Computers in Biology and Medicine 99 (2018) 24–37

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals

Kostas M. Tsiouris [a,b], Vasileios C. Pezoulas [b], Michalis Zervakis [c], Spiros Konitsiotis [d], Dimitrios D. Koutsouris [a], Dimitrios I. Fotiadis [b,e,*]

[a] Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, GR15773, Athens, Greece
[b] Unit of Medical Technology and Intelligent Information Systems, Dept. of Material Science and Engineering, University of Ioannina, GR45110, Ioannina, Greece
[c] Digital Image and Signal Processing Laboratory, School of Electrical and Computer Engineering Technical University of Crete, Chania, Greece
[d] Dept. of Neurology, Medical School, University of Ioannina, GR45110, Ioannina, Greece
[e] Dept. of Biomedical Research, Institute of Molecular Biology and Biotechnology, FORTH, GR45110, Ioannina, Greece

ARTICLE INFO

Keywords:
EEG
Epilepsy
Seizure prediction
LSTM model
Deep learning

ABSTRACT

The electroencephalogram (EEG) is the most prominent means to study epilepsy and capture changes in electrical brain activity that could declare an imminent seizure. In this work, Long Short-Term Memory (LSTM) networks are introduced in epileptic seizure prediction using EEG signals, expanding the use of deep learning algorithms with convolutional neural networks (CNN). A pre-analysis is initially performed to find the optimal architecture of the LSTM network by testing several modules and layers of memory units. Based on these results, a two-layer LSTM network is selected to evaluate seizure prediction performance using four different lengths of preictal windows, ranging from 15 min to 2 h. The LSTM model exploits a wide range of features extracted prior to classification, including time and frequency domain features, between EEG channels cross-correlation and graph theoretic features. The evaluation is performed using long-term EEG recordings from the open CHB-MIT Scalp EEG database, suggest that the proposed methodology is able to predict all 185 seizures, providing high rates of seizure prediction sensitivity and low false prediction rates (FPR) of 0.11–0.02 false alarms per hour, depending on the duration of the preictal window. The proposed LSTM-based methodology delivers a significant increase in seizure prediction performance compared to both traditional machine learning techniques and convolutional neural networks that have been previously evaluated in the literature.

1. Introduction

The electrical activity of the brain of patients suffering from epilepsy has been considered to change significantly over time while moving towards seizures onset, with various methodologies being proposed to isolate preictal EEG patterns and rhythmic changes that are able to declare an incoming seizure [1]. As our understanding in preictal EEG activity grows, the progress in epileptic seizure prediction could revolutionize patient treatment and disease management. However, currently, the study of preictal activity has yet to provide conclusive evidence on critical aspects such as actual duration, morphology of preictal EEG patterns and relative variations in brain states. Accounting for the differences in preictal EEG semiology and time-variations of preictal patterns from patient to patient, the seizure prediction problem increases in difficulty and complexity, justifying the lack of robust methodologies despite the increased attention in the field over the past few decades.

In the early stages, researchers focused on the analysis of intracranial EEG recordings (or ECoG) since they offer better signal-to-noise ratio, better localization over the epileptogenic area and are less susceptible to noise and artifacts. The aim was to eventually include the signal processing unit in closed-loop systems that could dynamically adjust the electrical activity to prevent or control an incoming seizure [2,3]. On the other hand, scalp EEG recordings have significant disadvantages in terms of EEG signal quality, presence of artifacts (e.g. body motion and electrodes movements, normal muscle activity, electromagnetic interference, etc.) and lower spatial discretization of the brain's electrical activity, which are outweighed by the fact that it is a non-invasive technique.

* Corresponding author. Unit of Medical Technology and Intelligent Information Systems, Dept. of Materials Science and Engineering, University Campus of Ioannina, GR 45110, Ioannina, Greece.
E-mail address: fotiadis@cc.uoi.gr (D.I. Fotiadis).

https://doi.org/10.1016/j.compbiomed.2018.05.019
Received 7 March 2018; Received in revised form 7 May 2018; Accepted 16 May 2018

Thus, for everyday patient monitoring and seizure alert generation, scalp EEG recordings have higher potential in terms of applicability and ease of use compared to intracranial electrodes.

In its core, a seizure prediction methodology relies in two distinctive approaches. In the first approach, a preictal duration is pre-selected to categorize EEG signals in preictal and interictal states and a binary classifier is trained to exploit differences between the two states. The ictal segments cannot contribute to seizure prediction and are discarded from the analysis; including postictal segments, if they are considered as a separate state. The classification can be performed directly on the raw EEG signals or after feature extraction. The second approach consists of threshold-based methodologies where the analysis is focusing on identifying increasing/decreasing trends in the values over some feature(s) during the preictal state. When the value of the examined feature exceeds the activation threshold an alarm is raised to declare an incoming seizure [4]. Predictive features include measures that show a sustained increase or decrease of EEG channel synchronization and phase locking during the preictal state and towards a seizure [5–7]. A disadvantage of such methodologies is that they are usually limited into a single feature to declare seizures, which can be a limiting factor in seizure prediction performance considering the complexity of preictal activity.

Machine learning has revolutionized the seizure prediction field offering tools to address the high complexity of EEG signals and allow multivariate analysis and feature spaces of higher order to be easily evaluated to distinguish hidden preictal characteristics. Up until the past few years, traditional machine learning techniques (i.e. non-deep learning algorithms) have been the only viable option in EEG analysis and in fact continue to be extensively used combined with various feature extraction and feature selection algorithms [8–11]. In a relative newer trend, deep learning algorithms have found applications in medical image and signal processing, due to the advancements and availability of computational power and big data, showing high potential and significant impact as, in most cases, their performance exceeds the rates that have been previously achieved with traditional machine learning techniques [12]. In seizure prediction, convolutional neural networks (CNNs) have predominantly attracted researchers interest, probably because CNNs have been extensively used in image processing [13] and, thus, are better established and more familiar in the research community.

In this study, we introduce for the first time Long Short-Term Memory (LSTM) deep learning models in the EEG analysis for epileptic seizure prediction research. The LSTM model was proposed by Hochreiter and Schmidhuber in 1997 [14] and is considered as an evolution over the Recurrent Neural Networks (RNN) that has been previously used in EEG analysis [15,16]. The innovative part of LSTM networks compared to traditional RNNs is the inclusion of "gates" to address the vanishing gradient problem and allow the algorithm to control more precisely what information needs to be kept in its memory and what must be removed [17,18]. By controlling the learning rate with the three gates (i.e. input gate, forget gate and output gate), the LSTM network can better adjust to large sequences of data series, compared to RNNs and other deep learning techniques. Considering that EEG signals are essentially highly dynamic, non-linear time series data, LSTM networks have by design an advantage over CNN in isolating temporal characteristics of brain activity during different states as reported in various applications, such as emotion recognition, confusion estimation and decision making prediction [19–22]. Despite their inherent advantage in EEG analysis, LSTM models have not gained the appropriate attention in seizure prediction.

The lack of open EEG data, is another limiting factor since the large volumes of continuous EEG recordings required in deep learning algorithms are scarce. The proposed LSTM-based methodology is evaluated using the public CHB-MIT Scalp EEG Database, which is the only open database available that contains continuous long-term scalp EEG recordings to allow for seizure prediction evaluation. Extending the results of previous studies in the literature, which only partially used the database, its entire contents are used in the evaluation of our methodology. In addition, the proposed LSTM network is compared with traditional machine learning algorithms, namely SVM, Decision Trees and the Repeated Incremental Pruning to Produce Error Reduction (RIPPER) algorithm.

The rest of this study is organized as follows. Section 3.1 describes the EEG dataset used to evaluate the proposed methodology, which is presented in detail in section 3.2. A pre-analysis is performed in section 3.4 to find the most efficient LSTM network architecture, using EEG features extracted as described in section 3.3. The evaluation results of the LSTM network are shown in section 4, including four different lengths of preictal window and various sizes of LSTM input data. The comparison with traditional machine learning algorithms is shown in section 4.4. The key aspects of this work are discussed in section 5 including a comparison with the literature.

2. Literature review

A moving window analysis is usually performed to split the raw EEG data into segments of smaller duration that are then used for feature extraction [23]. Feature extraction offers dimensionality reduction and more complex, higher order feature spaces which can increase the discriminative power of the classification algorithm used to isolate preictal EEG segments.

2.1. EEG features

Spectral power estimation of each EEG channel is one of the most popular univariate features. The most popular algorithms to transform the EEG signals from the time domain to the frequency domain are the Fourier Transform [24,25] and the Wavelet transform [23,26]. Both are usually used in their discrete form (i.e. Fast Fourier Transform, FFT, and Discrete Wavelet Transform, DWT) to lower the computational cost required to analyze large volumes of EEG data, although some seizure prediction studies have also used continuous wavelet transform in EEG spectrum extraction [27,28]. Besides estimating the total EEG signal power per channel, the signal's energy is also estimated in the standardized EEG frequency bands or rhythms: delta ($\leq$3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz) and gamma (>30 Hz). Relative power differences in these frequency bands have shown great potential in identifying preictal variations in scalp and intracranial signals [29–31].

The predictive value of morphological EEG characteristics has also been investigated leading to some promising results. The rate at which spikes emerge in the EEG signals of epileptic patients in particular has shown to change significantly [32,33]. Spikes consist of a pointed peak that is clearly distinguishable from the background EEG activity with a duration ranging between 20 and 70 ms. In Ref. [33] intracranial EEG signals from 21 patients were analyzed for spike rate estimation and the results suggested that the mean spike rate varied significantly from the interictal to preictal, ictal and postictal states with an average value of 1.7, 4.98, 18.97, and 2.36, respectively. Besides spikes, zero crossings of EEG signals have also been studied for seizure prediction [34,35]. Both negative and positive zero crossings were evaluated as features providing similar levels of seizure prediction performance. On average, a sensitivity of 88.34% with 0.155 false predictions/hour was achieved using scalp EEG signals from 20 patients [34], while 86.6% and 0.067 F P/h, respectively, was reported using intracranial signals form 6 patients [35].

The most common bivariate feature is cross-correlation that estimates the dependence between a pair of EEG channels, accounting for time delays by shifting one of the two signals. The maximal cross-correlation value is retained as the most informative feature. Secondary features can also be extracted from cross-correlation values such as spatiotemporal eigenvalues extracted from between channels correlation and covariance matrices [36]. These features resulted in a mean sensitivity of 86% with 0.03 false prediction/hour using intracranial EEG signals from 19 patients. Between signals phase synchronization measures have also been proposed for EEG analysis, including phase coherence or phase locking value [37]. Another major field in EEG analysis includes graph theoretic

features. The graph theory framework has been introduced to study functional and anatomical connections in the brain using connectivity graphs [38]. The graphs can be extracted using cross-correlation, phase lag index or lagged linear coherence as connectivity measures across the EEG channels [39,40]. Once constructed, the graphs are used to extract secondary features including, among the most commonly used, degree, diameter, characteristic path length, betweenness centrality and eccentricity [41–43].

2.2. Preictal-interictal classification

Traditional classification algorithms have been extensively used in seizure prediction studies to discriminate preictal EEG segments [8,9,11]. Cho et al. proposed a methodology that combined multivariate empirical mode decomposition to calculate phase locking values between scalp EEG signals and classify short EEG segments as interictal or preictal using SVM classifier [44]. Their methodology was evaluated using 65 seizures from the CHB-MIT EEG database and reached 82.44% sensitivity and 82.76% specificity, respectively. Zhang and Parhi used spectral analysis of scalp EEG signals to extract an initial set of features that was then used to isolate the most relevant electrodes and features that were fed into an SVM classifier to distinguish preictal from interictal EEG segments [30]. The EEG dataset used for evaluation consisted of 78 seizures from 17 cases from the same database and the proposed methodology reached an average seizure prediction sensitivity of 98.68% with a false prediction rate of 0.05 false positives per hour. The k-Nearest Neighbors (k-NN) classifier has also been evaluated for seizure prediction using an EEG dataset consisting of multiday recordings from 10 patients, but provided a slightly lower seizure prediction performance with 73% sensitivity and 67% specificity [45]. In our preliminary work [46] a Neural Network (NN), an SVM and a C4.5-based Decision Tree classifier were compared for their preictal classification performance using all 24 cases of the CHB-MIT database. The SVM and the NN provided similar sensitivity measures at 86.75% and 86%, respectively, with the SVM showing better specificity at 84.75%. The Decision Tree classifier was found to underperform with the sensitivity and specificity values being about 3–4% lower.

Regarding deep learning algorithms, Convolutional Neural Networks (CNN) have attracted the most interest in seizure prediction. A CNN architecture consisting of 3 blocks, each one including a normalization, a convolution and a max pooling layer was used in Ref. [47]. The spectral information form the raw EEG signals was extracted in 30-s windows and the spectrum was then fed as input to the CNN. Testing with 64 seizures from 13 cases of the CHB-MIT database, the average seizure prediction sensitivity reached 81.2% with a false prediction rate of 0.16 false predictions/hour. Khan et al. also used a CNN architecture with six convolutional to extract features that were used to assign EEG segments into preictal, interictal and ictal classes [48]. As input, the CNN received the output of the wavelet transform of every EEG channel calculated at various scales. The same database was used for evaluation with the authors using 18 seizure and 50 interictal recordings from 15 cases, which resulted in an average false prediction rate of 0.142 false predictions/hour, while three seizures were not predicted. Bivariate features that measure EEG synchronization in intracranial recordings from the Freiburg database were also used to evaluate the seizure prediction performance of 5-layer CNN and showed that it was able to predict all testing seizures with no false alarms in 15 of 21 patients [49]. A deep multi-layer CNN have also been used in the evaluation of motor imagery tasks from EEG signals using spatial filters (linear combinations of EEG channels) that enhance class-discriminative band power features contained in the EEG signals [50].

In contrast, LSTM networks have not yet been used in seizure prediction from EEG signals. However, LSTM networks have recently found applications in studies focusing on other areas of EEG analysis. An LSTM model was used for EEG-based behavioral microsleep and lapse detection system [51,52] using the spectral energy across the fundamental frequency bands as input to the LSTM network. LSTM networks have been also used in automated epileptic seizure detection from scalp EEG recordings. A recurrent convolutional neural network was designed in Ref. [53], consisting of a 4-layer CNN that inputs the spectral information of each EEG channel and an LSTM network with a single layer to perform the final classification of ictal EEG segments. A variety of classification algorithms were also evaluated in Ref. [54], combining LSTM, CNN, hidden Markov Models (HMM), stacked autoencoders and multi-layer perceptron (MLP) in various configurations. Using the TUH EEG dataset, seizure detection sensitivity for all configurations was found relative low ranging from 30 to 39% which to some extend is expected for a patient-independent approach, with the CNN-LSTM model providing the highest specificity at 96.86% with 7 false alarms/hour. A mixed model combining MLP and LSTM networks was also evaluated in automated sleep stage classification outperforming SVM, Random Forest and single MLP classifiers [55]. A two-layer LSTM network also managed to increase the classification accuracy workload classification using the four statistical moments as features to assess cognitive states from EEG signals, compared to SVM and feedforward Neural Network models from 83% to 93.0% with the deep LSTM model [56].

3. Materials and methods

3.1. EEG data

The EEG data used in this study are obtained from the CHB-MIT scalp EEG database [57] which is available with open access at PhysioNet.org [58]. The EEG signals were collected from 23 pediatric subjects with intractable seizures (17 females, ages 1.5–19; 5 males, ages 3–22; 1 missing gender/age data) at the Children's Hospital Boston in order to evaluate their potential for surgical intervention. The recordings are organized in 24 cases and were obtained with a sampling frequency of 256 samples per second and 16-bit resolution from electrodes placed according to the International 10–20 electrodes position system. In most of the 24 cases, the EEG signals are segmented in 1-h long epochs, while epochs of 2–4 h in duration can also be found. The total duration of the available EEG recordings is about 983 h. For the EEG epochs that contain ictal activity, the seizure onset and offset time intervals were manually annotated by clinical experts after visual inspection, leading to a total of 198 identified seizures. More details about the EEG dataset can be found in Table 1.

As the annotation files of the database suggest, in most of the 24 cases there are frequent changes in the EEG signal recording montage, with channels being added or removed from one epoch to another during the recording process. Since all EEG epochs are interchangeably used for training and testing through cross-validation, only the channels that are constantly available throughout the entire duration in each case are selected for analysis, to mitigate the effect of data heterogeneity. In total, 18 channels are consistent across all 24 cases including: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ and CZ-PZ. From the entire EEG volume of the database, only three epochs from Case 12 are disregarded (i.e. files "*chb12_27–29.edf*"), as the channel montage within these epochs does not contain any of the above channels. Thus, without the removed epochs, the final dataset for the evaluation of the proposed methodology consists of about 980 h of EEG signals and 185 seizures from 24 cases.

3.2. Seizure prediction methodology

An outline of the proposed methodology is presented in Fig. 1. Besides, isolating the 18 bipolar EEG channels that are available in the files of each case of the dataset, there is no further pre-processing applied such as noise/artifact rejection techniques. Thus, the EEG signals are retained to their initial raw format. For each case of the dataset the raw EEG signals are, initially, split into 5-s long segments, with no overlapping between adjacent segments. Each segment is assigned to ictal, preictal or

Table 1
Description of the CHB-MIT EEG dataset. Gender: Female (F), Male (M). Duration of Recordings: Total duration of EEG recordings. # of EEG channels: Number of channels included.

Case	Gender	Age (years)	# of EEG channels	# of seizures	Duration of Recordings (hh:mm:ss)
1	F	11	18	7	40:33:08
2	M	11	18	3	35:15:59
3	F	14	18	7	38:00:06
4	M	22	18	4	156:03:54
5	F	7	18	5	39:00:10
6	F	1.5	18	10	66:44:06
7	F	14.5	18	3	67:03:08
8	M	3.5	18	5	20:00:23
9	F	10	18	4	67:52:18
10	M	3	18	7	50:01:24
11	F	12	18	3	34:47:37
12[a]	F	2	18	27	20:41:40
13	F	3	18	12	33:00:00
14	F	9	18	8	26:00:00
15	M	16	18	20	40:00:36
16	F	7	18	10	19:00:00
17	F	12	18	3	21:00:24
18	F	18	18	6	35:38:05
19	F	19	18	3	29:55:46
20	F	6	18	8	27:36:06
21	F	13	18	4	32:49:49
22	F	9	18	3	31:00:11
23	F	6	18	7	26:33:30
24	–	–	18	16	21:17:47
Total:				**185**	**979:56:07**

[a] Files "*chb12_27–29.edf*" are not used due to montage inconsistency with the rest of the recordings.

interictal class, using the annotation files included in the CHB-MIT database. The ictal segments that contain the actual seizure activity are not required for seizure prediction and, thus, they are discarded, leading to a binary classification problem between preictal and interictal classes. The EEG segments are then passed through the feature extraction module, where a wide variety of the most commonly used EEG features is extracted to jointly produce a 643×1 feature vector from each segment. The feature extraction module includes units for time and frequency EEG analysis, estimation of between-signals cross-correlation and various measures from graph theory to provide the most informative feature space possibly.

The final module shown in Fig. 1 consists of the LSTM network. The features that have been previously extracted are standardized and then provided as input to the LSTM network. The learning potential of the LSTM model is used to assess the significance of each feature in the final decision making, as there is no separate feature selection step prior to classification. The feature space is internally evaluated for each patient by the LSTM model that learns the most informative features and adopts to patient-specific particularities in order to produce a more effective, personalized seizure predictor. In addition, instead of classifying each EEG segment individually, the LSTM network receives a larger input of multiple EEG segments and assigns the entire sequence of the EEG segments into either the preictal or interictal class. The number of the EEG segments provided as input to the LSTM network is adjusted for each case from 5 to 50 segments during training. By using a larger number of EEG segments as input, the LSTM network can extract timely distant preictal variations in the EEG signals. To avoid overfitting, the EEG segments inside the two classes that contain the preictal and interictal segments from the entire duration of EEG recordings available in each case, respectively, are shuffled in order to randomly select the sequences of 5–50 segments that are fitted as input to the LSTM network. Furthermore, the seizure prediction performance is evaluated using stratified 10-fold cross-validation. By shuffling the EEG segments that are used as input, the LSTM network is forced to learn more generic preictal patterns as each sequence consists of random, non-adjacent preictal segments that not only come from various locations with different time distances from the actual seizure onset, but also from the preictal activity of different seizures.

As in almost every application of seizure prediction from long-term EEG recordings, there is the issue of high data imbalance between preictal and interictal classes, as seizure frequency during an EEG recording session can be low (e.g. Case 4, 156 h of EEG with 4 seizures, Table 1). To overcome class imbalance issues but still use the entire volume of the available EEG data for a more complete evaluation, the interictal EEG signals are split into smaller subgroups that match the duration of the preictal class. Should the division of the interictal class into subgroups results in a remainder, EEG segments are randomly selected from the completed interictal subgroups to fill the proportion of missing data that are required to use the remaining EEG segments (i.e. oversampling for the last interictal subgroup). The total seizure prediction performance is estimated using the mean value of the classification results between the preictal class and every subgroup of the interictal class.

3.3. Feature extraction

Each case of the EEG dataset is analyzed individually as the proposed LSTM network is trained to predict seizures in a patient-specific approach. A variety of the most common features used in EEG signal analysis [59] is extracted from every 5-sec long EEG segment (i.e. 1280 EEG data samples) in order to be used as input to the LSTM classifier. The extracted features are presented below and are summarized in Table 2.

3.3.1. Time domain features

As far as the time domain features are concerned, the four statistical moments are computed (i.e., mean value, variance, skewness and

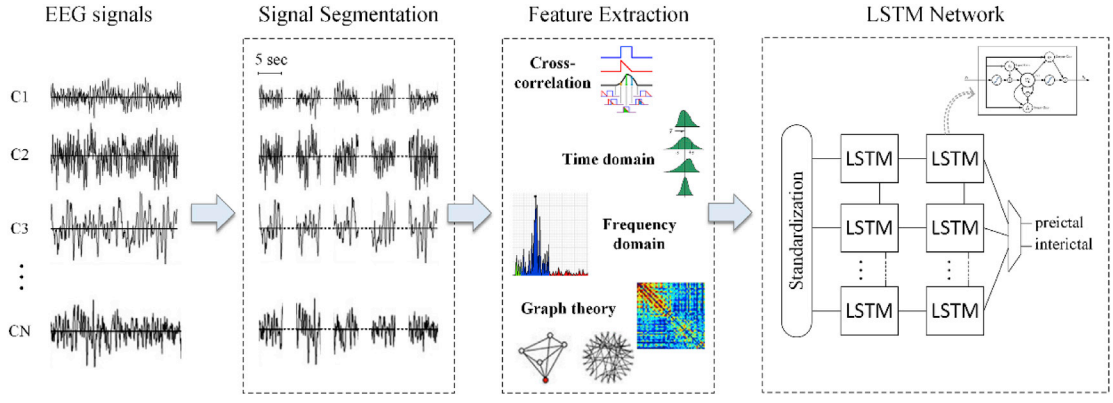

Fig. 1. Schematic representation of the proposed seizure prediction methodology. Features are extracted from 5-s long EEG segments and are fed into the LSTM network for classification.

Table 2
The complete list of the features extracted from each 5-s long EEG segment.

Time domain	Statistical moments	Mean value, variance, skewness, kurtosis
	Standard deviation	Square root of variance
	Zero crossings	Number of sign changes
	Peak-to-peak voltage	Difference between highest/lowest amplitude
	Total signal area	Absolute area under signal
Frequency domain	Fast Fourier Transform – Power Spectral Density	Total EEG signal energy Energy percentage at: • delta band (≤3 Hz) • theta band (4–7 Hz) • alpha band (8–13 Hz. • beta band (14–30 Hz) • gamma1 band (30–55 Hz) • gamma2 band (65–110 Hz)
	Discrete Wavelet Transform	7-level decomposition coefficients
Correlation	Cross-correlation	Max absolute coefficient
	Autocorrelation	Decorrelation time
Graph theory	Local measures	Clustering coefficient, local efficiency, betweenness centrality, eccentricity
	Global measures	Global efficiency, diameter, radius, characteristic path

kurtosis) [56], as well as the total signal area, peak-to-peak value (i.e. minimum to maximum), number of zero crossings and decorrelation time (i.e. time of first zero-crossing of the autocorrelation function). Despite the simplicity of these measures, significant variations have been previously reported when entering the preictal state, with variance and decorrelation time decreasing compared to their respective interictal values, while kurtosis was found to increase towards seizure onset [60].

3.3.2. Frequency domain features

In addition, the spectral information of the EEG signals is also taken into consideration, since various frequency domain features are also extracted including the total energy spectrum and the energy percentage across the fundamental rhythmic bands (i.e., delta: 1–3 Hz, theta: 4–7 Hz, alpha: 8–13 Hz, beta: 14–30 Hz, gamma1: 31–55 Hz and gamma2: 65–110 Hz).The EEG signal energy in each frequency band is extracted using the Discrete Fourier Transform (DFT). Along with DFT, the Discrete Wavelet Transform is also applied using a 7-level decomposition and the Daubechies 4 (db4) as the mother wavelet, to extract the detail (64–128 Hz, 32–64 Hz, 16–32 Hz, 8–16 Hz, 4–8 Hz, 2–4 Hz, 1–2 Hz) and approximation coefficients (<1 Hz). The 7-level decomposition is selected based on the 256 Hz sampling frequency used when assembling the CHB-MIT database, as it is the minimum depth required to cover the fundamental frequency bands and allow for the separation of bellow the 1 Hz that is most predominantly occupied by artifacts.

3.3.3. Brain connectivity

The maximal absolute cross correlation value is employed as a functional connectivity measure and was further calculated for each pair of EEG channels to quantify the similarity between any two EEG signals. Its advantage lies on the fact that it takes into consideration the potential time delays which have been introduced between two spatially distant signals. The maximal cross correlation C_{xy} between two signals $x(t)$, $y(t)$. is given as:

$$C_{xy} = \max_{\tau}\left(\left|\frac{C_{xy(\tau)}}{\sqrt{C_x(0)\cdot C_y(0)}}\right|\right), \quad (1)$$

where τ are the time delays, defined as $\tau \in [-5, 5]$.

3.3.4. Graph theoretic measures

In this study, graph theory is employed as a conventional network analysis approach for describing the functional connectivity between the EEG channels using weighted and undirected graphs based on a specified functional connectivity measure. The latter are defined as a set of nodes (i.e. channels) and weighted edges, where the weight w_{ij} connects the nodes i and j, respectively. Each graph is represented as an adjacency symmetric matrix $W \in R^{nxn}$, where n is the number of nodes with the cell (i,j) being equal to w_{ij}. The graph is constructed using the lag indexed cross-correlation values as connectivity measure since these have been already computed in 3.3.3 and using more computational resources to estimate alterative measures (e.g. phase lag index, lagged linear coherence) and build the graph from scratch would not be ideal. Thus, having essentially a pre-estimated graph a set of global and local features can be extracted as described below.

3.3.4.1. Local measures. The extracted local measures are computed and described for a node i as follows: (a) clustering coefficient which measures the cliquesnness of the node i, i.e., the density of the neighboring connections around that particular node [61], (b) local efficiency which is another segregation measure that is defined as the average inverse shortest path length on node's i neighbourhoods, (c) betweenness centrality which is defined as the number of shortest paths between any two nodes k and l that pass through i divided by the total number of shortest paths between the two nodes [61] and (d) eccentricity which is the longest of the paths from node i to any other node in the network.

3.3.4.2. Global measures. The extracted global graph measures include the characteristic path, the global efficiency, diameter and radius of the graph. The characteristic path length is defined as the average shortest path length in the network, where short path lengths indicate better functional integration [61]. Global efficiency measures the overall efficiency of the network. The diameter is the maximum eccentricity (longest shortest path) of the nodes and has an upper bound of $n - N_{leaf} + 1$ where N_{leaf} is the number of leaf nodes in the network (i.e., nodes without any connections). On the other hand, radius is the minimum eccentricity (small radius reflects more central network topology).

3.4. LSTM architecture evaluation

Since LSTM models have not been used for seizure prediction, there are no references in the literature regarding an optimal internal architecture. Thus, a pre-analysis is performed in this section testing the three different internal architectures shown in Fig. 2, moving from simple to more complex networks. In the LSTM_1 architecture, which consists the simplest approach, the network is composed of a single layer with 32 memory units. The number of memory units is increased to 128 in the second architecture design, LSTM_2, maintaining the single layer approach. Finally, in LSTM_3 the number of memory units is retained at 128 but another layer of equal dimension is included as well. All networks are followed by a fully connected layer with an output of 30 units using the "relu" activation function and a final dense layer that outputs the binary classification one hot encoded result (i.e. preictal or interictal), using the "softmax" activation. The batch size is set to 10 and the cross-entropy loss function is selected as the cost function, using the Adaptive Moment Estimation (Adam) optimizer (learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, decay = 0). The LSTM networks were built using Keras 2.0.9 [62] upon Tensorflow 1.4.0 backend [63] in Python 3.6.

The LSTM architectures are evaluated using EEG data from three randomly selected cases of the CHB-MIT EEG database; Case 2, 14 and 1. The entire duration of all the available EEG recordings per case is used to estimate the total classification accuracy in assigning the EEG segments into the preicral or interictal class. The duration of the preictal window is set at 15 min. The effect of the LSTM input size is also evaluated by testing 10 different numbers of EEG segments as input, ranging from an input sequence consisting of 5 EEG segments, with an LSTM input size of 643 × 5 features, and up to 643 × 50 features when a sequence of 50 EEG

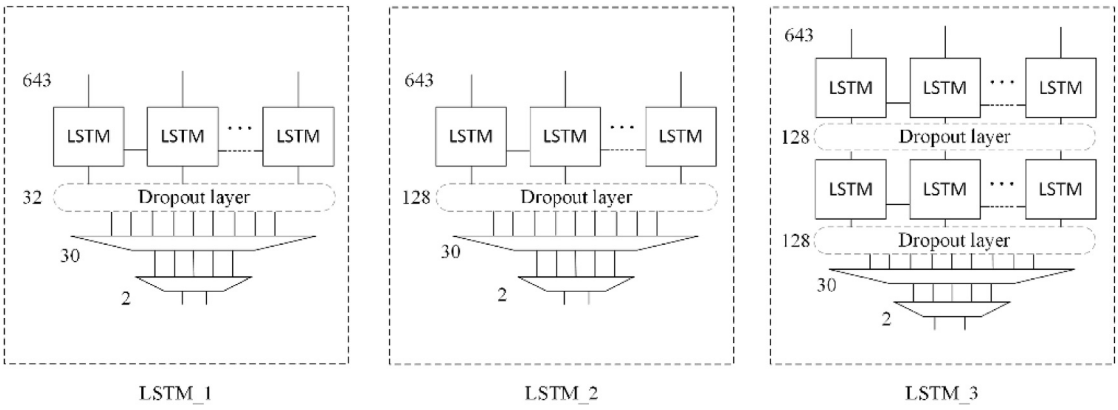

Fig. 2. The architecture of the three LSTM networks. The dropout layers are presented in dot lines as the evaluation is performed with and without including them.

segments is provided. The classification accuracy results of each network architecture per LSTM input size are shown in Fig. 3. The reported classification accuracy for the three experiments is the average across all the interictal subgroups, to avoid class imbalance, using 10-fold cross-validation for each subgroup, following the evaluation process described in section 3.2.

The graphs in Fig. 3 provide helpful insights on how the LSTM architecture affects the classification accuracy (except "CHB_01" where all variations have a ~1% margin). To begin with, the results confirm the expected advantage of using a more complex 2-layer network, as LSTM_3 provides the best classification accuracy, overall, and particularly in Cases 2 and 14. Regarding the size of LSTM input, as it can be seen in all three cases, increasing the number of EEG segments that are provided as input to the network is not necessarily beneficial if the network is not large enough to handle the extra volume of information, as LSTM_1 and LSTM_2 are found to be quickly saturated, resulting in reduced classification accuracy. In the same context, when a small number of EEG segments is used as input to the LSTM networks (i.e. less than 20 segments), increasing the network complexity in the form of LSTM_3, does not have a significant impact in discovering the hidden connections since there is a limited amount of EEG data considered. It is only when the number of input segments increases to around 20–30 EEG segments, that the more complex architecture of LSTM_3 network is able to show its full potential and provide higher classification accuracy. The effect of the LSTM input size will be further investigated in Section 4, as larger preictal windows are evaluated therein.

The inclusion of an intermediate dropout layer as shown in Fig. 2, has not significantly affected the classification performance (results not shown in Fig. 3 for clarity). The evaluation included dropout rates of both 0.2 and 0.5. In the case of LSTM_3, the classification accuracy is found to be mostly unaffected with a mean variation across the three cases 0.02% and 0.03% for 0.5 and 0.2 dropout rates, respectively, compared to the results without dropout layers shown in Fig. 3. For the LSTM_2 network architecture the dropout layers have a marginally higher impact with a mean of 0.6% and 0.3% for 0.5 and 0.2 dropout rates, respectively. The variation in accuracy for the smaller LSTM_1 network is also found at 0.06% for both dropout rates, suggesting that EEG segments shuffling prior to classification helped in mitigating overfitting making the intermediate dropout layers less important. Based on these results the in architecture of the LSTM_3 network will be finally used for our analysis in section 4 (from now on simply referred as LSTM network), the dropout layers are discarded to reduce the complexity of the network. Finally, performing cross-validation with smaller number of folds at 5-fold and 3-fold using the 3 cases of the pre-analysis, resulted in similar performance with the total classification accuracy values being, on average, within 0.46% and 0.09% of their respective values shown in Fig. 3 for 10-fold cross-validation.

3.4.1. *Feature learning compared to feature extraction*

According to the literature review above (Section 2), feature extraction prior to classification seems to be more preferable compared to directly using the raw EEG samples as input to the classifier. However, there are also some recent studies where feature extraction was not performed and the deep learning models had been trained with EEG signals instead [64–66]. Thus, in this section, the classification accuracy of the proposed LSTM network is also evaluated using as input the raw EEG signals, since an evaluation with both approaches has not been previously investigated. The most effective LSTM architecture from the analysis above is used (i.e. LSTM_3) using the same evaluation process (i.e. the same three cases, balanced classes and 10-fold cross-validation). Single segments are also evaluated for both raw EEG and feature extraction approaches with the number of LSTM input segments being 1–50. The results are presented in Fig. 4.

As it can be seen in Fig. 4, when using the raw EEG samples of each 5-s

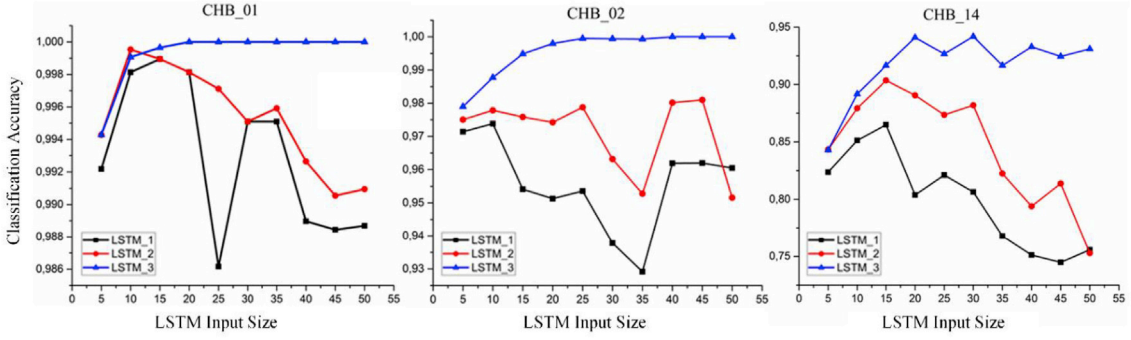

Fig. 3. Classification accuracy of the three LSTM networks using EEG data from Cases 1, 2 and 14 and 15 min of preictal window duration. LSTM Input Size: Number of EEG segments used as input.

segments instead of the extracted feature vector the LSTM network is massively underperforming being unable to extract more or equally meaningful features internally. The advantage of feature extraction is that it provides an easier and straightforward solution to obtain hidden information regarding the signal's frequency contents or brain areas connectivity from the available channels, compared to the information gain of using the EEG signals as time series. In theory, better EEG signal representation could be learned if the size of LSTM network was substantially increased, by adding more layers and memory units, to compensate for the increased input size of directly providing the EEG signals (i.e. 1280 samples x 18 EEG channels per 5-sec segment). However, the computational cost of training larger LSTM networks increases rapidly requiring more training time or using arrays of GPUs, that in our case are not available making the CPU-based training time required to reach multi-day duration. Even if computational cost was not a problem, this approach would require even more EEG data to effectively train the millions of network parameters. Thus, training the LSTM network with the extracted vector of features instead, allows our seizure prediction methodology to operate within more manageable conditions, by essentially simulating the implementation of a deeper and more complex LSTM model without the added time and data training restrictions.

4. Results

The proposed methodology is evaluated using the CHB-MIT database (Table 1). Four different preictal durations are tested, ranging from 15 min to 2 h. The impact of the size of the LSTM input in the prediction efficiency of an incoming seizure for each patient is examined, along with the evaluation of global input size for every case in section 4.2. Based on the pre-analysis results of section 3.4, the seizure prediction sensitivity is expected to be quite high, so that the main focus in section 4.1 is to minimize the false prediction rate for every case of the database to false alarms. As mentioned before, class imbalance is solved by splitting the EEG segments from the interictal class into smaller subgroups with size equal to the preictal class, while the reported evaluation results declaring the mean value from every subgroup using 10-fold cross-validation, respectively.

In order to enhance the scope of the evaluation, results are provided for both segment-based and event-based performance. To estimate seizure prediction performance, the true positive rate (TP) is defined by the number of segments that are correctly identified as preictal EEG segments, true negatives (TN) are defined by the number of EEG segments that are correctly classified as interictal, false positives (FP) are defined by the number of EEG segments that are incorrectly classified as preictal and false negatives (FN) as the segments that are incorrectly classified as interictal. For segment-based evaluation in Table 3, the sensitivity is calculated as TP/(TP + FN) and the specificity as TN/(TN + FP). For the event-based evaluation, sensitivity is defined as the number of successfully predicted seizures by detecting at least one preictal EEG segments and false prediction rate (FPR) declares the false alarm generation rate as the number of FPs per hour of EEG recordings.

4.1. Patient-specific LSTM input size evaluation

The evaluation is performed in the four preictal windows with a duration of 15, 30, 60 and 120 min, respectively, and the size of the EEG segments provided as input to the LSTM network is evaluated separately for each case of the dataset in order to minimize false predictions. The results in Table 3 show the LSTM input size required to reach the lowest FPR possible, per preictal window and for each one of the 24 cases. In the segment-based evaluation ("Segments" in Table 3), the EEG signals are classified with a mean sensitivity and specificity of 99.28% using the 15-min preictal window, while similar levels are also obtained using the 30-min preictal window, with the classification performance reaching 99.37% and 99.6% for sensitivity and specificity, respectively. In the event-based evaluation ("Events" in Table 3), every seizure across the 24 cases is successfully detected for both preictal window durations, providing 100% seizure prediction sensitivity, on average. The FPR is found to be significantly lower for the 30-min preictal window, reaching 0.06 F P/h compared to 0.11 F P/h for the 15-min preictal window, respectively. Furthermore, perfect seizure prediction rates with 100% sensitivity and 0 F P/h are achieved in 13 out of the 24 cases using the 15-min window and in 11 cases when using the 30-min preictal window despite its lower average FPR.

Increasing the preictal window to 60 min, the results in Table 3 suggest that the same levels of seizure prediction can be achieved with lower rates of false alarms per hour of EEG recordings. In the segment-based evaluation, segments are classified with an average 99.63% sensitivity and 99.78% specificity. A minor increase is found when using the 120-min preictal duration with the classification performance reaching 99.84% and 99.86% for sensitivity and specificity measures, respectively. Finally, in the event-based evaluation, all seizures are successfully detected for both preictal windows, providing an average seizure prediction sensitivity of 100%. By doubling the preictal duration to 1 h from 30 min, the FPR is reduced to half (i.e. 0.03 F P/h from 0.06 F P/h) while the number of cases with zero false predictions increased to 16 out of the total 24. Almost the same performance is obtained using the 2-h long preictal window with an average FPR of 0.02 F P/h and 17 cases with zero false alarms, suggesting that even remote EEG segments from the seizure onset can still contain valuable information.

4.2. Global LSTM input size evaluation

In this section a potential global number of EEG segments provided as input to the LSTM network is examined, by evaluating the effect of using a predefined number of LSTM input segments on prediction performance. The analysis of FPR demonstrates that false positives are strongly affected by the LSTM input size. As it can be seen in Fig. 5, LSTM input size is inversely proportional to FPR (i.e. higher input size, lower false predictions). The y axis in Fig. 5 shows the mean FPR value as estimated from the 24 cases of the database for each LSTM input size and preictal window. The minimum FPR is obtained using a sequence of 45 EEG

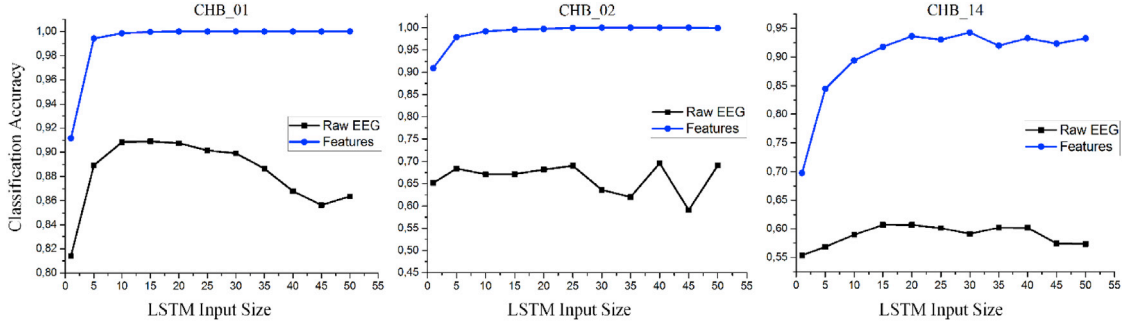

Fig. 4. Comparison of feature extraction (Features) and feature learning (Raw EEG) for Cases 1, 2 and 14. Preictal window duration: 15 min. LSTM Input Size: Number of EEG segments used as input.

Table 3
Seizure prediction results for the 24 cases of the CHB-MIT EEG Database. Preictal window duration: 15, 30, 60 and 120 min. Segments: Segment-based evaluation. Events: Event-based evaluation. SEN=Sensitivity, SPEC=Specificity, FPR=False Prediction Rate, LSTM input size: Number of EEG segments provided as input.

Case	# of seizures	EEG duration (hrs)	Preictal window: 15 min					Preictal window: 30 min					Preictal window: 60 min					Preictal window: 120 min				
				Segments		Events			Segments		Events			Segments		Events			Segments		Events	
			LSTM input size	SEN (%)	SPEC (%)	SEN (%)	FPR (h^{-1})	LSTM input size	SEN (%)	SPEC (%)	SEN (%)	FPR (h^{-1})	LSTM input size	SEN (%)	SPEC (%)	SEN (%)	FPR (h^{-1})	LSTM input size	SEN (%)	SPEC (%)	SEN (%)	FPR (h^{-1})
1	7	40.5	20	100	100	100	0	15	100	100	100	0	20	100	100	100	0	15	100	100	100	0
2	3	35	30	100	100	100	0	30	100	100	100	0	30	100	100	100	0	25	100	100	100	0
3	7	38	50	100	98.11	100	0.26	50	99.63	100	100	0	40	99.21	100	100	0	15	100	100	100	0
4	4	156	45	99.96	99.05	100	0.15	40	98.87	99.31	100	0.12	45	99.96	99.92	100	0.01	35	99.97	99.97	100	0
5	5	39	50	99.28	97.85	100	0.31	35	99.74	99.49	100	0.10	45	99.64	100	100	0	45	99.65	100	100	0
6	10	66.5	50	98.18	97.33	100	0.37	40	99.06	99.06	100	0.16	50	99.37	98.85	100	0.16	50	99.87	98.80	100	0.13
7	3	67	40	100	100	100	0	50	100	99.89	100	0.01	50	100	99.89	100	0.01	30	99.94	99.75	100	0.06
8	5	20	20	100	100	100	0	10	100	100	100	0	30	100	100	100	0	30	99.75	100	100	0
9	4	68	40	98.85	99.21	100	0.13	45	99.5	99.5	100	0.07	30	99.86	99.86	100	0.03	30	100	100	100	0
10	7	50	30	100	100	100	0	20	99.77	99.82	100	0.06	40	99.29	99.88	100	0.02	30	99.91	99.91	100	0.02
11	3	35	20	100	100	100	0	15	100	100	100	0	25	100	100	100	0	25	100	100	100	0
12[a]	27	21	45	100	100	100	0	30	99.14	100	100	0	20	100	100	100	0	35	99.71	100	100	0
13	12	33	45	100	100	100	0	30	100	99.87	100	0.03	35	100	100	100	0	25	99.89	100	100	0
14	8	26	40	90.97	95.6	100	0.73	45	94.54	96.17	100	0.54	45	95.88	98.08	100	0.27	40	98.72	99.36	100	0.08
15	20	40	50	98.24	98.83	100	0.15	45	98.21	98.75	100	0.17	40	99.03	98.55	100	0.22	45	99.50	100	100	0
16	10	19	35	99.42	99.71	100	0.05	45	98.79	99.6	100	0.05	40	99.68	100	100	0	40	99.68	99.68	100	0.05
17	3	21	25	100	100	100	0	10	99.93	100	100	0	15	100	100	100	0	15	99.89	100	100	0
18	6	35.5	50	99.19	99.19	100	0.11	35	99.17	99.86	100	0.03	35	99.57	99.71	100	0.06	40	99.83	99.49	100	0.08
19	3	30	25	100	100	100	0	25	100	100	100	0	35	100	100	100	0	35	100	100	100	0
20	8	27.5	15	100	100	100	0	10	100	100	100	0	10	100	100	100	0	5	100	100	100	0
21	4	33	50	100	99.78	100	0.03	35	99.85	100	100	0	30	99.74	100	100	0	35	100	100	100	0
22	3	31	40	100	100	100	0	40	100	100	100	0	30	99.86	100	100	0	30	100	100	100	0
23	7	26.5	25	99.58	100	100	0	40	100	99.78	100	0.04	15	100	100	100	0	10	100	100	100	0
24	16	21	45	99	98	100	0.28	30	98.74	99.24	100	0.14	40	100	100	100	0	25	99.78	99.78	100	0.05
Average:				**99.28**	**99.28**	**100**	**0.107**		**99.37**	**99.60**	**100**	**0.063**		**99.63**	**99.78**	**100**	**0.032**		**99.84**	**99.86**	**100**	**0.02**

[a] Files chb12_27–29.edf are not used due to montage inconsistency with the rest of the recordings.

segments as input to LSTM network, with 0.141 F P/h, 0.092 F P/h and 0.038 F P/h for the 15-min, 30-min and 120-min window, respectively. Only for the 60-min preictal duration the lowest FPR is obtained using 40 EEG segments instead, but the difference between the two is almost negligible at 0.061 F P/h for the 40 input segments compared to 0.062 F P/h for 45 segments. Thus, a sequence of 45 EEG segments can provide the lowest FPR if patient-specific adaptations for the LSTM input size are not performed during the training phase, in order to reduce the computation time required. Finally, only the 15-min preictal window failed to provide a FPR value smaller than 0.125 F P/h (i.e. 3 false alarms per day), which is considered as qualitative threshold of accepted performance in the seizure prediction literature. In contrast, any number of input EEG segments larger than 35 is able to provide a FPR below 0.125 F P/h for the other three preictal windows.

The equivalent results in terms of sensitivity and specificity for the segment-based analysis are presented in Fig. 6. Overall, the classification performance for every preictal window increases with increasing the number of EEG segments provided as input to the LSTM network, as before. For all four preictal windows the highest sensitivity reached similar levels with 99.42%, 99.55%, 99.68% and 99.91% for the 15, 30, 60 and 120-min preictal windows, respectively, within a range of 30–45 EEG segments as input to the LSTM network, depending on the preictal window. As it is shown in the right side of Fig. 6, the maximum values of specificity are also obtained in the same 30–45 LSTM input size range, reaching 99.1%, 99.4%, 99.63% and 99.7% for the 15, 30, 60 and 120 min window, respectively. Overall, the graphs in Fig. 5 suggest that sensitivity and specificity values reach a plateau above an LSTM input size of 30 segments, with diminishing returns in classification performance improvement. Considering that the variation in both sensitivity and specificity is within a 0.2% range of the maximum value in the 30–45 input size range, any selection of segments within this range is expected to provide nearly identical performance.

4.3. LSTM versus traditional machine learning algorithms

To demonstrate the advantages of introducing deep learning into seizure prediction a comparison with popular traditional classification algorithms is performed in this section, using the same set of features as in the LSTM-based methodology. The evaluation includes: (a) Decision Trees, DT, based on the C4.5 algorithm, (b) the Repeated Incremental Pruning to Produce Error Reduction (RIPPER) algorithm, which is a rule

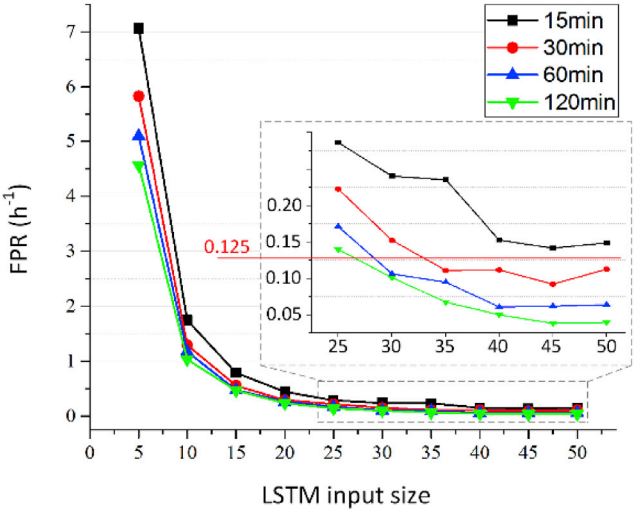

Fig. 5. Mean values of False Prediction Rate, FPR (FP/h) per LSTM input size. Average FPR across the 24 cases of the CHB-MIT database is shown per LSTM input size. Red line denotes the 0.125 F P/h value. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

inducting classification algorithm, and (c) Support Vector Machines, SVM, which is probably the most popular algorithm. Feature selection is performed prior to classification using filtering [67] to reduce the amount of features used in each fold of the 10-fold cross-validation. For all classifiers, the feature selection and classification algorithms are composed using WEKA's build-in functions [68] to assign the EEG segments into preictal or interictal class, using the same preictal windows as before.

For this comparison the LSTM network is tested using global LSTM input size as in section 4.2 and the number of input segments is set to 35. The results for the 15 and 30-min preictal windows are shown in Fig. 7. The LSTM network is able to outperform the traditional classification algorithms across all 24 cases. For the 15-min preictal window, the LSTM classifier produced an average sensitivity of 99.3% across all 24 cases compared to 88.68%, 89.36% and 89.37% that is extracted from the DT, RIPPER and SMV classifier, respectively. The margins in terms of specificity are similar, with the LSTM network reaching 98.79%, while DT reached 89.02%, RIPPER 89.9% and SVM scoring slightly higher at 91.55%, on average. Overall, the LSTM network is found to offer a 7.2–10.6% increase in sensitivity and specificity values compared to traditional classifiers. Setting the preictal duration to 30 min does not affect the performance of LSTM classifier, which produced an average sensitivity and specificity of 99.55% and 99.4%, respectively. A drop in classification performance is noted, however, for the rest of the algorithms with average sensitivity and specificity reducing to 87.26% and 87.15% for DT, 88.3% and 88.2% for RIPPER, 88% and 88.77% for SVM, respectively. The lower performance of the traditional classifiers led to a better sensitivity of 11.2–12.3% and specificity of 10.6%–12.3%, respectively, for the LSTM network.

The ability of the LSTM network to handle the chaotic nature of EEG data is further highlighted in Fig. 8 where the evaluation results with the 60 and 120-min preictal windows are presented. Setting the preictal duration to 60 min, the average sensitivity and specificity values of DT dropped to 85.92% and 85.48%, of RIPPER algorithm to 87.46% and 86.32% and of SVM to 86.93% and 85.87%, respectively. In contrast, the LSTM's performance remained at similar levels with an average sensitivity of 99.62% (i.e. 12.2–13.7% higher) and average specificity of 99.49% (i.e. 13.2–14% higher). Using the 120-min preictal duration window, the LSTM network pulls further ahead reporting 99.86% sensitivity and 99.59% specificity, on average across the 24 cases, compared to 84.54%, 85.83% and 83.51% sensitivity and 83.29%, 83.95% and 83.57% specificity for the DT, RIPPER and SVM, respectively, offering up to 16.35% higher performance.

Finally, in order to provide a point of reference for the classification performance achieved with the above classification algorithms, two basic and simplistic classifiers from Weka's build in functions were also evaluated, namely the Naïve Bayes and OneR,. Their respective classification performance was found on average 10.4–13.54% lower than the three traditional classification algorithms (with Naïve Bayes being about 4% better than OneR) decreasing as the duration of preictal widnow increased, and up to 26.7% behind the classification performance of the LSTM network.

5. Discussion

The evaluation results suggest that the proposed seizure prediction methodology is able to provide a solid performance being able to predict all 185 seizures (events) which are included in the annotation files of the CHB-MIT database, while maintaining a very low rate of false alarms per hour of EEG signals examined. If the duration of the training phase is not an issue, the results in Table 3 shows that the proposed LSTM architecture is able to provide an average FPR of 0.11–0.02 F P/h, with false alarms reducing as the preictal duration window increases from 15 min to 2 h prior to a seizure. Furthermore, the LSTM classifier is able to provide zero false alarms in 11–17 of the 24 cases of the database, depending on the duration of the preictal window. To provide this level

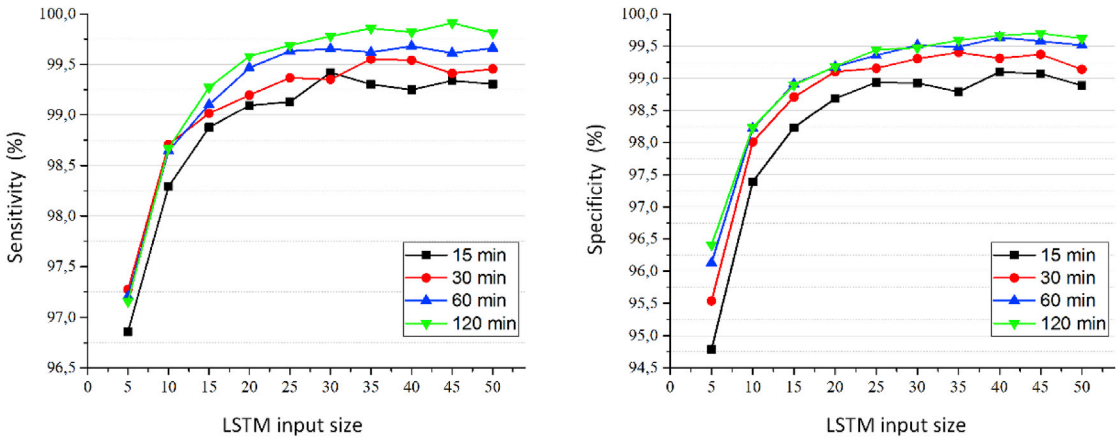

Fig. 6. Average sensitivity (left) and specificity (right) of the LSTM network using an input size of 5–50 EEG segments. Values represent mean sensitivity and specificity across the 24 cases of the CHB-MIT database.

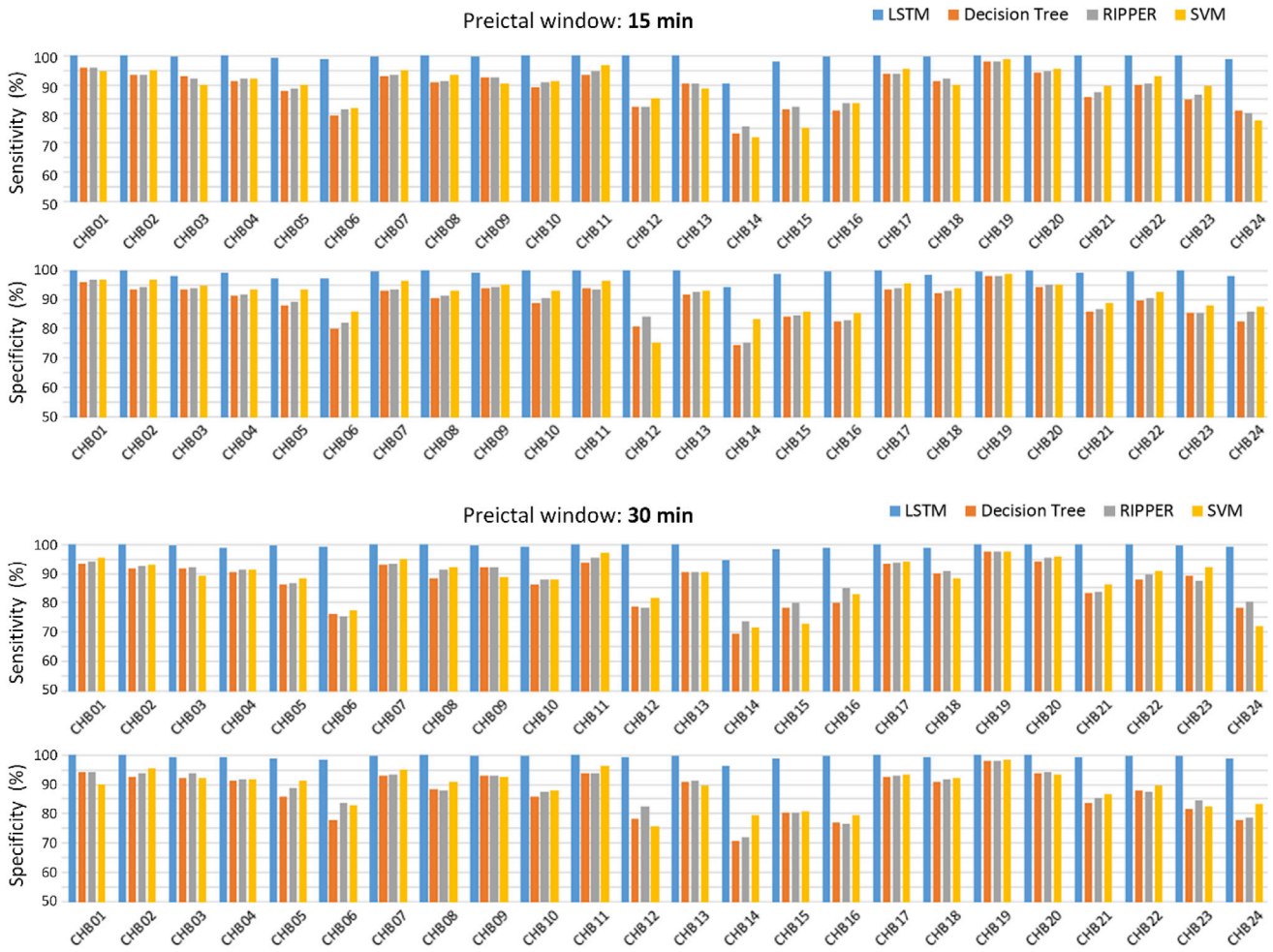

Fig. 7. Comparison of the proposed LSTM network's classification performance with Decision Trees, RIPPER and SVM in terms of sensitivity and specificity. Preictal window duration: 15 and 30 min.

of performance the proposed methodology requires on average an input sequence of 30–45 EEG segments or about 150–225 s of EEG activity (i.e. each segment's duration is 5 s). If patient-specific adaptations are not performed, the optimal number of EEG segments input is around 35–45 segments and, particularly, the seizure prediction performance is unaffected by the duration of the preictal window. Thus, based on the evaluation results, the proposed methodology is expected to predict all seizures with a false prediction rate of 0.038–0.141 with no further LSTM network architecture tuning or intervention required during the training phase for every new patient.

As became evident by the evaluation with the four different preictal windows, ranging from 15 min to 2 h, an advantage of the proposed methodology is that it can provide similar levels of seizure prediction performance even as preictal duration increases. In addition, considering

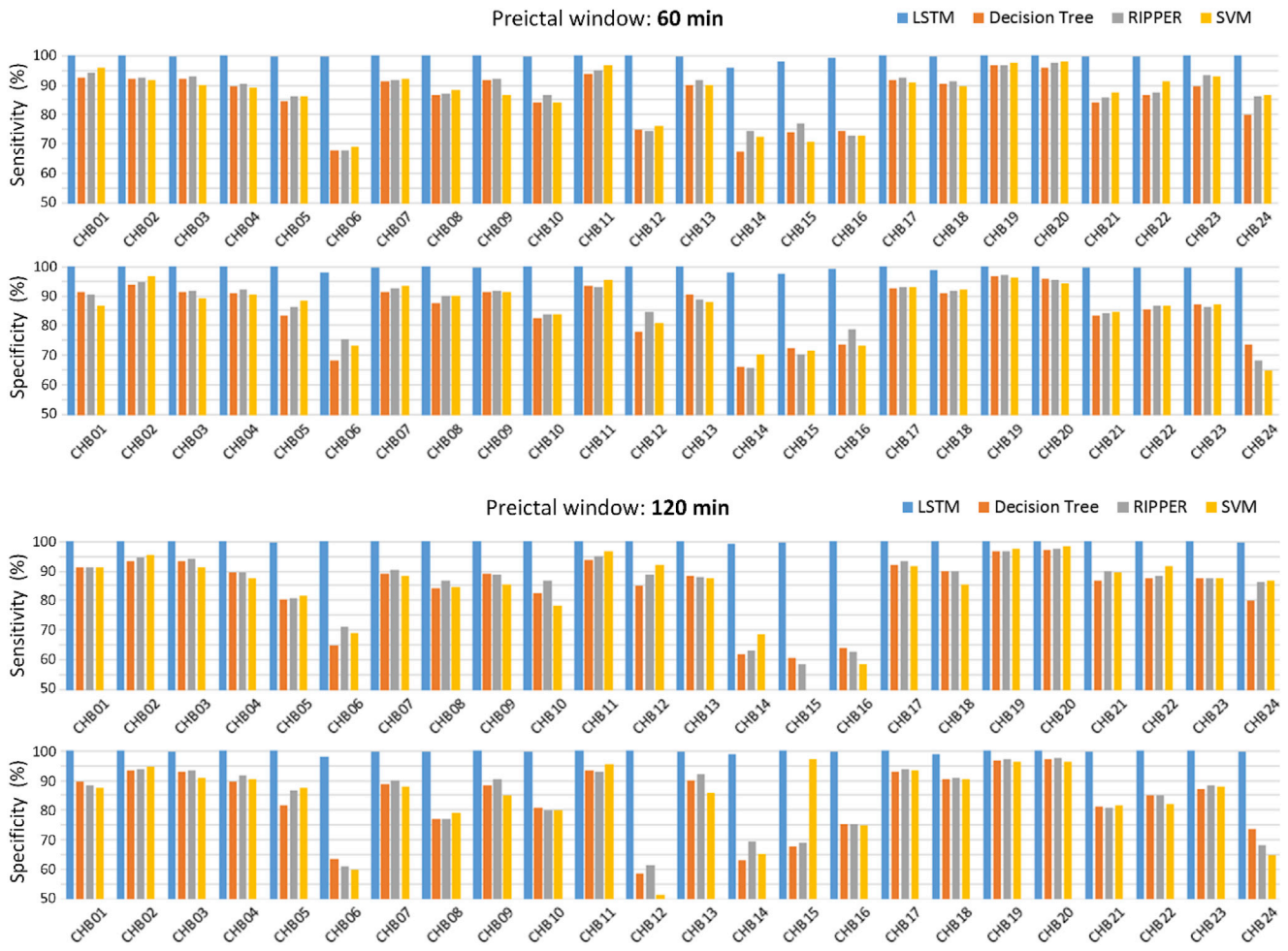

Fig. 8. Comparison of the proposed LSTM network's classification performance with Decision Trees, RIPPER and SVM in terms of sensitivity and specificity. Preictal window duration: 60 and 120 min.

that zero rates of false predictions have been achieved for up to 17 of the 24 cases used in the evaluation, there is a strong indication that the seizure prediction is very accurate with few false alarms. Furthermore, as preictal window duration increases, the LSTM network is found to provide lower rates of false predictions, which dropped from 0.107 F P/h in the 15 min window to 0.02 using the 120 min window (Table 3), accompanied by slightly higher rates of sensitivity and specificity values of up to +0.6%, respectively. In addition, similar levels of seizure prediction performance are reported regardless the number of available seizures in each case, which directly affects the number of EEG samples that can be used for training and testing. Based on these facts, the proposed methodology is found to be able to achieve high levels of seizure prediction under different conditions, increasing its robustness and capability of handling seizure prediction in real-life clinical environment.

Regarding the actual duration of the preictal state, the results from the four precital windows evaluated in this work indicate that EEG activity related to an imminent seizure might start hours before the actual seizure onset. This is evident by the increase in seizure prediction performance as preictal window grows, suggesting that using longer preictal windows leads to capturing the majority of the preictal-related variations in EEG activity that is exhibited prior to each seizure, leaving the interictal signals less contaminated with similar preictal-related activity. As a results the LSTM network was able to effectively classify preictal and interictal segments with fewer false predictions as the two classes become more distinguishable when a longer preictal duration is considered.

The impact of the proposed methodology, and that of deep learning algorithms in general, is further enhanced when compared to traditional machine learning techniques such as Decision Trees, rule inducers and SVM. The analysis with these three popular classification algorithms demonstrated that the LSTM network is able to provide a significant increase in classification performance, which the results showed to be from +7.2% and up to +16.4%. Besides pure classification performance gains, however, the biggest achievement of the LSTM classifier is its ability to sustain above 99% sensitivity and specificity that qualitatively translate to very low FPR while every seizure is also being successfully predicted. For example, a lower average of 97% would still be a sizeable difference from what could be achieved with traditional classification algorithms, but would not be enough to provide the desired false prediction rate, forcing a more strict approach that could eventually lead to seizures missing a preictal alarm and, thus, are not predicted.

A comparison of the proposed methodology with deep learning and other classification algorithms from the literature is presented in Table 4. The comparison focuses on studies that had been also evaluated with the CHB-MIT EEG database, since currently is the only public database consisting of long-term recordings, allowing for easier external validation. The proposed LSTM classifier is able to provide better seizure prediction performance than any of the previous methodologies that have been proposed and evaluated with the same EEG dataset so far, using similar windows of preictal duration. The features used in each study are also presented in Table 4. With the exception of graph theoretic features, the rest of the extracted features have been already successfully used for epileptic seizure prediction in previous studies and, according to our results, their combination to create a more informative feature space offers a significant advantage in classification performance. In fact this is

Table 4
Comparison with the literature, including studies that used the open CHB-MIT EEG database for evaluation as in this work. SEN=Sensitivity, SPEC=Specificity, FPR=False Prediction Rate.

Study	EEG dataset	# of cases	# of seizures	EEG duration (hrs)	Features	Classifier	SEN (%)	SPEC (%)	FPR (h^{-1})	Preictal duration (min)
Cho et al., 2017 [44]	CHB-MIT	21	65	10.83	Phase locking value	SVM	82.44[a]	82.76	–	5
Zhang and Parhi, 2016 [30]	CHB-MIT	17	78	647	Absolut/Relative spectral power	SVM	98.68	–	0.046	60
Myers et al., 2016 [70]	CHB-MIT	10	31	62	Phase/Amplitude locking value	–	77	–	0.17	60
Zandi et al., 2013 [34],	CHB-MIT	3	18	273	Zero crossings, similarity/dissimilarity index	–	83.81	–	0.165	40
	Own dataset	17	68	289	Zero crossings, similarity/dissimilarity index	–	91.2	–	0.07	40
Chu et al., 2017 [71]	CHB-MIT	13	125	434.6	Fourier Transform coefficients, PSD	–	83.33	–	0.392	86
	Own dataset	3	18	148.4	Fourier Transform coefficients, PSD	–	77.78	–	0.483	86
Truong et al., 2017 [47]	CHB-MIT	13	64	311.4	STFT spectral images	CNN	81.2	–	0.16	5
Khan et al., 2017 [48]	CHB-MIT	15	18	70.5	Wavelet Transform coefficients	CNN	83.33	–	0.147	10
	Own dataset	12	15	24.25	Wavelet Transform coefficients	CNN	93.33	–	0.128	10
Alotaiby et al., 2017 [69]	CHB-MIT	24	170	982.9	Common spatial pattern statistics	LDA	81 87 89	61 51 37	0.47 0.4 0.39	60 90 120
This work	CHB-MIT	24	185	979.9	Statistical moments, zero crossings, Wavelet Transform coefficients, PSD, cross-correlation, graph theory	LSTM	100/ 99.28[a] 100/ 99.38[a] 100/ 99.63[a] 100/ 99.84[a]	99.28 99.60 99.78 99.86	0.11 0.06 0.03 0.02	15 30 60 120

[a] Segment-based evaluation.

the first time that the entire volume of the CHB-MIT database is utilized for a complete evaluation including all EEG recordings from the 24 cases, since, besides [69], only a small proportion of the available recordings have been used in the majority of the remaining studies. As mentioned before, this is the first study to introduce LSTM networks for seizure prediction applications as only convolutional neural networks (CNN) have been studied before in Refs. [47,48], presenting lower levels of seizure prediction performance. These results, demonstrate the advantage of LSTM networks in EEG based seizure prediction, extending the lead of LSTM over CNN that had been previously reported in different, no seizure prediction related EEG applications [19–22]. Finally, the CHB-MIT database seems more challenging for seizure prediction, as the results in Refs. [34] and [48] were significantly better when the same methods were evaluated with private EEG datasets, respectively.

A limitation of this study lies within the framework of the evaluation process, since the limited number of preictal EEG segments led us to apply segment shuffling to reduce the impact of overfitting and force the LSTM model to extract more generic preictal information from the entire duration of the preictal state instead. This approach does not allow to evaluate the average prediction time (i.e. the duration from the first preictal segment identification to the actual seizure onset). Towards this direction, a similar seizure prediction system that would isolate timely preictal patterns and provide different alarms in multiple time zones as the time to seizure onset decreases (e.g. 1–2 h before seizure, 30 min to 1 h, 15–30 min, etc.) would help to more accurately predict seizures and reduces false predictions and is planned for our future work.

6. Conclusions

As the EEG data increase in volume and complexity, deep learning algorithms begin to demonstrate their capabilities in handling the chaotic nature of the EEG signals and open new opportunities in challenging biomedical applications such as epileptic seizure prediction. Long short-term memory networks (LSTM) were introduced in seizure prediction in this work and proven to be an ideal tool for the analysis of preictal EEG signals. The proposed methodology was able to provide solid seizure prediction performance with seizure prediction and false alarm rates that have not been reached before, in every preictal duration tested from 15 min to 2 h. The proposed methodology needs to be extensively tested in clinical practice with more EEG data, as the CHB-MIT database consists of mostly pediatric subjects, but the results of this study provide strong indications of its usefulness as an effective prediction tool for epilepsy and its timely intervention for individuals.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] F. Mormann, R.G. Andrzejak, C.E. Elger, K. Lehnertz, Seizure prediction: the long and winding road, Brain 130 (2006) 314–333.
[2] M.J. Cook, T.J. O'Brien, S.F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, et al., Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study, Lancet Neurol. 12 (2013) 563–571.
[3] R. Fisher, V. Salanova, T. Witt, R. Worth, T. Henry, R. Gross, et al., Electrical stimulation of the anterior nucleus of thalamus for treatment of refractory epilepsy, Epilepsia 51 (2010) 899–908.
[4] A. Aarabi, B. He, Seizure prediction in patients with focal hippocampal epilepsy, Clin. Neurophysiol. 128 (2017) 1299–1307.
[5] L.D. Iasemidis, D.S. Shiau, P.M. Pardalos, W. Chaovalitwongse, K. Narayanan, A. Prasad, et al., Long-term prospective on-line real-time seizure prediction, Clin. Neurophysiol. 116 (2005) 532–544.

[6] A. Schad, K. Schindler, B. Schelter, T. Maiwald, A. Brandt, J. Timmer, et al., Application of a multivariate seizure detection and prediction method to non-invasive and intracranial long-term EEG recordings, Clin. Neurophysiol. 119 (2008) 197–211.

[7] P. van Mierlo, M. Papadopoulou, E. Carrette, P. Boon, S. Vandenberghe, K. Vonck, et al., Functional brain connectivity from EEG in epilepsy: seizure prediction and epileptogenic focus localization, Prog. Neurobiol. 121 (2014) 19–35.

[8] S. Ramgopal, S. Thome-Souza, M. Jackson, N.E. Kadish, I.S. Fernández, J. Klehm, et al., Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy, Epilepsy Behav. 37 (2014) 291–307.

[9] T.N. Alotaiby, S.A. Alshebeili, T. Alshawi, I. Ahmad, F.E.A. El-Samie, EEG seizure detection and prediction algorithms: a survey, EURASIP J. Appl. Signal Process. (2014) 183.

[10] K.M. Tsiouris, A.T. Tzallas, S. Markoula, D. Koutsouris, S. Konitsiotis, D.I. Fotiadis, "A Review of Automated Methodologies for the Detection of Epileptic Episodes Using Long-term EEG Signals," Handbook of Research on Trends in the Diagnosis and Treatment of Chronic Conditions, vol. 231, 2015.

[11] K. Gadhoumi, J.-M. Lina, F. Mormann, J. Gotman, Seizure prediction for therapeutic devices: a review, J. Neurosci. Meth. 260 (2016) 270–282.

[12] S. Min, B. Lee, S. Yoon, Deep learning in bioinformatics, Briefings Bioinf. 18 (2017) 851–869.

[13] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[14] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (1997) 1735–1780.

[15] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, D. Wunsch, Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG, Neurocomputing 30 (2000) 201–218.

[16] A. Graves, Neural networks, in: Supervised Sequence Labelling with Recurrent Neural Networks, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 15–35.

[17] F.A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: continual prediction with LSTM, in: 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470) vol. 2, 1999, pp. 850–855.

[18] S. Hochreiter, The vanishing gradient problem during learning recurrent neural nets and problem solutions, Int. J. Uncertain. Fuzziness Knowledge-Based Syst. 6 (1998) 107–116.

[19] Z. Ni, A.C. Yuksel, X. Ni, M.I. Mandel, L. Xie, Confused or not Confused?: disentangling brain activity from EEG data using bidirectional LSTM recurrent neural networks, in: Presented at the Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics, Boston, Massachusetts, USA, 2017.

[20] M.M. Hasib, T. Nayak, Y. Huang, A hierarchical LSTM model with attention for modeling EEG non-stationarity for human decision prediction, in: 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2018, pp. 104–107.

[21] S. Tripathi, S. Acharya, R.D. Sharma, S. Mittal, S. Bhattacharya, Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset, in: AAAI, 2017, pp. 4746–4752.

[22] S. Alhagry, A.A. Fahmy, R.A. El-Khoribi, Emotion recognition based on EEG using LSTM recurrent neural network, Emotion 8 (2017).

[23] B. Direito, F. Ventura, C. Teixeira, A. Dourado, Optimized feature subsets for epileptic seizure prediction studies, in: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, 2011, pp. 1636–1639.

[24] J. Rasekhi, M.R.K. Mollaei, M. Bandarabadi, C.A. Teixeira, A. Dourado, Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods, J. Neurosci. Meth. 217 (2013) 9–16.

[25] A. Greaves, A. Raghuvanshi, K.-Y. Neo, Predicting Seizure Onset with Intracranial Electroencephalogram (EEG) Data-project Report, 2014.

[26] N. Moghim, D. Corne, Evaluating bio-inspired approaches for advance prediction of epileptic seizures, in: Nature and Biologically Inspired Computing (NaBIC), 2011 Third World Congress on, 2011, pp. 540–545.

[27] K. Gadhoumi, J.-M. Lina, J. Gotman, Discriminating preictal and interictal states in patients with temporal lobe epilepsy using wavelet analysis of intracerebral EEG, Clin. Neurophysiol. 123 (2012) 1906–1916.

[28] L. Wang, C. Wang, F. Fu, X. Yu, H. Guo, C. Xu, et al., Temporal lobe seizure prediction based on a complex Gaussian wavelet, Clin. Neurophysiol. 122 (2011) 656–663.

[29] Y. Park, L. Luo, K.K. Parhi, T. Netoff, Seizure prediction with spectral power of EEG using cost-sensitive support vector machines, Epilepsia 52 (2011) 1761–1770.

[30] Z. Zhang, K.K. Parhi, Low-complexity seizure prediction from iEEG/sEEG using spectral power and ratios of spectral power, IEEE Trans. Biomed. Circuits Syst. 10 (2016) 693–706.

[31] M. Bandarabadi, C.A. Teixeira, J. Rasekhi, A. Dourado, Epileptic seizure prediction using relative spectral power features, Clin. Neurophysiol. 126 (2015) 237–248.

[32] W. Truccolo, J.A. Donoghue, L.R. Hochberg, E.N. Eskandar, J.R. Madsen, W.S. Anderson, et al., Single-neuron dynamics in human focal epilepsy, Nat. Neurosci. 14 (2011) 635.

[33] S. Li, W. Zhou, Q. Yuan, Y. Liu, Seizure prediction using spike rate of intracranial EEG, IEEE Trans. Neural Syst. Rehabil. Eng. 21 (2013) 880–886.

[34] A.S. Zandi, R. Tafreshi, M. Javidan, G.A. Dumont, Predicting epileptic seizures in scalp EEG based on a variational bayesian Gaussian mixture model of zero-crossing intervals, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 60 (2013) 1401–1413.

[35] S.M.R. Miri, A.M. Nasrabadi, A new seizure prediction method based on return map, in: 2011 18th Iranian Conference of Biomedical Engineering (ICBME), 2011, pp. 244–248.

[36] J.R. Williamson, D.W. Bliss, D.W. Browne, J.T. Narayanan, Seizure prediction using EEG spatiotemporal correlation structure, Epilepsy Behav. 25 (2012) 230–238.

[37] J.P. Lachaux, E. Rodriguez, J. Martinerie, F.J. Varela, Measuring phase synchrony in brain signals, Hum. Brain Mapp. 8 (1999) 194–208.

[38] D.B. West, Introduction to Graph Theory Vol. 2: Prentice Hall Upper Saddle River, 2001.

[39] C.J. Stam, G. Nolte, A. Daffertshofer, Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources, Hum. Brain Mapp. 28 (2007) 1178–1193.

[40] R.D. Pascual-Marqui, D. Lehmann, M. Koukkou, K. Kochi, P. Anderer, B. Saletu, et al., Assessing interactions in the brain with exact low-resolution electromagnetic tomography, Philos. Trans. A Math. Phys. Eng. Sci. 369 (2011) 3768–3784.

[41] E. van Dellen, H. de Waal, W.M. van der Flier, A.W. Lemstra, A.J. Slooter, L.L. Smits, et al., Loss of EEG network efficiency is related to cognitive impairment in dementia with lewy bodies, Mov. Disord. 30 (2015) 1785–1793.

[42] E. van Diessen, W.M. Otte, K.P.J. Braun, C.J. Stam, F.E. Jansen, Improved diagnosis in children with partial epilepsy using a multivariable prediction model based on EEG network characteristics, PLoS One 8 (2013) e59764.

[43] S. Su, D. Yu, J. Cheng, Y. Chen, X. Zhang, Y. Guan, et al., Decreased global network efficiency in young male smoker: an EEG study during the resting state, Front. Psychol. 8 (2017) 1605.

[44] D. Cho, B. Min, J. Kim, B. Lee, EEG-based prediction of epileptic seizures using phase synchronization elicited from noise-assisted multivariate empirical mode decomposition, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (2017) 1309–1318.

[45] S. Wang, W.A. Chaovalitwongse, S. Wong, Online seizure prediction using an adaptive learning approach, IEEE Trans. Knowl. Data Eng. 25 (2013) 2854–2866.

[46] K.M. Tsiouris, V.C. Pezoulas, D.D. Koutsouris, M. Zervakis, D.I. Fotiadis, Discrimination of preictal and interictal brain states from long-term EEG data, in: 2017 IEEE 30th International Symposium on Computer-based Medical Systems (CBMS), 2017, pp. 318–323.

[47] N.D. Truong, A.D. Nguyen, L. Kuhlmann, M.R. Bonyadi, J. Yang, O. Kavehei, A Generalised Seizure Prediction with Convolutional Neural Networks for Intracranial and Scalp Electroencephalogram Data Analysis, 2017 arXiv preprint arXiv:1707.01976.

[48] H. Khan, L. Marcuse, M. Fields, K. Swann, B. Yener, Focal onset seizure prediction using convolutional networks, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. (2017), https://doi.org/10.1109/TBME.2017.2785401.

[49] P. Mirowski, D. Madhavan, Y. LeCun, R. Kuzniecky, Classification of patterns of EEG synchronization for seizure prediction, Clin. Neurophysiol. 120 (2009) 1927–1940.

[50] R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, et al., Deep learning with convolutional neural networks for EEG decoding and visualization, Hum. Brain Mapp. 38 (2017) 5391–5420.

[51] P.R. Davidson, R.D. Jones, M.T. Peiris, Detecting behavioral microsleeps using EEG and LSTM recurrent neural networks, Conf. Proc. IEEE Eng. Med. Biol. Soc. 6 (2005) 5754–5757.

[52] P.R. Davidson, R.D. Jones, M.T.R. Peiris, EEG-based lapse detection with high temporal resolution, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 54 (2007) 832–839.

[53] P. Thodoroff, J. Pineau, A. Lim, Learning robust features using deep learning for automatic seizure detection, in: Presented at the Proceedings of the 1st Machine Learning for Healthcare Conference, Proceedings of Machine Learning Research, 2016.

[54] M. Golmohammadi, S. Ziyabari, V. Shah, S.L. de Diego, I. Obeid, J. Picone, Deep Architectures for Automated Seizure Detection in Scalp EEGs, 2017 arXiv preprint arXiv:1712.09776.

[55] H. Dong, A. Supratak, W. Pan, C. Wu, P.M. Matthews, Y. Guo, Mixed neural network approach for temporal sleep stage classification, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (2018) 324–333.

[56] R.G. Hefron, B.J. Borghetti, J.C. Christensen, C.M.S. Kabban, Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation, Pattern Recogn. Lett. 94 (2017) 96–104.

[57] CHB-mit scalp EEG database, Physionet.org, [Online]. Available: https://www.physionet.org/pn6/chbmit. (Accessed 4 February 2018).

[58] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (2000) E215–E220.

[59] B. Direito, J. Duarte, C. Teixeira, B. Schelter, M.L. Van Quyen, A. Schulze-Bonhage, et al., Feature selection in high dimensional EEG features spaces for epileptic seizure prediction, IFAC Proc. Vol. 44 (2011) 6206–6211.

[60] F. Mormann, T. Kreuz, C. Rieke, R.G. Andrzejak, A. Kraskov, P. David, et al., On the predictability of epileptic seizures, Clin. Neurophysiol. 116 (2005) 569–587.

[61] M. Rubinov, O. Sporns, Complex network measures of brain connectivity: uses and interpretations, Neuroimage 52 (2010) 1059–1069.

[62] F. Chollet, Keras [Online]. Available: https://github.com/fchollet/keras. (Accessed 4 February 2018).

[63] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, et al., Tensorflow: Large-scale Machine Learning on Heterogeneous Distributed Systems, 2016 arXiv preprint arXiv:1603.04467.

[64] R. Hussein, H. Palangi, R. Ward, Z.J. Wang, Epileptic Seizure Detection: a Deep Learning Approach, 2018 arXiv preprint arXiv:1803.09848.

[65] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Comput. Biol. Med. (2017), https://doi.org/10.1016/j.compbiomed.2017.09.017.
[66] I. Ullah, M. Hussain, E.-u.-H. Qazi, H. Aboalsamh, An Automated System for Epilepsy Detection Using EEG Brain Signals Based on Deep Learning Approach, 2018 arXiv preprint arXiv:1801.05412.
[67] M.A. Hall, L.A. Smith, Practical feature subset selection for machine learning, in: C. McDonald (Ed.), Computer Science '98 Proceedings of the 21st Australasian Computer Science Conference ACSC'98, Perth, 4-6 February, 1998, pp. 181–191.
[68] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, ACM SIGKDD Explor. Newsl. 11 (2009) 10–18.
[69] T.N. Alotaiby, S.A. Alshebeili, F.M. Alotaibi, S.R. Alrshoud, Epileptic seizure prediction using CSP and LDA for scalp EEG signals, Comput. Intell. Neurosci. (2017) 11, 2017.
[70] M.H. Myers, A. Padmanabha, G. Hossain, A.L. de Jongh Curry, C.D. Blaha, "Seizure prediction and detection via phase and amplitude lock values, Front. Hum. Neurosci. 10 (2016) 80.
[71] H. Chu, C.K. Chung, W. Jeong, K.H. Cho, Predicting epileptic seizures from scalp EEG based on attractor state analysis, Comput. Meth. Progr. Biomed. 143 (2017) 75–87.