Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Querétaro
Departamento de Computación y Mecatrónica

Analysis and Classification of Raw Electroencephalograms for Epileptic Seizure Prediction using Long-Short Term Memory Networks

by

Enrique Posada Lozano

Business Solution Development Capstone Project

Ingeniería
en
Sistemas Computacionales

Advisor: Dr. José Antonio Cantoral Ceballos

Santiago de Querétaro, Querétaro, México
24/05/2021

"True prevention is not waiting for bad things to happen, it's preventing things from happening in the first place."

Don McPherson

Abstract

When dealing with epileptic seizures, sleep patterns or other mental disorders, electroencephalogram's (EEG) can provide vital information regarding electrical brain activity. People suffering from epilepsy suffer uncontrolled electrical discharges in the brain, causing uncontrolled body movements, confusion and other symptoms. Therefore, offering the possibility of anticipating and predicting seizures may minimize symptoms or the risks involved. EEG's are the easiest and most economic method for obtaining information form the brain, being cost-efficient and non-invasive procedure. Moreover, an EEG requires extensive knowledge of signal processing in order for medical practitioners to correctly diagnose a patient. Machine Learning, particularly Deep Learning has shown great promise in the analysis of EEG signals due to the capacity learning good feature representation, allowing a model to work on feature extraction instead of having a person doing it, making deep learning a more desired "end-to-end" approach to problems. Anticipating seizures can help warn the patient to take necessary precautions and measures, however raw EEG data can greatly vary between individuals because every person carries distinct bio-metrics. A Long-Short Term Memory Neural Network (LSTM) is proposed for analyzing EEG signals in a predefined period of minutes before an incoming seizure. Recurrent Neural Networks are able of retaining past data, making them suitable for anticipating seizures based on current brain activity. Whereas conventional machine learning process an input through feature extraction in order to then be fed to the model, deep learning can work directly with raw data for feature extraction, in this case for epilepsy detection. For this research, a methodology for segmenting and preparing data using a predefined period of pre-ictal activity and as for deep learning models, an Artificial Neural Network (ANN) is used for testing on raw EEG data as a pre-ictal classifier, and 4 LSTM architectures are used for the task, one being a state-of-the-art LSTM network replicated from literature in the area. Models are tested using different periods of pre-ictal activity of 30 minutes and 10 minutes, each tested with subwindow sizes of 30 and 60 seconds. While 30 second subwindows demonstrate little to no learning, using 60 second subwindows demonstrated that deep learning networks are able to learn and predict from raw EEG up to some extent, reaching an average sensitivity of 62% and a specificity of 72%. Without a pre-processing step, RNN architectures are able to make predictions up to some extent, yet data variability is an issue both in test results and in production because each patient, even though it may have epilepsy, does not indicate the same brain patterns, having an impact on the way the patient can be diagnosed and take early precautions.

Contents

Abbreviations

EEG	**E**lectro**e**ncephalo**g**ram
RNN	**R**ecurrent **N**eural **N**etwork
LSTM	**L**ong - **S**hort **T**erm **M**emory Network
GRU	**G**ated **R**ecurrent **U**nit Network
ML	**M**achine **L**earning
DL	**D**eep **L**earning
CNN	**C**onvolutional **N**eural **N**etwork
ANN	**A**rtificial **N**eural **N**etwork
STFT	**S**hort **T**ime **F**ourier **T**ransformation

Chapter 1

Introduction

> "Doctors can't be expected to instantly recall and correctly apply every piece of clinical information from every study they have ever read. An intelligent system that can assess a patient's data and suggest a potential solution is a logical next step." Greig Millar [1]

Epilepsy is a problem that affects around 50 million people worldwide [2] and is the second most common brain disorder, just behind migraine [3]. Epilepsy is the frequent occurrence of unprovoked seizures, for which a seizure is a sudden aberration in the brain's electrical activity, producing disruptive symptoms such as uncontrolled body movements, confusion, rapid eye blinking, among other symptoms [4]. For all this, electroencephalograms are able to help treat people suffering from epilepsy by monitoring brain activity for medical experts to analyze its data. An electroencephalogram (EEG) is a multi-channel recording reflecting the electrical activity of the brain [4]. Aside from epilepsy, it is used for monitoring sleep patterns, emotion, epilepsy, among other mental disorders [5]. Currently, EEGs are the preferred method for obtaining information from the brain due to the method being the most economic source for brain information, because other methods require some sort of invasive surgery in order to retrieve information, which are also expensive whereas an EEG consists of placing electrodes on the scalp of a person and is more cost-efficient [6]. Each channel from an EEG reflects activity from different regions of the brain [7], so when a non-invasive electrode is placed on the scalp of an individual, it is referred as a scalp EEG, on the other hand, when EEG data is measured using electrodes placed on the surface of the brain, it is referred to as an intracranial EEG [4].

For epilepsy, there are three phases or states that depending on every patient may be easy to see or barely distinguishable at all [8]. The first state is known as the ictal state, which simply is the period of time when a seizure occurs. Then there's the preictal state, being the period of time before a seizure occurs. After reviewing some of the literature in the area, there has yet to be defined a window of time for preictal, meaning there is no known duration of the preictal state. Because of this, the preictal can preictal period could last from minutes up to days [9], so this period is usually used as a warning for an upcoming seizure. Finally, there's the interictal state, which is the period between seizures. However, given that epileptic patients suffer recurrent seizures, the interictal state of the brain can also be understood as the period of "normal" brain activity.

One of the most common cases for using an EEG is for epilepsy, particularly for detecting if a patient has the condition and can also be used for acknowledging and determining the type of seizure a person has [10], thus an EEG can further help a medical expert treat someone with epilepsy. Since an EEG registers brain activity, as such, abnormal patterns can be detected when a seizure occurs.

Normally, people tend to associate a seizure with epilepsy, and this is mostly equivocal due to the simple fact that having a seizure does not indicate having epilepsy [11]. Epilepsy itself is having recurrent seizure episodes whereas a single seizure can occur to anybody, up to 10% of the human population is estimated to experience a seizure on some point of their lives [2]. Seizures can be controlled up to the point of being free of them, however in order to minimize the risk associated by a seizure, predicting an oncoming seizure can lead to taking control and preparing to minimize possible damage that may be damaged by a seizure and medicate oneself for the better [2].

With this in mind, it is really important to note that every person has distinct bio-metrics, whether they are fingerprints, an iris pattern, or their voice. However, brain activity is another unique type of bio-metric, being different among each human. A human brain's electrical activity can differ in each individual for the same activity, whether it is making a body movement or even just thinking about it, making the task more challenging since although there are patterns that individuals may share, it does not standardize or mean it will be the same thing for every other person. This can be further seen in 1.1 and 1.2, and this is shown in order to demonstrate that in some cases, detecting seizures can be something trivial such as 1.1, where brain activity decreases once a seizure starts. But with 1.2, detecting a seizure can become more challenging since the pattern from Patient A 1.1 is not present, but also because brain activity patterns can also be similar to those of a seizure, so being able to recognize the preictal activity can be a more challenging task. Also, EEG data from patients can vary due to a number of reasons, and besides it can also be detrimental to exclude factors caused by noise, sweat, body movement, among others. With a deep learning neural network, the task may be suitable in working with EEG data including up to some extent the noise it can include since deep learning models have the capability of learning and extracting features where conventional ML models would require people to prepare and pre-process the data just so the model can understand the data and find patterns.

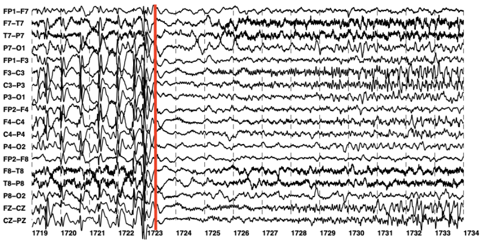

FIGURE 1.1: Seizure EEG of Patient A

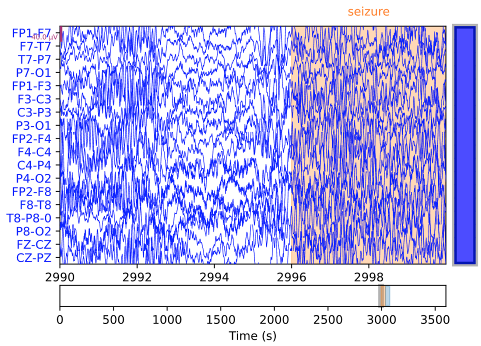

FIGURE 1.2: Seizure EEG from patient B

Chapter 2

State-of-the-Art

This chapter focuses on the fundamental aspects for understanding how and why deep learning is useful in analyzing EEG data for the prediction of seizures by providing an overview of previous research in the area. Understanding previous state-of-the-art models from people like Tsiouris [12], Hussein [13], Bongiorni [14], Usman [15], among others are necessary in order to comprehend what has been done in the past, and how it can then be contributed or improved upon.

Many or most state-of-the-art research on epilepsy are focused on seizure detection rather than seizure prediction. Whereas seizure detection can aid medical experts in a faster diagnosis of epilepsy and from there provide the required treatment for a patient, seizure prediction aims to warn the patient and the doctor before an oncoming seizure within a time defined window in order to reduce the risks involved and apply the required precautions and measures for the patient. As such, medical experts usually analyze EEG data in search of patterns in order to appropriately diagnose the disease.

ML models have been effectively used for the detection of epileptic seizures [6]. However, developing a model that can provide an accurate prediction of an oncoming seizure has remained an unsolved and challenging problem as of now [11]. Seizure detection helps medical experts use an approach to improve diagnostic capabilities. Seizure prediction has the objective of improving a patient's quality of life. Upon inspecting a systematic review on applying deep learning for EEGs, around 16% of the studies conducted on epilepsy are focused on seizure prediction, whereas the rest have focused on seizure detection tasks [5].

Even though seizure prediction has remained a challenge as of now, work in the area has grown over the last years due to availability or easiness to obtain information from the internet, plus the fact that there is now more computational power available that has recently enabled a bigger number of people to start developing deep neural network models for complex tasks. With a deep learning model, tasks that work with the prediction of signals, generating new images, or just making predictions in a more complex manner have proven to be more accurate than a conventional machine learning approach. For the case of seizure prediction, a model can be created in order to receive an EEG signal data, transform the data into a spectrogram and analyze images with a CNN. Another approach would be analyzing EEG signals as a time series, where an RNN is able to focus on that approach with the advantage of "remembering" past inputs, since past information is taken alongside current information in order to make predictions. These are the two most popular approaches for this task in contrast to other types of deep learning networks [5]. Over the last few years there has been an increase in the amount of research on applying deep learning EEGs for epilepsy, having grown more than double over the last couple of years [6]. As mentioned above, CNNs and RNNs are the two most popular approaches to these types of tasks, with CNNs being the most popular approach with a popularity of 44% and RNNs a popularity of 13%.

2.1 Previous Knowledge / Context of the Area

When talking about seizure prediction methodology, as of now there seems to be two distinctive approaches. The first approach consists of selecting a preictal window duration, such as 15, 30, 60, or a given number of minutes as an example and afterwards defining each segment as either preictal or interictal and from there a classifier is used to train to distinguish between the two states. The second approach is focused on threshold methodologies where the analysis is focused on the detection of increasing and decreasing trends in values during the preictal state, so when a value exceeds the threshold established, an alarm is activated in order to inform of an oncoming seizure [11].

Compared to the second approach, one distinguishable advantage from the first one is that this approach can be performed directly on raw EEG signals or even after having done some feature extraction, which aligns with the purpose of this research and is thus the approach chosen to work with [12].

Another point to disclose in this paper is not having to deal with feature extraction. Whereas conventional ML models require pre-processing the data in order for the model to work as desired, deep learning can directly work and learn features while working with the data. However, this comes at the cost of requiring vast amounts of data to work with and is computationally demanding, compared to other types of ML models [16].

2.2 Literature Review

Literature in the area has demonstrated that deep learning models are able to provide better results in contrast to conventional machine learning models. This is because a traditional ML model can't be anticipated without a good generalized model since it requires more time in analyzing and extracting the necessary features, which in turn demands a lot more time working on it whereas DL techniques are more capable of learning patterns with more precision from large amounts of data with the use of multi-layer neural networks [11]. When talking about the use of deep learning for seizure related tasks, [6], there is more research and interest in using CNNs and RNNs for this particular task mostly because they are the most commonly used deep neural networks [6]. Another important point to mention is how the majority of studies have a pre-process step of the data, where about 15% would not perform a pre-processing step such as filtering and would rather work on raw EEG data [5].

Most of the research was focused on two aspects, the first was looking for papers that made similar models such as an LSTM or what is expected to be developed for the research. The other point being for those papers to have worked using the CHB-MIT dataset, since this is one of the best EEG datasets but it also allows to inspect what kind of processing has been made to the data in order to determine an entry point for this paper as for working with data as raw as possible.

For EEG related tasks as well as other medical related tasks using deep learning techniques, relevant metrics to take into account to evaluate a model's performance are accuracy, sensitivity and specificity. While accuracy is a standard measure for a ML performance, having a higher accuracy indicates a model is able to make correct predictions. On the other hand, sensitivity and specificity analysis are more often used for medical diagnoses where sensitivity refers to the true-positive rate of predictions, while sensitivity refers to the false-positive rate of predictions. Sensitivity itself refers to the percentage of test seizures identified correctly [4], measuring how often a correct result is given for people being tested for something like epilepsy. A high sensitivity means that a model or test can diagnose a condition correctly for everyone checking for such a condition [17]. Specificity refers to the number of times over the course of 24 hours where false activity was detected, meaning a prediction was made of an onset seizure in the absence of a real seizure [4]. This means the specificity measures the ability to correctly provide a negative result for people that do not suffer the condition being tested for. This means that having a high specificity will provide correct results for people that do not have the condition being tested for [17].

For this work, the literature or state-of-the-art work can be grouped by their model type, being an RNN (Recurrent Neural Networks), CNN (Convolutional Neural Networks), and a third focused on other types or conventional methods such as DL hybrid architectures, SVM (support vector machine), Decision Trees, Random Forests, among others. Just to note, CNNs are excellent for working with image data, while an RNN is great for working with time-series data as well as audio and speech, which we will delve deeper into below.

1. **RNN**

   In the first group there are recurrent neural network models. In short words, an RNN is an ANN that has an internal memory, acting as a loop in order to work with past output in order to make predictions, also known as the internal state. Different to other types of ANN, RNNs can use the internal state as an ability to remember or work with past input in order to generate an output, meaning the current input worked on takes previous outputs as part of the input to generate a new output. Because of this, RNNs are applicable to work with tasks such as speech recognition, text generation and in itself, time-series data [18]. Having mentioned this, RNNs can be great for analyzing sequence data such as EEG, which could be used for seizure prediction tasks. Whereas past outputs from a neural network are independent from each other, a RNN looks at inputs as being related to one another. However, one important disadvantage from the RNN is the vanishing

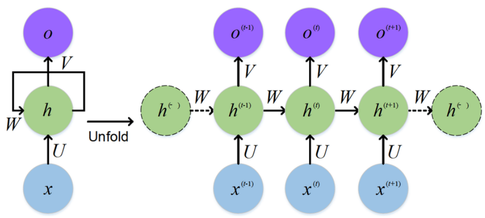

FIGURE 2.1: RNN Architecture [19]

gradient problem, and this affects by not remembering or modifying the initial gradients of the model in a way that it only focuses on the last inputs and when back propagating, the first gradients will have a minimal or no gradient change, resulting in a lack of learning and therefore, a loss in accuracy. To solve this problem, there are two types of RNNs, the LSTM and the GRU networks. These model architectures are good for this task because they remember past input and can work well with signal processing such as audio, stocks, and in overall data provided as a time-series. Because this paper focuses on using an LSTM architecture, it is important to note the three gates that exist in this network, the input gate, the forget gate and the output gate which helps determine what is important from the input, what data can be discarded and not remembered and it then provides an output [18].

Kostas M. Tsiouris proposed a 2-layer LSTM neural network to evaluate seizure prediction with 4 different predefined window lengths of pre-ictal segments of 15, 30, 60 (1 hour) and 120 minutes (2 hours) by using the CHB-MIT Scalp EEG open database [12]. For each window, the segment is subdivided into smaller windows of 5 seconds, which are then passed for feature extraction procedures by taking each 5-second EEG segment which extracts features such as the time-domain, frequency domain, correlation and graph theory. His research delves into using RNNs for the prediction of seizures since at that moment there was little to no literature in the area with this type of model. Having said so, he tests with three different LSTM architectures, each one with a greater number of memory units except the third model, which has added an extra layer of the same dimensionality, using the relu activation function for each layer and having a final dense layer using the softmax activation function for classifying the segment as either pre-ictal or interictal in a one-hot encoded manner. The resulting model is able to predict epileptic seizures with an overall accuracy of over 99% in all defined pre-ictal windows.

For his research, Hazrat Ali employs a Bi-directional LSTM since at the time of writing there was no research in using this type of architecture for seizure prediction [20]. Ali chooses an RNN because EEG data is a time-series, for which a RNN has been successful because it uses past data in order to make predictions, instead of a CNN working with data as images and therefore not being able to capture time dependencies. For his work, Ali used the Kaggle seizure prediction challenge dataset from the American Epilepsy Society, which contains EEG recordings from 7 subjects, 5 being dogs and 2 humans and was sampled at a frequency of 400Hz. Data is divided into 2 sets in order to compare and contrast results and model performance. The LSTM proposed uses a defined 10 minute window for the pre-ictal segments and has a 5 minute interval before a seizure in order to warn a person to take precaution measures 5 minutes before the attack. Afterwards each window is subdivided into 30 second windows in order to provide more data to the neural network. Results demonstrate an average accuracy of 82% for the two sets of data.

Luciano Bongiorni used unprocessed EEG data and proposed two LSTM model architectures for evaluating seizure predictions using different time windows [14] with the CHB-MIT dataset. The goal for his research is to classify 1 second EEG segments as either preictal or as interictal, detecting the preictal stage in order to take appropriate actions to mitigate events from the seizure. Since the sampling rate is of 256Hz, 1 second windows mean 256 samples of data which brings around 3600 batches per EEG recording. In order to prepare data as unprocessed as possible, Bongiorni reduced the number of total channels by 25%, focusing only on the following: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FZ-CZ, CZ-PZ, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8. And since data is limited, in order to obtain more data, windows are overlapped by 50% of the EEG segments taken in order to rely on more data to train on.

As seen on 2.2, spectral analysis using the STFT is then applied to the data in order to obtain spectrograms and then applies a Hanning Window to obtain the overlapped data. Having done so, the spectrogram data is normalized using the MinMax technique in order to preserve a relationship with the original data. The first LSTM architecture proposed consists of 2-LSTM layers, the first one with 30 units and the second with 4 units, and then an output layer with 2 units using the softmax activation function, which calculates the probability of each class for a segment. The second architecture only adds an additional LSTM layer of 30 units, while leaving the rest of the architecture intact. Each of the models uses different defined preictal windows, dividing them into 4 classes. The first one ranges from 5 up to 60 seconds. The second class ranges from 5 up to 120 seconds (2 minutes). The third class has preictal windows of 5 seconds up to 180 seconds. Finally, the fourth class consists of 5 seconds to 240 seconds. From these predefined

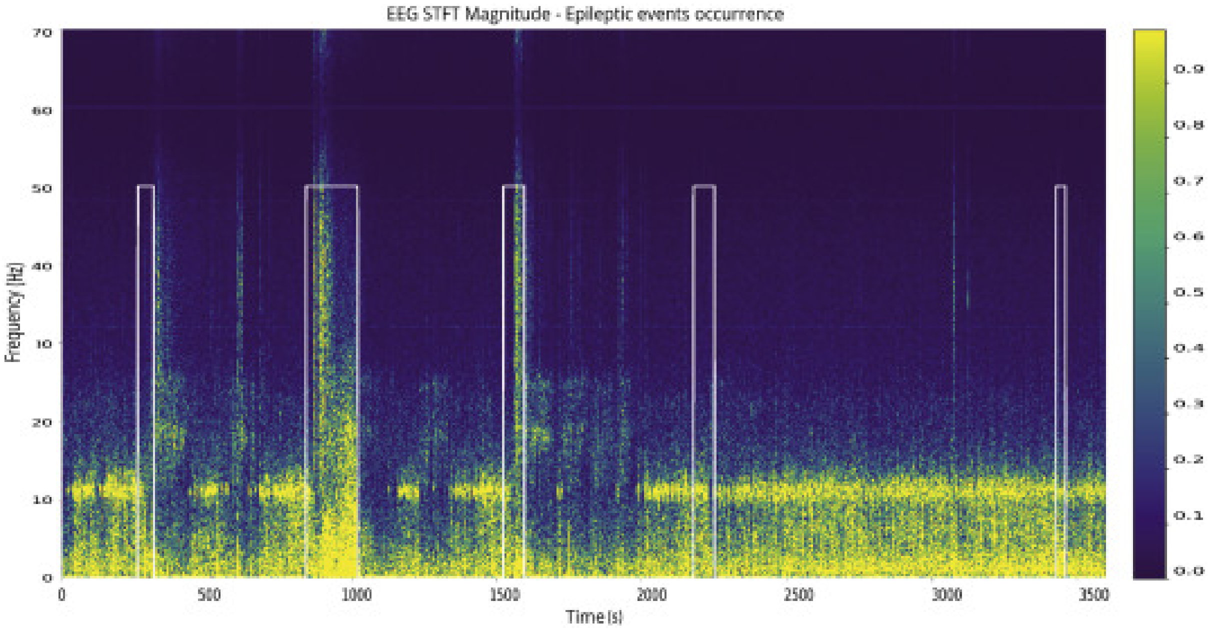

FIGURE 2.2: STFT applied to EEG data containing seizures [14]

windows, the sliding window technique is applied in order to obtain 1 second segments in the ranges mentioned previously. Each of the models anticipated epileptic seizures using different second intervals from 1 second up to 40 seconds of anticipation time, having a low sensitivity whereas it had a really high specificity. Results vary in the different classes defined, but sensitivity had an average of 19% with an average specificity of 98%, the best anticipated time before a seizure was 2 seconds before with a sensitivity of 50% and a specificity of 99%. One of the conclusions from the research was that overlapping data using different window sizes had an impact on model performance, meaning there was an optimal value, in this case of 1 second, as well as channel reduction for an improvement on model accuracy and being less computationally expensive.

2. **CNN**

The second group consists of convolutional neural networks. In contrast to an RNN, this model architecture works best with images and the way it approaches this problem is by preprocessing them by converting it to a spectrogram from which the model analyses the image for patterns. CNNs are considered relevant because they have proven to be very effective with tasks like image recognition and classification [21]. This type of DL model takes images as input and looks for patterns to differentiate images from others. Unlike traditional ML methods that work with images and just like, if not all DL models, CNNs require a lower level of pre-processing data since this NN has the ability to learn certain characteristics when training instead of someone transforming the data for the model to work appropriately [7]. There are two important layers when talking about CNNs, the first is the Convolution Layer, also known as the Kernel and the Pooling Layer. The Convolution Layer applies convolution operations through the image in order to extract the high-level features like edges from an image. So with each added convoluted layer, higher-level features are learned in a similar manner to how we see images. On the other hand, the Pooling Layer takes care of reducing the spatial size of the convolved feature, which reduces the amount of computations made to process data through dimensionality reduction.

Ramy Hussein has a paper focused on using a CNN for an efficient prediction of epileptic seizures using intracranial EGG (iEEG) [13], particularly for people with drug-resistant epilepsy, meaning that treatment is not particularly effective in treating or diminishing seizures. Therefore, he provides a DL model with the intention of warning patients so they can avoid activities that can put them at risk by taking necessary precautions. Hussein's research uses the 2016 Kaggle seizure prediction challenge dataset, containing data from 3 patients with epilepsy. To prepare data for the model, Hussein employs and tests different techniques in order to see how to reduce data while having little to impact on model accuracy. First, Hussein applies principal component analysis (PCA) to EEG data. However, one important point from his research is the importance or effectiveness of applying PCA to iEEG data because normally, PCA would be able to reduce the dimensionality of data, normally used in seizure detection tasks. However, upon testing the results suggest that this method is not reliable for iEEG data reduction. He then reduces the amount of channels in order to only consider data from channels that may have an impact on accuracy. Having said so, the data focuses on extracting 65 minutes of preictal data, then each segment is divided into six 10 minute segments, followed by a 5 min offset before each seizure. Moreover, each segment obtained is separated by a 10 second gap from each other. In a similar manner, the interictal data is segmented by extracting 61 minute long segments that started in a minimum time-span of 3 hours before the seizure, and 4 hours after the

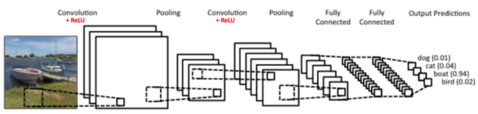

FIGURE 2.3: Simple CNN architecture [21]

seizure occurred, having each segment divided into six 10 minute recordings as well. Since each of the 10 minute recordings is sampled at a frequency rate of 400 Hz, data is resampled at 100Hz in order to diminish data points per second by a factor of 4. The resulting amount of data points per each 10 minute segment is 6,000 data points. Afterwards, each of these downsampled clips are divided into 60 second segments, which are then transformed into images as a spectrogram since a CNN performs best when working with image formats, so he applies a STFT in order to obtain a 2-dimensional representation of the iEEG data which provides a spectrogram for each EEG segment. As for the CNN itself, the model consists of feeding spectrogram images into a combination of CNNs of the same filter size (1x1), and then passes the output to CNNs of a bigger filter size (3x3), which are all then concatenated into a single vector passed to a 3-layered fully connected network, having 2 hidden layers and an output layer with 2 nodes using softmax as the activation function. With the data prepared, the model provided a performance as far as 82% of sensitivity, but when data was resampled at a frequency rate of 200Hz instead of 100Hz, the model performance improved up to a sensitivity of 87%. However, one important conclusion from the research was from working with an imbalanced dataset, having 4,314 samples of interictal data and only 733 samples of preictal data, which then employed a SMOTE (Synthetic Minority Over-sampling Technique) to generate more data from the minority class, because this can indicate that model may be biased towards classifying data as interictal more often than preictal, but applying such technique usually lead the model to overfit the dataset.

Finally, Truong makes use of a CNN by applying a Fast Fourier Transform in order to anticipate epileptic seizures 5 minutes before [22]. Unlike other literature, Truong defines multiple seizures less than 30 minutes apart as a single combined seizure, caring only for the pre-ictal section before consequent seizures. For his work, Truong experiments with three different datasets, two of them containing intracranial EEG data (iEEG) and one with scalp EEGs (sEEG). These datasets are from the Freiburg Hospital, the CHB-MIT dataset and the dataset from the Kaggle competition for epilepsy seizure prediction challenge. Before feeding data into a two-dimensional CNN, data is first preprocessed by converting the raw EEG data into a matrix that gives an image-like format, having the converted data bale of retaining the most important information from the recorded events such as seizure events and data before (the preictal stage). To achieve this, Truong applies a short-term Fourier transformation in order to translate signals into a two-dimensional matrix of frequency along the time axis, and then data is divided into 30 second long windows. Apart from the point mentioned above, frequencies detected in ranges of 50 Hz for the Freiburg Hospital dataset and 60 Hz for the CHB-MIT dataset are removed since those frequencies usually are power-line frequency noise. Aside from sensitivity and specificity, other metrics employed in the paper were SOP and SPH. SOP was defined as the interval from which a seizure is expected to occur while SPH signifies the interval or period between an alarm and the beginning of an SOP. As for results, prediction sensitivity was 81% for Freiburg Hospital dataset, 81% for CHB-MIT dataset and 75% for the Kaggle seizure prediction challenge.

3. **Other ML architectures**

   This third group focuses on other approaches to this problem using traditional/conventional ML models or other types of deep learning models for the task. According to a review of research on seizure prediction using Machine Learning, the majority of research is carried out using SVM models [11]. These ones, although they show some promise and provide a certain degree of accuracy, are not capable of achieving the same degree of accuracy as DL because they are more simple models, or at least less complex than DL models.

   Despite the previous point, among some of the work in the area within this group of models are hybrid architecture DL models.

   Aside from using a CNN or an RNN for tackling the challenge, another approach with deep learning is using a hybrid architecture. A hybrid architecture basically consists of having two or more DL architectures such as a CNN, RNN, ANN, DBN (Deep Belief Network), among others. One important paper using this architecture is a CNN-GRU (CGRNN) with a 35 preictal window proposed by Abir Affes [8]. This hybrid architecture model predicted a seizure 5 to 35 minutes in advance, meaning the preictal window itself is of 30 minutes but starts 35 minutes before the seizure and stops 5 minutes before the oncoming seizure and makes predictions between that time-period. As for the interictal windows, these are defined as those occurring 1 hour before the seizure starts and 1 hour after the seizure ended. Each window is subdivided into smaller segments of 30 seconds and makes use of a sliding window, meaning data is overlapped. Affes proposes this architecture as a way of taking advantage of the best of both RNN and CNN architectures, with a model composed of a CNN and a GRU. This model is formally a Convolutional Gated Recurrent Neural Network, which is presented to predict seizures from extracted features from EEG data using the CHB-MIT EEG dataset. The CNN plays the role of a feature extractor, whereas the GRU network learns from the extracted CNN features seizure patterns in order to make a prediction. In order to feed data into the model, data is prepared and transformed into an image format for the CNN, which receives image data and reduces it into an easier form to process, without losing essential features and passes it to the GRU to make decisions as said before. In the end, this model achieved an accuracy of 75% with a sensitivity of 89%.

As for traditional ML methods.
Syed Muhammad Usman proposed an algorithm for predicting seizures 34 minutes before it can occur by detecting when a patient has entered the preictal state [15]. For his research, Usman used the CHB-MIT Scalp EEG dataset, working only with EEG files that contained seizures. His approach focuses on establishing a threshold on features such as skewness, kurtosis, variance, entropy, maximum value, standard deviation, mobility, complexity, among other features. For the preictal state, Usman observed the values of kurtosis increase during the preictal state whereas the levels of variance decrease just like skewness, mobility and complexity. Therefore, these last features are used for the threshold in order to anticipate a seizure and warn the patient. He then makes use of three different models, a K-Nearest Neighbor (KNN), a Naive Bayes classifier, and a SVM, and compares their results. Out of the three models, the SVM achieves the best with an accuracy of 97%, a sensitivity of 88% and a specificity of 97%, the KNN comes second with an accuracy of 98%, a sensitivity of 80% and a specificity of 99%, and lastly the Naive Bayes classifier had an accuracy of 91% with a sensitivity of 80% and a specificity 91%.

Another SVM was proposed by Yanli Yan, whose model had an average prediction horizon of 61 minutes using the Freiburg Hospital dataset, which contained data from 19 patients with epilepsy [23]. When preparing the data, power-line interference was removed in order to reduce data noise. Yang then proposes using permutation entropy in order to track changes in the human brain's activity. The model was tested using three different sliding windows, a starting window of 5 seconds, a short window of 2 minutes and a long window of 6 minutes which is then fed to the model to classify the segment as interictal or preictal. These windows are used in combined matter in order for each small window to aggregate results and help determine a larger window if a segment is preictal or interictal, meaning a long window of 6 minutes contains 3 segments of 2 minutes, and each 2 minute short window has 24 segments of 5 second sliding windows. After training the model, there is a post-processing phase where the output is transformed into an alarm generator that is raised whenever the output of the model is preictal. So, when at least two short windows are preictal, the alarm is raised. This results in an achieved average sensitivity of 94%.

2.3 Advances in epilepsy seizure prediction

After reviewing some of the most relevant and available literature in the area, the most relevant papers for this research come from Tsiouris [12], Bongiorni [14] and Hussein [13] since they contribute on several aspects that provide insight into anticipating seizures, showing pros and cons of different approaches. On one hand, Tsiouris's research proves that a deep learning architecture is capable of predicting seizures using several predefined windows, meaning the task at hand can be achieved and with the same dataset and same model architectures (LSTM) that is going to be used for this research. Hussein also demonstrates the same point, but with a shorter time-window and focuses on spectrograms with a CNN, trying different techniques to obtain and segment data such as channel reduction, PCA, and resampling. Moreover, his research indicates an approach towards always having classes balanced, since applying oversampling strategies such as SMOTE can impact a models performance, plus his research indicates that using more data can significantly increase accuracy. Finally, Bongiorni's research, although not providing a model with stellar performance, provides great insight into Hussein's point of working with unbalanced datasets, suggests working with a larger dataset, and trying on working with raw EEG data. Moving on, all of the three papers applied techniques such as channel reduction in order to be consistent with results, having Bongiorni and Tsiouris using the same channels. Also, most papers with a good performance approached the problem by having predefined windows of a given number of minutes, instead of using a threshold methodology as mentioned at the beginning of the chapter.

2.4 Conclusion

Seeing some of the most relevant and important research in the area, it can be seen that deep learning models are in fact capable of providing better performant models than using a conventional ML approach, and since deep learning is an end-to-end approach, it means they no longer require a certain degree of pre-processing from the developer, working and functioning adequately with the data it is given. And even though some of the research concludes with not being able to provide satisfactory results, it opens up new opportunities as well as a deeper understanding of what can be done today in order to improve upon what has been done before. And as mentioned previously, there is little to no work available to review yet working on raw EEG data. Moreover, most of the literature contains visible patterns or approaches between one another that from a certain point of view paves a starting point for approaching the problem, such as using a predefined window approach, instead of figuring that sort of knowledge throughout the research. Also, even though some of the model experiments are proven as being successful, all seem to incline towards a heavy use of big amounts of EEG data in order for the model to be capable of actually making correct predictions, which is something common in deep learning. Finally, even though most of the research on epilepsy has been carried out using CNN models, RNNs have been becoming more relevant and popular to use on seizure related tasks over the last decade [5], opening new research opportunities in the area. This could be an indication that these neural network architectures may be able of acting as a better classifier for recognizing preictal brain activity and thus, be able to provide a tool to help patients with epilepsy.

Chapter 3

Problem Statement

Every person is different, and for this case, each individual has distinct bio-metrics, meaning that their brain activity is different from other people, even when doing or even thinking about the same thing. With cases like this, recognizing a seizure can vary among individuals, making it difficult to detect and monitor but even harder to predict in a given time window. Oftentimes this demands a person with an extensive knowledge of the area, and monitoring brain activity can be time consuming, often varying from a couple hours to even days [24] to review EEG recordings. This is where AI tools can help ease the burden of work, handling tasks that would require a complete human intervention or supervision to guarantee results. For medicine, this can mean improving the precision results make, or handling interpretation with both a high level of accuracy while having a performance that allows a fast or even real time prediction of events, such as the anticipation of seizures with 10 minutes in advance.

As mentioned before, EEGs are the most cost-effective way of obtaining information from the brain due to being low cost in comparison to other kinds of procedures and this is also because it is a non-invasive procedure. Other methods used for obtaining brain information require some sort of invasive surgery, and have a risk involved and a greater effort needed for such a task. However, one of the downsides with the use of EEGs are noise, or artifacts that may impact data, and these can be things such as sweat, body movements, sleep, among others.

Having said so, after reviewing some of the most relevant literature in the area, it is to be noted that even though there has been more research in classifying EEG signal data for epilepsy, less than 25% of the research is focused on seizure prediction, because the majority has rather approached seizure detection [5] tasks. Alas, it can be seen that not much of the work is focused on using RNNs for time-domain data like an EEG, where an RNN can be useful because it can remember previous information for future outputs. Furthermore, it is proven that deep learning does not require feature extraction, so little or even no data pre-processing is required for models to work [5]. Because of this, raw data can be used to feed a deep learning model, yet there is little to no research in the area working only with raw data without a pre-processing step.

Many of the papers discussed on the previous sections prepare data by pre-processing it before feeding it into a model, and seizure prediction being a current challenge opens up opportunities for exploring distinct approaches that may or may not provide a better outcome, but it will certainly offer more insight and information in the area of EEG analysis using deep learning architectures.

Chapter 4

Methodology

4.1 EEG Dataset

For this research, work will be done using the CHB-MIT scalp EEG dataset obtained from Physionet [25], which can be seen in the figure below, has recordings from 23 patients, from which two instances come from the same patient in different time periods. Following the work of Kostas M. Tsiouris [12] and Luciano Bongiorni [14], this work will also focus in using the channels "FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ and CZ-PZ" as shown in 4.1, since their work focused on making advances with these channels, and Tsiouris had a great advance in the area, reaching a high accuracy following this data. Each of these channels represents data from an electrode placed on the scalp of each patient, thus the name represents the position the electrode is receiving electrical brain information. All EEG data was recorded using the european data format, or ".edf". Furthermore, choosing these channels is desired because they are mostly consistent throughout most of the recordings of all patients, with the exception of one patient "chb12", who for a couple of files contain none of the channels and patient "chb13" has recording where not all 18 channels are being used, and its data is discarded. As seen on table 4.1, the dataset contains around 980 hours of EEG data and is sampled at a frequency of 256 Hz, meaning 256 data points per second, bringing a total of over 900,000,000 data points per EEG channel, having 16 billion data points as the total for all 18 channels desired in the dataset.

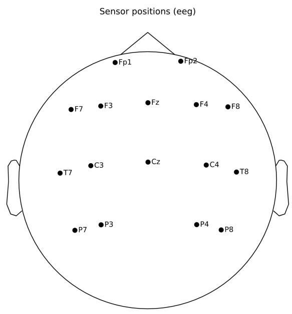

Figure 4.1: Scalp Representation of Desired EEG Channels

Subject	Gender	Age	# of Seizures	Total Ictal Time (seconds)	Duration of EEG Recordings (hours)
chb01/chb21	F	11, 13	11	641	73:22:57
chb02	M	11	3	172	35:15:59
chb03	F	14	7	402	38:00:06
chb04	M	22	4	378	156:03:54
chb05	F	7	5	558	39:00:10
chb06	F	1.5	10	153	66:44:06
chb07	F	14.5	3	325	67:03:08
chb08	M	3.5	5	919	20:00:23
chb09	F	10	4	276	67:52:18
chb10	M	3	7	447	50:01:24
chb11	F	12	3	806	34:47:37
chb12	F	2	27	989	20:41:40
chb13	F	3	12	535	33:00:00
chb14	F	9	8	169	26:00:00
chb15	M	16	20	1992	40:00:36
chb16	F	7	10	84	19:00:00
chb17	F	12	3	293	21:00:24
chb18	F	18	6	317	35:38:05
chb19	F	19	3	236	29:55:46
chb20	F	6	8	294	27:36:06
chb22	F	9	3	204	31:00:11
chb23	F	6	7	424	26:33:30
chb24	-	-	16	511	21:17:47
TOTAL	**-**	**-**	**185**	**11125**	**979:56:07**

Table 4.1: CHB-MIT Dataset summary

The whole EEG dataset consists of 685 files from the 23 (or a given number) patients, which would be working with vast amounts of data, and thus would demand serious computational power and time, which may affect this research by taking more time processing the data instead of modifying and tuning a model for best results. Therefore, the work proposed focuses on working only with EEG files that contain seizures. From there, they way data is separated is by defining a window from which to predict a seizure, which in works like Kostas is testing with segments of 15, 30, 60, and 120 minutes and each segment is then segmented into smaller segments of 5 seconds before doing feature extraction/pre-processing. For this research, the first approach consists of dividing the EEG data into segments of 30 minutes, which for example, if a recording lasts two hours, then there'll be four 30 min segments. Now each segment will be classified as either preictal, meaning before a seizure, or inter-ictal, which is the period between two consequent seizures. The ictal regions are discarded in order to avoid mixing preictal data with seizure data since the point of this research is predicting and classifying when a person has entered the pre-ictal phase and therefore prevent risks associated with being unaware of oncoming seizure and prepare the person with medication or put them away from risk. Doing so allows us to identify a seizure segment, remove ictal information and work only with the data needed. However, one important downside to this approach would be that of overfitting the model due to only working with seizure EEGs. Since the dataset to train must be balanced in order to avoid a model inclining always towards the majority class, in the event of having more preictal data, a small probability such as 10% will be used to extract interictal data from non-seizure files, minimizing the difference between classes and making it more suitable for following a sampling strategy such as undersampling or oversampling. The resulting amount of data is aimed towards using more than 160 hours of recordings, covering files with seizures and using little of non-seizure files for a balanced dataset.

4.2 Time-Series Features

As mentioned before, due to the vast amount of data available, there is not enough computational power to work with all data and thus, the data primarily focused on are those that contain seizures. The most relevant data for the research are the preictal and interictal EEG segments, which are the two classes the model will evaluate and predict in order to anticipate a seizure, by detecting preictal activity. All of the data worked with will remain in a raw format, in order to determine if seizures can be anticipated solely based on raw brain data from a patient. In order to obtain these segments, the approach to follow is by having a predefined window as mentioned in the state-of-the-art section, where each predefined segment is classified as either preictal or interictal and is the data used to train the model. For this, the data is mostly worked on segments of 30 minutes, particularly for the preictal state and also for the interictal segments that are within the same file with seizures. The way this works is by identifying the ictal state, and then having the preictal be all data available within a 30 minute window up to the starting point of the seizure. Then the interictal data is taken by having an offset window of 1 hour from the start of the preictal segment and the end of ictal state, in order to obtain a difference in data since currently there is no specific or proven duration of the preictal phase, which then also takes a time segment of 30 minutes of EEG data. Since each segment will be further divided into smaller segments of 60 seconds (1 minute), this approach aims to get as much information as possible without caring how small an interictal or preictal segment is, since its information can be crucial in distinguishing the preictal state before a seizure. However, it is to be noted that the initial approach consisted of only obtaining consistent 30 minute windows, resulting in less but consistent data windows.

As previously mentioned, after obtaining each segment and having it classified to a label, the data is divided into separate segments of 60 seconds, resulting in an 18x15360 data points per segment. For the use of an RNN architecture, data has to be further divided or reshaped into a given number of timesteps for the model to work with. In this case, the approach followed

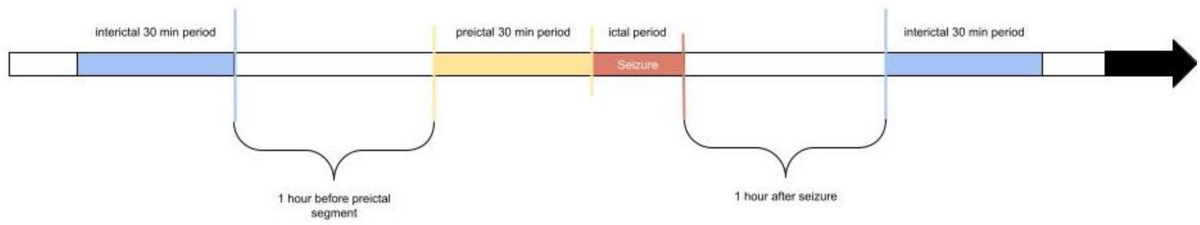

FIGURE 4.2: EEG Segment Analysis

was to have each 60 second segment divided into 12 timesteps, meaning each timestep is of 5 seconds. This is one of two approaches to segmenting data, the other consists of working with the full segment, and then partitioning data into a given number of timesteps. At the end, the dataset used consists of 10,000 segments of 60 seconds each with the desired 18 channels, which is then reshaped into 12, 5 second segments, leaving an array shape of (10000, 12, 23040).

4.3 Seizure Prediction Methodology

With the information established previously, an outline of the methodology can be observed in 4.3, where once the desired EEG segments are obtained using any preictal time window such as 30 min or 10 min, windows are then subdivided into segments of a predefined size, such as 60 seconds or 30 seconds. These segments are the samples which will be given to train the models, and each one is reshaped to a given number of timesteps, for which this paper focuses on using 5 second subsegments as the timestep duration, giving 12 timesteps per sample. The prepared data is then fed into several LSTM architectures including the one from Tsiouris [12], which will be shown in the next chapter 5.

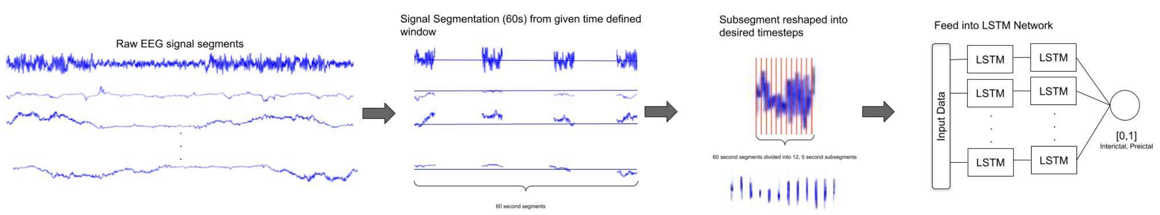

FIGURE 4.3: Representation of EEG Seizure Prediction Methodology

Chapter 5

Proposed Solution

In order to evaluate EEG segments as preictal or interictal, a couple of ANN architectures are employed to observe how precise and good it can anticipate seizures but mostly how well it can provide satisfactory results by working with raw data from several patients. Even though it has already been mentioned, deep learning models are able to work with raw data and provide good results because it is able to extract features, whereas conventional ML architectures would require preparing and pre-processing beforehand for the model to understand the data.

In order to test if raw data can actually be used to train a model and be capable of making predictions, data is fed to a simple ANN, each time becoming a little more complex in order to observe if there is any progress with working with the data and if it can actually train with raw data. The first ANN consists of 3-layers, having 2 hidden layers each 128 nodes using the "relu" activation function, and an output layer with 1 node using the "sigmoid" activation function.

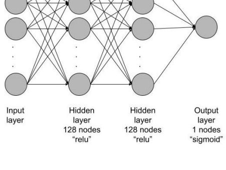

Figure 5.1: Initial ANN architecture

When it comes to RNN architectures, there are different types of an RNN, these being one-to-one, one-to-many, many-to-one and many-to-many. For this task, we feed the subsegments of a given window to the RNN in order to classify whether a window is pre-ictal or not, meaning the suitable approach is a many-to-one RNN. That being said, the RNN architecture chosen for this paper is the LSTM, a more advanced RNN employed in the research of Tsiouris [12], first proposed by Hochreiter & Schmidhuber in 1997 [26]. It is important to mention that the original proposed apporach was to start using a simple RNN and then move into using a more advanced variant such as the GRU or the LSTM. However, because RNNs suffer from the exploding gradients problem, not being able to modify earlier layers weight might have a negative impact on results, especially if the model is planned to work soley on raw EEG data from patients. So from using an ANN to test data and validate that training can be done on a model without underfitting, the proposed solution is by using LSTM architectures to anticipate seizures using different preictal time windows.

With this said, 3 models are created and tested, being contrasted with one another and using the architecture of Tsiouris [12] as the main point of reference, comparison and improvement between models. The first model is comprised of a 1-layered LSTM, with an output layer containing 1 neuron to predict if the segment is preictal or not (interictal). The second model consists of a 2-layer LSTM, with the same output layer of a single neuron. Finally, the third model consists of a 3-layer LSTM with a single output neuron just like the other model architectures.

Moving on, Tsiouris's approach to anticipating seizures consists of having 2 LSTM layers, each followed by a dropout layer with a rate varying between 0.2 and 0.5, followed by a fully connected layer with 32 units and an output layer of 2 nodes using the "softmax" activation function. Besides the RNN, his LSTM architecture is employed and contrasted with a similar proposed LSTM architecture, which consists of 2 LSTM layers, one with 128 nodes and the second layer with half the units, then it employs only one dropout layer with a 0.5 dropout rate because the research from Tsiouris suggests an intermediate dropout layer does not impact accuracy, and an output layer of 1 node using the "sigmoid" activation function.

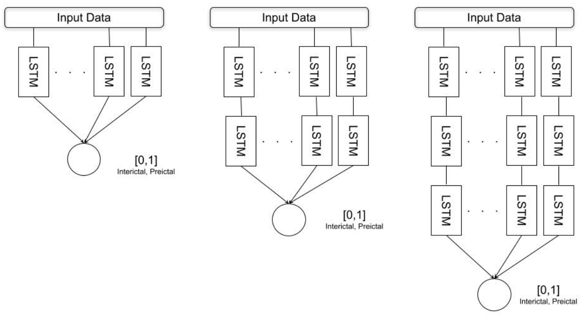

FIGURE 5.2: LSTM architectures

Chapter 6

Results

6.1 Initial Approach

Before using RNN networks, the first approach was testing data using an ANN. For this, features are flattened into a 1-dimensional array which is then fed to a neural network with 3-layers composed of 2 hidden layers of 128 nodes using "relu" as the activation function and an output layer with 1 node using the "sigmoid" activation function. This initial model with raw data under delivers results by overfitting by a big margin, up to the point that it "memorizes" data but even when training with 90% of the dataset, the remaining 10% used for testing provides a low accuracy of 50%, indicating that the model besides having been overfitted, seems to not understand data enough to make predictions, rather making guesses of which classes are preictal and which are interictal. Having a more complex NN by expanding the number of layers from 2 to 5 does not indicate a development in progress. However, plotting predictions using a confusion matrix indicates that the model seems to learn patterns for preictal data, but fails to do so for interictal data, predicting data as interictal when it should have been preictal (True Negative). Therefore, this initial approach demonstrated that a neural network is capable of learning with the raw data and shows promising results. However, the model seemed to be overfitting the training data and due to an imbalance in data, the model seems to be biased towards one class. Furthermore, just as using neural networks for other tasks, increasing the number of layers makes the model more complex, but it can bring better results. In the case of this research, improvement was minimal but it was enough for indicating that a different or more suitable neural network architecture like an RNN might be able to work better with raw EEG data for predicting 60 second windows as preictal or interictal.

The first implementation of an RNN consisted of a 2-layer LSTM, each with 128 nodes, with an output layer of 1 node using the “sigmoid” as activation function to classify a segment as pre ictal or interictal. Even though results demonstrate a high sensitivity, this means it can learn and recognize EEG activity as preictal, yet it fails to properly acknowledge when a segment is not a seizure and is within the interictal or normal state of the brain. Since data was unbalanced, results indicate the model was biased towards classifying most segments as preictal. However, when the same LSTM architecture is tested with a balanced dataset, the model’s performance increases, having an overall accuracy around 60% with having only 700 samples to train with and 80 samples to test with. Moreover, one important thing from this approach was that in order to balance the dataset, data was balanced by either undersampling, or oversampling. This experiment undersampled the class with the least amount of segments, reducing the amount of samples to train with to 700 minutes, leaving 80 minutes to test. Then, the experiment was repeated using dropout as a regularization technique, but results did not improve. For these results, data was undersampled in order to balance the classes at the cost of having less data to train and test with.

Next up was replicating the architecture used by Tsiouris, having a 2-layer LSTM of 128 units, each followed by a dropout layer, then a fully connected layer of 32 nodes, and an output layer of 2 nodes using the softmax activation function. The reason as to why results are not so effective can be because of two reasons, the first one might be that due to a high variance, the model may need more data since regularization techniques does not seem to have much of an impact on model accuracy. The second reason might be the data does need to be preprocessed in order to deliver results.

Model	Sensitivity	Specificity
2-Layer LSTM	84%	40%
3-Layer LSTM	60%	59%
Tsiouris [12] LSTM	60%	64%

Table 6.1: Results using a 30 min preictal period, initial approach and testing

The results seen in 6.1 demonstrate a poor test performance, and since training accuracy was ¿90%, performance results may indicate overfitting, which may be due to the simple fact of having neural networks learn from raw EEG data with little to no pre-processing done. This may be due due to the variability that exists between each persons raw brain activity. The second point may mean that the models may need to train with more data in order to learn better patterns and improve model accuracy. Also, many data segments are lost with the previous approach mainly because data was only taken if it was a 30 minute window, being discarded if the window was of a smaller size.

To improve the model, data was segmented using all the data available. With this approach, the model seems to have increased its performance substantially both with the validation and test datasets, even reaching an accuracy above 80% 6.1. However, sensitivity and specificity was poor because data was not balanced and it had 4 times more preictal data than interictal, leading to the model mostly predicting preictal rather than being more generalized.

6.2 30 min. preictal period

6.2.1 60 second subwindows

After following both experiments, now the most suitable approach was to balance the dataset without or minimizing as much as possible having to undersample or oversample in order to balance the dataset. Since with a 30 minute window there are only around 4000 segments, the way interictal data is increased and balanced is by randomly selecting EEG data from non-seizure files, and by using a 10% possibility of using the data, the dataset obtains around 6000 interictal segments, with 4000 of preictal which can both be oversampled or undersampled without impacting a models accuracy by having little amounts of data. To keep results as "real" as possible, the majority class is undersampled, leaving 2 classes each with 4300 segments, leaving a total of 8600 segments and over 2 billion data points to train and test with. As far as splitting the dataset, the training consisted of 90% of the dataset, having 5% for validation and 5% for testing. After obtaining each set of data, the training dataset is undersampled as well as the test dataset in order to balance classes and when testing, observe how well the models are at classifying EEG segments. The first experiment consisted of a 1-layer LSTM with an output layer containing 1 node using "sigmoid" as activation function just like previous experiments. The model achieves an accuracy of 62%, a sensitivity of 60% and a specificity of 64% while having a training accuracy of 92%.

The next experiment aims to see if a model can perform better when the network is focused on depth rather than height, and in order to do so, this model consists of 3-LSTM layers each with 32 units, and an output layer with 1 neuron.

Model	Model Accuracy	Sensitivity	Specificity
1-Layer LSTM	62%	60%	64%
3-Layer LSTM	60%	60%	59%
Replication of Tsiouris LSTM	62%	60%	64%
3-Layer ANN	69%	71%	68%

Table 6.2: Results using a 30 min preictal period with 60 second subwindows

The last experiment was replicating the LSTM architecture of Tsiouris. With a training accuracy of 92%, the model performance is similar as before, with an accuracy of 62%, a sensitivity of 60% and a specificity of 64%.

Lastly, a side experiment made with this approach was using only data from one channel in particular. In this case, the models used were the 1-layer LSTM, the 3-layer LSTM and the Tsiouris architecture. After training for over 60 epochs for each model, it was clear the models were not learning and the models were underfitting with an accuracy no more than 50% and a constant loss of 0.69. Therefore, even though some channels might contain higher quality data, it is the conglomerate of multiple channels that are able to provide the required data to train a ML model, since the chosen channels are distributed across the brain as seen on 4.1, and are able of providing data from seizures occurring in different parts of the brain across the 23 patients of the dataset.

Results in 6.2 are not satisfactory in either of the models, even though they are similar, experimentation was stopped using this approach since perhaps fine tuning the model could at most improve by 10%, yet the models would still be unsatisfactory for the task. One takeaway from these three experiments was the similarity in test results. Even though from 512 minutes of testing data, 60% are accurate, it is indication that the model is actually learning, yet perhaps the approach of having 1 minute windows divided in 12 timesteps of 5 seconds might not be what's needed. This is a clear indication that all 3 models are overfitting, but even then since data is in the "most desirable" state in which it is balanced, more data would require some sort of greatly undersampling or even oversampling data, which could misguide the accuracy of models. Moving on, if one model stands out more than others, it is the ANN shown at the end of 6.2, having the best results than the LSTM architectures. Even though the ANN overfitted in a small number of epochs, it was able of having similar results with the validation and test datasets, meaning that even though models are not able to provide a satisfactory rate of predictions, they are able to predict from raw data up to some extent, and this includes the

LSTM architectures. Also, this ANN architecture shows more promise than the initial approach, meaning that the initial way data was obtained was hurting model performance.

Another point to note is how easily models are overfitting, which can be seen by plotting loss and accuracy from each of the datasets. Everything can be seen going from a normal, progressive state to starting to overfit and having a gap increase difference between validation and training. An example of this is shown in **??**, showing how the model overfits quickly by diverging from the train validation, but model accuracy increased steadily up to a certain point.

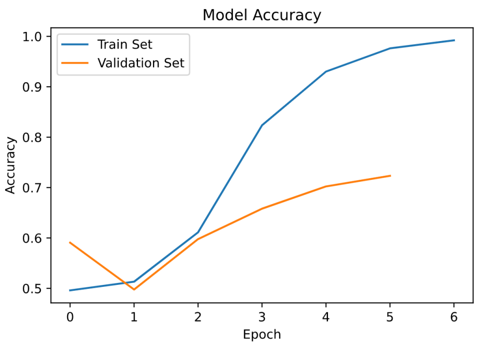

FIGURE 6.1: Model Accuracy

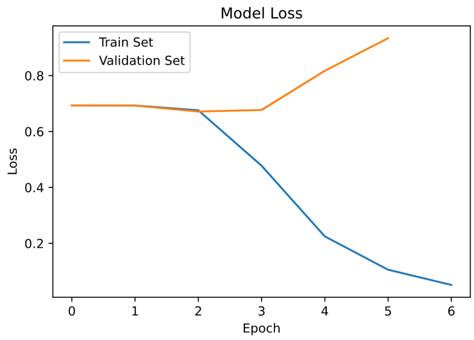

FIGURE 6.2: Model Loss

FIGURE 6.3: 3-Layer ANN performance

6.2.2 30 second subwindows

Using the same preictal period, data was subdivided into smaller segments of 30 seconds, which when having a second size for the timestep, gives 6 timesteps per sample, half of what it was in the previous experiment. Subdividing these windows into subsegments of 30 seconds did not provide better results, showing that even though model architecture was intact, all models including Tsiouris showed underfitting, and learning little to nothing over the span of 15-20 training epochs. Even though the amount of samples was doubled, results demonstrate that decreasing the subwindow size could have an impact on model performance. Further seen on figure **??**, the best performance was obtained from a 32 unit, 1-layer LSTM with an accuracy of 60%, a sensitivity of 61% and a specificity of 59%, while Tsiouris was biased towards predicting all data as interictal even though the dataset was balanced 50-50.

Another observation to note is the difference both in performance and complexity of model depth and height. Since results are worse, a more complex model or even deeper one with more layers does not guarantee accuracy.

Model	Model Accuracy	Sensitivity	Specificity
1-Layer LSTM	60%	61%	59%
2-Layer LSTM	57%	63%	55%
3-Layer LSTM	50%	0%	50%
Replication of Tsiouris LSTM	50%	0%	50%

Table 6.3: Results using a 30 min preictal period with 30 second subwindows

6.3 10 min. preictal period

6.3.1 60 second subwindows

With a 10 minute preictal window, in order to avoid losing as minimal data from both preictal or interictal data, the resulting dataset has a smaller size of 2500 segments, each consisting of 60 second windows partitioned into 12 timesteps of 5 seconds each one.

Using the 1-layer LSTM with an LSTM layer of 32 neurons, without using regularization, the experiment was repeated 2 times, one using sigmoid as the activation function chose while the second experiment used tanh. Using tanh, the model achieved an accuracy of 58%, a sensitivity of 59% and a specificity of 58%. In comparison, when using sigmoid the model apparently diverged faster, overfitting in less epochs than before. Early stopping was used to detect this, by having a patience of 8 epochs. However, one interesting point to highlight was the performance the 1-layer LSTM had on the test dataset. Even though its validation accuracy was low, it achieved an accuracy of 71%, a sensitivity of 72% and specificity of 70%, which although being good, does not validate the models performance.

Moving into a 2-layer LSTM, the network was tested using a first layer of 64 units, a second layer of 32 units, an output of 1 single neuron with "sigmoid" as the activation function. Then the experiment was repeated using a recurrent dropout rate of 0.2 on the second LSTM layer. Surprisingly, using recurrent dropout in an intermediate layer resulted in improving the models performance. Having trained for 30 epochs, the model just reached a a 64% training accuracy, having had a performance accuracy of 65%, a sensitivity of 61%, and a specificity of 70% on test data. Seen in the following figures, model accuracy and loss did not diverge much from one another. This could mean that the model assimilated and generalized much better than previous experiments. But upon training for more epochs, the model eventually diverges and overfits, suggesting this approach, although good might require fine tuning to improve results.

Model	Model Accuracy	Sensitivity	Specificity
1-Layer LSTM	71%	72%	70%
2-Layer LSTM	67%	61%	91%
2-Layer LSTM w/ recurrent dropout	65%	61%	71%
3-Layer LSTM	64%	59%	80%
3-Layer LSTM w/ recurrent dropout & dropout	69%	67%	72%
3-Layer LSTM w/ recurrent dropout & dropout + 128 unit Fully-Connected Layer	62%	58%	75%
Replication of Tsiouris LSTM	67%	62%	87%

TABLE 6.4: Results using a 10 min preictal period with 60 second subwindows

Repeating the experiment using 0.4 as the dropout rate for 60 epochs, showed a slower training and results were worse due to the model deviating and overfitting.

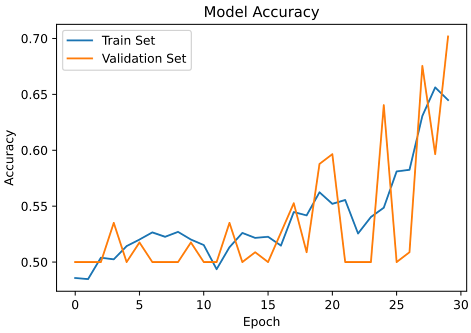

FIGURE 6.4: Model Accuracy

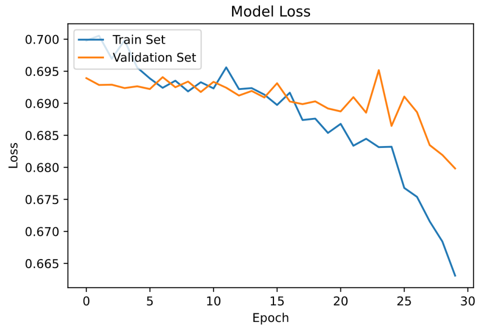

FIGURE 6.5: Model Loss

FIGURE 6.6: 2-Layer LSTM performance using recurrent dropout

Lastly, the next experiment was with a 3-layer LSTM, each with 64 units and a single node in the output layer using "sigmoid" as its activation function. Because of the results obtained with the 2-layer model, the 3-layer LSTM was tested without regularization and then again, using regular dropout and recurrent dropout, both with dropout rate of 0.2, reaching an accuracy of 69%, a sensitivity of 67% and a specificity of 72%. The meaning behind these results show that even though overfitting is a present problem in the research, raw data can be handled up to some extent to make accurate predictions from raw EEG data, while also showing that depth and height are two factors that can impact performance. In this case, with the exclusion of the 1-layer model because of its performant results not being near similar to validation results, the 2-layer model and 3-layer model without regularization techniques show differences both in consumption of resources through training time and model accuracy.

6.3.2 30 second subwindows

When working with the same periodo of 10 minutes of preictal, using 30 second windows once again showed to give unsatisfactory results. As seen on 6.5, results are similar to those on the

30 minute preictal period, having models completely going to predict one class, and having an accuracy no greater than 60%. Having said so, results from 30 second windows are consistently worse than working with bigger, 60 second windows.

Model	Model Accuracy	Sensitivity	Specificity
1-Layer LSTM	60%	61%	59%
2-Layer LSTM	57%	63%	55%
3-Layer LSTM	50%	0%	50%
Replication of Tsiouris LSTM	50%	0%	50%

TABLE 6.5: Results using a 10 min preictal period with 30 second subwindows

6.4 Experiment Results

Model	Model Accuracy	Sensitivity	Specificity
30 min, 60 sec			
1-Layer LSTM	62%	60%	64%
3-Layer LSTM	60%	60%	59%
Replication of Tsiouris LSTM	62%	60%	64%
3-layer ANN	**69%**	**71%**	**68%**
30 min, 30 sec			
1-Layer LSTM	**60%**	**61%**	**59%**
2-Layer LSTM	57%	63%	55%
3-Layer LSTM	50%	0%	50%
Replication of Tsiouris LSTM	50%	0%	50%
10 min, 60 sec			
1-Layer LSTM	71%	72%	70%
2-Layer LSTM	67%	61%	91%
2-Layer LSTM w/ recurrent dropout	65%	61%	71%
3-Layer LSTM	64%	59%	80%
3-Layer LSTM w/ recurrent dropout & dropout	**69%**	**67%**	**72%**
3-Layer LSTM w/ recurrent dropout & dropout + 128 unit Fully-Connected Layer	62%	58%	75%
Replication of Tsiouris LSTM	67%	62%	87%
10 min, 30 sec			
1-Layer LSTM	56%	54%	60%
2-Layer LSTM	**58%**	**62%**	**57%**
3-Layer LSTM	50%	0%	50%
Replication of Tsiouris LSTM	50%	50%	0%

TABLE 6.6: Results from different time windows

6.5 Discussion

When working with raw EEG data, one of the important factors that affected model performance was overfitting. Over the course of all experiments, the models proposed including the one from Tsiouris showed satisfactory results up to some extent, before completely diverging from the validation dataset results. Using regularization techniques such as dropout, recurrent dropout or both proved to have a positive effect on model performance, these 2 types of regularization showed up to some extent more effectiveness than not applying regularization to the model, results would improve less than 10%. Moreover, other regularization techniques were tested like layer normalization, L2 and L1L2, yet its results did not indicate that the model was learning, so these techniques were not focused or effective for this task.

Another concerning issue affecting model performance is the imbalance of data in the CHB-MIT dataset. When mentioning an imbalance, this in no way means that the quality of data is bad or poor, since the data itself has shown its effectiveness for training ML models and other tasks in much of the literature of the area. However, the issue with ML models in classification tasks is that by having an imbalanced dataset, this means there is more data of one class than there is of the other class, which when training can make the model become biased towards one class, which for this case can be particularly shown in its sensitivity and specificity metrics. So classes are balanced in a way that having 50% of one class and 50% of the other helps the model work with an equal amount of data. Moreover, epileptic seizures is a minor event, meaning that for some patients, there'll be a small amount of seizures whereas some patients may experience a greater or more frequent amount of seizures. This was the initial reasoning towards working with an unbalanced dataset in the initial approach, thinking perhaps with deep neural networks the models may be able of accurately classifying by having more interictal data than preictal, since out of 980 hours of recordings, only 3 hours are ictal data, which in turn affects the amount of preictal data available to work with, which also depends on the size of the predefined windows. For example, with a 30 minute preictal period, the dataset for the model to train with was approximately 160 hours long out of the 980 hour total.

Another important aspect to consider when segmenting data is the subwindow size, where reducing the window size affects model performance more often by underfitting or even in many cases not learning as seen on model results for 30 seconds as the subwindow size 6.6. Also, while this reduced the amount of data by splitting the subwindow size from 60 to 30 seconds, one consistent issue throughout the experiments were memory allocation issues for a given preictal period. Because the amount of data to work with is big, the total amount of features would be around 20 GB for a preictal period of 30 minutes. Segmenting and other operations that prepare the data for training would often cause memory issues, which would require the process to be restarted since there would be no memory left to work with, and for all experiments the total amount of data points would often be over a billion. Due to the limited time to work on the research, perhaps having a greater preictal period and a smaller amount of minutes to compensate in size could improve model performance, yet it is necessary to work with the most data as possible for better results.

When observing the impact network height and depth can have for model performance, results from using recurrent dropout on a 2-layer LSTM on 6.4 and the 3-layer LSTM on 6.4 are able to show that height and depth are 2 important factors both in performance of the model and in the consumption of computer resources, with the 2-layer LSTM needing less time to train while the 3-layer model took longer to train, yet the 3-layer LSTM did not provide much better performance, and having both models stop early to avoid overfitting. It is reasonable to think that having a bigger number of units in a layer can be better than having more layers because of a greater number of nodes per layer can help deal with more complex data such as raw EEGs while also taking less time in comparison to having more layers that just bring more complexity to the model [27].

With all this, results suggest that data may need to have a pre-processing step for an RNN architecture to provide more accurate results, like the research from Tsiouris [12]. Because the data worked on is in a raw format, data variability impacted model performance because EEG activity is different in every patient. So having a distinct or radically different brain activity can complicate ML methods both in a seizure state and non-seizure state [4]. Because by working over raw data, a ML model may not be able to generalize well and predict seizures because of a person's distinct brain activity. Therefore, applying a pre-processing step would improve results by generalizing the EEG data of patients, but at the cost of not having real-time or faster predictions when in production since more work is applied to data.

Chapter 7

Conclusions

With the results from the previous chapter, using an LSTM architecture alone without a pre-processing step on raw data is not able to fully provide accurate predictions to be used on patients. However, models were able to learn and provide good results up to some extent, perhaps meaning that DL models are able to work with the raw data successfully, yet an RNN alone is not a suitable approach for the task. The LSTM architectures showed that a pre-processing step is needed for an improvement in sensitivity and specificity, so approaching the problem using the same methodology but with a hybrid architecture could be used in order to extract relevant information and then take advantage of the LSTM architecture to use past information to anticipate seizures solely on raw data. Another suitable approach to see the extent to which the proposed solutions' effectiveness is by comparing results using different dataset, such as the Kaggle seizure prediction challenge dataset, Freiburg dataset, Bonn University dataset, TUH EEG dataset, among others and see if model performance increases or only works with data under specific conditions or taken in a specific format.

As mentioned consistently throughout the research, data variability is a concerning issue for raw data because each person has distinct biometrics. Since ML models need to generalize well in order to provide good predictions, DL architectures with the exclusion of a hybrid approach would need a pre-processing step in order to prepare and transform data into something the model can work better with. Even though results are not as satisfactory, the methodology used offers a method for segmenting and preparing EEG data using a predefined period approach and seeking to balance the datasets for the models to work with, both for ANNs and for a given number of timesteps for RNN architectures.

Finally, even though seizures are minor events in brain activity, anticipating seizures requires always a vast amount of data and data to be balanced in order to accurately distinguish one class from the other, in this case the precital state from the interictal state. But designing the models architecture would first need a focus on height rather than depth, using depth to add more complexity to the model and compare results, finding an optimal point from which to improve upon.

Bibliography

[1] Greig Millar. Artificial Intelligence in Healthcare | Spark Blog, November 2016.

[2] WHO. Epilepsy, 202. publisher: WHO.

[3] Ramy Hussein, Hamid Palangi, Rabab Ward, and Z. Jane Wang. Epileptic Seizure Detection: A Deep Learning Approach. *arXiv:1803.09848 [eess]*, March 2018. arXiv: 1803.09848.

[4] Ali Hossam Shoeb. *Application of machine learning to epileptic seizure onset detection and treatment.* Thesis, Massachusetts Institute of Technology, 2009.

[5] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5):051001, August 2019.

[6] Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of Neural Engineering*, 16(3):031001, June 2019.

[7] Sumit Saha. A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way, December 2018.

[8] Abir Affes, Afef Mdhaffar, Chahnez Triki, Mohamed Jmaiel, and Bernd Freisleben. A Convolutional Gated Recurrent Neural Network for Epileptic Seizure Prediction. In José Pagán, Mounir Mokhtari, Hamdi Aloulou, Bessam Abdulrazak, and María Fernanda Cabrera, editors, *How AI Impacts Urban Living and Public Health*, Lecture Notes in Computer Science, pages 85–96, Cham, 2019. Springer International Publishing.

[9] Epilepsy Foundation. What you need to know about different types and different stages of seizures, 2009.

[10] S. J. M. Smith. EEG in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(suppl 2):ii2–ii7, June 2005.

[11] Khansa Rasheed, Adnan Qayyum, Junaid Qadir, Shobi Sivathamboo, Patrick Kwan, Levin Kuhlmann, Terence O'Brien, and Adeel Razi. Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review. *arXiv:2002.01925 [cs, eess, q-bio]*, February 2020. arXiv: 2002.01925.

[12] Kostas M. Tsiouris, Vasileios C. Pezoulas, Michalis Zervakis, Spiros Konitsiotis, Dimitrios D. Koutsouris, and Dimitrios I. Fotiadis. A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals. *Computers in Biology and Medicine*, 99:24–37, August 2018.

[13] Ramy Hussein, Mohamed Osama Ahmed, Rabab Ward, Z. Jane Wang, Levin Kuhlmann, and Yi Guo. Human Intracranial EEG Quantitative Analysis and Automatic Feature Learning for Epileptic Seizure Prediction. *arXiv:1904.03603 [cs, q-bio]*, April 2019. arXiv: 1904.03603.

[14] Luciano Bongiorni and Alexandre Balbinot. Evaluation of recurrent neural networks as epileptic seizure predictor. *Array*, 8:100038, December 2020.

[15] Syed Muhammad Usman. Efficient Prediction and Classification of Epileptic Seizures Using EEG Data Based on Univariate Linear Features. *Journal of Computers*, pages 616–621, 2018.

[16] How Deep Learning is changing machine learning AI in EEG data processing, April 2020.

[17] Understanding medical tests: sensitivity, specificity, and positive predictive value. publisher: HealthNewsReview.

[18] Aditi Mittal. Understanding RNN and LSTM, October 2019.

[19] Weijiang Feng, Naiyang Guan, Yuan Li, Xiang Zhang, and Zhigang Luo. Audio visual speech recognition with multimodal recurrent neural networks. pages 681–688, 05 2017.

[20] Hazrat Ali, Feroz Karim, Junaid Javed Qureshi, Adnan Omer Abuassba, and Mohammad Farhad Bulbul. Seizure Prediction Using Bidirectional LSTM. *arXiv:1912.06385 [cs, eess, stat]*, 1137:349–356, 2019. arXiv: 1912.06385.

[21] Reuban Zacker. What are Convolutional Neural Networks and why are they important?, July 2019.

[22] Nhan Duy Truong, Anh Duy Nguyen, Levin Kuhlmann, Mohammad Reza Bonyadi, Jiawei Yang, Samuel Ippolito, and Omid Kavehei. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Networks*, 105:104–111, September 2018.

[23] Yanli Yang, Mengni Zhou, Yan Niu, Conggai Li, Rui Cao, Bin Wang, Pengfei Yan, Yao Ma, and Jie Xiang. Epileptic Seizure Prediction Based on Permutation Entropy. *Frontiers in Computational Neuroscience*, 12, 2018.

[24] Yunyuan Gao, Bo Gao, Qiang Chen, Jia Liu, and Yingchun Zhang. Deep Convolutional Neural Network-Based Epileptic Electroencephalogram (EEG) Signal Classification. *Frontiers in Neurology*, 11, 2020.

[25] Ali Shoeb. CHB-MIT Scalp EEG Database, 2010. type: dataset.

[26] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, November 1997.

[27] python - RNN Regularization: Which Component to Regularize?