2016 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 13–16, 2016, SALERNO, ITALY

DEEP LEARNING FOR EPILEPTIC INTRACRANIAL EEG DATA

Andreas Antoniades, Loukianos Spyrou, Clive Cheong Took, and Saeid Sanei

Department of Computer Science
University of Surrey, Guildford, United Kingdom
{a.antoniades, l.spyrou, c.cheongtook, s.sanei}@surrey.ac.uk

ABSTRACT

Detection algorithms for electroencephalography (EEG) data typically employ handcrafted features that take advantage of the signal's specific properties. In the field of interictal epileptic discharge (IED) detection, the feature representation that provides optimal classification performance is still an unresolved issue. In this paper, we consider deep learning for automatic feature generation from epileptic intracranial EEG data in the time domain. Specifically, we consider convolutional neural networks (CNNs) in a subject independent fashion and demonstrate that meaningful features, representing IEDs are automatically learned. The resulting model achieves state of the art classification performance, provides insights for the different types of IEDs within the group, and is invariant to time differences between the IEDs. This study suggests that automatic feature generation via deep learning is suitable for IEDs and EEG in general.

Index Terms— EEG, Deep Learning, Epilepsy, Convolutional Neural Networks.

1. INTRODUCTION

Neural networks are a popular class of machine learning algorithms typically used for classification and regression, and have been extensively used in many applications over the past few decades [1]. Considered as blackboxes by many, neural networks create non-linear combinations of the original input to identify its most informative characteristics. Deep learning is possibly one of the most significant advances in neural networks and has found applications mainly in dimensionality reduction and hierarchical multilayer feature learning [2]. Deep neural networks are now considered state of the art algorithms and have won numerous contests in pattern recognition and machine learning, and most importantly can achieve better than human accuracy on classification of benchmark datasets [3].

In a statistical context, deep learning algorithms can be seen as recursive generalised linear models based on regression [4]. Statistical methods that employ architectures with one or no hidden layers, perform the so-called shallow learning [5]. On the other hand, deep learning methods make use of architectures of many hidden layers. A number of deep learning algorithms have been proposed over the past years. Stacked Denoising Autoencoders learn useful representations of data by attempting to reconstruct the input signal at each layer while adding artificial noise [6]. Deep Belief Networks operates on stacked Restricted Boltzmann machines and follow a stochastic approach for optimisation [7]. Finally, Convolutional Neural Networks (CNNs) were designed to automatically learn optimal filters that generate high level features when applied to images [8].

Deep learning can benefit electroencephalography (EEG) studies, since neural processes exhibit a hierarchical structure. The electrical response of the brain to a stimulus, task, or physiological state usually comprises of a number of subresponses which are combined to provide the full response. Deep learning has been used for classification in EEG [9], feature extraction [10] and in the epilepsy field for seizure prediction [11]. All these established findings can make sense in the interpretation of EEG signals. This is what our work addresses, making sense of epileptic EEG data without the need for expert domain knowledge.

Interictal epileptiform discharges (IEDs) are transients of electrical activities that appear in EEG recordings of patients with epilepsy. Their accurate detection and localisation is important to the diagnosis and treatment of epilepsy. Many classification algorithms have been developed for the purpose of distinguishing IED from non-IED activity. A review of algorithms is provided in [12, 13]. The common denominator of all these methods lies in the statistical description of an IED signal, which can be obtained either through modeling or through using a similarity-based algorithm. This is often facilitated by obtaining useful representations of the signal that can better exploit its structure. The feature representation useful in IED detection has remained a rather unresolved issue [12].

In this work we exploit deep learning to generate an optimal set of features from the EEG signals of 25 patients. We provide a comprehensive analysis making the CNN-based

This work has been supported by the EPSRC, UK, Grant No. EP/K005510/1.

978-1-5090-0746-2/16/$31.00

trained filters suitable for IED feature extraction. That is because IEDs of different patients and channels may originate from slightly different temporal locations and also in some cases there might be multiple IEDs in a single segment. In Section 2, we provide initial insight into the epileptic dataset and the preprocessing methodology for the intracranial EEG. Section 3 describes in detail the CNN structure and parameters, while Sections 4 and 5 show the results and conclude the paper.

2. EPILEPTIC EEG

An epilepsy expert from King's College London Hospital visually inspected the intracranial data and marked the timing information where the intracranial data exhibited visually noticeable epileptic discharges. Classifier training was enabled by slicing the raw EEG data by a $\pm 160ms$ window centered on the intracranial timing scores and baselined on the preceding $160ms$ with the resulting signal finally being linearly detrended to remove undesired drifts. An example IED is provided in Fig. 1. Non-IED segments were also obtained from time segments where there were no scored IEDs. For each subject, the number of sliced IED and non-IED segments was chosen to be the same. The spectrogram method was used to convert the time-domain (TD) signal into time frequency (TF) features with a Hanning-tapered window length of $80ms$ and an overlap of 50% between windows. The window was chosen so as to capture the morphology of IED spikes whose sub-components have a duration of less than $80ms$. These TF features were obtained for each IED and non-IED segment. The TD features were of size 12 channels $\times$ 65 samples where the TF features were of size 12 channels $\times$ 9 frequency bins $\times$ 7 time points. Table 1 summarises the information on the epileptic EEG data.

Table 1. Dataset information

Subject	No. of trials	Subject	No. of trials
1	650	14	236
2	1906	15	448
3	1658	16	50
4	1098	17	100
5	424	18	146
6	1696	19	76
7	398	20	520
8	330	21	96
9	316	22	46
10	684	23	1212
11	682	24	228
12	944	25	1084
13	634		
No. of segments			
IED	7831	non-IED	7831

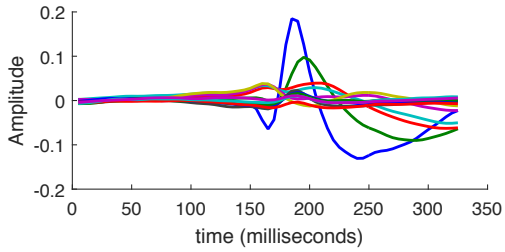

Fig. 1. Example IEDs for one subject. Channels are superimposed on the same plot and the signals were averaged over all IED segments of different time instances.

3. FEATURE GENERATION USING DEEP NEURAL NETWORKS

3.1. Preliminaries

3.1.1. Kernel Convolution

Kernel convolution is widely used in image processing and should not be confused with the traditional linear convolution $\mathbf{x}(n) * \mathbf{h}(n) = \sum_{\tau=-\infty}^{\infty} \mathbf{x}(\tau)\mathbf{h}(n-\tau)$. Although, kernel convolution involves a filter $\mathbf{F}$ *convolved* with an image $\mathbf{I}$ to create an image $\mathbf{C}$ for sharpening or blurring purposes, it is closer to correlation than linear discrete convolution. For clarity, an example of kernel convolution is given below.

$$\begin{matrix} & \mathbf{I} & & & \mathbf{F} & & \mathbf{C} \end{matrix}$$

$$\begin{pmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} & \mathbf{I}_{13} \\ \mathbf{I}_{21} & \mathbf{I}_{22} & \mathbf{I}_{23} \\ \mathbf{I}_{31} & \mathbf{I}_{32} & \mathbf{I}_{33} \end{pmatrix} * \begin{pmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix}$$

where $\mathbf{C}_{12} = \mathbf{I}_{12}\mathbf{F}_{11} + \mathbf{I}_{13}\mathbf{F}_{12} + \mathbf{I}_{22}F_{21} + \mathbf{I}_{23}\mathbf{F}_{22}$. This simple example is instructive in two ways. First, observe that the kernel convolution does not involve the time reversal involved in discreet linear convolution. In fact, the kernel convolution captures the correlation between filter coefficients and the image samples, and can be generalised as

$$\mathbf{C}_{ij} = \sum_{k=1}^{K}\sum_{l=1}^{L} \mathbf{I}[i+k-1, j+l-1]\mathbf{F}_{kl} \qquad (1)$$

where K and L are the width and height of the filter $\mathbf{F}$. Second, the resulting convolved signal (image $\mathbf{C}$) is actually smaller than the actual signal (image $\mathbf{I}$) in kernel convolution, whereas in linear discreet convolution, the resulting convolved signal is generally longer than the actual signal.

3.1.2. Convolutional Neural Networks for EEG processing

Unlike traditional fully connected networks, CNNs offer a mechanism to exploit *local connectivities* such that the learnt 'filters' lead to the strongest response to a spatially local input patterns in the image. During training, the learnt coefficients

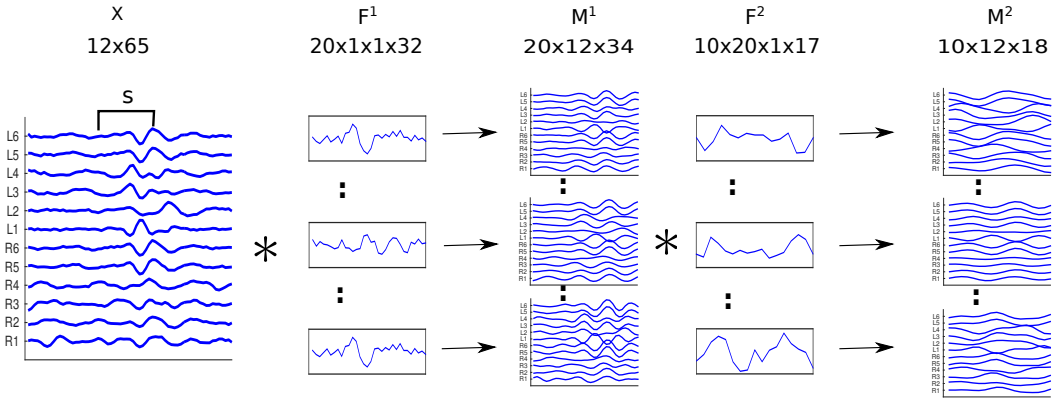

Fig. 2. Convolutional model for EEG processing, where $\mathbf{M}^n$ are the feature maps generated at convolutional layer n, $\mathbf{F}^n$ are the filters at layer n and $*$ is the convolution operator. $\mathbf{M}$ represents a tensor of dimension [Filters $\times$ Channels $\times$ Time].

are used to produce an input for the next layer, called a feature map. Each feature map can detect a different component of interest [2], provided *different* sets of coefficients are learnt. In practice, these different sets of learnt coefficients can be achieved by making use of a wide spectrum of initialisation values of those coefficients.

In the context of EEG processing, 2-d kernel convolution (1) needs to be adapted, for the spatial y-axis is not relevant for time series data. As such, 1-d variant of (1) is considered and the proposed EEG-oriented CNN framework is illustrated in Fig. 2. To learn the coefficients of each filter $\mathbf{f}$, our proposed CNN sweeps through the data of all channels one-by-one (from Channel L6 to Channel R1). Once the coefficients for one filter are initialised, these are convolved with each channel of $\mathbf{X}$ to produce a channel in the feature map smaller than the support of the actual channel in the input signal $\mathbf{X}$. For instance, the support of the input signal is $N = 65$ in layer 1, whereas length of the feature map in layer $\mathbf{M}^1$ is $J = N - L + 1 = 34$, where $L = 32$ is the filter length. All these convolved signals produce a new multichannel time series, which is the feature map. In Fig. 2 for layer $\mathbf{M}^1$, the $[i, j]$ sample of the kth feature map is generated from the learnt coefficients of the kth filter $\mathbf{f}_k$ as follows

$$\mathbf{M}_{kij} = \sum_{l=1}^{L} \mathbf{X}[i, j + l - 1]\mathbf{f}_{kl} \quad \forall k = 1, ..., K \tag{2}$$

where K is the total number of filters considered, e.g. $K = 20$ in layer $\mathbf{M}^1$, and $K = 10$ in layer $\mathbf{M}^2$ in Fig. 2. These sets of learnt coefficients can describe and detect different patterns in the EEG data irrespective of their temporal location within the data segment as shown in Section 4. For each CNN layer, new feature maps are produced (via convolutions) according to different sets of learnt filter coefficients of the previous layer. However, the linear model in (2) needs to be adapted to the n^{th} layer of the neural networks, which leads (2) to

$$\mathbf{G}^n_{ki} = \mathbf{M}^{n-1}_{ki} * \mathbf{f}^n_k + b^n_i \tag{3}$$

$$\mathbf{M}^n_{ki} = \tanh(\mathbf{G}^n_{ki}) \tag{4}$$

where the subscript i and k denote the corresponding indices for i^{th} channel of the feature map and k^{th} filter respectively, and the superscript n denotes the n^{th} layer of the neural network. The hyperbolic tangent function $\tanh(\cdot)$ is adopted as the nonlinear function, as it can model intracranial EEG successfully [11]. Due to limited space, we provide a sketch derivation of the backward propagation for our model, focusing on convolution operation. Given a cost function J, the update of the weight coefficients can be computed using the steepest descent as

$$\mathbf{f}^n_k[l + 1] = \mathbf{f}^n_k[l] - \mu \frac{\partial J[l + 1]}{\partial \mathbf{f}^n_k[l + 1]} \tag{5}$$

where the gradient component $\frac{\partial J}{\partial \mathbf{f}^n_k}$ is given as

$$\frac{\partial J}{\partial \mathbf{f}^n_k} = \frac{\partial J}{\partial \mathbf{M}^n_{ki}} \frac{\partial \mathbf{M}^n_{ki}}{\partial \mathbf{G}^n_{ki}} \frac{\partial \mathbf{G}^n_{ki}}{\partial \mathbf{f}^n_k} \tag{6}$$

The time indices are omitted for brevity and conciseness. It is straightforward to determine the derivatives $\frac{\partial J}{\partial \mathbf{M}^n_{ki}}$ and $\frac{\partial \mathbf{M}^n_{ki}}{\partial \mathbf{G}^n_{ki}}$, as the derivative of the cross-entropy error function and nonlinear function $\tanh(\cdot)$ are well documented in the literature. Instead, we focus on the term $\frac{\partial \mathbf{G}^n_{ki}}{\partial \mathbf{f}^n_k}$ which involves the kernel convolution operation. Using the property linear convolution $\frac{\partial (a*b)}{\partial c} = \frac{\partial a}{\partial c} * b$, we can exploit this property for the kernel convolution and (3) as

$$\begin{aligned} \frac{\partial \mathbf{G}^n_{ki}}{\partial \mathbf{f}^n_k} &\approx \frac{\partial \mathbf{f}^n_k}{\partial \mathbf{f}^n_k} * \mathbf{M}^{n-1}_{ki} + \frac{\partial b^n_i}{\partial \mathbf{f}^n_k} \\ &\approx \mathbf{M}^{n-1}_{ki} \end{aligned} \tag{7}$$

where it is clear that it is the function $\mathbf{f}_k^n$ is continuously differentiable with respect to itself, whereas the function $\mathbf{M}_{ki}^{n-1}$ can be approximated as an arbitrary locally integrable function to enable the derivative in (7).

3.2. Training procedure

We have used the leave-subject-out method for our experiments. This method uses the data of a subject for testing and the other 24 for training. This procedure is repeated until all subjects are used for testing. Initial experimentation has revealed that using very deep networks of more than two convolutional layers, does not improve the classification rate. Two CNN topologies were considered, CNN1 comprised of a single convolution layer whereas CNN2 comprised of two convolutional layers. Both networks were followed by a single fully-connected sigmoidal hidden layer and a logistic regression model to perform classification on the generated features. The network was trained using backpropagation until convergence. The parameters for the two CNNs are summarised in Table 2. We compared the proposed CNN architecture with a state-of-the-art classification algorithm [13] that only uses a single logistic regression layer but with handcrafted optimised features. State-of-the-art performance is achieved with TF features (see Section 2) with a logistic regression classifier. We include four comparisons in this study: TD features with logistic regression, TF with logistic regression, and CNN1/CNN2 with TD features. Table 3 summarises the classification accuracy of each method.

Table 2. Training parameters for the all the considered methods

Parameter/Method	TD [13]	TF [13]	CNN1	CNN2
Convolution layers	-	-	1	2
No. of filters	-	-	126	20, 10
Filter order	-	-	32	32, 17
Sigmoidal layer size	-	-	500	500
Network parameters	-	-	4532	4540
Classifier features	780	756	500	500
Total layers	1	1	3	4

4. MAKING SENSE OF EPILEPTIC EEG IN MACHINE LEARNING

To address the 'black-box' nature of neural networks, we now illustrate how the learning process of CNNs captures features of IEDs. Fig. 3 shows the evolution of learning weights pertaining to the first layer of CNN2 through training and how it correlates with an averaged IED. The correlation of layer 1 was 42% at convergence. We must note that this correlation was between a single filter and the average of all IEDs. At the second layer, a correlation of 52% was observed. From Fig. 4, it is clear that from the top-left most subplot to the bottom right-most subplot that these learning coefficients were converging towards the waveform of an IED, as shown previously in Fig 1. Also, notice that two epileptic spikes were captured through the weights.

In Fig. 4 the first layer filters of CNN2 (leftmost subplots) are compared with its second layer convolved filters (rightmost subplots). Observe that the background EEG occurred at the start of each subplot, followed by the occurrence of epileptic spikes, which were then followed again by background EEG. The second layer learning coefficients (rightmost subplots) captured more complex and well-defined epileptic shapes than those of first layer (leftmost subplots). In other words, background EEG was made much more *distinct* from those spikes. This enhanced distinction between background activity and epileptic spikes should improve the detection of those spikes for classification from shallow learning (CNN1) to deep learning (CNN2), as confirmed by the classification accuracy provided in Table 3.

The left plot of Fig. 5 illustrates the inputs to layer 1 of the IED class, whereas that of Fig. 6 shows the inputs to layer 1 pertaining to the non-IED class. On the other hand, the right plots of both figures show the outputs of the first layer, resulting from the convolution with the learnt filters in Fig. 4 (leftmost subplots) depicting epileptic waveforms. Notice that the high magnitude activations in the output signals in Fig. 5 is much more synchronised in the neighbouring electrodes than those in Fig. 6. This again indicates that the epileptic patterns were successfully learnt by our proposed method.

Although shallow learning of CNN1 yielded poorer results than those in our previous work [13], deep learning of CNN2 had similar performances as in [13]. This was confirmed by McNemars statistical test to assess the significance difference between the CNN1 method and the other three. TD-CNN2 and CNN1-CNN2 had significant difference with $p < 0.01$, while TF-CNN2 was not significant. In other words, our approach CCN2 provided similar performance results as those in our previous work [13]. Yet, the advantage of the proposed methodology was to circumvent the use of time-frequency analysis, facilitating the interpretation of the EEG data.

5. CONCLUDING REMARKS

We have demonstrated that CNNs can learn the intracranial IED waveform patterns. This has important healthcare implications as the shape of the IED can assist the diagnosis of epilepsy. The results of this paper are instructive and suggest that automatic feature generation based on deep learning is a potentially useful tool for IED detection and EEG data in general. Future works include deeper learning with additional hidden layers in CNNs and a closer examination of the different kinds of epileptic waveforms via neural network learning.

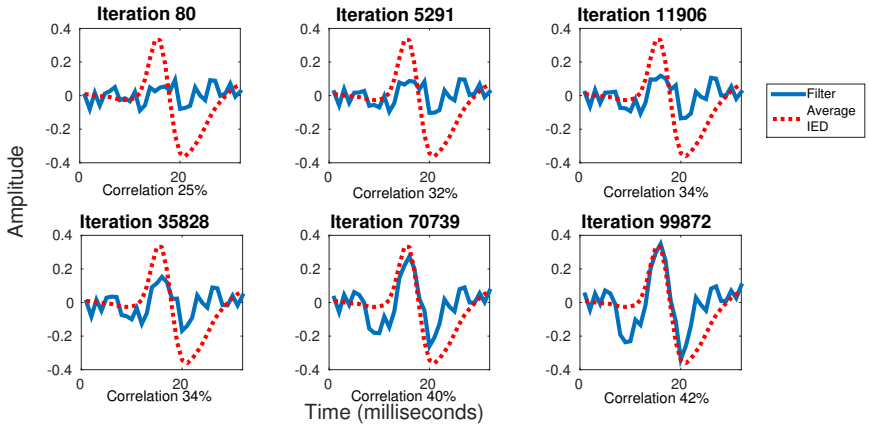

Fig. 3. Evolution of learnt coefficients towards the morphology of an epileptic waveform. Correlation does not exceed 42% as we are comparing a single filter with the average of different IEDs.

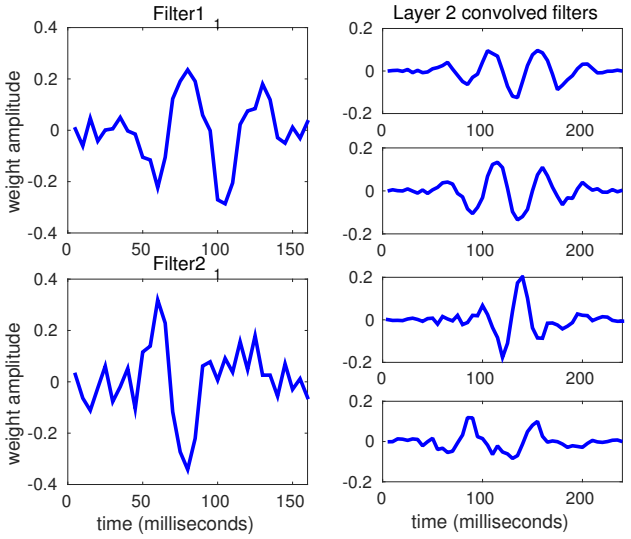

Fig. 4. Learnt filter coefficients for layer 1 for simple IED waveforms (left) and convolved learnt filters for both layers for more complex shapes (right).

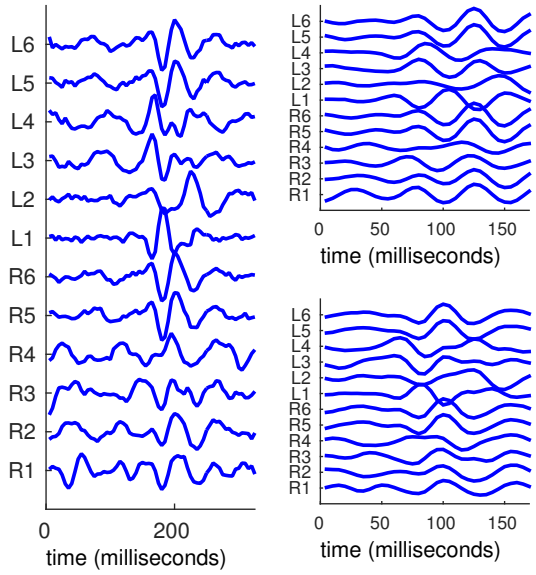

Fig. 5. Example segment of the IED class and its layer 1 feature map (multichannel) for the two filters shown on the L.H.S of Fig. 4.

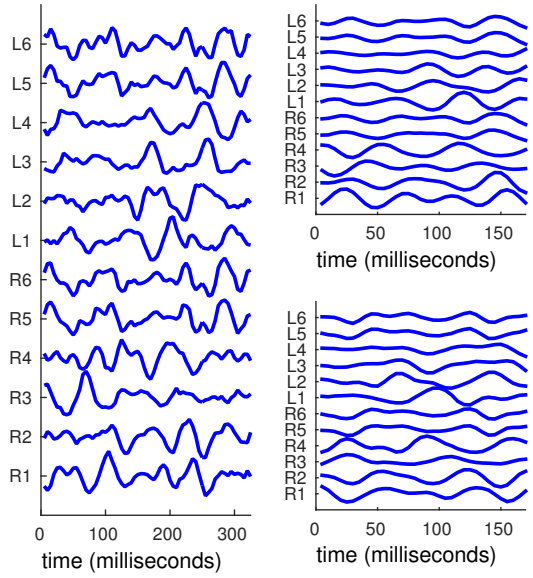

Fig. 6. Example segment of the non-IED class and its layer 1 feature map (multichannel) for the two filters of Fig. 4.

6. REFERENCES

[1] K. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators", in Neural Networks, vol.2, pp. 359–366, 1989.

[2] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", Science Magazine, vol 313, pp. 504-507, 2006.

[3] J. Schmidhuber, "Deep learning in neural networks: An overview", arXiv preprint arXiv:1404.7828, 2014.

[4] A. C. Davison, "Statistical models", Cambridge series in statistical and probabilistic mathematics. vol. 11, Cambridge: Cambridge University press, 2008.

[5] N. G. Polson, M. Heidari and B. Willard, "A statistical theory of deep learning via proximal splitting", arXiv 1509.06061v1, 2015.

[6] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion", in Proceedings of the 27th International Conference on Machine Learning, pp. 3371-3408, ACM, 2010.

[7] G.E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets", Neural Computation, pp. 1527–1554, 2006.

[8] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella and J.Schmidhuber, "Flexible, high performance convolutional neural networks for image classification", in IJCAI, pp. 1237–1242, 2001.

[9] D. F. Wulsin ,J. R. Gupta ,R. Mani , J. A. Blanco and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement", J. Neural Eng. 8(3):036015, Jun. 2011.

[10] R. Yuanfang and Y. Wu, "Convolutional deep belief networks for feature extraction of EEG signal", International Joint Conference on Neural Networks (IJCNN), Beijing, China, 2014.

[11] P. W. Mirowski, Y. LeCun, D. Madhavan, R. Kuzniecky, "Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG" in Machine Learning for Signal Processing, pp.244-249, 16-19 Oct. 2008.

Table 3. Classification accuracy for different approaches (%)

Subject	Temporal	Time-Frequency	CNN1	CNN2
1	78.92	83.54	81.70	87.54
2	89.83	89.75	78.55	87.31
3	83.37	89.75	78.59	85.53
4	69.24	88.18	77.42	84.80
5	70.19	88.81	78.31	89.85
6	81.37	91.04	85.48	90.73
7	76.00	89.68	81.91	86.69
8	77.89	85.68	82.13	90.91
9	80.91	88.79	79.12	90.19
10	67.72	87.66	83.34	90.94
11	75.73	91.52	81.19	88.27
12	79.03	87.83	80.00	89.62
13	76.91	91.00	82.18	89.12
14	80.80	92.41	80.81	90.85
15	89.83	89.75	82.00	92.00
16	94.00	89.00	84.00	93.00
17	85.21	94.37	82.88	92.47
18	61.27	84.00	71.05	88.16
19	49.62	47.12	50.39	50.43
20	84.37	87.50	81.25	86.31
21	59.01	81.37	84.79	91.31
22	83.58	88.61	78.80	87.30
23	68.86	91.23	80.27	87.29
24	86.80	81.35	75.47	85.75
25	79.66	88.98	86.45	92.83
Mean	77.20	86.67	79.48	87.51

[12] T. A. Tzallas, M. G. Tsipouras, D. G. Tsalikakis , E. C. Karvounis, L. Astrakas, S. Konitsiotis and M. Tzaphlidou, "Automated epileptic seizure detection methods : a review study" in Epilepsy - Histological, Electroencephalographic and Psychological Aspects, Dr. Dejan Stevanovic (Ed.), InTech, DOI: 10.5772/31597, 2012.

[13] L. Spyrou, D. Lopez, A. Valentin, G. Alarcon, S. Sanei. "Detection of intracranial signatures of interictal epileptiform discharges from concurrent scalp EEG.", DOI: 10.1142/S0129065716500167, International Journal of Neural Systems, 26(4):1650016 , Feb 2016.