Expert Systems With Applications 107 (2018) 61–71

Contents lists available at ScienceDirect

Expert Systems With Applications

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

An automated system for epilepsy detection using EEG brain signals based on deep learning approach

Ihsan Ullah[a], Muhammad Hussain[b,*], Emad-ul-Haq Qazi[b], Hatim Aboalsamh[b]

[a] Insight Centre for Data Analytics, National University of Ireland, Galway, Ireland
[b] Visual Computing Lab, Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

ARTICLE INFO

Article history:
Received 8 July 2017
Revised 13 April 2018
Accepted 14 April 2018
Available online 18 April 2018

Keywords:
Electroencephalogram (EEG)
Epilepsy
Seizure
Ictal
Interictal
1D-CNN

ABSTRACT

Epilepsy is a life-threatening and challenging neurological disorder, which is affecting a large number of people all over the world. For its detection, encephalography (EEG) is a commonly used clinical approach, but manual inspection of EEG brain signals is a time-consuming and laborious process, which puts a heavy burden on neurologists and affects their performance. Several automatic systems have been proposed using traditional approaches to assist neurologists, which perform well in detecting binary epilepsy scenarios e.g. normal vs. ictal, but their performance degrades in classifying ternary case e.g. ictal vs. normal vs. inter-ictal. To overcome this problem, we propose a system that is an ensemble of pyramidal one-dimensional convolutional neural network (P-1D-CNN) models. Though a CNN model learns the internal structure of data and outperforms hand-engineered techniques, the main issue is the large number of learnable parameters, whose learning requires a huge volume of data. To overcome this issue, P-1D-CNN works on the concept of refinement approach and it involves 61% fewer parameters compared to standard CNN models and as such it has better generalization. Further to overcome the limitations of the small amount of data, we propose two augmentation schemes. We tested the system on the University of Bonn dataset, a benchmark dataset; in almost all the cases concerning epilepsy detection, it gives an accuracy of $99.1 \pm 0.9\%$ and outperforms the state-of-the-art systems. In addition, while enjoying the strength of a CNN model, P-1D-CNN model requires 61% less memory space and its detection time is very short (< 0.000481 s), as such it is suitable for real-time clinical setting. It will ease the burden of neurologists and will assist the patients in alerting them before the seizure occurs. The proposed P-1D-CNN model is not only suitable for epilepsy detection, but it can be adopted in developing robust expert systems for other similar disorders.

1. Introduction

Epilepsy is a neurological disorder affecting about fifty million people in the world (Megiddo et al., 2016). Electroencephalogram (EEG) is an effective and non-invasive technique commonly used for monitoring the brain activity and diagnosis of epilepsy. EEG readings are analyzed by neurologists to detect and categorize the patterns of the disease such as pre-ictal spikes and seizures. The visual examination is time-consuming and laborious; it takes many hours to examine one-day EEG recording of a patient, and it requires the services of an expert. As such, the analysis of the EEG brain signals of patients puts a heavy burden on neurologists and reduces their efficiency. These limitations motivated efforts to design and develop automatic systems to assist neurologists in classifying epileptic and non-epileptic EEG brain signals.

Recently, a lot of research work has been carried out to detect the epileptic and non-epileptic signals as a classification problem (Gardner, Krieger, Vachtsevanos, & Litt, 2006; Meier, Dittrich, Schulze-Bonhage, & Aertsen, 2008; Mirowski, Madhavan, LeCun, & Kuzniecky, 2009; Sheb & Guttag , 2010). From the machine learning (ML) point of view, recognition of epileptic and non-epileptic EEG signals is a challenging task. Usually, there is a small amount of epilepsy data available for training a classifier due to infrequently happening of seizures. Further, the presence of noise and artifacts in the data creates difficulty in learning the brain patterns associated with normal, ictal, and non-ictal cases. This difficulty increases further due to inconsistency in seizure morphology among patients (McShane, 2004). The existing automatic seizure detection techniques use traditional signal processing (SP) and ML techniques. Many of these techniques show good accuracy for one problem but fail in performing accurately for others e.g.

* Corresponding author.
E-mail addresses: Ihsan.ullah@insight-centre.org (I. Ullah), mhussain@ksu.edu.sa (M. Hussain), qulhaq@ksu.edu.sa (E.-u.-H. Qazi), hatim@ksu.edu.sa (H. Aboalsamh).

https://doi.org/10.1016/j.eswa.2018.04.021
0957-4174/

they classify seizure vs. non-seizure case with a good accuracy but show poor performance in case of normal vs. ictal vs. inter-ictal (Zhang, Chen, & Li, 2017). It is still a challenging problem due to three reasons, i) a generalized model does not exist which can classify binary as well as a ternary problem (i.e. normal vs. ictal vs. inter-ictal), ii) less available labeled data, and ii) low accuracy. To help and assist neurologists, we need a generalized automatic system that gives good performance even with fewer training samples (Andrzejak et al., 2001; Sharmila & Geethanjali, 2016).

Exiting methods for the detection of seizures use hand-engineered techniques for feature extraction from EEG signals. Some methods use spectral (Tzallas et al., 2012) and temporal aspects of information from EEG signals (Shoeb, 2009). An EEG signal contains low-frequency features with long time-period and high-frequency features with a short time period (Adeli, Zhou, & Dadmehr, 2003) i.e. there is a kind of hierarchy among features. Deep learning (DL) is a state-of-the-art ML approach that automatically encodes hierarchy of features, which are not data dependent and adapt to internal structure of the data; it has shown promising results in many applications. Moreover, features extracted using the DL models have shown to be more discriminative and robust than hand-designed features (LeCun & Bengio, 1995). In order to improve the accuracy in the classification of epileptic and non-epileptic EEG signals, we propose a method based on DL.

The recent emergence of DL techniques show significant performance in several application areas. The variants of deep CNN i.e. 2D CNN such as AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), VGG (Simonyan & Zisserman , 2014) etc. or 3D networks such as 3DCNN Ji, Xu, Yang, & Yu, 2013), C3D (Tran, Bourdev, Fergus, Torresani, & Paluri, 2015) etc. have shown outstanding performance in many fields. Recently, 1D-CNN has been successfully used for text understanding, music generation, and other time series data (Cui, Chen, & Chen, 2016; Ince et al., 2016; LeCun, Bottou, Bengio, & Haffner, 1998; Zhang & LeCun, 2015). The end-to-end learning paradigm of DL approach avoids the selection of a proper combination of feature extractor and feature subset selector for extracting and selecting the most discriminative features that are to be classified by a suitable classifier (Andrzejak et al., 2001; Hussain, Aboalsamh, Abdul, Bamatraf, & Ullah, 2016; Sharmila & Geethanjali , 2016; Zhang et al., 2017). Although the traditional approach is fast in training as compared to DL approach, it is far slower at test time and does not generalize well. Trained deep models can test a sample in a fraction of a second, and are suitable for real-time applications; the only bottleneck is the requirement of a large amount of data and its long training time. To overcome this problem, an augmentation scheme needs to be introduced that may help in using a small amount of available data in an optimal way for training a deep model.

As an EEG recording is a 1D signal, we propose a pyramidal 1D-CNN (P-1D-CNN) model for detecting epilepsy, which comprises of far fewer number of learnable parameters. The amount of available data is small, therefore, to train a P-1D-CNN, we propose two augmentation schemes. Using trained P-1D-CNN models as experts, we design a system as an ensemble of P-1D-CNN models, which employs majority vote strategy to fuse the local decisions for detecting epilepsy. The proposed system takes an EEG signal, segment it with fixed-size sliding window, and pass the sub-signals to base P-1D-CNN models (Fig. 2) that process them and give the local decisions to the majority-vote module. In the end, the majority-vote module takes the final decision (Fig. 1). It outperforms the state-of-the-art techniques for different problems concerning epilepsy detection. The main contributions of this study are: (1) data augmentation schemes, (2) an automatic system based on an ensemble of P-1D-CNN deep models for binary as well as ternary EEG signal classification, (3) a new approach for structuring deep 1D-CNN model and (4) thorough evaluation of the augmentation schemes and the deep models for detecting different epilepsy cases.

The rest of the paper is organized as follows: In Section 2, we present the literature review. Section 3 describes in detail the proposed system. Model selection, data augmentation schemes, and training of P-1D-CNN model are discussed in Section 4. Section 5 presents results; Section 6 discusses the results and compares them with those by the state-of-the-art methods. In the end, Section 6 concludes the paper and present the future directions.

2. Literature review

The recognition of epileptic and non-epileptic EEG signals is a classification problem. It involves extraction of the discriminative features from EEG signals and then performing classification. In the following paragraphs, we gave an overview of the related state-of-the-art techniques, which use different feature extraction and classification methods for classification of epileptic and non-epileptic EEG signals.

Almost all existing methods for epilepsy detection are based on hand-engineered feature extraction techniques. Chua, Chandran, Acharya, and Lim (2011) used Higher Order Spectra (HOS) and power spectrum based features for the automated detection of epilepsy. The authors used the Gaussian Mixture Model (GMM) as a classifier and obtained the classification accuracies of 93.11% and 88.78% with HOS and power spectrum based features, respectively. In another study, Chua, Chandran, Acharya, and Lim (2009) used SVM classifier with HOS based features and achieved an accuracy of 92.67%. Acharya, Vinitha Sree, and Suri (2011) used cumulants for the automated detection of epilepsy. They extracted the HOS cumulants from Wavelet Packet Decomposition (WPD) coefficients and obtained an accuracy of 98.5% with SVM classifier.

Subasi (2007) proposed a method to classify normal vs epileptic EEG brain signals. In this method, EEG brain signals are decomposed into different frequency sub-bands using the discrete wavelet transform (DWT). Four statistical features are extracted from DWT coefficients and are passed to a modular neural network (called Mixture of Experts-MEs) for classification. They reported a sensitivity of 95%, specificity of 94% and accuracy of 94.5%. In another study (Acharya et al., 2012), the authors used SampEn, ApEn and two-phase entropies and a Fuzzy classifier; they reported a specificity of 100%, accuracy of 98.1% and sensitivity of 99.4%. Martis et al. (2013) used features derived from intrinsic Time-Scale decomposition (ITD) and decision tree classifier. This method achieved an accuracy of 95.67%, a specificity of 99.50% and a sensitivity of 99%. In (Acharya et al., 2013), authors proposed a method for the automated classification of EEG brain signals into three different classes, i.e., ictal, normal and interictal. They used Continuous Wavelet Transform (CWT) for feature extraction and SVM as a classifier. Results indicate that this method obtained an accuracy of 96%.

Swami, Gandhi, Panigrahi, Tripathi, and Anand (2016) extracted hand-crafted features such as Shannon entropy, standard deviation, and energy. They employed the general regression neural network (GRNN) classifier to classify these features and achieved maximum accuracy, i.e., 100% and 99.18% for A-E (non-seizure vs. seizure) and AB-E (normal vs. seizure) cases, respectively on Bonn dataset. However, maximum accuracy for other cases like B-E, C-E, D-E, CD-E, and ABCD-E is 98.4%. In another study, Guo, Rivero, Dorado, Rabunal, and Pazos (2010) achieved the accuracy of 97.77% for ABCD-E case on the same dataset. They used artificial neural network classifier (ANN) to classify the line length features that were extracted by using discrete wavelet transform (DWT). Nicolaou and Georgiou (2012) extracted the permutation entropy feature from EEG signals. They employed support vector machine (SVM) as a classifier and achieved an accuracy of 93.55% for A-

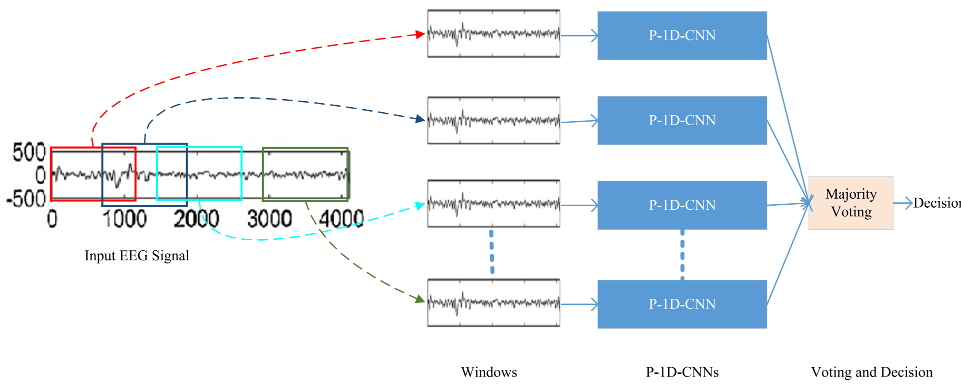

Fig. 1. Architecture of the proposed Automatic System for Epilepsy Detection based on the Ensemble of P-1D-CNN models.

E case on the University of Bonn dataset. The maximum accuracy for other cases such as B-E, C-E, D-E, and ABCD-E was 86.1%. Gandhi, Panigrahi, and Anand (2011) extracted the entropy, standard deviation and energy features from EEG signals using DWT. They used SVM and probabilistic neural network (PNN) as a classifier and reported the maximum accuracy of 95.44% for ABCD-E case. Gotman, Ives, and Gloor (1979) used sharp wave and spike recognition technique. They further enhanced this technique in Gotman (1982,1999), Koffler and Gotman (1985) and Qu and Gotman (1993). Shoeb (2009) used SVM classifier and adopted a patient-specific prediction methodology; the results indicate that a 96% accuracy was achieved. In most of the works, common classifier used to distinguish between seizure and non-seizure events is support vector machine (SVM). However, in Khan, Rafiuddin, and Farooq (2012) linear discriminant analysis (LDA) classifier was used for classification of five subjects consisting of sixty-five seizures. It achieved 91.8%, 83.6% and 100% accuracy, sensitivity, and specificity, respectively. Acharya et al. (2012) focused on using entropies for EEG seizure detection and seven different classifiers. The best-performing classifier was the Fuzzy Sugeno classifier, which achieved 99.4% sensitivity, 100% specificity, and 98.1% overall accuracy. The worst performing classifier was the Naive Bayes Classifier, which achieved 94.4% sensitivity, 97.8% specificity, and 88.1% accuracy. Nasehi and Pourghassem (2013) used Particle Swarm Optimization Neural Network (PSONN), which gave 98% sensitivity. Yuan, Zhou, Liu, and Wang (2012) used extreme learning machine (ELM) algorithm for classification. Twenty-one (21) seizure records were used to train the classifier and sixty-five (65) for testing. The results showed that the system achieved on average 91.92% sensitivity, 94.89% specificity and 94.9% overall accuracy. Patel, Chua, Fau, and Bleakley (2009) proposed a low-power, real-time classification algorithm, for detecting seizures in ambulatory EEG. They compared Mahalanobis discriminant analysis (MDA), quadratic discriminant analysis (QDA), linear discriminant analysis (LDA) and SVM classifiers on thirteen (13) subjects. The results indicate that the LDA show the best results when it is trained and tested on a single patient. It gave 94.2% sensitivity, 77.9% specificity, and 87.7% overall accuracy. When generalized across all subjects, it gave 90.9% sensitivity, 59.5% specificity, and 76.5% overall accuracy. Acharya, Faust, Kannathal, Chua, and Laxminarayan (2005) used Recurrence Quantification Analysis (RQA) features for the three-class classification of EEG signals to detect epilepsy. The authors employed the SVM as a classifier and achieved an accuracy of 95.60% by using the RQA parameters as features. Further, a detailed list of feature extractors and classifiers used for binary (e.g. epileptic vs. non-epileptic) and ternary (ictal vs. normal vs. interictal) scenarios is given in Sharmila and Geethanjali (2016) and Zhang et al. (2017).

The overview of the state-of-the-art given above indicates that most of the feature extraction techniques are hand-engineered, which do not extract the discriminative information from EEG signals through learning the internal structures of data; their performance depends on the tuning of various parameters and do not generalize well. In order to improve the accuracy and generalization of an epilepsy detection system, DL approach can be used to avoid the need for hand-engineered feature extractors and classifiers. To the best of our knowledge, so far no one used DL approach for epilepsy detection, perhaps the reason is the small amount of available data, which is not enough to train a deep model. As such, we felt motivated to employ DL technique for proposing a deep model that involves a small number of learnable parameters and classifies efficiently EEG brain signals as epileptic or non-epileptic. However, DL has been recently applied for similar problems. Acharya et al. (2017a,b,c) recently applied DL for arrhythmia, myocardial infraction and coronary artery detection from ECG signals. They used a deep convolutional neural network model with eleven layers. In Acharya et al. (2017a), the authors proposed two CNN models: A and B. Model A takes a window of ECG signal consisting of 500 samples as input whereas the input to Model B is a window of size 1250 samples. Our CNN model is different from these models from two aspects: (i) it involves fewer number of layers (only 5 layers), (ii) it is based on pyramid architecture, which reduces the number of parameters significantly. Our system is an ensemble of CNN models, which analyze different parts of an input signal and in this way add diversity and results in better detection performance.

3. The proposed system

The proposed automatic system for epilepsy detection using EEG brain signals based on an ensemble of P-1D-CNN models is shown in Fig. 1. It consists of three main modules: (i) input module, which takes an input EEG signal, splits it into sub-signals using fixed-size overlapping windows and passes them to base P-1D-CNN models, (ii) an ensemble where the sub-signals are classified by base P-1D-CNN models, and (iii) fusion and decision, the local decisions are fused using majority vote to take the final decision.

A standard deep CNN model needs a huge amount of data for training, but for epilepsy detection problem, the amount of data is limited. To tackle this issue, we introduce data augmentation schemes in Section 4 that are used to create data for training base P-1D-CNN models.

Once a P-1D-CNN model is trained, using its copies as base models, we build a deep ensemble classifier, where each based model plays the role of an expert examining a different part of the input signal. When an input signal is passed to the ensemble for classification, it is split into overlapping windows for the sake

Table 1
Key terms and their acronyms that will be used throughout the paper.

Name	Abbreviation	Name	Abbreviation
Accuracy	*Acc*	No of Kernels	*K*
Accuracy with Voting	*Acc_V*	Size of Receptive field	*Rf*
Fully connected	*FC*	Batch Normalization	*BN*
Rectifier Linear activation Unit	*ReLU*	Dropout	*DO*
Specificity	*Spe*	Sensitivity	*Sen*
Geometric Mean	*G_M*	F-Measure	*F_M*
10-fold Validation	K-number	Standard Deviation	*std*

of diversity (keeping in view the augmentation approach), which are passed to different base P-1D-CNN models in the ensemble, as shown in Fig. 1, i.e. different parts of the signal are assigned to different experts (base models) for its local analysis. After local analysis, each model provides a local decision; lastly, these decisions are fused using majority vote for final decision. The number of P-1D-CNN models (experts) in the ensemble depends on the number of windows. For example, if an input EEG signal is divided into n windows (sub-signals), the ensemble will consist of n base P-1D-CNN models. We tested the system with $n = 3$ and $n = 5$ and found that $n = 3$ gives better results.

The core component of the system is a P-1D-CNN model. It is a deep model, which consists of three main types of layers: convolution (*Conv*), batch normalization (*BN*), and fully connected (*FC*) layers. For *Conv* and *FC* layers, ReLU is used as an activation function; dropout technique is used for regularization during training. The ReLU and dropout are applied using separate layers and accordingly are shown in the model. In the following section, we present the detail of this deep model. For compactly describing the ideas, key terms and their acronyms are given in Table 1.

3.1. P-1D-CNN architecture

A deep CNN model (LeCun et al., 1998; Simonyan & Zisserman, 2014) learns structures of EEG signals from data automatically and performs classification in an end-to-end manner, which is opposite to the traditional hand-engineered approach, where first features are extracted, a subset of extracted features is selected and finally passed to a classifier for classification. The main component of a CNN model is a *Conv* layer, which consists of many channels (feature maps). The output of each neuron in a channel is the outcome of a convolution operation with a kernel (which is shared by all neurons in the same channel) of a fixed receptive field on the input signal or feature maps (1D signals) of the previous *Conv* layer. In this way, CNN analyses a signal to learn a hierarchy of discriminative information. In CNN, the kernels are learned from data unlike hand-engineered approach, where kernels are predefined e.g. wavelet transform.

CNN with its novel idea of shared kernels has the advantage of a significant reduction in the number of parameters over fully connected architectures. The recent emergence of making CNN deeper has given rise to a very large number of parameters that add to its complexity and are potential cause of overfitting when the available dataset is small. Available EEG dataset for epilepsy detection is small in size, we handled this problem using two different strategies i.e. novel data augmentation schemes and a memory efficient deep pyramidal CNN model that involves a small number of parameters.

EEG signal is a 1D time series; as such for its analysis, we propose a pyramidal 1D-CNN model, which we call P-1D-CNN and its generic architecture is shown in Fig. 2, it is an end-to-end model. Unlike traditional CNN models, it does not include any pooling layer; the redundant or unnecessary features are reduced with the help of bigger strides in *Conv* layers. The *Conv* and *FC* layers learn a hierarchy of low to high-level features from the given input signal. The high-level features with semantic representation are passed as input to the softmax classifier in the last layer to predict the respective class of the input EEG signal.

A CNN model is commonly structured by adopting course to fine approach, where low-level layers have a small number of kernels, and high-level layers contain a large number of kernels. But this structure involves a huge number of learnable parameters i.e. its complexity is high. Instead, we adopted a pyramid architecture similar to the one proposed by Ullah and Petrosino (2016) for deep 2D CNN, where low-level layers have a large number of kernels and higher level layers contain a small number of kernels. This structure significantly reduces the number of learnable parameters, avoiding the risk of overfitting. A large number of kernels are taken in *Conv-1* layer, which are reduced by a constant number in *Conv-2* and *Conv-3* layers e.g. models M5 and M6, specified in Table 3, contain *Conv-1, Conv-2* and *Conv-3* layers with 24, 16, and 8 kernels, respectively. The idea is that low-level layers extract a large number of microstructures, which are combined by higher level layers into higher level features, which are small in number but discriminative, as the network gets deeper i.e. this model does the feature selection implicitly.

To show the effectiveness of P-1D-CNN model, we considered eight models with different configurations, four of which are based on pyramid architecture. Table 3 shows detailed specifications of these models and also gives the number of parameters to be trained in each model. The last fully connected layer has two or three neurons depending on whether the EEG brain signal classification problem is two class (e.g. epileptic and non-epileptic) or three class (normal vs. ictal vs. interictal). With the help of these models, we show how a properly designed model can result in equal or better performance despite fewer parameters, which has less risk of overfitting. The models based on pyramid architecture involve significantly fewer number of learnable parameters, see Table 3; for example, model M5, which has pyramid architecture, has 61% fewer parameters than M1, a similar 1D-CNN model.

The detail of a deep P-1D-CNN model (M5) is shown in Fig. 2. The input signal is normalized to zero mean and unit variance i.e. using z-score normalization. This normalization helps in faster convergence and avoiding local minima. The normalized input is processed by three convolutional blocks, where each block consists of three layers: *Conv* layer, Batch normalization layer (*BN*) and non-linear activation layer (*ReLU*). The number of kernels for *Conv-1* is 24 and receptive field of each kernel is 5 (i.e.1×5); the number of kernels for *Conv-2* is 16, and receptive field of each kernel is 3 (i.e.1×3) and depth is 24; the number of kernels for *Conv-3* is 8 and receptive field of each kernel is 3 (i.e. 1×3) and the depth is 16. The output of the third block is passed to the first FC layer (*FC*1) that is followed by a *ReLU* layer and another FC layer (*FC*2). The number of neurons in *FC1* is 20. In order to avoid overfitting, we use dropout layer before *FC*2. The output of *FC*2 is given to a softmax layer, which serves as a classifier and predicts the class of the input signal. The number of neurons in FC2 is either 2 or 3 depending on the number of classes. At test time, the model

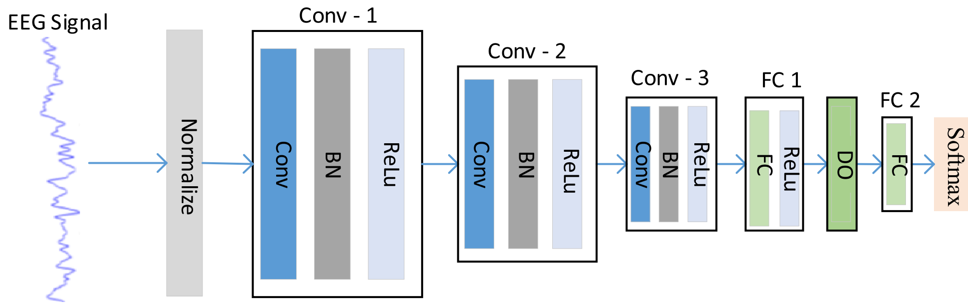

Fig. 2. The proposed Deep Pyramidal 1D-CNN Architecture (P-1D-CNN).

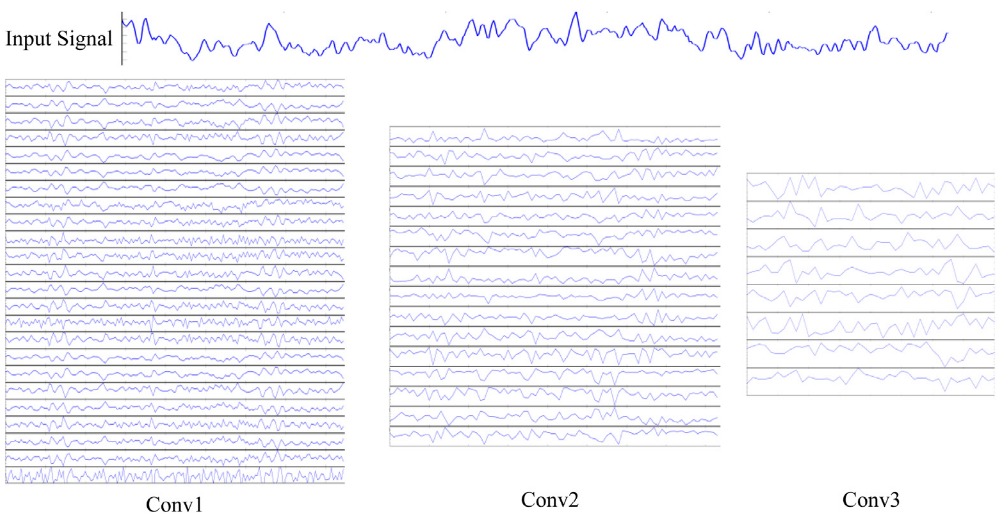

Fig. 3. Input Signal (first row), and activations of *Conv-1* (24 channels), *Conv-2* (6 channels) and *Conv-3* (08 channels) of P-1D-CNN model.

does not use BN and DO. The specifications of other models are given in Table 3. In the following subsections, we will briefly explain the main layers i.e. 1D-Conv, BN and FC layers. The details can be found in LeCun et al. (1998) and Ioffe and Szegedy (2015).

a) Convolution Layers

The 1D convolution operation is used to filter 1D signals (e.g. time series) for extracting discriminative features. A *Conv* layer is generated by convolving the previous layer with K kernels of receptive field *Rf* and depth c that is equal to the number of channels or feature maps in the previous layer. Formally, convolving the layer $X=\{x_{ij}: 1 \le i \le c, 1 \le j \le z\}$, where c is the number of channels in the layer and z is the number of neurons in each channel, with K kernels $k^l, l=1, 2, \ldots, K$ each of receptive field $Rf$and depth c yield the convolution layer $Y=\{y_{lm}: 1 \le l \le K, 1 \le m \le K\}$, where

$$y_{lm} = \sum_{d=1}^{c}\sum_{e=1}^{Rf} k^l_{d,e}\, x_{d,\ e+m}, \tag{1}$$

m is the number of neurons in each channel of the layer and K is the total number of channels in the layer. Note that the number of channels in the generated *Conv* layer is equal to the number of kernels. Different kernels extract different types of discriminative features from the input signal. The number of kernels varies as the network goes deeper and deeper. The low-level layer kernels learn micro-structures whereas the higher level layer kernels learn high-level features. In the proposed model, maximum number of kernels are selected in first *Conv* layer that is reduced by 33% in subsequent layers to maintain a pyramid structure. The activations (channels) of three *Conv* layers are shown in Fig. 3.

a) Batch Normalization

During training, the distribution of feature maps changes due to the update of parameters, which forces to choose small learning rate and careful parameter initialization. It slows down the learning and makes the learning harder with saturating nonlinearities. Ioffe and Szegedy (2015) called this phenomenon as internal covariate shift and proposed batch normalization (BN) as a solution to this problem. In BN, the activations of each mini batch at each layer are normalized, the detail can be found in Ioffe and Szegedy (2015). It is now very common to use BN in neural networks. It helps in avoiding special initialization of parameters, yet provides faster convergence. In the proposed model, we use *BN* after every Conv layer during training only.

a) Fully Connected

After convolutional layers, each model has two fully connected (FC) layers. All the neurons in *Conv3* layer are connected to each neuron is the first fully connected layer FC1. In different models, the number of neurons in FC1 different, the detail is given in Table 3. The second fully connected layer has either 2 or 3 neurons depending on the detection problem, e.g. for the problem normal vs epileptic, which is a two class problem, the number of neurons in FC2 is 2 and for the problem normal vs inter-ictal vs ictal, which is a three class problem, FC2 contains 3 neurons.

4. Model selection and parameter tuning

First, we present the detail of data, and the proposed data augmentation schemes. Then, we give evaluation measures that were used to validate the performance of the proposed system. After this, the training procedure is elaborated. Finally, the best data

Table 2
University of Bonn epilepsy dataset details.

A	B	C	D	E
Non-Epileptic Eyes opened	Non-Epileptic Eyes closed	Epileptic Interictal	Epileptic Interictal	Epileptic Ictal

augmentation scheme and P-1D-CNN model are figured out by analyzing the results with different ways of data augmentation, and different 1D-CNN models.

4.1. Dataset and data augmentation schemes

The data set used in this work was acquired by a research team at University of Bonn (Andrzejak et al., 2001) and have been extensively used for research on epilepsy detection. The EEG signals were recorded using standard 10–20 electrode placement system. The complete data consists of five sets (A to E), each containing 100 one-channel instances. Sets A and B consists of EEG signals recorded from five healthy volunteers while they were in a relaxed and awake state with eyes opened (A) and eyes closed (B), respectively. Sets C, D, and E were recorded from five patients. EEG signals in set D were taken from the epileptogenic zone. Set C was recorded from the hippocampal formation of opposite hemisphere of the brain. Sets C and D consist of EEG signals measured during seizure-free intervals (interictal), whereas, the EEG signals in Set E were recorded only during seizure activity (ictal) (Andrzejak et al., 2001). The detail is given in Table 2.

The number of instances collected in this dataset are not enough to train an efficient deep model. Acquiring a large number of EEG signals for this problem is not practical and their labeling by expert neurologists is not an easy task. We need an augmentation scheme that can help us in increasing the amount of the data that is enough for training a deep generalized CNN model, which requires large training data for better performance. The available EEG data is small that can learn the model but overfitting is evident. To overcome this problem, we propose two data augmentation schemes for training our model. The data is augmented by splitting the given full length EEG signals into small signals using a fixed size window; each small signal is used as an independent instance for learning CNN models. To split an EEG signal into small signals is a standard procedure which has been adopted in the existing methods (Sharmila & Geethanjali, 2016; Zhang et al., 2017, Zhang et al., 2017).

Each record in the University of Bonn dataset consists of 4097 samples. For generating many instances from one record, we adopted the sliding window approach similar to the one presented in Refs. Sharmila and Geethanjali (2016) and Zhang et al. (2017). In Zhang et al. (2017), the authors adopted a window size of 512 with a stride of 480 (93.75% of 512); each record is segmented into 8 equal EEG sub-signals, discarding the last samples. In this way, a total of 800 data instances are obtained for each dataset from 100 single-channel records, but this amount is not enough for learning the deep model. However, this approach indicates that the large stride is not helpful and smaller strides can be used for creating enough data. Based on the window size and stride, we propose two data augmentation schemes.

Scheme-1

The available signals are divided into disjoint training and testing sets, which consist of 90% and 10% of total signals, respectively. Data is augmented using training set. Choosing a window size of 512 and a stride of 64 (12.5% of 512 with an overlap of 87.5%), each signal of length 4097 in the training set is divided into 57 sub-signals, each of which is treated as an independent signal instance S^{tr}. In this way, a total of 5130 instances are created for each category (class), which are used to train the P-ID-CNN model.

For testing, each signal of length 4097 in the testing set is divided into 4 sub-signals S^{ts}, each of length 1024; these sub-signals are treated as independent signal instances for testing. When a signal instance $S^{ts}$of length 1024 is passed to the system, it is divided into three sub-signals with a window of size 512 and 50% overlap i.e. $S_i^{ts},\ i = 1,\ 2,\ 3$, each of size 512, which are passed to three trained base P-ID-CNN models in the ensemble and majority vote is used as a fusion strategy to take the decision about the input signal instance S^{ts}. Each base model in the ensemble plays the role of an expert, which analyses a local part of the signal instance S^{ts} independently and the global decision is given by the system with the help of fusing the local decisions.

Scheme-2

This method is similar to scheme-1. In this case, the window size is 512 with an overlap of 25% (i.e. stride of 128) for creating training instances S^{tr}.

For testing, when an input signal instance S^{ts} of length 1024 is passed to the system, it is divided into five sub-signals with a window of size 512 and 75% overlap i.e. $S_i^{ts},\ i = 1,\ 2,\ 3,\ 4,\ 5$, each of size 512, which are passed to five trained base P-ID-CNN models in the ensemble and majority vote is used as a fusion strategy to take the decision about the signal instance S^{ts}.

4.2. Performance measures (evaluation procedure)

For evaluation, we adopted 10-fold cross validation for ensuring that the system is tested over different variations of data. The 100 signals for each class divided into 10 folds, each fold (10%), in turn, is kept for testing while the remaining 9 folds (90% signals) are used for learning the model. For each fold, training examples are created using the proposed augmentation schemes and 90% training data, and the testing examples are created from the held-out 10% testing data. The average performance is calculated for 10 folds. The performance was evaluated using well-known performance metrics such as accuracy, specificity, sensitivity, precision, f-measure, and g-mean. Most of the state-of-the-art systems for epilepsy also employ these metrics, the adaptation of these metrics for evaluating our system helps in fair comparison with state-of-the-art systems. The definitions of these metrics are given below:

$$Accuracy\ (Acc) = \frac{TP\ + TN}{Total\ Samples} \tag{2}$$

$$Specificity\ (Spe) = \frac{TN}{TN\ +\ FP} \tag{3}$$

$$Sensitivity\ (Sen) = \frac{TP}{FN\ +\ TP} \tag{4}$$

$$F-Measure\ (F\_M)) = \frac{2\ *\ Precision\ *\ Sensitivity}{Precision\ +\ Sensitivity} \tag{5}$$

$$G-Mean\ (G\_M)\ =\ \sqrt{Specificity * Sensitivity} \tag{6}$$

where *TP* (true positives) is the number of abnormal cases (e.g. epileptic), which are predicted as abnormal, *FN* (false negatives) is the number of abnormal cases, which are predicted as normal, *TN* (true negatives) is the number of normal case that is predicted as normal and *FP* (false positives) is the number of normal cases that are identified as abnormal by the system.

Training of P-1D-CNN Model

Training of P-1D-CNN needs the weight parameters (kernels) to be learned from the data. For learning these parameters, we used the traditional back-propagation technique with cross entropy loss function and stochastic gradient descent approach with Adam optimizer (Kingma & Ba, 2014). Adam algorithm has six hyper-parameters: learning rate (0.001), beta1 (0.9), beta2 (0.999),

Table 3
The specifications of 8 1D-CNN models and their mean performance using 10-fold cross-validation for the AB vs. CD vs. E case. The (-) means this layer is an operator laery.

Layers\Model		*M*1	*M*2	*M*3	*M*4	*M*5	*M*6	*M*7	*M*8	
*Conv*1	*K*	8	8	8	8	24	24	24	24	
	Rf	5	5	5	5	5	5	5	5	
	St	3	3	3	3	3	3	3	3	
BN	–	–	–	–	–	–	–	–	–	
ReLU	–	–	–	–	–	–	–	–	–	
*Conv*2	*K*	16	16	16	16	16	16	16	16	
	Rf	3	3	3	3	3	3	3	3	
	St	2	2	2	2	2	2	2	2	
BN	–	–	–	–	–	–	–	–	–	
ReLU	–	–	–	–	–	–	–	–	–	
*Conv*3	*K*	24	24	24	24	8	8	8	8	
	Rf	3	3	3	3	3	3	3	3	
	St	2	2	2	2	2	2	2	2	
BN	–	–	–	–	–	–	–	–	–	
ReLU	–	–	–	–	–	–	–	–	–	
*FC*1	–	20	20	40	40	20	20	40	40	
DO	–	0	0.5	0	0.5	0	0.5	0	0.5	
*FC*2 (Out)	–	2/3	2/3	2/3	2/3	2/3	2/3	2/3	2/3	
Parameters		21,366/21,387	41,106/41,147	**8326/8347**	**14,946/14,987**					
										Avg ± std
AB vs CD vs E Aug. Scheme-1	*Acc*	96.23	96.00	96.03	95.92	96.27	96.18	96.12	96.45	96.45 ± 0.13
	Acc_V	99.10	98.95	98.95	99.15	99.10	98.95	99.05	98.75	99.00 ± 0.08
	std	0.02	0.01	0.01	0.01	0.02	0.01	0.02	0.02	
	Sen	0.96	0.96	0.96	0.96	0.95	0.96	0.96	0.97	0.960 ± 0.003
	Spe	0.98	0.98	0.97	0.98	0.98	0.97	0.98	0.96	0.975 ± 0.004
	$G-M$	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.968 ± 0.000
	$F-M$	0.95	0.96	0.95	0.96	0.96	0.96	0.96	0.95	0.956 ± 0.004
AB vs CD vs E Aug. Scheme-2	*Acc*	94.88	95.78	95.10	95.55	94.95	95.67	95.28	96.00	95.40 ± 0.35
	Acc_V	98.85	98.90	99.00	98.85	99.00	99.05	98.85	98.95	98.93 ± 0.08
	std	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.02	
	Sen	0.95	0.97	0.95	0.96	0.96	0.96	0.95	0.97	0.958 ± 0.007
	Spe	0.97	0.97	0.98	0.98	0.97	0.97	0.98	0.97	0.973 ± 0.005
	$G-M$	0.96	0.97	0.96	0.97	0.96	0.97	0.96	0.97	0.965 ± 0.005
	$F-M$	0.96	0.97	0.96	0.97	0.96	0.97	0.96	0.97	0.965 ± 0.005

epsilon (0.00000001), use locking (false) and name (Adam); we used default values of all these parameters (given in parentheses) except learning rate, which we set to a very small number of 0.00002. Although BN normally allows higher learning rate, a small learning rate is needed to control the oscillation of the network and to avoid any local minima problem when using Adam optimizer. The model is trained with a different number of iterations depending on the size of the dataset. In dropout, a probability value of 0.5 is used in all the experiments. The model was implemented in TensorFlow (TensorFlow, 2017), a freely available DL library from Google. The number of iterations varies for each experiment–depending on the number of datasets used in the experiment. For example, when we used two datasets i.e. A vs E or D vs E, we trained the model with 50k iterations; when we used three sets among five (i.e. A, B, C, D, or E) in an experiment e.g. AB vs C, we set maximum iterations to 150k. Whereas, when we are used four or all of the five available signal sets, we trained our model with 300k iterations. Although the model trains much faster, still we train it to a maximum number of assigned iterations for better generalization of the model.

4.2.1. Selection of best model and data augmentation scheme

For selecting the best model, we considered eight CNN models in our initial experiments, as is shown in Table 3. For best model selection, we need to address two questions: (a) which data augmentation scheme is the most suitable one? (b) does pyramidal architecture result in better generalization than the traditional model, where the number of kernels increases as the network goes deeper and deeper? To answer these questions, we performed in-depth experiments using 10-fold cross validation with all the eight models only on three class problem: non-epileptic (AB) vs epileptic inter-ictal (CD) vs epileptic ictal (E), which is the most challenging problem. These experiments led us to select the best model and the data augmentation scheme, which we used for other classification problems. It should be noted that all the 10-fold cross-validation sets are created randomly forcing to include all samples in training (90%) and testing (10%).

The models were trained and tested using data augmentation schemes 1 and 2. Models *M*1 to *M*4 are designed using the traditional concept of increasing *K* (the number of filters or kernels) in each higher layer as the network goes deeper, whereas models *M*5 to *M*8 (pyramid models) are designed using the concept of course to fine refinement approach i.e. reduce *K* (the number of filters or kernels) by ratio of 33% in this case as the network goes deeper. The pyramid models involve a fewer number of parameters than traditional models, and as such are less prone to overfitting and generalize well.

The average performance results obtained using 10-fold cross-validation of different models and the data augmentation schemes are given in Table 3. First, the average accuracies (over all models) along with their standard deviations are 96.45 ± 0.13 and 95.40 ± 0.35 using data augmentation schemes 1 and 2, respectively; almost similar results can be observed in terms of other performance measures. It indicates that augmentation scheme 1 results in better performance than scheme 2. Based on this observation, scheme 1 is adopted for all other experiments in the paper.

Secondly, based on overall results it can be observed that pyramid model (*M*5 to *M*8) show results, which are better than or equal to those by traditional models with both augmentation schemes. Further, in most of the cases, the best result is given by pyramid model *M*5 with dropout 0.5 and 20 neurons in the fully connected

Table 4
The (%) accuracies of three-class problem (AB vs CD vs E) using model M5 and 10-fold cross validation.

Fold	$K1$	$K2$	$K3$	$K4$	$K5$	$K6$	$K7$	$K8$	$K9$	$K10$	Mean Acc
Acc	96.5	97.2	96.7	94.7	97.8	96.7	95.5	96.2	96	93.8	96.1
Acc_V	99	100	99	97	100	100	100	98	99	99	**99.1**

Table 5
Confusion matrix for the three-class problem (AB vs CD vs E) using model M5, the values are the mean numbers of classified signals in 10-folds.

	Normal (AB)	Interictal (CD)	Ictal (E)
Normal (AB)	234	5	1
Interictal (CD)	23	217	0
Ictal (E)	0	5	115

layer; it works better with 20 neurons rather than 40 in the fully connected layer. It is obvious that *M5* is the best model, its gives slightly higher or similar performance but involves the minimum number of parameters among all; such a model is easy to deploy on low-cost chips with limited memory as compared to the models with more parameters ($M1 - M4$). In all onward experiments, we use model $M5$ with augmentation scheme 1.

5. Results

After model selection, i.e. *M5* with augmentation scheme 1, we present and discuss the results for different experiment cases related to epilepsy detection. We considered three experiment cases: (i) normal vs inter-ictal vs ictal (AB vs CD vs E), (ii) normal vs epileptic (AB vs CDE and AB vs CD), (iii) seizure vs non-seizure (A vs E, B vs E, A+B vs E, C vs E, D vs E, C+D vs E). The comparison is done with state-of-the-art on 16 experiments: AB vs. CD vs. E, AB vs. CD, AB vs. E, A vs. E, B vs. E, CD vs. E, C vs. E, D vs. E, BCD vs. E, BC vs. E, BD vs. E, AC vs. E, ABCD vs. E, AB vs. CDE, ABC vs. E and ACD vs. E. Among 16 experiments, 14 have been frequently considered in most of the studies e.g. Sharmila and Geethanjali (2016). The remaining 2 experiments have rarely or never been tested. All experiments were performed using 10-fold cross validation.

5.1. Experiment 1: normal vs ictal vs interictal classification (AB vs CD vs e)

Zhang et al. (2017) pointed out that almost 100% accuracy was achieved by several recent research works for normal vs epileptic or non-seizure vs seizure EEG signals classification. However, less work has been devoted to normal vs interictal vs ictal signals classification. They proposed a system targeting specifically for this three-class problem, which achieved an accuracy of 97.35%.

Using M5 model, we achieved a mean accuracy of 96.1% with single P-1D-CNN model and 99.1% with an ensemble of 3 P-1D-CNN models, outperforming (Zhang et al., 2017) by 1.7% and (Bhattacharyya, Pachori, Upadhyay, & Acharya, 2017) by 0.5%. The detailed analysis of the performance for this problem is given in Table. 3, which shows the average results for all the models and the augmentation schemes. However, Tables 4 and 5 show the 10-fold cross-validation results and the confusion matrix for this problem. Table 5 indicates that the main confusion arises between normal and inter-ictal or inter-ictal and ictal.

5.2. Experiment 2: normal vs epileptic classification (AB vs CDE and AB vs CD)

This case involves two types of experiments involving binary classification problems: (i) normal (AB) vs non-seizure epileptic (CD), and (ii) normal (AB) vs non-seizure and seizure epileptic (CDE); the 10-fold cross-validation results are shown in Table. 6. The mean accuracy of the proposed system for AB vs CD is 98.2% with single P-1D-CNN model, while 99.8% with the ensemble of 3 P-1D-CNN models. Similarly, the mean sensitivity and specificity are 98% and 99%, respectively. In the case of AB vs CDE, the mean accuracies are 98.1% and 99.95% with single model and ensemble, respectively, whereas both the mean sensitivity and specificity are 98%. The results indicate that the proposed system has better generalization and outperforms the state-of-the-art method reported in Sharma, Pachori, and Acharya (2017) and Sharmila and Geethanjali (2016). Also, it points out that ensemble of P-1D-CNN models performs better than single P-1D-CNN model, the reason is that in ensemble each model works as an expert, which analyses a local part of the signal, and finally local decisions are fused using majority vote to take the final decision.

5.3. Experiment 3: normal or non-seizure vs seizure classification (A vs E, B vs E, A+B vs E, C vs E, D vs E, C+D vs E)

Third set of experiments involves six binary class problems ((i) normal (A) vs seizure (E), (ii) normal (B) vs seizure (E), (iii) normal (AB) vs seizure (E), (iv) non- seizure (C) vs seizure (E), (v) non-seizure (D) vs seizure (E), and (vi) non- seizure (CD) vs seizure (E)). We tested all these combinations in order to check the robustness of the proposed system. Table. 7 reports the results. The mean accuracies given by single P-1D-CNN model varies from 99.9% to 97.4, whereas those given by ensemble varies from 100% to 98.5% for all the above problems. For all normal vs seizure problems, the accuracy is almost 100% with ensemble. For the problem C vs E, mean accuracy is 98.1% with single P-1D-CNN model and 98.5% with ensemble; in this case, there is a little improvement with ensemble, it indicates that in this case, almost all experts (P-1D-CNN models)

Table 6
Performance Results of Normal vs Epileptic case using model $M5$ with 10-fold cross validation, here Ki means ith fold and all results are in (%).

		$K1$	$K2$	$K3$	$K4$	$K5$	$K6$	$K7$	$K8$	$K9$	$K10$	Mean
AB vs CD	*Acc*	99.2	96.3	98.8	97.9	98.3	99.6	97.3	99.2	96.9	98.5	98.2
	Sen	100	95	98	97	98	100	95	100	96	99	98
	Spe	98	97	100	99	98	99	99	99	98	98	99
	Acc_V	100	100	99.4	100	100	100	98.8	100	99.4	100	**99.8**
AB vs CDE	*Acc*	95.8	99.3	98.5	99.7	96.8	96.7	98.7	99.2	97.8	98.2	98.1
	Sen	98	99	99	100	96	99	99	99	98	97	98
	Spe	92	100	98	99	98	93	98	99	98	100	98
	Acc_V	99.5	100	100	100	100	100	100	100	100	100	**99.95**

Table 7
The (%) accuracies of 10-folds for Normal or Seizure vs Non-Seizure using model $M5$.

		$K1$	$K2$	$K3$	$K4$	$K5$	$K6$	$K7$	$K8$	$K9$	$K10$	Mean
A vs E	*Acc*	99.2	100	100	100	100	100	100	99.6	100	100	99.9
	Acc_V	100	100	100	100	100	100	100	100	100	100	100
B vs E	*Acc*	100	95.4	98.8	100	97.1	98.8	100	100	100	100	99
	Acc_V	100	97.5	100	100	98.8	100	100	100	100	100	99.6
AB vs E	*Acc*	99.2	96.4	97.5	98.9	96.7	100	100	99.7	100	100	98.8
	Acc_V	100	99.2	98.3	100	99.2	100	100	100	100	100	99.7
C vs E	*Acc*	95	99.2	99.6	92.9	100	99.2	100	100	97.9	97.1	98.1
	Acc_V	95	98.8	100	93.8	100	100	100	100	100	97.5	98.5
D vs E	*Acc*	98.8	100	97.5	99.6	93.8	94.6	97.9	97.9	98.8	95	97.4
	Acc_V	100	100	100	100	98.8	98.8	100	100	100	95	99.3
CD vs E	*Acc*	99.4	98.9	100	99.4	97.5	100	99.2	98.9	98.1	96.1	98.8
	Acc_V	100	100	100	100	100	100	100	100	100	96.7	99.7

Table. 8
Performance Comparison of the proposed system based on model M5 with Scheme 1 and 10-fold cross validation.

Data Sets Combination	Methodology	10-fold CV	Stat-of-the-Art	Acc(%)	Our Acc (%)
AB vs CD vs E	VMD+AR+RFTQWT	Yes	Zhang-17	97.4	**99.1**
	+KNN-Entropy	Yes	**Bhattacharyya –17**	98.6	
AB vs CD	ATFFWT+LS-SVM	Yes	Sharma-17	92.5	**99.9**
AB vs E	ATFFWT+LS-SVM	Yes	Sharma-17	**100**	99.8
	DTCWT+GRNN	Yes	Swami-16	99.2	
A vs E	DWT+NB/K-NN	No	Sharmila-16	100	**100**
	TF+ANN	Yes	Tzallas-12	100	
	DTCWT+GRNN	Yes	Swami-16	100	
	TQWT+KNN-Entropy	Yes	Bhattacharyya –17	100	
B vs E	ATFFWT+LS-SVM	Yes	Sharma-17	**100**	99.8
	DTCWT+GRNN	Yes	Swami-16	98.9	
	TQWT+KNN-Entropy	Yes	Bhattacharyya –17	**100**	
CD vs E	DWT+NB/K-NN	No	Sharmila-16	98.8	**99.7**
	ATFFWT+LS-SVM	Yes	Sharma-17	98.7	
	DTCWT+GRNN	Yes	Swami-16	95.2	
C vs E	ATFFWT+LS-SVM	Yes	Sharma-17	99	99.1
	DTCWT+GRNN	Yes	Swami-16	98.7	
	TQWT+KNN-Entropy	Yes	Bhattacharyya –17	**99.5**	
D vs E	ATFFWT+LS-SVM	Yes	Sharma-17	98.5	**99.4**
	DTCWT+GRNN	Yes	Swami-16	93.3	
	TQWT+KNN-Entropy	Yes	Bhattacharyya –17	98	
BCD vs E	DWT+NB/K-NN	No	Sharmila-16	96.4	**99.3**
BC vs E	DWT+NB/K-NN	No	Sharmila-16	98.3	**99.5**
BD vs E	DWT+NB/K-NN	No	Sharmila-16	96.5	**99.6**
AC vs E	DWT+NB/K-NN	No	Sharmila-16	99.6	**99.7**
ABCD vs E	ATFFWT+LS-SVM	Yes	Sharma-17	99.2	**99.7**
	TF+ANN	Yes	Tzallas-12	97.7	
	DWT+MLP	No	Orhan-11	99.6	
	FT+MLP	No	Samiee-15	98.1	
	DTCWT+GRNN	Yes	Swami-16	95.2	
	TQWT+KNN-Entropy	Yes	Bhattacharyya –17	99	
AB vs CDE	DWT+NB/K-NN	No	Sharmila-16	–	**99.5**
ABC vs E	DWT+NB/K-NN	No	Sharmila-16	98.7	**99.97**
ACD vs E	DWT+NB/K-NN	No	Sharmila-16	97.3	**99.8**

have the same decision and it does not have a significant impact. For other two non-seizure vs seizure problems, the mean accuracies are 99.3% and 99.7%, which show that these are relatively easier problems than C vs E.

6. Discussion

Many methods have been proposed for the classification of EEG signals in binary (Normal vs Epileptic and seizure vs non-seizure) and ternary (Normal vs Interictal vs Ictal) classification problems. A comparison with state-of-the-art methods is given in Table 8: Zhang-17 ((Zhang et al., 2017), Bhattacharyya-17 (Bhattacharyya et al., 2017), Sharma-17 (Sharma et al., 2017), Swami-16 (Swami et al., 2016), Sharmila-16 (Sharmila & Geethanjali, 2016), Samiee-15 (Samiee et al., 2015), Orhan-11 (Orhan et al, 2011), Tzallas-12 (Tzallas et al., 2012). According to our knowledge, DL approach has not been used for this problem so far. Recently, a fusion technique using variational mode decomposition (VMD) and an auto-regression based quadratic feature extraction technique have been proposed in Zhang et al. (2017). Random forest classifier has been used to classify the extracted features into three categories. Despite using multiple complex techniques, it achieved 97.35% accuracy for three class problem, our system achieves 1.7% higher accuracy i.e. 99.1%.

The technique proposed in Bhattacharyya et al. (2017) computes the quality factor (Q) based multi-scale entropy measures using tunable-Q wavelet transform (TQWT) and uses as features. The performance of the method is based on the tuning of Q and the redundancy parameter (R) of TQWT. It was tested on A vs. E, B vs. E, C vs. E, D vs. E, ABCD vs. E and AB vs. CD vs. E. Our method outperforms it in D vs. E, ABCD vs. E, and AB vs. CD vs. E. However, in two cases B vs. E and C vs. E, it gives better results, but the difference is not significant, see Table 8. The downside of this method is that it needs manual tuning of parameters and is data dependent.

The method proposed in Sharmila and Geethanjali (2016) employs discrete wavelet transform (DWT) for feature extraction and naïve Bayes (NB) and k-nearest neighbors (k-NN) for the classification of epileptic and non-epileptic signals. The results reported for this method are without 10-fold cross validation. Table 8 shows that the proposed system evaluated with 10-fold cross validation overall outperforms the best performing methods (Sharmila & Geethanjali, 2016; Bhattacharyya et al., 2017). It is interesting to observe that the methods by Sharmila and Geethanjali (2016) and Bhattacharyya et al. (2017) are not consistent i.e. there is significant difference in their performance for different cases. It indicates that these systems do not generalize well over different cases and depend on the data. On the other hand, the proposed system shows consistent performance on all cases i.e. for different cases, the accuracies varies from 99.1% to 99.97%, the slight variation is due to the nature of problems; it implies that it does not heavily depend on data, is robust and has better generalization than state-of-the-art methods. The mean accuracy of the proposed system is 99.6% for all the sixteen cases (shown in Table 8 last column), which validates the generalization power of the proposed system.

All existing systems are based on hand-engineered feature extraction techniques, which need the tuning of parameters, their performance heavily depends on the selection of hyper-parameters and the data; they do not learn the internal structure of the data. As such, they do not generalize well across different datasets i.e. different cases. In addition, they involve laborious designs i.e. first features are extracted and selected and then passed to a classifier, all these stages involve hyper-parameters whose joint tuning is laborious. In contrast, the proposed system is an end-to-end system, which is based on the deep learning theory; it takes input signal and gives the decision; there is no need of any kind of signal preprocessing, manual feature extraction and selection and laborious parameter tuning. It learns the discriminative information automatically from the data and the learning process is fully automatic. The only weakness of the proposed system in comparison to traditional systems based on hand-engineered techniques is that the learned model must be kept all the times. It is to be noted that our design requires minimum memory space as compared other CNN models. The P-1D-CNN models based on pyramid design (M5, M6, M7) involve the lowest number of parameters as compared to similar standard CNN models for three class case. The best pyramid based P-1D-CNN model (M5) contains 8347, which is 39% of 21,387 parameters contained in similar standard CNN model (M1) i.e. M5 contains at least 61% fewer parameters. The small number of parameters not only ensures better generalization, but also results in less memory overhead.

We trained P-1D-CNN model on a laptop with Intel Core i7-6700HQ CPU @ 2.60 GHz having 16 GB RAM, 4 GB Nvidia Geforce GTX 965 M Graphics Card. For training P-1D-CNN model, we used 22,400 EEG signals, each consisting of 512 samples; one epoch (training and validation) took 4.33 s. In the testing scenario, the system accepts an EEG signal of 1024 samples as an input instance and predicts its class. The prediction time of one instance is 0.000142 s with GPU acceleration whereas 0.000481 s without GPU acceleration for three class case AB vs. CD vs. E. This indicates that the proposed system is suitable for deployment in small FGPAs; the only weakness of the system in this scenario can be the memory and storage requirements. Due to high accuracy and real-time performance, the system can be deployed in a clinical setting to assist neuroscientists.

7. Conclusion

An automatic system for epilepsy detection has been proposed, which deals with binary detection problems (epileptic vs. non-epileptic or seizure vs. non-seizure) and ternary detection problem (ictal vs. normal vs. interictal). The proposed system has been designed as an ensemble of a memory efficient and simple pyramidal one-dimensional deep convolutional neural network (P-1D-CNN) models, which takes an EEG signal as input, passes it to different base P-1D-CNN models and finally fuses their decisions using majority vote. To overcome the issue of small dataset, two data augmentation schemes have been introduced. Due to fewer parameters and augmentation scheme, the model is easy to train having limited data as well as easy to deploy on chips where memory is limited. It will assist neurologists in detecting epilepsy, and will greatly reduce their burden and increase their efficiency.

There are many future directions related to the proposed work. Though the proposed system gives good performance on a benchmark dataset, its clinical validation and to examine its suitability for deployment in clinical setting is a future work. Another possible direction is its incorporation in a wearable device for epilepsy patients. Though P-1D-CNN takes about 61% less memory and storage, its storage and memory requirements for wearable devices might be a problem. This issue needs further research to reduce further the memory and storage requirements. In addition, the system can be deployed in a centralized cloud environment to quickly access through mobile devices without the use of specific wearable devices. The small size EEG signal required as input and light P-1D-CNN model make it suitable for cloud deployment. The small signal can be easily transmitted to cloud for processing in real-time where it can generate a warning alarm to alert doctors/patients if necessary. In this regard, privacy protection and data loss are challenges while transmitting data to cloud. These issues will be addressed in future work. The proposed system is general in nature and can be employed to develop expert systems for similar disorders involving EEG brain signals such as workorder stress detection. Further, the deep model can be extended to design a more generalized and powerful model by increasing the depth of the model, and enhancing the diversity of based CNN models.

Acknowledgements

This work was supported by the Research Center of College of Computer and Information Sciences, King Saud University. The authors are grateful for this support.

References

Acharya, U. R., Fujita, H., Lih, O. S., Hagiwara, Y., Tan, J. H., & Adam, M. (2017a). Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Information Sciences, 405*, 81–90.

Acharya, U. R., Fujita, H., Lih, S., Hagiwara, Y., Hong, J., & Adam, M. (2017b). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences, 415–416*, 190–198.

Acharya, U. R., Fujita, H., Lih, O. S., Adam, M., Tan, J. H., & Chua, C. K. (2017c). Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. *Knowledge-Based Systems, 132*, 62–71.

Acharya, U. R., Molinari, F., Sree, S. V., Chattopadhyay, S., Ng, K.-H., & Suri, J. S. (2012a). Automated diagnosis of epileptic EEG using entropies. *Biomedical Signal Processing and Control, 7*(4), 401–408.

Acharya, U. R., Molinari, F., Vinitha Sree, S., Chattopadhyay, S., Kwan-Hoong, Ng., & Suri, J. S. (2012b). Automated diagnosis of epileptic EEG using entropies. *Biomed. Signal Process. Control, 7*(4), 401–408.

Acharya, U. R., Faust, O., Kannathal, N., Chua, T. J., & Laxminarayan, S. (2005). Dynamical analysis of EEG signals at various sleep stages, Comput. *Methods Programs Biomed, 80*(1), 37–45.

Acharya, U. R., Vinitha Sree, S., & Suri, J. S. (2011). Automatic detection of epileptic EEG signals using higher order cumulant features. *Int. J. Neural Syst., 21*(5), 1–12.

Acharya, U. R., Yanti, R., Zheng, J. W., Krishnan, M. M., Tan, J. H., Martis, R. J., et al. (2013). Automated diagnosis of epilepsy using CWT, HOS and texture parameters. *Int J Neural Syst, 23*(3).

Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of neuroscience methods, 123*(1), 69–87.

Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E, 64*(6), 061907.

Bhattacharyya, A., Pachori, R., Upadhyay, A., & Acharya, U. (2017). Tunable-Q Wavelet Transform Based Multiscale Entropy Measure for Automated Classification of Epileptic EEG Signals. *Applied Sciences, 7*(4), 385.

Chua, K. C., Chandran, V., Acharya, U. R., & Lim, C. M. (2011). Application of higher order spectra to identify epileptic EEG. *Journal of Medical Systems, 35*(6), 1563–1571.

Chua, K. C., Chandran, V., Acharya, U. R., & Lim, C. M. (2009). Automatic identification of epileptic EEG signals using higher order spectra,. *Journal of* Engineering and Medicine*, 223*(4), 485–495.

Cui, Z., Chen, W., & Chen, Y. (2016). Multi-scale convolutional neural networks for time series classification. arXiv preprint arXiv:1603.06995.

Gandhi, T., Panigrahi, B. K., & Anand, S. (2011). A comparative study of wavelet families for EEG signal classification. *Neurocomputing, 74*(17), 3051–3057.

Gardner, A. B., Krieger, A. M., Vachtsevanos, G., & Litt, B. (2006). One-class novelty detection for seizure analysis from intracranial EEG. *Journal of Machine Learning Research, 7*(Jun), 1025–1044.

Gotman, J. (1982). Automatic recognition of epileptic seizures in the EEG. *Electroencephalography and Clinical neurophysiology, 54*(5), 530–540.

Gotman, J. (1999). Automatic detection of seizures and spikes. *Journal of Clinical Neurophysiology, 16*(2), 130–140.

Gotman, J., Ives, J., & Gloor, P. (1979). Automatic recognition of inter-ictal epileptic activity in prolonged EEG recordings. *Electroencephalography and Clinical Neurophysiology, 46*(5), 510–520.

Guo, L., Rivero, D., Dorado, J., Rabunal, J. R., & Pazos, A. (2010). Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *Journal of Neuroscience Methods, 191*(1), 101–109.

Hussain, M., Aboalsamh, H., Abdul, W., Bamatraf, S., & Ullah, I. (2016). An intelligent system to classify epileptic and non-epileptic eeg signals. paper presented at the signal-image technology & internet-based systems (SITIS). *2016 12th international conference on*.

Ince, T., Kiranyaz, S., Eren, L., Askar, M., & Gabbouj, M. (2016). Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Transactions on Industrial Electronics, 63*(11), 7067–7075.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning, Lille, France: 37* JMLR: W&CP.

Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 221–231.

Khan, Y. U., Rafiuddin, N., & Farooq, O. (2012). Automated seizure detection in scalp EEG using multiple wavelet scales. Paper presented at the Signal Processing. *Computing and control (ISPCC), 2012 IEEE international conference on*.

Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980

Koffler, D., & Gotman, J. (1985). Automatic detection of spike-and-wave bursts in ambulatory EEG recordings. *Electroencephalography and Clinical Neurophysiology, 61*(2), 165–180.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks* Paper presented at the Advances in neural information processing systems.

LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks, 3361*(10) 1995.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324.

Martis, R. J., Acharya, U. R., Tan, J. H., Petznick, A., Tong, L., Chua, C. K., et al. (2013). Application of intrinsic time-scale decomposition (ITD) to EEG signals for automated seizure prediction. *International Journal of Neural Systems October, 23*(5), 1350023.

McShane, T. (2004). A clinical guide to epileptic syndromes and their treatment. *Archives of Disease in Childhood, 89*(6), 591.

Megiddo, I., Colson, A., Chisholm, D., Dua, T., Nandi, A., & Laxminarayan, R. (2016). Health and economic benefits of public financing of epilepsy treatment in India: An agent-based simulation model. *Epilepsia, 57*(3), 464–474.

Meier, R., Dittrich, H., Schulze-Bonhage, A., & Aertsen, A. (2008). Detecting epileptic seizures in long-term human EEG: A new approach to automatic online and real-time detection and classification of polymorphic seizure patterns. *Journal of Clinical Neurophysiology, 25*(3), 119–131.

Mirowski, P., Madhavan, D., LeCun, Y., & Kuzniecky, R. (2009). Classification of patterns of EEG synchronization for seizure prediction. *Clinical Neurophysiology, 120*(11), 1927–1940.

Nasehi, S., & Pourghassem, H. (2013). Patient-specific epileptic seizure onset detection algorithm based on spectral features and IPSONN classifier. *Paper presented at the communication systems and network technologies (CSNT), 2013 international conference on*.

Nicolaou, N., & Georgiou, J. (2012). Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Systems with Applications, 39*(1), 202–209.

Orhan, U., Heki, M., & Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications, 38*, 13475–13481.

Patel, K., Chua, C.-P., Fau, S., & Bleakley, C. J. (2009). Low power real-time seizure detection for ambulatory EEG. *Paper presented at the pervasive computing technologies for healthcare, 2009. pervasivehealth 2009. 3rd international conference on*.

Qu, H., & Gotman, J. (1993). Improvement in seizure detection performance by automatic adaptation to the EEG of each patient. *Electroencephalography and Clinical Neurophysiology, 86*(2), 79–87.

Samiee, K., Kovcs, P., & Gabbouj, M. (2015). Epileptic seizure classification of EEG timeseries using rational discrete short-time Fourier transform. *IEEE Transactions on Biomedical Engineering, 62*(2), 541–552.

Sharma, M., Pachori, R. B., & Acharya, U. R. (2017). A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recognition Letters*.

Sharmila, A., & Geethanjali, P. (2016). DWT based detection of epileptic seizure from EEG signals using naive Bayes and k-NN classifiers. *IEEE Access, 4*, 7716–7727.

Sheb, A., & Guttag, J. (2010). Application of machine learning to epileptic seizure detection. *Paper presented at the 2010 the 27th international conference on machinelearning*.

Shoeb, A. H. (2009). *Application of machine learning to epileptic seizure onset detection and treatment*. Massachusetts Institute of Technology.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Swami, P., Gandhi, T. K., Panigrahi, B. K., Tripathi, M., & Anand, S. (2016). A novel robust diagnostic model to detect seizures in electroencephalography. *Expert Systems with Applications, 56*, 116–130.

Subasi, A. (2007). EEG Signal classification using wavelet feature extraction and a mixture of expert model, Expert Syst. *Appl, 32*(4), 1084–1093.

TensorFlow. (June 15, 2017). Retrieved June 25, 2017, from https://www.tensorflow.org/.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3d convolutional networks. In *Paper presented at the Proceedings of the IEEE international conference on computer vision*.

Tzallas, A. T., Tsipouras, M. G., Tsalikakis, D. G., Karvounis, E. C., Astrakas, L., Konitsiotis, S. et al. (2012). Automated epileptic seizure detection methods: A review study Epilepsy-histological, electroencephalographic and psychological aspects: InTech.

Ullah, I., & Petrosino, A. (2016). About pyramid structure in convolutional neural networks. *Paper presented at the neural networks (IJCNN), 2016 international joint conference on*.

Yuan, Q., Zhou, W., Liu, Y., & Wang, J. (2012). Epileptic seizure detection with linear and nonlinear features. *Epilepsy & Behavior, 24*(4), 415–421.

Zhang, T., Chen, W., & Li, M. (2017). AR based quadratic feature extraction in the VMD domain for the automated seizure detection of EEG using random forest classifier. *Biomedical Signal Processing and Control, 31*, 550–559.

Zhang, X., & LeCun, Y. (2015). Text understanding from scratch. arXiv preprint arXiv:1502.01710.