Application of Conv-1D and Bi-LSTM to Classify and Detect Epilepsy in EEG Data

Chetana R*, A Shubha Rao, Mahantesh K
Department of Electronics and Communication Engineering,
SJB Institute of Technology, Bangalore, Karnataka, India

***Abstract*—EEG is used to study the electrical changes in the brain and can derive a conclusion as epileptic or not, using an automated method for accurate detection of seizures. Deep learning, a technique ahead of machine learning tools, can self-discover related data for the detection and classification of EEG analysis. Our work focuses on deep neural network architecture to visualize the temporal dependencies in EEG signals. Algorithms and models based on Deep Learning techniques like Conv1D, Conv1D + LSTM, and Conv1D + Bi-LSTM for binary and multiclass classification. Convolution Neural Networks can spontaneously extract and learn features independently in the multichannel time-series EEG signals. Long Short-Term Memory (LSTM) network, with its selective memory retaining capability, Fully Connected (FC) layer, and softmax activation, discover hidden sparse features from EEG signals and predicts labels as output. Two independent LSTM networks combine to form Bi-LSTM in opposite directions and appreciate added visibility to upcoming information to provide efficient work contrary to previous methods. Long-term EEG recordings on the Bonn EEG database, Hauz Khas epileptic database, and Epileptic EEG signals from Spandana Hospital, Bangalore, assess performance. Metrics like precision, recall, f1-score, and support exhibit an improvement over traditional ML algorithms evaluated in the literature.**

Keywords—*1D CNN; bidirectional LSTM; dataset (DS); deep learning; electroencephalogram (EEG); LSTM*

I. Introduction

Epilepsy is a neural sickness portrayed by a sudden attack called seizures due to strange initiation by the networks of neurons [1]. The abrupt behavior of electrical movement causing disorder inside mind is due to abnormalities, lack of oxygen during labor, and reduction in blood sugar. A seizure is a time of irregular excitation of neurons lasting from seconds to a minute [2] and upsets the body. These seizures are not quickly perceived, which is a significant issue. Now researchers are exploring and assessing seizures in the beginning phase utilizing Electroencephalogram (EEG). The strange enactment is the voltage alteration due to the flow of current by the ions in the neurons, demonstrating the cerebrum's bioelectric phenomena [3] converted to electrical action and looked through electroencephalography (EEG). The recording is done to gauge the voltage motions in brain and changed to time series data called signals, characterized by spikes, sharp waves, or a combination of both. EEG signals are preferred in the frequency domain since they are convenient and give clarity [4]. Diagnosing epilepsy with EEG signals is tedious and arduous, and human mistakes are a possibility, so that a machine-based determination would be better.

Therefore data-preprocessing is done by normalizing the input variables. Features are extracted and selected from EEG signals in time, frequency, or in the time-frequency domain, like spectral, amplitude, entropy, wavelet, statistical, non-linear features, etc., and passed to the classification process. Because EEG patterns are exceptionally unique and may be unsuccessful for slight differences, time-series information is considered for dynamic examination since methodologies based on domain features have impediments.

Machine learning and deep learning strategies are predominant for learning, to prove the model with complex real-world information. We achieve crucial data collection by creating robust features [5], so the deep neural network can distinguish between seizure and non-seizure events. It concentrates on computational models and learns through non-linear transformations like neural networks. Initially, neural networks required more calculation time. Subsequently, they didn't get consideration, yet presently, due to enormous datasets and complex Graphic Processing Units (GPUs), it has given scientists an economical and robust arrangement, permitting them to examine deep learning models. Without prior knowledge of the dataset, neural networks have improved their boundaries repeatedly.

Work here demonstrates a one-dimensional Convolution Neural Network (1D-CNN) model to learn high-level representations from filtered EEG signal data for seizure detection and classification after reviewing the available research. However, increasing the convolutional layers can eventually obtain strong and conclusive features, with simplicity and efficiency being the most important advantages of this type of network. 1D-CNNs are naturally apt for handling biological signals like EEG for seizure detection [6] by using pooling and convolutional layers. In addition to that, signals are 1D in nature, and using preprocessing methods there is no information loss.

Next, the one-dimensional Convolution Neural Network Long Short-Term Memory (1D CNN-LSTM) model is proposed, with preprocessing applied to the raw EEG signal and normalized features effectively extracted by 1D-CNN. The obtained characteristics handled by LSTM layers extract temporal features and passed to fully connected layers before conclusion as epileptic or not. Results obtained demonstrate the proposed model exhibits identification recognition correctness in classifying epileptic seizure recognition tasks as binary and multiclass, respectively. The 1D CNN-LSTM model comprises one input layer, six convolutional layers, three pooling layers, two LSTM layers, one fully connected (FC) layer, and three dropout layers.

The modified version of the Recurrent Neural Network (RNN) is LSTM, and it is tough to train standard RNNs because of vanishing and exploding gradient problems [7]. The identity function of derivative 1 happens as the activation function, thus preventing the gradient from vanishing or exploding. The Bi-LSTM architecture selected consists of 64 forward and 64 backward LSTM cells per layer. Bidirectional long short-term memory (Bi-LSTM) network explores seizure detection and classification in this research. Bi-LSTM evolved considering the merits of LSTM and Bi-RNN [8]. Processing happens in two opposite directions, thereby improving performance. When compared with CNN models and Bi-LSTM models on time series data, the time dependencies of the signal are described poorly in CNN models but well in Bi-LSTM models.

II. Literature Survey

Numerous procedures are employed to obtain EEG signal features for seizure detection. In [9], integrating with extreme learning machine (ELM), features like approximate entropy and sample entropy are employed. In [10], non-subsampled wavelet–Fourier features are incorporated for seizure detection, with a considerable quantity of continuous EEG recordings being the limitation. Combining wavelet decomposition with directed transfer function (DTF) for feature extraction is used in [11]. Still, the limit here is the existence of muscle artefacts in scalp EEG recordings. However, better results can be expected if an intracranial electrocorticogram (ECoG) uses subdural grid electrode implementation. In [12] authors suggested a unique feature as a matrix determinant for EEG analysis. For noise removal, researchers proposed a Bandpass filter to enhance SNR in intracranial EEG signals to obtain a sensitivity more significant than 80% and specificity ranging between 75% and 88%. Correspondingly, sensitivity, specificity, and accuracy of 77.10%, 71.63%, and 75.07% are obtained [13], demonstrating weak execution. Linear Discriminant Analysis (LDA) [14] and Bayesian classifier [15] comprise the machine learning classifiers and Convolutional Neural Networks (CNN) [16] as deep learning classifiers with no perfect prediction available. Positive results are obtained using Long Short-Time Memory Units (LSTMs) [17]. However, the investigation by collecting additional experimental data and fusing it to develop new AI algorithms improves upon existing applications.

Robust features [18] with single-channel epileptic EEG signals automatically learn using machine learning and deep learning techniques. Research should focus on algorithms capable of handling complex multichannel epileptic EEG signals. Using discrete wavelet transform (DWT) and K-means with multilayer perceptron (MLP) for classification in [19] is implemented. Though deep CNN-based architecture obtains prominent features from raw EEG data to detect seizures, overlapping among seizure and non-seizure events happens. It becomes tedious to construct a generic technique to obtain high sensitivity [20]. In [21] a hybrid ensemble learning framework that systematically combines pre-processing methods with ensemble machine learning algorithms specifically, principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) combined along k-means clustering followed by ensemble learning such as extreme gradient boosting algorithms (XGBoost) and random forest is considered. However, in [22], using 13 layers, deep CNN architecture is considered.

Nevertheless, the drawback is the lack of a vast EEG database. Researchers demonstrated the deep belief nets (DBN) mechanism for modelling EEG data [23]. The training time using K Nearest Neighbor (KNN) and Support Vector Machine (SVM) took a few hours to a few days, but with DBNs, it took a few days to more than a week.

Using two parallel 1D-CNN blocks, a stacked 1D-CNN model is implemented with a random selection and data augmentation (RS-DA) strategy to overcome sample imbalance in [24] but with Two-Dimensional Convolution Neural Network (2D-CNN) and LSTM, collectively with RS-DA, thorough assessments with statistical, entropies, frequency, or time-frequency domain features, etc., can be derived and combined to 1D-CNN model as input. A generic auto-detection method, robust to noise, is used in [25]. Inputs are the digital version of the EEG recordings to the model, which aids the neurologists in detection. The limitation is the SNR value decreases the classification accuracy. A key reason for using Bi-directional LSTM in [26] is they look after the time dependencies both in a forward and backward direction. The authors in [27] use spectral feature-based two-layer long short-term memory (LSTM) model. The segments considered are in the frequency domain. In [28], an automated epilepsy detection system implementing wavelet decomposition and a 1D- CNN, along with Bi-LSTM, is incorporated. But the limitation is its inability to detect the occurrence of seizure at 512 Hz as the sample rate. However, the decimation of samples can enable the model at 256 or 512 Hz sampling rate to identify epileptic seizures.

III. Datasets

Before we begin the experiment with the results and analysis, discussion on the various datasets being used in the work is being dealt with.

A. *Bonn EEG – UCI Machine Learning Repository: Epileptic Seizure Recognition Dataset*

We expect to characterize the different classes of the Bonn EEG dataset into five categories named class1, class2, class3, class4 and class5, each having 100-single channel sections of EEG. Every single channel is 23.6s recording at a sampling frequency of 173.61 Hz. The comparing time series inspects 4097 data focused on separating and rearranging into 23 pieces, each containing 178 data of interest every second. The data of interest is the EEG recording at the alternate moment. The recording is of both healthy and epileptic patients. Class 1 contains EEG signals from epileptic seizure sections, and EEG signals originating from the tumor zone belong to Class 2. Class 3 has signals from the healthy brain area of the tumor found in the brain. Class 4 contains EEG information on healthy volunteers with closed eyes. EEG data of subjects with open eyes belong to Class 5 (Fig. 1).

The 178 information are X1, X2, X3… .X177, X178 the logical factors with various classes labeled y (Fig. 2). For 500 patients, we get 11500 columns (23X500= 11500). Each of the 178 pieces of information is put in sections as columns and 11500 examples as lines or rows and named the information from [1-5] as the last segment (segment y). People other than one category, i.e., 2,3,4,5, classes are non-epileptic.

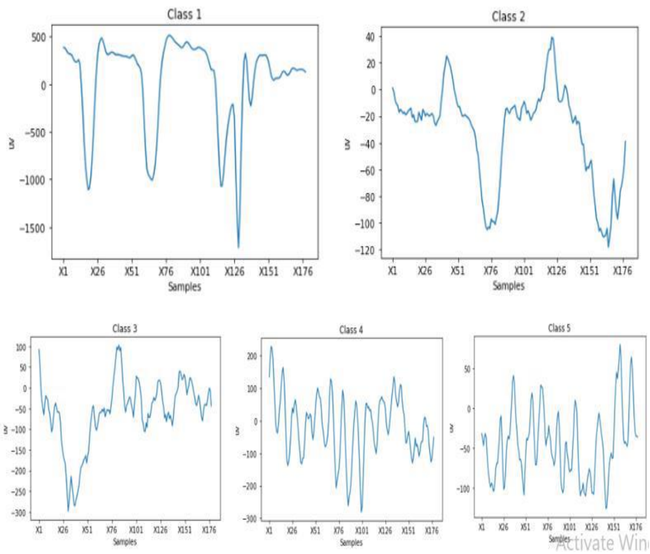

Fig. 1. Plot of each class.

	Unnamed: 0	X1	X2	X3	X4	X5	X6	X7	X8	X9	...	X170	X171	X172	X173	X174	X175	X176	X177	X178	y
0	X21.V1.791	135	190	229	223	192	125	55	-9	-33	...	-17	-15	-31	-77	-103	-127	-116	-83	-51	4
1	X15.V1.924	386	382	356	331	320	315	307	272	244	...	164	150	146	152	157	156	154	143	129	1
2	X8.V1.1	-32	-39	-47	-37	-32	-36	-57	-73	-85	...	57	64	48	19	-12	-30	-35	-35	-36	5

3 rows × 180 columns

Fig. 2. Datapoints with label.

B. *Neurology and Sleep Centre, Hauz Khas, New Delhi*

The EEG recording uses the Comet AS40 EEG machine and 200 Hz as the sampling rate. Signals ranging between 0.5 to 70 Hz undergo filtering and are divided into pre-ictal, interictal and ictal stages. The duration of the EEG portion in this archive is 5.12s with 1024 examples.

There are three folders that are named according to the epileptic seizure stages and each folder contains fifty files of EEG time series. Each segment is considered as an instance. In total, 150 instances are considered belonging to each intended class.

C. *Spandana Nursing Home Dataset, Bangalore*

An ongoing EEG Data is being utilized for the location of epilepsy. Twenty EEG accounts of epileptic patients with 10 EEG signals during seizures and 10 EEG data from a sound volunteer with open eyes are considered. The sampling rate is 175 Hz. Filtering for signals ranging between 0.5 to 60 Hz is done. The universally perceived technique to portray areas of various electrodes on the scalp is utilized, which depends on the connection between the primary regions of the cerebral cortex. The number '10' - '20' suggests the distance between every terminal from each other is 10% or 20% of the absolute right-left or front-back space of the skull. A letter has been assigned to each site for lobe recognition, and a number is assigned to distinguish the cerebral hemisphere area. Even numbers 2, 4, 6, and 8 indicate the right of the brain for the electrode position of the brain, and odd numbers 1, 3, 5, and 7 mean terminals on the left part. The crude signals acquired are switched over entirely to ASCII design. EDF (European Data Format) Browser programming is utilized, an open source, universal viewer, multiplatform, and tool kit for conversion. The classifier tool is in such a way if the output is '0', it is epileptic or abnormal EEG, and '1' indicates normal EEG or non-epileptic EEG.

IV. Proposed Methodology: Deep Learning Algorithms and Models

The proposed technique introduces three different methodologies for adequate recognition and classification of an epileptic seizure. EEG as input is a Comma Separated Values (CSV) document. When the input document is perused and switched over entirely to a python data frame, the information is standardized, split for training, validation, and testing in the proportion of 6:2:2 and labels are changed over into One Hot Encoded design. The architecture's performance is analyzed for Multiclass (1,2,3,4,5) and Binary Classification (1/0). All the presented models are examined by training for up to 40 epochs using Categorical Cross Entropy as a loss function.

A. *Method-1: Based on Conv1D*

In the proposed method 1, the EEG information is examined by applying Convolution, and a deep learning method is prominently used to analyze time series data. Direct and quicker design models are presented because the boundaries are low. Pooling and convolutional layers with bigger size are utilized in 1D models and, when applied, produces a kernel of determined size(m) which is convolved with the input(x) to create the filtered output(y) whose dimensionality will be equivalent to the number of kernels(n). Conv1D is fit for learning features (w) concealed in the series of time sequence data.

$$y_i = \sum_{-m}^{m} x_{i-k} w_k \tag{1}$$

In the output expressed as the above equation, k is the counter value ranging from -m to +m, covering the length of the kernel. Initially, considering 1D data of the EEG signal with the feature vector, it is convolved along with the filter to acquire a feature map. 1D data is ordered along a single line data organized by time and fits on a 1D line. Convolution takes a kernel (internal weights) of a filter and a sliding dot product with the signal. The process of multiplying each aligned pair of points and adding all products is called the dot product.

Since we are sliding, the data gets overlapped, and the representation is as below.

$$x = \{x_0, x_1, x_2, \dots \dots \dots x_{m-1}\} \tag{2}$$

$$w=\{w_0, w_1, w_2, \dots \dots \dots w_{n-1}\} \quad (3)$$

$$y = \{y_0, y_1, y_2, \dots \dots \dots y_{m-1}\} \quad (4)$$

Losses are overcome by backpropagation, and the above equations are explicitly differentiated concerning gradients through layers. The partial derivative of loss for y is propagated back to calculate the partial derivative of loss for x through every network using the chain rule and the loss to each input given by.

$$\frac{\partial L}{\partial x_i} = \sum_{j=0}^{m-1} \frac{\partial L}{\partial y_j} \frac{\delta y_j}{\delta x_i} \quad (5)$$

Therefore, input gradient = output gradient (W), where $W=\frac{\delta y_j}{\delta x_i}$ should be known and so the layer is differentiable. The architecture consists of a series of Conv1D layers followed by MaxPool layer. It reduces spatial size, number of parameters and computation while aggregating the dominant features, thereby reducing the dimensionality. The most well-known pooling strategy is max pooling. Max pooling alludes to getting maximum value after each pooling activity and the data is flattened. Flatten concatenates the results from the convolution layers to frame a flat structure taken as input to dense layer. A fully connected layer or Dense network helps to classify based on features. It is a dense network of neurons, and every neuron is connected to the previous and subsequent layers. If there are multiple dense layers, then the last layer has output as the same number of the classes or categories. The linearity principle is used in Dense Layer, where the outcome depends on every input. The activation function utilized is the SoftMax activation which adds learning capacity to neural networks by learning complex patterns and multiplying weights with the input features and concluding regarding firing. Activation functions make the network non-linear, else it becomes linear. For example, output relies linearly upon the input features. SoftMax activation is the most ordinarily busy work as final layer in neural network for multiclass classification, being a blend of different sigmoid which works out the general probabilities and standardizes neural network results to fit between 0 and 1. The SoftMax probabilities will constantly aggregate to 1. The architecture of the proposed Method-1 can be seen in Fig. 3. The results obtained using Method-1 to determine the metrics like Accuracy, Precision, Recall, F1- Score and Support with DS1, DS2 and DS3 for multiclass and binary classification are shown in Table I (A) and Table II (B) respectively.

$$S(Y)_i = \frac{e^{Y_i}}{\sum_{j=1}^{n} e^{Y_j}} \quad (6)$$

B. *Method-2: Based on Conv1D+LSTM*

The architecture with a mix of Convolution (Conv1D) and Long Short-Term Memory (LSTM) is proposed as method 2. Input is passed to Convolution, MaxPool, and dropout layer before executing with the next layer, where it invalidates a portion of the neurons towards the following layer by randomly setting the input units to 0, thus forestalling overfitting and consequently evades the network from depending on a single neuron. Typically, dropouts are put on fully connected layers. Dropout might be carried out on any hidden layer or input layer in the network, yet not utilized on the output layer.

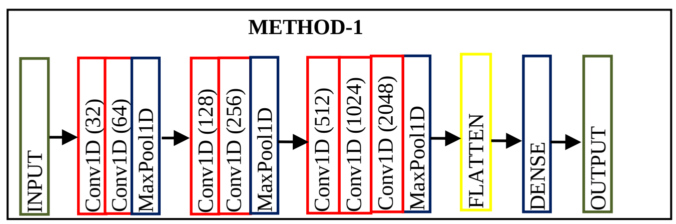

Fig. 3. Architecture of the proposed method-1.

TABLE I. (A) PROPOSED METHOD1 WITH METRICS FOR MULTICLASS CLASSIFICATION OF DATASET1 AND DATASET2

	Conv1D					
	Multiclass Classification	**Precision**	**Recall**	**F1-score**	**Support**	**Accuracy**
DS-1 Bonn EEG	Class1 (Epileptic)	0.98	0.97	0.97	457	**74.61**
	Class 2	0.65	0.66	0.66	477	
	Class 3	0.66	0.61	0.63	472	
	Class 4	0.72	0.79	0.76	422	
	Class 5	0.74	0.71	0.72	475	
DS -2 HauzKhas	0 (Ictal)	1.00	1.00	1.00	10	**64.52**
	1(Inter ictal)	0.80	0.29	0.42	14	
	2(preictal)	0.38	0.86	0.52	7	

TABLE II. (B) PROPOSED METHOD1 WITH METRICS FOR BINARY CLASSIFICATION OF DATASET1, DATASET2 AND DATASET3

	Conv1D					
	Binary Classification	**Precision**	**Recall**	**F1-score**	**Support**	**Accuracy**
DS-1 Bonn EEG	0(Nonepileptic)	0.99	0.99	0.99	1846	98.83
	1 (Epileptic)	0.98	0.96	0.97	454	
DS -2 Hauz Khas	0(Nonepileptic)	0.86	0.71	0.77	17	77.42
	1 (Epileptic)	0.71	0.86	0.77	14	
DS -3 Spandana	0(Nonepileptic)	1.00	0.67	0.80	3	75.00
	1 (Epileptic)	0.50	1.00	0.67	1	

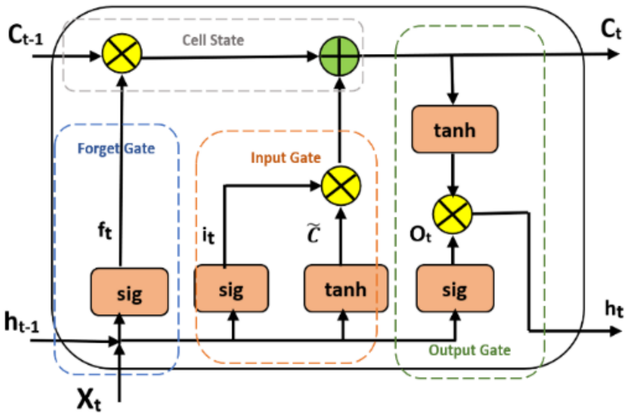

Fig. 4. Cell of LSTM.

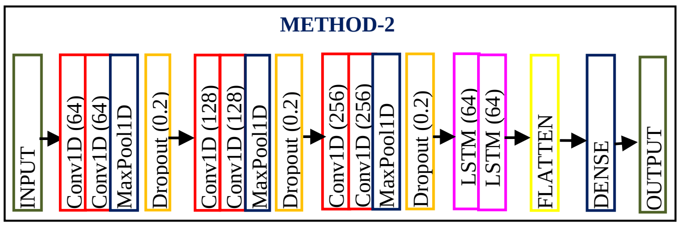

Fig. 5. Architecture of the proposed method-2.

LSTM is based on recurrent neural networks capable of learning, remembering, and processing information from a series of data distributed over time and, accordingly, is slow. LSTMs, with their specially designed gates, process the data linearly while deciding against retaining the learnt feature is good or forgetting it and moving forward. The LSTM gates use sigmoid activation (σ) as shown in Fig. 4. The architecture of the proposed method 2 is seen in Fig. 5. It has been observed in the conducted research that using many filters results in hindering the model from learning. Long Short-Term Memory (LSTM) networks are fit for learning to rely on the sequence and handle the disappearing gradient issue.

Sigmoid and tanh functions are two normalizing conditions utilized in LSTM. The sigmoid function implies a mechanism attempting to compute a bunch of scalars in the range of 0 and 1. The tanh function tells a system trying to change the information into a standardized data encoding between - 1 and 1. Inputs are multiplied by different frameworks of weights and added together. Feature extraction is done when the sigmoid function crushes the outcome between 0 and 1 when added with bias and applied. Though training is lengthy, it glances at the long sequence of inputs without expanding the network size. An LSTM network empowers to include sequence information in the network and makes forecasts relying on individual time stamps.

The LSTM cell is shown in Fig. 4. To replace memory, the Input gate finds the value. The second sigmoid function accepts current state x_t and previously hidden state h_{t-1} and concludes values to let through as 0 (critical) or 1(not critical). Furthermore, the tanh function gives weightage to the qualities which are passed to create a vector $\tilde{C}_t$ concluding their degree of significance from - 1 to 1.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{7}$$

$$\tilde{C}_t = tanh(W_c\,[h_{t-1}, x_t] + b_C) \tag{8}$$

where t = timestamp, i_t = input gate at t, W_i = Weight matrix of sigmoid operator between input gate and output gate, b_i = bias vector, $\tilde{C}_t$ = value generated by tanh, W_c = weight matrix of tanh operator between cell state information and network output, b_C = bias vector concerning W_c.

Based on the block's input and memory, the output gate result is chosen, and current and previous hidden state values are passed to the third sigmoid. The function tanh accepts new cell state generated, and outputs are multiplied point-by-point. The final value decides the hidden state to carry the information. Therefore, a new cell state and a new hidden state are passed to the next timestamp.

$$o_t = \sigma\left(W_0\,\left[h_{t-1,} x_t\right] + b_0\right) \tag{9}$$

$$h_t = o_t * tanh(C_t) \tag{10}$$

where t = timestamp, o_t = output gate at t, W_0 = Weight matrix of output gate, b_0= bias vector with respect to W_0 , h_t= LSTM output.

Related data from the earlier process is found by forget gate. The sigmoid function is passed with the current input x_t and hidden state $h_{t-1,}$ and value derived is implemented for point-by-point multiplication.

$$f_t = \sigma\left(W_f\,\left[h_{t-1,} x_t\right] + b_0\right) \tag{11}$$

where t = timestamp, f_t = forget gate at t, xt = input, h_{t-1} = previous hidden state, W_f = weight matrix between forget gate and output gate, b_0= bias at t.

The data needs to be stored from the new state in the cell to obtain the end output. The product of previous cell state C_{t-1} and forget vector f_t, if found to be 0, then values are eliminated, and point-by-point addition is performed to get a new cell state C_t.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{12}$$

where t =timestamp, C_t is cell state information, $f_t$is forget gate at t, C_{t-1} is previous time stamp, i_t is the input gate, $\tilde{C}_t$ is a value generated by tanh.

The boundaries in LSTMs are learning rates, information, and result predispositions. In forget gate, a duplicate of the time-stamp information is separated, and in input gate a copy is passed. Using the above method, various metrics like Accuracy, Precision, Recall, F1- Score and Support for various datasets DS1, DS2 and DS3 are calculated and demonstrated in Table III (A) and Table IV (B) for multiclass and binary classification.

TABLE III. (A) PROPOSED METHOD 2 WITH METRICS FOR MULTICLASS CLASSIFICATION OF DATASET1 AND DATASET2

	Conv1D+LSTM					
	Multiclass Classification	**Precision**	**Recall**	**F1-score**	**Support**	**Accuracy**
DS-1 Bonn EEG	Class1 (Epileptic)	0.97	0.98	0.97	454	77.52
	Class 2	0.72	0.59	0.65	477	
	Class 3	0.66	0.71	0.68	472	
	Class 4	0.82	0.76	0.79	422	
	Class 5	0.76	0.78	0.77	475	
DS -2 Hauz Khas	0 (Ictal)	1.00	1.00	1.00	10	74.19
	1(Inter ictal)	0.67	0.57	0.62	14	
	2(preictal)	0.33	0.29	0.31	7	

TABLE IV. (B) PROPOSED METHOD 2 WITH METRICS FOR BINARY CLASSIFICATION OF DATASET1, DATASET2 AND DATASET3

	Conv1D+LSTM					
	Binary Classification	**Precision**	**Recall**	**F1-score**	**Support**	**Accuracy**
DS-1 Bonn EEG	0(Nonepileptic)	0.99	1.00	0.99	1846	99.40
	1 (Epileptic)	0.98	0.97	0.97	454	
DS -2 Hauz Khas	0(Nonepileptic)	0.82	0.82	0.82	17	80.65
	1 (Epileptic)	0.79	0.79	0.79	14	
DS -3 Spandana	0(Nonepileptic)	1.00	1.00	1.00	3	1.00
	1 (Epileptic)	1.00	1.00	1.00	1	

TABLE V. (A) PROPOSED METHOD 3 WITH METRICS FOR MULTICLASS CLASSIFICATION OF DATASET1 AND DATASET2

	Conv1D+Bi-LSTM					
	Multiclass Classification	**Precision**	**Recall**	**F1-score**	**Support**	**Accuracy**
DS-1 Bonn EEG	Class1 (Epileptic)	0.97	0.98	0.97	454	**80.43**
	Class 2	0.70	0.72	0.71	477	
	Class 3	0.72	0.62	0.67	472	
	Class 4	0.74	0.83	0.78	422	
	Class 5	0.78	0.72	0.75	475	
DS -2 Hauz Khas	0 (Ictal)	1.00	1.00	1.00	10	**77.42**
	1(Inter ictal)	0.73	0.79	0.76	14	
	2(preictal)	0.50	0.43	0.46	7	

TABLE VI. (B) PROPOSED METHOD 3 WITH METRICS FOR BINARY CLASSIFICATION OF DATASET1, DATASET2 AND DATASET3

	Conv1D+Bi-LSTM					
	Binary Classification	**Precision**	**Recall**	**F1-score**	**Support**	**Accuracy**
DS-1 Bonn EEG	0(Nonepileptic)	0.99	0.99	0.99	1846	**99.40**
	1 (Epileptic)	0.98	0.96	0.97	454	
DS -2 Hauz Khas	0(Nonepileptic)	0.92	0.71	0.80	17	**80.65**
	1 (Epileptic)	0.72	0.93	0.81	14	
DS -3 Spandana	0(Nonepileptic)	1.00	1.00	1.00	3	**1.00**
	1 (Epileptic)	1.00	1.00	1.00	1	

C. Method-3: Based on Conv1D+BiLSTM

In the proposed method 3, a more meaningful output is produced by using a powerful tool for modeling the sequential dependencies in both directions. The architecture is planned with a blend of Convolution and Bidirectional Long Short-Term Memory (Bi-LSTM). It offers preferable expectations by two LSTMs. Every component of an input sequence computes the input arrangement from the reverse path to a hidden forward sequence and a backward hidden sequence. Concatenation of the final forward and backward outputs leads to an encoded vector. Thirty-two units of LSTM of 0.2 dropouts, are utilized in a bidirectional manner as depicted in Fig. 6. At every timestamp, each hidden layer yield is created alongside the memory cell state and passed to a 1D convolutional layer of 64 filters of kernel size four as shown in Fig. 7. The past LSTM network trails the remainder of the network. The results show that Bi-LSTM based modeling offers better predictions than regular LSTM based models. Accuracy, Precision, Recall, F1 Score and Support for datasets DS1, DS2 and DS3 for both multiclass and binary classification using Conv1D+ Bi LSTM are referred in Table V (A) and Table VI (B) accordingly.

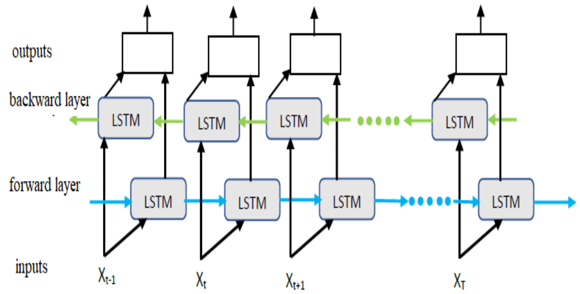

Fig. 6. Bidirectional LSTM.

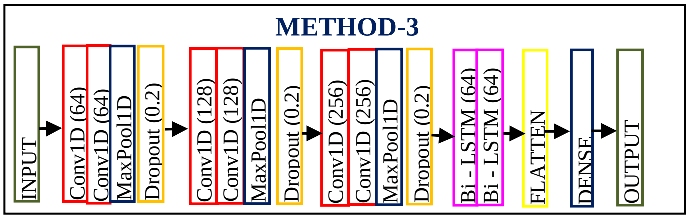

Fig. 7. Architecture of the proposed method-3.

V. Result and Discussion

Precision is defined as the quality of a correct prediction given by the model and is the number of true positives divided by the total positive predictions. Precision is how good the model is at predicting a specific category. It does not predict negative class, called false negatives [30-32].

$$\text{Precision} = TP / (TP + FP)$$

Recall is a measure computing the number of correct predictions from all positive predictions possible. In binary class, recall is computed as the number of true positives divided by the sum of true positives and false negatives.

$$\text{Recall} = TP / (TP + FN)$$

The F1 score or F-measure gives the harmonic average of precision and recall together to measure the efficiency of two classifiers.

$$\text{F-Measure} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Support: The support is several actual occurrences of the class in the specified dataset and got by summing the rows of the confusion matrix.

However, comparing three models with Bonn EEG Dataset, displayed in Table VII illustrates the proposed Deep learning methods with added layers have a higher score than simple CNN approaches, suggesting high classification accuracy.

TABLE VII. Comparison of Accuracy with Existing and Proposed Methods of Dataset1(Bonn), Dataset2(Hauz Khas) and Dataset3(Spandana) with Binary Classification

DL Algorithm and Models	Accuracy (%)	Proposed	Sensitivity	Proposed	Precision	Proposed	F1- Score	Proposed
Dataset1(Bonn)								
Conv1D [29]	88.70	98.83	95.00	98	90.00	96	--	97
Conv1D+ LSTM	---	99.04	---	97	---	98	--	97
Conv1D+ BiLSTM	---	99.40	---	96.5	---	98	---	97
Dataset2 (Hauz Khas)								
Conv1D [29]	---	**77.42**	---	**86**	---	**86**	--	**77**
Conv1D+ LSTM	---	**80.65**	---	**82**	---	**82**	--	**83**
Conv1D+ BiLSTM	---	**80.65**	---	**93**	---	**92**	---	**80**
Dataset3 (Spandana)								
Conv1D [29]	---	**75.00**	---	**100**	---	**100**	--	**80**
Conv1D+ LSTM	---	**100.00**	---	**100**	---	**100**	--	**100**
Conv1D+ BiLSTM	---	**100.00**	---	**100**	---	**100**	---	**100**

VI. Conclusion and Future Work

In Proposed Method 1 (Conv 1D), a CNN model is built on a 1D time series, and architecture consists of a series of Conv1D layers followed by MaxPool. Flatten forms a flat structure which acts as input to dense layer with SoftMax activation adding learning capacity to neural networks for classification. A fully connected layer or Dense network helps to classify based on features. The metrics are improved compared to ML algorithms. A dense or Fully Connected layer is used as a classifier based on extracted features. Generally, the performance of the CNN classifier can be improved by the right choice of parameters like pooling size, learning rate, activation function and optimizer. Our approach uses CNN to detect epileptic seizures and has improved the classification accuracy along with the generalization ability of the classifier. The tabulated results from Table I (A) and Table II (B) significantly shows the improvement compared to Machine Learning based algorithms.

In Proposed Method 2 (Conv1D + LSTM), the LSTM is based on recurrent neural networks capable of learning, remembering, and processing information from a time series data. LSTMs have gates that process the data and decide on retaining the known feature if it is sound, forgetting if imperfect, and moving ahead. Gates use sigmoid activation (σ) and are fit for learning order. Though the training time is lengthy, LSTM glances at a long sequence of inputs without expanding the network size. An LSTM network empowers to include sequence information in succession. It is evident from the measured metrics shown in Table III (A) and Table IV (B) that LSTM, with its selective memory, can successfully empower the network with the retained essential features. This model can also be implemented on different domain signals like frequency and time- frequency domain signals and can compare the performance accuracy. Furthermore, LSTM layers can implement the data in classifying with multiclass exclusively on the Bonn EEG epileptic dataset deeply and classifying better seizure states.

In Proposed Method 3 (Conv1D + Bi LSTM), the Bi-LSTM based model is much better than usual LSTM based models as per the results obtained. Bi- LSTM takes a step further ahead from LSTMs with its capability to view and understand the data in both directions. In contrast, they are utilizing the knowledge of the past and future data present in the time series. Bi-LSTMs can extract the best describing feature vectors from the data. While the proposed methods prove their ability with the Bonn and Hauz Khas dataset, it leads to overfitting with Spandana dataset due to its smaller size. As part of further work, improvement is made by enhancing the dataset from medical agencies, building deeper models, regularization with Batch normalization, augmentation techniques, and reinforcement-based learning remains unexplored. Moreover, when operated to multi-class classification, the present approach does not have a good recognition accuracy that is principally exceptional. Results prove that Bi-LSTMs are an ideal choice for time sequence data as demonstrated in Table V (A) and Table VI (B) correspondingly.

REFERENCES

[1] V. Gabeff et al., “Interpreting deep learning models for epileptic seizure detection on EEG signals,” Artificial Intelligence in Medicine, vol. 117, no. 1, pp. 102084, 2021.

[2] A. A. Hameed, and M. Bayoumi, “A Deep Learning Approach for Automatic Seizure Detection in Children with Epilepsy,” Frontiers in Computational Neuroscience, vol. 15, no. 1, pp. 650050, 2021.

[3] E. Niedermeyer and F. L. da Silva, “Electroencephalography: Basic Principles, Clinical Applications, and Related Fields,” Oxford University Press, 2017.

[4] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, “Automated EEG-based screening of depression using deep convolutional neural network,” Computer Methods and Programs in Biomedicine, vol. 161, no. 1, pp. 103–113, 2018.

[5] J. Birjandtalab, M. Heydarzadeh, and M. Nourani, “Automated EEG-Based Epileptic Seizure Detection Using Deep Neural Networks,” in proceedings of the IEEE International Conference on Healthcare Informatics (ICHI), pp. 552-555, 2017.

[6] A. Shoeibi et al., “A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in EEG signals,” Expert Systems with Applications, vol. 163, no. 1, pp. 113788, 2021.

[7] R. Pascanu, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks,” in proceedings of the International Conference on Machine Learning, vol. 28, no. 2, pp. 1310–1318, 2013.

[8] X. Hu, S. Yuan, F. Xu, Y. Leng, K. Yuan, and Q. Yuan, “Scalp EEG classification using deep Bi-LSTM network for seizure detection,” Computers in Biology and Medicine, vol. 124, no.1, pp. 103919, 2020.

[9] Y. Song, J. Crowcroft, and J. Zhang, “Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine,” Journal of Neuroscience Methods, vol. 201, no. 2, pp. 132-146, 2012.

[10] G. Chen, W. Xie, T. D. Bui, and A. Krzyżak, “Automatic epileptic seizure detection in EEG using nonsubsampled wavelet–Fourier features,” Journal of Medical and Biological Engineering, vol. 37, no.1, pp. 123-131, 2017.

[11] D. Wang et al., “Epileptic seizure detection in long-term EEG recordings by using the wavelet-based directed transfer function,” IEEE Transactions on Biomedical Engineering, vol. 65, no. 11, pp. 2591–2599, 2018.

[12] S. Raghu, N. Sriraam, A. S. Hegde, and P. L. Kubben, “A novel approach for classification of epileptic seizures using matrix determinant,” Expert Systems with Applications, vol. 127, no. 1, pp. 323–341, 2019.

[13] Y. Song, and J. Zhang, “Discriminating preictal and interictal brain states in intracranial EEG by sample entropy and extreme learning machine,” Journal of Neuroscience Methods, vol. 257, no.1, pp. 45-54, 2016.

[14] T. N. Alotaiby, S. A. Alshebeili, F. M. Alotaibi, and S. R. Alrshoud, “Epileptic seizure prediction using CSP and LDA for scalp EEG signals,” Computational Intelligence and Neuroscience, vol. 2017, no. 6, pp. 1-11, 2017.

[15] M. Behnam, and H. Pourghassem, “Real-time seizure prediction using RLS filtering and interpolated histogram feature based on hybrid optimization algorithm of Bayesian classifier and Hunting search,” Computer Methods and Programs in Biomedicine, vol. 132, no.1, pp. 115-136, 2016.

[16] N. D. Truong et al., “Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram,” Neural Networks, vol. 105, no.1, pp. 104–111, 2018.

[17] X. Wei, L. Zhou, Z. Zhang, Z. Chen, and Y. Zhou, “Early prediction of epileptic seizures using a long-term recurrent convolutional network,” Journal of Neuroscience Methods, vol. 327, no.1, pp. 108395, 2019.

[18] X. Wang, Y. Zhao, and F. Pourpanah, “Recent advances in deep learning,” International Journal of Machine Learning and Cybernetics, vol. 11, no. 1, pp. 747–750, 2020.

[19] U. Orhan, M. Hekim, and M. Ozer, “EEG signals classification using the K-means clustering and a multilayer perceptron neural network model,” Expert Systems with Applications, vol. 38, no.10, pp. 13475-13481, 2011.

[20] M. S. Hossain, S. U. Amin, M. Alsulaiman, and G. Muhammad, “Applying Deep Learning for Epilepsy Seizure Detection and Brain Mapping Visualization,” ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 15, no. 1s, pp. 1–17, 2017.

[21] C. Rachappa, M. Kapanaiah, and V. Nagaraju, “Hybrid ensemble learning framework for epileptic seizure detection using electroencephalograph signals,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 28, no. 3, pp. 1502-1509, 2022.

[22] 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,” Computers in Biology and medicine, vol. 100, no. 1, pp. 270–278, 2018.

[23] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, “Modelling electroencephalography waveforms with semi-supervised deep belief nets: Fast classification and anomaly measurement”, in Journal of Neural Engineering, vol. 8, no. 3, 2011.

[24] X. Wang, X. Wang, W. Liu, Z. Chang, T. Kärkkäinen, and F. Cong, “One dimensional convolutional neural networks for seizure onset detection using long-term scalp and intracranial EEG,” Neurocomputing, vol.12, no. 1, pp. 212-222, 2021.

[25] D. K. Thara, B. G. Premasudha, R. S. Nayak, T. V. Murthy, G. A. Prabhu and N. Hannon, “Electroencephalogram for epileptic seizure detection using stacked bidirectional LSTM_GAP neural network,” Evolutionary Intelligence, vol. 14, no. 2, pp. 823–833, 2021.

[26] H. Ali, F. Karim, J. J. Qureshi, A. O. Abu-assba, and M. F. Bulbul, “Seizure Prediction using Bidirectional LSTM,” Cyberspace Data and in proceedings of the International conference on Intelligence, and Cyber-Living, Syndrome, and Health, pp 349-356, 2019.

[27] K. Singh, and J. Malhotra, “Two-layer LSTM network-based prediction of epileptic seizures using EEG spectral features,” Complex and Intelligent Systems, vol. 8, no. 3, pp. 2405–2418, 2022.

[28] M. K. Alharthi, K. M. Moria, D. M. Alghazzawi, and H. O. Tayeb, “Epileptic Disorder Detection of Seizures Using EEG Signals,” Sensors, vol. 22, no. 17, pp. 6592, 2022.

[29] M. Woodbridge, B. Verma, and A. Haidar, “Autonomous deep feature extraction based method for epileptic EEG brain seizure classification,” Neurocomputing, vol. 444, no. 1, pp. 30-37, 2021.

[30] T. Rajendran, K. P. Sridhar, P. Vidhupriya, N. Gayathri, and T. Anitha, “Epileptic seizure: Classification using autoregression features,” International Journal of Current Research and Review, vol. 13, no. 4, pp. 121–131, 2021.

[31] T. Rajendran, and K. P. Sridhar, “An overview of EEG seizure detection units and identifying their complexity-a review,” Current Signal Transduction Therapy, vol. 15, no. 3, pp. 234–242, 2020.

[32] D. Yuvaraj, A. M. U. Ahamed, and M. Sivaram, “An Investigation into the Application of NLP in the Healthcare Sector,” Journal of Computational Science and Intelligent Technologies, vol. 3, no. 3, pp. 50-58, 2022.