IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 27, 2019, accepted November 24, 2019, date of publication December 13, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959234

Detection of Epilepsy Seizures in Neo-Natal EEG Using LSTM Architecture

MUHAMMAD U. ABBASI[1], (Member, IEEE), ANUM RASHAD[1], (Member, IEEE), ANAS BASALAMAH[2], (Member, IEEE), AND MUHAMMAD TARIQ[1], (Senior Member, IEEE)
[1]Department of Electrical Engineering, National University of Computer and Emerging Sciences, Peshawar 25000, Pakistan
[2]Computer Engineering Department, Umm Al Qura University, Mecca 24231, Saudi Arabia

Corresponding author: Muhammad Tariq (mtariq@princeton.edu)

ABSTRACT Epilepsy is the most unpredictable and recurrent disease among neurological diseases. Early detection of epileptic seizures can play a critical role in providing timely treatment to patients especially when a patient is in a remote area. This paper uses deep learning framework to detect epilepsy in the Electroencephalography (EEG) signal. The dataset used is publicly available and has a recording of three kinds of EEG signals: pre-ictal, inter-ictal (seizure-free epileptic) and ictal (epileptic with seizure). The proposed Long Short-Term Memory (LSTM) classifier classifies these three kinds of signals with up to 95% accuracy. For binary classification such as detection of inter-ictal or ictal only, its accuracy increases to 98%. The EEG signal is modelled as wide sense non-stationary random signal. Hurst Exponent and Auto-regressive Moving Average (ARMA) features are extracted from each signal. In this work, two different configurations of LSTM architecture: single-layered memory units and double-layered memory units are also modelled. After standardising the features, double-layered LSTM approach gives the highest accuracy in comparison to previously used Support Vector Machine (SVM) classifier and proved to be computationally efficient at Graphics Processing Unit (GPU).

INDEX TERMS Deep learning, neo-natal EEG, LSTM architecture, desnoising, biomedical signal processing.

I. INTRODUCTION

Epilepsy is a harmful disease which affects millions of people around the world. It is a brain disease which occurs due to a chronic neurological disorder of neurons producing an abnormal signal. Any process which affects the neuronal activity whether it is from illness or brain damage can cause a seizure. It is caused by a sudden disturbance which appears in the brain functions [1]. The seizures occur due to a sudden change in brain activity. EEG is an effective method commonly used for monitoring the brain activity diagnosis of epilepsy. An EEG signal is a band-limited frequency (0.1Hz to 60Hz) modeled and classified into five categories: delta (0.1Hz to 4Hz), theta (4Hz to 8Hz), alpha (8Hz to 12Hz), beta (12Hz to 30Hz) and gamma (30Hz to 60Hz) waves that record different brain activities [2]. The four main stages of transitions during the epileptic seizure cycle are (i) pre-ictal, (ii) ictal, (iii) post-ictal and (iv) inter-ictal. Pre-ictal is the period before the seizure, ictal is the interval during which the seizure occurs, post-ictal is the period at the end of a seizure and inter-ictal is the time between two seizures [3], [4]. The signal captured at the ictal stage represents the time in which the seizure occurs. Most neurologists can detect upcoming seizures by inspecting changes in EEG recordings [2]. These four transitional stages are calculated by the EEG signal. The ictal stage can be easily identified from the recorded intracranial EEG signal.

A. SEIZURE PREDICTION

The process of discrimination between the pre-ictal and inter-ictal states is known as seizure prediction. It is also called seizure forecasting. By detecting the appearance of a pre-ictal state the epileptic seizure is predicted. Machine learning approaches are used to predict the epileptic seizure which includes EEG signal acquisition, pre-processing, feature extraction and the classification of the seizure states.

Threshold-based techniques are used for the prediction of seizure where the analysis is targeting a high/low change in the values of some features during the pre-ictal stage.

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

The classification method can also be performed on the raw data of the EEG signal. The prediction of the seizure depends on the threshold value of the features. If the value of the estimated features exceeds the activation value (pre-ictal value), an alarm is raised to signal an incoming seizure [5].

The current machine learning techniques or deep learning approaches are commonly used for the prediction and treatment of epilepsy seizure. The Adaptive Neural Network (ANN) and SVM are both widely used by researchers [5]–[8]. In these schemes, the EEG signal data are classified by features which are extracted from the recorded EEG data signal. A binary classifier is trained to obtain the difference between the two (pre-ictal and inter-ictal) stages. The ictal and post-ictal segments cannot help in seizure prediction and are removed from the analysis [4]–[10] in most of the applications of deep learning algorithms used in medical image and signal processing. The main reason is the large computational power and big data, which shows the high impact and significance in most cases. Convolutional Neural Network (CNN) can be used for seizure prediction because CNN provides better results in the image processing field [7].

The seizure prediction algorithms are not only accurate but also employable in real-time. The pre-processing and feature extraction from an EEG signal plays an important role in improving true positive rates and prediction time. The noise and unwanted components from the signal are removed in the pre-processing of the EEG signal. The main advantage of pre-processing EEG signal is that it increases the Signal-to-Noise Ratio (SNR). Many filters are used by the researchers like Common Spatial Pattern (CSP) and a large Laplacian filter for the pre-processing of the EEG signal. Some machine learning approaches like Discrete Cosine Transform (DCT), Fast Fourier Transform (FFT) and graph theory are used for the feature extraction process. Classifiers like SVM, Least-Squares SVM (LS-SVM) were used to detect a seizure signal from the EEG signal [11].

B. RESEARCH HIGHLIGHTS

SVM classification performs well for binary classification problems but the non-stationary Gaussian EEG signals have variations at multiple stages before epilepsy. The Bonn University data for epilepsy detection is used which has five classes. The EEG signal in the data is recorded from a normal person with eyes open/close and persons who have epilepsy. So, we focused on multiclass problem and proposed a new architecture of deep learning network-LSTM. Two main highlights of this work are as under:

1) a new LSTM architecture is selected after two different LSTM layers group evaluation.
2) Hurst and ARMA features are extracted and a 20 features dimension for a channel are created for the EEG signal as the Brownian movement of neurons.
3) the extracted features are standardised as a pre-processing step before LSTM training. This is done to avoid the biasing in LSTM training towards any particular data class.
4) the experimenting is done for cases like pre-ictal vs ictal, pre-ictal vs inter-ictal, inter-ictal vs ictal, pre-ictal vs inter-ictal vs ictal.

II. RELATED WORK

As we mentioned in the previous section, epilepsy is one of the common neurological disorders characterized by a sudden malfunction of the brain which causes a seizure. 50 million people are affected by epilepsy worldwide [2]. EEG is generally used for the detection and treatment of epilepsy seizures. A Heart-Rate Variability (HRV) based epileptic seizure method [12] that predicts monitoring was used in which the interval between R-waves in the ECG (RRI) data recorded from patients was translated into HRV features and then monitored by Multi-variate Statistical Process Control (MSPC). The Pre-processing of data was done by Common Average Reference (CAR) filter [13], Independent Component Analysis (ICA) [14], Anti-aliasing filter sets [3], Sparse optimization [15], 10-fold cross-validation [10] and Butterworth (IIR) filter [16].

Mathematical and statistical methods like HRV [12], Empirical Mode Decomposition (EMD), Singular Value Decomposition (SVD), Discrete Wavelet Transform (DWT) [17], Autoregressive (AR) [2], Multi-rate filter bank and DCT filter [2], Power Spectral Density (PSD) [3], FFT, Butterworth bandpass filter and Cross Correlation (XCORR) [6], Averaged Instantaneous Envelope (AIE) and Averaged Instantaneous Frequency (AIF) [4] were used in the previous research for features extraction. In addition, graph theory was used to measure the local and global features for a cluster field [7].

The features are the sets of coefficients of a linear model built for each channel of EEG data. These features were used for classification purposes. The classifier generally used SVM [1], [2], [6], Multi-variate Statistic Process Control (MSPC) [12], Recursive and Sequential Multiple Model (RSMM) [18], LS-SVM [9], CNN [7], [10], LSTM [5], Multi-layer Perceptron (MLP) [8], K-nearest neighbor and naïve bays [19] in the previous research work. In most of the previous work, the SVM classifier was used. Deep learning methods had large applications as a classifier.

The EEG signals were preprocessed by the ICA method and Fuzzy Multichannel EEG Classifier (FMCEC) used for the classification and the feature extraction process [1]. Similarly, in [10], the data was pre-processed by Computer-Aided Design (CAD) and classified by CNN which detected normal, pre-ictal and seizure classes. The EEG data features were extracted by the Amplitude and Frequency Modulated (AM-FM) model in which the amplitude dominant signal and frequency dominant signal identify the features. The internal process like Averaged Instantaneous Envelope (AIE) and Averaged Instantaneous Frequency (AIF) parameters were produced. SVM tested the feature vectors and classified the

EEG signal [4], [20], [21]. A significant amount of research of seizure prediction was conducted in [5], [7]–[10], [16], [17], [19], [22]–[29]. The combination of multiple features into the feature vector for the epileptic seizure can be predicted. The EEG signal of brain activity was pre-processed by the EMD method and then features were extracted by the time-frequency domain. The extracted features were classified by the SVM classifier [19]. The EEG recorded signals are very reliable for the treatment of epilepsy seizure. A great number of uni-variate features were extracted [9], [24] but none of them obtained high performance when compared to the bi-variate techniques. The EEG signal analysis obtains large information from the signal. This can be done by the transformation used in [17]. The wavelet transform was used in previous research widely.

A method was proposed in [25] where the EEG and ECG data were analyzed by the DWT pre-processing method. Sequential Forward Selection (SFS) was provided for the features selection and K-nearest neighbour or Linear Bayes was implemented as the classifier [25]. The ECG signal could be a potential resource for predicting the epileptic seizure. A Short-Time Fourier Transform (STFT) was used in [26] for the transformation of EEG data into a matrix form, taken from Children's Hospital of Boston-Massachusetts Institute of Technology (CHB-MIT).

The output of STFT was provided to the CNN which extracted the features from the transformed data. A discrete Kalman filter was used as the classifier. K of n-filter provided the prediction of an epileptic seizure. A phase space representation method was proposed in [27] for the feature extraction of the EEG signal. The EEG signal was pre-processed by the EMD method. Propeller Shaft Rate (PSR) extracted the features from the EEG signal. LS-SVM was used as the classifier which predicts the epileptic seizure from EEG signals. A publically available dataset was used for the experiment. The weighted Extreme Learning Machine (ELM) was provided in [28] for the epileptic seizure from the imbalance EEG data set. The dataset contains the normal signal as well as the seizure signal. The wavelet transform was used to pre-process the seizure signal and Pattern Match Regularity Statistic (PMRS) was used for the non-seizure signal. The Notch filter was used in [29] for pre-processing the data. Shannon Entropy (PE) provided the feature's selection. Finally, the time series domain features were extracted by the PE method.

III. DATASET

The data used in our study is available on the EEG time series page of University of Bonn, Germany. All the data samples were recorded by the test of epilepsy patients in the Department of Epileptology at Bonn University [22]. The data contains five segments ($A - E$). Each segment contains 100 text files. Each text file consists of 4096 samples of one EEG time series in ASCII code. The EEG time-series signals were recorded with the same 128-channel amplifier system. The five sets of signals were recorded in this dataset. The data was converted into a digital form using an Analog to Digital Converter (ADC). After that, the data was continuously recorded on the disc at a sampling rate of 173.61 Hz. The filter used for the data acquisition process normally ranges from 0.53 to 40 Hz. Each signal from the A to E sets contained 100 signal channels with EEG segments of 23.6-second duration. After the visual inspection of contaminants like muscle activity, five segments were selected and separated from the continuous multi-channel EEG recordings. The segments A and B were taken out from the EEG recording surface. Both segments represented healthy person data. Segment A displayed the eyes-open state and segment B was concerned with the eyes-closed state of a healthy volunteer. The other three segments (C, D, and E) were related to the seizure condition of the brain activity. Five patients that were suffering from epileptic seizures were selected [22]. The sets C and D only contained seizure-free brain activity interval. The E segment contained the seizure activity. Only the ictal activity was represented in selected segments.

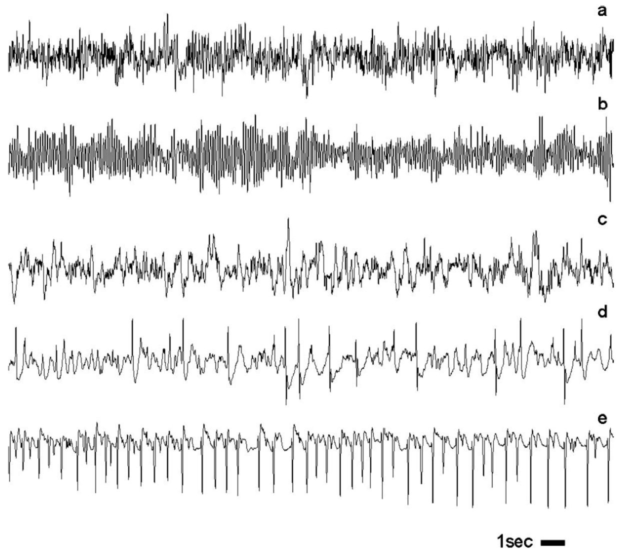


FIGURE 1. EEG time series representation for five sets (*A*, *B*, *C*, *D*, and *E*) [22].

The Figure 1 shows the electrical signal of the brain activity. These electrical signals were recorded on the EEG surfaces. Generally, the amplitude of the signal is in the form of μV. The inter-acranial EEG signal amplitude is approximately 100μV. In the seizure case, the voltage amplitude exceeds the value 100μV. The E signal amplitude exceeds the 100μV range which is ictal having almost periodic and high amplitude. The signals C and D are seizure-free sets which are recorded in the inter-ictal epileptic activities. A total of 4396 samples were cut out in the recordings based on time series spectral frequency components. At the beginning of each final segment, 4096 samples were selected. Each segment consists of 4096 samples of EEG data [22]. Figure 2 shows the number of samples in all recorded signals. Every object with a different signal condition, whether

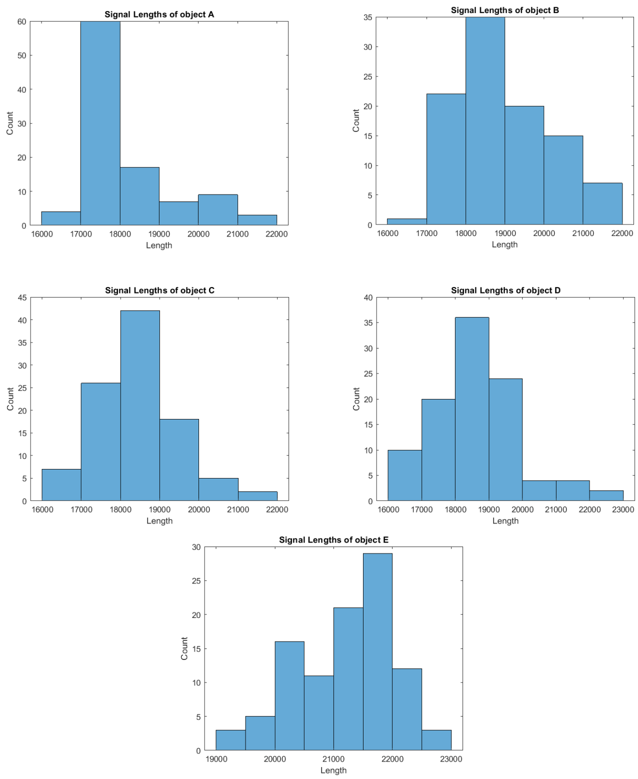


FIGURE 2. Histogram plot of the number of EEG samples distribution.

normal or elliptic, has a different number of samples. There are 17000 samples in Object A and 22000 samples in Object E.

In the deep learning training process, data is divided into mini-batches and signals are truncated or padded to match the size of a mini-batch. It may remove the required information in the signal and degrade the performance of classification/prediction. To avoid this, we segmented the data into a fixed number of 17000 samples. In this process, more than 17000 samples were rejected and fewer than 17000 were padded.

IV. LONG SHORT-TERM MEMORY ENCODER

A variation in the recurrent network by using memory blocks is called Long Short Term Memory blocks or LSTM. LSTM has the input layer, hidden layers and output layers. The hidden layers have memory blocks which are also called cells. Memory blocks contain information in a gated cell which is outside the normal flow of the recurrent network [30]. The three gates present in the LSTM are the input gate, the output gate and the forget gate. The memory cells contain the same input and the same output gate, forming a structure called memory cell block. All the data or information can be stored in a memory cell block as in computer memory. The functions, like data (read, write and erase), are performed by the cells through the open and close operation of gates. The gates are analogue, a sigmoids function multiplies element-wise and provides range 0-1. The gates have a differentiable property due to its analogue nature so it can be used for the

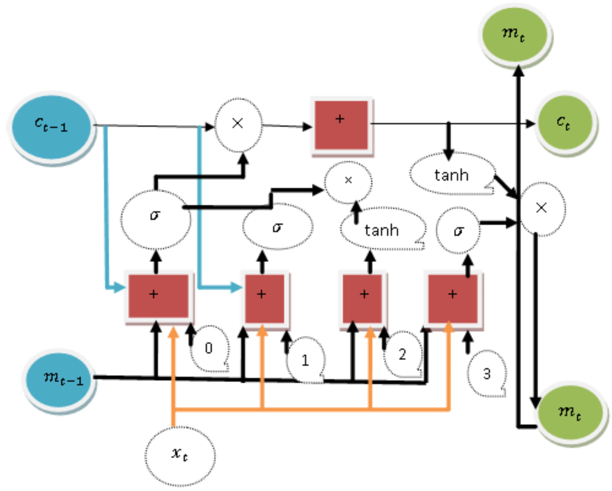


FIGURE 3. LSTM basic architecture.

back-propagation. The gates of LSTM perform a similar task like a node in Neural Network (NN) and they can pass or block the information of received signal by strength. We used the LSTM method to classify the EEG data. As a classifier, the LSTM approach provides better performance. Another approach, the SVM, can also be used as a classifier. The LSTM method is used for the large sequence time-series data but SVM is only used for small sequence time-series data. LSTM has more parameters and gating units that control the flow of information easily and provide higher accuracy in classification. Moreover, the multiplicative gates of LSTM allow the memory cells to access and store the information over a long period. Finally, the computational cost of LSTM is lower than the SVM [5].

A. LSTM ARCHITECTURE

The LSTM cell contains the following components: Forget gate (F) which is a neural network with a sigmoid function, Candidate layer (C) NN with *tanh*, Input gate (I) NN with a sigmoid, Output gate (O) NN with sigmoid, Hidden state (H) and Memory state (M) which is a vector. The basic structure of LSTM is shown in Figure 3. x_t is the input vector, $c_{(t-1)}$ memory from input blocks, $m_{(t-1)}$ previous output from the blocks. Similarly, m_t and c_t are the output and memory of the current blocks respectively. The network contains three inputs. x_t is the input current time step, $m_{(t-1)}$ is the output from the previous LSTM unit and $c_(t-1)$ represents the memory of the previous unit. Therefore, a single unit decides by considering current input, the previous output and the previous memory which generates a new output and alerts its memory.

The LSTM architecture contains the special blocks called memory blocks which are present in the hidden layers. Each memory block contains the input gate and the output gate. Both gates perform the control functions at input activation and output activation. Forget gate is added to the memory block latter. An LSTM network finds the mapping from input sequence $x = (x_1, x_2 \ldots x_T)$ to the output sequence

$y = (y_1, y_2 \dots y_T)$ by figuring out the network unit activations using following equations:

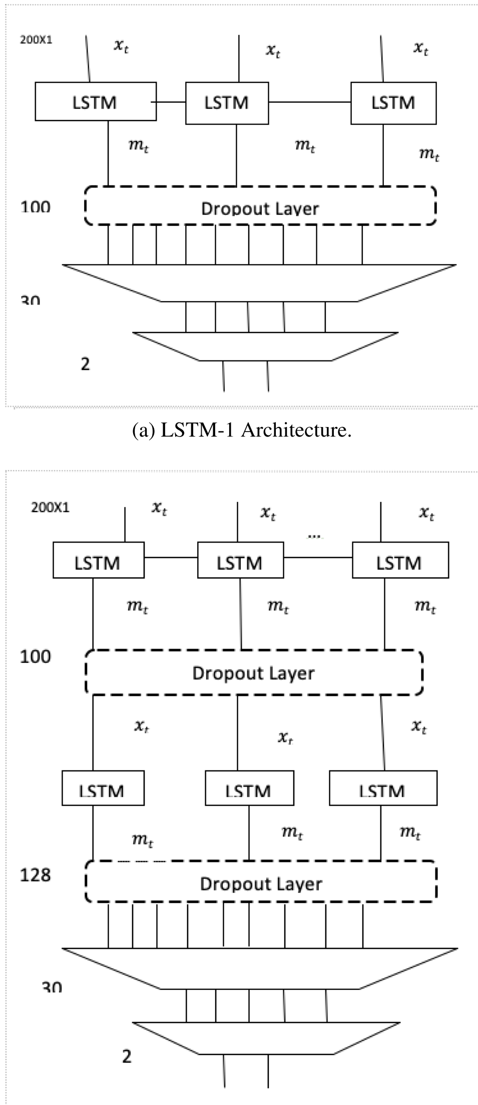
$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i), \tag{1}$$
$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f), \tag{2}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cm}m_{t-1} + b_c), \tag{3}$$
$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_{t-1} + b_o), \tag{4}$$
$$m_t = o_t \odot hc_t, \tag{5}$$
$$y_t = \phi(W_{ym}m_t + b_y). \tag{6}$$

In the above equations, W represents the weight matrices and W_ix is the maximum weight of the input gate to the input. W_ic, W_fc and W_oc are the diagonal weights of peep-holes connections. LSTM has the peep-holes connections in internal cells to the gates in the same cells. b_i is the input gate bias vector, σ is the sigmoid function i, f, o and c are the input gate, forget gate, output gate and cell activation vector. g, h are the cell input and cell output functions and ϕ is the network output activation function (softmax) [31]. The activation layer is used to learn the complex structure of the input data. The sigmoid (σ) and hyperbolic tangent ($tanh$) activation functions are used in the multilayer LSTM structure. Both activation fucntions sigmoid and hyperbolic tangent are transfomed the input value between 0 to 1 and -1 to 1. The sigmoid function is used as the gate input activation function and hyperbolic $tanh$ is used as the block input and output activation function. The non linear activation function allows LSTM to learn complex mapping functions.

B. LSTM ARCHITECTURE EVALUATION

The LSTM architecture varies with the change in the nature of data. The number of memory units in the hidden layers, the activation function and number of stacked layers are variables in any LSTM structure. The classification accuracy is affected by their various combinations. We test the two different layers of arrangement of LSTM for various memory unit size. One layer arrangement is a conventional pattern with a stacked sequence input layer, an LSTM layer, a fully connected layer, a softmax layer and finally, a classification layer. This arrangement is tested for different size of memory units $mu = [50 : 10 : 130]$ and we name it LSTM-1. This arrangement is shown in Figure 4a. In the second arrangement, the layer arrangements and memory units are the same as LSTM-1, the difference is that we made the LSTM deeper and added two LSTM layers instead of one, as shown in Figure 4b.

The previously mentioned LSTM architectures are evaluated by the EEG data. From the Bonn University EEG dataset, the normal person EEG recordings and the epileptic person recordings are used for classification purposes. It is a binary classification problem. The whole period of raw EEG recordings per case is used to find out the complete classification accuracy in the EEG segments into pre-ictal and inter-ictal classes. The five-segment EEG signal input of LSTM. The size of the LSTM input is 200 × 1, featuring the two sets of pre-ictal recordings, each of which contains 100-channel recordings. The complete raw data for each channel is the single feature. The structure of the LSTM-1 is not complex; it can be quickly saturated. So LSTM-1 provides a reduced classification accuracy as shown in Figure 5. A comparison of LSTM-1 and LSTM-2 with raw EEG signals is shown in Figure 5. LSTM-1 architecture does not give a robust conclusion for various memory units whereas, LSTM-2 is not showing significant changes with the increase in the number of memory units. So, we evaluate both architectures with features rather than raw data. The feature extraction process is discussed in the next section of this paper.



(a) LSTM-1 Architecture.

(b) LSTM-2 Architecture.

FIGURE 4. LSTM architectures.

C. METHODOLOGY

The whole work is divided into two major steps: features extraction from EEG signal and classification of EEG as epileptic or non-epileptic. Each EEG signal has five channels

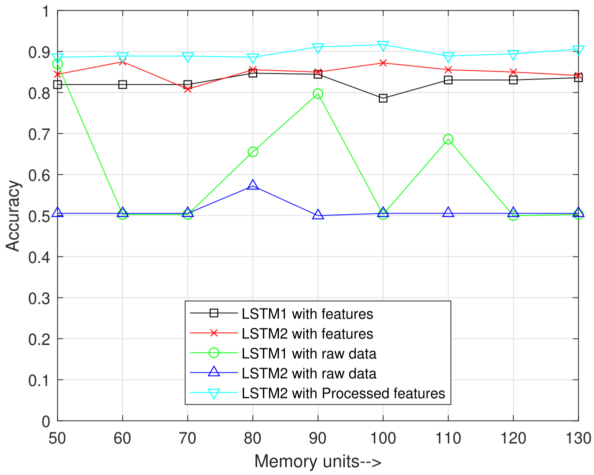


FIGURE 5. **Comparison of classification accuracy for two different LSTM architectures with raw EEG data and extracted features.**

with different frequency range: delta (0.1Hz-4Hz), theta (4Hz-8Hz), alpha(8Hz-12Hz), beta (12Hz-30Hz) and gamma (30Hz-60Hz). The features from each channel are extracted. To segment the signal into these five channels, a 2-level discrete cosine transform (DCT) is used which breaks the signal into two equal frequency range. The lower frequency range EEG signal is down-sampled further to achieve the channels. Since, the EEG signal is quite random-pattern in nature due to fractional brownian movement, Hurst exponent and ARMA features extracts the more similar nature of it. These hurst and ARMA features are extracted from the DCT segmented channels.

These features are trained and classified by the proposed LSTM architecture. In section 4.2 we rigorously tested the optimal number of memory units for LSTM to get the highest accuracy. In LSTM architecture, we have primarily two hyperparameters which can affect the epilepsy detection accuracy: number of memory blocks and LSTM layers. The Adam optimization in activation layers is widely accepted as best choice since its birth. The good learning rate with adam optimization is either 0.01 or 0.001. For image classification tasks, the learning rate of 0.001 for adam optimizer suits best but for our case 0.01 learning rate gives better accuracy. To select the memory blocks in each LSTM layer, we tested with different number of units as discussed in section 4.2. 100 units in LSTM showed the best performance for two LSTM layers. The final structure and hyperparameters after rigorous tuning of hyperparameters is tabulated in table 2. The complete methodology is shown in figure 6. Before final training of the data, it is divided into 80:20 ratio for testing and training. The complete features set is standardized also before LSTM training to avoid the overfitted results. The complete flow chart is shown in figure 7.

V. FEATURE EXTRACTION

Previously, we have seen that the accuracy of EEG signal classification is higher with features data than raw signals. In this section, we discuss the feature extraction process, considering the EEG signal as externally stimulated non-stationary signal. EEG signal is a non-stationary wide sense stationary Gaussian process [2], so its behaviour can be considered as fractional Brownian motion of multiple neurons in the brain. These are self-similar Gaussian processes as shown in the periodogram plot of a signal of set D in Figure 6. The periodogram of such a process has a mirror image. The self-similarity of these signals is measured by Hurst Exponent (H). The first order fractional Brownian motion *fBm* has $H \in (0, 1)$ [32], the second-order has similarity value $H \in [1, 2)$ and $H \in [m-1, m)$ is of m^{th} order Hurst Exponent [2]. Since the dataset collected is of external stimulus and can be modelled by *fBm*, we can use the Hurst Exponent of each EEG channel as a feature to predict epilepsy. The five sets in our Bonn University data are of 1^{st} and 2^{nd} order *fBm* as shown in Table 1.

TABLE 1. **Hurst exponent for each set in the considered dataset.**

Set	A	B	C	D	E
H	0.9159	1.2701	0.9042	0.9755	1.6509

Five brain waves are extracted from the EEG signals. These brain waves are *delta*, *theta*, *alpha*, *beta* and *gamma*. The frequency-domain of the following waves are 0.1-4Hz(*delta*), 4-8Hz (*theta*), 8-12Hz (*alpha*), 12-30Hz (*Beta*) and 30-60Hz (*gamma*). Since the EEG signal is a non-stationary random process to model the EEG signal as fractional Brownian motion, it is required to extract the un-correlated brain pattern. Multi-rate DCT filters can break the EEG into un-correlated sub-bands using the eigen-values of covariance of *fBm* process [33]. To break the signal into five sub-bands, it is passed through a 2-level DCT basis filter bank which breaks down the 0-60Hz signal to 0-30 Hz and 30-60 Hz band. The 0-30 Hz is further filtered down to 5-level decimation and higher three bands are combined to form β sub-band. The other two channels of 0-6Hz each are combined to make a channel of 0-12 Hz and again 3-level decimated filtered down to get the δ, θ, α sub-bands of 0-4Hz, 4-8Hz and 8-12Hz respectively. A block diagram of the DCT filter decimation is shown in Figure 9.

From these brain waves, the following features have been chosen: (i) Hurst Exponent (H) (ii) ARMA model.

A. HURST EXPONENT

The Hurst Exponent calculates the similarity index if the signal is as discussed earlier. We used the Maximum Likelihood (ML) method for H calculation but that method is limited only for the 1^{st} order Hurst Exponent, so it is extended for a higher-order [2]. A vector $y \in Y_m[N-1]$ of m *fBm* process is considered whose probability density function can be defined as:

$$p(y; H) = \frac{1}{2\pi^{N/2}|H|^{1/2}} e^{\frac{-(y^T k^{-1} y)}{2}}. \qquad (7)$$

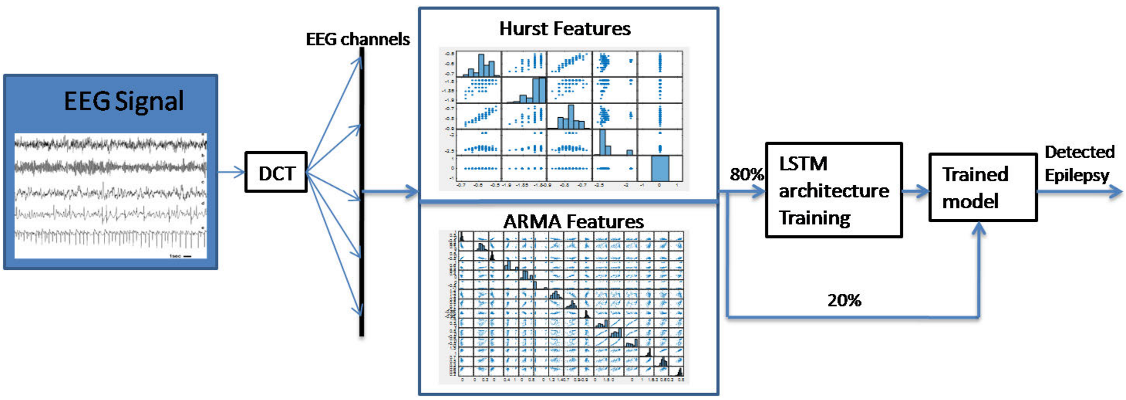


FIGURE 6. Epilepsy detection methodology by LSTM architecture.

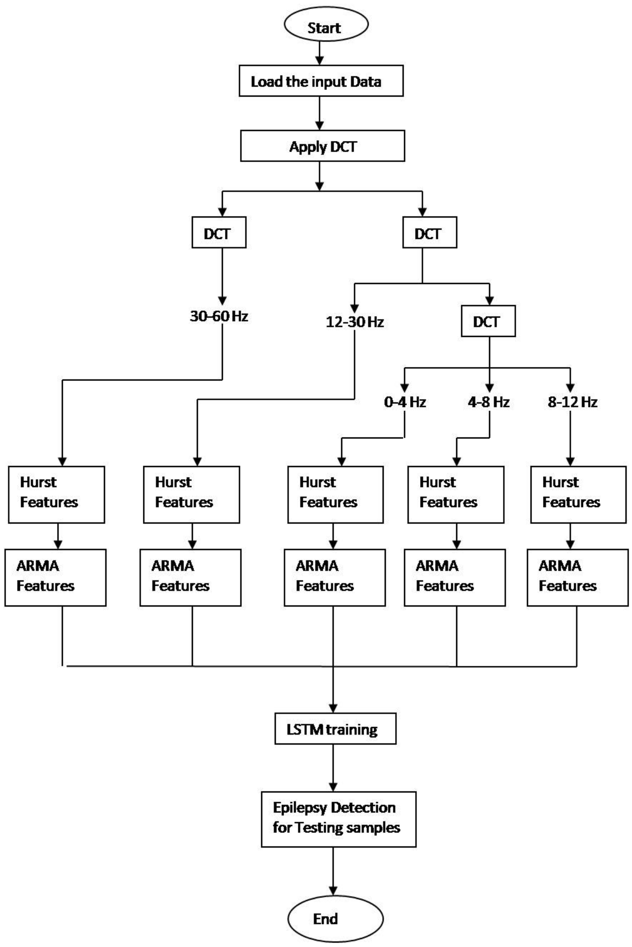


FIGURE 7. Flow chart of proposed work.

where $y = \{y_1, y_2 \ldots y_{N-1}\}^T$ is the data vector, K is the covariance matrix which is $E[xx]^T$. In extended version we used logarithm of $p(y; H)$ as:

$$\log p(y; H) = \frac{-N}{2} \log(2\pi) - \frac{1}{2} \log |K| + \frac{-(y^T k^{-1} y)}{2}. \quad (8)$$

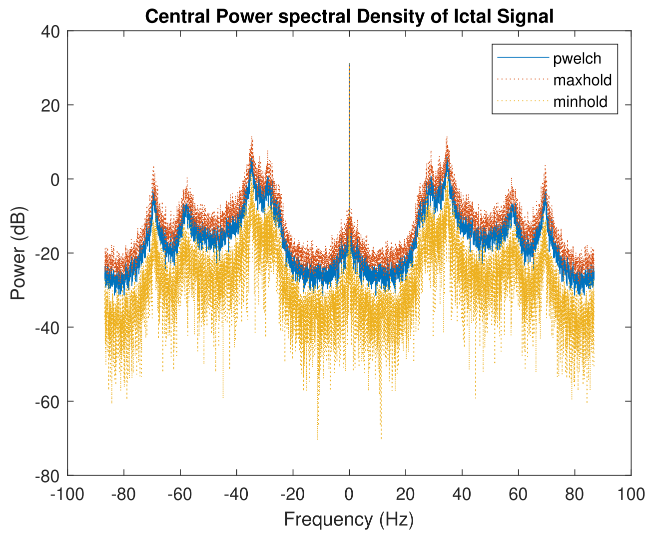


FIGURE 8. Power spectral density of wide sense stationary gaussian pre-ictal signal.

The elements of K are expressed as:

$$K_{ij} = k[i-j] = \frac{\sigma^2}{2}[|r+1|^{2h} - 2|r|^{2H} + |r-1|^{2H}]. \quad (9)$$

where $r = |i-j|$. There are two parameters to be estimated: H and σ^2. The covariance matrix K can also be represented as:

$$K = \sigma^2 K'. \quad (10)$$

Putting this value in above equation:

$$\log p(y; H) = \frac{-N}{2} \log(2\pi) - \frac{N}{2} \log \sigma_H^2 - \frac{1}{2} \log |K| + \frac{-(y^T k^{-1} y)}{2\sigma_H^2}. \quad (11)$$

To maximize this equation, derivative w.r.t H is taken and compared to zero, which gives

$$\hat{\sigma}_H^2 = \frac{-(y^T k'^{-1} y)}{2}. \quad (12)$$

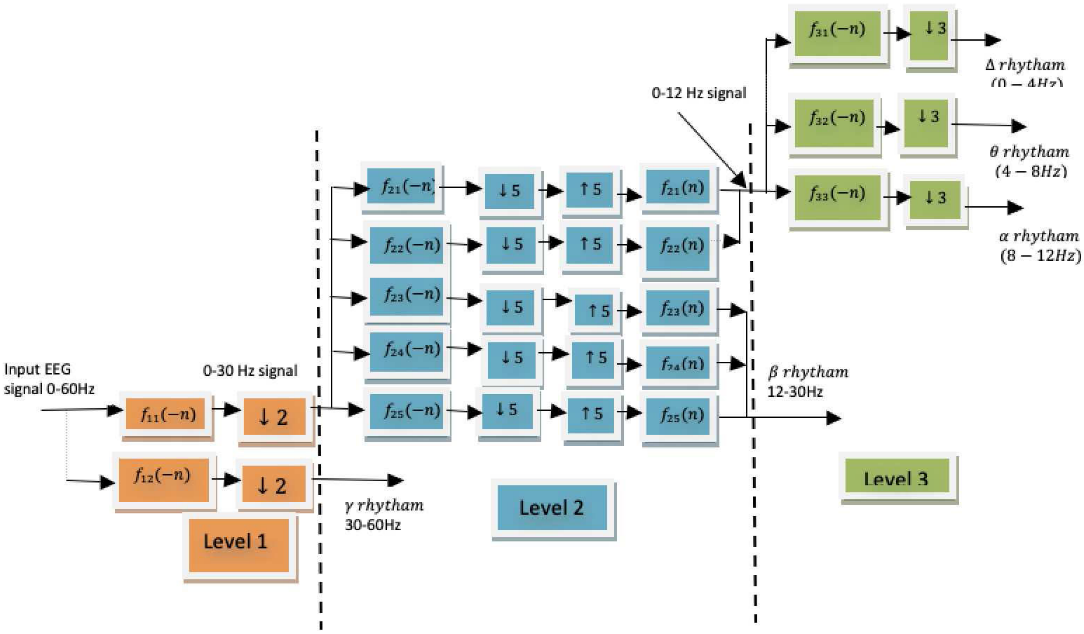


FIGURE 9. DCT basis vector based 3-level multirate filterbank structure for the extraction of EEG brain rhythms [2].

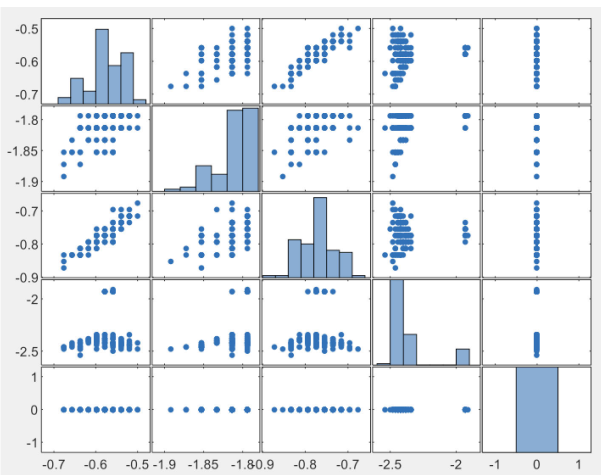


FIGURE 10. Five hurst exponent features for five sub-bands of Ictal EEG data of 100 single channels.

By putting this into (12), we get the maximization problem to get the Hurst Exponent of higher order:

$$\hat{H} arg\, max_{0<H<1}(-N \log y^T k'^{-1} y) - log|K'|) - m. \tag{13}$$

The maximization of this equation will give the $0 < H < 1$ for the 1^{st} order and $1 < H < 2$ for the 2^{nd} order Hurst Exponent. The Hurst Exponent for each sub-band will be calculated and we get the Hurst features of dimension 100×5. A scatter plot of each row to the column of features and histogram plot is shown in Figure 8 for set E.

B. ARMA FEATURE

The ARMA model predicts the nature of a random signal by using past values and present values. Let a vector represents the set k time-series at time t

$$Z_t = (Z_{t1}, Z_{t2}, Z_{t3} \dots Z_{tk})^T. \tag{14}$$

The vector of the theoretical means for Z_t given by $\mu = (\mu_1, \mu_2, \mu_3, \dots \mu_k)^T$, and T represents the transpose of the vector. If the auto-regressive order is p and the moving average is represented by q [34], then the equation of the k-dimensional model is:

$$AR = (Z_t - \mu) - \phi_1(Z_{t-1} - \mu) - \phi_2(Z_{t-2} - \mu) \dots \phi_p(Z_{t-p} - \mu). \tag{15}$$

$$MA = a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2} \dots - \dots \Theta_q a_{t-q}. \tag{16}$$

where Θ_1 is the i^{th} matrix of auto-regressive parameters order of $k \times k$ and $i = 1, 2, 3 \dots p$ and Θ_1 is the i^{th} moving average parameter matrix order of $k \times k$ where $i = 1, 2, 3 \dots q$. a_t is the k-dimensional vector of innovations for Z_t. A multivariate ARMA model is always preferred for the analysis purpose. This feature is extracted from the epileptic EEG dataset [34]. Each EEG signal sub-band gives 3-ARMA features, making a total of 15-ARMA features in an EEG signal channel.

VI. RESULTS

In the LSTM architecture evaluation section, we noticed that LSTM-2 with 100-memory units has the highest performance in Figure 5. Thus, we finalized the double LSTM layered architecture in our work. Table 2 lists all the selected parameters of the proposed LSTM architecture.

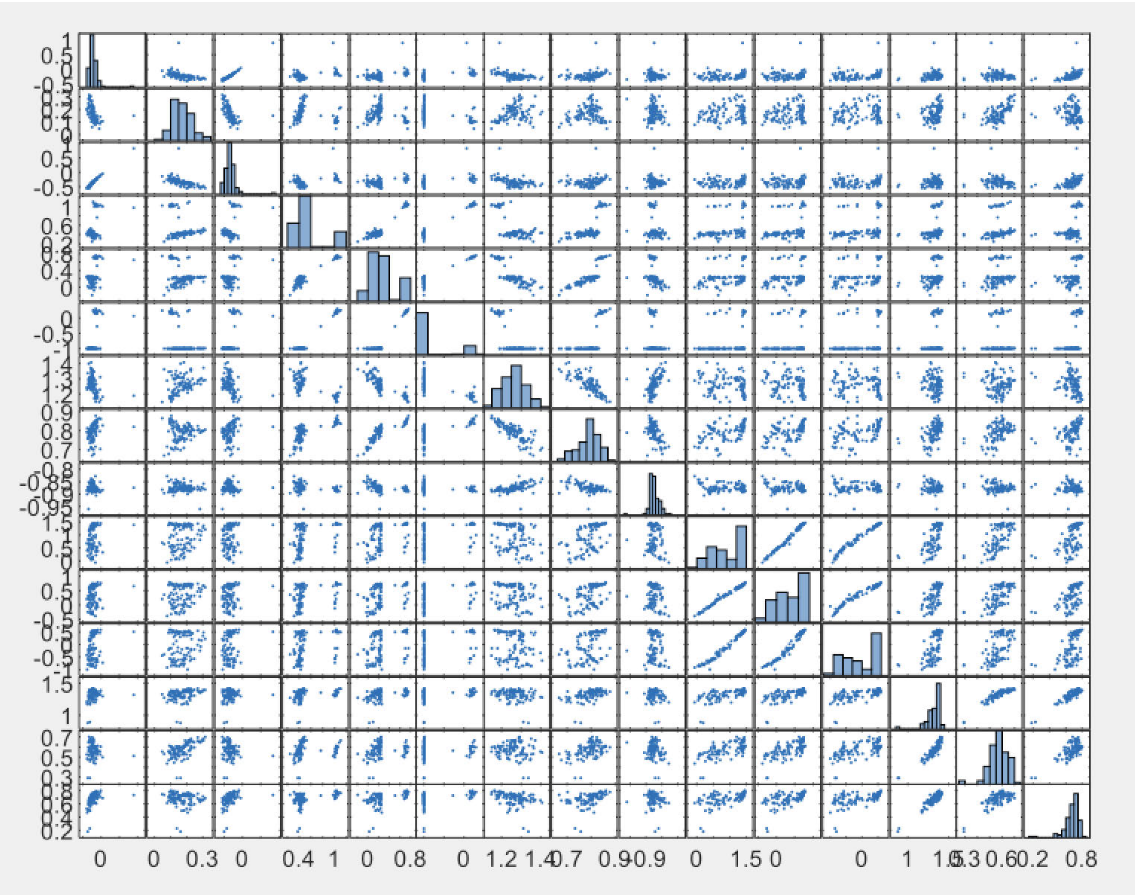


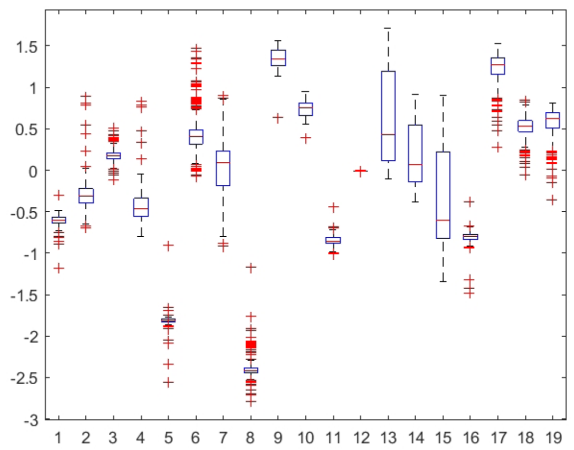
FIGURE 11. ARMA features of dimension 100 × 15 for five sub-bands of Ictal EEG data of 100 single channels.

TABLE 2. Proposed LSTM architecture parameters.

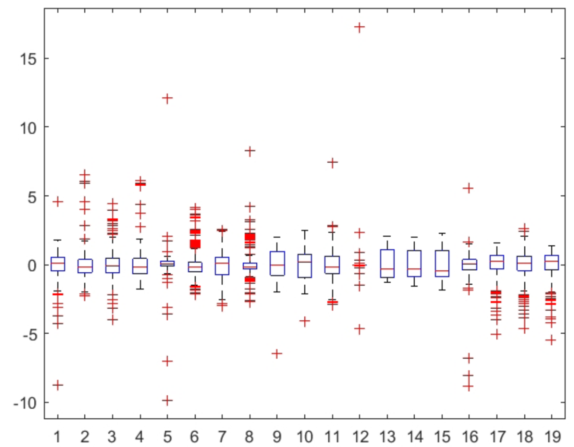
Parameters	Value
LSTM layers	2
Learning rate	0.01
Memory Units	100 for each LSTM layer
Activation Layer	adam
Epochs	100
Batch size	150
Input Features dimension	400x20

In machine learning algorithms, if the data used is highly different in their magnitudes, then training would be biased towards the high mean and hence overfitted results could be expected. The mean of Hurst feature for a signal of set E is -1.1706 whereas for ARMA features it is 0.3292 which creates the biasing of LSTM towards ARMA features. So, it is required to standardize the data. We used popular z-score which uses the mean (μ) and standard deviation (σ) approach to standardize the data ($z = (x - \mu)/\sigma$). This pre-processing of data increases the classification accuracy of an ictal signal as shown in Figure 7. We compared our results with the work of Gupta [2] which used a 10-fold SVM classification technique to predict the ictal EEG signals. Figure 12 shows the box plot for the raw featured training data and standardized data. It shows the difference between the median of every feature in the data. Every feature out of total 20 features is noisy as these differ in magnitudes but after pre-processing, the data is normalized and the LSTM training will not be biased now.

We tested the algorithm for different dataset cases which are: pre-ictal vs ictal (AB vs E), pre-ictal vs inter-ictal (A or B vs C or D), inter-ictal vs ictal (C or D vs E) and a multi-label classification problem of pre-ictal vs inter-ictal vs ictal (A vs C vs E). The linear SVM classification technique in [2] is limited to binary classification problems and can perform well only for those problem sets. However, our proposed LSTM architecture surpasses the SVM results for binary as well as achieves high accuracy for multi-label classification problems. The algorithm was tested on the single Nvidia GPU GeForce 940M for good computational efficiency. The proposed algorithm was very robust for each test case and outperformed the SVM. Table 3 lists the accuracy values of all data test cases with SVM comparison. The proposed LSTM showed a 99.17% accuracy for pre-ictal and ictal signal classifications whereas it is 97.27% for SVM [2]. Figure 11 compares the accuracy in each iteration for



(a) Training features distribution of EEG data



(b) Standardize training features

FIGURE 12. Boxplot for raw featured training data and standardize data.

TABLE 3. Comparative results for Proposed LSTM and SVM classifier for Ictal classification.

Data Set Cases	Proposed LSTM results			SVM results [2]		
	Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity
pre-ictal vs ictal (AB vs E)	0.9917	0.9945	0.9888	0.9727	0.9715	0.9740
pre-ictal vs inter-ictal (A vs C)	0.9778	0.9892	0.9570	0.9650	0.9600	0.9700
inter-ictal vs ictal (D vs E)	0.9778	0.9885	0.9677	0.9635	0.9620	0.9650
pre-ictal vs inter-ictal vs ictal (A vs C vs E)	0.9481	0.9943	0.9263	-	-	-

100-epochs for all test cases. The fastest convergence is obtained for pre-ictal vs inter-ictal (A vs C) classification. Binary classes are rarely available in real-life problems and so is the EEG signal. The proposed algorithm is also robust for multi-class EEG signal classification. We get a 94.1% accuracy for all three classes of classification in our data. It can help to detect epilepsy before it happens.

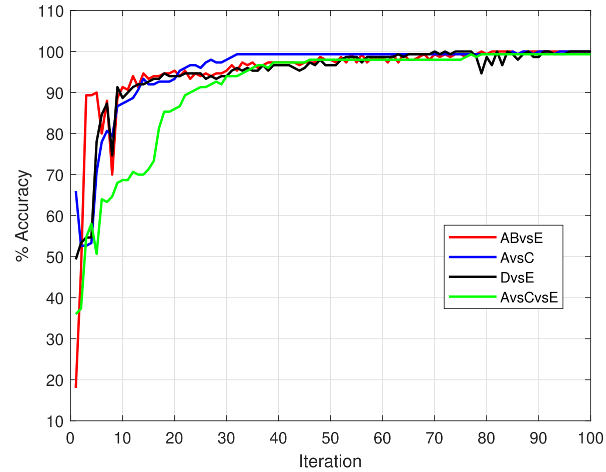


FIGURE 13. LSTM training iteration vs accuracy for each test cases of EEG signal classification for epilepsy prediction.

VII. CONCLUSION

This paper has contributed a novel LSTM architecture to classify the epileptic EEG signal. We modelled the EEG signal as Brownian movement of brain neurons and Hurst Exponent with ARMA features used to train the LSTM classifier. The novelty of this paper lies in the multi-class EEG signal detection. We can now detect epilepsy in EEG in its prior stage of occurrence i.e. pre-ictal. Our algorithm gives a 94% accurate detection in case of multi-class EEG detection for the data taken from Bonn University which has five data sets recorded at different stages of epilepsy from different sensor locations in the brain. The proposed approach has improved the binary classification accuracy by 2% from the previous SVM classifier. Each EEG signal is passed through a multi-rate DCT filter which divides it into five sub-bands of different bandwidths. Hurst and ARMA features are extracted for each sub-band which generate a total of 20-features for an EEG signal. These features, when classified by the proposed LSTM architecture, give an accuracy of 99.17% for pre-ictal vs ictal, 97.78% for pre-ictal vs inter-ictal, 97.78% for inter-ictal vs ictal and 94.81% for pre-ictal vs inter-ictal vs ictal is achieved.

REFERENCES

[1] P.-Y. Zhou and K. C. C. Chan, "Fuzzy feature extraction for multichannel EEG classification," *IEEE Trans. Cogn. Devel. Syst.*, vol. 10, no. 2, pp. 267–279, Jun. 2016.

[2] A. Gupta, P. Singh, and M. Karlekar, "A novel signal modeling approach for classification of seizure and seizure-free EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 5, pp. 925–935, May 2018.

[3] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan, "EEG-based neonatal seizure detection with support vector machines," *Clin. Neurophysiol.*, vol. 122, no. 3, pp. 464–473, Mar. 2011.

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

[5] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Comput. Biol. Med.*, vol. 99, pp. 24–37, Aug. 2018.

[6] H.-T. Shiao, V. Cherkassky, J. Lee, B. Veber, E. E. Patterson, B. H. Brinkmann, and G. A. Worrell, ''SVM-based system for prediction of epileptic seizures from iEEG signal,'' *IEEE Trans. Biomed. Eng.*, vol. 64, no. 5, pp. 1011–1022, May 2017.
[7] P. Fergus, A. Hussain, D. Hignett, D. Al-Jumeily, K. Abdel-Aziz, and H. Hamdan, ''A machine learning system for automated whole-brain seizure detection,'' *Appl. Comput. Inform.*, vol. 12, pp. 70–89, Jan. 2016.
[8] E. Alickovic, J. Kevric, and A. Subasi, ''Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction,'' *Biomed. Signal Process. Control*, vol. 39, pp. 94–102, Jan. 2018.
[9] M. Z. Parvez and M. Paul, ''Seizure prediction using undulated global and local features,'' *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 208–217, Jan. 2017.
[10] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, ''Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals,'' *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2017.
[11] D. Bhati, M. Sharma, R. B. Pachori, and V. M. Gadre, ''Time–frequency localized three-band biorthogonal wavelet filter bank using semidefinite relaxation and nonlinear least squares with epileptic seizure EEG signal classification,'' *Digit. Signal Process.*, vol. 62, pp. 259–273, Mar. 2017.
[12] K. Fujiwara, M. Miyajima, T. Yamakawa, E. Abe, Y. Suzuki, Y. Sawada, M. Kano, T. Maehara, K. Ohta, and T. Sasai-Sakuma, ''Epileptic seizure prediction based on multivariate statistical process control of heart rate variability features,'' *IEEE Trans. Biomed. Eng.*, vol. 63, no. 6, pp. 1321–1332, Jun. 2015.
[13] D. Flotzinger, G. Pfurtscheller, C. Neuper, J. Berger, and W. Mohl, ''Classification of non-averaged eeg data by learning vector quantisation and the influence of signal preprocessing,'' *Med. Biol. Eng. Comput.*, vol. 32, no. 5, pp. 571–576, 1994.
[14] B. S. Kim and S. K. Yoo, ''Motion artifact reduction in photoplethysmography using independent component analysis,'' *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 566–568, Mar. 2006.
[15] M. M. Baskaran and R. Bordawekar, ''Optimizing sparse matrix-vector multiplication on GPUs,'' Int. Bus. Mach., New York, NY, USA, IBM Res. Rep. RC24704, W0812-047, 2009.
[16] B. Direito, C. A. Teixeira, F. Sales, M. Castelo-Branco, and A. Dourado, ''A realistic seizure prediction study based on multiclass SVM,'' *Int. J. Neural Syst.*, vol. 27, no. 3, 2017, Art. no. 1750006.
[17] M. Z. Parvez and M. Paul, ''Prediction and detection of epileptic seizure by analysing EEG signals,'' Ph.D. dissertation, School Comput. Math., Charles Sturt Univ., Bathurst, NSW, Australia, 2015, pp. 35–36.
[18] Q. Zheng, F. Zhu, and P.-A. Heng, ''Robust support matrix machine for single trial EEG classification,'' *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 3, pp. 551–562, Mar. 2018.
[19] S. M. Usman, M. Usman, and S. Fong, ''Epileptic seizures prediction using machine learning methods,'' *Comput. Math. Methods Med.*, vol. 2017, Dec. 2017, Art. no. 9074759.
[20] P. W. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky, ''Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG,'' in *Proc. IEEE Workshop Mach. Learn. Signal Process.*, Oct. 2008, pp. 244–249.
[21] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, and B. Litt, ''Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: A report of four patients,'' *IEEE Trans. Biomed. Eng.*, vol. 50, no. 5, pp. 603–615, May 2003.
[22] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, ''Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state,'' *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 6, 2001, Art. no. 061907.
[23] I. Kiral-Kornek, S. Roy, E. Nurse, B. Mashford, P. Karoly, T. Carroll, D. Payne, S. Saha, S. Baldassano, and T. O'Brien, ''Epileptic seizure prediction using big data and deep learning: Toward a mobile system,'' *EBioMedicine*, vol. 27, pp. 103–111, Jan. 2018.
[24] M. Hasan, M. Ahamed, M. Ahmad, and M. Rashid, ''Prediction of epileptic seizure by analysing time series EEG signal using-κNN classifier,'' *Appl. Bionics Biomech.*, vol. 2017, Aug. 2017, Art. no. 6848014.
[25] K. Hoyos-Osorio, J. Castañeda-Gonzaiez, and G. Daza-Santacoloma, ''Automatic epileptic seizure prediction based on scalp EEG and ECG signals,'' in *Proc. XXI Symp. Signal Process., Images Artif. Vis. (STSIVA)*, 2016, pp. 1–7.
[26] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, S. Ippolito, and O. Kavehei, ''Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram,'' *Neural Netw.*, vol. 105, pp. 104–111, Sep. 2018.
[27] R. Sharma and R. B. Pachori, ''Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions,'' *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1106–1117, 2015.
[28] Q. Yuan, W. Zhou, L. Zhang, F. Zhang, F. Xu, Y. Leng, D. Wei, and M. Chen, ''Epileptic seizure detection based on imbalanced classification and wavelet packet transform,'' *Seizure*, vol. 50, pp. 99–108, Aug. 2017.
[29] Y. Yang, M. Zhou, Y. Niu, C. Li, R. Cao, B. Wang, P. Yan, Y. Ma, and J. Xiang, ''Epileptic seizure prediction based on permutation entropy,'' *Frontiers Comput. Neurosci.*, vol. 12, p. 55, Jul. 2018.
[30] J. Cheng, L. Dong, and M. Lapata, ''Long short-term memory-networks for machine reading,'' 2016, *arXiv:1601.06733*. [Online]. Available: https://arxiv.org/abs/1601.06733
[31] H. Sak, A. Senior, and F. Beaufays, ''Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition,'' 2014, *arXiv:1402.1128*. [Online]. Available: https://arxiv.org/abs/1402.1128
[32] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*. Boston, MA, USA: McGraw-Hill, 2000.
[33] A. Rubel, V. Lukin, M. Uss, B. Vozel, O. Pogrebnyak, and K. Egiazarian, ''Efficiency of texture image enhancement by DCT-based filtering,'' *Neurocomputing*, vol. 175, pp. 948–965, Jan. 2016.
[34] C. Yuan, S. Liu, and Z. Fang, ''Comparison of china's primary energy consumption forecasting by using arima (the autoregressive integrated moving average) model and gm (1, 1) model,'' *Energy*, vol. 100, pp. 384–390, Apr. 2016.

MUHAMMAD U. ABBASI (S'16–M'19) received the B.Sc. and M.Sc. degrees in electrical engineering from the CECOS University of IT and Emerging Sciences, Pakistan, in 2009 and 2014, respectively. He is currently pursuing the Ph.D. degree, on an exchange program, with the Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, U.K. He is also working as a Ph.D. Scholar with the Electrical Engineering Department, National University of Computer and Emerging Sciences (NUCES), Pakistan. He is also a Lab Engineer with the Department of Electrical Engineering, NUCES. Besides this, he has a passion to solve daily life problems with the use of technology. This provoked in becoming a Co-Founder and the Chief Technical Officer (CTO) of TrioTech Pvt., Ltd. TrioTech is specialized in solving Industrial problems with the help of technologies like the industrial Internet of Things, machine learning, big data, and artificial intelligence. He is also a Registered Engineer with the Pakistan Engineering Council. His research interests include de-noising, biomedical signal quality, and machine learning. He is a member of the Board of Studies and the CECOS University of IT and Emerging Sciences. He holds several international certifications in the Internet of Things, microcontroller interfacing, and embedded systems.

ANUM RASHAD received the bachelor's degree (Hons.) and the master's degree from FAST-NUCES, Pakistan, in 2010 and 2017, respectively. She is currently an Electrical Engineer by profession and also working as a Lab Engineer with the Electrical Engineering Department, FAST-National University of Computer and Emerging Sciences (NUCES), Pakistan. Her major area of research is power systems and detection of malicious data attacks on power systems. Her first research article was published in IEEE Asia Innovative Smart Grid Technologies (ISGT) Conference, in May 2018. She has also worked on digital systems design using field programmable gate array (FPGA). She holds the distinction of being nominated in dean's list of honours for her bachelor's degree.

ANAS BASALAMAH received the M.Sc. and Ph.D. degrees from Waseda University, Tokyo, in 2006 and 2009, respectively. He is currently an Associate Professor with the Computer Engineering Department, Umm Al Qura University, where he is also a Co-Founder and the Director of the Wadi Makkah Technology Innovation Center. He worked as a Postdoctoral Researcher with the University of Tokyo and the University of Minnestoa, in 2010 and 2011, respectively. He is also a Co-Founder of Hazen.ai and Averos. His areas of interests include embedded networked sensing, ubiquitous computing, participatory, and urban sensing.

MUHAMMAD TARIQ (S'08–M'12–SM'17) received the M.S. degree from Hanyang University, South Korea, the Ph.D. degree from Princeton University, under the supervision of Prof. H. V. Poor, in 2016, and the Ph.D. degree from Waseda University, Japan, in 2012. He is currently the Director of the FAST National University of Computer and Emerging Sciences (NUCES), Peshawar. Before taking charge as the Director, he was the Head of the Department of Electrical Engineering, NUCES. He is the author or coauthor of more than 50 research articles. He has presented his research work in various IEEE flagship conferences held around the world. He rendered his Technical Committees services in various IEEE flagship conferences and transactions. He has coauthored a book on *Smart Grids* (John Wiley and Sons, 2015) with leading researchers from Europe, China, Japan, and USA. In 2017, Chinese Government selected him as a High End Foreign Expert through the International Cooperation Project funded by the State Administration of Foreign Experts Affairs China. He received many awards for his work. He received the HEC Scholar for his M.S. degree and the Fulbright Scholar and the Japanese Government (MEXT) Scholar for his Ph.D. degree. He has delivered research talks as a Guest/Invited/Keynote Speaker at various forums and universities, Pakistan, China, Saudi Arabia, and USA. He is the programs' evaluator (PEV) of Pakistan Engineering Council. He is the Chair of the IEEE Peshawar Subsection and an Associate Editor of IEEE Access.