Detection of Interictal Discharges with Convolutional Neural Networks Using Discrete Ordered Multichannel Intracranial EEG

Andreas Antoniades, Loukianos Spyrou, David Martin-Lopez, Antonio Valentin, Gonzalo Alarcon, Saeid Sanei and Clive Cheong Took

Abstract—Detection algorithms for electroencephalography (EEG) data, especially in the field of interictal epileptiform discharge (IED) detection, have traditionally employed handcrafted features which utilised specific characteristics of neural responses. Although these algorithms achieve high accuracy, mere detection of an IED holds little clinical significance. In this work, we consider deep learning for epileptic subjects to accommodate automatic feature generation from intracranial EEG data, while also providing *clinical insight*. Convolutional neural networks are trained in a subject independent fashion to demonstrate how meaningful features are automatically learned in a hierarchical process. We illustrate how the convolved filters in the deepest layers provide insight towards the different types of IEDs within the group, as confirmed by our expert clinicians. The morphology of the IEDs found in filters can help evaluate the treatment of a patient. To improve the learning of the deep model, moderately different score classes are utilised as opposed to binary IED and non-IED labels. The resulting model achieves state of the art classification performance and is also invariant to time differences between the IEDs. This study suggests that deep learning is suitable for automatic feature generation from intracranial EEG data, while also providing insight into the data.

Index Terms—Convolutional neural networks, epilepsy detection, intracranial EEG, multi score class learning.

I. INTRODUCTION

E EG is the most popular tool for the investigation of the brain function. Its high temporal resolution property makes it ideal for a variety of fields that consider time varying signals. EEG has been extremely useful in medical sciences, since it complements other biomarkers for many types of brain diseases and conditions. Detection of epileptic seizure was the first application of EEG-based diagnosis by Hans Berger in 1924. Epilepsy is a chronic condition with the primary characteristic being the occurrence of seizures. Fast changing abnormal electrical activities that occur during seizures (ictal) and between seizures (inter-ictal) make the use of EEG ideal for diagnosis. These types of electrical activity are analysed in order to predict a seizure, detect the seizure onset location, or aid the clinical diagnosis by quantifying properties of the recorded signals. Although EEG has been employed for a number of years, *automated* feature and pattern detection have proved to be a challenging task, due to noise and multi-task brain activities.

Interictal epileptiform discharges (IEDs) originate from deep brain or hippocampus and in many cases are descriptive of the epileptic condition. They are the most reliable biomarkers and are widely used in clinical evaluations. Detection of IEDs has attracted interest from the machine learning and biomedical communities and a variety of algorithms have been developed [1]. These algorithms are based on methods such as template matching [2]-[5], classification [6]-[9], dictionary learning [10], differential operator [11], spike rate [12] and other methods common in the well-established field of spike detection [14], [15]. The common characteristic of all these methods is that a description of an IED/spike signal is obtained either through modeling or similarity measurement with features of interest. This is often facilitated by obtaining useful representations of the signal that can better exploit its structure. Finding the optimal features for IED detection is a challenging yet important problem [13].

Artificial neural networks are computational systems modeled after the human brain that simulate how interconnected neurons in the brain collect, process, and disseminate electrical signals [16], [17]. These models are used as universal model approximators for classification or regression tasks [18] and have seen many successes in the past few decades [19], [20]. Neural networks have been utilised in a wide range of fields including physiological signal analysis through the use of electromyography [21], [22] and EEG [23], [24] signals.

Recent advances in neural networks have been in the form of deep learning which operates based on a hierarchical squashing function to generate high level features at each layer by reducing the dimensionality of data [25]. Convolutional neural networks (CNNs) are deep learning models inspired by the animal visual cortex [26]. CNNs allow the extraction of higher level features from the original input. Traditionally, CNNs have been used for image processing and computer vision and have set new benchmarks for image classification problems [25], [27].

In contrast to their success in image processing, CNNs have been used sparingly in EEG processing. Notable contributions are in areas such as mental load classification [28], detection

A. Antoniades, S. Sanei and C. Cheong Took are with the Department of Computer Science, University of Surrey, GU2 7XH, UK (email: {a.antoniades, s.sanei, c.cheongtook}@surrey.ac.uk)

L. Spyrou is with the School of Engineering, University of Edinburgh, EH9 3FB, UK (email: l.spyrou@ed.ac.uk)

D. M. Lopez is with Kingston Hospital NHS FT, London, UK and King's College London, UK (email: david.martin_lopez@kcl.ac.uk)

A. Valentin is with King's College Hospital, London, UK and King's College London, UK (email: antonio.valentin@kcl.ac.uk)

G. Alarcon is with Hamad Medical Corporation, Doha, Qatar and King's College London, UK (email: galarcon@aol.com)

of P300 [29], prediction of epileptic seizure [30], [31]. The shortcoming in the literature is perhaps due to the blackbox nature of neural networks for clinicians and the limited availability of specialised EEG (such as IED data) to data scientists. To address these issues, we provide a comprehensive review of CNNs and demonstrate how clinically meaningful features manifest themselves within the structure of CNNs. In this way, a specialised clinician can make sense of the data due to the qualitative (and not quantitative) nature of the IED features. To our best knowledge, this is the first time that the IED morphology is detected within the filter weight coefficients in the convolutional layer of a CNN.

A recent study into deep learning has revealed that deep models benefit from out of distribution examples [32]. In other words, transforming the original input samples using a variety of filters in conjunction with separating the original classes into sub-classes enhances the classification performance of deep learning models. We therefore hypothesise that additional subclasses of IEDs can improve the detection of IEDs, over the methodology whereby we only consider the binary IED and non-IED classes. We expect that the increased number of IED classes and their clinical significance (see Section II-D) assist the classifier in learning the features of IEDs better, leading to improved classification rates. To enhance learning, we introduce discrete ordered score labels, as a means of training with moderately different classes. At the end of classification, a threshold layer is designed to categorise the predicted subclasses to IED and non-IED classes. This paper is summarised as follows: In Section II we describe the epileptic dataset and the preprocessing methodology for the intracranial EEG. The training algorithm for our selected models is derived in Section III and a thorough study of the proposed method and our hypothesis are detailed in Section IV. Section V provides the training procedure, parameter selection and empirical results. A discussion of the achieved results and their significance is provided in Section VI. Finally, our manuscript is concluded in Section VIII.

II. DATASET

A. Subjects

In this study, 18 subjects (11 males, 7 females, average age 25.2 years, range $13 - 37$) assessed for temporal lobe epilepsy with scalp and foramen ovale (FO) electrodes at King's College Hospital, London were included. Subjects were submitted for telemetry recording with FO electrodes when their medical history, interictal scalp EEG, neuroimaging and neuropsychological studies were not able to confidently determine the side of seizure onset or there were doubts about lateral temporal or extra-temporal seizure onset. In 10 subjects the seizure onset was located within mesial temporal structures, while in eight subjects it was located in the lateral temporal region. 13 EEG traces were entirely recorded during wakefulness and the remaining also included periods of slow wave sleep (stages I-II) (Subjects S2, S9, S10, S13, S15).

B. Electrode placement

The scalp EEG was obtained with the 'Maudsley' electrode placement system [33] using 20 standard chlorided silver cup electrodes. Each subject had two flexible bundles of six intracranial electrodes inserted through the subjects FOs under general anesthesia and fluoroscopic control. The procedure waw performed following the technique described by Stodieck [34] and their positions were radiographically confirmed. Each electrode consisted of a 0.1 mm insulated stainless steel wire. The three deepest electrodes had recording contacts of 3 mm long while for the most superficial electrodes the recording contacts were 5 mm long. The distance between the two most superficial recording contacts was 15 mm and 10 mm for the rest of the contacts in the electrode. For each FO electrode bundle, the two deepest electrodes were located close to mesial temporal structures.

C. Recording system

A Telefactor Beekeeper cable telemetry system of 32 channels was used. Data was digitised at 200 Hz and bandpass filtered ([0.3 - 70] Hz). The system input range was 2 mV and data were digitised using a 12 bit analog-to-digital converter (an amplitude resolution of 0.488 V). The FO data were recorded with respect to Pz as common reference. For each subject, a period of 20 min of scalp EEG and FO interictal recordings was used.

D. IED scoring

Scoring of the IED was performed by an expert epileptologist (DML) based on the spatial distribution and morphology of the observed waveforms. These were evaluated in the background context as described in [35], and following the standard definitions for epileptiform pattern, spike and sharp wave of the International Federation of Clinical Neurophysiology [36] and currently accepted EEG descriptions were taken into account for scoring. Each trial was 65 sample long which is equivalent to 325ms and was given a certainty score $(0 - 4)$ and categorised to one of the following:

0: Non-physiological and physiological artifacts, physiological 'sharpened/spiky' activities (vertex waves, K-complexes), and low amplitude irregularities barely distinguishable from the background activity and restricted to $1 - 2$ channels.

1: Sharp wave (restricted to at least 3 channels).

2: Broadly distributed sharp wave (> 3 channels).

3: Spike (restricted to at least three channels).

4: Broadly distributed spikes or spike and wave complexes (> 3 channels), interictal bursts of spikes or sharp waves.

A summary of the scored data can be found in Table I.

E. Notation

We treat each of the six channels of the intracranial EEG data as a univariate time sequence, as the data were sampled at successive points in time at uniform time intervals. As such, we denote each channel $\mathbf{c}_j$ as $\langle \mathbf{c}_j | j \in \text{Channels} \rangle$ where $\mathbf{c}_j = \{c_{j_{t_1}}, c_{j_{t_2}}, .., c_{j_{t_N}}\}$ and $\mathbf{c}_{j_k}$ is the kth sample in channel j. CNNs utilise filters that use windowing to sweep through the data of each channel and produce a new multichannel time series that captures the *temporal similarity*
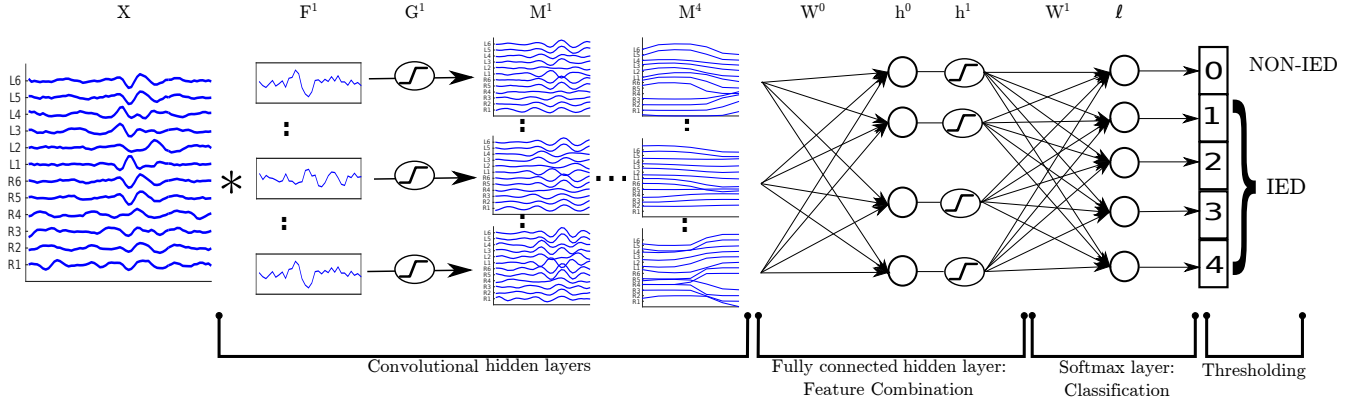
Fig. 1: Resulting network after initial optimisation of parameters. $\mathbf{X}$ is the input, $\mathbf{F}^j$ are the filters for layer j, $\mathbf{M}^j$ are the feature maps for layer j, $\mathbf{W}^0$ is the weight matrix for the fully connected hidden layer and, $\mathbf{h}^0$ and $\mathbf{h}^1$ represent the pre-activation and post-activation features for the fully connected hidden layer, $\mathbf{W}^1$ and ℓ are the weight matrix and activation of the logistic regressor.

TABLE I: Summary of the data scoring process.

Subject	No. of trials	Subject	No. of trials
S1	684	S10	448
S2	100	S11	1696
S3	144	S12	1906
S4	330	S13	1658
S5	316	S14	1082
S6	944	S15	520
S7	398	S16	1212
S8	634	S17	228
S9	682	S18	236
Score	No. of trials	Score	No. of trials
0	6609	3	2175
1	1389	4	915
2	2130		

between EEG and the filter. At each step, filter f_o is convolved with each subsequence $\mathbf{s}$ of the same length. For example, the first subsequence of the third channel is represented as: $\mathbf{s}_3^1 = \{c_3^1, c_3^2, .., c_3^{|f|}\}$, where $|\cdot|$ is the cardinality operator. Recall that we employ 1-d filters to avoid channel mixing as it is not relevant for time series.

III. CONVOLUTIONAL NEURAL NETWORKS FOR EEG LEARNING

Based on neural networks and the convolution operation, CNNs perform convolution at each convolutional hidden layer between the input and the weights. However, unlike normal neural networks, CNNs can have many sets of weights at each layer, also called filters. Each filter sweeps the entirety of the input space to produce feature maps via convolution. The feature maps of previous layers are used as input to the next layer, leading to a hierarchical feature learning. Fig. 1 illustrates the architecture of our network, whereby the filters operate within the convolutional hidden (CNH) layers and traditional weights combine the feature representation learnt in CNH layers in the fully connected hidden layer.

In image processing, these filters are typically 2-d by design and are used to capture the spatial information of pixels. In this work, we exploit these filters to capture the *temporal* information of the EEG. To do so, CNN convolution is performed by convolving each electrode signal with a 1-d filter and adding a bias term to generate the feature map. This can be expressed in two ways:

$$\mathbf{G}_o^k = \sum_{g=1}^{\Phi'} \mathbf{F}_o^k * \mathbf{M}_g^{k-1} + b^k \tag{1}$$

$$\mathbf{G}_{j,t+1,o}^k = \sum_{g=1}^{\Phi'} \sum_{i=1}^{L} \mathbf{F}_{i,o}^k \mathbf{M}_{j,i+t,g}^{k-1} + b^k$$
$$\forall \ t = \{0,..,T-L\}$$
$$\forall \ j = \{1,..,12\} \tag{2}$$
$$\forall \ o = \{1,..,\Phi\}$$

where $*$ is the convolution operator, L is the filter length, T is the length of the data segment, Φ and Φ' are the number of filters in the current and previous layer respectively (1 for the first layer), $\mathbf{F}_o^k$ is the oth filter at layer k, $\mathbf{G}^k$ are the feature maps at layer k before applying the non-linear function, $\mathbf{M}^{k-1}$ are the feature maps at layer $k-1$ after applying the non-linear function, t denotes the time index and j represents the jth channel.

To introduce non-linearity to the network, the result of Eq. (2) is passed through a non-linear activation function ϕ.

$$\mathbf{M}_o^k = \phi(\mathbf{G}_o^k) \tag{3}$$

A number of activation functions have been proposed, however, for the purposes of this work we consider the hyperbolic tangent function defined as:

$$\phi(\theta) = \tanh(\theta) = \frac{\sinh(\theta)}{\cosh(\theta)} = \frac{\exp(\theta) - \exp(-\theta)}{\exp(\theta) + \exp(-\theta)} \tag{4}$$

where $\exp(\cdot)$ is the exponential function. Allowing for a number of convolutional layers (as shown on the L.H.S of Fig. 1) enables CNNs to learn complex features as combinations of filters from the preceding layers. The generated features from

the convolutional layers are passed through a fully connected hidden layer, such that:

$$\mathbf{h}^0 = (\mathbf{W}^0 \text{vec}(\mathbf{M}^{K^\intercal}) + \mathbf{b}) \tag{5}$$

$$\mathbf{h}^1_j = \phi(\mathbf{h}^0_j) \tag{6}$$

where $(\cdot)^\intercal$ is the transpose operator, K is the last convolutional layer, $\mathbf{W}^0$ and $\mathbf{b}$ are the weight matrix and bias of the fully connected hidden layer respectively, see L.H.S of Fig. 1. Note that $\mathbf{M}^K$ is a 3-d tensor. In order to pass the generated features to the fully connected hidden layer the vec$(\cdot)$ function is used so that $\mathbf{M}^K$ is flattened to a vector. This is performed by repeatedly concatenating each column of the feature map matrices into a vector.

Finally, the output of the hidden layer is passed through a logistic regression model for classification of the intracranial EEG generated features. We use a *one vs all* approach so that a number of binary classification problems are solved in the case of discrete ordered score labels.

$$\boldsymbol{\ell} = \mathbf{W}^1 \mathbf{h}^1 + \mathbf{b} \tag{7}$$

where $\mathbf{W}^1$ and $\mathbf{b}$ are the weights and bias for the logistic model respectively. $\boldsymbol{\ell}$ is a vector containing the probabilities of a sample being in a particular class, normalised using the softmax function:

$$z_i = \text{softmax}(\ell_i) = \frac{\exp(\ell_i)}{\sum_{c=1}^C \exp(\ell_c)} \tag{8}$$

where $\mathbf{z}$ is the result of the softmax function normalisation of $\boldsymbol{\ell}$ and ℓ_i is the ith element of $\boldsymbol{\ell}$, corresponding to the probability of being in the ith class.

For classification the model outputs the index of the maximum value, from 0 (non-IED) to 4 (IED). The model is trained using the backward propagation algorithm. We consider the cross entropy as the error function for the training of the proposed model.

$$E = -\sum_{c=1}^C z_c \log(y_c) \tag{9}$$

where $\log(.)$ is the logarithm operator to the base e, C is the number of classes, z_c is the generated output and y_c the desired output for class c.

The gradient of a given cost function E can be computed by backpropagation for each component of our network. For the logistic model, the partial derivative $\frac{\partial E}{\partial \ell_c}$ is calculated as:

$$
\begin{aligned}
\delta^{\boldsymbol{\ell}}_c = \frac{\partial E}{\partial \ell_c} &= \frac{\partial E}{\partial z_i}\frac{\partial z_i}{\partial \ell_c} \\
&= -\sum_{i=1}^C y_i \frac{1}{z_i}\frac{\partial z_i}{\partial \ell_c} \\
&= -y_c(1-z_c) - \sum_{i \neq c}^C y_i \frac{1}{z_i}(-z_i z_c) \\
&= z_c - \sum_{i=1}^C y_i y_c = z_c - y_c
\end{aligned}
\tag{10}
$$

In terms of the fully connected hidden layer the partial derivative $\frac{\partial E}{\partial \mathbf{h}^0}$ is calculated as:

$$
\begin{aligned}
\delta^h_j = \frac{\partial E}{\partial \mathbf{h}^0} &= \frac{\partial E}{\partial \mathbf{z}}\frac{\partial \mathbf{z}}{\partial \ell_j}\frac{\partial \ell_j}{\partial \mathbf{h}^1}\frac{\partial \mathbf{h}^1}{\partial \mathbf{h}^0} \\
&= \delta^{\boldsymbol{\ell}}_j \mathbf{w}^{1^\intercal}_j \frac{1}{\cosh^2(\mathbf{h}^0)}
\end{aligned}
\tag{11}
$$

For the last convolutional layer the partial derivative $\frac{\partial E}{\partial \mathbf{G}^K_{j,i,o}}$ is derived as follows:

$$
\begin{aligned}
\boldsymbol{\Delta}^K_o = \frac{\partial E}{\partial \mathbf{G}^K_o} &= \frac{\partial E}{\partial \mathbf{h}^0}\frac{\partial \mathbf{h}^0}{\partial \mathbf{M}^K_o}\frac{\partial \mathbf{M}^K_o}{\partial \mathbf{G}^K_o} \\
&= \sum_j \Big(\underbrace{\delta^{h^\intercal} * \text{ROT180}(\mathbf{W}^0_{[:,s(o-1)+1:so]})}_{\rho} \Big)_j \; \frac{1}{\cosh^2(\mathbf{G}^K_o)}
\end{aligned}
\tag{12}
$$

where the function ROT180$(\cdot)$ is a rotation at 180 degrees and equivalent to time reversal used in 1-d linear convolution. ':' denotes all rows of $\mathbf{W}^0$, s refers to the length of each vectorised map (in this case $7 \times 12 = 84$); the formula $s(o-1)+1$ corresponds to the start index and so corresponds to the end index - which effectively selects the appropriate columns (weight coefficients) that relate to the feature map of the oth filter. The vector ρ (of size $s \times 1$) is summed up to yield the sum of contributions of all elements in the oth feature map. Note that $\boldsymbol{\Delta}^K_o$ is a matrix of the same size as $\mathbf{G}^K_o$. This can be generalised for all convolutional layers as:

$$
\begin{aligned}
\boldsymbol{\Delta}^{k-1}_o = \frac{\partial E}{\partial \mathbf{G}^k_o}\frac{\partial \mathbf{G}^k_o}{\partial \mathbf{M}^{k-1}_o}\frac{\partial \mathbf{M}^{k-1}_o}{\partial \mathbf{G}^{k-1}_o} \\
= \sum_i \sum_j \Big(\boldsymbol{\Delta}^k_o * \text{ROT180}(\mathbf{F}^k_o) \Big)_{ji} \; \frac{1}{\cosh^2(\mathbf{G}^{k-1}_o)}
\end{aligned}
\tag{13}
$$

The weights for each component at every iteration are updated as follows:

$$\text{Logistic: } \mathbf{w}^1_c(n+1) = \mathbf{w}^1_c(n) - \alpha \delta^{\boldsymbol{\ell}}_c \mathbf{h}^1 \tag{14}$$

$$\text{Hidden: } \mathbf{w}^0_i(n+1) = \mathbf{w}^0_i(n) - \alpha \delta^h_i \text{vec}(\mathbf{M}^{K^\intercal}) \tag{15}$$

$$\text{Convolutional: } \mathbf{f}^k_o(n+1) = \mathbf{f}^k_o(n) - \alpha \boldsymbol{\Delta}^k_o * \mathbf{M}^{k-1}_o \tag{16}$$

where the transpose of $\mathbf{w}^1_i$ corresponds to the ith row of the matrix $\mathbf{W}^1$, the transpose of $\mathbf{w}^0_i$ corresponds to the ith row of the matrix $\mathbf{W}^0$, $\mathbf{f}^k_o$ is the oth filter/column of the filter matrix $\mathbf{F}^k$ at layer k and α is the learning rate. Observe that for the last convolutional layer, the convolution between $\boldsymbol{\Delta}^k_o$ (of size 12×7) and $\mathbf{M}^{k-1}_o$ (of size 12×12) results in a vector of size (1×6). This is because convolution of two matrices $\mathbf{A}$ (of size $a1 \times a2$) and $\mathbf{B}$ (of size $b1 \times b2$) results in a matrix of size $((b1 - a1 + 1) \times (b2 - a2 + 1))$.

Following each convolutional layer, a pooling layer is generally utilised to provide a squashing function. Specifically, the pooling function maps an area (also called window) of the generated feature map to a scalar value. Pooling has proven to achieve invariance to transformation (such as rotation) and more compact representation of images by reducing the image size. A number of pooling strategies have been proposed with
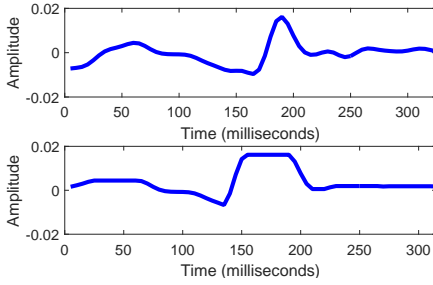
Fig. 2: Max pooling with a window size of 3 samples for an IED example. The top plot represents the original IED segment while the bottom plot represents the *deformed* segment after max pooling.

the most prominent one being max-pooling where $\max(\mathbf{u})$ returns the maximum sample in the vector or matrix $\mathbf{u}$. Although pooling has proven to be beneficial to deep learning, it also modifies the input-output space. In the case of IEDs, this alters the morphology of an epileptic spike, see Fig. 2, which could degrade the performance. This observation was confirmed by simulations.

IV. BINARY VS MULTICLASS IED DETECTION

Our method challenges the assumption that binary labeling is sufficient to represent the complexity of the data for deep learning [32]. Examples that lie close to the threshold of a logistic regressor have a high probability to be misclassified. To support our hypothesis, we provide two methods to demonstrate the proximity of IED-1 and IED-2 segments to non-IED compared to IED-3 and IED-4 segments. The first method provides a *quantitative* measure while the second provides a visual, *qualitative* measure on the proximity of IED and non-IED segments.

First, a special case of the Minkowski distance metric called Chebyshev distance is used to show the distance between the different scored IEDs, emphasising that IED-1 class is much closer to the non-IED class than any other IED class. The Chebyshev distance is the generalised L_p-norm of the difference between two vectors when p moves towards infinity:

$$d : (\mathbf{x}, \mathbf{y}) \rightarrow ||\mathbf{x} - \mathbf{y}||_\infty = \lim_{p \to \infty} \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (17)$$
$$= \max_i |x_i - y_i|$$

where $\mathbf{x}$ and $\mathbf{y}$ are vectors and $|| \cdot ||$ is the norm of a vector. This is more appropriate in our comparison as the p value is set.

Since the scored classes do not have the same number of examples, see Table I, the first 914 examples from each class were selected in order to calculate the Chebyshev distance. The distance between the examples of different scored classes are presented in Table II. Although examples in the IED-1 class are classified as IED, it is clear that these segments are closer to the non-IED class than any other class due to the smaller distance.

Second, to provide a visualisation of the closeness between the non-IED and IED-1 classes, we used t-distributed

TABLE II: Chebyshev Distance of data segments.

	Non-IED	IED-1	IED-2	IED-3	IED-4
Non-IED	0	**0.0583**	0.4774	0.5756	0.6594
IED-1	**0.0583**	0	0.4809	0.5785	0.6620
IED-2	0.4774	0.4809	0	0.3215	0.4549
IED-3	0.5756	0.5785	0.3215	0	0.3218
IED-4	0.6594	0.6620	0.4549	0.3218	0

Stochastic Neighbour Embedding (tSNE)[37]. As a dimensionality reduction technique, tSNE is useful for visualising high dimensional data. It is a supervised method that takes as input a dataset X, and returns a 2-d or 3-d projection of the data. Given a dataset $X = [x_1, x_2, .., x_n]$, it implements a function $f(X) \rightarrow Y$ where Y can be plotted in a 2-d or 3-d space. Based on Stochastic Neighbour Embending, tSNE calculates and converts the high dimensional Euclidean distances between the data points into probabilities and is defined as:

$$p_{j|i} = \frac{\exp(-||x_i - x_j||^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-||x_i - x_k||^2 / 2\sigma_i^2)} \quad (18)$$

where σ_i is the variance of the Gaussian centered at x_i. For the low-dimensional space Y, Gaussian neighbourhoods with a fixed variance, in this case $\frac{1}{\sqrt{2}}$, are used. Therefore, for the low-dimensional y_i and y_j counterparts of the high-dimensional datapoints x_i and x_j, a conditional probability $q_{j|i}$ can be computed as follows:

$$q_{j|i} = \frac{\exp(-||y_i - y_j||^2)}{\sum_{k \neq i} \exp(-||y_i - y_k||^2)} \quad (19)$$

For our visualisation we have set the dimensionality of Y to 2, since a 2-d plot is more straightforward to interpret. The distributions in Eq. (18) and Eq. (19) are matched by minimising a cost function of Kullback-Leibler divergence over neighbours for each object.

$$C_2 = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} \quad (20)$$

For more information about tSNE, we refer the reader to [37].

The resulting visualisation from tSNE is presented in Fig. 3. To show the closeness of IED-1, IED-2 and non-IED segments the figure is divided as follows:
Fig 3(a) presents the visualisation for IED and non-IED segments with a different marker for IED-1 segments which overlap clearly with the class of non-IED segments.

Finally, Fig. 3(b) presents the feature space for all scored classes; we can also observe the overlap between IED-2 and non-IED segments.

Both the quantitative analysis based on Chebyshev distance and the qualitative t-SNE analysis have shown the proximity of IED-1 and IED-2 to the non-IED class. This can introduce some bias in the binary classification of IEDs. As such, we consider multiclass IEDs to minimise the interference of IED-1 and IED-2 on the classification task.

(a) Three class feature space.
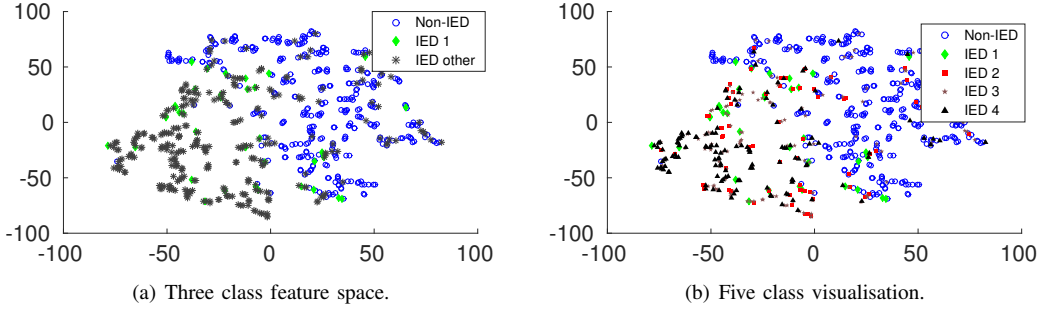
(b) Five class visualisation.
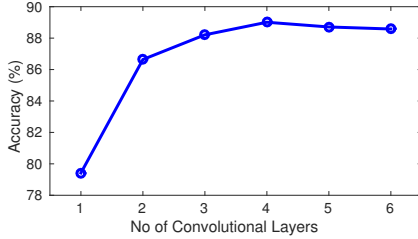
Fig. 3: Data feature space using tSNE.

Fig. 4: Initial simulations with different number of convolutional layers. The greatest increase in accuracy was between 1 and 2 layers, while the performance plateaued with 4 layers.

V. EXPERIMENTS

A. Training procedure

The leave-subject-out training methodology along with two-fold validation was used for all the simulations. For each simulation, the data of a single subject was used for testing. The remaining 17 subjects were divided into two folds, training and validation sets. Using early stopping, competing algorithms were trained until the validation error stopped improving. Once the first fold was completed, the validation set was used for training and vice versa. This procedure was repeated until all the subjects were used as test data for both folds. The results presented in Section V were averaged over the two folds for each subject. Both the binary labelled (IED and non-IED) and scored data were considered for all competing algorithms using the above training procedure. The models were built using the Theano framework [38] and run on a GTX 980ti.

B. Parameter selection

A number of simulations were conducted initially to obtain the optimal neural network configuration. Fig. 4 shows the performance of CNNs with different topologies. It is apparent that using more than 4 convolutional layers was not beneficial as the performance no longer improved. Therefore, the additional computational complexity entailed to more than 4 layers cannot be justified; as a result a CNN with 4 convolutional layers was selected. Additionally, pooling was omitted as the networks trained with max-pooling had on average lower accuracy than those trained without, possibly due to the loss of information as the result of the change in the morphology of the IED segments, see Fig. 2. The length of the first layer's

filters was set to 32 samples which corresponds to $160ms$; this duration was adequate to capture the main part of an IED waveform. For the following layers we gradually reduced the length of the filters to capture finer details of IEDs [39].

The resulting optimal model consisted of 4 convolutional layers followed by a fully connected hidden network and a logistic regressor. We introduced a threshold layer at the end of our model so that the discrete ordered score subclasses were mapped to the binary non-IED and IED classes. The topology of the proposed model is depicted in Fig. 1. Note that the output of the final convolutional layer was unfolded from a 3-d tensor to a 1-d vector. For clarity, the training parameters of all competing algorithms are summarised in Table III while information on the dimensionality of our data structures are presented in Table IV.

TABLE III: Training parameters for the considered methods.

Parameter/Method	Wavelet [41]	TD [35]	TF [35]	CNN
Convolution layers	-	-	-	4
No. of filters	-	-	-	40, 30, 20, 10
Filter size	-	-	-	32, 16, 8, 6
Sigmoidal layer size	-	-	-	500
Network parameters	-	-	-	1980
Classifier features	240	780	756	500
Total hidden layers	0	0	0	5

TABLE IV: Dimensions of variables.

Type	Size
$\mathbf{X}$	12 channels $\times$ 65 samples
$\mathbf{F^1}$	32 coefficients $\times$ 40 filters
$\mathbf{M^1}, \mathbf{G^1}$	12 channels $\times$ 34 samples $\times$ 40 maps
$\mathbf{F^2}$	16 coefficients $\times$ 30 filters
$\mathbf{M^2}, \mathbf{G^2}$	12 channels $\times$ 19 samples $\times$ 30 maps
$\mathbf{F^3}$	8 coefficients $\times$ 20 filters
$\mathbf{M^3}, \mathbf{G^3}$	12 channels $\times$ 12 samples $\times$ 20 maps
$\mathbf{F^4}$	6 coefficients $\times$ 10 filters
$\mathbf{M^4}, \mathbf{G^4}$	12 channels $\times$ 7 samples $\times$ 10 maps
$\mathbf{W^0}$	500 neurons $\times$ 840 coefficients
$\mathbf{h^0}, \mathbf{h^1}$	500 features $\times$ 1 sample
$\mathbf{W^1}$	5 classes $\times$ 500 neurons
$\boldsymbol{\ell}$	5 classes $\times$ 1 probability

C. Dataset balancing

Initial experimentation revealed that due to the unbalanced dataset, classifiers trained with the scored data became biased
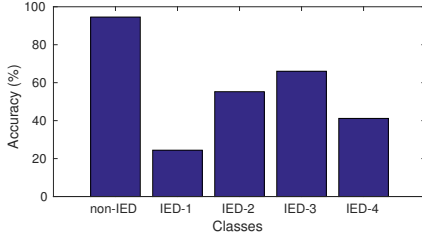
Fig. 5: Classification accuracy for CNN using the multiclass score labels.

Fig. 6: ROC curves for the different methods.

towards non-IED segments as they represented the majority of the available data. In an effort to balance the dataset, we employed undersampling for the non-IED class using random sampling, this method is known to be the most effective approach for bioinformatic data [40]. Reducing the number of non-IED waveforms should not deteriorate the classification rate on the IED segments. The same subsets of undersampled non-IED segments were used by the competing algorithms to ensure objectivity of the presented results.

VI. DISCUSSION

First, it is instructive to observe the classification accuracy for the CNN trained with the scored data but without the threshold layer, as presented in Fig. 5. It is clear that IED-1 is the most contested class as it represents a transition state between IED and non-IED segments; this is consistent with our hypothesis, see Section IV for more details. An index of the methods considered in this section is available in Table V. The achieved accuracy for each subject and method is presented in Table VI.

TABLE V: Model Notations.

Notation	Description
Wavelet	A logistic regression model trained with wavelet data and binary labels.
TD	A logistic regression model trained with time-domain data.
TF	A logistic regression model trained with handcrafted time-frequency features.
CNN	A convolutional neural network trained with time-domain data.
Bin	A model trained with binary labelled data.
Multi	A model trained with discrete ordered score data.

Remark 1: It is clear from Table VI, that the CNN Multi model trained using our approach has, on average, outperformed all other methods. For rigour, a statistical analysis for the performance of each method is also provided in Table VII.

Remark 2: Models trained with binary labels were less likely to misclassify a segment as a false positive IED, whereas models trained with discrete ordered score labels were superior in terms of minimising false negatives, see Table VII.

Remark 3: The wavelet approach produced good results for a minority of the subjects, i.e. S8 and S17, but had the lowest overall accuracy. In terms of the competing algorithms, the TF Bin model achieved the best performance, but still did not outperform the CNN models.

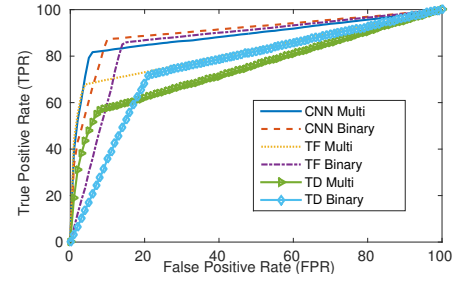**Remark 4**: The CNN Multi model seems to hold a balance

by retaining a relatively small number of TP and FP classifications. The Receiver Operating Characteristic (ROC) analysis presented in Fig. 6 also supports this.

Remark 5: Observing the ROC curve leads to the conclusion that models trained with the binary labels leaned towards the True Positive Rate (TPR) while models trained with discrete ordered scores leaned towards the False Positive Rate (FPR). The deep CNN Multi model provided the best balance between the two and greatest area under the curve (**0.900**).

The analysis showed that certain types of IED presented sharp waves which were closer to non-IED segments than broad IED spikes over 3 channels, therefore, they should not share the same class as IED segments of greater amplitude in terms of training. To allow the deep model to benefit from this insight, we have replaced binary labels with ordered ordinal score labels. The network learns the IEDs of different scales and amplitudes as individual IED subclasses. Finally, a threshold layer was utilised to convert the predicted subclasses to binary IED and non-IED. Using this methodology, we have observed a 3% increase in accuracy from our CNN, when compared to a model trained with the binary labels. Convolved filters generated at the deepest layers showed prominent features, beneficial for the classification of IED segments, as verified by our expert clinicians. Specifically, prominent IED segments for Subjects 2 and 11 were compared with the learnt filters at different layers in Fig. 7 and Fig. 8 respectively. The differences in IEDs are evident and may express different neural generation mechanisms [42], [43]. Traditionally, spikes and sharp waves are assigned equal etiologic and clinical significance in the evaluation of a subject with epilepsy [44]. However, it is well known that the morphology of an IED can provide clinical insights to the subject's condition [45]-[47]. Consistent changes in amplitude, duration and sharpness have been described in different sleep stages [48], during pharmacological treatment [49], [50] and, furthermore, after achievement of seizure control, revealing significant decrease in spike amplitude, duration and sharpness [51]. Therefore, the ability to discriminate IEDs based on their morphology has an additional value in the evaluation of treatment efficacy.

VII. COMPUTATIONAL COMPLEXITY

It is instructive to provide a comparative analysis on the computational complexity of the competing methodologies. As such, we have considered the three main stages of each

TABLE VI: Classification accuracy per subject for different approaches.

Subj.	Wavelet	TD Bin	TD Multi	TF Bin	TF Multi	CNN Bin	CNN Multi
1	0.6798	0.7558	0.8048	0.9137	0.9137	0.9371	**0.9510**
2	0.8200	0.8800	0.8500	0.8850	0.8950	0.8700	**0.9600**
3	0.7431	0.7674	0.7812	0.9167	0.9375	0.9375	**0.9653**
4	0.7970	0.7879	0.7758	0.8788	0.8470	0.8424	**0.9076**
5	0.7278	0.7358	0.7358	0.8339	0.8196	0.8639	**0.8940**
6	0.6441	0.7505	0.7903	0.8957	0.8798	0.8962	**0.9237**
7	0.6608	0.8028	0.7902	0.8580	0.7977	0.8266	**0.8832**
8	0.8486	0.7390	0.7256	0.8904	0.8604	**0.9290**	0.8856
9	0.6862	0.7273	0.7603	0.8702	0.8350	0.8651	**0.9098**
10	0.7679	0.8013	0.7667	0.9129	0.9498	0.9040	**0.9576**
11	0.7382	0.7765	0.7798	0.8818	0.8113	**0.8945**	0.8750
12	0.7833	0.7977	0.7972	0.8547	0.7985	**0.8762**	**0.8762**
13	0.6628	0.6674	0.7099	0.8477	0.7796	**0.8613**	0.8589
14	0.5878	0.6945	0.6945	0.8549	0.8068	0.8429	**0.8946**
15	**0.5058**	0.5038	0.4942	0.4837	0.4971	0.4827	0.4990
16	0.7624	0.7954	0.7987	0.8886	0.8630	0.8705	**0.9134**
17	0.8816	0.7632	0.7412	0.8772	0.9211	0.8465	**0.9364**
18	0.7203	0.8347	0.7903	0.8771	0.8347	0.9195	**0.9301**
Mean	0.7232	0.7545	0.7548	0.8567	0.8360	0.8592	**0.8901**

TABLE VII: Statistical results for different approaches.

Method	TP	FP	FN	TN	Precision	Sensitivity	f-measure	Specificity	AUC
Wavelet	9778	3428	4196	9034	0.74	0.70	0.72	0.72	0.733
TD Bin	11311	1927	1919	11279	0.85	0.85	0.85	0.85	0.755
TD Multi	12714	4313	516	8893	0.75	0.96	0.84	0.67	0.751
TF Bin	10417	3740	2813	9466	0.74	0.78	0.76	0.72	0.857
TF Multi	12204	5766	1026	7440	0.68	0.92	0.78	0.56	0.825
CNN Bin	11874	1729	1356	11477	0.87	0.90	0.88	0.87	0.887
CNN Multi	12463	2472	767	10734	0.83	0.94	0.88	0.81	**0.900**
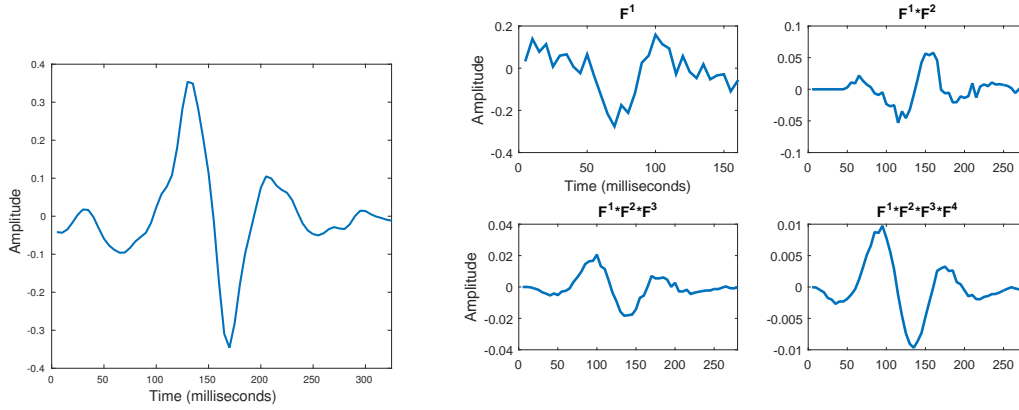
(a) The morphology of a prominent IED segment.

(b) Learnt filter coefficients for each convolutional layer.

Fig. 7: Similarity between an IED segment and learnt filters for Subject S2.

algorithm, i.e. feature generation, feature combination, and classification. The computational complexities of these three stages are shown in Table VIII. As expected, TD has the lowest computational cost, whereas our proposed CNN model has the highest. Moreover, the theoretical time complexity is approximately in agreement with the experimental running time. For example, TD is approximately 40 times less expensive than the CNN ($423/10$), while the theoretical time complexity analysis indicates that the TD method is 50 times less complex as shown below:

$$\frac{O(TJN\Phi)}{O(TJNC)} = \frac{O(\Phi)}{O(C)} = \frac{100}{2} = 50$$

This is not surprising, as the theoretical Big Oh notation reflects the upper limit in terms of time complexity. The wavelet method provides a middle ground in terms of computational complexity, but is the lowest in terms of accuracy. The higher computational cost of our proposed CNN is compensated by a 14% increase in accuracy. Moreover, our approach circumvents the need to handcraft features in EEG data, a time-consuming and difficult task for clinical experts [52], [53].

VIII. CONCLUSIONS

Indeed, deep learning models benefit from moderately different classes, and IED detection is no exception. Using
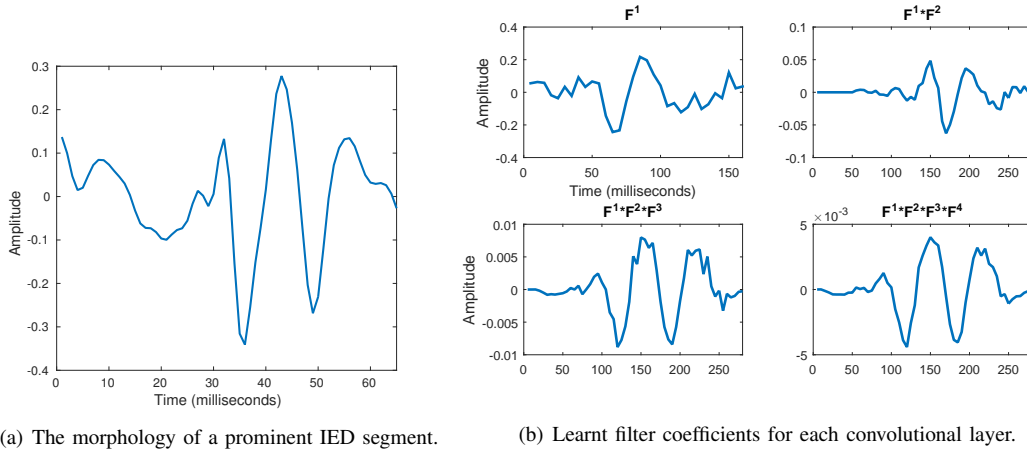
(a) The morphology of a prominent IED segment.

(b) Learnt filter coefficients for each convolutional layer.

Fig. 8: Similarity between an IED segment and learnt filters for Subject S11.

TABLE VIII: Computational Complexity Analysis.

Method	Feature Generation	Feature Combination	Classification	Big Oh	Running Time (m)	Accuracy
Wavelet	T^2NJ	N/A	HNC	$O(T^2NJ)$	48	0.7232
TD Bin	N/A	N/A	$TJNC$	$O(TJNC)$	10	0.7545
TD Multi	N/A	N/A	$TJNC$	$O(TJNC)$	11	0.7548
TF Bin	$(TlogT)JN$	N/A	HNC	$O(HNC)$	11	0.8567
TF Multi	$(TlogT)JN$	N/A	HNC	$O(HNC)$	13	0.8360
CNN Bin	$(T-t+1)JN\Phi$	$(V+H)N$	HNC	$O(TJN\Phi)$	423	0.8592
CNN Multi	$(T-t+1)JN\Phi$	$(V+H)N$	HNC	$O(TJN\Phi)$	436	0.8901

Average computational complexity for each method per subject. T is the number of samples per trial, J is the number of channels, N is the number of trials, H is the number of neurons in the hidden layer, V is the number of vectorised features, C is the number of classes and Φ is the number of filters.

linearly ordered score labels we trained a deep convolutional network and found that our model outperformed a deep model trained with binary labels. This approach involved training the model for different subclasses of IEDs and then converting the predictions to binary IED and non-IED classes. We have also shown how certain learnt filters converge towards prominent IED segments found in the subject group. Our contribution is three-fold, first we demonstrated that deep learning can be used for automatic feature generation and match the accuracy of models trained using features *handcrafted* by experts. Second, we established that deep learning can assimilate a richer collection of information within complex EEG data, i.e. linear ordered score classes. Finally, the automatic discrimination of IED morphology in convolved filters can be beneficial for evaluating treatment effectivity of epileptic subjects, as discussed in Section VI and confirmed by expert clinicians.

REFERENCES

[1] S. Sanei, "Adaptive Processing of Brain Signals", 1st ed., Wiley, May 2013.

[2] S. S. Lodder and M. J. A. M. van Putten, "A selfadapting system for the automated detection of inter-ictal epileptiform discharges ", PLoS ONE 9, p. e85180, 2014.

[3] F. Grouiller, R. Thornton, K. Groening, L. Spinelli, J. Duncan, K. Schaller, M. Siniatchkin, L. Lemieux, M. Seeck, C. Michel and S. Vulliemoz, "With or without spikes: localization of focal epileptic activity by simultaneous electroencephalography and functional magnetic resonance imaging", IBrain 134, pp. 2867–2886, 2011.

[4] J. Zhanfeng, T. Sugi, S. Goto, X. Wang and M. Nakamura, "Multichannel template extraction for automatic EEG spike detection", International Conference on Complex Medical Engineering, pp. 179–184, 2011.

[5] K. Vijayalakshmi and A. M. Abhishek, "Spike detection in epileptic patients EEG data using template matching technique", International Journal of Computer Applications 2, pp. 5-8, 2010.

[6] J. Zhou, R. J. Schalkoff, B. C. Dean and J. J. Halford, "Morphology-based wavelet features and multiple mother wavelet strategy for spike classification in EEG signals", international conference of the IEEE Engineering in Medicine and Biology society, pp. 3959–3962, 2012.

[7] L. Yinxia, Z. Weidong, Y. Qi and C. Shuangshuang, "Automatic Seizure Detection Using Wavelet Transform and SVM in Long-Term Intracranial EEG", IEEE Trans. Neural Syst. Rehabil. Eng., vol. 20, no. 6, pp. 749–755, November 2012.

[8] O. Makeyev, X. Liu, H. Luna-Munguia, G. Rogel-Salazar, S. Mucio-Ramirez, Y. Liu, et al., "Toward a noninvasive automatic seizure control system in rats with transcranial focal stimulations via tripolar concentric ring electrodes", IEEE Trans. Neural Syst. Rehabil. Eng., vol. 20, pp. 422–431, 2012.

[9] S. Ghosh-Dastidar, H. Adeli and N. Dadmehr, "Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection", IEEE Trans. on Biomedical Engineering, vol. 54(9), pp. 1545–1551, 2007.

[10] L. Spyrou and S. Sanei, "Coupled dictionary learning for multimodal data: An application to concurrent intracranial and scalp EEG", IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2349–2353, 2016.

[11] K. Majumdar and P. Vardhan, "Automatic seizure detection in ECoG by differential operator and windowed variance", IEEE Trans. Neural Syst. Rehabil. Eng., vol. 19, no. 4, pp. 356–365, Aug. 2011.

[12] S. Li, W. Zhou, Q. Yuan and Y. Liu, "Seizure prediction using spike rate of intracranial EEG", IEEE Trans. Neural Syst. Rehabil. Eng., vol. 21, no. 6, pp. 880–886, Oct. 2013.

[13] A. T. 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", Electroencephalographic and Psychological Aspects, ed. D. Stevanovic 2009. Available from: http://www.intechopen.com/books/epilepsyhistological-electroencephalographic-andpsychological-aspects/automated-epileptic-seizuredetection-methods-a-review-study

[14] S. B. Wilson, C. A. Turner, R. G. Emerson and M. L. Scheuer, "Spike

detection II: automatic, perceptionbased detection and clustering", Clinical Neurophysiology, vol. 110, pp. 404–411, 1999.

[15] S. B. Wilson and R. Emerson, "Spike detection: a review and comparison of algorithms", IEEE Trans. Neural Networks, vol. 113, pp. 1873–1881, 2002.

[16] V. N. Vapnik, "An overview of statistical learning theory ", IEEE Trans. Neural Networks, vol. 10, pp. 988–999, 1999.

[17] D. Mandic, J. Chambers, "Recurrent Neural Networks for Prediction: Learning Algorithms Architectures and Stability", 2001, Wiley.

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

[19] J. A. Benediktsson , J. A. Palmason and J. R. Sveinsson,, "Classification of hyperspectral data from urban areas based on extended morphological profiles", IEEE Trans. Geosci. Remote Sens., vol. 43, no. 3, pp. 480-491, 2005.

[20] G. B. Huang , L. Chen and C. K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes", IEEE Trans. Neural Networks, vol. 17, no. 4, pp. 879–892, 2006.

[21] Z. Li, M. Hayashibe, C. Fattal and D. Guiraud, "Muscle fatigue tracking with evoked EMG via recurrent neural network: Toward personalized neuroprosthetics", IEEE Comput. Intell. Mag., vol. 9, no. 2, pp. 38-46, 2014.

[22] L. Wang and T. S. Buchanan, "Prediction of joint moments using a neural network model of muscle activations from EMG signals", IEEE Trans. Neural Syst. Rehabil. Eng., vol. 10, no. 1, pp. 30-37, 2002.

[23] M. Arvaneh, C. Guan, K. K. Ang and C. Quek, "Optimizing spatial filters by minimizing within-class dissimilarities in electroencephalogram-based brain-computer interface", IEEE Trans. Neural Netw. Learn. Syst., vol. 24, no. 4, pp. 610-619, 2013.

[24] D. Garrett, D. A. Peterson, C. W. Anderson and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification", IEEE Trans. Neural Syst. Rehabil. Eng., vol. 11, no. 2, pp. 141–144, 2003.

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

[26] D. Hubel and T. Wiesel , "Receptive fields and functional architecture of monkey striate cortex" , J. Physiology, vol. 195, pp.215–243, 1968.

[27] S. Ji, W. Xu, M. Yang, K. Yu,"3D convolutional neural networks for human action recognition", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.35, no.1, pp. 221–231.,Jan. 2013.

[28] P. Bashivan, I. Rish, M. Yeasin and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks", ICLR, 2016.

[29] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces" , IEEE International Conf. on Acoustics, Speech and Signal Processing , vol. 33, pp. 433–445, Mar. 2011.

[30] P. Mirowski, Y. LeCun, D. Madhavan and R. Kuzniecky, "Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG", IEEE Workshop on Machine Learning for Signal Processing (MLSP), pp. 244–249, Oct. 2008.

[31] A. Antoniades, L. Spyrou, C. C. Took and S. Sanei, "Deep learning for epileptic intracranial EEG data", IEEE Machine Learning for Signal Processing (MLSP), pp. 1–6, Sept. 2016.

[32] Y. Bengio, "Deep learners benefit more from out-of-distribution examples", Proc. Worshop. Artif. Intell. Stat., pp. 164-172, 2011.

[33] J. Fernandez Torre, G. Alarcon, C. Binnie and C. Polkey, "Comparison of sphenoidal, foramen ovale and anterior temporal placements for detecting interictal epileptiform discharges in presurgical assessment for temporal lobe epilepsy", Clin. Neurophysiol. vol. 110, pp. 895–904, 1999.

[34] S. Stodieck, H. G. Wieser and C. E. Elger, "The foramen ovale electrode: A new recording method for the preoperative evaluation of subjects suffering from mesio-basal temporal lobe epilepsy", Electroencephalogr. Clin. Neurophysiol. vol. 61, pp. 314–322, 1985.

[35] 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, 2016.

[36] S. Noachtar, C. Binnie, J. Ebersole, F. Mauguiere, A. Sakamoto and B. Westmoreland, "A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the EEG findings. The International Federation of Clinical Neurophysiology", Electroencephalogr. Clin. Neurophysiol. Suppl., vol. 52, pp. 21–41, 1999.

[37] L. J. P. van der Maaten and G. E. Hinton, "Visualizing data using t-SNE", Journal of Machine Learning Research, vol. 9, pp. 2431–2456, Nov 2008.

[38] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley and Y. Bengio., Theano: A CPU and GPU Math Expression Compiler, Proceedings of the Python for Scientific Computing Conference (SciPy) Jun 2010.

[39] Y. LeCun, L. Bottou, G.B. Orr, and K.-R. Mller. "Efficient backprop", Neural Networks: Tricks of the Trade, pages 950. Springer, 1998.

[40] D. J. Dittman, T. M. Khoshgoftaar, R. Wald, and A. Napolitano, "Comparison of data sampling approaches for imbalanced bioinformatics data", in Proceedings of FLAIRS, pp. 268–271, 2014.

[41] J. J. Halford, R. J. Schalkoff, J. Zhou, S. R. Benbadis, W. O. Tatum, R. P. Turner, S. R. Sinha, N. B. Fountain, A. Arain, P. B. Pritchard, E. Kutluay, G. Martz, J. C. Edwards, C. Waters, B. C. Dean, "Standardized database development for EEG epileptiform transient detection: EEGnet scoring system and machine learning analysis", J. Neurosci. Meth., vol. 212, no. 2, pp. 308-316, 2013.

[42] A. Keller, W. Truccolo, J. T.Gale, "Heterogeneous neuronal firing patterns during interictal epileptiform discharges in the human cortex", Brain, vol. 133, pp. 1668–1681, 2010.

[43] B. S. Naze, C. Bernard, V. Jirsa, "Computational modeling of seizure dynamics using coupled neuronal networks: factors shaping epileptiform activity", PLoS Comput Biol., vol. 11, no. 5, 2015.

[44] H. Jaseja, "EEG spike versus EEG sharp wave: Differential clinical significance in epilepsy", Epilepsy & Behavior, vol. 25, no. 1, p. 137, 2012.

[45] M. Curtis, J.G. Jefferys, M. Avoli , "Interictal epileptiform discharges in partial epilepsy: complex neurobiological mechanisms based on experimental and clinical evidence", in Jasper's Basic Mechanisms of the Epilepsies, National Center for Biotechnology Information 2012.

[46] F. Wendling, F. Bartolomei, F. Mina, C. Huneau and P. Benquet, "Interictal spikes, fast ripples and seizures in partial epilepsies - combining multi-level computational models with experimental data", in European journal of Neuroscience, vol. 6, pp. 2164–2177, 2012.

[47] M. E. Andraus, C. F. Andraus and S. V. Alves-Leon, "Periodic EEG patterns: importance of their recognition and clinical significance", Arq Neuropsiquiatr., vol. 70, no. 2, pp. 145–151, 2012.

[48] J. D. Frost, R. A. Hrachovy, D. G. Glaze, M. I. McCully, "Sleep modulation of interictal spike configuration in untreated children with partial seizures", Arq Neuropsiquiatr., Epilepsia, vol. 32, pp. 341–346, 1991.

[49] L. Bustamante, H. Lueders, V. Pippenger and E. S. Goldensohn, "The effects of phenytoin on the penicillin-induced spike focus", Electroencephalogr. Clin. Neurophysiol., vol. 48, no. 1, pp. 90–97, 1980.

[50] L. Bustamante, H. Lueders, V. Pippenger and E. S. Goldensohn, "Quantitative evaluation of anticonvulsant effects on penicillin-induced spike foci in cats", Neurology, vol. 31, no. 9, pp. 1163–1166, 1981.

[51] J. D. Frost, P. Kellaway, R. A. Hrachovy, D. G. Glaze and E. M. Mizrahi, "Changes in epileptic spike configuration associated with attainment of seizure control", Ann Neurology, vol. 20, no. 6. pp. 723–726, 1986.

[52] L. Guo, D. Rivero, J. Dorado, C. R. Munteanu, A. Pazos, "Automatic feature extraction using genetic programming: An application to epileptic EEG classification", Expert Syst. Appl., vol. 38, no. 8, pp. 10425–10436, 2011.

[53] N. Acir, I. Oztura, M. Kuntalp, B. Baklan, C. Guzelis, "Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks", IEEE Trans. Biomed. Eng., vol. 52, pp. 30–40, Jan. 2005.