CNN based framework for detection of epileptic seizures

Mustafa Sameer[1] · Bharat Gupta[1]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Epilepsy is a common neurological disease that uses electroencephalogram (EEG) data for its detection purpose. Neurologists make the diagnosis by visual inspection of EEG reports. As it is time-consuming and due to the shortage of specialists worldwide, researchers have proposed automated systems to detect the disease. In the past decade, most of the systems were designed using hand-engineered features. However, identifying appropriate features is always a challenging task in the development of a seizure detector system. Deep learning networks eliminate the problem of selecting the best features but suffer from long training time, generally days or weeks. To overcome this problem, the authors have proposed a new 1D convolutional neural network (CNN) that automatically extracts features at an average of seven epochs, only followed by traditional machine learning (ML) classifier. 1D CNN architectures are intrinsically suitable for the processing of EEG time-series data. The proposed model doesn't require any preprocessing of EEG signal and results in approximately 94% reduced training time than end-to-end deep learning models. Different ML techniques have been applied to extracted features to check the robustness of the proposed 1D CNN. Maximum accuracy of 99.83% has been achieved by most of the classifiers to detect between healthy and seizure patients. The reduced number of processing steps and epochs makes it suitable for real-time clinical applications.

✉ Mustafa Sameer
   mustafa.ec17@nitp.ac.in

   Bharat Gupta
   bharat@nitp.ac.in

1 Department of Electronics and Communication Engineering, National Institute of Technology Patna, Ashok Rajpath, 800005 Patna, India

1 Introduction

Epileptic seizures are one of the most common brain disorders, and according to World Health Organization (WHO) reports, approximately 50 million people worldwide are suffering from it. Mainly middle income and low-income countries people are affected by this disease. Seizures results due to excessive discharge of neurons in the brain. It can result in a sudden fall if a person is standing or an accident if a person is driving. Individuals detected with epilepsy have a higher death rate as compared to a healthy person.

Generally, an EEG machine is used to capture the electrical activity of the brain for diagnosis purposes. The significant advantage of EEG is its high temporal resolution, noninvasiveness in nature, and low cost. EEG is obtained by placing electrodes on the human scalp. The widely recognized 10–20 method is used for this work. Highly skilled neurologists are required to analyze EEG readings; still, the wrong prediction is probable. So, an automated system to assist clinicians is high in demand. Observing the dreadful situation of this disease, researchers have done a lot of work in the last decade.

1.1 Related work

Researchers have proposed many hand-engineered feature extraction techniques for epileptic seizure detection systems in the last twenty years. Time analysis [13], frequency analysis [41], and time-frequency(t-f) analysis [32] have been used to extract features followed by machine learning classifier. Discrete wavelet transform (DWT) is one of the most widely used techniques for this purpose [25]. Different variants of entropy [26], hurst exponent [24], Djorth parameters, statistical averages, average frequency, relative spike amplitude are the most important features calculated in previous works. Support vector machine(SVM) [28], K-Nearest Neighbour (KNN), Naïve Bayes, Random Forest [40], artificial neural network (ANN) [38] are some of the most important classifiers used for the detection purpose. The major disadvantage of these automated systems is they require a specialist to select the optimum set of features.

After the advent of deep learning, various groups have started working on it. Deep learning has the advantage of automatic feature selection. Ulah et al. [39] presented a pyramidal structure of CNN and used a large number of epochs to reach acceptable accuracy. As well as data augmentation has also been done in their work to increase the number of instances which is an additional stage in the automated system. In [23], authors have used a combination of 1D and 2D CNN network followed by autoencoder for detection in mobile multimedia framework. They have achieved 99.02% accuracy on CHB-MIT dataset and the total training time reported is around 7 h. Zhou, Tian, and Cao have fed time domain and frequency domain signals in the CNN network for seizure classification purposes [44]. In [11], authors have used plot images of EEG as input, to CNN to classify seizures and non-seizures. The model took approximately 11 h to train on GPU system for 50 epochs. In another study, the authors have incorporated three convolutional layers to learn the seizure properties from data then fed into the Nested Long Short Term Memory (NLSTM) to predict the output label [20]. San-Segundo et al. [35] introduced an algorithm where the outputs of Fourier transform, wavelet transform, and empirical mode decomposition are fed to CNN, which comprises two convolutional layers and three fully connected layers. A hybrid model of CNN and LSTM is used for seizure detection and tumor as well as identification of eye status [21]. In [5], authors have converted EEG data into saliency encoded spectrograms and uses an ensemble of CNN for classification.

However, all the above-mentioned deep learning methodologies give good accuracy; still, complexity is increased in terms of the number of layers, number of stages, number of epochs, and hybrid approach. The high complexity of the model burdens the computational requirements. In addition to that, all deep learning networks are very time-consuming in the training process, and usually, a GPU is required [15]. Also, some methodologies convert 1-D EEG signals to 2-D images to fed into CNN which may result in loss of important information. So, real-time implementation would be a little cumbersome task. Hence, there is a need for a useful model to overcome these issues.

1.2 Contribution of the work

This main objective of this work is to make an automated seizure detection system using the CNN framework having the least computational requirements and training time. Along with that, the authors have also focused on the robustness of the extracted features. The main challenge is getting a high detection rate in terms of less complexity and reduced training time. The proposed 1D CNN model directly takes raw data; it doesn't require any data splitting or augmentation steps, and fewer epochs are needed to extract features from the model. These things reduce the complexity of the model and make it suitable for real-time clinical monitoring. The features extracted from the proposed CNN model in this paper are merged with Logistic Regression (LR), Support vector machine (SVM), Random Forest (RF), K-Nearest Neighbour (KNN), Gaussian Naïve Bayes (GNB), Decision Tree (DT), Adaboost. The performance comparison depicts the sturdiness of the proposed CNN model.

The rest of the paper is structured as follows. Section 2 presents the dataset used, and Section 3 describes the proposed model. Results have been shown in Section 4 and discussion in Section 5. Finally, the conclusion has been presented in Section 6. Figure 1 shows the schematic of the proposed methodology. The detailed description of each block is described in the methodology section.

2 Dataset description

The experiments have been performed on a publicly available EEG dataset of the University of Bonn, Germany [3]. The EEG dataset has five different groups, namely Z, O, N, F and S, and
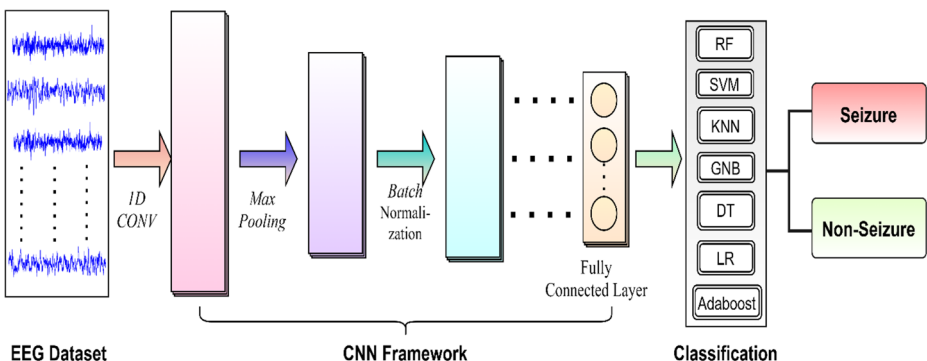

Fig. 1 Proposed methodology using 1D CNN framework and ML classifier

it is collected with the help of 128 channels. Each group of the dataset has 100 segments. The duration of each segment is 23.6 s and has the sampling frequency of 173.61 Hz.

The Z and O group segments were acquired from healthy person with eyes open and closed, respectively, using 10–20 system. The data collected in groups N and F are from hippocampal formation and epileptogenic zone, respectively, during seizure-free periods of epileptic patients. Segments of group S were taken from epileptic patients during the seizures period. Figure 2 presents EEG segment of Z, O, N, F and S group.

3 Methodology

3.1 Feature extraction

In this paper, features for epileptic seizure detection system are extracted using CNN. The advantage of using CNN is there is no need to select features manually. CNN is a type of ANN whose concept is inspired by the animal visual cortex. It is also known as shift-invariant artificial neural networks (SIANN). In recent years, handmade feature extraction techniques have been replaced by deep learning methods. CNN is a subset of deep learning that has attracted the research community, especially image classification problems. It has been used for fundus images [8], x-ray [10], CT scan [42], and other medical images in recent years. Many variants of CNN have also been used in object detection [27, 43]. CNN uses convolution operation to learn higher-order features from the data. It consists of a convolutional layer, pooling layer, and activation function. The proposed technique uses CNN on one-dimensional EEG signals for the best results in the terms of computational complexity, training time, and feature extraction. Table 1 shows the architecture of the proposed CNN model and the description of layers is as follows.

One dimensional convolutional layer It comprises kernels (filters) that are convolved with the input EEG signal. A kernel is a matrix that slides across the EEG signal to perform convolution operation after the operation feature map is generated.

$$f_m = \sum_{n=0}^{N-1} x_n l_{m-n} \qquad (1)$$

where x is EEG data, l is filter, and N is the number of elements in x. The subscript denotes the nth element of the vector, and the output vector is f.

Pooling layer There are different types of pooling layers. Maximum pooling, average pooling, global maximum pooling and global average pooling. In this work, one maximum pooling layer and one global average pooling layer are explored for better results. Pooling is used for downsampling operation. Maximum pooling helps in reducing the dimension of output neurons from the convolutional layer by selecting only the maximum value in each feature map, which prevents overfitting and reduces computational intensity. Global average pooling layer calculates the average of the previous layer feature map, and it is used in place of fully connected layers. Pooling layers do not have any trainable parameters.

Rectified linear activation unit (ReLu) Activation function is used to activate the node's summed input after every convolutional layer. In this work, ReLu activation function is used in
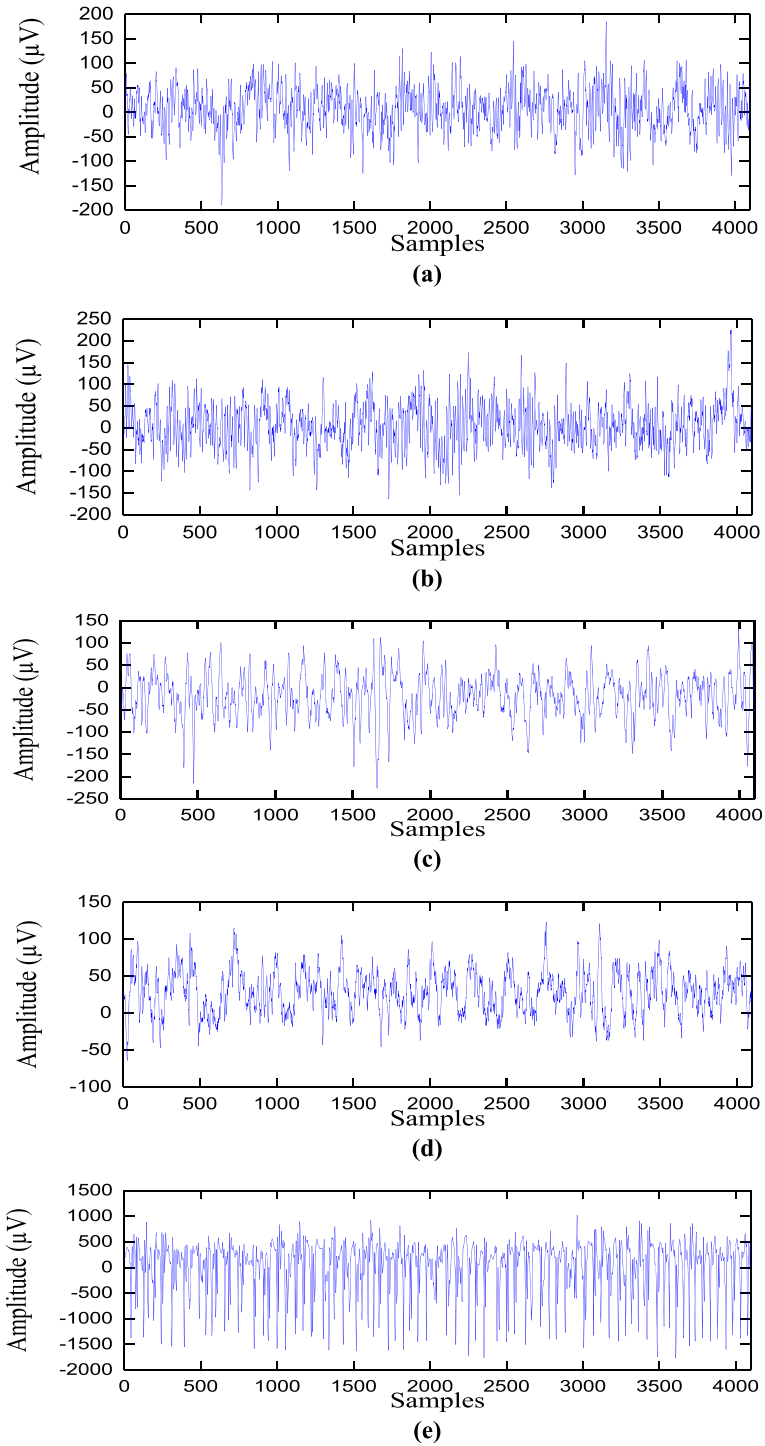
Fig. 2 Signal traces of the five groups (**a**) Z, (**b**) O, (**c**) N, (**d**) F, (**e**) S

Table 1 CNN structure for feature extraction

	Layer Info	Output Shape
Input		(4097,1)
Feature Extraction	Convolution 1D	(4095,128)
	Batch Normalization	(4095,128)
	Convolution 1D	(4093,128)
	Max Pooling	(1364,128)
	Batch Normalization	(1364,128)
	Convolution 1D	(1362,256)
	Batch Normalization	(1362,256)
	Convolution 1 D	(1360,256)
	Global Average Pooling	(256)
	Dropout	(256)
	Fully Connected	(100)

which it assigns zero for all negative values and has a linear identity for all positive values. The advantage of using Relu is the model converges fast and takes less time to train.

Batch Normalization (BN) Each layer of the neural network tries to correct its output due to an update in weights and biases during the training process. Due to which one has to initialize the parameters carefully and has to choose a small learning rate. This phenomenon is known as the internal covariate shift. To overcome this problem, batch normalization has been proposed [17]. Batch normalization normalizes the activation layer's value in mini-batches at each layer's input, due to which neural network stability increases. It helps in avoiding the special initialization of parameters yet provides faster convergence. In the proposed model, three BN layers have been used.

Fully connected layer Fully connected layers connect every neuron in one layer to every neuron in another layer. It accepts flattened output from the convolutional layers.

3.2 Classification

The end layer of the CNN is replaced by machine learning classifier and seven classification techniques have been taken into consideration, (a) Random Forest, (b) Support Vector Machine, (c) K-Nearest Neighbour, (d) Gaussian Naïve Bayes, (e) Decision Tree, (f) Logistic Regression, and (g) Adaboost. The brief elucidation of these classifiers is as follows:

(a) Random Forest - It is an ensemble learning method made out of various decision trees. In this classifier, the outputs of all decision trees are aggregated for classification [6]. The simplicity and flexibility of DT are combined in RF, which results in improvement of accuracy. The operation of RF algorithm is as follows: It creates bootstrapped dataset by selecting samples randomly from the original dataset and having the same dimensions as the original. The same sample can be selected more than once. After that, a DT is build using bootstrapped dataset but using only a random subset of variables at each step. Now again, it creates a new bootstrapped dataset and builds a DT in the same manner. It continues this process a number of times resulting in a wide variety of DT; this makes RF more effective than individual DT. Predictor values run down to all trees, and it belongs

to a class that has more votes. The parameters used in the experimental work are: number of estimators = 300, maximum depth = 100, and minimum sample split = 3.

(b) Support Vector Machine - This algorithm classifies the data points by finding a hyperplane in an N-dimensional space, where N is the number of features. To maximize the margin between data points of different classes is the main aim of this classifier so that future data points can be classified with more confidence.

(c) K-Nearest Neighbour - It is a non-linear, non-parametric, and one of the simplest classifiers. It classifies the data based on the similarities between the sample [19]. Similarity can be termed as closeness, proximity, or distance. The most popular method of calculating distance is Euclidian distance (ED).

$$ED = \sqrt{\sum\nolimits_{i=1}^{n} (y_{1i} - y_{2i})^2} \qquad (2)$$

where, $y_{1i} = (y_{11}, y_{12}y_{1n})$ and $y_{2i} = (y_{21}, y_{22}y_{12})$ are the data points, while n is the number of dimensions. The value of K decides the performance of this classifier. It is one of the widely used classifiers in the industry due to its less calculation time. K is set to be 5 in this work.

(d) Gaussian Naïve Bayes - This classifier has a simple probabilistic model which utilizes Bayes theorem for classification purpose. It works on the theory of class conditional independence, which assumes that a particular class's attribute is entirely independent of others [22]. The main advantage of GNB classifier is it needs less amount of training data compared to other classifiers. In this work, it is assumed that the data has a normal (Gaussian) distribution.

(e) Decision Tree - It is a tree-structured classifier. It has two types of nodes: a decision node and another is a leaf node. The decision node performs a test or takes the decision based on specific rules applied to features, and each leaf node shows the outcome of this test [30]. It is simple to understand, interpret, and visualize. In this study, gini index is used for the selection of features at each level selection. The main disadvantage of DT is there is a high probability of overfitting.

$$Gini = 1 - \sum\nolimits_{i} P_i^2 \qquad (3)$$

where P_i is the probability of ith class.

(f) Logistic Regression - It is a special case of ordinary linear regression (OLR) models. It imposes less strict requirements than OLR. LR shows a non-linear relationship between the explanatory variables and response. LR determines the changes in the logarithm of odds of the response variable, and it utilizes the sigmoid function for classification.

(g) Adaboost - It is a sequential ensemble classifier that focuses on combining results of weak classifiers to get a strong output. [12]. It assigns proper weights to weak classifiers at any level during training to get a high classification rate. Misclassified data is assigned to higher weights to be adequately classified by the next classifier, and the number of estimators in this experiment is set to 3.

4 Results

Raw EEG data has been directly fed into the proposed model, and the dataset is split randomly ten times into 70:30 ratios for training and testing, respectively. In Table 2, the average results have been reported, and nine experiments have been performed in the experimental setup. Classification accuracy, sensitivity, and specificity are calculated to check the performance of the proposed model. The metrics are given as

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \times 100\% \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% \tag{6}$$

where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

Table 2 Results obtained for different machine learning classifiers

Data Cluster	Metrics	RF	SVM	KNN	GNB	DT	LR	Adaboost
Z-S	Accuracy	**99.83**	**99.83**	**99.83**	98.67	99.67	**99.83**	99.67
	Sensitivity	**100**	**100**	**100**	97.65	**100**	**100**	**100**
	Specificity	**99.70**	**99.70**	**99.70**	**99.70**	99.35	**99.70**	99.39
O-S	Accuracy	98.67	98.67	98.67	98.67	**98.83**	**98.83**	98.67
	Sensitivity	98.66	98.63	**99.67**	99.65	98.98	99.32	98.68
	Specificity	98.64	98.64	97.72	97.72	**98.75**	98.69	**98.75**
ZO-S	Accuracy	**99.78**	99.67	99.67	99.22	99.67	99.67	99.67
	Sensitivity	**99.84**	**99.84**	99.68	99.03	99.68	**99.84**	99.68
	Specificity	**99.71**	99.35	**99.71**	**99.71**	**99.71**	99.35	**99.71**
N-S	Accuracy	**99.17**	98.99	**99.17**	99.00	98.67	**99.17**	98.33
	Sensitivity	**99.00**	**99.00**	**99.00**	98.69	98.35	**99.00**	98.35
	Specificity	**99.35**	98.99	**99.35**	**99.35**	99.03	**99.35**	98.35
F-S	Accuracy	98.17	98.00	98.17	97.45	**98.50**	98.00	97.50
	Sensitivity	96.92	97.17	97.87	96.54	**98.55**	97.23	96.59
	Specificity	**99.35**	98.70	98.41	98.41	**98.39**	98.70	98.74
NF-S	Accuracy	98.77	98.88	98.44	98.44	**98.99**	98.88	98.77
	Sensitivity	99.50	**99.68**	98.35	98.33	99.49	**99.68**	99.50
	Specificity	98.05	98.05	98.72	**99.08**	98.04	98.05	98.05
ZONF-S	Accuracy	98.33	98.13	98.06	97.73	97.93	**98.47**	98.07
	Sensitivity	98.68	98.62	98.02	97.52	98.53	**98.77**	98.36
	Specificity	96.83	96.03	98.36	**98.48**	95.49	97.29	96.89
ZO-NFS	Accuracy	**98.73**	98.47	98.33	92.80	98.40	98.67	98.40
	Sensitivity	99.30	99.30	**99.82**	98.80	98.77	99.30	98.39
	Specificity	98.38	**98.57**	97.38	89.13	98.16	98.29	98.37
ZO-NF	Accuracy	98.25	**98.75**	98.16	94.83	97.17	98.58	96.91
	Sensitivity	98.26	**99.31**	99.27	93.88	97.06	98.96	97.80
	Specificity	**98.14**	**98.14**	97.05	95.88	97.13	**98.14**	95.82
ZO-NF-S	Accuracy	97.07	97.87	95.73	93.47	95.87	**97.93**	86.73
	Sensitivity	97.03	97.94	95.22	92.15	95.71	**98.01**	84.06
	Specificity	97.15	97.55	97.77	98.88	96.43	**97.55**	98.72

The bold symbols signifies the highest value among different classifiers

LR, RF, SVM, and KNN show an accuracy of 99.83% between Z (healthy person- eyes open) and S (seizures). All classifiers have achieved 100% sensitivity except GNB. To balance the imbalanced data clusters ZO-S, NF-S, ZONF-S, ZO-NFS, and ZO-NF-S Adaptive Synthetic (ADASYN) sampling method is used [16]. Among the three classes, maximum accuracy of 97.93% is achieved by LR classifier. Tables 3 and 4 show the comparison of metrics for different classifiers. It can be seen from both tables that there is a maximum 1.18% difference among different classifiers for detection between seizures and non-seizures. This shows the robustness of the extracted features using the proposed CNN model.

To compare our CNN-based framework, we have removed the last fully connected layer with 100 neurons and passed it through softmax function for classification. It is a type of activation function and usually the last layer in the deep learning classification model. This function gives the probability distribution of different output classes. For testing and training purposes K-fold cross-validation scheme has been used. In this method, data has been split into K folds, and at some point, each part is used as a testing set. K is set to 10 in this experimental setup; hence the data cluster is divided into 10 folds. 10 iterations took place, and in each one, different folds are carried as test data. The final output is taken as an average of the results obtained from each iteration. Training of the model is done using Tensorflow, a deep learning library on Anaconda Software. Cross-entropy loss function and Adam as optimizer have been used in this paper. Learning rate in adam optimizer is taken 0.00000001, and the other five parameters are set to their default values. A small learning rate is taken to avoid local minima and control the oscillation of the network. Batch size of 10 has been selected, and different data clusters have been trained using different number of epochs. Minimum epoch is 200 for O-S, F-S, ZO-NF-S. It can be seen from Table 5 in classification between normal and ictal, data cluster O-S shows the highest classification accuracy of 99.5% and 100% sensitivity and specificity while Z-S gives 99% and ZO-S gives 98% accuracy. The classification between interictal and ictal groups shows a maximum of 98.5% accuracy and a minimum of 98%.

Table 3 Comparison of metrics for data cluster Z-S

		RF	SVM	KNN	GNB	DT	LR	Adaboost
RF	Accuracy	-	0.00	0.00	-1.18	-0.16	0.00	-0.16
	Sensitivity	-	0.00	0.00	-2.41	0.00	0.00	0.00
	Specificity	-	0.00	0.00	0.00	-0.35	0.00	-0.31
SVM	Accuracy	0.00	-	0.00	-1.18	-0.16	0.00	-0.16
	Sensitivity	0.00	-	0.00	-2.41	0.00	0.00	0.00
	Specificity	0.00	-	0.00	0.00	-0.35	0.00	-0.31
KNN	Accuracy	0.00	0.00	-	-1.18	-0.16	0.00	-0.16
	Sensitivity	0.00	0.00	-	-2.41	0.00	0.00	0.00
	Specificity	0.00	0.00	-	0.00	-0.35	0.00	-0.31
GNB	Accuracy	1.16	1.16	1.16	-	1.00	1.16	1.00
	Sensitivity	2.35	2.35	2.35	-	2.35	2.35	2.35
	Specificity	0.00	0.00	0.00	-	-0.35	0.00	-0.31
DT	Accuracy	0.16	0.16	0.16	-1.01	-	0.16	0.00
	Sensitivity	0.00	0.00	0.00	-2.41	-	0.00	0.00
	Specificity	0.35	0.35	0.35	0.35	-	0.35	0.04
LR	Accuracy	0.00	0.00	0.00	-1.18	-0.16	-	-0.16
	Sensitivity	0.00	0.00	0.00	-2.41	0.00	-	0.00
	Specificity	0.00	0.00	0.00	0.00	-0.35	-	-0.31
Adaboost	Accuracy	0.16	0.16	0.16	-1.01	0.00	0.16	-
	Sensitivity	0.00	0.00	0.00	-2.41	0.00	0.00	-
	Specificity	0.31	0.31	0.31	0.31	-0.04	0.31	-

Table 4 Comparison of metrics for data cluster N-S

		RF	SVM	KNN	GNB	DT	LR	Adaboost
	Accuracy	-	-0.18	0.00	-0.17	-0.51	0.00	-0.85
RF	Sensitivity	-	0.00	0.00	-0.31	-0.66	0.00	-0.66
	Specificity	-	-0.36	0.00	0.00	-0.32	0.00	-1.02
	Accuracy	0.18	-	0.18	0.01	-0.32	0.18	-0.67
SVM	Sensitivity	0.00	-	0.00	-0.31	-0.66	0.00	-0.66
	Specificity	0.36	-	0.36	0.36	0.04	0.36	-0.65
	Accuracy	0.00	-0.18	-	-0.17	-0.51	0.00	-0.85
KNN	Sensitivity	0.00	0.00	-	-0.31	-0.66	0.00	-0.66
	Specificity	0.00	-0.36	-	0.00	-0.32	0.00	-1.02
	Accuracy	0.17	-0.01	0.17	-	-0.33	0.17	-0.68
GNB	Sensitivity	0.31	0.31	0.31	-	-0.35	0.31	-0.35
	Specificity	0.00	-0.36	0.00	-	-0.32	0.00	-1.02
	Accuracy	0.50	0.32	0.50	0.33	-	0.50	-0.35
DT	Sensitivity	0.66	0.66	0.66	0.34	-	0.66	0.00
	Specificity	0.32	-0.04	0.32	0.32	-	0.32	-0.69
	Accuracy	0.00	-0.18	0.00	-0.17	-0.51	-	-0.85
LR	Sensitivity	0.00	0.00	0.00	-0.31	-0.66	-	-0.66
	Specificity	0.00	-0.36	0.00	0.00	-0.32	-	-1.02
	Accuracy	0.85	0.67	0.85	0.68	0.34	0.85	-
Adaboost	Sensitivity	0.66	0.66	0.66	0.34	0.00	0.66	-
	Specificity	1.01	0.65	1.01	1.01	0.69	1.01	-

5 Discussion

Many feature extraction techniques followed by different classifiers have been proposed for epileptic seizure detection. A comparison of different methodologies has been represented in Table 6. In previous work, authors used only alpha band to detect seizures and achieved maximum accuracy of 98%. Short-time Fourier transform (STFT) was used to extract features, and four time-frequency statistical features were fed to different classifiers to analyze the performance [31]. In another work, Haralick features were extracted from the gamma band, followed by a decision tree classifier and 96% success rate to distinguish between seizures and healthy persons [33]. There is always a possibility that some important information is partially or fully missed in selecting the features in classic feature detectors. However, in the proposed method, since the features are extracted directly from EEG data, no preprocessing steps, transformations, or any feature selection technique are involved, so the maximum information is potentially extracted.

Table 5 Results obtained for softmax classification

Data cluster	Accuracy (%)	Sensitivity (%)	Specificity (%)
Z-S	99	100	99
O-S	99.5	100	100
ZO-S	98.33	100	96
N-S	98.5	100	98
F-S	98.5	96	97
NF-S	98	96.5	94
ZONF-S	96	98.75	83
ZO-NFS	95	96	94
ZO-NF	97.25	97	98.5
ZO-NF-S	94.4	90	93

Table 6 Comparison of the proposed algorithm with other studies on Bonn EEG dataset

Reference	Classification task	Methodology	Accuracy (%)
[34]	Z-S	MT rational DSTFT and MLP	99.80
[37]		Hermite transform and LS-SVM	99.50
Proposed work		CNN+LR	**99.83**
[45]	O-S	weighted horizontal visibility graph and KNN	97
Proposed work		CNN+LR	**98.83**
[9]	ZO-S	Weighted complex networks and SVM	96.40
[28]		Matrix determinant and MLP	97.10
[37]		Hermite transform and LS-SVM	98.67
[7]		TQWT, entropy features, Hjorth parameter and HMM	99.58
Proposed work		CNN+LR	**99.67**
[34]	N-S	MT rational DSTFT and MLP	98.50
[37]		Hermite transform and LS-SVM	98.50
[14]		CEEMDAN and Adaboost	99
Proposed work		CNN+LR	**99.17**
[34]	F-S	MT rational DSTFT and MLP	94.90
[37]		Hermite transform and LS-SVM	97.50
[36]		DWT based WL, ZC, SSC and NB&SVM	97.75
Proposed work		CNN+LR	**98.00**
[9]	NF-S	Weighted complex networks and SVM	94.50
[37]		Hermite transform and LS-SVM	98
[18]		GModPCA and SVM	95.80
[7]		TQWT, entropy features, Hjorth parameter and HMM	97.50
Proposed work		CNN+LR	**98.88**
[34]	ZONF-S	MT rational DSTFT and MLP	98.10
[37]		Hermite transform and LS-SVM	97.60
[7]		TQWT, entropy features, Hjorth parameter and HMM	97
[36]		DWT based WL, ZC, SSC and NB&SVM	97.90
[2]		SEA based DNN	97.17
Proposed work		CNN+LR	**98.47**
[1]	S-ZO-NF	13 layer CNN	88.70
[14]		CEEMDAN and Adaboost	97.60
Proposed work		CNN+LR	**97.93**

HMM: Hidden Markov Model, TQWT: Tunable Q wavelet transform, CEEMDAN: Complete ensemble empirical mode decomposition with adaptive noise, MT: Malmquist-Takenaka, DSTFT: Discrete Short-time Fourier Transform, MLP: Multilayer Perceptron, WL: Waveform length, SSC: Number of slope sign changes, ZC: Number of zero crossings, LS-SVM: Least square support vector machine

The bold value represents highest value in comparison with other existing studies

Acharya et al. [1] were the first to use CNN for epileptic seizure detection. They presented a network with 13 layers, which gives 88.7% accuracy for three-class classification on Bonn EEG dataset. They trained the model using 150 epochs. In [4], the authors presented a CNN structure having 18 layers for feature extraction followed by an RF classifier to detect neonatal seizures and achieved 77% accuracy. Raghu et al. [29] used pretrained networks such as Alexnet, Vgg16, Vgg19, and other landmark models on Temple University dataset. They presented two approaches: (a) transfer learning and (b) extracting features from pretrained networks followed by SVM classifier. Transfer learning approach achieved 82.85% accuracy and 88.30% accuracy using extracting feature approach.

Almost all deep learning networks are very time-consuming in the training process. The advantages of the proposed methodology regarding other deep learning techniques are that it does not require any preprocessing step or conversion of 1-D EEG signal to 2-D. On average,

only 7 epochs are required to train the model and per epoch takes approximately 10 s. The simulations were carried on Intel(R) Core(TM) i7-8700 CPU@ 3.2 GHz having 8GB RAM. The total training time is less than 2 min. The reduced number of stages and epochs makes this model suitable for portable/ wearable devices.

The limitations of the proposed methodology need to be acknowledged. The parameters of the model i.e., the number of filters and number of convolutional layers, etc. have been chosen using trial and error method. The experiments are performed on a small dataset. To generalize the results, we will implement them on a large dataset in the future.

6 Conclusions

A CNN-based framework for feature extraction followed by a machine learning classifier was introduced in this paper. The proposed methodology does not require a hand-engineered feature extraction process. It automatically extracts the features optimally based on training data. Few epochs (average seven in the proposed technique) make it faster than other algorithms, a prime condition for real-time application. All classifiers show almost the same accuracy to classify among different data clusters, making the extracted features robust in nature. The model was also compared by terminating the classifier part with the softmax function. It is concluded that machine learning classifiers give more accuracy in few epochs than pure deep learning model. The proposed detection system will be helpful for a neurologist for making decisions correctly. Both cloud-based and standalone systems can be developed using the proposed model.

References

1. Acharya UR, Oh SL, Hagiwara Y et al (2018) Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput Biol Med. https://doi.org/10.1016/j.compbiomed.2017.09.017
2. Akyol K (2020) Stacking ensemble based deep neural networks modeling for effective epileptic seizure detection. Expert Syst Appl 148. https://doi.org/10.1016/j.eswa.2020.113239
3. Andrzejak RG, Lehnertz K, Mormann F et al (2001) 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 Physics, Plasmas, Fluids, Relat Interdiscip Top. https://doi.org/10.1103/PhysRevE.64.061907
4. Ansari AH, Cherian PJ, Caicedo A et al (2019) Neonatal seizure detection using deep convolutional neural networks. Int J Neural Syst 29:1–20. https://doi.org/10.1142/S0129065718500119
5. Asif U, Roy S, Tang J, Harrer S (2019) SeizureNet: a deep convolutional neural network for accurate seizure type classification and seizure detection. ArXiv
6. Breiman L (2001) Randomforest2001. Mach Learn. https://doi.org/10.1017/CBO9781107415324.004
7. Dash DP, Kolekar MH (2020) Hidden Markov model based epileptic seizure detection using tunable Q wavelet transform. J Biomed Res 34:170–179. https://doi.org/10.7555/JBR.34.20190006
8. Diaz-Pinto A, Morales S, Naranjo V et al (2019) CNNs for automatic glaucoma assessment using fundus images: An extensive validation. Biomed Eng Online. https://doi.org/10.1186/s12938-019-0649-y
9. Diykh M, Li Y, Wen P (2017) Classify epileptic EEG signals using weighted complex networks based community structure detection. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2017.08.012
10. Dong Y, Pan Y, Zhang J, Xu W (2017) Learning to read chest x-ray images from 16000 + examples using CNN. In: Proceedings – 2017 IEEE 2nd International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2017

11. Emami A, Kunii N, Matsuo T et al (2019) Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images. NeuroImage Clin. https://doi.org/10.1016/j.nicl.2019.101684
12. Freund Y, Schapire RE (1996) Experiments with a new boosting algorithm. In Proceedings of the Thirteenth International Conference on Machine Learning Theory. San Francisco, CA: Morgan Kaufmann Publishers Inc. https://doi.org/10.5555/3091696.3091715
13. Gotman J (1990) Automatic seizure detection: improvements and evaluation. Electroencephalogr Clin Neurophysiol. https://doi.org/10.1016/0013-4694(90)90032-F
14. Hassan AR, Subasi A, Zhang Y (2020) Epilepsy seizure detection using complete ensemble empirical mode decomposition with adaptive noise. Knowl-Based Syst 191:105333. https://doi.org/10.1016/j.knosys.2019.105333
15. He K, Sun J (2015) Convolutional neural networks at constrained time cost. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp 5353–5360
16. He H, Bai Y, Garcia EA, Li S (2008) ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: Proceedings of the International Joint Conference on Neural Networks
17. Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: 32nd International Conference on Machine Learning, ICML 2015
18. Jaiswal AK, Banka H (2017) Epileptic seizure detection in EEG signal with GModPCA and support vector machine. Biomed Mater Eng. https://doi.org/10.3233/BME-171663
19. Kotsiantis SB (2007) Supervised machine learning: a review of classification techniques. In: Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies. IOS Press, Amsterdam, pp 3–24
20. Li Y, Yu Z, Chen Y et al (2020) Automatic seizure detection using fully convolutional nested LSTM. Int J Neural Syst. https://doi.org/10.1142/S0129065720500197
21. Liu Y, Huang YX, Zhang X et al (2020) Deep C-LSTM neural network for epileptic seizure and tumor detection using high-dimension EEG signals. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2976156
22. Marsland S (2014) Machine learning: An algorithmic perspective (2nd edn). New York: Chapman & Hall
23. Muhammad G, Masud M, Amin SU et al (2018) Automatic seizure detection in a mobile multimedia framework. IEEE Access. https://doi.org/10.1109/ACCESS.2018.2859267
24. Namazi H, Kulish VV, Hussaini J et al (2016) A signal processing based analysis and prediction of seizure onset in patients with epilepsy.Oncotarget. https://doi.org/10.18632/ONCOTARGET.6341
25. Ocak H (2009) Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2007.12.065
26. Pravin Kumar S, Sriraam N, Benakop PG, Jinaga BC (2010) Entropies based detection of epileptic seizures with artificial neural network classifiers. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2009.09.051
27. Quan Q, He F, Li H (2020) A multi-phase blending method with incremental intensity for training detection networks. Vis Comput. https://doi.org/10.1007/s00371-020-01796-7
28. Raghu S, Sriraam N, Hegde AS, Kubben PL (2019) A novel approach for classification of epileptic seizures using matrix determinant. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2019.03.021
29. Raghu S, Sriraam N, Temel Y et al (2020) EEG based multi-class seizure type classification using convolutional neural network and transfer learning. Neural Netw. https://doi.org/10.1016/j.neunet.2020.01.017
30. Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern. https://doi.org/10.1109/21.97458
31. Sameer M, Gupta B (2020) Detection of epileptical seizures based on alpha band statistical features. Wirel Pers Commun. https://doi.org/10.1007/s11277-020-07542-5
32. Sameer M, Gupta AK, Chakraborty DC, Gupta DB (2019) Epileptical Seizure Detection: Performance analysis of gamma band in EEG signal Using Short-Time Fourier Transform. In: International Symposium on Wireless Personal Multimedia Communications, WPMC
33. Sameer M, Gupta AK, Chakraborty C, Gupta B (2020) ROC analysis for detection of epileptical seizures using haralick features of gamma band. In: 26th National Conference on Communications, NCC 2020
34. Samiee K, Kovács P, Gabbouj M (2015) Epileptic seizure classification of EEG time-series using rational discrete short-time fourier transform. IEEE Trans Biomed Eng. https://doi.org/10.1109/TBME.2014.2360101
35. San-Segundo R, Gil-Martín M, D'Haro-Enríquez LF, Pardo JM (2019) Classification of epileptic EEG recordings using signal transforms and convolutional neural networks. Comput Biol Med. https://doi.org/10.1016/j.compbiomed.2019.04.031
36. Sharmila A, Geethanjali P (2020) Evaluation of time domain features on detection of epileptic seizure from EEG signals. Health Technol (Berl) 10:711–722. https://doi.org/10.1007/s12553-019-00363-y

37. Siuly S, Alcin OF, Bajaj V et al (2019) Exploring Hermite transformation in brain signal analysis for the detection of epileptic seizure. IET Sci Meas Technol. https://doi.org/10.1049/iet-smt.2018.5358
38. Tzallas AT, Tsipouras MG, Fotiadis DI (2009) Epileptic seizure detection in EEGs using time-frequency analysis. IEEE Trans Inf Technol Biomed. https://doi.org/10.1109/TITB.2009.2017939
39. Ullah I, Hussain M, Qazi E, ul H, Aboalsamh H (2018) An automated system for epilepsy detection using EEG brain signals based on deep learning approach. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2018.04.021
40. Wang G, Deng Z, Choi KS (2017) Detection of epilepsy with Electroencephalogram using rule-based classifiers. Neurocomputing. https://doi.org/10.1016/j.neucom.2016.09.080
41. Widman G, Schreiber T, Rehberg B et al (2000) Quantification of depth of anesthesia by nonlinear time series analysis of brain electrical activity. Phys Rev E - Stat Physics, Plasmas, Fluids, Relat Interdiscip Top. https://doi.org/10.1103/PhysRevE.62.4898
42. Xie H, Yang D, Sun N et al (2019) Automated pulmonary nodule detection in CT images using deep convolutional neural networks. Pattern Recognit. https://doi.org/10.1016/j.patcog.2018.07.031
43. Zhang S, He F, Ren W (2020) NLDN: Non-local dehazing network for dense haze removal. Neurocomputing 410:363–373. https://doi.org/10.1016/j.neucom.2020.06.041
44. Zhou M, Tian C, Cao R et al (2018) Epileptic seizure detection based on EEG signals and CNN. Front Neuroinform. https://doi.org/10.3389/fninf.2018.00095
45. Zhu G, Li Y, Wen PP (2014) Epileptic seizure detection in EEGs signals using a fast weighted horizontal visibility algorithm. Comput Methods Programs Biomed. https://doi.org/10.1016/j.cmpb.2014.04.001