2018 17th IEEE International Conference on Machine Learning and Applications

Semi-supervised Deep Learning System for Epileptic Seizures Onset Prediction

Ahmed M. Abdelhameed and Magdy Bayoumi, *Fellow, IEEE*
The Center for Advanced Computer Studies,
University of Louisiana at Lafayette,
Louisiana 70503, USA
Emails: {amb7251, mab0778}@louisiana.edu

***Abstract*— The advance prediction of seizures before its onset has been a challenging task for scientists for a long time. It is still the epileptic patients' hope to find an effective way of preventing seizures to improve the quality of their lives. In this paper, using an innovative mixing of unsupervised and supervised deep learning techniques, we propose a novel epileptic seizure prediction system using electroencephalogram (EEG) recordings from the human brains. The proposed system is built upon classifying between the interictal and the preictal brain states. The proposed system uses two-dimensional deep convolutional autoencoder for learning the best discriminative spatial features from the multichannel unlabeled raw EEG recordings. A Bidirectional Long Short-Term Memory recurrent neural network is used for classification based on the temporal information. To help achieve faster learning and reliable convergence for our system, the transfer learning technique is used for initializing the weights for the patient-specific networks. Within, up to one hour of prediction window, our system achieved an average sensitivity of 94.6% and average low false prediction alarm rate of 0.04FP/h which makes it one of the most efficient among state-of-the-art methods.**

Keywords—EEG signals; epileptic seizure prediction; convolutional autoencoders; bidirectional long short-term memory; transfer learning; classification.

I. INTRODUCTION

Epilepsy is a persistent neurological disorder arising from anomalies of the electrical activity in the brain. Over 60 million individuals of all ages are diagnosed with epilepsy, which makes it one of the most widespread neurological diseases worldwide [1]. Epilepsy is characterized by its recurring seizures, which are transient events of partial or whole abnormal involuntary body movements, that might be also associated with loss of consciousness. Even though the epileptic seizures are occurring very infrequently in each patient, its subsequent impacts on the mental, psychological and physical communications of the patients make epileptic seizures diagnosis is of ultimate significance.

Electroencephalograms (EEGs) [2] has for a long time been broadly used by physicians for epilepsy diagnosis because of feasible reasons like its accessibility, effortlessness, and low cost. By positioning several electrodes along the scalp, EEG can measure the voltage oscillations coming from the ionic current flows inside the brain. These voltage oscillations that correspond to brain neural activities are then translated into time series which are called signals. EEG is a powerful diagnostic tool that can be utilized precisely to capture and denote the epileptic signals in the events of seizures. EEG signals help draw a relatively clear distinction between different epileptic brain states.

Epilepsy patients can be treated either using anti-epileptic drugs or using surgical procedures by removing the affected brain part in case the patient's epilepsy has anti-drug resistance [3]. However, scientists are still seeking more effective seizure management treatments. The early prediction of epileptic seizures guarantees enough time before it really happens. That is extremely helpful since the seizure attack can be treated or completely avoided by using the suitable medication.

Recent studies have increased the scientists believe in the existence of a transitional state that can last from minutes to hours prior to epileptic seizure onsets [4]. The detection of this state would be very helpful for early seizure prediction. Using EEG signal, we can differentiate between four different states of the epileptic seizures: the preictal state, that appears before the seizure onset, the ictal state that begins with the the onset of the seizure, the postictal state that lasts for a short time after ictal state, and interictal state that starts after the postictal state of the first seizure and ends before the start of preictal state of consecutive seizure. Fig.1 shows epileptic brain states using four-channels EEG recording.

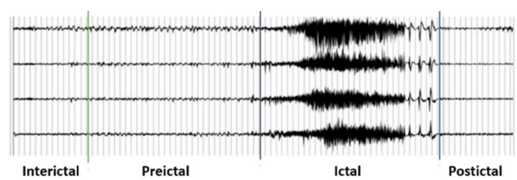


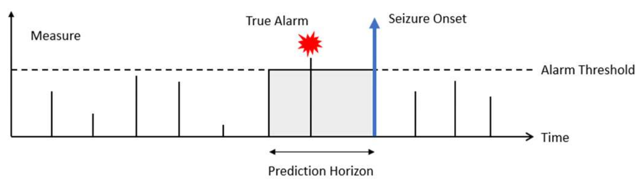
Figure 1. Seizure states

There is still no agreement about the exact duration of each of the brain states for epileptic patients. The determination of the starting and the ending time of each state is subjective as it is usually decided by human cogitation. It is believed that upcoming seizure prediction can be achieved by successfully detecting the beginning of the preictal state and distinguishing it

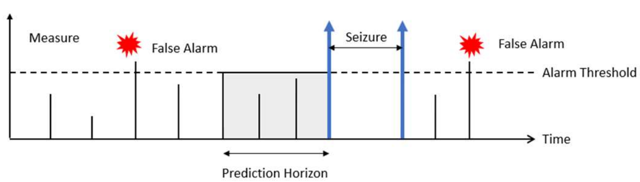
978-1-5386-6805-4/18/$31.00 ©2018 IEEE
DOI 10.1109/ICMLA.2018.00191

from the interictal state. This problem is somehow complicated because there is no specific definition for the transition between interictal state and preictal state of the EEG signal.

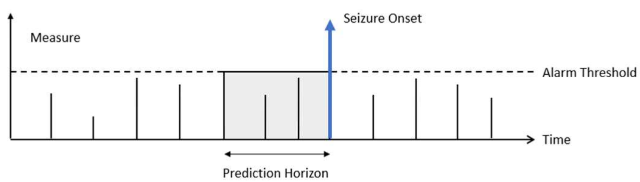
EEG-based prediction techniques mostly work by calculating and analyzing characteristic measure(s) in small windows of EEG data. A window size can range from 1 second to 30 seconds. This measure can be used to characterize a single EEG channel (univariate) or relations between two or more EEG channels (multivariate). A typical prediction algorithm starts by analyzing an EEG window that subsequently moves towards the beginning of the seizure. This window's measure is quantified and then tested against a certain threshold. Based on this test, the algorithm then issues a warning in the form of a false or true alarm. To judge this alarm, a prediction horizon or time must be predefined. This prediction horizon is allocated just before the seizure onset and may range from several minutes to a few hours. If the alarm fires within the prediction horizon period, this it is a true alarm (true positive). If the alarm fires before the starting of the prediction horizon or after the end of the seizure, it is a false alarm (false positive). If the alarm didn't fire within the prediction horizon, that means the seizure is not predicted at all (false negative). Fig.2 show different scenarios for seizures prediction alarms.



a) True Positive Alarm



b) False Positive Alarm



c) False Negative Alarm

Figure 2. Seizures prediction alarms

So, seizures prediction process can be considered as a classification problem between EEG signals to differentiate the interictal state from the preictal state. The desired goal is to get a true alarm as soon as possible before the start of the prediction horizon and at the same time avoid false prediction alarms.

In the literature, seizure prediction methods usually consist of two main phases, each composed of different steps. In the first phase (feature extraction phase), measures or features are computed from preprocessed raw EEG data signals over time. These extracted features should be the most representative or characterizing features that will be used later in the second phase. In the second phase (classification phase), a classification algorithm is applied to the extracted features to differentiate between preictal and interictal states.

Due to the non-stationary nature of the EEG signals, features extraction usually requires considerable work and a great deal of mastery to analyze the signal either in the time domain or frequency domain or even in time and frequency domains simultaneously [5]. Therefore, this task needs a lot of knowledge in different areas of expertise to ideally extract and select the best-representing features from different kinds of epileptic brain states in various subjects [6]. Many successful epileptic seizure prediction studies that use linear and non-linear extracted features are reported in the literature. Some of these methods are using autoregressive coefficients, cross-correlation, spike rates, wavelet analysis, largest Lyapunov exponent and phase synchronization [7]-[12]. Even with achieving very good results in the literature, it is not predominantly ensured that the manually extracted features will result in an optimum classification or prediction system. In this manner, it would be much better if we managed to design a reliable and productive system which could automatically extract the best representative features from the EEG raw signals with minimal preprocessing. This system; at the same time; would carry out seizures' states classification very dependably and efficiently.

The recent advances in machine learning techniques and especially the breakthrough of deep learning strategies in building unsupervised and supervised learning models have been proved to outperform the human engineered features extraction and selection in numerous disciplines such as computer vision, speech recognition, natural language processing and biomedical applications [13]. Different deep learning architectures have been utilized in the classification of EEG brain signals. Shallow and deep artificial neural networks (ANNs) were applied only as classifiers after performing the features' extraction task using other different techniques. In a different way of application, deep convolutional neural networks (CNNs) have been applied as features' extractors but only after transforming raw EEG signals using wavelets or STFT [14]-[15]. Applying CNNs solely as features' extractors doesn't guarantee to achieve high classification accuracies as it is only capable of learning excellent spatial patterns but still doesn't exploit the EEG signal's temporal patterns.

In this work, a semi-supervised deep learning system is proposed for predicting seizures onset from raw EEG data directly without the need to apply any of the features' extraction methods. The system is capable of automatically classifying the epileptic EEG signals into interictal and preictal with a minimum false positive rate. Due to the variations between EEG signals among the epileptic patients, the proposed system will be patient-specific. Our prediction system uses a two-dimensional (2D) deep convolutional autoencoder for learning spatial features from the EEG multichannel data and bidirectional long short-term memory (Bi-LSTM) recurrent neural network for making use of the long-term temporal dependencies that exist in the EEG time signals. Our system is semi-supervised because it combines learning from unlabeled and labeled data in two separate stages.

One of the advantages of our system is that there is no need to retrain the whole system again from scratch to make it work with each different patient. Our system takes advantage of one of the best features of deep learning which is transfer learning. So, after training and testing the system on a combined dataset (Interictal and Ictal) from two selected patients, a transfer learning mechanism involving using the weights of the pre-trained autoencoder is used to initialize the training for a similarly structured network for other patients. This method of transfer learning offers a very good way to accurately and fast train another supervised classification network instead of initializing its training with random weights. By applying this methodology, this work managed to obtain state of the art results for the prediction task

The remaining sections of this paper are organized as follows: in section two, we introduce the system architecture, the dataset and then explain the different methodologies used in our proposed work. Section three describes and discusses the results of our experiments compared to previous work. Finally, in section four, a review and conclusion of our work are provided.

II. METHODOLOGY

A. System Architecture

The objective of this work is to build up an efficient and reliable patient-specific epileptic seizure prediction system. This can be achieved by classifying raw EEG signals' segments into two classes of brain states which are interictal and preictal. The system keeps away from the burden resulting from preprocessing and features extraction. Two-dimensional convolutional autoencoder consists of several layers is used for learning inherent signals features from the raw EEG signals. An autoencoder consists of two parts: an encoder and a decoder. The encoder is used for compressing signals into low dimensional representation and the decoder is used in a reverse way to reconstruct the original signal. Our convolutional autoencoder is trained in an unsupervised way using unlabeled EEG signals from two preselected patients. After training, the encoder and decoder subnetworks have their optimum weights. The trained encoder subnet is then augmented into a bidirectional LSTM recurrent neural network for building a classifier. This LSTM part of the network together with only later layers of the encoder are retrained in a supervised way using class-labeled training EEG dataset and the network's accuracy for assigning the correct signal class label is measured using a testing dataset. Fig.3 shows a block diagram of all system's components.

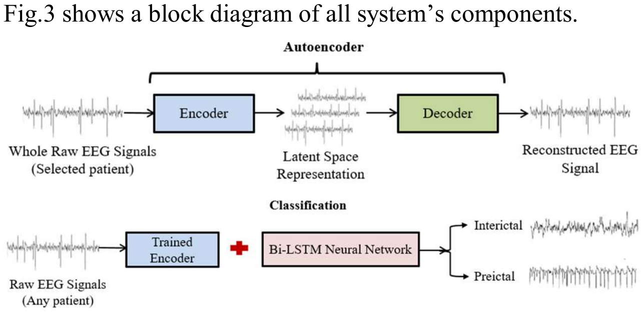


Figure 3. Block diagram of the seizure prediction system

After training and testing the classification and prediction accuracy of the system on the selected patients, a transfer learning using the pre-trained encoder part with its saved weights is utilized to initialize another similar supervised network that can be used to do the classification task and prediction for other patients.

B. Dataset

In this work, patients' data is obtained from the CHB-MIT EEG Database [16], which consist of scalp EEG recordings from pediatric patients with intractable seizure. All selected signals are captured using up to 23 electrodes and sampled at 256 Hz using the International 10-20 system of EEG electrode positions. For the analysis, the selected patients have a considerable number of seizures that are adequately distributed over time. For each patient and seizure, data segments have been split from the raw EEG recordings containing interictal and preictal data. The dataset for the preictal state was selected such that it will end just at the onset of the seizure and start at most one hour before it. When the seizure starts earlier than this one-hour limit in the raw EEG data, all recordings until the seizure's onset are considered as the prediction horizon. The interictal periods are determined based on the condition that they must be at least 4 hours away before either next seizure's beginning or after the previous seizure terminates.

C. 2D Convolutional Autoencoder

An autoencoder is an unsupervised neural network based learning algorithm that can be used for data compression and dimensionality reduction. The autoencoder takes input information and compresses (encodes) it and then attempts to reconstruct (decode) it using minimal possible representation. The compression in autoencoders is achieved by training the network until the loss function between the original input and the reconstructed one is minimal. The 2D convolutional autoencoder applied in our system is a regular deep autoencoder stacked with convolution and pooling layers instead of fully connected layers.

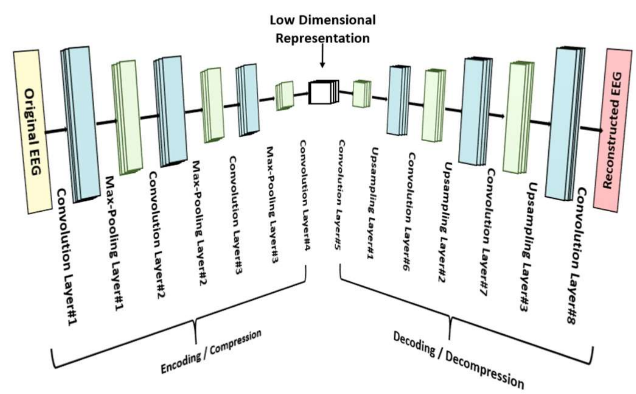


Figure 4. 2D Convolutional Autoencoder

The encoder part of the autoencoder consists of convolutional layers and max-pooling layers used interchangeably. These layers are responsible for learning EEG signals spatial features and dimension reduction by downsampling. The result of the last convolutional layer of the encoding network is a low dimensional representation called latent space representation or bottleneck. The decoding network consists of two types of layers: convolutional layers and upsampling layers that are used interchangeably too. The upsampling layers do the opposite work of the max-pooling layers. The result of the last convolutional layer in the decoding network is a reconstructed version of the original input EEG segment as shown in Fig.4. Our proposed 2D convolutional autoencoder is trained using two selected patients' unlabeled EEG segments containing both interictal and preictal data. The encoder weights are saved to be used later for training other networks for the classification task.

The activation function used in the convolutional and pooling layers is the rectifier linear unit (RELU) [17] defined by (1):

$$f(x) = \begin{cases} x & if\ x > 0 \\ 0 & if\ x < 0 \end{cases} \qquad (1)$$

The (RELU) function is used for its nonlinearity to ensure the robustness of the proposed system against noise in the input signals. The autoencoder is optimized using the RMSprop optimizer and uses the binary cross entropy loss function defined by (2).

$$l(y, \hat{y}) = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})] \qquad (2)$$

where y is the calculated output and $\hat{y}$ is the desired output.

A batch normalization layer is used between the convolutional and the pooling layers of the autoencoder to speed up the training process and to avoid overfitting in the final system. The batch normalization transform [18] is defined as in (3):

$$BN_{\gamma,\beta}(x_i) = \gamma \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \qquad (3)$$

where x_i is the vector to be normalized in a mini-batch $B=\{x_1, x_2, \ldots, x_m\}$. μ_B and σ_B^2 are the expectation and variance of the current mini-batch of x_i, respectively. ϵ is a constant added to the mini-batch variance for numerical stability, and β is a parameter to shift the normalized values.

D. Bidirectional LSTM

The recurrent neural networks (RNNs) are a different type of neural networks that can save states between different sequential inputs. RNNs can process a temporal series of data without the need to show the complete sequence to the network at once. An RNN is a composition of identical feedforward neural networks (RNN cells), one for each step in time. LSTMs are one of the variations of the traditional RNNs that have been implemented to overcome typical RNNs training problems like exploding and vanishing gradients and information morphing. LSTMs introduced the idea of memory cells (units) that have controlling gates. LSTMs managed to maintain gradients values during training and at the same time keep long-term temporal dependencies between inputs. Fig.5 shows the structure of a single LSTM cell. Note that x_t is the input at time t. f_t, $\tilde{c}_t$ and o_t are the forget, input and output gates. h_{t-1} and c_{t-1} are the previous shadow and cell states. h_t and c_t are the next states transferred to the next cell [19].

An LSTM block may contain one or many LSTM cells. One LSTM block typically processes a sequence's time instances consecutively in one direction. At a single point in time, in a layer of an LSTM network, the LSTM block calculates an output in addition to new state information based only on the current input and the previous state. This information is then feedbacked into the same LSTM block and therefore the process continues until the last input is processed and the last output is calculated. LSTM networks consisting of this type of blocks are called unidirectional LSTMs.

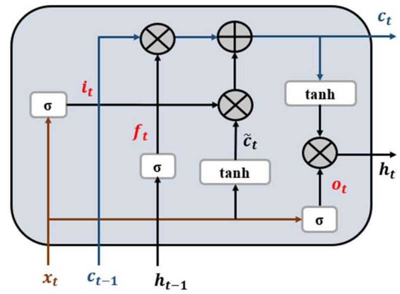


Figure 5. A single LSTM Cell

The bidirectional LSTM (Bi-LSTM) network architecture is quite like that of the unidirectional LSTMs. The difference is that in a Bi-LSTM, a single layer consists of two LSTM blocks instead of one. Both LSTM blocks simultaneously process a sequential input in two opposite directions as shown in Fig.6. Each block calculates its own output. At each instance of the input, the two outputs generated from both LSTM blocks are combined to calculate the final output at this instance.

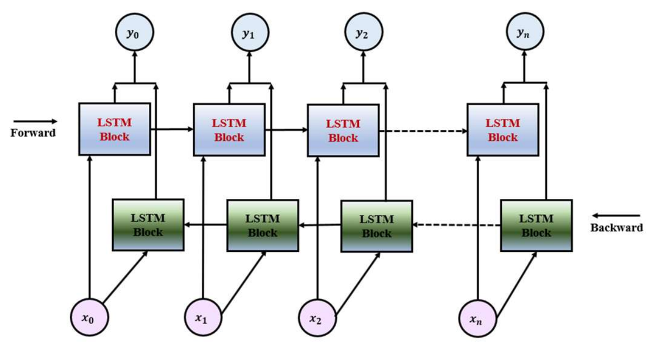


Figure 6. Bidirectional LSTM

The advantage of using a Bi-LSTM network in comparison to the unidirectional one is that it takes into attention the temporal dependencies between the current input instance and its previous and subsequent equivalents. This advantage makes it very promising for achieving higher classification results for sequenced inputs. The inputs to the Bi-LSTM are the smaller size sequence vectors generated from the encoder last layer. The average of the outputs computed after passing the whole input sequences in two opposite directions is then used for classification.

This work deploys a Bi-LSTM network; consisting of one layer; after the encoder network. To overcome overfitting while training the final network, the Dropout regularization technique is used. The Adam algorithm is used as an optimizer. The

Sigmoid activation function is used to predict the final class label for the EEG input signal.

E. Data Preparation and Training

(1) Data Selection and Preparation

In this work, EEG data of 12 patients with a total number of 56 seizures, are chosen for experiments. These patients' data share the exact electrodes placement and number of channels while being recorded. The specific interictal and preictal data segments, selected from all patients, are further divided into smaller ones such that each subsegment represents 4 seconds of EEG recordings. The dimension of each of these segments is (1024 x 23) because data is sampled at 256 points per second and the number of channels is 23. Two patients are randomly selected for the training of the autoencoder.

(2) 2D Convolutional Autoencoder Training

The convolutional autoencoder has two parts. The first part is the encoder which consists of four 2D convolutional layers and three max-pooling layers. The second part is the decoder which consists of four 2D convolutional layers and three upsampling layers. For the selected patients' data, both unlabeled interictal and preictal are shuffled together to make one dataset. While training the autoencoder, 80% of the dataset is selected randomly and used as a training set and the remaining 20% of the dataset is used as a testing test. The autoencoder is trained for 100 epochs with batch size equals to 10. The weights of the encoder part are saved to be used for weights initialization while training other patients' networks.

(3) Bi-LSTM Training

The encoder part of the 2D convolutional autoencoder is augmented by the Bi-LSTM network. The new network is then trained in a supervised way using labeled interictal and preictal datasets. As this new network is patient-specific, so during the training process, some of the early layers of the encoder network may or may not have its weights fixed. The new network is trained for 50 or100 epochs with batch size equals to 10 or 20. These variations are experimented based on the selected patient.

(4) Classification

To evaluate the robustness and accuracy of the system, a repeated cross-validation approach is used. If a patient has K seizures, then he will have K preictal segments. A whole one of the K preictal segments is kept out for testing, while the other (K-1) preictal segments are used for training and validation. Interictal segments are also randomly divided into K parts such that (K-1) parts are used for training and validation and the rest for testing. Both (K-1) interictal and preictal parts are mixed together to create a single dataset. This dataset is divided into training and validation. The system is trained to optimize the validation accuracy and minimize a loss function and then tested against the left-out preictal and ictal segments. The above procedure is repeated for every patient for all interictal and preictal data segments. Fig.7 shows samples for the classification accuracies and loss functions' curves while training and testing one of the patient's system.

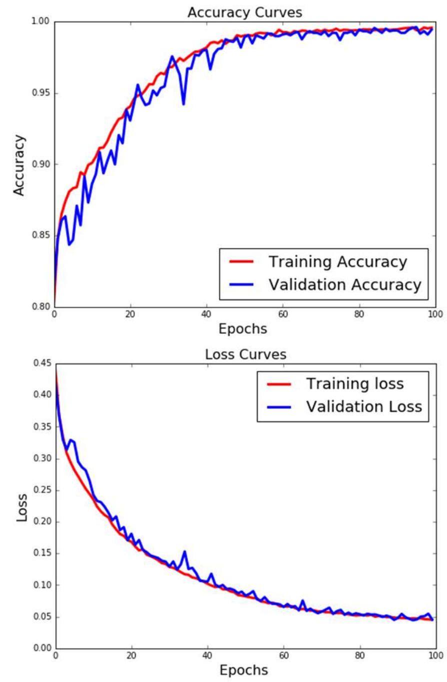


Figure 7. Accuracy and loss curves against the number of epochs

(5) Seizure Prediction Measures

Because of the strictness nature of this prediction problem and the large time span of the interictal period data, there must be a way to handle the possibility of excluding outlying false positives. So, this work will consider that one false positive alarm will be counted only in case there exists at least 2 positive predictions for each 75 interictal data predictions or 300 seconds (A prediction is done every 4 seconds). That means 8 false positive predictions for every 1200 seconds or 20 minutes. This assumption is more restricted compared to the one in [15] where a false positive alarm rises if there are at least 8 prediction every 300 seconds. Different evaluation measures like sensitivity, specificity, are used to judge the system performance. The sensitivity of a prediction system is calculated as the number of seizures with a minimum of one true alarm within the specified prediction horizon preceding the seizure divided by the total number of seizures. The specificity of a prediction system is measured as the number of false predictions per hour.

III. RESULTS

The result of the proposed seizure prediction system is summarized in Table I.

TABLE I. PERFORMANCE MEASURES OF THE PROPOSED SYSTEM

Patient	No. of Seizures	Sensitivity	Specificity FPR/h
Patient01	5	100%	0
Patient03	6	83.3%	0
Patient04	2	50%	0.02

Patient05	5	80%	0
Patient09	4	100%	0.1
Patient10	7	100%	0
Patient14	5	100%	0.2
Patient18	5	100%	0.08
Patient19	3	100%	0
Patient20	5	100%	0.05
Patient21	4	100%	0.04
Patient23	5	100%	0
Total	**56**	**94.6%**	**0.04**

The table shows that using up to one hour of prediction window, our proposed system managed to successfully predict 53 out of the 56 studied seizures (a %94.6 sensitivity rate) with only an average false prediction rate of 0.04FP/h for the interictal data.

Comparing with previous work that have used various features extraction methods applied on the same dataset, authors in [20] used common spatial pattern (CSP) based feature extraction and linear discriminant analysis classifier, they achieved a relatively high false prediction rate (FPR) equals 0.39 while achieving a good sensitivity (89%). Authors in [15] used Short-Time Fourier Transform for features extraction and convolutional neural network (CNN) as a classifier and got a better FPR equals to 0.16 but a lower sensitivity (81.2%). Finally, authors in [15] used a Wavelet Transform for features extraction and CNN as a classifier and got a better FPR equals to 0.142 and a better sensitivity (87.8%). Based on the previous comparison, our system has achieved the highest sensitivity with minimal false positive rate compared to some of the state-of-the-art systems.

Table II summarizes the comparison between our proposed system and the previous work on the same dataset.

TABLE II. COMPARISON BETWEEN THE PROPOSED SYSTEM AND OTHERS USING THE SAME DATASET

Authors	Features Extraction	Classifier	FPR h^{-1}	Sen %
Alotaiby et al., 2017 [20]	Common spatial pattern statistics	LDA	0.39	89
Truong et al 2017 [15]	Short-Time Fourier Transform	CNN	0.16	81.2
Khan et al 2018 [14]	Wavelet Transform	CNN	0.142	87.8
This work	**Automatic (Convolutional Autoencoder)**	**Bi-LSTM**	**0.04**	**94.6**

IV. CONCLUSION

A novel semi-supervised deep learning system for seizure onset prediction based on EEG signals recordings is proposed. The system uses a deep 2D convolutional autoencoder, trained on unlabeled data, for features learning. A Bidirectional Long Short-Term recurrent neural network, augmented to the encoder part of the autoencoder, is used for classification. The proposed system is semi-supervised because it combines learning from unlabeled and labeled data in two separate stages. Transfer learning is applied for initializing the weights for the patients-specific networks for faster training and better convergence. A comparison between our approach and different prediction systems showed that our proposed system outperforms the previous work in terms of sensitivity and false prediction rate per hour despite working on raw EEG signals.

V. REFERENCES

[1] WHO media center, (2018, September)."Epilepsy". Fact Sheets [Online]. Available: http://www.who.int/en/news-room/fact-sheets/detail/epilepsy

[2] Niedermeyer E., da Silva F. H. L. (2005). Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Philadelphia, PA: Lippincott Williams & Wilkins.

[3] Elger CE, Schmidt D, "Modern management of epilepsy: A practical approach," Epilepsy & Behavior 12: 501–539, 2008.

[4] F. Mormann, T. Kreuz, C. Rieke, R. Andrzejak, A. Kraskov, P. David, C. Elger, and K. Lehnertz, "On the predictability of epileptic seizures, J. Clinical Neurophysiology," 116 (2006) 569-587.

[5] U. R. Acharya, S. V. Sreec, G. Swapna, R. J. Martisa, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," Knowl. Based Syst., vol. 45, pp. 147–165, Jun. 2013.

[6] F. Wendling, P. Benquet, F. Bartolomei, and V. Jirsa, "Computational models of epileptiform activity," Journal of neuroscience methods, vol. 260, pp. 233–251, 2016.

[7] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio, and F.Fuggetta, "Real-time epileptic seizure prediction using, AR models and support vector machines," IEEE Transactions on Biomedical Engineering 57 (2010) 1124-1132.

[8] F. Mormann, R. G. Andrzejak, C. E. Elger & K. Lehnertz, "Seizure prediction: the long and winding road," Brain 130, 314–333 (2007)

[9] S. Li et al., "Seizure prediction using spike rate of intracranial EEG," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 21, no. 6, pp. 880–886, 2013

[10] K. Gadhoumi, J.-M. Lina, and J. Gotman, "Discriminating preictal and interictal states in patients with temporal lobe epilepsy using wavelet analysis of intracerebral EEG," Clinical Neurophysiology, vol. 123, no. 10, pp. 1906–1916, Oct 2012

[11] L. D. Iasemidis, D.-S. Shiau, W. Chaovalitwongse et al., "Adaptive epileptic seizure prediction system," IEEE Transactions on Biomedical Engineering, vol. 50, no. 5, pp. 616–627, 2003

[12] D. Cho et al., "EEG-Based Prediction of Epileptic Seizures Using Phase Synchronization Elicited from Noise-Assisted Multivariate Empirical Mode Decomposition," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 8, pp. 1309–1318, 2017

[13] Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time-series. The Handbook of Brain Theory and Neural Networks, page 3361, MIT Press, 1995

[14] Khan H., Marcuse L., Fields M., Swann K., Yener B., "Focal onset seizure prediction using convolutional networks," IEEE Transactions on Biomedical Engineering, PP (99) (2017), p. 1

[15] Truong, N. D. et al. "Convolutional neural network for seizure prediction using intracranial and scalp electroencephalogram". Neural Netw. 105, 104–111 (2018).

[16] A. H. Shoeb, "Application of machine learning to epileptic seizure Onset detection and treatment," Ph.D. Thesis, Massachusetts Institute of Technology, Sep. 2009.

[17] I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning", MIT Press, 2016. http://www.deeplearningbook.org.

[18] S. Ioffe, C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", Proc. Int. Conf. Mach. Learn., 2015.

[19] Sepp Hochreiter , Jürgen Schmidhuber, Long Short-Term Memory, Neural Computation, v.9 n.8, p.1735-1780, November 15, 1997.

[20] T. N. Alotaiby, S. A. Alshebeili, F. M. Alotaibi, and S. R. Alrshoud, "Epileptic Seizure Prediction Using CSP and LDA for Scalp EEG Signals," Computational Intelligence and Neuroscience, vol. 2017, p. 11, 2017