

# Deep Learning Enabled Automatic Abnormal EEG Identification

Subhrajit Roy<sup>\*1</sup>, Isabell Kiral-Kornek<sup>1</sup>, and Stefan Harrer<sup>1</sup>, *IEEE Senior Member*

**Abstract**—In hospitals, physicians diagnose brain-related disorders such as epilepsy by analyzing electroencephalograms (EEG). However, manual analysis of EEG data requires highly trained clinicians or neurophysiologists and is a procedure that is known to have relatively low inter-rater agreement (IRA). Moreover, the volume of the data and rate at which new data is acquired makes interpretation a time-consuming, resource hungry, and expensive process. In contrast, automated analysis offers the potential to improve the quality of patient care by shortening the time to diagnosis, reducing manual error, and automatically detecting debilitating events. In this paper, we focus on one of the early decisions made in this process which is identifying whether an EEG session is normal or abnormal. Unlike previous approaches, we do not extract hand-engineered features but employ deep neural networks that automatically learn meaningful representations. We undertake a holistic study by exploring various pre-processing techniques and machine learning algorithms for addressing this problem and compare their performance. We have used the recently released “TUH Abnormal EEG Corpus” dataset for evaluating the performance of these algorithms. We show that modern deep gated recurrent neural networks achieve 3.47% better performance than previously reported results.

**Index Terms**—Electroencephalogram, Deep Learning, Recurrent Neural Networks, Epilepsy.

## I. INTRODUCTION

Electroencephalogram (EEG) data i.e. recordings of electrical activity along the scalp, is often used for the diagnosis and management of various neurological conditions such as epilepsy, somniphathy, coma, and encephalopathies. The high temporal resolution of EEG recordings and the noninvasive nature and comparatively low cost of equipment contribute to the popularity of EEG data among physicians [1].

The diagnosis of a neurological disorder by interpreting EEG data typically involves the recording of multiple short sessions or long-term monitoring [1]. This is because the peculiarities of a disorder are not guaranteed to be present in EEG data during one short recording session. For example, only 50% of epileptic patients show interictal epileptiform discharges (IED) in their first recording [1]. This can lead to the generation of a large amount of data that then needs to be manually interpreted by expert investigators.

A typical EEG report contains information such as patient history, medications, interesting findings, and general peculiarities. This includes a first impression of whether the recorded activity is normal or abnormal, a decision based on the raw signal and the patient’s state of consciousness [2]. Based on this classification alone, it can be decided if further investigations are done or medication is prescribed.

<sup>\*</sup>Corresponding author: subhrajit.roy@au1.ibm.com

<sup>1</sup>Subhrajit Roy, Isabell Kiral-Kornek, and Stefan Harrer are with IBM Research – Australia, 60 City Road, Southbank, Victoria, 3006, Australia.

Since training and certification to medically interpret EEG data can take years, this task falls on a relatively low number of neurologists. Hence, delays between data acquisition and diagnosis can range from hours to weeks. Moreover, the EEG interpretation process is known to have low inter-rater agreement, which can lead to misdiagnoses [3].

A way to address these issues is the introduction of automated EEG interpretation, an approach that has recently gained in popularity. In a first instance, automatically labelled data could serve as an aid to neurologists, reducing heavy workload and delays.

This work, similar to [2], [4], is concerned with the automatization of the abnormal/normal EEG data classification.

In [2] and [4], the authors first extracted hand-engineered frequency features from the recordings. Next, traditional machine learning algorithms were used, such as k-nearest neighbour, random forests, and hidden markov models, and deep learning techniques such as convolutional neural networks were used. The authors [2], [4] released the corresponding dataset which is known as the “TUH Abnormal EEG Corpus” [5] and is the largest dataset containing normal/abnormal scalp EEG recordings.

We envision that modern deep learning algorithms such as deep gated recurrent networks or time-distributed neural networks might be capable of directly learning useful representations from the data without any explicit pre-processing and/or transformation step. Note that, to the best of our knowledge, this is the first time these two types of networks have been used for solving the abnormal EEG identification task. However, to obtain a holistic view, we did not constrain ourselves to these methods. Hence, we undertake a comprehensive study by exploring various data augmentation, data pre-processing, and machine learning techniques on the “TUH Abnormal EEG Corpus” for automated EEG identification. The primary contributions of this work are:

- We propose a novel data augmentation technique for the “TUH Abnormal EEG Corpus” which leads to a 11 fold increase in the training data. Note that this is crucial for modern deep learning techniques where performance typically tends to increase with more training data.
- We explore three pre-processing techniques: feeding the input EEG directly to a classifier with minimal pre-processing (only filtering), converting EEG data into spectrograms [6], and transforming EEG data into Gramian Angular Fields (GAF) [7].
- We present results for classical machine learning algorithms such as logistic regression and multi-layer perceptrons as well as more sophisticated deep learning algorithms such as 1D convolutional neural networks,

deep convolutional gated recurrent neural networks, and time-distributed convolutional recurrent neural networks. To the best of our knowledge, we have used deep recurrent neural networks on this dataset for the first time.

## II. BACKGROUND AND THEORY

In this section, we briefly discuss the various pre-processing techniques and learning algorithms we have used.

### A. Pre-processing techniques

1) *Filtering*: All EEG data considered in our study were highpass filtered using a 4<sup>th</sup> order Butterworth filter with a cutoff frequency of 0.01 Hz.

2) *Spectrograms*: Multiple previous studies have reported that analyzing EEG signals in the frequency domain can be effective for subsequent pattern recognition tasks. Taking inspiration from that, we converted the recorded EEG time series signal into spectrograms. Spectrograms are two-dimensional representation of the signal spectrum over time.

3) *Gramian Angular Field*: Gramian Angular Fields allow to encode time-series in a two-dimensional space. GAF images are generated by representing input time-series in a polar coordinate system where each element is the trigonometric sum between different time intervals.

### B. Learning algorithms

1) *Logistic Regression*: Logistic regression is a linear classifier that predicts the probability of occurrence of an event by fitting input data to a logit function. It is a simple and widely used machine learning technique [8].

2) *Multi-layer perceptron*: Multi-layer perceptron is a class of feed-forward neural networks which consist of at least three layers of non-linear (except the input layer) nodes. It is trained using the backpropagation learning rule [8] and has the capability to discriminate non-linearly separable data.

3) *1D convolutional neural network (1D-CNN)*: 1D-CNN is a multi-layered architecture where each layer consists of a few one-dimensional convolution filters. The filters operate on short consecutive subsequences of the input for extracting meaningful features. Hence it is well-suited for time-series classification [9].

4) *2D convolutional neural network (2D-CNN)*: 2D-CNN is a category of artificial neural networks that is particularly suited for processing image data. The primary idea is to train two-dimensional filters to extract local features in different level of hierarchies. Recently, 2D-CNNs have achieved impressive results in various different pattern recognition tasks [9].

5) *Deep 1D convolutional gated recurrent neural network (1D-CNN-RNN)*: Recurrent Neural Networks (RNN) are a family of neural networks for processing variable-length sequential data. Since RNNs have a tendency to overfit, efficient recurrent units have been proposed in recent times such as long short term memory (LSTM) [10] and the gated recurrent unit (GRU) [11]. In this paper, 1D-CNN-RNN denotes a combination of one-dimensional convolution

layers followed by stacked GRU layers. It is computationally efficient and particularly suited for time-series analyses.

6) *Time-distributed convolutional recurrent neural network (TCNN-RNN)*: Unlike 1D-CNN that processes sequences or one dimensional input and 2D-CNN which takes two dimensional images as input, TCNN-RNN processes one sequence of images at a time and classifies them. They are useful in scenarios where time-series data is converted into image representation (spectrograms, GAFs, etc.) and fed to a classifier.

## III. EXPERIMENTS

### A. Data selection

In this article, we have considered the TUH Abnormal EEG Corpus [4] that contains EEG records labelled as either clinically abnormal or normal. The TUH Abnormal EEG Corpus is the worlds largest publicly available dataset of its type and consists of 1488 abnormal and 1529 normal EEG sessions respectively. The dataset is demographically balanced with respect to gender and age of patients. For evaluation of automated systems, it was divided into a training set (1361 abnormal/1379 normal samples), and a test set (127 abnormal/150 normal samples).

### B. Data preparation and augmentation

First, we converted the EEG signal recorded by the electrodes to a set of differentials according to the transverse central parietal (TCP) montage system [4]. We did not extract any hand-crafted features from this dataset since we envisioned that the deep neural networks used in this paper will be able to automatically capture relevant features. Typically the sessions were recorded at a sampling frequency of 250 Hz. When this was not the case, the signal was resampled to 250 Hz.

In previous studies [2], [4], the authors noted that trained clinicians can accurately determine whether an EEG session is abnormal or normal by examining its first few minutes. This encouraged them to design automated systems that can classify EEG sessions by looking at only the first minute of data. Hence, to create the training and test set the authors extracted the first minute of data from the available EEG sessions. This protocol was adopted during the testing or inference phase to make a fair comparison of their system to human-level performance. On the other hand, using only the first minute of data during training was a design choice motivated by the fact that it might be most representative of the test set. Once the electrodes are placed on the scalp and data recording starts, the signal will gradually deviate from the first minute due to changing impedances. However, this method significantly constricts the amount of data that can be used for training. This impacts the performance of deep neural networks which typically gets better as more training data is used.

To alleviate this problem, our intuition was to include data beyond the first minute in the training set. To find out how much data can be used without performance degradation we performed the following experiment: we chose a random

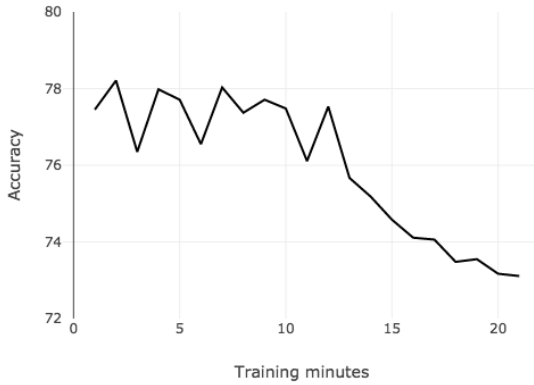


Fig. 1: The effect of each training minute on the validation accuracy.

subset of sessions from the original training data which was further divided into smaller training and validation sets. Separate models were trained for each minute of the training sets, allowing us to assess the performance during inference on the first minute of the validation sessions. Figure 1 shows the outcome of this experiment. It is evident from Figure 1 that performance starts to drop after the 11<sup>th</sup> minute thereby suggesting that we can use up to 11 minutes of data from the training EEG sessions. This led to a 11-fold increase in our training data as compared to previously reported methods [4].

### C. Results

We used the dataset described above to train combinations of pre-processing techniques and learning algorithms described in Section II. We combined algorithms that are suited to process two-dimensional data with GAFs or spectrograms. This includes 2D-CNN and TCNN-RNN. On the other hand, time-series data is directly fed to algorithms that are suited to handle one-dimensional data such as logistic regression, MLP, 1D-CNN, and 1D-CNN-RNN.

For logistic regression we used mini-batch gradient descent with a learning rate of 0.001. For MLP, we employed one hidden layer having 1000 sigmoidal neurons and used gradient descent with a learning rate of 0.01. The architecture of 1D-CNN and 1D-CNN-RNN used in this paper are shown in Figure 2. Moreover, we show the specific network for 2D-CNN and TCNN-RNN in Figure 3. The format we have used throughout the paper to describe the layers used in deep networks are (layer name, filter length, number of filters, stride size) whenever applicable. 1D-CNN, 2D-CNN, 1D-CNN-RNN, and TCNN-RNN were trained using the adaptive moment estimation optimization [12] algorithm with a learning rate of 0.001.

In Table I, we list the average results obtained through 5 repetitions of the above experiments. We find that logistic regression overfits on the input data and hence provides a test accuracy similar to a random classifier. A slightly better performance is achieved by MLP which reduces overfitting

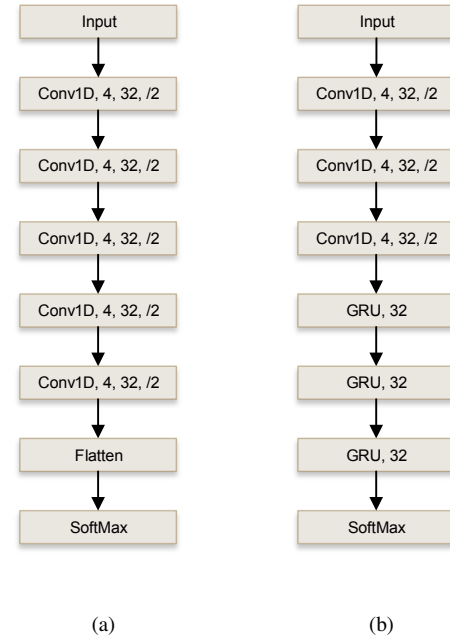


Fig. 2: (a) The 1D convolutional neural network is formed by stacking multiple 1D convolutional layers. (b) The deep 1D convolutional gated recurrent neural network (1D-CNN-RNN) stacks multiple 1D convolution layers followed by GRU layers. Both networks take sequences as input and hence time-series EEG data is fed into them with minimal pre-processing (filtering only).

and achieves a test accuracy slightly over chance. 1D-CNN achieves a training and test accuracy of 82.04% and 76.90% and therefore outperforms both logistic regression and MLP. Compared to 1D-CNN, 1D-CNN-RNN increases the test accuracy even further to 82.27%. However, due to the presence of the recurrent layers 1D-CNN-RNN overfits.

The second and third rows of the table show the results for the cases where time-series is converted to image representations. We find that for the classifiers (2D-CNN and TCNN-RNN) considered in this paper, spectrogram representation works better than GAFs. 2D-CNN and TCNN-RNN achieve test accuracies of 70.39% and 71.48% when input is represented as spectrograms.

Our findings suggest that the best performance is obtained by 1D-CNN-RNN directly operating on time-series data. We speculate that since minimal pre-processing is involved in this case, the deep network is automatically able to learn representative features from the data. Compared to previously reported best results [4], 1D-CNN-RNN shows a 3.47% increase in performance.

## IV. CONCLUSION

In this paper, we have focused on classifying an EEG record into either abnormal or normal type. This is one of the first steps in EEG data interpretation and successfully automating this procedure will not only substantially reduce

Pre. proc. \ Algo.	Log. Reg.		MLP		1D-CNN		2D-CNN		1D-CNN-RNN		TCNN-RNN	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Time-series	83.72%	49.09%	69.64%	54.15%	82.04%	76.90%	N/A	N/A	<b>99.16%</b>	<b>82.27%</b>	N/A	N/A
Spectrograms	N/A	N/A	N/A	N/A	N/A	N/A	86.31%	70.39%	N/A	N/A	95.22%	71.48%
GAF	N/A	N/A	N/A	N/A	N/A	N/A	79.46%	68.61%	N/A	N/A	92.55%	67.02%

TABLE I: Comparison of performance for combination of pre-processing techniques and learning algorithms.

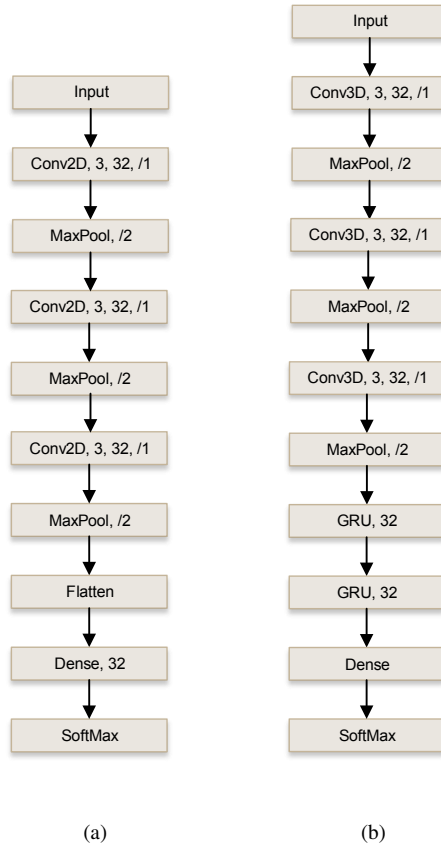


Fig. 3: (a) The 2D convolutional neural network (2D-CNN) combining Convolution2D, MaxPool, Flatten, and Dense layers. (b) The time-distributed convolutional neural network (TCNN-RNN) that combines Convolution3D, MaxPool, Flatten, and Dense layers. While both these networks take inputs represented as images only TCNN-RNN takes sequence of images as input.

time required to read EEG but also act as an aid to human investigators. For this study, we have used the TUH Abnormal EEG Corpus [5], [4]. We explored various pre-processing techniques, traditional machine learning algorithms, and modern deep neural networks to solve this problem. The best performance is achieved by directly feeding time-series data to a hybrid convolutional recurrent network architecture. This network denoted as 1D-CNN-RNN achieves a 3.47% increase in performance compared to previous reported best result. However, the network can overfit and in future we intend to develop custom recurrent neural networks to tackle this problem.

#### ACKNOWLEDGMENT

The authors would like to thank Benjamin Scott Mashford from IBM Research - Australia for valuable discussions and Joseph Picone and Iyad Obeid from Temple University for providing the dataset used in this paper.

#### REFERENCES

- [1] S. J. M. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, no. suppl 2, pp. ii2–ii7, 2005.
- [2] S. Lopez, G. Suarez, D. Jungreis, I. Obeid, and J. Picone, "Automated Identification of Abnormal Adult EEGs," *IEEE Signal Process Med Biol Symp*, vol. 2015, Dec. 2015.
- [3] H. Azuma, S. Hori, M. Nakanishi, S. Fujimoto, N. Ichikawa, and T. A. Furukawa, "An intervention to improve the interrater reliability of clinical EEG interpretations," *Psychiatry and clinical neurosciences*, vol. 57, no. 5, pp. 485–489, 2003.
- [4] S. Lopez, "Automated Interpretation of Abnormal Adult Electroencephalograms," *MS Thesis, Temple University*, 2017.
- [5] I. Obeid and J. Picone, "The Temple University Hospital EEG Data Corpus," *Front Neurosci*, vol. 10, May 2016. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4865520/>
- [6] I. Kiral-Kornek, S. Roy, E. Nurse, B. Mashford, P. Karoly, T. Carroll, D. Payne, S. Saha, S. Baldassano, T. O'Brien, D. Grayden, M. Cook, D. Freestone, and S. Harter, "Epileptic Seizure Prediction Using Big Data and Deep Learning: Toward a Mobile System," *EBioMedicine*, Dec. 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S235239641730470X>
- [7] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," in *Proceedings of the 24th International Conference on Artificial Intelligence*, ser. IJCAI'15. AAAI Press, 2015, pp. 3939–3945.
- [8] K. P. Murphy, *Machine learning : a probabilistic perspective*. MIT Press, 2013.
- [9] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning | The MIT Press," 2016. [Online]. Available: <https://mitpress.mit.edu/books/deep-learning>
- [10] A. Graves, "Generating Sequences With Recurrent Neural Networks," *arXiv:1308.0850 [cs]*, Aug. 2013, arXiv: 1308.0850. [Online]. Available: <http://arxiv.org/abs/1308.0850>
- [11] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches," *arXiv:1409.1259v2*, Sep. 2014. [Online]. Available: <https://arxiv.org/abs/1409.1259>
- [12] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Dec. 2014, arXiv: 1412.6980. [Online]. Available: <http://arxiv.org/abs/1412.6980>