 1: EEG Channel Selection Algorithm.
Input: Eight patients EEG preictal segments, the seizure
       prediction accuracy for each patient using all
       channels Acc[1 : 8]
Output: Ch_red [1 : 8] (array of reduced channels for each
        patient that give the same accuracy Acc)
Initialization: m = 8 (initial number of channels), done = 0
for patient ← 1 to 8 do
    m = 8, done = 0;
    for ch ← 1 to 23 do
        compute variance[ch];
        compute entropy[ch];
        compute variance[ch]*entropy[ch];
    end
    sort the 23 channels with highest variance entropy product
    first in Temp;
    while done ≠ 1 do
        select first m channels from Temp;
        train and test the model with m channels;
        compute the prediction accuracy Acc_new ;
        if Acc_new ≥ Acc[patient] then
            done ← 1;
            Ch_red [patient] ← Temp[1: m];
        else
            m ← m+1
        end
    end
end
```

where $p(x_c(i))$ is the probability mass function of the channel c having N samples.

In the channel selection algorithm, we chose the channels with the highest variance entropy product because we want to maximize both. We want to select the channel that has a high variance during the preictal interval and also provide the largest amount of information.

G. Training and Testing Method

In order to overcome the problem of the imbalanced dataset, we selected the number of interictal segments to be equal to the available number of preictal segments during the training process. The interictal segments were selected at random from the overall interictal samples. To ensure robustness and generality of the proposed models, we used the Leave-one-out cross validation (LOOCV) technique as the evaluation method for all of our proposed models. In LOOCV, the training is done N separate times, where N is the number of seizures for a specific patient. Each time, all seizures are involved in the training process except one seizure on which the testing is applied. The process is then repeated by changing the seizure under test. By using this method, we ensure that the testing covers all the seizures and the tested seizures are unseen during the training. The performance for one patient is the average across N trials and the overall performance is the average across all patients. 80% of the training data is assigned to the training set while 20% is assigned to the validation set over which the hyperparameters are updated and the model is optimized.

We evaluated the performance of our models by calculating some measures such as sensitivity, specificity, and accuracy on the test data. These measures are averaged across all patients. The prediction time of each model is recorded at the time of first preictal segment detection. The evaluation measures are defined as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{14}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{15}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

where TN, TP, FN and FP are the true negative, true positive, false negative and false positive respectively.

III. RESULTS

A. Performance Evaluation and Analysis

We evaluated our proposed patient-specific models on the selected patients by calculating some performance measures such as prediction accuracy, prediction time, sensitivity, specificity and false alarm per hour. The training time is also computed to evaluate our proposed channel selection algorithm. Table III shows the obtained values of these measures for the proposed four models which are MLP, DCNN + MLP, DCNN + Bi-LSTM, DCAE + Bi-LSTM. The fifth model, DCAE + Bi-LSTM + CS, is the same as the fourth one but with using the channel selection algorithm.

As could be noticed from Table III, MLP has the worst accuracy, sensitivity, specificity and false alarm rate among the proposed models and this is because the learning process in this model aims at updating the network parameters for the output to be close to the ground truth without extracting any features from the input data. The huge number of parameters in this model (around 9 million) is another drawback. The training time is moderate (7.3 min) due to network simplicity.

By introducing the DCNN as a front-end, we found around 10% enhancement in the accuracy, sensitivity and specificity and the false alarm rate is improved by 60%. This improvement is due to the ability of DCNN to extract the spatial features across different scalp positions to use it in discrimination between preictal and interictal brain states. On the other hand, the training time is increased by 5 min and this is due to the added computation complexity by the DCNN. The network parameters are drastically decreased because of the parameter sharing and sparse connectivity properties of the DCNN. In our third model, we used Bi-LSTM as the back-end along with DCNN and this model increase the accuracy to be 99.6%, the sensitivity to be 99.72% and the specificity to be 99.6%. The false alarm rate is enhanced a lot to reach 0.004 false alarm per hour. This improvement is due to using Bi-LSTM as a classifier instead of MLP. Bi-LSTM extracts temporal features from the input sequence which helps

TABLE III
PERFORMANCE EVALUATION OF THE PROPOSED MODELS

Proposed Model	Sensitivity	Specificity	Accuracy	False Alarm h^{-1}	Training Time (min)	No. of Parameters
MLP	84.67%	82.60%	83.63%	0.174	7.3	8,870,291
DCNN + MLP	95.41%	92.80%	94.10%	0.072	12.5	520,477
DCNN + Bi-LSTM	99.72%	99.60%	99.66%	0.004	14.2	27,657
DCAE + Bi-LSTM	99.72%	99.60%	99.66%	0.004	4.25	27,657
DCAE + Bi-LSTM + CS	99.72%	99.60%	99.66%	0.004	2.2	18,345

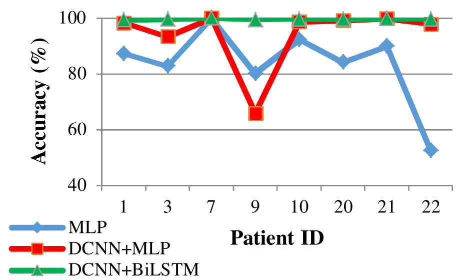

Fig. 11. The measured accuracy among three different proposed algorithms.

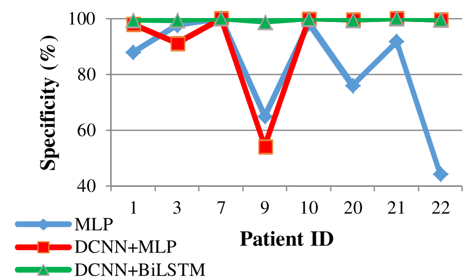

Fig. 13. The measured specificity among three proposed algorithms.

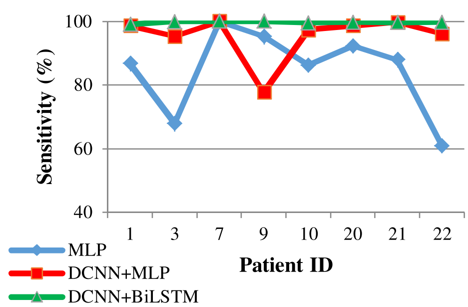

Fig. 12. The measured sensitivity among three different proposed algorithms.

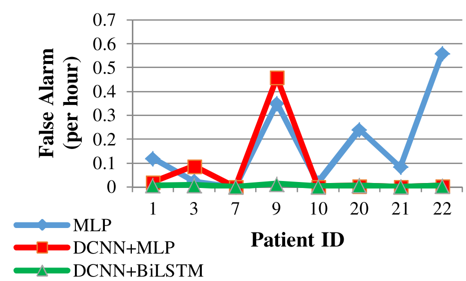

Fig. 14. The measured false alarm rate among three proposed algorithms.

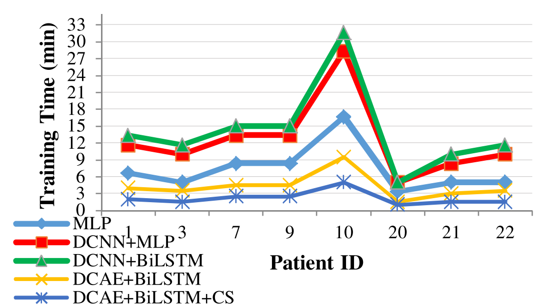

Fig. 15. The measured training time on the test set among five proposed algorithms: MLP, DCNN + MLP, DCNN + Bi-LSTM, DCAE + Bi-LSTM and DCAE + Bi-LSTM + CS.

in seizure prediction more accurately at the cost of training time which reached 14.2 min. The number of parameters is decreased by 94% by getting rid of the MLP. In the fourth model, DCAE is used to train the front-end part of our model. This improves the network optimization by starting the training with an initial set of parameters that makes the convergence process faster. As a result, the training time decreased to 4.25 min on average with the same highest performance. Utilizing the transfer learning technique reduces overfitting and generalizes better.

The proposed channel selection algorithm reduces the number of channels to 10 channels on average among all the selected patients instead of using all the channels which are 23 channels. Therefore, the computation complexity is reduced making the training time to reach 2.2 min on average with lowest number of parameters of around 18K which make this model suitable for real-time applications. All the obtained results are shown graphically for different models across the selected patients in (Figs. 11–15).

Regarding the prediction time, all the proposed models were able to accurately predict the tested seizures from the start of the preictal segments, thus the prediction time is one hour before the seizure onset or less in case of a shorter preictal segment.

B. Statistical Analysis

We performed Kruskal-Wallis test [32] as a nonparametric test statistic to compare the accuracy, sensitivity, specificity and false alarm rate of each model of the three basic models which are, MLP, DCNN + MLP, and DCNN + Bi-LSTM. The

TABLE IV
COMPARISON WITH OTHER SEIZURE PREDICTION METHODS APPLIED TO CHB-MIT DATASET

Ref.	Method		Data Selection	Sensitivity	Specificity	Accuracy	False Alarm h^{-1}	Prediction Time
	Feature Extraction	Classification						
[33]	ZC Interval	GMM	LPOCV	83.81%	N/A	N/A	0.165	19.8 min
[10]	ZC in WT	SVM	LPOCV	96%	90%	94%	N/A	N/A
[11]	Time, Freq., Graph	SVM	10-fold CV	85.75%	85.75%	85.75%	N/A	N/A
[13]	WT	CNN	10-fold CV	87.8%	N/A	N/A	0.147	5.8 min
[34]	Spectral Power	SVM	LOOCV	98.68%	N/A	N/A	0.046	42.7 min
Ours	**DCAE + Bi-LSTM**		**LOOCV**	**99.72%**	**99.60%**	**99.66%**	**0.004**	**1 hr**

Kruskal-Wallis test yielded (p-value < 0.05) for all the performance measures indicating statistical significance difference between the results among all the proposed models. For the accuracy (p-value $= 0.01$), for the sensitivity (p-value $= 0.006$), for the specificity (p-value $= 0.04$), and for the false alarm rate (p-value $= 0.04$).

C. Comparison With Other Methods

For further evaluation of our proposed method, we compared our achieved experimental results with previous work that have used the same dataset as shown in Table IV. While the same criterion to select the patients from the dataset is applied in this paper and [10], the other compared work employed different criteria which led to different selection of patients. In the presented previous work, some features were extracted like Zero-Crossing (ZC) interval in the EEG signals as in [33] and ZC of the Wavelet Transform (WT) coefficients of the EEG signals as in [10], WT of the EEG signals as in [13], spectral power as in [34] and set of features in time domain, frequency domain and from graph theory as in [11]. These studies used machine learning based classifiers like SVM or Gaussian Mixture Model (GMM). The authors in [13] used CNN as a classifier. The proposed method achieved the highest accuracy, sensitivity and specificity among others. Our prediction time is the earliest and the false alarm rate is the lowest.

IV. CONCLUSION

In this paper, a novel deep learning based patient-specific epileptic seizure prediction method using long-term scalp EEG data has been proposed. This method achieves a prediction accuracy of 99.6%, a sensitivity of 99.72%, a specificity of 99.60%, a false alarm rate of 0.004 per hour and prediction time of one hour prior the seizure onset. An important spatial and temporal feature from raw data are learned by the DCNN and Bi-LSTM networks respectively. DCAE based Semi-supervised learning approach is investigated with the transfer learning technique which led to reducing the training time. For the system to be suitable for real-time application, a channel selection algorithm is proposed which reduces the computational load and the training time. Using Leave-One-Out exhaustive cross-validation technique to test the proposed models proves the robustness and generality of our method against variation across various seizure types.

Our experimental results and the comparison with previous work demonstrate that the proposed method is efficient, reliable and suitable for real-time application of seizure prediction. This is by achieving accuracy higher than the state of the art with earlier prediction time to mitigate the potential life-threatening incidents for epileptic patients.

REFERENCES

[1] R. S. Fisher *et al.*, “ILAE official report: A practical clinical definition of epilepsy,” *Epilepsia*, vol. 55, no. 4, pp. 475–482, Apr. 2014.
[2] World Health Organization, *Neurological Disorders: Public Health Challenges*. Geneva, Switzerland: World Health Organization, 2006.
[3] C.-Y. Chiang, N.-F. Chang, T.-C. Chen, H.-H. Chen, and L.-G. Chen, “Seizure prediction based on classification of EEG synchronization patterns with on-line retraining and post-processing scheme,” in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Boston, MA, USA, 2011, pp. 7564–7569.
[4] “Epilepsy prevalence, incidence and other statistics,” Joint Epilepsy Council, Leeds, U.K., 2005.
[5] E. Bou Assi, D. K. Nguyen, S. Rihana, and M. Sawan, “Towards accurate prediction of epileptic seizures: A review,” *Biomed. Signal Process. Control*, vol. 34, pp. 144–157, Apr. 2017.
[6] A. Aarabi, R. Fazel-Rezai, and Y. Aghakhani, “EEG seizure prediction: Measures and challenges,” in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 1864–1867.
[7] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, “Epileptic seizure prediction using relative spectral power features,” *Clin. Neurophysiol.*, vol. 126, no. 2, pp. 237–248, Feb. 2015.
[8] L. D. Iasemidis, J. C. Sackellares, H. P. Zaveri, and W. J. Williams, “Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures,” *Brain Topography*, vol. 2, no. 3, pp. 187–201, 1990.
[9] M. Le Van Quyen, J. Martinerie, M. Baulac, and F. Varela, “Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings,” *Neuroreport*, vol. 10, no. 10, pp. 2149–2155, Jul. 1999.
[10] S. Elgohary, S. Eldawlatly, and M. I. Khalil, “Epileptic seizure prediction using zero-crossings analysis of EEG wavelet detail coefficients,” in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol.*, 2016, pp. 1–6.
[11] K. M. Tsiouris, V. C. Pezoulas, D. D. Koutsouris, M. Zervakis, and D. I. Fotiadis, “Discrimination of preictal and interictal brain states from long-term EEG data,” in *Proc. IEEE 30th Int. Symp. Comput.-Based Med. Syst.*, 2017, pp. 318–323.
[12] C. A. Teixeira *et al.*, “Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients,” *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 324–336, May 2014.
[13] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, “Focal onset seizure prediction using convolutional networks,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 2109–2118, Sep. 2018.

[14] H. G. Daoud, A. M. Abdelhameed, and M. Bayoumi, "Automatic epileptic seizure detection based on empirical mode decomposition and deep neural network," in *Proc. IEEE 14th Int. Colloq. Signal Process. Appl.*, 2018, pp. 182–186.

[15] "American epilepsy society seizure prediction challenge," Accessed: Jun. 17, 2018. [Online]. Available: https://www.kaggle.com/c/seizure-prediction

[16] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial: special issue on learning from imbalanced data sets," *SIGKDD Explorations*, vol. 6, pp. 1–6, Jun. 2004.

[17] H. Daoud and M. Bayoumi, "Deep Learning based Reliable Early Epileptic Seizure Predictor," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, Cleveland, OH, USA, 2018, pp. 1–4.

[18] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," M.Sc. Thesis, Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.

[19] 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. 2000.

[20] "CHB-MIT scalp EEG database," Accessed: May 2, 2019. [Online]. Available: https://physionet.org/pn6/chbmit/

[21] N. Siddique and H. Adeli, *Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing*. Hoboken, NJ, USA: Wiley, 2013.

[22] R. H. R. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit," *Nature*, vol. 405, no. 6789, pp. 947–951, Jun. 2000.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, Inc., 2012, pp. 1097–1105.

[24] Y. Bengio, Y. Lecun, and Y. Lecun, "Convolutional networks for images, speech, and time-series," in *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press, 1995.

[25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc.32nd Int. Conf. Mach. Learn.*, Feb. 2015, vol. 37, pp. 448–456.

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[28] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2005, vol. 4, pp. 2047–2052.

[29] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proc. ICML Workshop Unsupervised Transfer Learn.*, 2012, pp. 37–49.

[30] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," May 2014, *arXiv:1312.6114*.

[31] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

[32] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *J. Amer. Statist. Assoc.*, vol. 47, pp. 583–621, Dec. 1952.

[33] A. Shahidi Zandi, R. Tafreshi, M. Javidan, and G. A. Dumont, "Predicting epileptic seizures in scalp EEG based on a variational Bayesian Gaussian mixture model of zero-crossing intervals," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 5, pp. 1401–1413, May 2013.

[34] 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.

Hisham Daoud received the B.Sc. and M.Sc. degrees from Cairo University, Giza, Egypt, and the Ph.D. degree from Ain Shams University, Cairo, Egypt, in 2004, 2007, and 2014, respectively, all in electronics and communications engineering.

Since 2004, he has held multiple positions in both industry and academia. He is currently with the University of Louisiana, Lafayette, LA, USA. His research interest includes biomedical signal processing, machine learning, deep learning, and neuromorphic computing.

Dr. Daoud was the recipient of the IEEE CASS Student Travel Award, Best Paper Award in the 14th IEEE Colloquium on Signal Processing and its Applications Conference (CSPA'2018). He has served as a Reviewer for several IEEE conferences and journals.

Magdy A. Bayoumi (LF'16) received the B.Sc. and M.Sc. degrees in electrical engineering from Cairo University, Giza, Egypt, in 1973 and 1977, respectively, the M.Sc. degree in computer engineering from Washington University, St. Louis, MO, USA, in 1981, and the Ph.D. degree in electrical engineering from the University of Windsor, ON, Canada, in 1984.

He is currently the Department Head of W. H. Hall Department of Electrical and Computer Engineering. He is the Hall Endowed Chair in computer engineering. He was the Director of the Center for Advanced Computer Studies and the Department Head of Computer Science Department. He was also, the Loflin Eminent Scholar Endowed Chair in computer science, all at the University of Louisiana at Lafayette where he has been a faculty member since 1985. He has graduated about 100 Ph.D. and 150 M.Sc. students, authored/coauthored about 600 research papers and more than 10 books. He was the Guest/Coguest Editor of more than 10 special journal issues, the latest was on Machine to Machine Interface. He has served the IEEE Computer, Signal Processing, and Circuits & Systems (CAS) societies. He is currently the Vice President of Technical Activities of IEEE RFID council and he is on the IEEE RFID Distinguished Lecture Program (DLP). He is a member of IEEE IoT Activity Board. He was the recipient of the many awards, among them; the IEEE CAS Education award and the IEEE CAS Distinguished Service award. He was on the IEEE DLP programs for CAS and Computer societies. He was on the IEEE Fellow Selection Committee. He has been an ABET Evaluator and he was an ABET Commissioner and Team Chair. He has given numerous keynote/invited lectures and talks nationally and internationally. He was the General Chair of IEEE ICASSP 2017 in New Orleans. He, also, chaired many conferences including ISCAS 2007, ICIP 2009, and ICECS 2015. He was the Chair of an international delegation to China, sponsored by People-to-People Ambassador, 2000. He received the French Government Fellowship, University of Paris Orsay, 2003–2005 and 2009. He was a Visiting Professor with the King Saud University. He was a United Nation Visiting Scholar. He has been an advisor to many EE/CMPS Departments in several countries. He was on the State of Louisiana Comprehensive Energy Policy Committee. He was the Vice President of Acadiana Technology Council. He was on the Chamber of Commerce Tourism and Education committees. He was a member of several delegations representing Lafayette to international cities. He was on the Le Centre International Board. He was the General Chair of SEASME (an organization of French Speaking cities) conference in Lafayette. He is a member of Lafayette Leadership Institute; he was a founding member of its executive committee. He was a Column Editor for Lafayette Newspaper; the Daily Advertiser.