Master's in Informatics Engineering
Dissertation
Final report

Seizure prediction based on Long Short Term Memory Networks

Rodrigo Pinto
rsp@student.dei.uc.pt

Advisors:
César Teixeira
António Dourado

September 5, 2017

FCTUC DEPARTAMENTO
DE ENGENHARIA INFORMÁTICA
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

This project was developed with the cooperation of

Center for Informatics and Systems of the University of Coimbra

Acknowledgements

This dissertation could not have been finished without the help and support from my advisors, Dr. César Teixeira and Dr. António Dourado. I would like to thank them for their excellent guidance.

I would also like to thank my parents, Claudia and Luís, for always being supportive and for providing me with everything I need.

A thanks to all my friends and to Inês.

Abstract

Epilepsy is a neurological disease affecting millions of people worldwide. About one third of them are pharmaco-resistant and cannot be submitted to surgery; their disease is called refractory epilepsy, and a seizure can happen any time, anywhere. These refractory patients would benefit from seizure prediction devices, but the current methods for seizure prediction are not good enough for clinical applications.

We present and evaluate the capacity of two types of deep artificial neural networks architectures to learn how to predict seizures with data extracted from electroencephalogram (**EEG**): Long Short Term Memory (**LSTM**) and Convolutional Long Short Term Memory (**C-LSTM**). To demonstrate clinical usefulness of our models, they are evaluated using long and continuous out of sample records. The study considers 105 patients from the European Epilepsy Database, 87 with scalp recordings and 18 with invasive recordings. The data includes: 1087 seizures, from which 203 are used for out-of-sample evaluation, and a total recording duration of 19959 hours, from which 3991 hours are used for out-of-sample evaluation. We extracted 22 univariate features based on 5 second windows from the EEG signal.

For all patients, scalp and invasive, our LSTM models achieved an average sensitivity of 29.28% and an average FPR of 0.58/h. We predicted 52 out of 203 (25.62%) seizures on the testing set. It is observed that for 5 out of 105 (4.8%) patients, optimal test performance with sensitivity $\geq 50\%$ and FPR $\leq 0.15h^{-1}$ was achieved. Perfect performance with 100% sensitivity and 0 h^{-1} FPR was achieved for 2 (1.9%) patients.

For all patients, scalp and invasive, our C-LSTM models achieved an average sensitivity of 28.17% and an average FPR of 0.64/h. We predicted 54 out of 203 (26.60%) seizures on the testing set. It is observed that for 2 out of 105(1.9%) patients, optimal test performance with sensitivity $\geq 50\%$ and FPR $\leq 0.15h^{-1}$ was achieved. Perfect performance with 100% sensitivity and 0 h^{-1} FPR was achieved for 0 (0.0%) patients.

The results were not satisfactory, we achieved worse results when making a comparison with a study using the same database but based on support vector machines (**SVM**s). Given that, in theory, the LSTMs are a method expected to perform better than the SVMs when sequential data is involved, we were expecting the opposite.

In the future we expect to experiment with raw EEG signal. We expect that the LSTMs will be able to capture better temporal dependencies on raw signal, since compacting 5 seconds of information into a single value leads to a great loss on information. A better approach for selecting training data, in a way that it covers a full day/night cycle to better capture intra day variations, will also be considered.

Contents

List of Figures

List of Tables

Abbreviations

ANN	**A**rtificial **N**eural **N**etwork
CNN	**C**onvolutional **N**eural **N**etwork
C-LSTM	**Convolutional L**ong **S**hort **T**erm **M**emory
ECG	Electrocardiogram
EEG	Electroencephalogram
FP	**F**alse **P**ositive
FPR	**F**alse **P**ositive **R**ate
LSTM	**L**ong **S**hort **T**erm **M**emory
MDADH	**M**aximum **D**ifference in **A**mplitude **D**istribution **H**istograms
MPC	**M**ean **P**hase **C**oherence
PCA	**P**rincipal **C**omponent **A**nalysis
RNN	**R**ecurrent **N**eural **N**etwork
SOP	**S**eizure **O**ccurrence **P**eriod
SVM	**S**upport **V**ector **M**achines

Chapter 1

Introduction

1.1 Motivation

Epilepsy is one of the most common neurological diseases, characterized by the constant occurrence of spontaneous seizures. Epilepsy affects around 50 million people, from which 30% of the patients suffer from drug refractory epilepsy, that is, their epilepsy is not completely controlled by drug administration.[1] Of those, only the ones that have a well localized focus in an area outside of the eloquent cortex are good candidates for surgical treatment. [2] An alternative method for seizure prevention, such as electric stimulation before the seizure onset, could be explored if there was some way to predict seizures. Alternatively, warning devices could be manufactured since the quality of life of a patient could substantially increase just by being alerted of the upcoming seizures.

Seizure prediction is traditionally based on time series data, therefore it is wise to choose a model for prediction able to learn temporal dependencies. Recurrent Neural Networks (**RNN**) are able of such feat. Long Short Term Memory networks [3] (**LSTM**), a special kind of RNNs, are better at capturing long-term dependencies than the standard RNNs, and they are the current state-of-the-art for many difficult problems involving sequential data. This includes handwriting recognition [4] and generation [5], language modeling [6] and translation [7], acoustic modelling of

speech, [8], speech synthesis [9], analysis of audio [10], and video data [11] among others.[12] LSTMs have also been successfully applied to seizure detection.[13] Being the current state of the art for many problems involving sequential data, we choose LSTMs to solve our problem. Zhou et al. (2015) [14] presented the concept of combining a one dimensional convolution network with a LSTM network, with the purpose of extracting higher-level representations of the sequential inputs. They obtained satisfactory results on a text classification problem involving sequential data. This motivated us to experiment with a Convolutional Long Short Term Memory architecture.

1.2 Goals

The goals of this project are:

- development of models, specialized in seizure prediction, based on two types of deep learning architectures: Long Short Term Memory and Convolutional Long Short Term Memory. The models will be trained with data from the EPILEPSIAE database.
- Presentation of results obtained on realistic conditions, i.e., on out of sample data and extensive continuous EEG recordings.
- Integration of a LSTM module on the EPILAB toolbox, which will enable its users to train LSTMs.

1.3 Document structure and organization

Chapter 2 exposes the reader to the background concepts related to epilepsy and the methods used to predict seizures.

In Chapter 3 an overview is made of the methods that have been applied to seizure prediction and the results they provide.

Chapter 4 contains information about the data used for this project. It also explains the experimental methods used, including which programming languages and toolboxes were considered.

In Chapter 5 we present and discuss the obtained results.

Chapter 6 contains the conclusion and and perspectives for future work.

Chapter 2

Background Concepts

2.1 Epilepsy

Epilepsy is a complex symptom caused by a variety of pathological processes in the brain. It is characterized by the occasional but recurrent occurrence of seizures. From the physiological point-of-view, a seizure it is characterized by an excessive and disorderly discharging of neurons.

Based on clinical and EEG information, seizures can be classified in two types:

- Focal seizures - Start in a particular region in the brain, affecting the part of the body controlled by that brain region.
- Generalized seizures - Occur in the whole brain, therefore affecting the whole body.

Epilepsy, depending on the type of seizures suffered by the patient, can also be classified in the categories:

- Focal epilepsy
- Generalized epilepsy

- Focal and generalized epilepsy

In our study we will use data from patients suffering from focal epilepsy, i.e., their epilepsy is characterized by the occurrence of focal seizures.

Focal epilepsy can also be categorized according to the region of the brain where the seizures occur:

- Focal lobe epilepsy
- Temporal lobe epilepsy
- Central lobe epilepsy
- Occipital lobe epilepsy
- Parietal lobe epilepsy

Typical treatments for epilepsy are drug administration and surgical treatment. About 30% of the patients with epilepsy do not benefit from drug administration for seizure contention. Those are said to suffer from refractory epilepsy. For those suffering from refractory epilepsy, surgical treatment is another option, but only 10% of those meet the requirement for such treatment,i.e., a well localized focus in an area outside of the eloquent cortex. [15]

2.2 Electroencephalogram (EEG)

EEG is a monitoring method that measures voltage fluctuations resulting from ionic current within the neurons of the brain. This information can be captured by electrodes that can either be placed along the scalp (non-invasive) or in direct contact with the brain (invasive). To ensure that the naming and location of electrodes is consistent across laboratories, naming and location are typically specified by the International 10-20 system (figure 2.1).

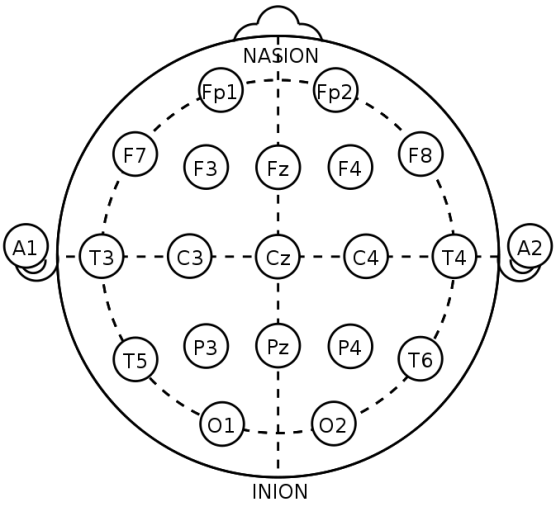

FIGURE 2.1: Electrode placement according to the 10-20 system.

Signal sampling is typically done at 256-2000 Hz and the EEG signal, for humans, has an amplitude ranging from 10 μV to 100 μV for scalp recordings.[16]

EEG can be used to diagnose and characterize neurological diseases, including epilepsy, which causes abnormalities in EEG readings (figure 2.2).The EEG is a helpful diagnostic tool in the investigation of a seizure disorder. It confirms the presence of abnormal electrical activity, gives information regarding the type of seizure disorder, and discloses the location of the seizure focus.

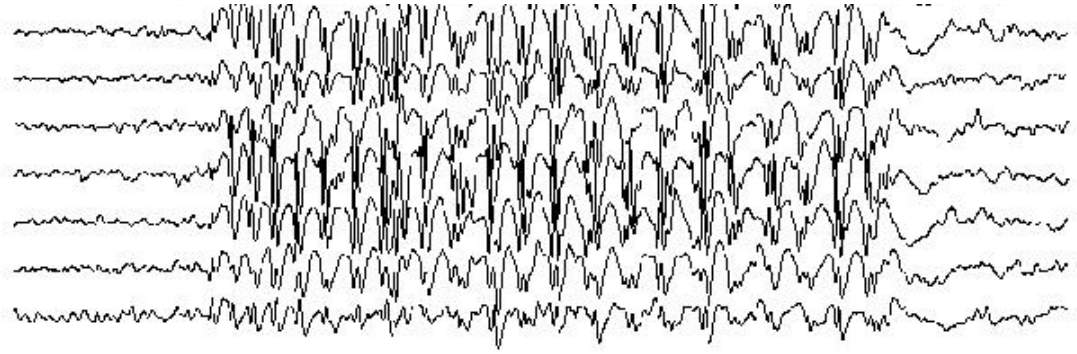

FIGURE 2.2: Abnormalities on EEG recording during the occurrence of a seizure. [17]

2.3 Seizure Prediction

Seizure prediction is an active area of research because researchers hypothesize that specific EEG patterns occur during the pre-ictal period. [18]

If such hypothesis is true, then those patterns might be detected in EEG recordings, signalling a future seizure. Seizure prediction algorithms make use of features extracted from EEG data. Prediction can be made by generating alarms when a given feature crosses a certain threshold, or by using multiple features as the input for a discriminative classifier.

In the literature, different names have been attributed to the different brain periods of an epileptic brain. The ictal period is the period during which the seizure actually occurs. The post-ictal period is the time immediately after a seizure, and the preictal the one immediately before. The interictal period is the time between seizures, during which the brain is well-functioning.

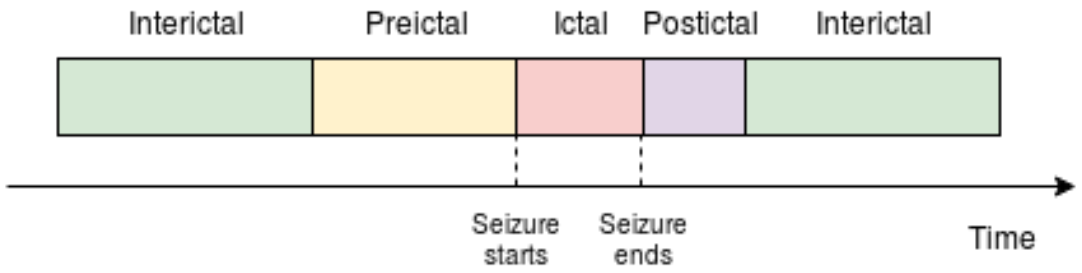

FIGURE 2.3: Different brain periods of an epileptic patient.

If the pre-ictal state does exist, then its duration might vary from patient to patient and even from seizure to seizure. Its duration is usually chosen in order to maximize the predictors performance. This period is often called the seizure occurrence period (SOP).

2.4 Artificial Neural Networks

Artificial Neural Networks (**ANNs**) are a learning algorithm which had in the anatomy of the brain the inspiration for its origin. Its basic computational unit is

the artificial neuron. Figure 2.4 depicts the similarities between the real(left) and the artificial(right) neurons. The biological neuron works as follows, information from axon terminals of several other neurons is received into its dendrites, which is in turn processed and outputted through its axon terminals into other neurons. A synapse is a structure that permits a signal to be passed between neurons. In the artificial model, the different signals received from other neurons are called the inputs (e.g. $\boldsymbol{x}$) and the synapses are the weights (e.g $\boldsymbol{w}$). All the information received by one neuron can then be combined into one value using equation 2.1.

$$n = \boldsymbol{w}^T \boldsymbol{x} + b \tag{2.1}$$

where b is a constant, named bias, intrinsic to each neuron. This value n is then the argument of a function f:

$$a = f(b) \tag{2.2}$$

producing a, the neuron output. This function f is called the activation function, is intrinsic to each neuron, and it can take many forms. Some commonly used activation functions are the sigmoid and hyperbolic tangent function, given by equations 2.3 and 2.4, respectively.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.3}$$

$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.4}$$

2.4.1 Feedforward ANN Architectures

Neurons can be stacked to form one layer of neurons, and ANN architectures can have multiple layers. The most basic ANN architecture is the feedforward ANN,

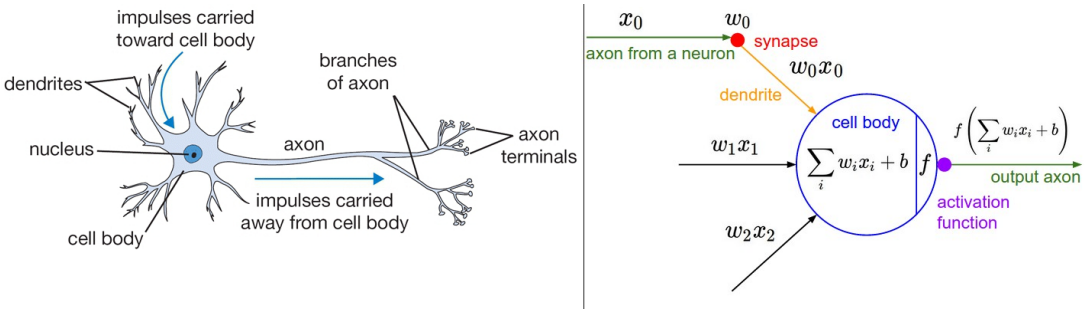

FIGURE 2.4: Biological neuron(left), Artificial neuron(right). [19]

depicted in figure 2.5. The input layer (blue) is an array defined as $\boldsymbol{p} \in R^d$, where d is the number of inputs (features) of the network. The middle layers (green) are called the hidden layers while the last one (red) is the output layer, it produces the final output of the network. Formally, each layer j can be characterized by the following items:

- $\boldsymbol{W}^{j,j-1} \in R^{s \times r}$, where s is the number of neurons of that layer and r is the number of neurons of the previous layer, or, if $j = 1$, the number of inputs of the network. $\boldsymbol{W}$ contains in each row, the weights associated with each neuron.
- $\boldsymbol{b}^j \in R^s$, where s is the number of neurons of that layer. $\boldsymbol{b}$ contains in each row, the bias associated with each neuron.
- f^j, the activation function.

The output of layer j is then given by the equations:

$$\boldsymbol{a}^j = f^j(\boldsymbol{W}^{j,j}\boldsymbol{p} + \boldsymbol{b}^j) = f^j(\boldsymbol{n}^j), \text{ if } j = 1 \tag{2.5}$$

$$\boldsymbol{a}^j = f^j(\boldsymbol{W}^{j,j-1}\boldsymbol{a}^{j-1} + \boldsymbol{b}^j) = f^j(\boldsymbol{n}^j), \text{ if } j > 1 \tag{2.6}$$

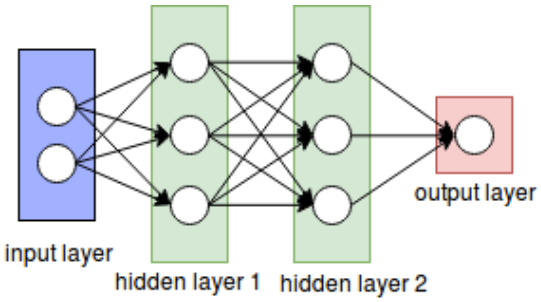

FIGURE 2.5: Feedforward Artificial Neural Network Architecture.

2.4.2 ANN for Classification

When performing classification, the number of neurons of the output layers is typically equal to the number of classes of the problem, so we can associate each class to one output neuron. By analysing the outputs of the network for a single case(pattern), we can make a classification. What the values of the outputs represent depend on the activation function of the output layer. If the activation function used is, for example, the Softmax, then the value of an output neuron represents the probability of a given pattern belonging to the class associated with it, which is why the Softmax function is typically used in the context of classification.

Mathematically, the Softmax function is given by equation 2.7m where $\sigma(\boldsymbol{n})_j$ is the probability of a given pattern belonging to the class associated with output neuron j and K is the number of output neurons (classes).

$$\sigma(\boldsymbol{n})_j = \frac{e^{n_j}}{\sum_{k=1}^{K} e^{n_k}} \tag{2.7}$$

2.4.3 Learning

The word learning, in the context of ANNs, means finding the set of weights and biases that deliver the optimal performance. Performance is measured by a loss function, and, the optimal performance is achieved at its global minimum. An

example of a loss function is the mean squared error (2.8), where N is the number of cases, $\hat{Y}$ is the predicted value and Y is the true value.

$$L = MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{Y}_i - Y_i)^2 \tag{2.8}$$

Any combination of weights will be associated with a particular loss measure, then, the task of training an ANN is simply finding the set of weights that minimizes the loss. This task can be accomplished with the backpropagation algorithm [20], one of the most used algorithms for training. The goal of the backpropagation is to compute, for every training pattern i, the partial derivatives, $\frac{\partial L_i}{\partial w}$ and $\frac{\partial L_i}{\partial b}$, of the loss function L_i with respect to any weight w or bias b in the network, which can be averaged over all the training patterns and used to update of the weights and bias. The backpropagation algorithm can be performed any number of times (epochs).

The updates are performed, in the simplest case, by means of a gradient descent approach:

$$w_{t+1} = w_t - \eta\frac{\partial \bar{L}}{\partial w} \tag{2.9}$$

$$b_{t+1} = b_t - \eta\frac{\partial \bar{L}}{\partial b} \tag{2.10}$$

where $\bar{L}$ is the average loss over all patterns, t is the time step and η is the learning rate, a parameter that has to be tuned for each problem in order to achieve loss convergence. More complex functions can be used to update the parameters.

2.4.4 Recurrent Neural Networks

Recurrent Neural Networks(RNNs) are a class of ANNs containing loops in their structure, as shown in fig 2.7. They are typically good at solving problems when sequential data is present, like a time series for example. They are good at those

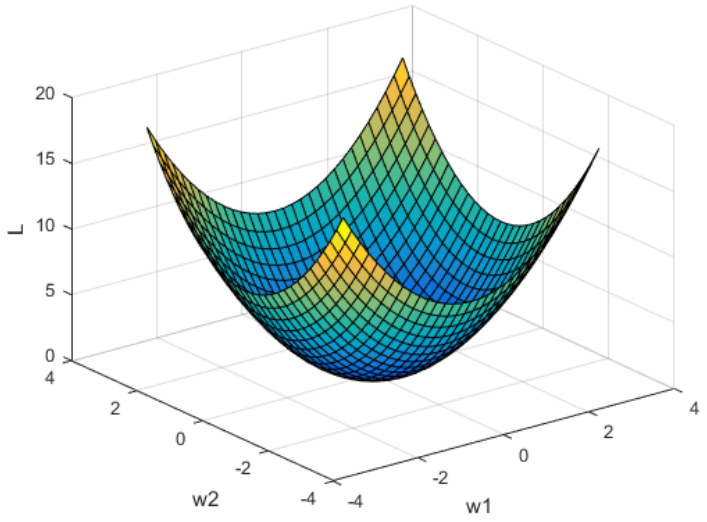

FIGURE 2.6: Loss function represented in the weight space. In this case the loss function is convex with a global minimum, but generally the function contains many local minimums.

kind of problems because since they have inner loops they can pass information from pattern to pattern, making the network able to learn temporal dependencies. The vanilla RNN is depicted in fig 2.7. This network has the so-called state h state $\boldsymbol{h}$. As seen in equation 2.11, this state, at each timestep t, is a function of the current inputs, $\boldsymbol{x}_t$ and the state at the previous time step, $\boldsymbol{h}_{t-1}$, which have associated weight matrices $\boldsymbol{W}$ and $\boldsymbol{U}$, respectively. The network state can be used to compute the output at any given time step (2.12), $\boldsymbol{V}$ is the weight matrix associated with the state $\boldsymbol{h}$ when computing the output.

When a RNN is unrolled in time it becomes similar to a standard ANN with many layers, except that for the RNNs the weights are shared across layers.

$$\boldsymbol{h}_t = f(\boldsymbol{W}\boldsymbol{h}_{t-1} + \boldsymbol{U}\boldsymbol{x}_t) \tag{2.11}$$

$$\boldsymbol{a}_t = f(\boldsymbol{V}\boldsymbol{h}_t) \tag{2.12}$$

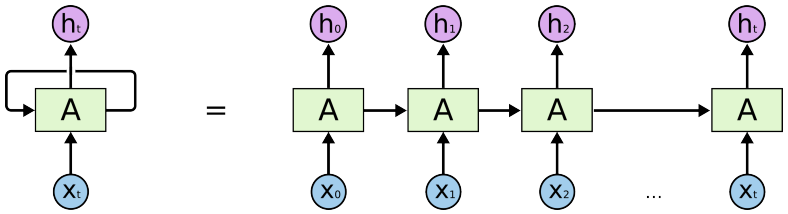

FIGURE 2.7: Unrolled Recurrent Neural Network.[21]

2.4.5 Vanishing Gradient Problem

To update the weights, for each timestep the error has to be backpropagated across layers until we reach the layer corresponding to the beginning of time. The act of backprogatating from layer to layer implies the multiplication of the derivatives of the activation functions of each layer. If we look at one common activation function, the sigmoid function, given by $\sigma(x) = \frac{1}{1+e^{-x}}$, its derivative has the shape shown in figure 2.8, it is observed that its maximum value is 0.25. Since the contribution to the weights update of one layer located n time steps behind is proportional to $(\frac{d\sigma}{dx})^n$, assuming that the sigmoid function is used in every layer, and since $max(\frac{d\sigma}{dx}) = 0.25$, we can conclude that the contribution from previous layers approaches zero the more we go deeper. This is called the Vanishing Gradient Problem. [22]

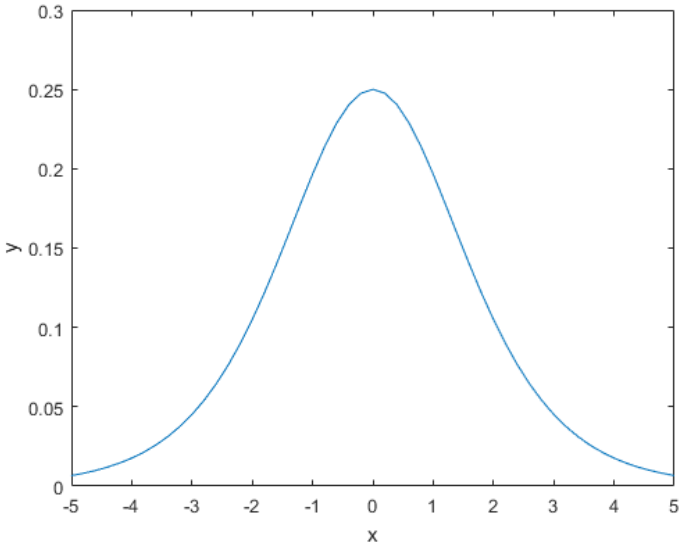

FIGURE 2.8: Derivative of sigmoid function.

2.4.6 Long Short Term Memory (LSTM)

To overcome the Vanishing Gradient Problem, Sepp Hochreiter and Jürgen Schmidhuber [3] proposed an innovative RNN architecture. These networks are capable of learning long-term dependencies.

Traditional RNNs have the aspect shown in figure 2.9. Typically the state $\boldsymbol{h}_t$ is a function of $\boldsymbol{h}_{t-1}$ and $\boldsymbol{x}_t$, the previous state and the current inputs. The architecture of the LSTMs is seen in figure 2.10. The main thing to notice is that instead of existing just one type of state $\boldsymbol{h}_t$ that passes information from one timestep to another, now there is another type of state called the cell state $\boldsymbol{C}_t$. The cell state acts like a conveyor belt that allows information to just flow along without changes, if needed. Information can be removed or added to the cell state by means of structures called gates. Mathematically, the LSTMs are represented by equations 2.13-2.18. The operator $\circ$ represents point-wise multiplication.

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_f.[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_f) \tag{2.13}$$

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_i.[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i) \tag{2.14}$$

$$\widetilde{\boldsymbol{C}}_t = tanh(\boldsymbol{W}_C.[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_C) \tag{2.15}$$

$$\boldsymbol{C}_t = \boldsymbol{f}_t \circ \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \circ \widetilde{\boldsymbol{C}}_t \tag{2.16}$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_o.[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_o) \tag{2.17}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \circ tanh(\boldsymbol{C}_t) \tag{2.18}$$

Taking a closer look at 2.16, the equation for the computation of the cell state, we observe two terms. The first one is responsible for removing information from the cell state, while the second one is responsible for adding information.

The equation 2.18 is the one responsible for the computation of the network state $\boldsymbol{h}_t$.

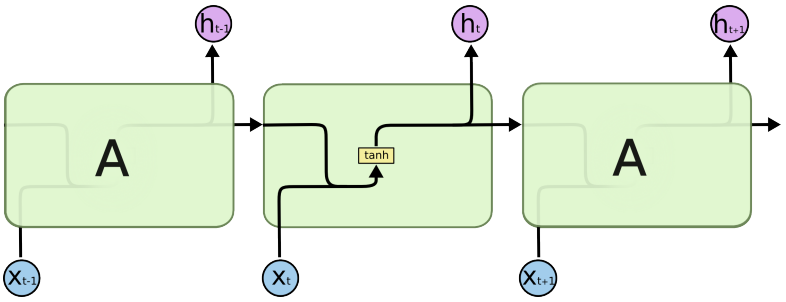

FIGURE 2.9: Simple RNN architecture.[21]

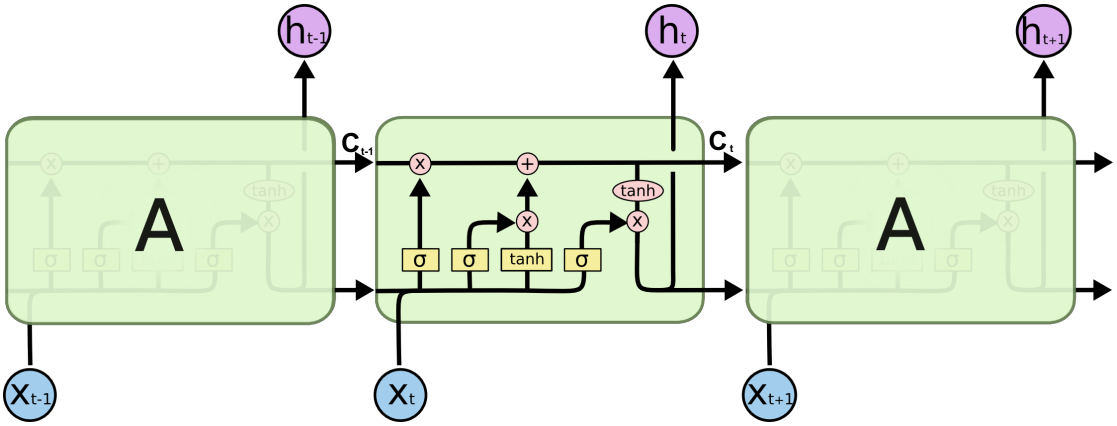

FIGURE 2.10: LSTM architecture.[21]

2.4.7 Many to one Classification

It has been stated that RNNs have memory, but it is not feasible to to retain information since the beginning of time. On the training phase it would imply that as time increases there would be an increment in the number of layers through which the algorithm has to backpropagate. This would translate in a infeasible training time. Moreover, retaining information since information since the beginning of time might not bring any improvement to our model. In our project we are dealing with long-term EEG recordings, the information from one day might not be useful in the next. These are the reasons why we have to create finite sequences from our data. We define a sequence length and at any timestep t we only backpropagate through the sequence. Figure 2.11 exemplifies how one output is computed from a sequence. Creating sequential inputs implies the addition of another dimension to the feature matrix. Instead of having a feature matrix with dimensions $n \times d$ we will have a matrix with dimensions $n \times l \times d$, where n is the number of patterns, l is the sequence length and d is the number of features.

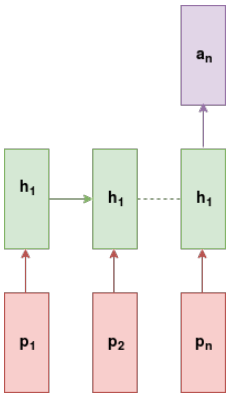

FIGURE 2.11: Many to one classification: The red rectangle represent sequential inputs and the purple rectangle represents the output. The green rectangles represent the state of the network, during each time step the state of the network depends on the previous state as well as on the current input. The output depends only on the state at the last time step.

2.5 One dimensional convolution

One dimensional convolution can be used to extract higher-level features from a sequence. It involves a filter vector sliding over a sequence and detecting features at different positions. Let $\boldsymbol{x} \in R^{l \times d}$ denote the input sequence where l is the sequence length and d the number of features; $\boldsymbol{x}_i \in R^d$ denote the i^{th} input of the sequence; $\boldsymbol{m} \in R^{k \times d}$ denote the filter of length k for the convolution operation. For positions $1, \ldots, l-k+1$ in the sequence, we have a window vector $\boldsymbol{w}_j \in R^{k \times d}$ denoted as

$$\boldsymbol{w}_j = [\boldsymbol{x}_j, \boldsymbol{x}_{j+1}, ..., \boldsymbol{x}_{j+k-1}] \tag{2.19}$$

where commas represent row vector concatenation. A filter $\boldsymbol{m}$ convolves with the window vectors in a way to generate a feature map $\boldsymbol{c} \in R^{l-k+1}$; each element c_j of the feature map for window vector $\boldsymbol{w}_j$ is produced as follows

$$c_j = f(\boldsymbol{w}_j * \boldsymbol{m} + b) \tag{2.20}$$

where $b \in R$ is a bias term, * is matrix convolution and f is a non-linear transformation function.

For m filters with the same length, the generated m feature maps can be rearranged as feature representations for each window w_j,

$$\boldsymbol{W} = [\boldsymbol{c}_1; \boldsymbol{c}_2; ...; \boldsymbol{c}_m] \tag{2.21}$$

Here, semicolons represent column vector concatenation and $\boldsymbol{c}_i$ is the feature map generated with the i-th filter. The resulting $\boldsymbol{W} \in R^{(l-k+1)\times m}$ is a higher-level sequence of features that can be fed into the LSTM.

Chapter 3

State of the Art

3.1 Introduction

Up until 1975, neuroscientists believed that the transition between normal brain function and an epileptic seizure was instantaneous, that is, they thought that an epileptic seizure appeared with no previous traces of its incoming. Viglione and Walsh (1975) [23] had visionary ideas and believed that maybe there was a transition state, between the normal and epileptic states, that could be detected. A detection of that state would imply the ability to predict seizures. Using linear approaches, such as pattern detection and spectral analysis, they performed experiments with five patients, yielding 90% of correct separation between pre-seizure and non-pre-seizure EEG epochs in the training set. They even patented an electronic warning device, but it produced many false-positive results. Eventually, they stopped research because they were very limited by the computational resources available at the time.

With the advance of technology, many experiments have been made in the field, and even thought each experiment may use different methods, most of the published studies were based on the use of the EEG. Features can be linear, non-linear, univariate (computed using a single channel) or multivariate (computed using multiple channels). Furthermore, features can be categorized in the time, frequency,

or time-frequency domains. It should also be noted that seizure prediction can be made through feature thresholding or machine learning methods.

3.2 Databases

Most of the studies we are going to review use data from the Freiburg or EPILEPSIAE databases. We will briefly outline their contents.

3.2.1 Freiburg database

The Freiburg database contains invasive EEG recordings from 21 patients suffering from medically intractable focal epilepsy. The EEG data was acquired using a Neurofile NT digital video EEG system with 128 channels, 256 Hz sampling rate, and a 16 bit analogue-to-digital converter.

Although this database provided data of a higher quality than ever before it has flaws: it contains discontinued data and only contains 50 minutes of preictal data per patient, which might not be enough to represent the intra-patient pre-ictal state variations.

3.2.2 EPILEPSIAE database

The EPILEPSIAE database was the first database to provide continuous long-term records, turning the Freiburg database obsolete. It contains recordings of 275 patients collected at the University Hospital Freiburg, Germany, of the University Hospital of Coimbra,Portugal, and of the Hopital de la Pitié-Salpêtrière in Paris, France. To be more precise, it contains surface recordings from 225 patients and invasive EEG recordings from 50 patients. The scalp EEG data was recorded with scalp electrodes placed according to the international 10-20 system with a referential montage, and the invasive EEG was recorded with stereo-tactically implanted depth electrodes, subdural grids and/or strips. The recording of the

electrocardiogram (ECG) is also present for all the patients. This database contains almost 2031 days of EEG recordings including 2702 seizures. [24] [25]

3.3 Past studies on Seizure Prediction

Sackellares et al., (2006) [26] used convergence in short-term maximum Lyapunov exponent(STLmax), a measure of the local chaoticity in a dynamical system, to track preictal changes. For a prediction horizon of 30min, the average sensitivity reached 80% and the false prediction rate was 0.56/h, however, the value dropped to 0.12/h when the prediction horizon was set to 150min.

Schelter et al. (2006) [27] applied time series analysis techniques for seizure prediction. Using invasive EEG data of four representative patients suffering from epilepsy, they demonstrated the performance of a seizure prediction method based on a quantity measuring phase synchronization. They used a random predictor to decide if their method provides prediction better than chance, and only for two of the four patients the performance of their method was superior to a random predictor.

Mirowski et al. (2008) [28] wrote one of the first articles where machine learning techniques were applied to epileptic seizure prediction. The methods used were logistic regression, SVMs and Convolutional Neural Networks. They made use of four different types of bivariate features computed based on 5s windows. They claim that, for each patient on the Freiburg EEG dataset, at least one method predicts 100% of the seizures on average 60 minutes before the onset, with no false alarm. All results were statistically validated.

Chisci et al. (2010) [29] applied Auto-Regressive models with SVMs for seizure prediction. SVMs had already been applied in the past but they proposed the association of the standard SVM with a Kalman Filter to regularize the continuous variable used for classification, and thus, to significantly reduce the number of false

alarms. Their methods exhibited 100% sensitivity and high values of specificity on the Freiburg EEG dataset.

Kuhlmann et al. (2010) [30] investigated a single feature, the bivariate synchrony. They analysed all available channels from long-term invasive recordings from 6 patients with intractable focal epilepsy. Four threshold techniques using fixed and dynamic threshold values were used to generate alarms. For each patient, the best sensitivity ranged from 50% to 88%, while the false prediction rates ranged from 0.64/h to 4.69/h.

Park et al. (2011) [31] proposed an algorithm for seizure prediction using multiple EEG power spectral features and cost-sensitive SVMs, which they trained and tested using data from 18 patients from the Freiburg database. A Fisher discriminant kernel analysis was used to determine the best feature set for testing. The output of the classifier was smoothed using a Kalman filter to decrease the number of false positives. They achieved an average sensitivity of 98.3% and an average false positive rate of 0.27/h. Prediction above chance was attained for all the 18 patients. This study demonstrated that low computational cost linear features can provide good results. It was also the first study to demonstrate above-chance performance in out-of-sample data.

Williamson et al. (2012) [32] presented a method based on a patient-specific SVM classifier. They used a total of 488 features based on the eigenspectra of space-delay correlation and covariance matrices computed at multiple delays. The feature space was reduced using a principal component analysis(**PCA**), where the top 20 components were used as inputs to the SVMs. The algorithm achieved an average sensitivity between 86% and 95% with the proportion of time spent under false warning between 9% and 3%.

Aarabi and He (2012) [33] presented a patient-specific algorithm based on the combination of non-linear univariate and bivariate features. The authors used invasive EEG datasets from 11 patients selected from the Freiburg database. The univariate features used were: correlation dimension, correlation entropy, noise level, Lempel-Ziv complexity and largest Lyapunov exponent, while a single bivariate

measure was used, the nonlinear interdependence. Alarms were generated using 4 threshold values and a set of combination and integration rules. For two seizure prediction horizons, 30 and 50min, an average sensitivity of 79.9% and 90.2%, an average false prediction rate of 0.17 and 0.11/h were achieved, respectively. The results were statistically validated. By analysing the performance of each feature individually, the authors concluded that the combination of univariate and bivariate features contribute to a better algorithm performance.

Li et al. (2013) [34] investigated the rate of epileptic spikes in invasive EEG as a measure for seizure prediction on the Freiburg dataset. They found that the spike rate was significantly different for the interictal, preictal, ictal and postictal segments. Alarms were generated when the spike rate of any channel crossed a patient-specific threshold. The algorithm achieved 56% sensitivity and a false prediction rate of 0.15/h for a prediction horizon of 30 min and 72.7% sensitivity and a false prediction rate of 0.11/h for a prediction horizon of 50 min. All these results were demonstrated to be above chance level.

Gadhoumi et al. (2013) [35] presented an algorithm based on measures of similairty between preictal and interictal states derived from wavelet entropy and energy. The algorithm was tested on long-term intracerebral EEG from 17 patients from the Montreal Neurological Hospital. For 7 out of the 17 patients, the algorithm predicted seizures above chance. The average sensitivity was higher than 85% and the average false prediction rate was below 0.1/h.

Teixeira et al. (2013) [36] reported the first largest study on long-term EEG recordings. The data used was the outcome of the European project EPILEPSIAE (http://www.epilepsiae.eu) [24] [25], which was created with the purpose of providing the research and clinical communities with data that eliminates the problems of short-term discontinuous data segments and the low number of patients. In this study they reported the performance of three different models on the new dataset: multilayer perceptron neural networks(**MLP-ANN**), radial basis neural networks(**RBF-ANN**) and SVM. Only six electrodes were used for the computation of twenty-two univariate features, using 5s windows without overlap.

One innovation presented was the concept of firing power, a filter used to reduce the number of false alarms. For each patient, the best predictor was chosen by its proximity to the optimal predictor (100% of sensitivity, $0h^{-1}$ of FPR). They achieved an average sensitivities of 74% and 68%, and average false positive rates of 0.28 and 0.37 for scalp and invasive patients, respectively.

Eftekhar et al. (2014) [37] presented a method for detecting and predicting seizures using the Freiburg database. The method was based on N-gram modelling. The N-gram algorithm looks for repeating patterns using a 1-min sliding window. The sequence of the repeating patterns counts was used to generate alarms through threshold crossing. Three channels corresponding to the seizure onset were analysed. The authors performed what appears to be a cross-validation on the dataset and reported the average performance across iterations and channels.The authors reported a sensitivity of 67% and a false prediction rate of 0.04/h for temporal lobe cases, and a sensitivity of 72% and a false prediction rate of 0.61/h for frontal lobe cases. All results were statistically significant, by comparison with a random predictor.

Zheng et al. (2014) [38] used a method based on the mean phase coherence (**MPC**), originally proposed by Mormann et al.(2000). The original MPC predictive power did not exceed random levels. A bivariate empirical mode decomposition of EEG channels was then used to improve performance. Data from 10 patients of the Freiburg database was used to optimize and test the method. The performance was assessed by using a cross-validation on the dataset. For a maximum false prediction rate of 0.15/h, the average sensitivity ranged between 25 and 70%. Performance was higher than the performance presented by a random predictor.

Bandarabadi et al. (2015) [39] presented an algorithm based on spectral power rations. They used data from 24 patients of the EPILEPSIAE database. Channels from 3 electrodes in the seizure onset area and other 3 not related with the focus were analysed. For each channel, the normalized spectral power in the conventional EEG frequency bands(Alpha, Beta, Delta, Theta and Gamma) was computed based on 5s windows. The authors introduced the ratio between normalized spectral

powers to track preictal changes. The ratio was computed for all combinations of channels and frequency bands, giving a total of 435 features. A new feature selection method based on the maximum difference in amplitude distribution histograms (**MDADH**) of preictal and non-preictal samples were considered(Teixeira et al., 2011). The prediction was made using a SVM classifier, using the first 3 seizures for training and the remaining ones for testing. Four seizure prediction horizon values were tested: 10, 20, 30 and 40 minutes. A method based on the "firing power" filter (Teixeira et al., 2011) was applied to the classifier's output, aiming to generate alarms before seizures and at the same time minimizing the number of false alarms. For the 8 patients with invasive EEG, an average sensitivity of 78.36% and a false positive rate of 0.15/h were achieved. The results for the scalp data were similar, however, the authors reported that the number of features selected was smaller for invasive than for scalp data, concluding that invasive EEG contains clearer and more localized epileptogenic information than scalp EEG.

Direito et al. (2016) [40] presented a study based on SVMs, using 216 patients from the EPILEPSIAE database. Six electrodes were used for the computation of twenty-two univariate features, using 5s windows without overlap. The "firing power" filter was used to reduce the number of false alarms. The validation procedure included a k-fold cross validation procedure that lead to the attainment of a best model that was used in out-of-sample data. They achieved an average sensitivity of 38.47% and an average false positive rate of 0.2/h. The results were statistically for 11% of the patients.

3.4 Deep learning approaches for seizure detection

Another topic of importance is how the features are extracted from the EEG data. One study has shown that features extracted from EEG signal using convolutional networks might outperform human engineered features for seizure detection.

Thodoroff et al. (2016) [13] proposed a recurrent convolutional architecture designed to capture spectral, temporal and spatial patterns representing a seizure. First, they project the multi-channel EEG signal into an image representation, second, a recurrent convolutional neural network is trained to predict whether or not the corresponding image contains a seizure. Their image-based representation of the EEGs is a method that makes use of the knowledge of the electrodes spatial montage. The recurrent convolutional neural network used can be separated into two types of neural networks: convolutional and recurrent. The architecture of the convolutional neural network (CNN) used was inspired by a model that achieved state of the art on the ImageNet competition [41]. The CNN has the ability to automatically extract features, thus, replacing the need for complex feature engineering that was used in previous works on seizure detection and prediction. The recurrent neural network (RNN), which feeds on the automatically extracted features by the CNN, is one of the type Long Short Term Memory (LSTM), a type of RNNs which have the ability to retain information for longer periods of time than normal RNNs [3]. They claim that their model has obtained state of the art performance on patient specific detectors, and, unlike earlier models, it has the ability to learn a general representation of a seizure, which leads to significant improvement in cross-patient detection performance.

3.5 Discussion

Almost all studies presented in this review used the Freiburg database, however, as mentioned previously, it contains discontinued data and only 50 minutes of preictal data per patient. Lehnertz et al. (2007) [42] stated that to demonstrate the clinical usefulness of a seizure prediction method, it should be evaluated on long term, continuous and uninterrupted EEG data covering all possible states and conditions of the patient. Having this in mind, we will use data from the EPILEPSIAE database, which overcomes all the limitations of its predecessor.

Since one of the goals of our study is to compare the performance of LSTMs with other methods, we will use the set of features used by Direito et al. (2016) in their work, which was based on SVMs. We choose to compare our work with the one by Direito et al. (2016) because we consider it to be the most realistic up to date, in a way that it provides out-of-sample results for 216 patients of the EPILEPSIAE database.

Thodoroff et al. (2016) have successfully applied feature extraction algorithms to improve the results of seizure detection, but the same has not yet been applied to seizure prediction. This motivated us to also experiment one dimensional convolution, a type of network able to extract features from data.

In table 3.1 are summarized the results of the reviewed studies.

TABLE 3.1: Summary of results of the reviewed seizure prediction studies.

Year	Author	Database	Patients	Seizures	EEG(h)	Sensitivity(%)	fpr/h
2016	Direito et al.	EPILEPSIAE	216	1206	-	38.47	0.20
2014	Bandarabadi et al.	EPILEPSIAE	24	183	3565	78.36	0.15
2014	Eftekhar et al.	Freiburg	21	87	623	67 (temporal lobe) 72(frontal lobe)	0.04 (temporal lobe) 0.61 (frontal lobe)
2014	Aarabi et al.	Freiburg	21	87	596	79.9-90.2	0.17-0.11
2014	Zheng et al.	Freiburg	10	50	221.1	25-70	-0.15
2013	Teixeira et al.	EPILEPSIAE	278	2702	48744	74 (scalp) 68 (intra)	0.28 (scalp) 0.37 (intra)
2013	Li et al.	Freiburg	21	-	-	75.8	0.09
2013	Gadhoumi et al.	Montreal Neurological Institute	17	175	1565	>85	<0.1/h
2012	Williamson et al.	Freiburg	19	83	448.3	86-95	-
2012	Aarabi and He	Freiburg	11	49	267	79.9-90.2	0.17-0.11
2011	Park et al.	Freiburg	18	80	433.2	98.3	0.27
2010	Kuhlmann et al.	Freiburg/Melbourne	6	73	597.6	50-88	0.64-4.69
2010	Chisci et al.	Freiburg	-	-	-	100	-
2009	Mirowski et al.	Freiburg	21	-	-	100	0.0
2006	Sackellares et al.	Gainesville	10	130	2100	80	0.12

Chapter 4

Data and Methods

4.1 Data

From the EPILEPSIAE database, we selected data from 87 patients with scalp EEG recordings, and from 18 patients with invasive recordings. The data includes: 1087 seizures, from which 203 are used for test, and a total recording duration of 19959 hours, from which 3994 are used for test.

Detailed patient characteristics are presented in Appendix B.

4.1.1 Scalp data

The scalp data includes: 874 seizures, from which 161 are used for test, and a total recording duration of 14867 hours, from which 3089 are used for testing.

The distribution of gender, age, seizure localization and lateralization, and number of testing seizures, for the scalp patients, is available in figures 4.1, 4.2, 4.3, 4.4 and 4.6, respectively.

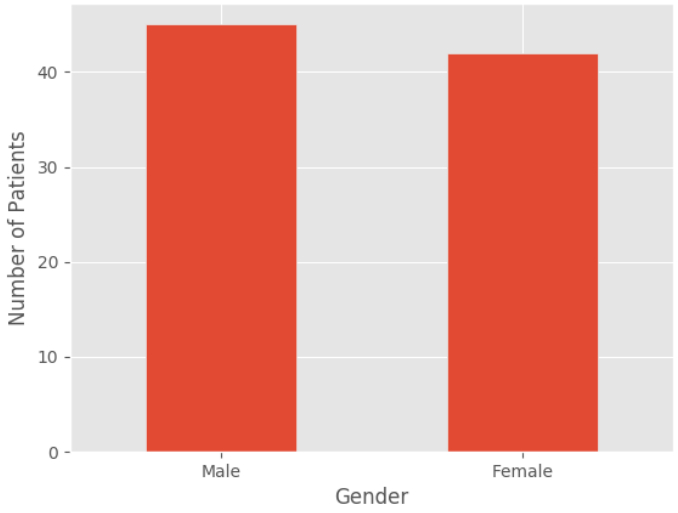

FIGURE 4.1: Histogram of genders for scalp patients.

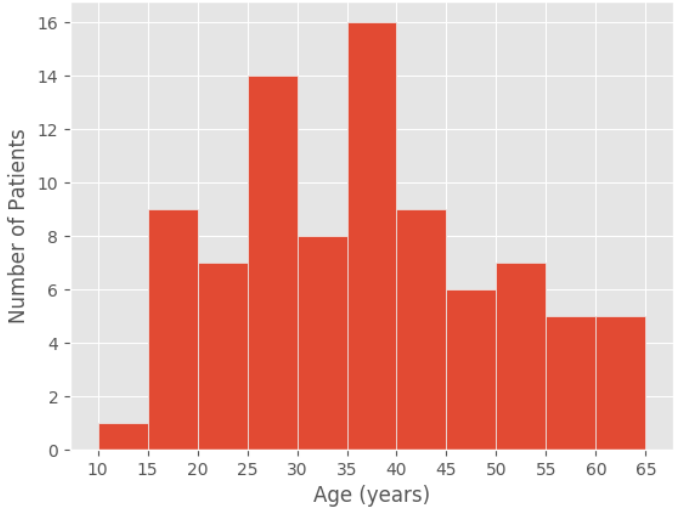

FIGURE 4.2: Histogram of ages for scalp patients.

4.1.2 Invasive data

The invasive data includes: 213 seizures, from which 42 are used for test, and a total recording duration of 5092 hours, from which approximately 905 are used for testing.

The distribution of gender, age, seizure localization and lateralization, and number of testing seizures, for the invasive patients, is available in figures 4.7, 4.8, 4.9, 4.10

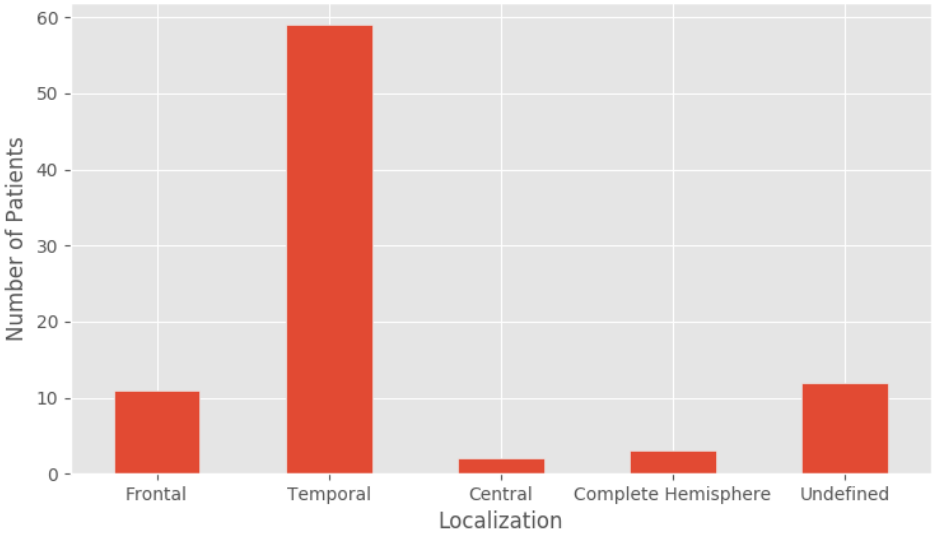

FIGURE 4.3: Histogram of seizure localization for scalp patients.

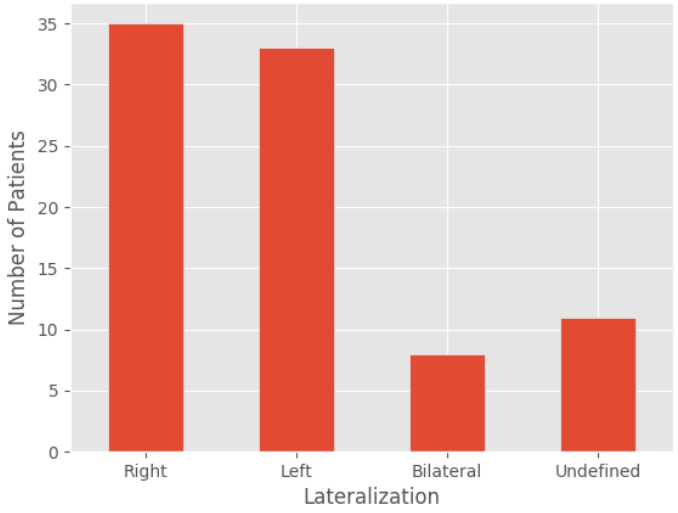

FIGURE 4.4: Histogram of seizure lateralization for scalp patients.

and 4.12, respectively.

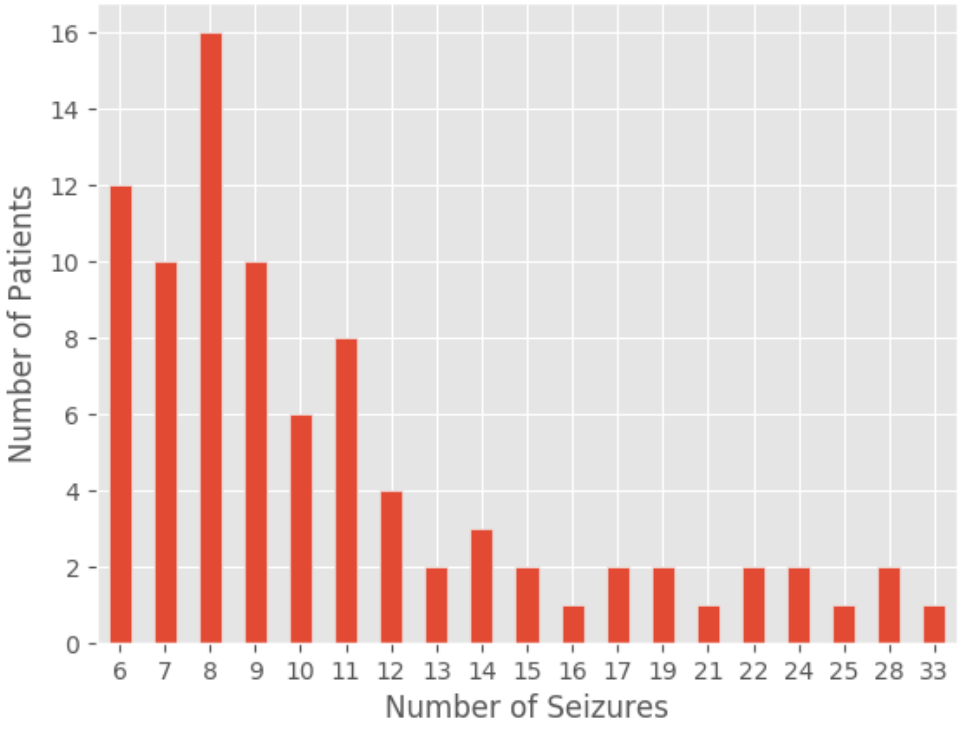

FIGURE 4.5: Histogram of number of seizures for scalp patients.

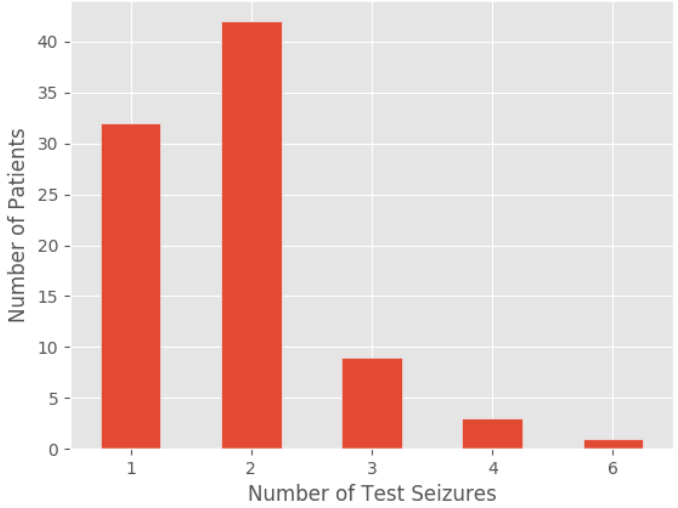

FIGURE 4.6: Histogram of number of test seizures for scalp patients.

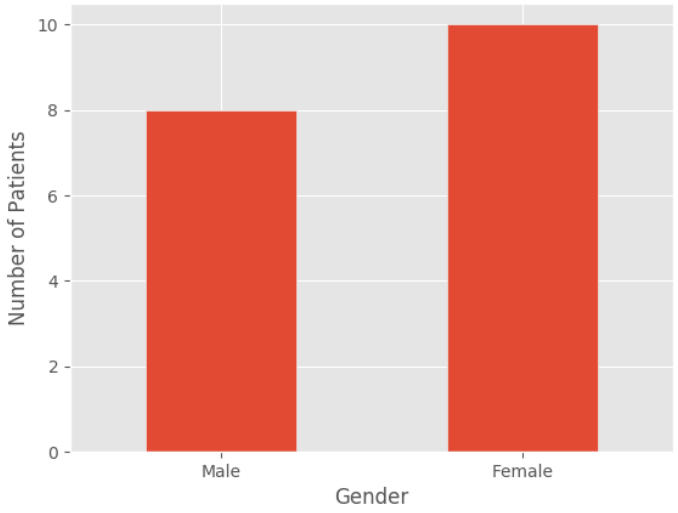

FIGURE 4.7: Histogram of genders for invasive patients.

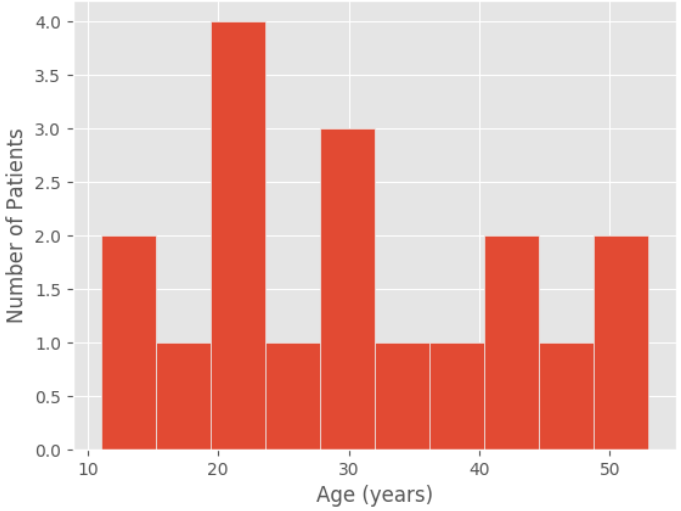

FIGURE 4.8: Histogram of ages for invasive patients.

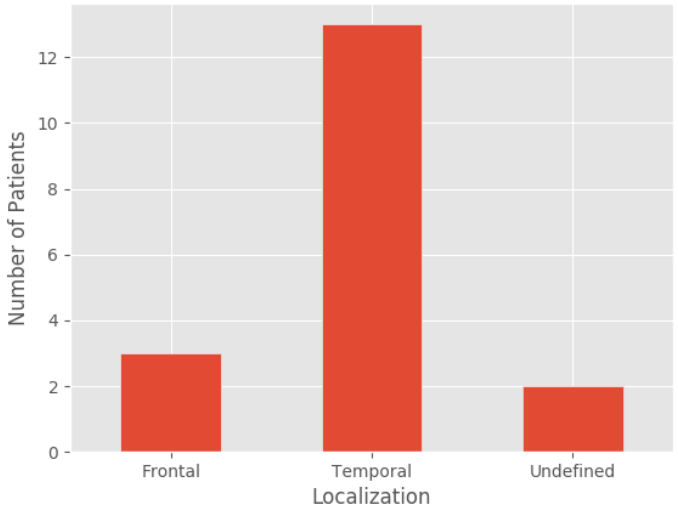

FIGURE 4.9: Histogram of seizure localization for invasive patients.

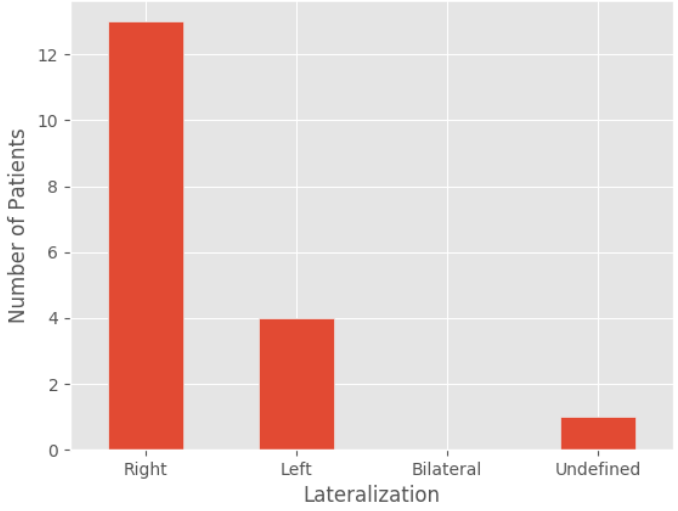

FIGURE 4.10: Histogram of seizure lateralization for invasive patients.

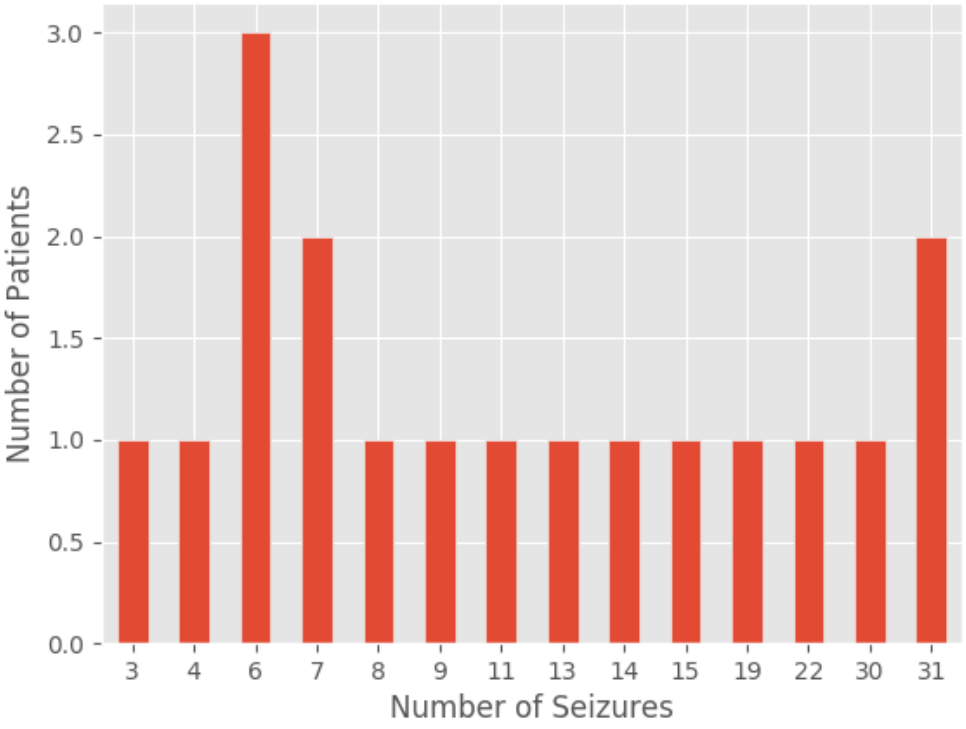

FIGURE 4.11: Histogram of number of seizures for invasive patients.

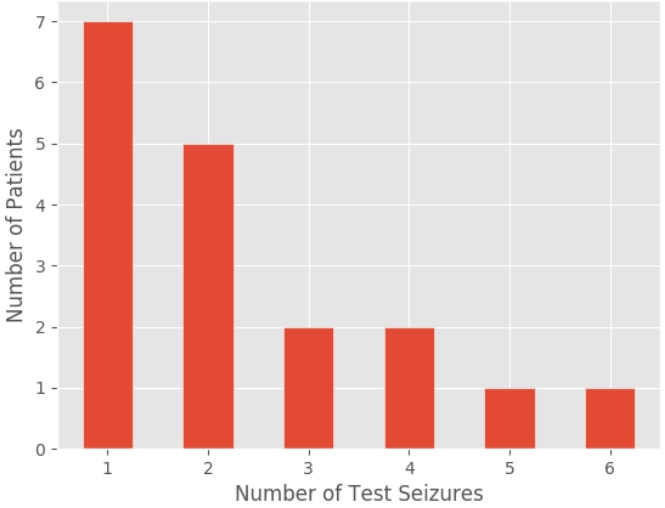

FIGURE 4.12: Histogram of number of test seizures for invasive patients.

4.2 Raw Data Pre-Processing

Following a standard procedure for EEG pre-processing [43], we used a 48-52 Hz, fourth order Notch Butterworth infinite impulse response (IIR) filter to remove the noise related to the power line.

4.3 Feature Extraction

For each EEG channel, 22 univariate features were computed based on 5s windows with no overlap. We choose this set of features because they are computationally efficient, with potential for real-time implementation.[40] They were also used in [40], which is our work of reference.

AR Modelling predictive errors

An autoregressive model was used to predict the EEG signal. The mean squared error between the predicted and original signals was then used as a feature. The prediction error has been suggested for seizure prediction [44], it is claimed that as seizures approach, the EEG signals become more predictable.

The autoregressive model of order p is given by 4.1

$$Y_t = c + \sum_{i}^{p} \phi_i Y_{t-i} + \epsilon_t \tag{4.1}$$

where Y_t is the predicted value for instant t, $\phi_1, ..., \phi_p$ are the parameters of the model, c is a constant, and ϵ_t is white noise. An AR model with order 10 and the Burg maximum entropy method was used.

Decorrelation time

The decorrelation time corresponds to the first zero crossing of the autocorrelation function and is an indicator of how data differs from random noise, the bigger

the decorrelation time, the more random is the time series. The autocorrelation function of time series x_n, $n = 1, ..., N$ can be defined as 4.2.

$$c_{xx_k} = \frac{\sum_{i=1}^{N-k} x_i x_{i+k}}{(N-1)\sigma^2} \tag{4.2}$$

where x_i is the i^{th} value of the time series and σ is the standard deviation of the time series.

A decrease on the power of the lower frequencies during the preictal period results in an increase on the decorrelation time. [45]

Hjorth Mobility and Complexity

Hjorth Mobility [46] is a quantification of the variance of the slopes of a time series normalized by the variance of its amplitude distribution.

Hjorth Complexity [46] is the variance of the rate of the slope changes having an ideal sinusoid as their reference.

Mathematically, Hjorth Mobility is given by 4.3 and Hjorth Complexity by4.4.

$$H_M = \sqrt{\frac{\sigma^2(x'_n)}{\sigma^2(x_n)}} \tag{4.3}$$

$$H_C = \sqrt{\frac{\sigma^2(x''_n) \times \sigma^2(x_n)}{(\sigma^2(x'_n))^2}} \tag{4.4}$$

x'_n and x''_n are the first and second derivatives of the time series x_n. σ represents the standard deviation.

A decrease on the power of the lower frequencies during the preictal period implies an increase in the Hjorth Mobility and Complexity.[45]

Relative spectral power

Considering the main spectral frequency bands defined in classical EEG analysis: δ (0.5-4Hz), θ (4-8Hz), α (8-13Hz), β (13-30Hz) and γ (>30Hz), the relative spectral power in each frequency band corresponds to the average of the square fast Fourier transform (FFT) coefficients in that frequency band divided by the total spectral power. It was reported that during the preictal period, a shift of EEG power from the lower to the higher frequencies occurs.[45]

Spectral edge frequency and power

Most of the spectral power is comprised in the 0.5-40Hz band. We define spectral edge frequency as the frequency below which 50% of the total power of the signal is located. The spectral edge power is the value of the power existing below the spectral edge frequency.

Statistical moments

Four statistical measures were used as a feature: mean, variance, skewness and kurtosis, given by equations 4.5, 4.6, 4.7 and 4.8, respectively. It was reported that the variance and kurtosis vary significantly in the preictal phase. A decrease in variance and an increase in kurtosis were observed in the preictal time when compared with interictal data [47].

$$\mu = \frac{\sum_{i=1}^{N} Y_i}{N} \tag{4.5}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} Y_i - \mu}{N} \tag{4.6}$$

$$S = \frac{\sum_{i=1}^{N} (Y_i - \mu)^3}{N s \sigma^3} \tag{4.7}$$

$$K = \frac{\sum_{i=1}^{N} (Y_i - \mu)^4}{N \sigma^4} \tag{4.8}$$

In the above equations, N is the number of samples of the time series, Y_i is the value at time i and σ is the standard deviation.

Accumulated Energy

Accumulated Energy is measure of the signal which has shown promising results[48]. For a time series x_n, $n = 1, ..., N$ it is defined as the mean of the squared values. 4.9.

$$Energy = \frac{\sum_{i=1}^{N} x_i^2}{N} \tag{4.9}$$

Energy of wavelet coefficients

The signal was decomposed into six decomposition levels using the wavelet transform. The EEG signal was decomposed in six levels by considering a Daubechies-4 wavelet (DB-4).[49] DB4 signal decomposition was successfully used in the past for seizure detection. [50]

The energy (4.9) of the wavelet coefficients of the various levels was used as a feature.

4.4 Feature Pre-Processing

4.4.1 Normalization

To prevent numerical errors that might arise during computations due to round-off and truncation errors, each dimension of the data went through a unit variance and zero mean normalization, also known as z-score normalization.

4.4.2 Electrode Selection

In this study, for each patient only data from 6 electrodes was used to train the models. The reason for the choice of a small number of electrodes is that only a

small percentage of patients would consider using an ambulatory device with a complete array of scalp electrodes [51]. Other studies have also suggested that six electrodes might be appropriate for seizure prediction.[52][53][27] With this in mind we expect that six electrodes are enough to provide enough information while at the same time providing comfort for the patients. For each patient we selected three electrodes as close as possible to the focal region and other three in regions far from the focal region.

4.5 Labelling

We considered seizure prediction as a two-class problem. Every sample in a certain period preceding a seizure, called the seizure occurrence period (SOP), was labelled as pre-ictal, while all the other samples are labelled as non-pre-ictal, which include all the samples that belong to the interictal, ictal and postictal states.

4.6 Data division

Each patient dataset was divided in three blocks: train, validation and test. The train block contains 60% of the seizures, while the validation and test sets contain 20% each. A visual representation of the division can be seen in figure 4.13.

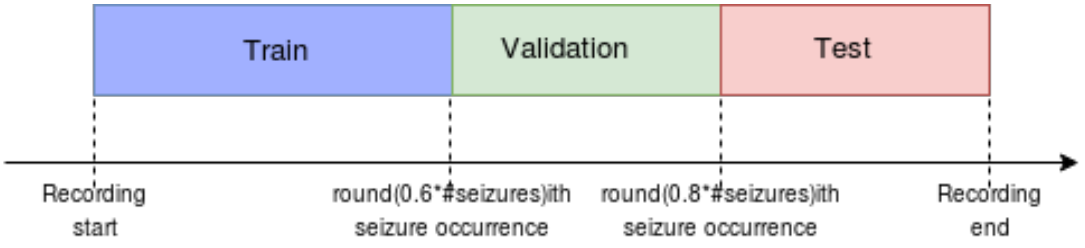

FIGURE 4.13: Data division method.

Furthermore, to prevent class bias when training due to class imbalance, we randomly removed non-preictal patterns from the training set until the number of patterns of the non-preictal and preictal classes was equal.

4.7 Training

For each patient, several models were trained, the differences being in the choice of architecture and parameters of the models. The diferent free parameters are presented in section 4.7.5.

4.7.1 Architectures

For each patient, we experimented with two types of architectures: Long Short Term Memory (LSTM) and Convolutional Long Short Term Memory (C-LSTM). The LSTM architecture as the name implies, is a succession of LSTM layers. The C-LSTM architecture consists of the inputs passing through a one dimensional convolutional network which feeds its outputs into a LSTM network.

4.7.2 Overfitting Control

The methods used to avoid overfitting were: L2 regularization, dropout and early-stopping.

4.7.2.1 L2 Regularization

Regularization techniques are used to reduce overfitting. One of the most commonly used regularization technique, and which used to train the models is the L2 regularization. L2 regularization works by adding a term to the loss function, resulting in equation 4.10.

$$E = \frac{1}{N} \sum_{i} L_i + \lambda \sum_{w} w^2 \tag{4.10}$$

The first term of 4.10 is the average loss over all patterns, the second term is the regularization loss. In the second term, the sum is done for all the weights of the

model, and the parameter λ is called the regularization strength, it has to be tuned for each problem. The effect of regularization is to make the network learn small weights evenly spread, instead of having a portion of weights with large values while another portion has small values. It has been shown that regularization reduces overfitting. [54].

4.7.2.2 Dropout

Dropout [55] is another technique used to reduce overfitting. The idea is that during each training epoch, each neuron has a probability d of not being used for the output computation, having no contribution during the backpropagation of the errors. This implies that while using dropout multiple architectures are explored during training. Because the weights are shared between the different architectures, they becomes regularized.

4.7.2.3 Early Stopping

Early stopping keeps track of the loss on the validation set as the learning proceeds, and stops training as validation loss does not decrease after n epochs. The weights of the model are then reverted to the ones that provided the lowest loss.

In our set up, we kept n equal to 7.

4.7.3 Optimizer

Adaptive Moment Estimation (Adam) [56] was the optimization algorithm used to update the networks weights at each iteration.

Adam is a method that computes adaptive learning rates for each parameter. It stores an exponentially decaying average of past gradients m_t (first moment), and an exponentially decaying average of past squared gradients v_t (second moment).

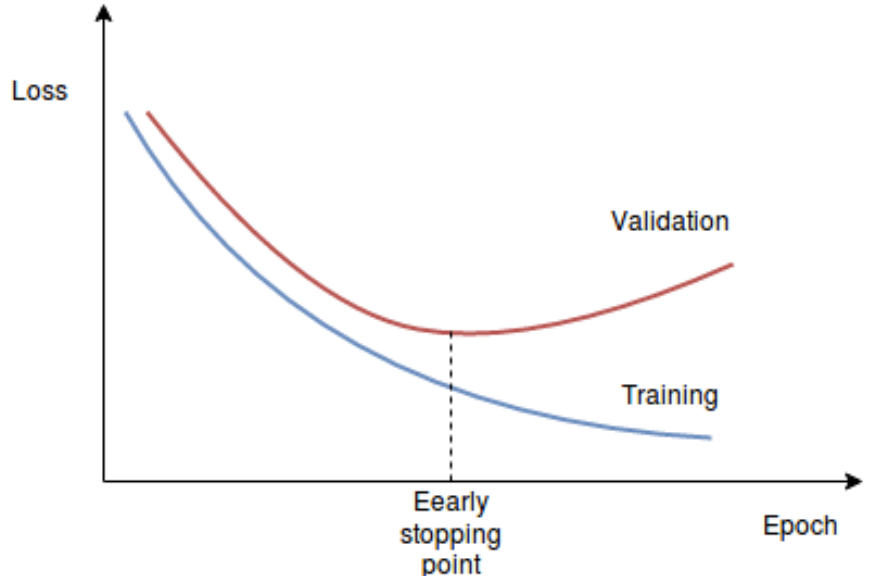

FIGURE 4.14: Early stopping: the red line represents the validation loss and the blue line the training loss, during each training epoch. The weights of the model are reverted back to the early stopping point.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{4.11}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{4.12}$$

where g_t are the actual gradients. m_t and v_t are initialized as vectors of 0's, which causes them to be biased towards zero during the initial time steps and when the decay rates β_1 and β_2 are small. To correct these biases, a second computation of the moments is performed:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1} \tag{4.13}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2} \tag{4.14}$$

$\hat{m}_t$ and $\hat{v}_t$ are then used to update the network weights and bias:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t \tag{4.15}$$

$$b_{t+1} = b_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t \tag{4.16}$$

The authors propose default values of 0.9 for β_1, 0.999 for β_2, and 10^{-8} for ϵ. We choose this optimizer because it was shown empirically that it compares favourably to other optimizers. [56]

4.7.4 Loss function

We selected the cross entropy as the loss function to minimize during training. It is given by

$$L = -\sum_{i=1}^{N} \frac{1}{N} log_2(\hat{y}_i) \tag{4.17}$$

where N is the number of samples and $\hat{y}_i$ is the probability, given by the network, that sample i belongs to its true class.

4.7.5 Parameters Overview

Overall our models had the following free parameters:

- d - dropout probability
- λ - regularization strength
- n - number of epochs, without a decrease of the validation loss, necessary for the termination of training
- η - learning rate

- SOP - seizure occurrence period
- l - length of the sequence of inputs
- units - number of neurons per hidden layer
- layers - number of hidden layers of the network
- k - number of filters of the one dimensional convolution network.

4.8 Post-Processing

Considering binary classifiers that classify electroencephalogram epochs in preictal vs non-preictal, the concept of firing power is applied to reduce the number of false positives. The equation of the firing power [36] is presented in 4.18.

$$fp[n] = \frac{\sum_{k=n-\tau}^{n} o[k]}{\tau} \tag{4.18}$$

where $o[k]$ is the classifier output for sample k. If $fp[n] = 1$ it means that the classifier output for the previous τ patterns was positive, in the opposite way, if $fp[n] = 0$ all the outputs are zero. Each pattern is classified as pre-ictal only if $fp[n]$ is over a certain threshold θ.

When an alarm is generated, we consider that no alarm can occur for the following seizure occurrence period (SOP).

4.9 Performance Measures Model Selection

Several models are tarined for each patient. Afterwards, for each model,the firing power filter (4.18) was applied to the output of each model on the validation set, using different combinations of the parameters θ and γ. θ is the threshold for $fp[n]$ at which an alarm is generated. γ is the threshold for $\sigma(\boldsymbol{n})_{ictal}$ (2.7) at which a pattern is classified as preictal.

After applying the firing power filter multiple times to each model, we select the model and combination of θ and γ, for each patient, that perform the best on the validation set. Performance is measured by the distance d to the optimal prediction performance (4.19), which is 100% sensitivity and 0 FPR.

$$d = \sqrt{FPR^2 + (1 - sensitivity)^2} \tag{4.19}$$

FPR is computed by dividing the number of false positives by the non-pre-ictal time minus the number of false positives times the seizure occurrence period.

$$FPR = \frac{FP}{t_{non-pre-ictal} - FP \times SOP} \tag{4.20}$$

The reason why we subtract $FP \times SOP$ from the denominator is because when an alarm is generated we consider that a seizure is approaching, therefore, for a period SOP no false alarms can occur.

Sensitivity is simply the ration between the number of correctly predicted seizures and the total number of seizures.

$$sensitivity = \frac{\#predicted\_seizures}{\#seizures} \tag{4.21}$$

4.10 Model Evaluation

After selecting the best model for each patient, the final performance of each model was accessed on the test set. The final performance is measured by the sensitivity (4.21) and false positive rate (4.20).

4.11 Statistical Validation

To statistically validate our results, we follow a method proposed by Schelter et al. (2006) [27]. The method computes the sensitivity of a random predictor that generates alarms following a Poisson process. It is then possible to statistically validate the results of a model by comparing its sensitivity with the one of the random predictor.

Considering K seizures in the test set, k the number of predicted seizures and z the number of independent models selected for testing. The probability P_k of predicting k seizures by means of at least one of the d models is

$$P_k = \{k; K; P_{SOP}\} = 1 - \left(\sum_{j<k} \binom{K}{j} P_{SOP}^{j} (1 - P_{SOP})^{K-j} \right)^{d} \tag{4.22}$$

P_{SOP} is the probability of emitting an alarm within the SOP.

Considering a significance level α, the upper sensitivity of the random predictor can be defined as

$$upper = \max_{k}(P_k\{k; K; P_{SOP}\} > \alpha) \times 100\% \tag{4.23}$$

A model is statistically validated only if the obtained sensitivity is superior the the upper sensitivity of the random predictor.

In our study only one model is evaluated on the testing set, therefore we used d equal to 1. We considered a significance level α of 5%.

4.12 Implementation

A framework for training and evaluating models was written in Python. The framework has the following functionalities:

- Read csv files containing the features and the targets.
- Normalize the data.
- Create sequential inputs. (2.4.7)
- Separate the data by choosing the percentage of seizures included in each set.
- Train models and save models. The number of layers and neurons per layer are a free parameter.
- Perform a grid search for the regularization and dropout.
- Evaluate models on the validation and test sets.

The framework is available on https://github.com/rodrigomfw/framework

Training of the LSTMs was done using the Keras library, a high-level neural networks API, written in Python and capable of running on top of Theano, a Python library which allows the use of GPU acceleration.

The computer used to train the models had the following specifications:

- RAM - 20 Gb
- CPU - Intel i7 975 @ 3.33GHz
- GPU - NVIDIA GeForce GTX 1080 8Gb

We was also integrated a module for training LSTMs in the EPILAB toolbox (appendix C).

Chapter 5

Results & Discussion

In this chapter we present and discuss the obtained results using both architectures: LSTM and C-LSTM.

Detailed results are presented in appendix A

5.1 LSTM

We trained a patient-specific model for each combination of parameter values presented in table 5.1. Preliminary experiments were done to obtain the best interval for each parameter.

TABLE 5.1: Values, for each parameter, used during training.

Parameter	Values
η	0.001
p	0.6, 0.7, 0.8
λ	0.001, 0.01, 0.1
n	7
sop(min)	10, 20, 30, 40
l	50
units	200, 400, 600
layers	2, 3

We selected, for each patient, the best model by its proximity to the optimal predictor on the validation set (4.19).

We analysed the influence of a variety of factors on performance (sensitivity and FPR). The factors are: recording type (scalp or invasive), gender, age, seizure localization, seizure lateralization, recording sampling frequency, number of test seizures and SOP.

We used the Kruskal-Wallis test of significance to statistically validate the differences in performance. The choice of this test, a non-parametric one, comes from the fact that the results do not come from a normal distribution. We verified normality of results using the DAgostino Pearson test.

In all our tests we considered a significance level of 5%, and our null hypothesis is that a factor has no influence on performance.

For all patients, scalp and invasive, our LSTM models achieved an average sensitivity of 29.28% and an average FPR of 0.58/h. We predicted 52 out of 203 (25.62%) seizures on the testing set. It is observed that for 5 out of 105 (4.8%) patients, optimal test performance with sensitivity $\geq 50\%$ and FPR $\leq 0.15h^{-1}$ was achieved. Perfect performance with 100% sensitivity and 0 h^{-1} FPR was achieved for 2 (1.9%) patients.

Performance was superior to a random predictor for 7 (6.67%) patients (5 with scalp recording and 2 with invasive).

In table 5.4 it is present the results obtained for each factor mentioned. It is also included the p-values of the Kruskal-Wallis test.

Although scalp patients present a higher average sensitivity and a lower average FPR than invasive patients, these differences were not statistically significant according to a Kruskal-Wallis test.

We observe that the average values for sensitivity and FPR are very similar for both genders. We also verify that the box plots for each gender are very similar (Figure

5.2). The statistical test confirms that the influence of gender on performance is not significant.

Those with ages superior to 40 have a higher average sensitivity and a lower average FPR, when comparing with the other two groups of an inferior age. However, those differences in performance are not statistically significant.

Seizure localization induce significant changes only in FPR, therefore affecting performance. Looking at figure 5.4 it is observed that those with an undefined seizure focus have higher FPRs, which can be explained by the fact that our models use data from 6 electrodes, 3 of them in the focal region. If there is no focus then it is not possible to select 3 electrodes in the focal region, which should contain essential information for prediction. It is also observed that for the patients with frontal lobe seizures the average sensitivity is significantly higher comparing with other focal regions, while having a low average FPR of 0.21/h.

Analysing figure 5.5 we verified that the box plots of the sensitivities are almost identical for all the seizure lateralizations, however, observing the box plots of the FRPs it is clear that those with an undefined lateralization have higher FPRs. This can also be explained by the fact that by having an undefined lateralization it is difficult to place 3 electrodes in the seizure focus, loosing valuable information. The differences in the FPR were, however, not statistically significant.

Comparing the sampling frequencies of 1024Hz and $\leq$256Hz we observe that the former has higher sensitivity and a similar FPR, therefore providing a better performance. This behaviour is expected since a higher sampling frequency implies a better signal quality. However, when comparing the sampling frequencies of 1024Hz and 2500Hz we find that the latter performs worse, but considering that only 6 patients have a sampling frequency of 2500Hz, we can not make a solid conclusion about this fact. Observing the p-values of the Kruskal-Wallis test we found that none of the differences induced by the sampling rate were statistically significant.

The number of test seizures seems to have no impact on performance. Even thought the group with only one test seizure has a slightly higher sensitivity it comes at the expense of a higher FPR, therefore no conclusions can be made. The statistically test also proves that the differences in performance are not significant.

In our study we considered four SOP values: 10, 20, 30, 40. After choosing the best SOP for each patient we observe an average SOP of 30.57 minutes, with the upper values, 30 and 40, being selected for 71% of the patients. This indicates that higher SOPs lead to better results, which is in agreement with previous studys [36] [40]. It is also observed that a SOP of 40 minutes produces the highest sensitivity and the lowest FPR. The differences in the sensitivity were found to be statistically significant.

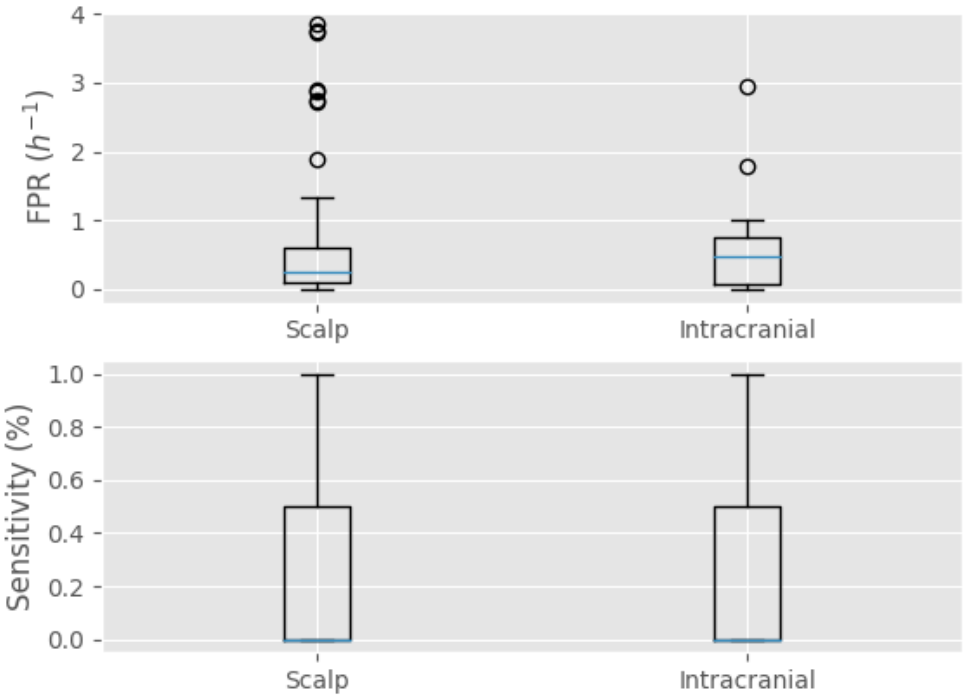

FIGURE 5.1: Boxplots of FPR and sensitivity for scalp and invasive patients.

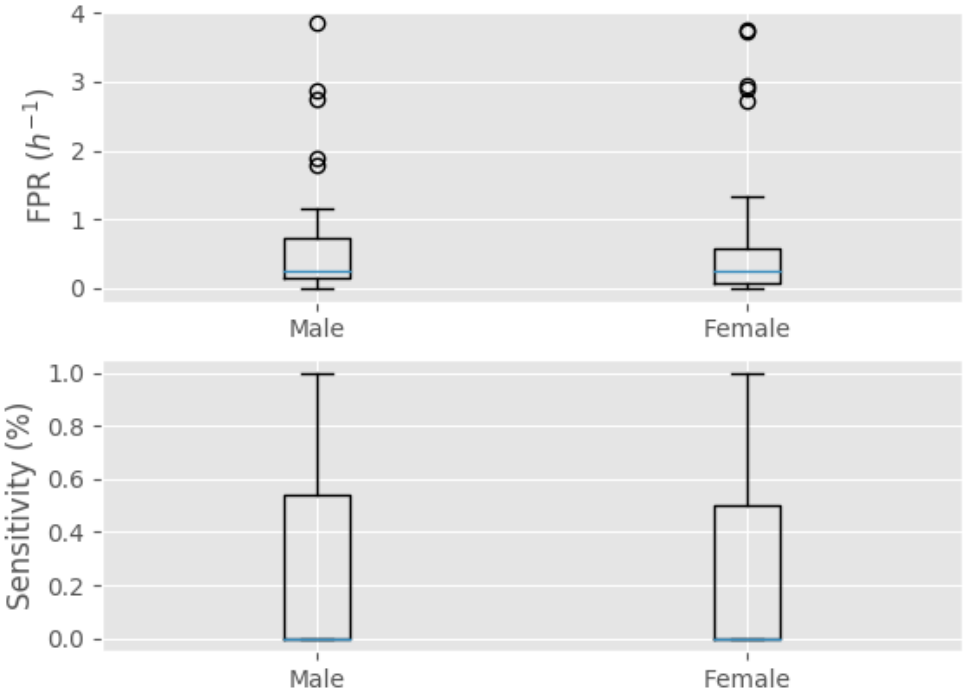

FIGURE 5.2: Boxplots of FPR and sensitivity for all patients grouped by gender.

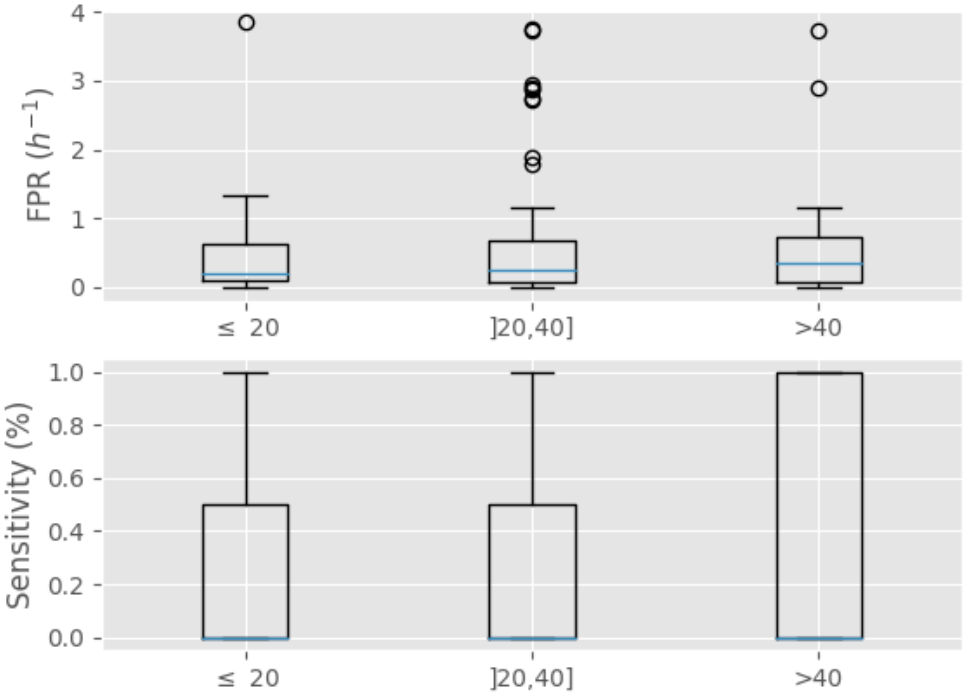

FIGURE 5.3: Boxplots of FPR and sensitivity for all patients grouped by age.

TABLE 5.2: Influence of the different factors on performance. "Rec. Type" is the recording type, scalp or invasive; "Gen." is the gender, "M" is male and "F" is female; "Age" is the patient age during the recording; "Loc." is the seizure localization: "F" - frontal, "T" - temporal, "C" - central, "H" - whole hemisphere, "Und" - undefined localization; "Lat." is the seizure lateralization; "S.F" is the sampling frequency of the recording; "SOP" is the best seizure occurrence period found for the patient; "# Pat." is the number of patients; "Avg. SS" is the average sensitivity; "Std. SS" is the standard deviation of the sensitivity; "Avg. FPR" is the average false positive rate; "Std.FPR" is the standard deviation of the false positive rate; "p-value" is the p-value obtained using the Kruskal-Wallis test.

		#Pat.	Avg. SS(%)	Std. SS(%)	Avg.FPR(h^{-1})	Std.FPR(h^{-1})
Rec. Type	Scalp	87	29.59	40.82	0.57	0.88
	Invasive	18	27.78	42.78	0.62	0.75
	p-value		0.78		0.59	
Gen.	M.	53	28.28	40.64	0.59	0.92
	F.	52	30.29	41.66	0.57	0.79
	p-value		0.58		0.68	
Age	≤ 20	13	29.46	43.11	0.62	1.04
	]20,40]	55	26.2	38.5	0.58	0.88
	> 40	37	33.78	44.36	0.56	0.76
	p-value		0.80		0.78	
Loc.	F	14	42.86	47.46	0.21	0.29
	T	72	26.03	40.27	0.58	0.83
	C	2	25.0	35.36	0.1	0.13
	H	3	16.67	28.87	0.33	0.06
	Undefined	14	35.71	41.81	1.09	1.23
	p-value		0.60		0.01	
Lat.	Right	48	27.42	42.17	0.47	0.7
	Left	37	31.97	41.08	0.46	0.66
	Bilateral	8	28.13	41.05	0.89	1.29
	Undefined	12	29.17	40.31	1.19	1.3
	p-value		0.91		0.54	
S.F.(Hz)	≤ 256	45	22.96	36.29	0.5	0.8
	400	13	34.62	47.37	0.77	1.01
	512	18	36.11	44.74	0.62	0.81
	1024	23	32.96	44.84	0.58	0.88
	2500	6	30.5	40.02	0.63	1.13
	p-value		0.83		0.72	
# T. Sz.	1	39	33.33	47.76	0.65	1.05
	[2,3]	58	26.71	36.27	0.55	0.73
	[4,5,6]	8	28.13	41.05	0.46	0.62
	p-value		0.98		0.46	
# SOP(min)	10	12	20.83	39.65	0.61	1.08
	20	18	27.78	42.78	0.86	1.22
	30	27	8.63	23.27	0.52	0.8
	40	48	43.56	43.71	0.5	0.63
	p-value		0.003		0.55	

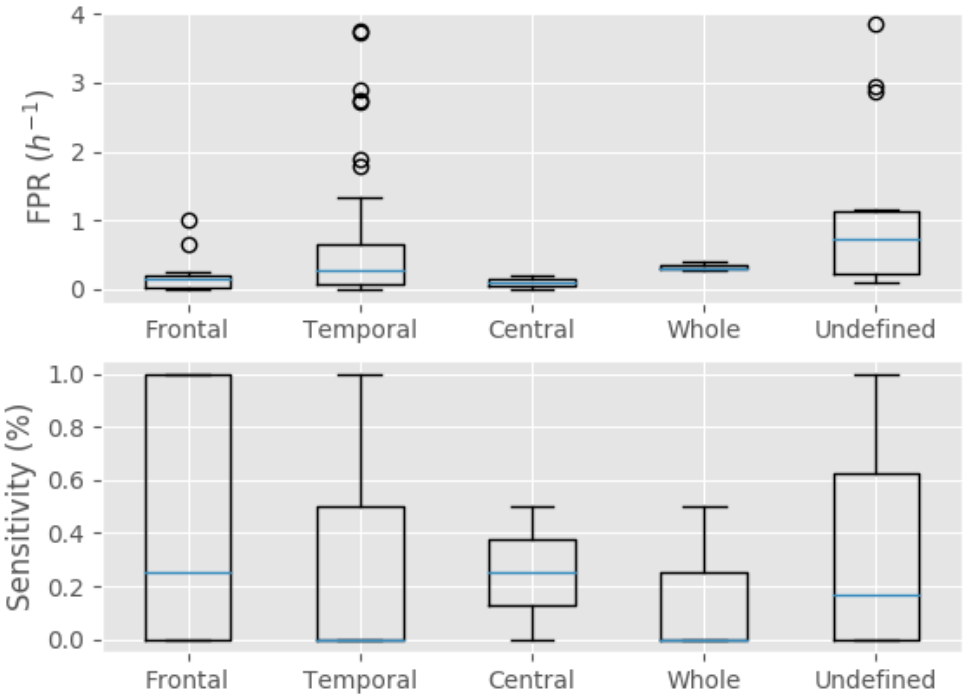

FIGURE 5.4: Boxplots of FPR and sensitivity for all patients grouped by seizure localization.

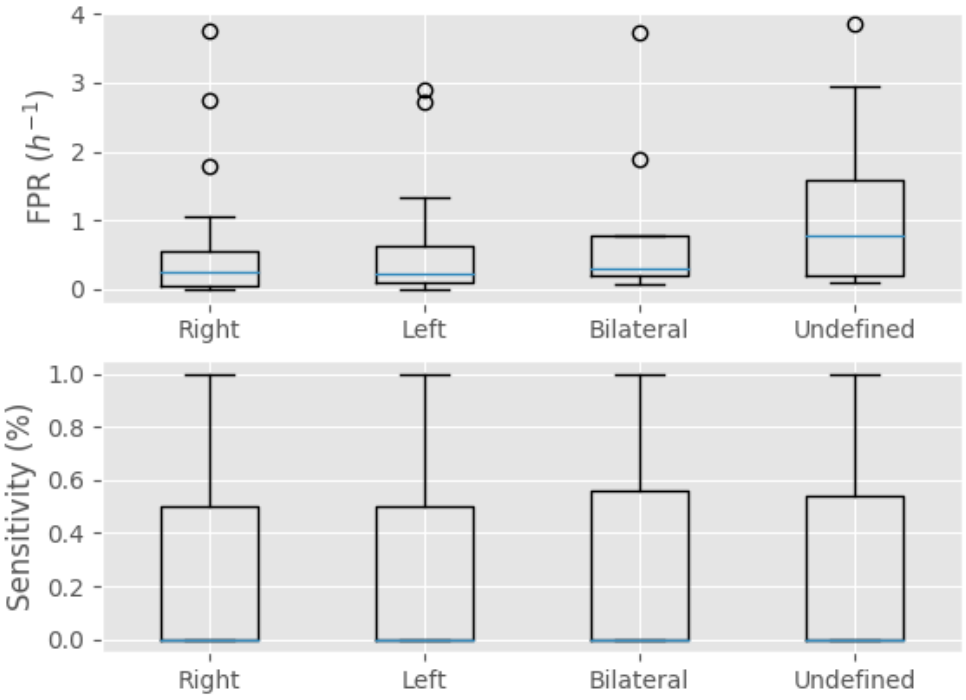

FIGURE 5.5: Boxplots of FPR and sensitivity for all patients grouped by seizure lateralization.

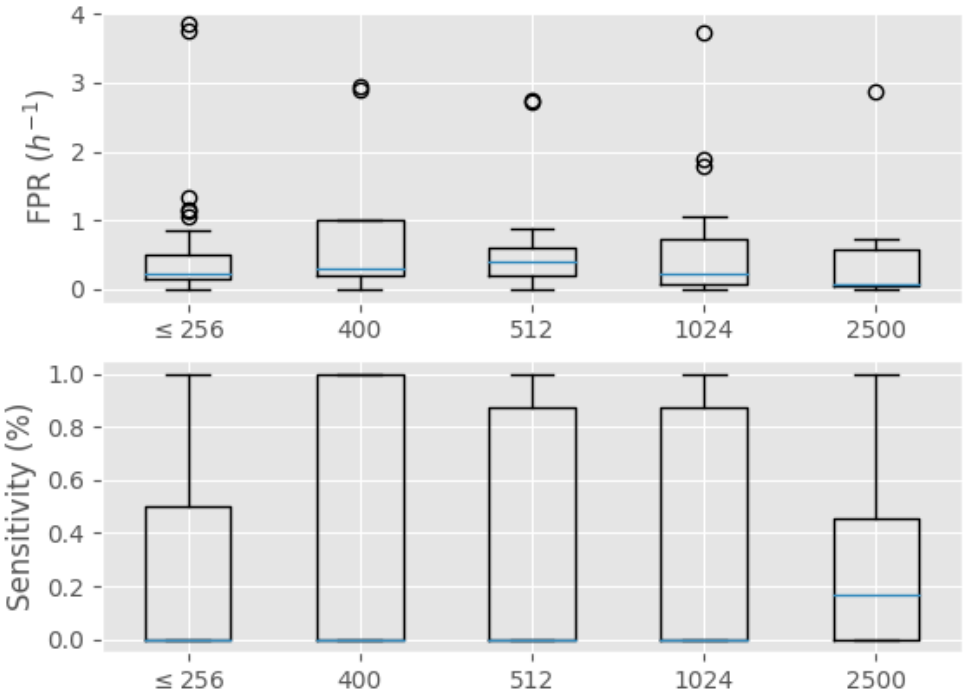

FIGURE 5.6: Boxplots of FPR and sensitivity for all patients grouped by sampling frequency.

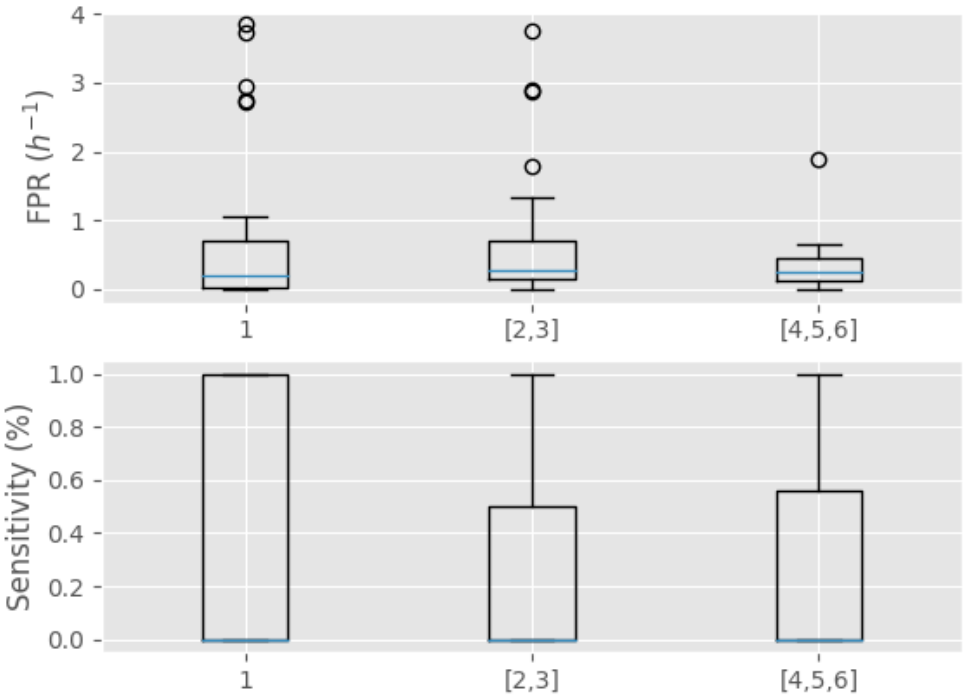

FIGURE 5.7: Boxplots of FPR and sensitivity for all patients grouped by number of test seizures.

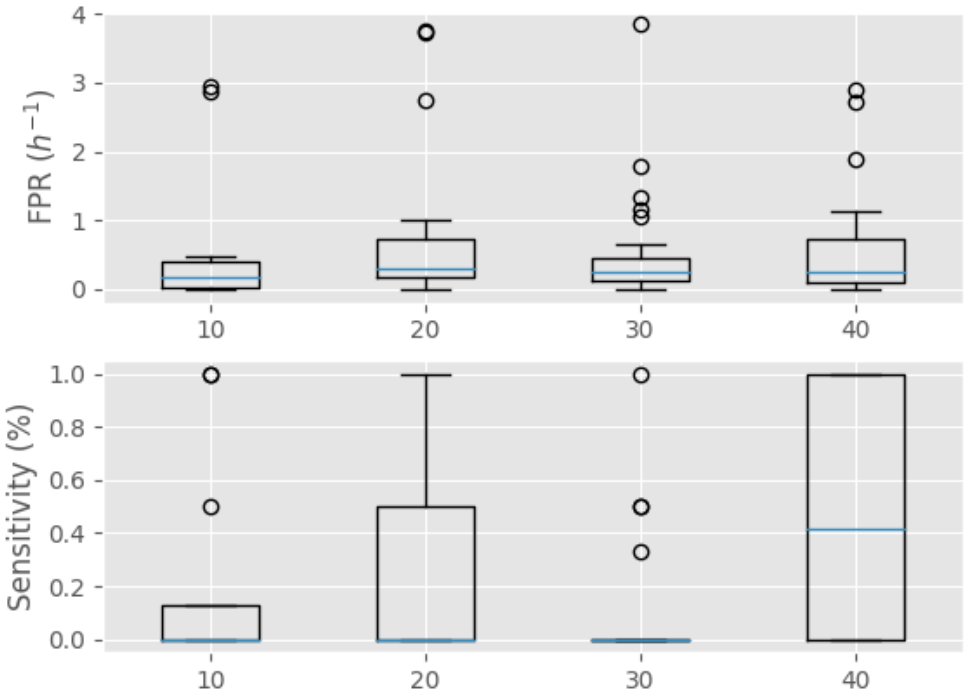

FIGURE 5.8: Boxplots of FPR and sensitivity for all patients grouped by SOP.

5.2 C-LSTM

We used a one dimensional convolution neural network to feed a LSTM network. While maintaining the best parameters found for the LSTM part of the network, we performed a grid search to tune the parameters the convolutional network (5.3).

TABLE 5.3: Values, for each parameter, used during training.

Parameter	Values
d	0.6, 0.7, 0.8
λ	0.01, 0.1, 1, 10
layers	2, 3
units	200
k	10,20,30

We selected, for each patient, the best model by its proximity to the optimal predictor on the validation set (4.19).

For all patients, scalp and invasive, our C-LSTM models achieved an average sensitivity of 28.17% and an average FPR of 0.64/h. predicted 54 out of 203 (26.60%) seizures on the testing set. It is observed that for 2 out of 105(1.9%) patients, optimal test performance with sensitivity $\geq 50\%$ and FPR $\leq 0.15h^{-1}$ was achieved. Perfect performance with 100% sensitivity and 0 h^{-1} FPR was achieved for 0 (0.0%) patients.

Performance was superior to a random predictor for 6 (5.7%) patients.

In table 5.4 is the sensitivity and FPR obtained for the LSTM and C-LSTM networks. LSTMs presents higher sensitivities and lower FPRs, although the difference is not statistically significant. However, optimal and perfect test performances, as well as performance above random, was achieved for fewer patients, concluding that using the C-LSTM architecture might actually decrease performance.

TABLE 5.4: Influence of architecture type on prediction performance. "Avg. SS" is the average sensitivity; "Std. SS" is the standard deviation of the sensitivity; "Avg. FPR" is the average false positive rate; "Std.FPR" is the standard deviation of the false positive rate.

		Avg. SS(%)	Std. SS(%)	Avg.FPR(h^{-1})	Std.FPR(h^{-1})
Arch. Type	LSTM	29.28	40.96	0.58	0.85
	C-LSTM	28.17	40.23	0.64	1.11
p-value		0.51		0.86	

5.3 Studies comparison

In Table 5.5 we have a comparison between our results using the LSTM architecture and the results of the study by Direito et al. (2016), which was based on SVMs. There is a clear difference in performance, they achieved a significantly higher sensitivity and lower FPR. They also achieved a higher percentage of patients with statistically significant results. This indicates that the LSTMs do not provide better results than SVMs. We were expecting the opposite since LSTMs are, in theory, a method expected to perform better than the SVMs when sequential data is involved. However, this is not a solid conclusion as the experimental conditions were not exactly the same. Even though we used data from the same database we only experimented with 105 patients, less than half the patients they used. Our training methods were also different, while they followed a cross-validation approach, we used one single block for training and another for validating.

TABLE 5.5: Comparision of performances achieved in our study and the study by Direito et al.(2016). "# Pat." is the number of patients considered in the study; "Avg. SS" is the average sensitivity; "Std. SS" is the standard deviation of the sensitivity; "Avg. FPR" is the average false positive rate; "Std.FPR" is the standard deviation of the false positive rate; "# Stat. Sig" is the number of patients with statistically significant results.

Study	#Pat.	Avg.SS(%)	Avg.FPR(h^{-1})	# Stat. Sig.
Direito et al.	216	38.47	0.20	24 (11%)
Ours	105	29.28	0.58	7 (6.67%)

Chapter 6

Conclusion and Future Work

In our study we achieved performance above chance for 6.67% of the patients. Although this is a small percentage, if other studies achieve statistically significant for the same subset, it is sign that they could be users of a clinical application.

We analysed the influence of a variety of factors on prediction performance, but only the seizure localization and SOP were found to have a statistically significant impact on performance.

Using the C-LSTM architecture as an attempt to extract a higher-level sequence of features came down as a failed experience, achieving worse results than with the LSTM architecture.

We achieved worse results when comparing our results with a study using the same database but based on SVMs. We conclude that it is not worth to use LSTMs, at least with current features, since they are a more computationally expensive model that does not provide better results. However, the number of patients and training methodologies considered were not the same. In a future study, experimental conditions should be identical to make a solid conclusion.

In the future we expect to experiment both the LSTM and C-LSTM architectures with raw EEG signal. We expect that the LSTMs will be able to capture better temporal dependencies on raw signal, since compacting 5 seconds of information into

a single value leads to a great loss on information. Furthermore, the convolutional networks are also typically used on raw data to extract features. A better approach for selecting training data will also be considered, instead of selecting the first 60% seizures we will select data in a way that covers a full day/night cycle, to better capture intra day variations.

Appendix A

Results

A.1 LSTM

TABLE A.1: Results obtained, in the testing set, for scalp patients and using the LSTM architecture. "ID" is the patient identification; "# Test Seiz." is the number of seizures present in the testing set; "Test Duration" is the total duration of the EEG recording used for testing; "True Alarms" is the number of correctly raised alarms; "FPR" is the false positive rate; "Sens." is the sensitivity (percentage of correctly predicted seizures); "C. Sens." is the critical sensitivity of the random predictor; "p-value" is the p-value of the test that determines if performance is above chance

ID	SOP	# Test Seiz.	Test Duration	# True Alarms	FPR(h^{-1}	Sens.(%)	C. Sens.(%)	p-value
70702	20	1	43.1	0	0.0	0.0	0.0	1.0
1306903	30	2	38.7	0	0.0	0.0	0.0	1.0
6700	40	1	7.1	1	1.07	100.0	100.0	0.51
102	40	2	24.8	1	0.19	50.0	50.0	0.22
70202	40	1	9.7	0	0.0	0.0	0.0	1.0
1314803	40	2	10.7	1	0.25	50.0	50.0	0.28
32702	20	1	23.4	0	0.18	0.0	100.0	1.0
71202	30	2	17.8	0	0.2	0.0	50.0	1.0
111902	40	2	71.3	0	0.01	0.0	0.0	1.0
30802	30	2	28.4	0	0.25	0.0	50.0	1.0
72602	10	2	50.7	0	0.47	0.0	50.0	1.0
1235103	40	1	41.4	0	0.19	0.0	100.0	1.0
71802	40	1	17.1	0	0.73	0.0	100.0	1.0
70502	30	2	31.0	0	1.34	0.0	100.0	1.0
73002	10	1	28.1	0	0.07	0.0	0.0	1.0
93902	40	2	21.2	0	0.05	0.0	50.0	1.0
98202	30	2	8.6	0	0.5	0.0	50.0	1.0

70302	40	3	56.8	0	0.08	33.0	33.33	0.15
1316503	20	1	6.2	0	0.62	0.0	100.0	1.0
1330203	30	1	83.5	1	0.16	100.0	100.0	0.08
400	40	2	39.3	2	0.59	100.0	100.0	0.11
72702	30	2	41.7	0	1.17	0.0	100.0	1.0
72402	30	3	38.9	0	1.07	0.0	100.0	1.0
114902	40	2	85.6	2	0.62	100.0	100.0	0.11
110602	40	2	17.9	2	0.86	100.0	100.0	0.19
5800	20	2	21.9	0	0.26	0.0	50.0	1.0
11002	30	2	28.3	0	0.4	0.0	50.0	1.0
51002	40	2	16.9	1	1.13	50.0	100.0	0.78
71102	10	2	22.4	0	0.0	0.0	0.0	1.0
70402	10	1	39.7	0	0.21	0.0	0.0	1.0
53402	40	2	5.3	0	0.31	0.0	50.0	1.0
8100	40	1	53.8	0	0.32	0.0	100.0	1.0
109502	40	1	65.9	0	0.08	0.0	100.0	1.0
72502	40	2	29.4	2	0.75	100.0	100.0	0.15
1329503	20	1	5.6	1	0.2	100.0	100.0	0.06
101702	40	1	53.9	1	0.02	100.0	0.0	0.01
1314703	40	1	16.0	0	2.73	0.0	0.0	1.0
7300	40	2	78.7	2	0.1	100.0	50.0	0.0
1316403	30	2	19.2	1	0.4	50.0	50.0	0.33
115102	30	2	42.3	0	0.02	0.0	0.0	1.0
81102	30	3	46.7	0	0.31	33.0	66.67	0.37
70602	20	2	48.6	0	3.76	0.0	100.0	1.0
5900	40	1	67.0	0	0.0	0.0	0.0	1.0
1307503	30	3	9.7	0	0.28	0.0	33.33	1.0
1306003	40	1	33.3	1	0.54	100.0	100.0	0.3
202	20	2	31.8	1	0.24	50.0	50.0	0.15
103802	30	1	7.4	0	3.85	0.0	100.0	1.0
6200	30	2	13.7	0	0.27	0.0	50.0	1.0
52302	40	1	71.6	0	0.23	0.0	100.0	1.0
1234303	30	1	29.6	0	0.0	0.0	0.0	1.0
1321803	20	1	7.3	1	2.75	100.0	100.0	0.6
1306203	10	3	29.9	0	0.03	0.0	0.0	1.0
81402	40	2	8.2	1	0.16	50.0	50.0	0.19
1325103	30	2	40.1	1	0.33	50.0	50.0	0.28
7800	20	1	43.0	1	3.74	100.0	100.0	0.71
1325903	40	2	69.4	2	1.01	100.0	100.0	0.24
3600	30	6	14.9	0	0.09	0.0	16.67	1.0
26102	20	2	70.6	1	0.35	50.0	50.0	0.21
6000	30	1	59.3	0	0.19	0.0	100.0	1.0
1300003	10	1	12.8	0	0.0	0.0	0.0	1.0
7302	10	2	61.6	0	0.15	0.0	0.0	1.0
94402	40	1	89.1	0	0.18	0.0	100.0	1.0

70902	10	2	59.9	1	2.88	50.0	100.0	1.0
1312703	10	1	27.7	1	0.39	100.0	100.0	0.06
1321903	40	1	6.1	1	0.89	100.0	100.0	0.45
200	20	2	35.1	0	0.15	0.0	50.0	1.0
70102	40	1	41.7	1	0.05	100.0	0.0	0.03
6500	40	3	42.1	0	0.23	33.0	66.67	0.37
2300	20	2	37.0	0	0.59	0.0	50.0	1.0
112402	30	2	21.0	0	0.16	0.0	50.0	1.0
80702	40	2	41.5	0	0.2	0.0	50.0	1.0
70802	40	3	24.4	2	0.84	67.0	100.0	0.39
1330903	40	2	65.5	0	0.3	0.0	50.0	1.0
95202	40	2	7.9	0	0.67	0.0	100.0	1.0
1307103	40	2	93.7	1	0.24	50.0	50.0	0.27
72302	40	2	24.4	0	0.14	0.0	50.0	1.0
113902	40	4	7.0	0	0.37	0.0	50.0	1.0
100002	30	4	22.9	0	0.15	0.0	25.0	1.0
72202	40	3	70.6	0	0.11	33.0	33.33	0.2
1328603	40	3	10.8	0	2.89	0.0	100.0	1.0
72802	10	1	9.5	1	0.22	100.0	0.0	0.04
71002	10	1	75.3	0	0.0	0.0	0.0	1.0
6100	40	4	17.0	3	1.9	75.0	100.0	0.68
96002	30	2	78.2	0	0.23	0.0	50.0	1.0
1313003	20	1	29.6	0	0.0	0.0	0.0	1.0
22602	40	1	4.3	1	0.0	100.0	0.0	0.0
2900	30	2	27.4	0	0.08	0.0	50.0	1.0

TABLE A.2: Results obtained, in the testing set, for invasive patients and using the LSTM architecture. "ID" is the patient identification; "# Test Seiz." is the number of seizures present in the testing set; "Test Duration" is the total duration of the EEG recording used for testing; "True Alarms" is the number of correctly raised alarms; "FPR" is the false positive rate; "Sens." is the sensitivity (percentage of correctly predicted seizures); "C. Sens." is the critical sensitivity of the random predictor; "p-value" is the p-value of the test that determines if performance is above chance

ID	SOP	# Test Seiz.	Test Duration	# True Alarms	FPR(h^{-1}	Sens.(%)	C. Sens.(%)	p-value
1259203	20	1	28.8	1	0.75	100.0	100.0	0.22
1308903	30	2	85.7	0	1.79	0.0	100.0	1.0
1310403	30	1	44.2	0	0.0	0.0	0.0	1.0
1235903	20	1	48.1	0	0.25	0.0	100.0	1.0
13902	30	1	16.7	0	0.66	0.0	100.0	1.0
1311203	40	3	146.6	3	0.73	100.0	100.0	0.06
52303	10	1	10.8	0	2.94	0.0	100.0	1.0
107702	20	2	35.9	0	0.06	0.0	0.0	1.0
38402	40	2	22.8	2	0.0	100.0	0.0	0.0
112502	40	3	112.4	0	0.75	0.0	100.0	1.0

97002	40	4	31.5	2	0.29	50.0	50.0	0.14
54802	40	6	11.3	0	0.0	0.0	0.0	1.0
1322903	20	1	29.5	0	1.01	0.0	100.0	1.0
44202	40	2	39.3	1	0.88	50.0	100.0	0.69
92202	40	4	13.5	4	0.21	100.0	50.0	0.0
1324503	20	5	135.0	0	0.66	0.0	60.0	1.0
62002	40	1	59.8	0	0.07	0.0	0.0	1.0
1272703	30	2	33.3	0	0.03	0.0	0.0	1.0

A.2 C-LSTM

TABLE A.3: Results obtained, in the testing set, for scalp patients and using the C-LSTM architecture. "ID" is the patient identification; "# Test Seiz." is the number of seizures present in the testing set; "Test Duration" is the total duration of the EEG recording used for testing; "True Alarms" is the number of correctly raised alarms; "FPR" is the false positive rate; "Sens." is the sensitivity (percentage of correctly predicted seizures); "C. Sens." is the critical sensitivity of the random predictor; "p-value" is the p-value of the test that determines if performance is above chance

ID	SOP	# Test Seiz.	Test Duration	# True Alarms	FPR(h^{-1}	Sens.(%)	C. Sens.(%)	p-value
70702	20	1	43.1	0	0.0	0.0	0.0	1.0
1306903	30	2	38.7	0	0.17	0.0	50.0	1.0
6700	40	1	7.1	0	0.0	0.0	0.0	1.0
102	40	2	24.8	1	1.12	50.0	100.0	0.78
70202	40	1	9.7	0	0.0	0.0	0.0	1.0
1314803	40	2	10.7	1	0.84	50.0	100.0	0.67
32702	20	1	23.4	0	0.29	0.0	100.0	1.0
71202	30	2	17.8	0	0.06	0.0	50.0	1.0
111902	40	2	71.3	0	0.0	0.0	0.0	1.0
30802	30	2	28.4	2	0.39	100.0	50.0	0.03
72602	10	2	50.7	0	0.0	0.0	0.0	1.0
1235103	40	1	41.4	0	0.37	0.0	100.0	1.0
71802	40	1	17.1	0	1.03	0.0	100.0	1.0
70502	30	2	31.0	0	1.53	0.0	100.0	1.0
73002	10	1	28.1	0	0.0	0.0	0.0	1.0
93902	40	2	21.2	0	0.17	0.0	50.0	1.0
98202	30	2	8.6	0	0.31	0.0	50.0	1.0
70302	40	3	56.8	2	0.96	67.0	100.0	0.46
1316503	20	1	6.2	0	2.0	0.0	100.0	1.0
1330203	30	1	83.5	0	0.61	0.0	100.0	1.0

400	40	2	39.3	1	0.75	50.0	100.0	0.63
72702	30	2	41.7	1	1.3	50.0	100.0	0.73
72402	30	3	38.9	0	0.59	0.0	66.67	1.0
114902	40	2	85.6	0	0.32	0.0	50.0	1.0
110602	40	2	17.9	2	0.72	100.0	100.0	0.15
5800	20	2	21.9	1	1.1	50.0	100.0	0.52
11002	30	2	28.3	0	0.0	0.0	0.0	1.0
51002	40	2	16.9	0	0.78	0.0	100.0	1.0
71102	10	2	22.4	0	0.0	0.0	0.0	1.0
70402	10	1	39.7	0	0.0	0.0	0.0	1.0
53402	40	2	5.3	1	0.0	50.0	0.0	0.0
8100	40	1	53.8	0	0.32	0.0	100.0	1.0
109502	40	1	65.9	0	0.0	0.0	0.0	1.0
72502	40	2	29.4	0	0.0	0.0	0.0	1.0
1329503	20	1	5.6	1	0.2	100.0	100.0	0.06
101702	40	1	53.9	1	0.32	100.0	100.0	0.19
1314703	40	1	16.0	0	0.07	0.0	0.0	1.0
7300	40	2	78.7	2	1.06	100.0	100.0	0.26
1316403	30	2	19.2	2	0.57	100.0	100.0	0.06
115102	30	2	42.3	0	0.02	0.0	0.0	1.0
81102	30	3	46.7	0	2.07	0.0	100.0	1.0
70602	20	2	48.6	0	0.76	0.0	100.0	1.0
5900	40	1	67.0	1	1.46	100.0	100.0	0.62
1307503	30	3	9.7	0	0.0	0.0	0.0	1.0
1306003	40	1	33.3	1	0.54	100.0	100.0	0.3
202	20	2	31.8	0	0.0	0.0	0.0	1.0
103802	30	1	7.4	1	5.44	100.0	100.0	0.93
6200	30	2	13.7	0	0.17	0.0	50.0	1.0
52302	40	1	71.6	0	0.8	0.0	100.0	1.0
1234303	30	1	29.6	0	0.04	0.0	0.0	1.0
1321803	20	1	7.3	0	0.0	0.0	0.0	1.0
1306203	10	3	29.9	0	0.0	0.0	0.0	1.0
81402	40	2	8.2	1	0.16	50.0	50.0	0.19
1325103	30	2	40.1	1	0.44	50.0	50.0	0.36
7800	20	1	43.0	1	6.86	100.0	100.0	0.9
1325903	40	2	69.4	2	0.51	100.0	100.0	0.08
3600	30	6	14.9	1	0.0	33.0	0.0	0.0
26102	20	2	70.6	0	0.0	0.0	0.0	1.0
6000	30	1	59.3	1	0.54	100.0	100.0	0.24
1300003	10	1	12.8	0	0.0	0.0	0.0	1.0
7302	10	2	61.6	0	0.0	0.0	0.0	1.0
94402	40	1	89.1	0	0.32	0.0	100.0	1.0
70902	10	2	59.9	0	0.0	0.0	0.0	1.0
1312703	10	1	27.7	0	0.0	0.0	0.0	1.0
1321903	40	1	6.1	1	1.48	100.0	100.0	0.63

200	20	2	35.1	0	0.0	0.0	0.0	1.0
70102	40	1	41.7	0	0.0	0.0	0.0	1.0
6500	40	3	42.1	0	0.37	33.0	66.67	0.52
2300	20	2	37.0	0	0.0	0.0	0.0	1.0
112402	30	2	21.0	0	0.11	0.0	50.0	1.0
80702	40	2	41.5	0	0.26	0.0	50.0	1.0
70802	40	3	24.4	2	0.47	67.0	66.67	0.18
1330903	40	2	65.5	1	0.53	50.0	100.0	0.51
95202	40	2	7.9	1	1.6	50.0	100.0	0.88
1307103	40	2	93.7	0	0.16	0.0	50.0	1.0
72302	40	2	24.4	0	0.2	0.0	50.0	1.0
113902	40	4	7.0	1	0.37	25.0	50.0	0.63
100002	30	4	22.9	0	0.0	0.0	0.0	1.0
72202	40	3	70.6	0	0.16	33.0	33.33	0.27
1328603	40	3	10.8	0	2.89	0.0	100.0	1.0
72802	10	1	9.5	0	0.0	0.0	0.0	1.0
71002	10	1	75.3	0	0.0	0.0	0.0	1.0
6100	40	4	17.0	3	1.27	75.0	100.0	0.43
96002	30	2	78.2	0	1.14	0.0	100.0	1.0
1313003	20	1	29.6	0	0.0	0.0	0.0	1.0
22602	40	1	4.3	1	1.81	100.0	100.0	0.7
2900	30	2	27.4	0	0.31	0.0	50.0	1.0

TABLE A.4: Results obtained, in the testing set, for invasive patients and using the C-LSTM architecture. "ID" is the patient identification; "# Test Seiz." is the number of seizures present in the testing set; "Test Duration" is the total duration of the EEG recording used for testing; "True Alarms" is the number of correctly raised alarms; "FPR" is the false positive rate; "Sens." is the sensitivity (percentage of correctly predicted seizures); "C. Sens." is the critical sensitivity of the random predictor; "p-value" is the p-value of the test that determines if performance is above chance

ID	SOP	#Test Seiz.	Test Duration	#True Alarms	FPR	Sens.(%)	C. Sens.(%)	p-value
1259203	20	1	28.8	1	0.86	100.0	100.0	0.25
1308903	30	2	85.7	0	2.32	0.0	100.0	1.0
1310403	30	1	44.2	0	0.0	0.0	0.0	1.0
1235903	20	1	48.1	0	0.15	0.0	0.0	1.0
13902	30	1	16.7	1	1.03	100.0	100.0	0.4
1311203	40	3	146.6	3	0.26	100.0	66.67	0.0
52303	10	1	10.8	0	1.39	0.0	100.0	1.0
107702	20	2	35.9	0	0.0	0.0	0.0	1.0
38402	40	2	22.8	2	0.05	100.0	50.0	0.0
112502	40	3	112.4	0	0.33	0.0	66.67	1.0

97002	40	4	31.5	3	0.58	75.0	75.0	0.1
54802	40	6	11.3	0	0.0	0.0	0.0	1.0
1322903	20	1	29.5	1	5.92	100.0	100.0	0.86
44202	40	2	39.3	1	0.7	50.0	100.0	0.61
92202	40	4	13.5	4	0.21	100.0	50.0	0.0
1324503	20	5	135.0	0	1.68	0.0	80.0	1.0
62002	40	1	59.8	0	0.56	0.0	100.0	1.0
1272703	30	2	33.3	0	0.0	0.0	0.0	1.0

Appendix B

Patients characteristics

Table B.1: Patients characteristics for scalp patients. 'ID' is the patient identification; Gender - 'm' = male, 'f' = female; Age is the patient age at the time of the recording; Electrodes is the number of EEG electrodes used during the recording; 'Loc.' is the seizure localization - 'f' = frontal region, 't' = temporal region, 'c' = central region, 'o' = occipital region, 'p' = parietal region, 'h' = complete hemisphere, '-' = impossible to define a cerebral region; 'Lat.' is the seizure lateralization - 'r' = right hemisphere, 'l' = left hemisphere, 'b' = bilateral, '-' = impossible to define a lateralization; 'Seiz.' is the number of seizures on the recording; 'Rec. Duration' is the recording duration in hours; 'Samp. Freq.' is the sampling frequency in Hertz.

ID	Gender	Age	Eletrodes	Loc.	Lat.	Seiz.	Rec. Duration (h)	Samp. Freq. (Hz)
70702	m	26	33	f	r	8	168.0	250
1306903	m	42	27	t	l	12	292.02	512
6700	m	35	29	t	r	6	117.52	1024
102	m	36	24	c	l	11	165.41	256
70202	f	29	25	t	l	6	138.0	250
1314803	f	23	27	f	l	9	214.55	512
32702	f	62	25	t	l	6	164.43	256
71202	m	41	25	t	l	8	101.0	250
111902	m	42	28	c	l	8	158.99	256
30802	m	28	25	t	r	9	159.28	256
72602	m	55	24	t	r	9	96.0	250
1235103	m	19	27	t	b	7	173.79	400
71802	f	46	24	-	-	6	72.95	2500
70502	m	18	35	t	l	11	138.0	250
73002	m	32	25	t	l	6	162.12	2500
93902	m	50	24	t	r	9	406.9	256
98202	m	39	28	t	r	9	171.87	256
70302	m	18	33	t	r	14	164.28	2500
1316503	m	36	27	t	r	7	283.91	512

1330203	f	31	27	t	r	6	161.05	400
400	f	36	27	t	l	8	115.54	512
72702	f	56	22	-	-	11	163.0	250
72402	f	36	25	-	-	17	99.0	250
114902	f	16	24	t	l	12	161.64	256
110602	m	56	24	t	r	8	136.76	256
5800	f	26	29	t	b	9	117.25	1024
11002	m	41	22	t	r	8	620.33	256
51002	f	46	24	t	l	9	133.88	256
71102	f	31	25	t	r	8	93.22	2500
70402	f	37	25	t	r	7	185.0	250
53402	m	39	24	t	r	8	95.24	256
8100	m	16	29	t	b	6	189.88	1024
109502	m	50	32	t	r	10	161.12	256
72502	f	60	24	t	l	11	126.0	250
1329503	f	18	27	f	r	7	93.5	400
101702	m	52	24	t	l	6	96.61	256
1314703	m	32	27	t	l	6	59.62	512
7300	m	18	29	-	-	12	118.13	1024
1316403	m	27	27	h	b	11	240.02	512
115102	m	31	24	f	r	15	163.98	256
81102	m	41	24	t	r	13	98.35	256
70602	m	35	33	t	r	8	162.0	250
5900	f	27	29	t	r	7	117.71	1024
1307503	f	24	27	h	r	17	48.02	512
1306003	m	29	27	t	r	7	218.44	512
202	m	24	24	-	l	8	95.45	256
103802	f	17	22	-	-	10	98.0	256
6200	f	37	29	t	l	11	118.66	1024
52302	f	61	24	t	l	7	161.02	256
1234303	m	30	27	t	l	7	234.59	400
1321803	f	27	27	t	r	7	91.01	512
1306203	m	47	27	f	r	14	238.92	512
81402	f	65	32	f	l	15	214.24	256
1325103	f	37	27	t	r	8	284.16	400
7800	m	55	29	t	b	6	118.22	1024
1325903	m	41	27	f	r	9	646.77	400
3600	m	10	29	-	-	28	90.32	1024
26102	m	65	23	t	l	8	134.93	256
6000	m	16	29	t	l	7	97.42	512
1300003	f	22	27	t	r	6	313.37	512
7302	f	52	24	f	l	8	448.24	256
94402	f	37	32	t	r	11	243.72	256
70902	f	28	32	-	-	8	110.95	2500
1312703	f	54	27	t	r	6	70.39	512

1321903	m	41	27	t	l	11	215.25	512
200	f	53	27	t	r	12	120.93	1024
70102	f	46	22	t	r	8	208.46	2500
6500	m	27	27	-	-	19	118.76	1024
2300	m	49	27	t	l	13	117.22	512
112402	f	32	30	f	r	16	164.8	256
80702	f	22	32	t	b	9	115.72	256
70802	f	45	25	-	-	19	168.0	250
1330903	m	36	27	h	l	10	231.66	400
95202	f	50	24	t	l	14	163.04	256
1307103	f	37	27	t	l	10	336.14	512
72302	m	21	37	t	l	28	92.0	250
113902	f	29	24	t	r	25	121.97	256
100002	m	35	24	-	-	24	185.71	256
72202	f	27	25	t	l	33	162.0	250
1328603	m	42	27	t	l	22	191.71	400
72802	f	36	33	-	-	10	95.48	250
71002	f	30	25	t	r	10	164.0	250
6100	f	27	29	t	b	24	107.01	1024
96002	m	58	24	t	r	9	160.39	256
1313003	f	44	27	t	l	8	257.59	512
22602	m	26	24	f	l	22	115.47	256
2900	f	21	27	f	b	21	119.16	1024

TABLE B.2: Patients characteristics for invasive patients. 'ID' is the patient identification; Gender - 'm' = male, 'f' = female; Age is the patient age at the time of the recording; Electrodes is the number of EEG electrodes used during the recording; 'Loc.' is the seizure localization - 'f' = frontal region, 't' = temporal region, 'c' = central region, 'o' = occipital region, 'p' = parietal region, 'h' = complete hemisphere, '-' = impossible to define a cerebral region; 'Lat.' is the seizure lateralization - 'r' = right hemisphere, 'l' = left hemisphere, 'b' = bilateral, '-' = impossible to define a lateralization; 'Seiz.' is the number of seizures on the recording; 'Rec. Duration' is the recording duration in hours; 'Samp. Freq.' is the sampling frequency in Hertz.

ID	Gender	Age	Eletrodes	Loc.	Lat.	Seiz.	Rec. Duration (h)	Samp. Freq (Hz)
1259203	m	30	47	-	l	3	328	400
1308903	f	29	64	t	r	8	475	1024
1310403	m	26	74	t	r	4	503	1024
1235903	f	21	58	t	r	7	472	400
13902	f	53	37	t	l	6	172	512
1311203	f	48	66	t	r	13	503	1024
52303	m	36	40	-	-	6	302	400

107702	f	29	121	t	r	9	185	1024
38402	f	50	94	f	r	15	139	1024
112502	f	11	62	t	r	14	158	1024
97002	m	15	98	t	r	19	204	256
54802	m	17	84	t	l	31	143	1024
1322903	f	23	47	t	r	6	553	400
44202	m	21	70	t	r	22	81	1024
92202	m	39	84	f	l	30	114	1024
1324503	f	43	57	f	r	31	358	1024
62002	m	42	38	t	r	7	249	1024
1272703	f	23	49	t	r	11	153	400

Appendix C

EPILAB

EPILAB is a MATLAB toolbox that enables its users to train and validate seizure prediction models. We have incorporated a LSTM module into the EPILAB neural network section (figure C.1).

With our implementation, users are able to train and evaluate LSTM networks trough the EPILAB toolbox.

The module is available at The framework is available on https://github.com/rodrigomfw/EPILAB

When training LSTMs using our module, the following parameters can be modified :

- Optimization algorithm
- Number of layers
- Neurons per layer
- Sequence length
- Learning rate
- Regularization strength

- Dropout
- Number of epochs
- SOP
- Early detection time - minimum time necessary between an alarm and seizure for the seizure to be classified as predicted.
- Number of classes ([preictal, ictal, postictal, interictal] or [preictal, non-preictal])

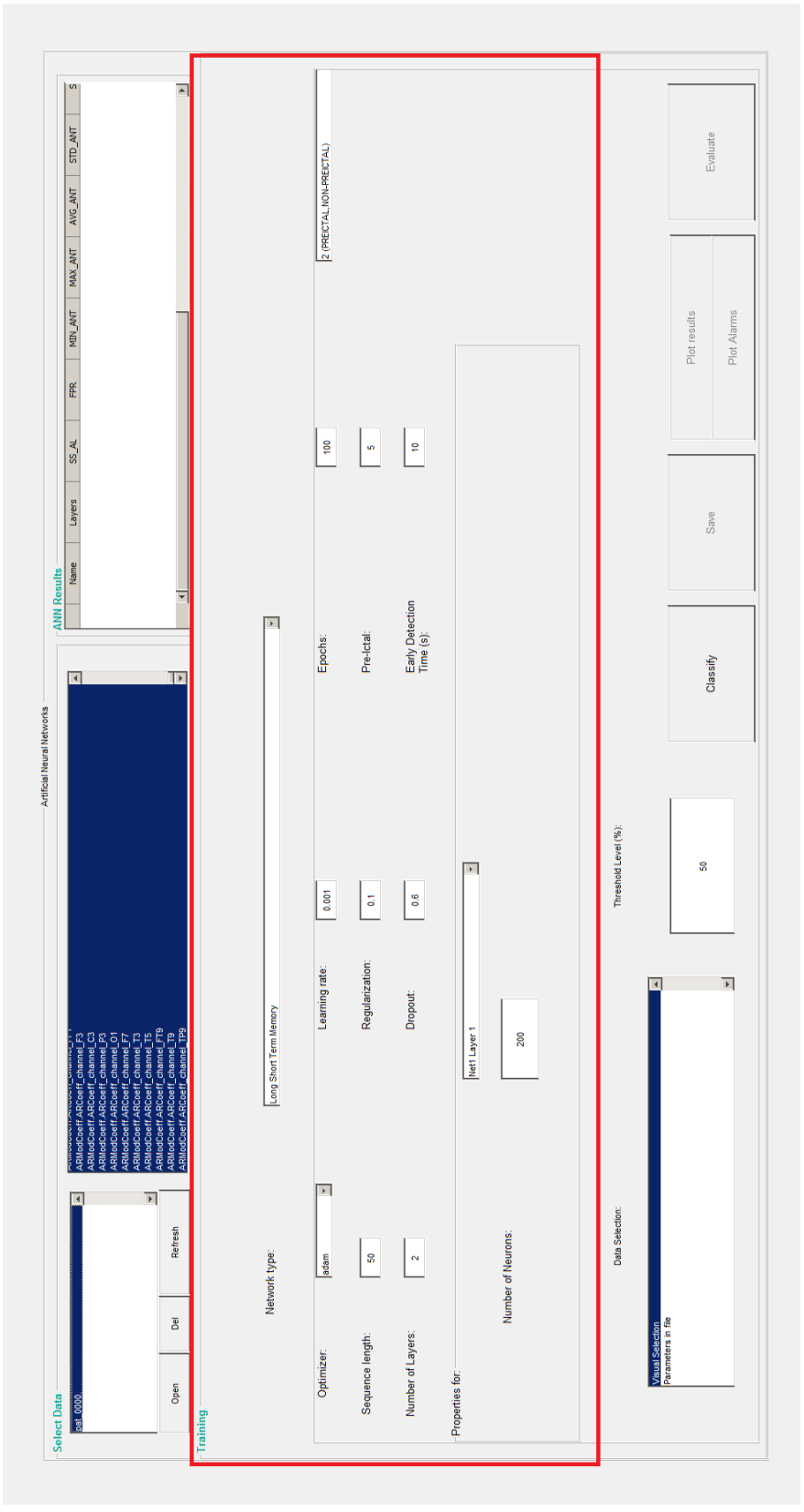

FIGURE C.1: epilab.

Bibliography

[1] Patrick Kwan, Alexis Arzimanoglou, Anne T Berg, Martin J Brodie, W Allen Hauser, Gary Mathern, Solomon L Moshé, Emilio Perucca, Samuel Wiebe, and Jacqueline French. Definition of drug resistant epilepsy: consensus proposal by the ad hoc task force of the ilae commission on therapeutic strategies. *Epilepsia*, 51(6):1069–1077, 2010.

[2] Vivek Nagaraj, Steven Lee, Esther Krook-Magnuson, Ivan Soltesz, Pascal Benquet, Pedro Irazoqui, and Theoden Netoff. The future of seizure prediction and intervention: Closing the loop. *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, 32(3):194, 2015.

[3] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[4] Patrick Doetsch, Michal Kozielski, and Hermann Ney. Fast and robust training of recurrent neural networks for offline handwriting recognition. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 279–284. IEEE, 2014.

[5] Alex Graves. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013.

[6] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*, 2014.

[7] Minh-Thang Luong, Ilya Sutskever, Quoc V Le, Oriol Vinyals, and Wojciech Zaremba. Addressing the rare word problem in neural machine translation. *arXiv preprint arXiv:1410.8206*, 2014.

[8] Hasim Sak, Andrew W Senior, and Françoise Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *Interspeech*, pages 338–342, 2014.

[9] Yuchen Fan, Yao Qian, Feng-Long Xie, and Frank K Soong. Tts synthesis with bidirectional lstm based recurrent neural networks. In *Interspeech*, pages 1964–1968, 2014.

[10] Erik Marchi, Giacomo Ferroni, Florian Eyben, Leonardo Gabrielli, Stefano Squartini, and Bjorn Schuller. Multi-resolution linear prediction based features for audio onset detection with bidirectional lstm neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 2164–2168. IEEE, 2014.

[11] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2625–2634, 2015.

[12] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 2016.

[13] P. Thodoroff, J. Pineau, and A. Lim. Learning robust features using deep learning for automatic seizure detection. *arXiv preprint arXiv*, 1608(00220), 2016.

[14] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. A c-lstm neural network for text classification. *arXiv preprint arXiv:1511.08630*, 2015.

[15] John S Duncan, Josemir W Sander, Sanjay M Sisodiya, and Matthew C Walker. Adult epilepsy. *The Lancet*, 367(9516):1087–1100, 2006.

[16] H Aurlien, IO Gjerde, JH Aarseth, G Eldøen, B Karlsen, H Skeidsvoll, and NE Gilhus. Eeg background activity described by a large computerized database. *Clinical Neurophysiology*, 115(3):665–673, 2004.

[17] SJM Smith. Eeg in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(suppl 2): ii2–ii7, 2005.

[18] Michel Le Van Quyen, Vincent Navarro, Jacques Martinerie, Michel Baulac, and Francisco J Varela. Toward a neurodynamical understanding of ictogenesis. *Epilepsia*, 44(s12):30–43, 2003.

[19] Andrej Karpathy. http://cs231n.github.io/neural-networks-1/.

[20] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533536, 1986.

[21] Christopher Olah. Understanding LSTM Networks. http://colah.github.io/posts/2015-08-Understanding-LSTMs/, 2015.

[22] S. Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116, 1998.

[23] Viglione SS and Walsh GO. Proceedings: Epileptic seizure prediction. *Electroencephalography and clinical neurophysiology*, 39(4):435, 1975.

[24] M. Ihle, H. Feldwisch-Drentrup, C. A. Teixeira, A. Witon, B. Schelter, J. Timmer, and A. Schulze-Bonhage. Epilepsiaea european epilepsy database. *Computer methods and programs in biomedicine*, 106(3):127–138, 2012.

[25] J. Klatt, H. Feldwisch-Drentrup, M. Ihle, V. Navarro, M. Neufang, C. A. Teixeira, C. Adam, M. Valderrama, C. Alvarado-Rojas, A. Witon, and M. Le Van Quyen. The epilepsiae database: An extensive electroencephalography database of epilepsy patients. *Epilepsia*, 53(9):1669–1676, 2012.

[26] J Chris Sackellares, Deng-Shan Shiau, Jose C Principe, Mark CK Yang, Linda K Dance, Wichai Suharitdamrong, Wanpracha Chaovalitwongse, Panos M Pardalos, and Leonidas D Iasemidis. Predictability analysis for an automated seizure prediction algorithm. *Journal of Clinical Neurophysiology*, 23(6):509–520, 2006.

[27] Björn Schelter, Matthias Winterhalder, Thomas Maiwald, Armin Brandt, Ariane Schad, Andreas Schulze-Bonhage, and Jens Timmer. Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(1): 013108, 2006.

[28] P. W. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky. Comparing svm and convolutional networks for epileptic seizure prediction from intracranial eeg. *Machine Learning for Signal Processing*, pages 244–249, 2008.

[29] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio, and F. Fuggetta. Real-time epileptic seizure prediction using ar models and support vector machines. *IEEE Transactions on Biomedical Engineering*, 57 (5):1124–1132, 2010.

[30] Levin Kuhlmann, Dean Freestone, Alan Lai, Anthony N Burkitt, Karen Fuller, David B Grayden, Linda Seiderer, Simon Vogrin, Iven MY Mareels, and Mark J Cook. Patient-specific bivariate-synchrony-based seizure prediction for short prediction horizons. *Epilepsy research*, 91(2):214–231, 2010.

[31] Y. Park, L. Luo, K. K. Parhi, and T. Netoff. Seizure prediction with spectral power of eeg using cost-sensitive support vector machines. *Epilepsia*, 52(10): 1761–1770, 2011.

[32] James R Williamson, Daniel W Bliss, David W Browne, and Jaishree T Narayanan. Seizure prediction using eeg spatiotemporal correlation structure. *Epilepsy & Behavior*, 25(2):230–238, 2012.

[33] Ardalan Aarabi and Bin He. A rule-based seizure prediction method for focal neocortical epilepsy. *Clinical Neurophysiology*, 123(6):1111–1122, 2012.

[34] Shufang Li, Weidong Zhou, Qi Yuan, and Yinxia Liu. Seizure prediction using spike rate of intracranial eeg. *IEEE transactions on neural systems and rehabilitation engineering*, 21(6):880–886, 2013.

[35] Kais Gadhoumi, Jean-Marc Lina, and Jean Gotman. Seizure prediction in patients with mesial temporal lobe epilepsy using eeg measures of state similarity. *Clinical Neurophysiology*, 124(9):1745–1754, 2013.

[36] C. A. Teixeira, B. Direito, M. Bandarabadi, M. L .V. Quyen, M. Valderrama, B. Schelter, A. Schulze-Bonhage, V. Navarro, F. Sales, and A. Dourado. Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients. *Computer methods and programs in biomedicine*, 114(3):324–336, 2013.

[37] Amir Eftekhar, Walid Juffali, Jamil El-Imad, Timothy G Constandinou, and Christofer Toumazou. Ngram-derived pattern recognition for the detection and prediction of epileptic seizures. *PloS one*, 9(6):e96235, 2014.

[38] Yang Zheng, Gang Wang, Kuo Li, Gang Bao, and Jue Wang. Epileptic seizure prediction using phase synchronization based on bivariate empirical mode decomposition. *Clinical Neurophysiology*, 125(6):1104–1111, 2014.

[39] César Alexandre Teixeira, Bruno Direito, Mojtaba Bandarabadi, Michel Le Van Quyen, Mario Valderrama, Bjoern Schelter, Andreas Schulze-Bonhage, Vincent Navarro, Francisco Sales, and António Dourado. Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients. *Computer methods and programs in biomedicine*, 114 (3):324–336, 2014.

[40] B. Direito, C. A. Teixeira, F. Sales, M. Castelo-Branco, and A. Dourado. A realistic seizure prediction study based on multiclass svm. *International Journal of Neural Systems*, 1750006, 2016.

[41] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[42] Klaus Lehnertz, Florian Mormann, Hannes Osterhage, Andy Müller, Jens Prusseit, Anton Chernihovskyi, Matthäus Staniek, Dieter Krug, Stephan Bialonski, and Christian E Elger. State-of-the-art of seizure prediction. *Journal of Clinical Neurophysiology*, 24(2):147–153, 2007.

[43] Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[44] Pooja Rajdev, Matthew P Ward, J Rickus, R Worth, and PP Irazoqui. Real-time seizure prediction from local field potentials using an adaptive wiener algorithm. *Computers in biology and medicine*, 40(1):97–108, 2010.

[45] Florian Mormann, Thomas Kreuz, Christoph Rieke, Ralph G Andrzejak, Alexander Kraskov, Peter David, Christian E Elger, and Klaus Lehnertz. On the predictability of epileptic seizures. *Clinical neurophysiology*, 116(3): 569–587, 2005.

[46] Bo Hjorth. Eeg analysis based on time domain properties. *Electroencephalography and clinical neurophysiology*, 29(3):306–310, 1970.

[47] Yaseen M Arabi, Ahmed A Arifi, Hanan H Balkhy, Hani Najm, Abdulaziz S Aldawood, Alaa Ghabashi, Hassan Hawa, Adel Alothman, Abdulaziz Khaldi, and Basel Al Raiy. Clinical course and outcomes of critically ill patients with middle east respiratory syndrome coronavirus infection. *Annals of internal medicine*, 160(6):389–397, 2014.

[48] Brian Litt, Rosana Esteller, Javier Echauz, Maryann D'Alessandro, Rachel Shor, Thomas Henry, Page Pennell, Charles Epstein, Roy Bakay, Marc Dichter, et al. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron*, 30(1):51–64, 2001.

[49] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7):909–996, 1988.

[50] Ivan Osorio, Mark G Frei, Jon Giftakis, Tom Peters, Jeff Ingram, Mary Turnbull, Michele Herzog, Mark T Rise, Scott Schaffner, Richard A Wennberg, et al. Performance reassessment of a real-time seizure-detection algorithm on long ecog series. *Epilepsia*, 43(12):1522–1535, 2002.

[51] Andreas Schulze-Bonhage, Francisco Sales, Kathrin Wagner, Rute Teotonio, Astrid Carius, Annette Schelle, and Matthias Ihle. Views of patients with epilepsy on seizure prediction devices. *Epilepsy & behavior*, 18(4):388–396, 2010.

[52] Piotr Mirowski, Deepak Madhavan, Yann LeCun, and Ruben Kuzniecky. Classification of patterns of eeg synchronization for seizure prediction. *Clinical neurophysiology*, 120(11):1927–1940, 2009.

[53] Yun Park, Lan Luo, Keshab K Parhi, and Theoden Netoff. Seizure prediction with spectral power of eeg using cost-sensitive support vector machines. *Epilepsia*, 52(10):1761–1770, 2011.

[54] A. Krogh and J. A. Hertz. A simple weight decay can improve generalization. *NIPS*, 4:950–957, 1991.

[55] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, 15(1):1929–1958, 2014.

[56] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.