FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Detection of Epilepsy in EEGs using sequence models

Miguel Filipe Saraiva Marques

MESTRADO INTEGRADO EM BIOENGENHARIA - ENGENHARIA BIOMÉDICA

Supervisor: Luís Filipe Pinto de Almeida Teixeira, Assistant professor, FEUP | INESC TEC

Co-supervisor: Catarina da Silva Lourenço, PhD student, University of Twente

July 12, 2018

Detection of Epilepsy in EEGs using sequence models

Miguel Filipe Saraiva Marques

MESTRADO INTEGRADO EM BIOENGENHARIA - ENGENHARIA BIOMÉDICA

July 12, 2018

Abstract

Epilepsy is a neurological disorder characterized by the predisposition to generate seizures. Nowadays, the detection of epilepsy is done by visual assessment of Electroencephalography (EEG) recordings, as the primary method of diagnosis. From all EEG recordings, ictal EEGs, i.e., the part of the signal corresponding to seizure activity, are the principal patterns to ensure the presence of epilepsy. However, the occurrence of seizures is rare and spontaneous, so the presence of these patterns in EEG recordings is poor. On the other hand, interictal epileptiform discharges (IEDs) are electrical activity present in the EEG of epileptic patients that are used as an alternative to diagnose epilepsy. Even though visual assessment is the primary process to detect epilepsy, it is a difficult method that requires long periods of time and a high level of concentration. This appeals to the development of algorithms that automate the detection of epilepsy through analysis of EEG records, increasing assertiveness and reducing time spent on the diagnosis. Deep learning algorithms can be an alternative to visual assessment, making the process faster and more reliable. Over the years, several surveys were conducted to assess the performance of deep models in IED detection, revealing promising results and justifying the bet on automation of EEG analysis.

The work developed in this dissertation explored various deep sequence and attention-based models, with the use of two-second EEG epochs containing normal, focal and generalized epilepsy EEG samples, organized in four different datasets (A, B, C and D). Each dataset was divided into six different parts and, then, five-fold cross-validation was done to the training sets and evaluation was done with an independent test set. Performance evaluation was made by calculating Sensitivity, Specificity, True Positive and False Positive per hour rates at several decision thresholds and AUC value was calculated and corresponding average ROC curves were plotted.

From all the experiments performed, a 2D hybrid convolutional recurrent neural network reached the best performance. This model, named CNN+LSTM4, reached an overall AUC value of 0.96. At 99% specificity criteria, test sensitivity value reached 26.4%. At sensitivity equal specificity condition, a total value of 90.6% was achieved, with true and false positive per hour rates of 106.52 and 157.85, respectively. Aside from this model, other deep networks obtained a good performance, more concretely CNN+LSTM3 and LSTM5, with AUC values of 0.95 and 0.94, respectively.

The work developed shows that deep learning and, more concretely, deep sequence models can be very effective in the analysis of EEG signals for IED detection. Even though several deep learning algorithms have been already applied in the diagnosis of epilepsy, according to the literature, that is not verified in IED detection. In addition to this, methods to surpass data imbalance also showed a positive effect in the improvement of overall deep models.

Resumo

A epilepsia é uma doença neurológica caracterizada pela predisposição para gerar convulsões. Actualmente, a detecção da epilepsia é feita através da avaliação visual dos registos electroencefalográficos (EEG), como o método principal de diagnóstico. Dentro do sinal de EEG, os padrões ictais, ou seja, a parte do sinal correspondente à actividade de convulsão, são os principais padrões para assegurar a presença de epilepsia. No entanto, a ocorrência de convulsões é rara e espontânea, pelo que a presença destes padrões nos sinais de EEG é muito baixa. Por outro lado, as descargas interictais de epilepsia (IEDs) são a actividade eléctrica presente no EEG de doentes epilépticos que são utilizadas como alternativa para diagnosticar a epilepsia. Embora a avaliação visual seja o processo primário para a deteção da epilepsia, é um método difícil que requer longos períodos de tempo e um elevado nível de concentração. Isto apela ao desenvolvimento de algoritmos que automatizam a detecção da epilepsia através da análise dos registos de EEG, aumentando a assertividade e reduzindo o tempo gasto no diagnóstico. Algoritmos de *deep learning* podem ser uma alternativa à avaliação visual, tornando o processo mais rápido e mais fiável. Ao longo dos anos, foram realizados vários estudos para avaliar o desempenho de modelos profundos na detecção de EEG, revelando resultados promissores e justificando a aposta na automatização da análise de EEG.

O trabalho desenvolvido nesta dissertação explorou vários modelos de sequência e baseados em atenção , com a utilização de *epochs* de dois segundos contendo amostras de padrão normal de EEG e padrões de epilepsia focal e generalizada, organizadas em quatro conjuntos de dados diferentes (A, B, C e D). Cada conjunto de dados foi dividido em seis partes diferentes e, em seguida, foi feita uma *five-fold cross-validation* para os conjuntos de treino e validação e a avaliação foi feita com um conjunto de testes independente. A avaliação do desempenho foi feita através do cálculo da Sensibilidade, Especificidade, Verdadeiro Positivo e Falso Positivo por hora , com diferentes limiares de decisão, do cálculo do valor de AUC e as curvas ROC correspondentes foram traçadas.

De todas as experiências realizadas, a rede neuronal *2D hybrid convolutional recurrent* atingiu o melhor desempenho. Este modelo, denominado CNN+LSTM4, atingiu um valor global de AUC de 0.96. Com critérios de especificidade de 99%, o valor da sensibilidade de teste atingiu 26.4%. Para o critério de sensibilidade igual à especificidade, foi alcançado um valor de 90.6%, com taxas de verdadeiros e falsos positivos por hora de 106.52 e 157.85, respectivamente. Para além deste modelo, outras redes profundas obtiveram um bom desempenho, mais concretamente CNN+LSTM3 e LSTM5, com valores AUC de 0.95 e 0.94, respectivamente.

O trabalho desenvolvido mostra que o *deep learning*, mais concretamente, os modelos sequenciais podem ser muito eficazes na análise dos sinais EEG para a detecção de IED. Embora vários algoritmos de *deep learning* já tenham sido aplicados no diagnóstico de epilepsia, isso não é verificado na detecção de IED, de acordo com a literatura. Além disso, os métodos para superar o não balanceamento dos dados das diferentes classes também mostraram um efeito positivo na melhoria dos modelos profundos globais.

Agradecimentos

Cinco anos volvidos, fecha-se, segundo dizem, os melhores anos da minha vida e, com certeza, existem pessoas a quem tenho de agradecer por terem feito parte deles.

Primeiro, agradecer ao Prof. Luís Teixeira e à Catarina Lourenço por todo o apoio e orientação prestados ao longo deste semestre, o resultado nunca seria o mesmo sem a presença de ambos.

Um agradecimento à minha familia, por estarem sempre presentes e que, independentemente do que acontecer, sei que nunca vou estar sozinho. Um agradecimento especial aos meus pais e irmão, que há 23 anos me aturam e me ajudam, devendo-lhes, em parte, tudo que daqui para a frente atingir.

A todos os meus amigos pois, mesmo sabendo como sou, optaram por me manter na vida deles. Desde os resistentes do Marco até ao OG que conheci na faculdade, tenho de agradecer por todas as experiências vividas e por terem tornado estes cinco anos ainda mais inesqueciveis. A amizade que nos une é certamente das coisas mais valiosas que guardo.

Uma menção especial à Inês, pois quem tiver uma Inês na vida, tem uma sorte muito rara.

Miguel Marques

“Tão bom, caraças!”

Nélson Fagundes, Hotel

Contents

List of Figures

List of Tables

Abbreviations

1D	Unidimensional
2D	Bidimensional
3D	Tridimensional
ABSZ	Absence Seizure
ACNN	Artificial Celular Neural Network
AE	Auto Encoder
AED	Antiepileptic Drugs
ANN	Artificial Neural Network
AP	Average Pooling
AUC	Area Under ROC Curve
BP	Band Pass
CBH-MIT	Children's Boston Hospital - Massachusetts Institute of Technology
CDRNN	Convolutional Densely Connected Gated Recurrent Neural Network
CNN	Convolutional Neural Network
CNS	Central Nervous System
CPSZ	Complex Partial Seizure
CRNN	Convolutional Recurrent Neural Network
DC	Direct Current
DCAE	Deep Convolutional Stacked Autoencoder
DCNN	Dense Convolutional Neural Network
EEG	Electroencephalography
FC	Fully Connected
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FN	False Negative
FNSZ	Focal Non-Specific Seizure
FP	False Positive
FPR	False Positive Rate
GNSZ	Generalized Non-Specific Seizure
GRU	Gated Recurrent Networks
HP	High Pass
HIC	High-income countries
ICRNN	Inception Convolutional Gated Recurrent Neural Network
IED	Interictal Epileptiform Discharge
IIR	Infinite Impulse Response
ILAE	International League Against Epilepsy
LFCC	Linear Frequency Ceptral Coefficient-based
LMIC	Low/middle-income countries

LSTM	Long Short-Term Memory
MCSZ	Myoclonic seizures
MLPNN	Multilayer Perceptron Neural Network
MP	Max Pooling
MRI	Magnetic Resonance Imaging
PSD	Power Spectral Density
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SDA	Sparse Denoising Autoencoder
SMOTE	Synthetic Minority Over-sampling Technique
SPSZ	Simple Partial Seizure
SSAE	Stacked Sparse Autoencoder
STFT	Short-Time Fourier Transform
TBI	Traumatic Brain Injury
TCSZ	Tonic Clonic Seizure
TN	True Negative
TNSZ	Tonic Seizure
TP	True Positive
TUH	Temple University Hospital
ViT	Vision Transformer

Chapter 1

Introduction

1.1 Motivation and Context

Epilepsy is a neurologic disorder characterized by an enduring predisposition to generate epileptic seizures and by the neurobiologic, cognitive, psychological, and social consequences of this condition. It is important to notice that epilepsy is the result of brain dysfunction that can have various causes. For the diagnosis of epilepsy in a patient, this patient must suffer from a seizure episode in association with an enduring disturbance of the brain capable of giving rise to other seizures. An epileptic seizure is a "transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain." [3]. This event can be expressed by various body reactions, depending on which part of the brain gets involved (such as loss of awareness with body shaking, confusion, and difficulty responding; visual or other sensory symptoms; isolated posturing of a single limb). The origin of seizure episodes can be provoked or spontaneous. If seizures occur spontaneously, they might be related to underlying epilepsy [4]. Besides that, many patients that have an unprovoked seizure rarely suffer another episode, but when this happens, the probability of occurrence of recurrent seizures is high [5]. If this happens within at least 24 hours, epilepsy is assured [6].

Nowadays, the diagnosis of epilepsy is done by visual assessment of electroencephalography recordings. This procedure helps to determine if attacks have an epileptic origin, allows to estimate the recurrence risk after a first seizure, aids in the diagnosis of the epilepsy syndrome and represents the gold standard in the pre-surgical evaluation of epilepsy [7]. Regarding the analysis of EEG recording, it is important to look for specific and relevant features that indicate abnormal brain activity and, consequently, the presence of epilepsy disorder. Interictal Epileptiform Discharge (IED) detection is one of the most prominent indicatives for the presence of epilepsy disorder. Between seizures, the brain of some patients with epilepsy generates pathological patterns of activity that are clearly distinguished from the activity observed during the seizure itself and normal brain activity. The detection of these patterns requires the visual analysis of EEG recordings, which corresponds to a prolonged and demanding process. In addition to this, subjectivity and intra/inter-variability are also factors that affect the diagnosis [8].

To improve the detection of interictal epileptiform discharges in EEG, diminishing time and increasing the accuracy of diagnosis, it is required the design of efficient algorithms that can surpass visual assessment. The EEG signal is characterized by a variety of properties and features, turning it into a very complex and irregular signal. Thus, Deep Learning models might be the solution needed to address this issue. Deep learning techniques have been employed in various subjects and, their application in the health field is becoming more usual and is gaining importance.

1.2 Research Goals

The IED detection task is an issue that has not been widely explored, even more when it is associated with deep sequence models. In this way, one of the main goals of this dissertation was to design and test different deep sequence architectures to automatize the detection of IEDs in EEG recordings, improving the efficiency of the EEG patterns classification. Moreover, the exploration of a new type of deep models, the attention-based models, was set as an important objective, as it is a recent type of framework with good potential to achieve remarkable results in this area. Furthermore, another important objective was to test different techniques to deal with class data imbalance, an obstacle common in EEG databases, in order to potentiate the predictive ability of the deep networks.

1.3 Contributions

The work developed at this dissertation might bring contributions to the current state of the art. Through the exploration of several distinct deep architectures, with different design and specifications between them all, it was possible to widely verify a range of deep models (sequence and attention-based) capable of detect efficiently IED samples and establish a comparison between all of them. This analysis allow to build a map with the various architectures that can deal effectively with EEG datasets, having distinct features, which can be significant to decide if a model can fit better the needs of a certain situation. Besides that, the implementation of attention-based models to deal with signal processing tasks allowed to understand the ability that these architectures can reach and the potential of these models might have to achieve good performances in IED detection task.

1.4 Dissertation Structure

This document is divided into 7 chapters. Chapter 1 is an introductory section to present the subject of this dissertation, the motives and context to perform with this scientific investigation. Chapter 2, named Fundamentals, with a background of the three main themes of this dissertation: Electroencephalography, Epilepsy and Deep Learning, more specifically deep sequence models. Chapter 3 outlines different technical strategies related to Deep Learning models in Epilepsy, like a description of public databases vastly experimented in this matter, acquisition and preparation of recordings, feature extraction methods, comparison of deep architectures and respective results. Afterwards, Chapter 4 covers the methods that were implemented and used during the dissertation, Chapter 5 describes the results obtained with the methods mentioned in the chapter before and, in Chapter 6, the discussion of the results are presented. Finally, Chapter 7 presents the main conclusions that can be drawn from this dissertation and describes several ideas and steps for future work that can be carried out in the future.

Chapter 2

Fundamentals

2.1 Epilepsy

Epilepsy is a chronic disease of the brain, entailing a predisposition for the generation of seizures and with consequences at a neurobiologic, psychological, cognitive and social level [3]. This neurological condition is very common and affects around 65 million people worldwide[9]. According to the International League Against Epilepsy [6], epilepsy is characterized by, at least, one of the following circumstances:

- A minimum of 2 unprovoked (or reflex) seizures occurring > 24 h apart;
- one unprovoked (or reflex) seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after 2 unprovoked seizures, occurring over the next 10 years;
- diagnosis of an epilepsy syndrome.

2.1.1 Epileptic Seizures

As mentioned before, epileptic seizures are the most relevant manifestation of epilepsy. But, even though every epileptic patient experiences seizures, not all individuals that undergo seizure episodes are epileptic patients [10]. A seizure is identified as an interval (seconds or minutes) of time with an abnormal, synchronous excitation of a region of the Central Nervous System, i.e., a period of disruption of the mechanisms that maintain the balance between excitation and inhibition of neurons [11]. This kind of event can be expressed by diverse body reactions, which are dependent on the region of the brain where the seizure is originated. Between body manifestations, it is known the loss of awareness, body shaking, confusion and difficulty responding to stimulus, visual or other sensory symptoms, between others. The occurrence of a seizure episode can be provoked or spontaneous, which can have implications in the presence or absence of epilepsy [4]. Provoked seizures are usually related to a reversible and isolated insult, like hypoglycemia or fever, so these types of seizures are not connected to the presence of epilepsy syndrome, thus it is a sporadic secondary condition and not a chronic state [12].

Regarding the seizure classification, this is done taking into consideration different parameters: the seizure's origin in the CNS, the type of body reactions resultant from the seizure episode and the subject's level of awareness. According to ILAE [13], seizure classification begins with the establishment of whether manifestations of the initial seizure are focal or generalized,i.e., if the origin of the brain's abnormal excitation occurs in just one hemisphere, in a specific region of the brain, or if it happens in both. On some occasions, the origin may not be encountered, classified as seizures with unknown onset. Concerning focal seizures, these can be sub-grouped taking into consideration the subject's level of awareness suffering from a seizure episode. If the person is aware of her/himself and the surrounding environment we have a "focal aware seizure", but if the subject gets unconscious and unaware of him/her surroundings, it is called "focal impaired awareness seizure". Besides that, the type of body manifestations allows division between the motor (body shaking) and non-motor (loss of consciousness) onset. For both generalized and unknown seizures, these are just divided depending on the type of body reactions.

2.1.2 Etiology, Incidence and Prevalence

The causes that originate epilepsy have various etiologies, systematized in three main groups. First, for nearly half of epilepsy cases, the origin is impossible to identify, while the other 50% of cases have a genetic and/or acquired origin[14], i.e, Idiopathic or Symptomatic epilepsy. Idiopathic epilepsy is defined by the absence of any brain lesion or other neurologic symptoms or signs. On the other hand, Symptomatic epilepsy is the result of one or more CNS lesions or malfunctions[15], such as traumatic brain injury, CNS infections, cerebrovascular disease, febrile seizures, among others [16].

Regarding the incidence of epilepsy[17], it is known that the incidence rate is around 64,7 per 100000 person-year, with significant differences between low/middle-income and high-income countries. This is due to the greater exposure the LMIC population has to perinatal risk factors, higher rates of CNS infections and traumatic brain injury. The same behavior is observed in the statistics of prevalence[17], with an overall lifetime prevalence of 7,6 per 1000 population, with a greater rate in LMIC countries. It is important to refer that these results are more variable among different countries, since the criteria to include an epilepsy patient differs between them, such as the number of seizures in a diagnostic, the local distribution of risk and etiologic factors, between others.

2.1.3 Interictal Epileptiform Discharges

As mentioned in Section 2.1.1, epileptic seizures (also termed ictal discharges) represent the critical events and the primary clinical burden of an active epileptic condition. Between seizures, the brain of patients with epilepsy generates pathological patterns of activity, designated as interictal epileptiform discharges (IEDs), that are clearly distinguished from the activity observed during the seizure itself. The correct identification of IEDs might be very useful for the diagnosis of epilepsy, due to its presence in more than 80% of patients with clinically confirmed epilepsy [18].

In addition to this, interictal epileptiform discharges are rarely found in healthy subjects and non-epileptic patients. In healthy people with no record of experiencing a seizure episode, IEDs are found in 0.5% of routine EEG recorded [19]. For children and non-epileptic patients, the incidence of interictal discharges was encountered in 2%-4% of EEG acquisitions. The presence of IEDs in clinic EEG records increase 10%-30% in cerebral pathologies like a tumor, prior head injury, cranial surgery, or congenital brain injury [20].

Regarding interictal discharges specifications, they are characterized by a large-amplitude rapid component lasting 50–100 ms that is usually followed by a slow wave, 200–500 ms in duration [21]. Several examples of IEDs can be seen in Figure 2.1

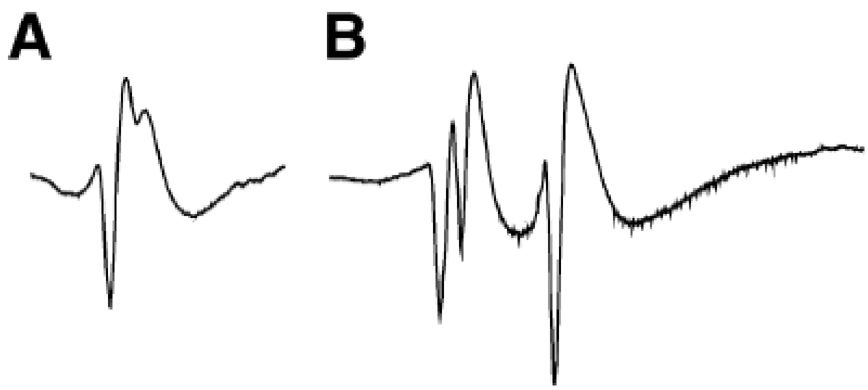


Figure 2.1: Interictal Epileptiform Discharges patterns. A) Interictal spike B) group of interictal spikes [22]

2.1.4 Diagnosis and Treatment Options

The diagnosis of epilepsy is not a simple process. Even though the pathophysiologies of epilepsy and other paroxysmal disorders are distinguishable, their clinical manifestations may be similar leading to difficulties in a precise diagnosis, making, unfortunately, misdiagnosis common, with 20%-30% of epileptic patients misdiagnosed [23; 24]. The diagnosis procedure combines the analysis of the patient's history, physical and neurological examination, laboratory tests if indicated, and electroencephalography (EEG) and neuroimaging (MRI) recordings. Clinical history should include information like events directly preceding the seizure, the number of seizures in the past 24 hours, length and description of the seizure and length of the postictal period. Regarding laboratory tests, these encompass blood tests (such as blood glucose or blood counts), electrolyte panels (specifically sodium), lumbar puncture in febrile patients, and urine toxicology. EEG recordings are useful to confirm, but not to exclude, an epilepsy diagnosis[25].

Concerning treatment options, these are dependent on the risk level for recurrent epileptic seizures associated with the patient. The key risk factors for recurrence are age-dependent. For adults, high-risk aspects include two unprovoked seizures that occurred more than 24 hours apart from each other, nocturnal seizures, abnormalities on EEG signals and MRI images [6; 26; 27]. Other conditions have been identified for children, such as abnormal electroencephalography results, the presence of a syndrome predisposing to seizures, and an etiology such as severe head trauma or cerebral palsy [28].

The first option of treatment is antiepileptic drugs. The AED therapy is used to decrease the risk of seizure recurrence, with the choice of medication being dependent on the seizure type. AED therapy is only an option if the patient presents a high risk of seizure recurrence, otherwise, delaying the use of antiepileptic drugs is advised[29]. In addition to this, other clinical treatments are available when AED does not present good results. The first treatment alternative is surgery. Patients with focal epilepsy may be candidates for epilepsy surgery if the seizure origin is in a part of the brain that can be safely resected. If a patient has not had seizure control after trials of two or more appropriately chosen AEDs at therapeutic doses, surgery may offer a higher chance of seizure freedom than additional medications. The other treatment alternative is neurostimulation. Patients that have seizures originated in a fundamental region of the brain, like primary language or primary motor areas, or suffer from generalized or unknown seizures, are candidates for treatment through neurostimulation.

2.2 Electroencephalography

Electroencephalography (EEG) is the term used to describe the measurement of human cerebral electrical activity. The first attempts to record this physiological signal date back to 1928 and 1935, by the German psychiatrist Hans Berger [30]. But, only around the 1990s, the EEG signal applications grew in interest and relevance with the improvements of technology and computerization techniques. The creation of digital systems allowed clinicians to record EEG in a computer, which could be analyzed and processed afterward, increasing the potential and utility of EEG records. Nowadays, EEG signals are applied in various branches of healthcare with great relevance [31].

2.2.1 EEG signal Analysis

The brain is a complex organ responsible for controlling the behavior of the body in response to internal/external motor/sensory stimuli. With electroencephalography, it is possible to assess the brain signals that are generated during its activity. Signals are categorized as alpha, beta, theta, delta and gamma, depending on the frequency, which can range from 0.5 to above 100 Hz. This classification is, naturally, related to a variety of factors, like age, stage of alertness, activity performed, brain's health, among others, which is translated in specific waveform shapes, signal's amplitude, frequency and the brain's lobe where the signal is produced. [32].

The correct characterization of brain activity is important to assess if there is no abnormal neurological behavior and to properly understand the origin and the possible causes of those abnormalities (ex: pathological events like epilepsy, tumor, cerebrovascular lesions, depression and problems associated with trauma). In the case of epilepsy or seizure activity, specific epileptiform abnormalities can be detected, presenting different characteristics depending on the seizure's onset. For example, high frequency (>13 HZ) ictal EEG and periodic spikes highly correlated with temporal lobe lesion [33] or also low frequency (<2 Hz) spikes with duration equal to or greater than 5 seconds from ictal EEG correspond to a selective neuronal loss in the CA 1 region of the hippocampus [34]. In Figures 2.2 and 2.3, an example of normal EEG and a seizure EEG records can be seen.

Figure 2.2: Normal EEG recording from an healthy adult [35]

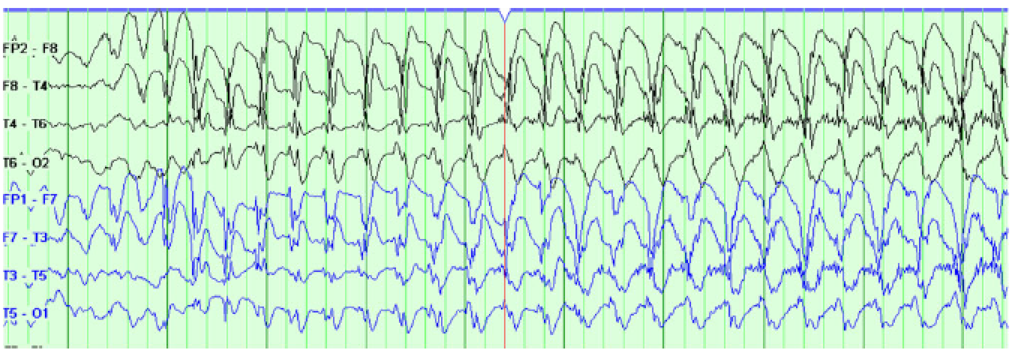


Figure 2.3: Scalp EEG recording from an ictal onset [35]

2.2.1.1 Artefacts

EEG signal recordings are always affected by the presence of artefacts and other interferences that compromise the quality of the acquisition. The characterization of artefacts is based on their origin, which can be physiological/biological or external/technical. Looking at patient-based artefacts, the most common onsets are small body movements, the presence of other physiological signals (such as electromyography and electrocardiography), or eye movements. Regarding external artefacts encountered in EEG signals, the most commons are related to the electrical components used to acquire the signals (cable movements or broken wire contacts), deficient electrode contact with the patient, power line interference (50/60Hz), between others. All these interferences have a specific frequency and amplitude range, which affects differently the raw EEG record. Thus, it is of the utmost importance to effectively remove these artifacts in order to improve the quality of the signal. Filtering the signal is essential to obtain better quality, but other actions can improve its acquisition, such as decreasing electrode impedance and by shorter electrode wires [36; 37]. In Figure 2.4, it is possible to see some of the aforementioned artefacts.

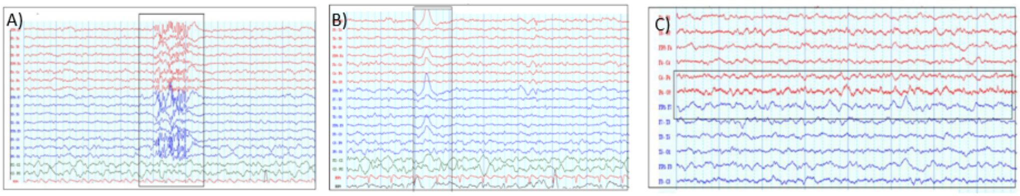


Figure 2.4: Examples of Artefacts in EEG records. A) Muscle Artefact B) Eye blink artefact C) 50Hz interference [36]

2.2.2 EEG Signal Acquisition

As mentioned before, the development of digital technologies transformed the acquisition of EEG signals, with a well-established system and recording strategy. This modern data collection system presents diverse components with specific functions, like data acquisition devices, processing devices, recording accessories, among others.

Regarding recording accessories, the electrodes are a very important part of the acquisition system, responsible for connecting the location where the signal is measured and the amplifier [31]. From all the available electrodes, three types are preferential for electrical signal acquisition: disposable Ag/Cl non-invasive electrodes, single-use subdermal needles and implanted invasive electrodes, each one of them with advantages and drawbacks. According to the Federation International of Clinical Neurophysiology [38], the standard electrode disposal for clinical EEG recordings is the 10-20 Electrode Placement.

The electrodes' leads are connected to a differential amplifier, responsible for amplifying the electrical signals to an amplitude range that is understandable in electronic circuits. The amplification factor is around 10000, converting from a 500 μV amplitude to a voltage of a few volts. After signal amplification, several digital filters are applied in order to remove noise and irrelevant signal frequencies to improve EEG signal quality, such as artefacts created by other physiological signals, movement of the body, or interference by the power line and another electric equipment surrounding (for further information, see Section 2.2.1.1 To achieve this goal, low-pass, high-pass, notch and anti-aliasing filters are applied to the recorded EEG signals [39]. In the end, the acquired physiological signals are submitted to an Analog-to-Digital converter for the signal to be stored and processed on a computer.

2.2.3 EEG applications in Healthcare

For the diagnosis of epilepsy, EEG analysis did not have the same protagonism during the time from the first EEG recording until nowadays. Up to the 1990s, EEG signal recording lost relevance due to the appearance of MRI images, which gain more protagonism in epilepsy diagnosis. After that time, with the development of computerization techniques, the use of EEG signals to correctly diagnose epilepsy became more relevant. This development allows more precise and secure localization of abnormalities, correct identification of artefacts and longer record times. Besides the diagnosis of epilepsy, EEG is also required to manage patients in coma or impaired cognitive state (children and adults in intensive care units) and for evaluation of patients for epilepsy surgical treatment [35]. EEG acquisition is also a valuable tool to assess other neurological diseases like Alzheimer [40], dementia [41], depression [42], among others. Aside from the detection of neurological pathologies, electroencephalography is used to perform sleep analysis and to monitors patients exposed to anesthesia [43; 44].

2.3 Deep Learning

Deep learning can be defined as a class of machine learning algorithms that allow computers to learn from experience and to understand the world as a hierarchy of concepts, i.e., AI algorithms that reach complicated concepts explained by the construction of a relation between simpler concepts[45]. In other words, deep learning uses a cascade of multiple layers of nonlinear progressing units for feature extraction and transformation. In each successive layer uses the output of the previous one as input, learning multiple levels of representations that correspond to various abstraction levels [46]. This hierarchy of layers, which build a deep learning model, is called an Artificial Neural Network. ANN is named as such because they were inspired by the biological neural networks present in human brains. More concretely, ANN is a group of different neural networks that diverge in complexity, dimension and type of data they are specialized to process. Inside this broad family of models, we have Convolutional and Recurrent Neural Networks.

2.3.1 Convolutional Neural Networks

Convolutional Neural Networks are a specialized group of neural network models that operate with data in a grid-like structure [47]. The main difference between these networks with standard ANN is related to the mathematical operation applied at activation functions, named convolution. Convolution is a special kind of linear operation defined as the integral of the multiplication between two sequences after one of them has been reversed and shifted [45]. To improve the performance of deep frameworks, convolution introduces three important concepts to achieve better execution: sparse interactions, parameter sharing and equivariant representations. Sparse interactions mean that not all output units from a layer are connected with all input units from the next layer. In the convolution operation, all the kernel elements are used in every input element, instead of using a separate group of parameters at each input instance, improving efficiency in memory and statistical requirements. Lastly, equivariant representations is a relevant feature of convolution operation. It means that, if the operation input changes its shape, the output will change similarly.

An important characteristic of Convolutional Neural Networks is that they can eliminate manual feature extraction. A given input image or signal goes through a set of filters in the convolutional layer, followed by a pooling layer. At each convolutional layer, feature maps are obtained through successive convolutions between the input and the weight vectors (filters). Normally, in the end, fully connected layers work as final learning phases, where the features are mapped into the predicted outputs [48].

2.3.2 Recurrent Neural Networks

Recurrent Neural Networks are a group of neural network models that are specialized in sequence data processing [49]. Due to its sequence-based representation, RNNs are able to scale to much longer sequences, when compared with other networks that do not have this specialization. An important feature of these deep architectures is the sharing of parameters. Unlike other neural networks, each member of the output is dependent on all previous outputs members and it is produced under the same update rule, resulting in parameter sharing through many layers[45]. Recurrent Neural Networks can be divided into three different designs. In the first design, an output is produced at each time step and recurrent connections are made between hidden units. Secondly, an output is also produced at each time step, but recurrent connections only happen between the output at one time step with the hidden units of the time step after. Lastly, another design is defined by the production of a single output with recurrent connections between the hidden units. In Figure 2.5, a representation of the three designs can be observed.

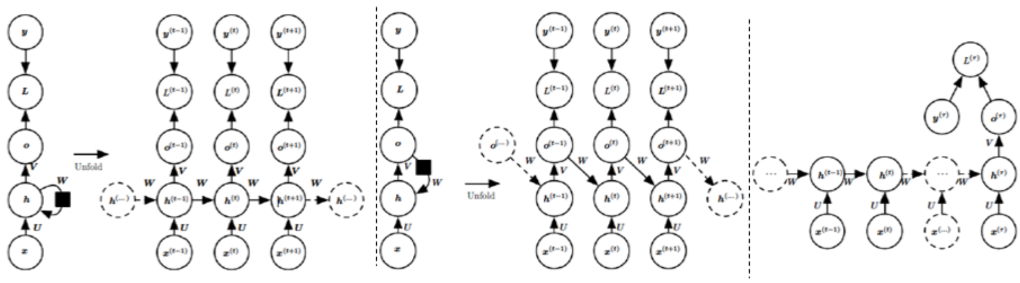


Figure 2.5: Recurrent Neural Network architectures [45]

2.3.2.1 Bidirectional RNN

For most RNN models, the output at a time step t only depends on the information of the past, i.e., the inputs from time step 1 until time step t. But, for some applications, it is imperative that the output at each time step is dependent on the whole input sequence, like speech or handwriting recognition. The development of bidirectional RNN models allows the use of the whole input sequence at each time step. This type of models results from the combination of two RNNs, one that moves forward through time starting at the beginning of the input sequence, while the other moves backward beginning at the end of the sequence, which allows that, at each time step, the respective output is affected by past and future information [50]. In Figure 2.6 an example of a bidirectional RNN can be seen.

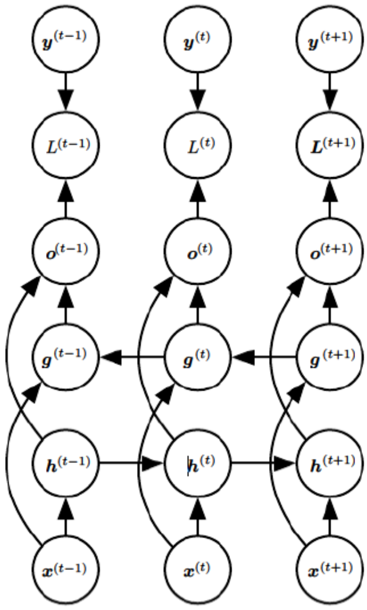


Figure 2.6: Bidirectional Recurrent Neural Network [45]

2.3.2.2 Encoder-Decoder or Sequence-to-Sequence Architectures

There are some applications, like speech recognition, machine translation or question answering, in which the length of the input sequence and the output sequence are not the same. To address this issue, Encoder-Decoder or Sequence-to-Sequence models were created [51; 52]. The workflow of this type of NN architectures is very basic. First, an input sequence is processed by the Encoder with the production of a variable, called context C, which summarizes the information present in the input sequence, normally as a function of its final hidden state. After that, the context C is sent to the Decoder, which will process the variable in order to produce the output sequence intended [45]. In Figure 2.7, a schematic of the architecture can be seen.

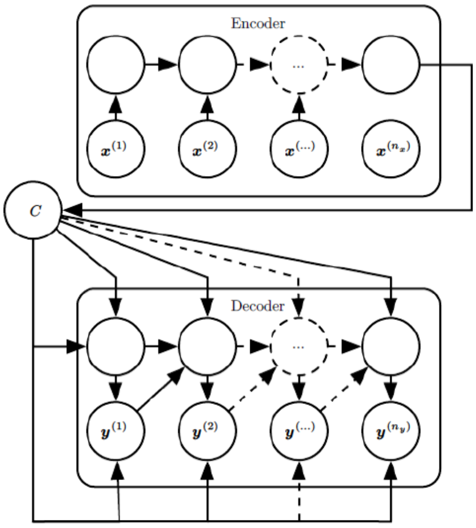


Figure 2.7: Encoder-Decoder or Sequence-to-Sequence Neural Network models[45]

2.3.2.3 Recursive Neural Network

Recursive Neural Network is a derivation of typical Recurrent Neural Networks. This type of architectures differ from other RNN models due to their tree-like structure, different from the chain-like shape, as can be seen in Figure 2.8. Recursive models were first introduced by [53] and, lately, successful experiments proved the capacity of Recursive NN architectures to work with natural language processing and computer vision [54; 55]. Recursive nets have a clear advantage that is the decrease in-depth for the same input sequence when compared with Recurrent nets[45].

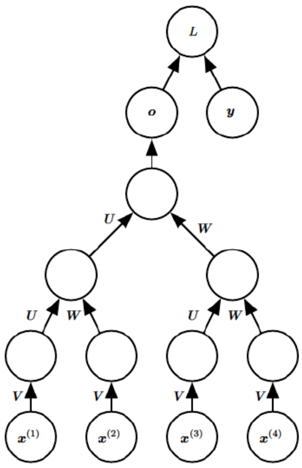


Figure 2.8: Recursive Neural Network tree-like structure[45]

2.3.2.4 Long Short-Term Memory Network

With the development of RNN models and with an increase of complexity, a new problem appeared, called long-term dependencies. It was claimed by some authors that RNN architectures are unable to identify and learn long-term dependencies in the sequence data for more than 10-time steps [56; 57; 58]. With the increase of networks in-depth, the optimization step became more difficult. In RNN models, the same operation is applied several times at each time step, so, the propagation of gradients over many time steps tends to vanish or explode, difficulting the optimization of the RNN model [59]. To overcome this problem, gated RNNs were created, like Long Short-Term Memory networks. Instead of a simple unit that applies an element-wise nonlinear function to the input, LSTM recurrent networks have a structure called "LSTM cell". This cell has the same inputs and outputs as conventional RNN units, but it also has more parameters and an organization of gating units that control the flow of information that enters and leaves the memory cell [45]. A standard architecture of a LSTM cell can be seen in Figure 2.9.

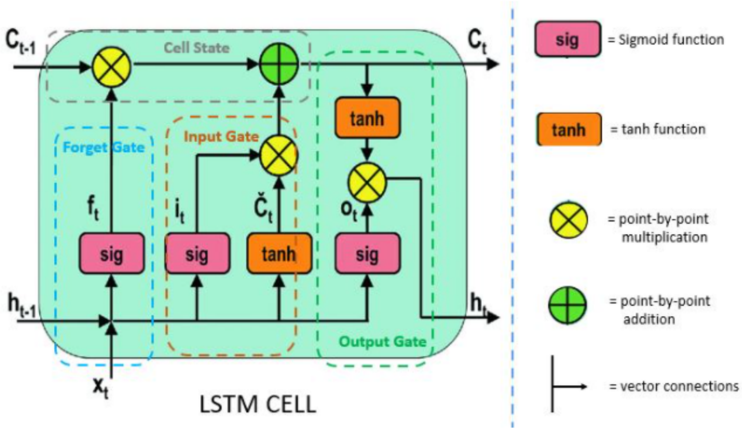


Figure 2.9: LSTM Cell structure [60]

The three gate units existent in a LSTM cell are called forget gate, input gate and output gate.

First, the forget gate f(t) is responsible for deciding which information is relevant and which is not. The data from the input sequence x(t) and the hidden state from the previous cell h(t-1) enters in a sigmoid function, generating a value between 0 and 1 (equation 2.1. This result is then applied to the output of the previous cell C(t-1), denominated cell state, in accordance with its importance [61; 60; 45]. In all the formulas presented below, W and b variables are the parameters of the deep models, respectively, weight and bias matrices.

$$f(t) = \sigma(W_f * [h(t-1), x(t)] + b_f) \tag{2.1}$$

Second, the input gate applies a sigmoid and tanh function to the previous hidden state h(t-1) and the input variable x(t), producing the outputs i(t) (equation 2.2 and $\tilde{C}$(t) (equation 2.3). The application of the sigmoid function has the same purpose as at the forget gate. The use of the tanh function is intending to regulate the network, producing a vector with values between -1 and 1. These two results are then combined through point-to-point multiplication. The output will update the new cell state C(t) [61; 60; 45].

$$i(t) = \sigma(W_i * [h(t-1), x(t)] + b_i) \tag{2.2}$$

$$\tilde{C}(t) = \tanh(W_c * [h(t-1), x(t)] + b_c) \tag{2.3}$$

Finally, the last is the output gate. The purpose of the output gate is to determine the hidden state h(t) that will be passed to the next LSTM cell. The output gate, three non-linear functions are applied: first, a sigmoid function is applied to the current hidden state h(t-1) and the input x(t), similar to the input and forget gate, producing the output variable o(t) (equation 2.4). At the same time, a tanh function is applied to the new cell state C(t), and this result is combined with o(t) by a point-by-point multiplication, producing the new hidden state h(t) (equation 2.5) [61; 60; 45].

$$o(t) = \sigma(W_o * [h(t-1), x(t)] + b_o) \tag{2.4}$$

$$h(t) = o(t) * \tanh(C(t)) \tag{2.5}$$

Cell State operation is a point-by-point addition between the multiplication of forget gate output f(t) and the old cell state C(t-1) and the result of the input gate unit. This operation is responsible for updating the cell state C(t) that will be passed to the next LSTM cell. Equation 2.6 shows the calculus to update cell state [61; 60; 45].

$$C(t) = f(t) * C(t-1) + i(t) * \tilde{C}(t) \tag{2.6}$$

To summarize, the forget gate determines which data from the prior steps is relevant. The input gate decides what relevant information can be added from the current step, and the output gate produces the next hidden state.

This new RNN approach became very successful in many applications, such as handwriting recognition and generation, speech recognition, image captioning, among others [62; 63; 52; 64].

2.3.2.5 Gated Recurrent Units

Gated Recurrent Units are a type of RNN architecture that evolved from the LSTM model. The difference between GRU and LSTM units is the number of gating units that exist at each cell. For GRU models, there are only two gates: update and reset gates. This simpler model is advantageous because it decreases computation time. In Figure 2.10 it is possible to see a scheme of GRU cell[45; 65].

Starting with update gate u(t), it is responsible for determining if the information present in the input x(t) and the hidden state from the previous cell h(t-1) are important to upgrade the next hidden state h(t). On the other hand, the reset gate r(t) is used to decide how much of input information should be neglected (equations 2.7 and 2.8). These two variables, together with the input and current hidden state, are used to produce the next hidden state h(t), passed to the following GRU cell. W and b are, respectively, the weight and bias matrices[45; 65].

$$u(t) = \sigma(W_u * [h(t-1), x(t)] + b_u) \tag{2.7}$$

$$r(t) = \sigma(W_r * [h(t-1), x(t)] + b_r) \tag{2.8}$$

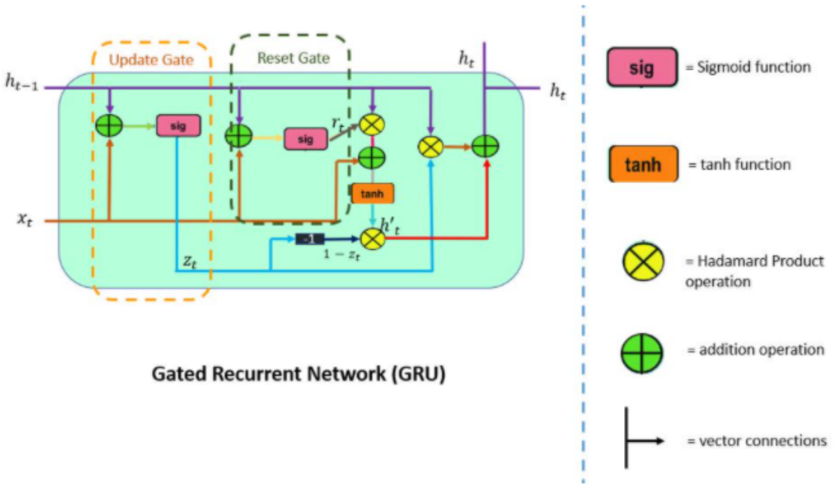


Figure 2.10: GRU Cell structure[65]

2.3.3 Transformer

Transformer is a recent sequence-to-sequence deep model that is based only on attention processes. It was introduced in 2017, by Vaswani et al [66]. This new architecture has the intent to surpass the disadvantages of recurrent and convolutional-based neural networks. The first implementations of this type of architecture were in natural language processing, but, nowadays, other approaches were explored, like image recognition and signal processing [67; 68; 69; 70].

This model has an encoder-decoder structure, with a Self-attention Module. The module is a group of stacked Multi-Head Self-attention and Feed Forward sub-layers. Attention is a function that maps a query and a set of key-value pairs to an output, which is computed as a weighted sum of the values, where the weight associated to each value corresponds to the result of a function between the query with the corresponding key [66]. The scaled dot-product attention function can be seen below:

$$Attention(Q,V,K) = softmax(\frac{Q * K^T}{\sqrt{d_k}}) * V \tag{2.9}$$

where Q, V and K are the query, value and key matrices. Multi-Head Attention allows the attention mechanism to be performed several times, combining those results to produce an output that is the result of concatenation of all attention output's. After that, the output is passed through a simple Feed Forward network, with the application of two similar linear transformations, with a ReLU function between them.

$$FFN(x) = max(0; x * W_1 + b_1) * W_2 + b_2 \tag{2.10}$$

where W1,W2 are the weights and b1,b2 are the biases of the linear transformations. In addition to this, a residual connection [71] and layer normalization [72] are applied to the output of the Multi-Head attention and Feed Forward sub-layers. This module structure is equal for both

Encoder and Decoder. This architecture was applied to Natural Language Processing tasks, but, for classification problems, the Decoder part was replaced by a Multi Layer Perceptron and a Softmax layer. The input of the Transformer suffers a preprocessing before entering the model. This preprocessing is a two-step processing, with an input embedding and a positional encoding mechanisms. Input embedding converts the input vectors in embedded representations of itself. It helps to convert the input dimensions to the desired ones. These embedded representations are then passed through a positional encoding function. This encoding of the input helps to surpass the inability of the attention models to remember the order of the sequence, as recurrent networks do. The order of the sequence is fundamental to discover the time dependencies of the input. The encoding mechanism can be performed in several ways [73]. One manner to perform positional encoding is through sine and cosine functions,

$$PE(pos;2i) = sin(\frac{pos}{10000^{2i/d_{model}}}) \tag{2.11}$$

$$PE(pos;2i+1) = cos(\frac{pos}{10000^{2i/d_{model}}}) \tag{2.12}$$

where pos is the position of each input element, d_{model} is the dimension of each embedding vector and i is the dimension of each vector embedding element. This function allows to assign a unique encoding to each time-step, with consistent distance between to any time-steps, independently of the distance between them. Another way to perform positional encoding is 1D standard learnable position embeddings. In Figure 2.11, the architecture of a Transformer is shown. These type of models show promising results to capture long time dependencies in sequences, something that, until now, Recurrent Neural Networks were the most advancing models to perform, bringing computational advantages.

2.3.4 Optimization

Nowadays, with the increase in depth and complexity of neural network architectures, and with the growth of the amount of data used to train and validate the model, the total computational time and cost of training the NN model, naturally, rises to values obstructive to the scientific development. In this way, optimization algorithms were created to diminish the total computational time, by altering the parameters of the neural network, to achieve a good performance, with a decay of the losses. Multiple optimization algorithms were created, with different configurations and objectives, like Gradient Descent, Momentum, RMSProp, Adam, among others [74].

Gradient descent is the most basic optimizer, but also one of the most used. Gradient Descent is a first-order optimization algorithm, i.e., it is dependent on the first derivative of the loss function to update the parameters of the NN model. This algorithm is easy to implement and understand. Although, required memory can reach a significant value when the dataset used to train the model has a lot of data because the update of the parameters is only made after the optimizer is applied to the whole dataset. In addition to this, with a dataset very large, finding the local minima would

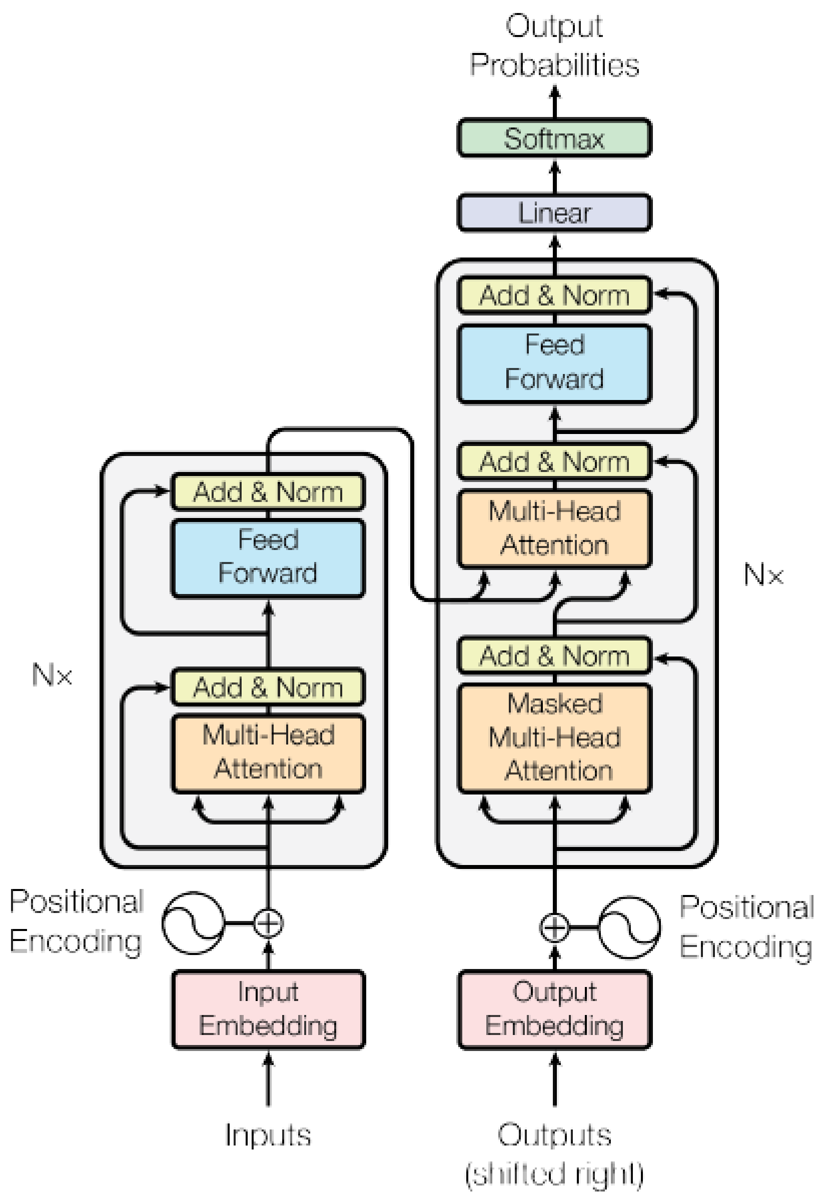


Figure 2.11: Transformer model structure[66]

take a long time. In equation 2.13, Gradient Descent calculus is made as follows,

$$\theta = \theta - \alpha * \nabla J(\theta) \tag{2.13}$$

where α is the learning rate, ∇J(.) is first derivative of loss function and θ is a generalization of the model's parameters (weights and biases).

Momentum [75] is one variation of Basic Gradient Descent, with the intent to accelerate the optimization task, by decreasing the variance associated with GD. In this optimizer, an exponentially weighted average is applied to the gradient, which allows the use of a higher value of the learning rate α, converging faster than GD. On the other hand, a new hyperparameter has to be introduced with this optimization algorithm, called momentum β_1, defined 0.9 as default.

$$V_\theta = \beta_1 * V_\theta + (1 - \beta_1) * \theta \tag{2.14}$$

$$\theta = \theta - \alpha * V_\theta \tag{2.15}$$

Another optimization algorithm used to surpass Gradient Descent limitations is named RMSProp [75; 76]. RMSProp was created with the same purpose as Momentum, but with a different implementation. The objective of RMSProp is to decrease the oscillations and variance of Gradient Descent, intending to increase the learning rate α. In equations 2.16 and 2.17, it is shown the RMSProp calculus, where β_2 is the new parameter, equal to 0.999 as default, ε is a constant (1^{-7} as default) for numerical stability.

$$S_\theta = \beta_2 * S_\theta + (1 - \beta_2) * \theta^2 \tag{2.16}$$

$$\theta = \theta - \alpha * \frac{\theta}{\sqrt{S_\theta} + \varepsilon} \tag{2.17}$$

The last optimizer referred is Adam (Adaptive Moment Estimation) [77], which is a combination of RMSProp and Momentum optimizers. The aggregation of these two optimization algorithms potentiates the two optimizers. In equations 2.18, 2.19 and 2.20 can be seen the formulas applied by the Adam optimizer, with default values for β_1,β_2 and ε of 0.9,0.999 and 1^{-7}, respectively.

$$S_\theta^{connected} = \frac{S_\theta}{1 - \beta_2} \tag{2.18}$$

$$V_\theta^{connected} = \frac{V_\theta}{1 - \beta_1} \tag{2.19}$$

$$\theta = \theta - \alpha * \frac{V_\theta^{connected}}{\sqrt{S_\theta^{connected}} + \varepsilon} \tag{2.20}$$

2.3.5 Recurrent Neural Network Applications in Healthcare

With the increase of biomedical engineering, biomedical research is increasing over time. As a result of this innovative research, various databases are being recorded, such as electronic health records, imaging databases, -omics, sensor data and text. This data is characterized by its high complexity, heterogeneity, poorly annotation and unstructured organization. The interpretation of these recorded data is, naturally, hard and time-consuming. Deep learning algorithms are an appealing alternative to deal with this kind of data and properly understand and systematize the information contained, creating a path to translate complex biomedical data into refined healthcare strategies. Even though deep frameworks show promising results, their versatility is not explored, being just applied in some specific areas. The examples of applications that will be discussed in this subsection are related to Recurrent Neural Network approaches [78]. RNN architectures have been employed to deal mostly with Electronic Health Records and Data Imaging databases.

Following the advances made in computer vision, RNN architectures are starting to be applied to analyze clinical images, more specifically Encoder-to-Decoder algorithms. *Liu et al* [79] applied a Stacked Sparse Auto Encoder model to analyze brain Magnetic Resonance Images for diagnosis of Alzheimer's disease. Besides that, *Cheng et al* [80] examined ultrasound images for diagnosis of breast nodules and various lesions.

Regarding applications to Electronic Health Records, Long Short-Term Memory, Gated Recurrent Unit and Stacked Auto Encoder algorithms were implemented to correctly interpret the information contained in EHR files. Electronic Health Records are digital files that contained patient-centered information, like a patient's medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory and test results [81]. For example, *Lipton et al* [82] describe the application of a LSTM model to help to recognize multivariate time patterns in-clinic medical data measurements from UCI patients. Aside from that, *Lasko et al* [83] designed a Stacked AE to survey sequences of serum uric acid measurements to identify multiple populations sub-types and to distinguish the uric-acid signatures of gout and acute leukemia. In addition to this, a GRU algorithm was created to predict possible diagnoses and treatments taking into consideration their medical history and clinical records. This last example was implemented by *Choi et al* [84].

Although the implementation of deep learning architectures shows promising results, it still exists several limitations and obstacles to overcome to fully integrate deep frameworks in the healthcare environment. A deep neural network requires a great amount of data to acquire a good generalization ability. Besides that, this data should be of good quality, i.e., well structured and without noise. In many branches of healthcare, still does not exist the required amount of data that allows the implementation of deep models. Apart from that, even when exists enough medical data to build an effective deep model, the quality of the data is not necessary to successfully implement these algorithms. The complexity and diversity of diseases halt the growth of knowledge of their causes and how they progress, which compromises the quality and outdates the information contained in medical data records. One last limitation is related to the trust given to NNs by health professionals. It is difficult to fully understand how deep neural networks interpret and adapt to the medical data that is provided, which creates insecurity between healthcare professionals concerning its real efficiency.

2.3.6 Performance

2.3.6.1 Evaluation Metrics

Evaluation metrics are used to measure the performance of a deep model given a certain dataset. Each one of them is used according to the goal to be achieved.

The first evaluation parameter is Accuracy, which assesses the number of correct predictions above all predictions. It is calculated by

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.21}$$

where TP is true positive, TN is true negative, FP is false positive and FN is false negative. A limitation of this metric is that should be only used when the classes for classification are balanced in the dataset. For the situations when classes are imbalanced, other performance metrics can be used, such as Sensitivity , Specificity or Precision,

$$Sens = \frac{TP}{TP+FN} \tag{2.22}$$

$$Spec = \frac{TN}{TN+FP} \tag{2.23}$$

$$Prec = \frac{TP}{TP+FP} \tag{2.24}$$

which evaluate the model's performance over each class separately. The sensitivity indicates the percentage of positive samples that were classified correctly, while specificity displays the percentage of negative samples that were classified as such. Besides that, Precision is an important measure that tells, from all the samples classified as positive, how many are actually positive.

Other two evaluation metrics are the ROC curve (Receiver Operating Characteristic curve) and the AUC (Area Under ROC Curve). ROC curve is a graph that shows the performance of the NN model for all the classification thresholds. This graph has, in the x-axis, the True Positive Rate and, in the y-axis, the False Positive Rate.

$$TPR = \frac{TP}{TP+FN} \tag{2.25}$$

$$FPR = \frac{FP}{FP+TN} \tag{2.26}$$

In addition to this, AUC is a performance measure that reflects the capacity of the model to distinguish between two separate classes. When a model classifies the samples 100% correctly, the AUC value is equal to 1.0.

2.3.6.2 Overfitting

The evaluation metrics mentioned in the previous subsection are not enough to evaluate the performance of a deep model. Usually, the performance measures calculated in the training set are better than the ones calculated in the test set. When the difference is too big, it means that the model performs well with observed data, but it fits poorly on unseen data. This phenomenon is called Overfitting [85]. This has three main causes: noise in the training data, the limited size of the training data and the complexity of the model.

To overcome this poor generalization, some strategies were designed, such as Early Stopping, Dropout and Data Augmentation. Early Stopping [86] is a technique that judges the behavior of an evaluation metric, applied to the training and validation sets (like accuracy, loss, among others). When the validation metric stops improving during some epochs,i.e., the training metric continues to improve, but the validation one stagnates, this means the model is reaching overfitting. With

Early Stopping, the training and validation metrics are monitored and when the behavior described above is observed, the training is interrupted, preventing the model from overfitting. Secondly, Dropout is a strategy to decrease the complexity of the network used, by randomly dropping units from hidden and visible layers, which deactivates, during training, all the connections made by those units. According to [87], Dropout prevents overfitting and, by randomly deactivate different neurons in the network, thus, at each training epoch, the model's architecture is different, applying model combination, which improves deep frameworks methods. Lastly, Data Augmentation is another technique to prevent overfitting. For the model to obtain a good classification performance, it needs to learn properly the principal features that characterize the observed data, which are also going to be present in the unseen data. For this to happen, it is necessary to provide a great amount of data, that is not always available. When this data is limited, some transformations can be applied to the data to increase it, like rescaling, shifting, randomly shuffling, cropping the data, between others.

Chapter 3

Deep Learning in Epilepsy: State-of-art

The application of Deep Learning algorithms in healthcare has been growing, as mentioned in section 2.3.5. A plausible implementation of deep frameworks in the automation of EEG analysis might facilitate the achievement of an accurate diagnosis, monitoring and treatment of various neurological diseases. Concerning the diagnosis of epilepsy through EEG patterns, the traditional assessment of EEG records is done visually by clinicians and, due to the duration of the recordings, the process of interpreting EEG patterns is very laborious and time-consuming, leading to a misdiagnosis of 23% of cases [88]. Thus, the automation of EEG analysis is of upmost importance, in order to achieve better diagnosis and monitoring performance. The following subsections will focus on the work that has been developed in this area, with the main interest given to the detection through deep sequence models, as it will be a keystone matter of this dissertation.

3.1 Acquisition Settings

Acquisition settings are the strategy of place electrodes to record EEG patterns. As already mentioned in section 2.2.2, some criteria need to be established in order to guarantee a good acquisition,i.e., the quality of the database's signal to improve performance assessment. Looking at the electrode placement, the standard 10-20 electrode disposal, according to the Federation International of Clinical Neurophysiology [38] was used throughout most of the research covered by this literature review. This system standardizes the location of electrodes in the scalp for EEG recording in order to allow replicability between different surveys' outcomes, for further analysis, according to the scientific method. The identification of the electrodes is made in agreement with the brain lobe where the electrode is placed (pre-Frontal (Fp), Frontal (F), Parietal (P), Occipital (O) and Temporal (T)). Besides that, two other electrodes are identified as "C" and "A" with distinguished meanings. The "C" electrodes are placed in the center of the system and not in a specific lobe. These electrodes usually record EEG activity from the frontal, temporal, and some parietal-occipital lobes. On the other hand, the "A" electrodes are placed in the outear bone process and they are used as a contralateral reference for the rest of the electrodes. Some electrodes are identified with an additional "z" (Cz, Fz, Pz), which refers to electrodes placed on the midline sagittal plane of the skull, to be used as "ground" references for all the EEG electrodes. Lastly, the "10" and "20" numbers present in the name that label this standard method are related to the distance between adjacent electrodes of 10 and 20% of the total front-back and left-right measurement of the skull, respectively [89; 90]. Besides this acquisition scheme, there is another one called 10-10 electrode system that is an extension of the aforementioned system, using more electrodes and with a subsequent decrease in the distance between adjacent electrodes, in order to cover more regions of the brain and to get a more detailed signal acquisition [90; 91].

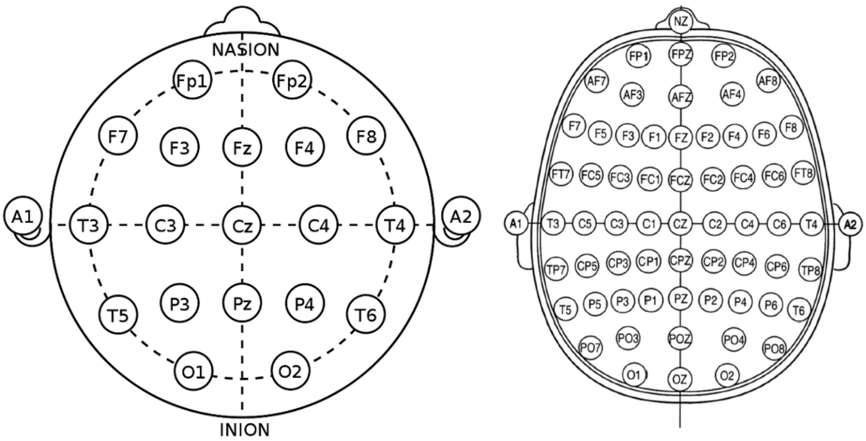


Figure 3.1: 10-20 and 10-10[90] Electrode Placement system, by the Federation International of Clinical Neurophysiology

Taking into consideration other acquisition specifications, the sampling rate is an important component of the EEG acquisition that influences the quality of the recorded data. The sampling rate corresponds to the number of data points per second used for the reconstruction of the analog/natural EEG signal. The choice of the sampling rate is based on the Nyquist-Shannon sampling theorem [92], which determines that the best sampling rate has to be, at least, twice the maximum frequency of the signal being sampled. Besides that, the introduction of noise from the amplifier and the ADC converter is directly related to the sampling rate established, according to [93]. The EEG signal is composed of various bands, with different frequency intervals and, also, it presents abnormalities in neurologic patients, which increases the difficulty of the determination of a maximum frequency. In this way, a high value for sampling rate is required for the decrease in the loss of information.

From literature review, different sampling rate values were established. The interval encountered varies between 100 Hz [94; 95] to 4096 HZ [96], which most common sampling frequencies are 173.6 Hz [97; 98; 99; 100; 101; 102; 103; 104; 105; 106; 107], 250 Hz [108; 109; 110; 111; 112; 113; 114] and 256 Hz [115; 116; 117; 118; 119; 120]. These sampling values correspond to the implementation of one of the three public databases mentioned in Section 3.4, with a uniform data recording. It is important to refer no data resampling was applied.

3.2 Preprocessing methods

As mentioned in Section 2.2.1.1, electroencephalography recordings are prone to be disturbed by noise and artefacts that affect the quality of the signal. Some of the most common artefacts present in an EEG signal are power line interference, other physiological signals, body movement, among others. Thus, to properly remove this noise interference and to improve the quality of data, there are some methods established to correctly perform the denoising task. In addition to this, other preprocessing methods are applied to EEG signals to prepare them for the feature extraction phase, to obtain an improved generalization performance. Two of the most common techniques applied to the raw signal are data normalization and signal segmentation.

3.2.1 Filtering

Filtering is a very simple procedure, but it shows great efficiency in denoising the EEG signal, maintaining only relevant information. For this purpose, bandpass is the preferential filter applied to remove high and low-frequency interferences and to maintain only EEG frequency bands (alpha, beta, theta, gamma, delta) and it was the preprocessing strategy most common from the literature review. Even though EEG signals present relevant information in a frequency interval between 0.5-100 Hz, high-frequency interferences, like the power line, can disrupt the signal. That is why the most recurrent cutoff interval is 0.53-40 Hz [97; 98; 99; 100; 101; 102; 103; 104; 105; 106; 107]. Besides that, other intervals were considered (0.3-45 Hz [94], 0.5-35 Hz [121] or 0.5-70 Hz [122]). Beyond this, in some cases, the high cutoff frequency used surpassed the 60 Hz [123; 124; 122], in order to obtain more gamma sub-band content, so a notch filter was applied to remove power line interference specifically. Application of low pass and high pass filters for noise removal was also implemented.

3.2.2 Data Normalization

Data Normalization is a technique to even the amplitude of all the EEG signals acquired. The power of the EEG signal recorded varied between subjects and acquisition sessions, due to various factors like attachment of the electrodes, electrode position on the skull, anatomical characteristics of different subjects, among others. For this reason, normalizing EEG records is fundamental to level the significance of all signals. With normalization, the signals are all rescaled and the baseline is aligned to the same DC offset (usually zero). Different normalization techniques were applied like Min-Max normalization, implemented by *Bongiorni et al* [115], *Huang* [98] and *Fukumori et al* [125], or Standard deviation normalization [114].

3.2.3 Signal Segmentation

Signal Segmentation is one of the most common techniques for the preparation of EEG signals for feature extraction. Ictal and Interictal events occur sporadically and spontaneously, being a small part of the whole EEG record. To guarantee the presence of these episodes, it is necessary to record several hours of EEG. For this purpose, in order to keep just relevant segments of the EEG, and to even the presence of normal and abnormal signal segments, for class balance, it is necessary to do signal segmentation. Besides that, it can also be applied as a data augmentation technique, through overlapping segmentation, using a sliding window, which allows increasing the number of samples of the class with fewer instances. The segmentation strategy is an important tool to find a trade-off between the computational load of the input sequences and the good performance of the model. For several articles reviewed, EEG signals were divided into small samples, from 1 to 5 second EEG segments [115; 116; 98; 99; 104; 121; 126; 96; 125; 118; 120; 94; 127; 128; 102]

3.2.4 Downsampling

Another technique to decrease data size to reduce computational time and load and to improve the deep model's efficiency is called downsampling. Downsampling a signal is similar to decreasing the sampling rate of the segment,i.e., the data points present in the total sample [129; 130]. The optimal downsampling frequency is reducing the amount of data that the deep network has to deal with, maintaining a good representation of the EEG patterns' main features. This has been implemented as an alternative to signal segmentation or as a complementary strategy. [115; 121; 122; 123; 124]

3.3 Feature Extraction Methods

EEG signals are abundant in features that characterize it, making it a very rich signal. So, even though deep architectures are capable of automatically extract features from the raw signals, the prior extraction and introduction of these features as input in deep frameworks has been vastly experimented. These features can be divided into two main domains, frequency and time domains. For the acquisition of these features, some procedures and processing algorithms were implemented. One of the objectives of feature extraction was related to the refinement of the data used as input, decreasing the amount that is introduced and increasing the relevance of the data extracted.

3.3.1 Time Domain Features

Several articles reviewed implemented algorithms for the extraction of time domain features. The features chosen for extraction vary between surveys. *Bongiorni et al* [115] and *Khorasani et al* [128] decided to calculate various features like the root mean squared error between the reference and each sample, standard deviation and the number of peaks above a certain threshold, average amplitude and duration of each EEG sample and the coefficient of variation. *Ahmedt-Aristizabal et al* [108] worked with several time features, such as statistical moments (Mean Value, Variance, skewness and kurtosis), Standard Deviation, Zero Crossings, Peak-to-Peak Voltage and Total Signal Area.

Cross and Auto Correlation is an important measure. From this technique, not only the outcome was used as an input feature, but also some features were extracted from it, like peak value of the cross correlogram, the peak value, centroid, equivalent width and mean square abscissa [108; 105]. The work developed by *Gayatri et al* [131] implemented a new time domain feature called ApEn [132], capable of classifying complex systems, providing a useful statistic to distinguish datasets in terms of their amounts of regularity, more concretely between normal and abnormal EEG. Aside from this, several articles used as input the raw EEG signals in the time domain, without any feature extraction [121; 133; 97; 98; 99; 101; 118; 134; 114; 102; 103; 96; 122; 119; 113].

3.3.2 Frequency domain features

Along with time domain features, various researches are considered as inputs frequency domain features of EEG signals. To obtain properties from the frequency domain, the signal must be transformed into this domain. Some feature extraction techniques were presented in the surveys reviewed, such as Fast Fourier Transform [115; 128; 108; 117; 119], Short Time Fourier Transform [112], Discrete Fourier Transform [105], Continuous Wavelet Transform [95], Discrete Wavelet Transform [125; 106; 108; 133], among others. In addition to this, some papers applied a feature extraction algorithm called standard linear frequency cepstral coefficient-based (LFCC), popularized in some applications such as speech recognition [135; 136]. This technique was implemented by *Von Weltin et al* [109], *Golmohammadi et al* [110] and *Shah et al* [111].

The application of Wavelet Transform had the goal of acquiring several frequency sub-bands, which were used as inputs or instances to obtain some basic statistics, like average power, mean, entropy and standard deviation of the wavelet coefficients in each sub-band. Another technique to obtain frequency signal sub-bands is the application of Fast Fourier Transform. In addition to this, with the application of Fast Fourier Transform is possible to compute the Power Spectral Density, a source of spectral features. Regarding PSD applications, *Ubeyli et al* [104] introduced standard PSD algorithms (Pisarenko, MUSIC and Minimization-Norm), calculated their logarithms and processed features values (the maximum, minimum, standard deviation and mean from the power levels of the PSD results).

3.3.3 Alternative methods

In this subsection, some algorithms to obtain features aside from time and frequency domains are presented. *Guler et al* [100] proposed the employment of Lyapunov exponents to EEG samples and, with the outcome of these exponents, some features are calculated, such as Mean of the absolute values, the maximum of the absolute values, average power and standard deviation of the Lyapunov exponents in each EEG segment. Another method proposed by *Kumar et al* [137] used Wavelet Entropy and Spectral Entropy as feature inputs. Wavelet Entropy [138] is a metric of the order-disorder degree of the EEG signal and Spectral Entropy [139] makes use of the power spectrum's amplitude components of a given signal as probabilities for entropy calculations.

3.4 Databases

As mentioned in sections 2.3.6.2 and 2.3.5, to obtain a fine tune of the deep model, with a good generalization ability and classification performance, it is necessary the supply a proper amount of data, well structured and clean. Thus, the creation of a proper database is fundamental for the success of the study. Through the literature review, it was possible to encounter some public databases that were used on a recurring basis for the automation of seizure detection. The advantage of using the same database in different research projects is that it allows the comparison between the results achieved.

The first public database is named CBH-MIT scalp EEG [140; 141], gathered by a group of Massachusetts Institute of Technology and Children's Hospital Boston. This database contains long term bipolar referenced scalp EEG records of 23 children with intractable epileptic seizures. Each record was obtained using the 10–20 International System for electrodes, using 256 samples per second with a 16 bits ADC. In most of the 23 cases, the EEG signals are segmented in 1-h long epochs, while epochs of 2–4 h in duration can also be found. The total duration of the available EEG recordings is about 983 h. This database was used to implement different Recurrent Neural Networks for epilepsy detection [115; 116; 117; 118; 119; 120].

The Department of Epileptology of the University of Bonn published an EEG database [142]. The database includes five sets (named from A to E) with a total of 100 EEG samples for each set. Each sample is a single-channel EEG recorded at 173.6 Hz of sampling rate, with 23.6 seconds of duration. Thus, the sample length of each sample is 4096. Sets A and B were recorded using scalp EEG from five healthy subjects, with eyes open for set A and eyes closed for set B. Set C, D and E, were recorded from five epileptic patients prior to surgery, which was recorded using depth electrodes implanted symmetrically into the hippocampal formation. Sets C and D were recorded during seizure-free intervals, where set D was recorded from the epileptogenic zone (Interictal state) and set C from the opposite brain hemisphere (Interictal state). Set E contains records of the epileptogenic zone during an epileptic seizure (Ictal state). This database was used by *Ahmedt et al* [97], *Huang et al* [98], *Hussein et al* [99], *Guler et al* [100],*Abdelhameed et al* [101], among others [102; 103; 104; 105; 106; 107].

Another public database was released by the Temple University Hospital, from Philadelphia [143]. From the TUH EEG database, a subset was considered, the TUH EEG Seizure Corpus, which was developed for seizure detection's investigation. The acquisition was made at 250Hz of sampling rate and contain the standard channels of a 10-20 configuration. The seizure corpus contains surface EEG data of 817 recording sessions, of which 305 are sessions that contain seizures, resulting in a total of 2012 seizures with different lengths and categorizes in eight classes: Focal Non-Specific Seizure (FNSZ), Generalized Non-Specific Seizure (GNSZ), Simple Partial Seizure (SPSZ), Complex Partial Seizure (CPSZ), Absence Seizure (ABSZ), Tonic Seizure (TNSZ), myoclonic seizures (MCSZ) and Tonic Clonic Seizure (TCSZ). Some seizures of the same patient are categorized with different seizure types. This database was used in several research projects, like *Ahmedt-Aristizabal et al* [108], *VonWeltin et al*, [109], *Golmohammadi et al* [110], *Shah et al*

[111] and *Liu et al* [112].

Besides all the public databases available, there are still various research projects that use a private database created in accordance with the goals of the study [121; 133; 144; 134; 114; 131; 95; 96; 122; 123; 124; 125].

3.5 Classification Architectures

After the EEG acquisition, preprocessing of signals and feature extraction, the architecture and hyperparameters of the deep neural network can play a fundamental role in the success of the EEG signals classification. Signal analysis can focus on automated detection of seizure patterns or in transient patterns in seizure-free segments, like interictal epileptiform discharges. The subsequent sections will focus on the analysis of different Recurrent Neural Network implementations for the automatic analysis of EEG patterns and consequent results and conclusions. The division of the studies is made according to the database used, for comparison purposes.

3.5.1 Proposed Architecture with CBH-MIT database

As mentioned in Section 3.4, the CBH-MIT database resulted from a cooperation between Children's Boston Hospital and Massachusetts Institute of Technology and contains EEG records from 23 epileptic children. This database was used, by all the research projects reviewed, in a two-class classification, between seizure and non-seizure, pre-ictal and interictal or ictal and interictal events. *Thodoroff et al* [117] and *Vidyaratne et al* [118] tested various deep architectures for the classification between seizure and non-seizure activity. The first implemented a hybrid Recurrent Convolutional Neural Network, concatenating 3D CNN layers, 3D Max Pooling Layers with a bidirectional Long Short-term Memory layer and a Fully Connected layer to capture spatial, temporal and frequency domain patterns that represent a seizure. On the other hand, the latter implemented a combination of an Artificial Cellular Neural Network [145], in which each cell of the network is composed of a bidirectional LSTM layer, with a Multilayer Perceptron Network. This ACNN allows a unique mapping of seizure EEG signal for efficient extraction of spatial and temporal features. For *Thodoroff et al*, the performance of the network was evaluated for patient-specific and cross patient detection, while *Vidyaratne et al* only performed patient-specific detection. The performance was assessed using sensitivity and false positive rate (per hour). Sensitivity, for patient-specific detection, achieved was 95-100% and 100.0% and false positive rate was $0.3h^{-1}$ and $0.08h^{-1}$, respectively.

Looking at pre-ictal vs interictal classification, *Tsiouris et al* [116] implemented three LSTM networks, with a different number of layers and recurrent units to exploit a wide range of features extracted before patient-specific classification, from time and frequency domain features, to cross-correlation between EEG channels and graph theoretic features. The three architectures were concatenated with Fully Connected and a final Dense layers. Besides that, various pre-ictal window sizes, from 15 min to 2 hours, and a segment-based and event-based performance were tested. For segment-based evaluation, average sensitivity and specificity from all networks were

99.3% and 99.3%, 99.4% and 99.6%, 99.6% and 99.8%, 99.8% and 99.9%, for 15min, 30min, 1 hour and 2 hours of pre-ictal window sizes, respectively. For event-based evaluation, sensitivity and false positive rate per hour were chosen as evaluation metrics. The results for 15 min, 30 min, 1 hour and 2 hours was 100.0% for all and 0.107, 0.063, 0.032 and 0.2 h^{-1}, respectively. On the other hand, *Daoud et al* [120] implemented different deep frameworks, two of them a Dense Convolutional Neural Network concatenated with a Bidirectional LSTM layer and a Deep Convolutional Stacked Autoencoder (DCAE) concatenated with a Bidirectional LSTM network. For patient-specific detection, sensitivity and false positive rate per hour were 99.7% and 0.004 h^{-1} for both methods, respectively.

Lastly, *Bongiorni et al* [115] implemented a Recurrent Neural Network composed by LSTM layers for the classification between ictal and interictal segments, with overall accuracy of 97.4% , 19.7% for sensitivity and specificity of 99.7%. The proposed network was not able to detect efficiently pre-ictal periods.

For all the works where the CBH-MIT database was implemented, only two-class classification was performed. All the networks implemented were a hybrid combination between Convolutional and Recurrent Neural Networks, combining the advantages of both types of architectures, resulting in overall high performance metrics. The combination of both types of deep architectures allows the capturing of spatial, temporal and frequency features that characterize a seizure EEG signal, which gives a wider feature space. In Table 3.1, it is possible to overview the results obtained with this database.

3.5.2 Proposed Architecture with Database from the Department of Epileptology, University of Bonn

The database from the Department of Epileptology is composed of 5 sets (A to E), two with non-seizure activity EEG samples (A and B), two with interictal patterns (C and D) and one with samples of seizure/ictal activity (E), as mentioned in Section 3.4. In this way, from the articles reviewed, two, three and five class classification was performed when this database was utilized.

Regarding exclusively two-class classification, *Ahmedt-Aristizabal et al* [97], *Huang et al* [98], *Srinivasan et al* [126] implemented three types of Recurrent Neural Networks, a Long Short-Term Memory network, a Stacked Sparse Autoencoder and an Elman Neural Network [146]. The first two studies implemented more complex architectures to automatically extract temporal features and learn more abstract and high level representations, while *Srinivasan et al* used techniques to prior extract frequency and temporal features to introduce on a simpler architecture. Both approaches obtained high results, with sensitivity, specificity and overall accuracy reaching >90% values. *Guler et al* [100] implemented also an Elman Neural Network to test a three-class classification (normal vs interictal vs ictal events) of this model, with overall Accuracy equals 96.8%, Specificity (healthy segments) equals 97.4% and Sensitivity for interictal segments and ictal segments equals to 96.9% and 96.1%, respectively. This study implemented Lyapunov exponents as input instances. Later in 2018, a Deep Convolutional Bidirectional LSTM Neural Network was applied by *Abdelhameed et al* [101] to perform two and three class classification on raw EEG data, with overall results (Accuracy, Sensitivity, Specificity) above 98.0%.

To perform a two, three and five class classification,*Hussein et al* [99] created a hybrid LSTM network concatenated with a Fully Connected, Average Pooling and a Dense layers, with excellent results at all classifications, achieving 100.0% for test accuracy, test sensitivity and test specificity. Besides that, it was studied the performance of the architecture with noise addition (eyes movement, muscle activities and white noise), to simulate imperfect acquisition conditions, with results traducing robustness against common EEG artifacts. To implement a five-class classification, *Ubeyli et al* [104] built an Elman Neural Network, which achieved an average accuracy of 98.2%, the sensitivity of 98.2% and specificity of 99.0%. Contrarily, *Dutta et al* [102] implemented two layer LSTM, GRU and Vanilla Recurrent Neural Network, but overall accuracy obtained was poor when compared with other studies, specifically 69.5%, 73.1% and 45.3% for each one of the networks.

The architectures applied to this database are simpler, most of them without hybrid combinations between RNN and CNN models. First applications of sequence models relied on simpler architectures with the implementation of feature extraction methods and, afterward, more complex models were introduced to deal with raw EEG data, relying on the deep frameworks the ability to learn high representations of the input data. Overall results correspond to a very good tuning of parameters and a generalization ability, which corroborate the capacity of Recurrent Neural Networks to deal with sequence time instances, being capable of distinguishing between the rich patterns present in an EEG signal. At Table 3.2 all results are presented.

3.5.3 Proposed Architectures with TUH EEG database

The TUH EEG database is the world's largest publicly available database of seizure recordings, released by Temple University Hospital. From this database, the subset TUH EEG Seizure Corpus was developed to allow the research on seizure detection. A particular characteristic of this subset is the presence of eight types of seizure, clearly labeled, so seizure type classification experiments are also possible to perform.

Most of the scientific studies reviewed performed a two-class classification, between non-seizure and seizure samples. *Shah et al* [111] and *VonWeltin et al* [109] test a variety of Deep Recurrent Convolutional Neural Networks. In addition to this, *Golmohammadi et al* [110] decided to implement, in parallel with a RCNN, a Hidden Markov Model strategy. HNN is a powerful statistical modeling tool to extract both time and frequency domain features [147]. This HNN procedure was concatenated with a Long Short-Term Memory and Sparse Denoising Autoencoder models. These studies reached similar results, with low values for sensitivity (around 30-35%), while specificity reached higher values (around 85-100%). Moreover, a false positive rate attains an interval of 6-38 per 24 hours. According to [148], Human performance on EEG classification is in the range of 65.0% sensitivity with a false alarm rate of 12 per 24 hours. Thus, the aforementioned architectures did not meet the clinical requests and it is needed the search for other training procedures to improve models' performance.

Regarding seizure type classification, *Ahmedt-Aristizabal et al* [108], implemented two deep frameworks. The first architecture was a hybrid LSTM model with the application of an external memory model. This external memory model is constituted by three controllers: a read controller to query the knowledge stored in the memory, an update controller to update the memory with new knowledge, and an output controller that controls what results are passed out from the memory [149]. The second model is a standard Recurrent Convolutional Neural Network. The weighted F1-score obtained for both networks is, respectively, 0.95 and 0.82, giving rise to an alternative for seizure type detection models, with the application of neural plasticity. Furthermore, *Liu et al* [112] developed three alternatives using bilinear models. The bilinear model consists of two feature extractors whose outputs are multiplied using matrix outer product at each location, then pooled to obtain a high level descriptor. The first option implements ConvLSTM cells in both feature extractors, the second a CNN layer and the third one use, at each feature extractor, ConvLSTM and CNN layers. The evaluation metric chosen was F1-score. For the three networks, this performance measure reached 96.9% for bilinear-RNN, 96.7% Bilinear-CNN and 97.4% for the hybrid bilinear architecture. Using both RNN types potentiate the performance of the architecture, by exploring a wider feature space that characterizes the EEG signals. All results can be seen in Table 3.3.

3.5.4 Proposed Architecture with Private Databases

Not all research projects introduced a public database for their development. Even so, they present alternatives of sequence model architectures that can be analyzed by their efficiency in automated seizure detection. In 1998, a Recurrent Multilayer Perceptron was implemented by *Petrosian et al* [133] to perform two-class classification between seizure and non-seizure signals. The database was built by records of two patients who were undergoing long-term electrophysiological recordings. The database consisted of four pre-ictal single-channel recordings (I1, I2, I3, I4) and three non-seizure recordings (N1, N2, N3). Performance assessment was made with intra and extracranial EEG signals, with poor results using raw data as input, but, using wavelet decomposed signals, performance quality increased.

Later on 2018, *Tjepkema-Cloostermans et al* [121] developed a CNN and hybrid CNN+LSTM architectures, with a different number of layers, to properly identify interictal epileptiform discharges from a database with 50 EEGs that were classified as normal and 50 EEGs that contained focal IEDs. The 2DCNN and 2D CNN+LSTM architectures, which reached the better results, obtained an AUC of 0.94. For the EEGs with IEDs, the average value of sensitivity for all architectures was 47.4% (range 18.8–78.4%), with the specificity of 98.0% (range 93.8– 99.8%), which is associated with a false positive rate of 0.60 min^{-1}. For the normal EEGs, the average specificity of all models was 99.9% (range 98.3–100.0%) and a false positive rate of 1–2 misclassified samples per hour.

With EEG records of six patients with benign epilepsy with centrotemporal Spikes, *Fukumori et al* [125] developed two RNN architectures, stacked layers of LSTM or GRU concatenated with a Fully Connected Layer. The performance assessment was made taking into consideration the type of gated RNN implemented and, beyond this, the feature extraction method was studied. The first method consisted of a bandpass filter and Discrete Wavelet Transform while the second comprised a CNN layer. Overall performance was evaluated with AUC values, which reach an interval of 0.80-0.85. Lastly, *Roy et al* [113] implemented some variations of Deep RCNN using a subset of TUH EEG database, TUH abnormal EEG corpus subset, with the inclusion of multiple filters of exponentially varying lengths in the 1D convolution layers and dense connections within the GRU layers. The accuracy obtained for all the architectures was above 80.0%, for both training and testing phases. All results can be seen in Table 3.4.

3.6 Summary and Conclusion

Through the analysis of forty-one collected research publications, it was possible to draw conclusions and guidelines to aid the development of the dissertation. First, an overview of three public databases that are largely used in automated epilepsy pattern detection is performed. The use of the same input instances allows the comparison of performance between different deep frameworks. Regarding acquisition settings, the standard recording strategy is explained and all the issues that compromise the success of EEG signal recording.

As for preprocessing methods, various denoising and signal preparation techniques are explained, giving attention to the advantages and drawbacks of the multiple strategies gathered. Regarding features, frequency and time domain attributes are widely explored achieving good results that justify the implementation of feature extraction methods for the development of input features. On the other hand, procedures like dimensionality reduction are underexplored and might be a method with promising performance. Lastly, looking at deep architectures explored throughout the various studies, the main focus is given to Recurrent-based neural networks, because it will be a cornerstone subject of this dissertation. Overall, results show a fair application of a variety of architectures, from more standard and gated RNNs, until more complex hybrid solutions, as the implementation of Hidden Markov Models or a neural memory controller, among others.

Table 3.1: Results of proposed approaches evaluated with the CBH-MIT database

Research	**Models' Architecture**	**Results**
Thodoroff [117]	RCNN - 2x (3DCNN + 3DCNN + 3DMaxPooling) + FV + Bi-LSTM + FC.	Patient Specific evaluation: • Average sensitivity of 95-100% ; • Average false positive rate 0.3 h^{-1}. Cross-patient evaluation: • Average sensitivity 85.0%; • Average false Positive Rate 0.8 h^{-1}.
Vidyaratne [118]	ACNN (Bi-LSTM at each cell) + MLPNN.	• Sensitivity of 100.0%; • False positive rate of 0.08 h^{-1}; • Detection Delay of 7 $\pm$2.7 seconds.

Tsiouris [116]	• 1^{st} Architecture - 1x LSTM (32 units) + FC + Dense; • 2^{nd} Architecture - 1x LSTM (128 units) + FC + Dense; • 3^{rd} Architecture - 2x LSTM (128 + 128 units) + FC + Dense.	Pre-ictal Window - 15 min: • Segment-based evaluation: Average sensitivity of 99.3% / Average specificity of 99.3%; • Event-based evaluation: Average sensitivity of 100.0% / Average FPR of $0.107h^{-1}$. Pre-ictal Window - 30 min: • Segment-based evaluation : Average sensitivity of 99.4% / Average specificity of 99.6%; • Event-based evaluation : Average sensitivity of 100.0% / Average FPR of $0.063h^{-1}$. Pre-ictal Window - 1 hour: • Segment-based evaluation : Average sensitivity of 99.6% / Average specificity of 99.8%; • Event based evaluation : Average sensitivity of 100.0% / Average FPR of $0.032h^{-1}$. Pre-ictal Window - 2 hour: • Segment-based evaluation : Average sensitivity of 99.8% / Average specificity of 99.9%; • Event based evaluation : Average sensitivity of 100.0% / Average FPR of $0.02h^{-1}$.

Daoud [120]	DCNN + Bi-LSTM - Dense Convolutional Neural Network concatenated with a Bidirectional LSTM; DCAE + Bi-LSTM - Deep Convolutional Stacked Autoencoder concatenated with Bidirectional LSTM.	DCNN + Bi-LSTM / DCAE + Bi-LSTM: • Sensitivity 99.7%; • Specificity 99.6%; • Accuracy 99.7%; • FPR $0.004h^{-1}$.
Bongiorni [115]	3x LSTM layer + Dense layer.	• Accuracy 97.4%; • Sensitivity 19.7%; • Specificity 99.7%.

Table 3.2: Results of proposed approaches evaluated with the database from University of Bonn

Research	**Models' Architecture**	**Results**
Ahmedt - Aristizabal [97]	LSTM layer + Dense layer; LSTM layer + LSTM layer + Dense layer.	• Average accuracy of 91.3%; • Average sensitivity of 91.8%; • Average specificity of 90.5%; • Average precision of 91.5%; • Average AUC of 0.96.
Huang [98]	3-hidden-layer Stacked Sparse Autoencoder.	• Average accuracy of 96.0%; • Average sensitivity of 98.7%; • Average specificity of 93.8%.
Guler [100]	Elman Recurrent Neural Network.	• Total Classification Accuracy - 96.8%; • Specificity (healthy segments) - 97.4%; • Sensitivity (Interictal segments) - 96.9%; • Sensitivity (Ictal segments) - 96.1%.

Srinivasan [126]	Elman Neural Network.	Training Scheme 1 (One feature at a time): • Sensitivity 99.4%; • Specificity 99.4%; • Accuracy 98.9%. Training Scheme 2 (Two features each training time): • Sensitivity 96.7%; • Specificity 96.3%; • Accuracy 96.5%. Training Scheme 3 (Three features each training time): • Sensitivity 98.4%; • Specificity 95.9%; • Accuracy 97.1%. Training Scheme 4 (Four features each training time): • Sensitivity 98.2%; • Specificity 95.6%; • Accuracy 96.9%. Training Scheme 5 (Four features at once): • Sensitivity 99.6%; • Specificity 94.4%; • Accuracy 97.0%.

Abdelhameed [101]	4x CNN + 3x MaxPooling layers + Bi-LSTM.	Two class classification (Normal vs Ictal EEG segments): • Sensitivity 100.%; • Specificity 99.0%; • Accuracy 99.5%. Three class classification (Normal vs Inter-Ictal vs Ictal EEG segments): • Sensitivity 97.8%; • Specificity 98.9%; • Accuracy 98.5%.
Hussein [99]	LSTM + FC + AP layers.	Two-class /Three-class /Five-class EEG classification: • Average Specificity 100.0%; • Average Sensitivity 100.0%; • Average Accuracy 100.0%.
Ubeyli [104]	Elman Neural Network.	• Average Specificity 99.5%; • Average Sensitivity 98.2%; • Average Accuracy 98.2%.

Dutta [102]	1^{st} Arch - 2x Vanilla RNN layers; 2^{nd} Arch -2x LSTM layers; 3^{rd} Arch -2x GRU layers.	• Vanilla RNN Accuracy 45.3%; • LSTM Accuracy 69.5%; • GRU Accuracy 73.1%.

Table 3.3: Results of proposed approaches evaluated with the TUH EEG seizure corpus database

Research	**Models' Architecture**	**Results**
Shah [111]	3x(2DCNN + 2DMaxPooling) + FC + 2xBi-LSTM.	• Specificity 85.5%; • Sensitivity 33.4%; • False Positive Rate (per 24hours) 38.19.
Ahmedt - Aristizabal [108]	1^{st} Arch - 2x LSTM + External Memory Model + Dense; 2^{nd} Arch - 2x (2DCNN + 2DMaxPooling) + FC + 2xLSTM + Dense.	1^{st} Architecture: • Weighted F1-score 0.95. 2^{nd} Architecture: • Weighted F1-score 0.82.

Von Weltin [109]	1^{st} Arch - 3x (2DCNN + 2DMaxPooling) + 1DCNN + 1MaxPooling + Bi-LSTM; 2^{nd} Arch - 3x (2DCNN + 2DMaxPooling) + 1DCNN + 1MaxPooling + Bi-GRU.	1^{st} Architecture (CNN/LSTM): • Sensitivity 97.1%; • Specificity 30.8%; • False Positive Rate (per 24hours) 6. 2^{nd} Architecture (CNN/GRU): • Sensitivity 91.5%; • Specificity 30.8%; • False Positive Rate (per 24hours) 21.
Liu [112]	1^{st} Arch - Bi-linear CNN model; 2^{nd} Arch - Bi-linear LSTM model; 3^{rd} Arch - Hybrid Bi-linear CNN/LSTM model.	Bi-linear CNN: • F1-score 96.7%. Bi-linear LSTM: • F1-score 96.9%. Hybrid Bi-linear CNN/LSTM: • F1-score 97.4%.

Golmohammadi [110]	1^{st} Arch - HNN + SDA; 2^{nd} Arch - HNN + LSTM; 3^{rd} Arch - 3x (2DCNN + 2DMaxPooling) + 1DCNN + 1MaxPooling + LSTM.	1^{st} Architecture (HNN/SDA): • Specificity 73.4%; • Sensitivity 35.4%; • False Positive Rate (per 24hours) 77. 2^{nd} Architecture (HNN/LSTM): • Specificity 80.5%; • Sensitivity 30.1%; • False Positive Rate (per 24hours) 60. 3^{rd} Architecture (CNN/LSTM): • Specificity 96.9%; • Sensitivity 30.8%; • False Positive Rate (per 24hours) 7.

Table 3.4: Results of proposed approaches evaluated with private databases

Research	**Database Description**	**Models' Architecture**	**Results**
Tjepkema - Cloostermans [121]	Database is constituted with 50 EEGs records classified as normal and 50 EEGs with focal interictal epileptiform discharges.	1^{st} Arch - 6x (2DCNN + 2DMaxPooling) + 3xFC; 2^{nd} Arch - 6x (LSTM + 2DMaxPooling) + 3xFC.	For the samples with IEDs: • Average specificity 47.4%; • False positive rate $0.60 min^{-1}$. For the normal EEG samples: • Average specificity 99.9%; • False positive rate 1–$2h^{-1}$.
Petrosian [133]	The EEG data used in this study were obtained from two patients who were undergoing long-term electrophysiological monitoring for epilepsy. Four preictal single-channel recordings (I1 , I2 , I3 , I4) and three non-seizure recordings (N1,N2,N3).	Discrete-time Recurrent Multilayer Perceptron.	Results for intracranial: • Performance was poor with raw data, but, using wavelet decomposed signals, the results were better. Results for extracranial: • Better network performance with using both high pass and low pass subband signals.

Fukumori [125]	EEG records from six patients with Benign Epilepsy with Centro-Temporal Spikes.	1^{st} Arch - 4xLSTM + FC; 2^{nd} Arch - 4xGRU + FC.	1^{st} Architecture (LSTM): • BP + DWT - Average AUC 0.85; • 1DCNN - Average AUC 0.83. 2^{nd} Architecture (GRU): • BP + DWT - Average AUC 0.82; • 1DCNN - Average AUC 0.83.
Roy [113]	TUH Abnormal EEG corpus subset.	1^{st} Arch - A Multilayer Neural Network with stacked 1DCNN and GRU layers (CRNN); 2^{nd} Arch - The Inception Convolutional Gated Recurrent Neural Network modifies the CRNN including multiple filters of exponentially varying lengths in the 1DCNN layers (ICRNN); 3^{rd} Arch - The Convolutional Densely Connected Gated Recurrent Neural Network architecture includes dense connections in the recurrent stage (CDRNN); 4^{th} Arch - Combination of ICRNN and CDRNN (ChronoNet).	CRNN: • Test Accuracy 82.3%. ICRNN: • Test Accuracy 84.1%. CDRNN: • Test Accuracy 83.9%. ChronoNet: • Test Accuracy 86.6%.

Chapter 4

Methods

4.1 Databases

4.1.1 EEG data

The original EEG data is comprised of two-second EEG epochs with two types of EEG patterns. Normal EGGs, from 67 healthy subjects without any epileptic episodes and interictal epileptiform discharges from focal (50 patients) and generalized (49 patients) epilepsy. The labeling of EEG data was done by experts, to correctly identify the presence or absence of IEDs in EEG records. Samples in which there was uncertainty concerning the presence of interictal epileptiform discharges were discarded, to ensure that all the annotations correspond undoubtedly to IED patterns. The EEG recordings were performed according to the 10/20 international electrode system, using twenty-one silver/silver chloride cup electrodes.

4.1.2 Pre-processing of data

As pre-processing steps, the data were filtered in the 0.5-35 Hz range to reduce artifacts and downsampled to 125 Hz to reduce input size (and consequently computational load). The acquisition of the EEG signals was performed under longitudinal bipolar electrode montage and, afterward, the signals were converted to the common average and source derivation montages. The three montages were used to correctly label the EEG records at visual assessment, increasing effectiveness in the search of IED patterns. After that, each recording was split into 2 s non-overlapping epochs, yielding an 18x250 (channels x time) matrix for each epoch.

4.1.3 Database Structuring

During this dissertation, several deep neural networks were experimented with to explore a two-class classification problem, for IED detection. The positive class, labeled as 1, was composed of Focal and Generalized epilepsy records and normal EEG samples as negative class elements. The first dataset created, Set A, was composed of original EEG records mentioned above. This set was used for almost all of the different trials performed during the dissertation. At a final stage, another set was structured, named Set B, constituted by the same signals present in Set A, but with the records from the three different electrode montages. This set was created as an alternative dataset with more positive samples, which means it is a less imbalanced dataset. Besides this alternative, other two datasets were created in order to surpass the data imbalance, Set C and Set D, with over-sampling of positive class samples by applying Synthetic Minority Over-sampling Technique (SMOTE)[150] technique and data overlap by 70%, throughout Sliding Window technique, respectively. All datasets are described in Table 4.1.

Table 4.1: Databases description (number of epochs for train, validation and test)

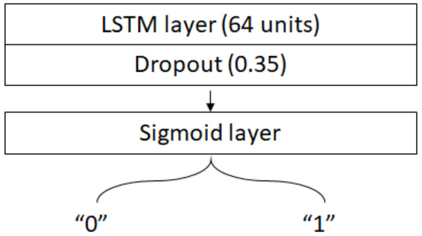
Dataset	Train & Validation			Test		
	Epochs		Duration (h)	Epochs		Duration (h)
	Negative	Positive		Negative	Positive	
Set A	43589		24.2	15716		8.7
	41369	2220		15414	302	
Set B	145709		80.9	21285		11.8
	139373	6336		19900	1385	
Set C	88320		49.1	15716		8.7
	44160	44160		15414	302	
Set D	50249		27.9	15716		8.7
	41369	8880		15414	302	

The data were randomly split into train, validation and test sets. 80% of the data was used to train and validate all the deep neural networks explored, and 20% was used to test the models. Afterwards, in order to decrease the difference between positive and negative samples, the negative samples present in the set for train and validation were divided into 5 different parts, while all the positive samples were repeated in all the 5 parts. In the end, five-cross validation was implemented to each part and the evaluation results were calculated as an average of the training to the five parts.

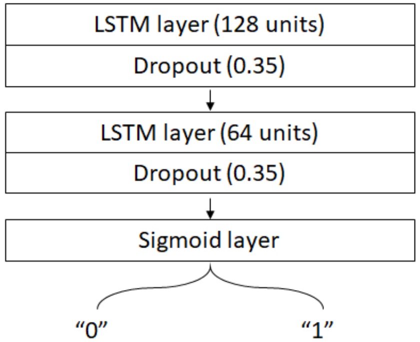
4.2 Deep Learning Models

4.2.1 Recurrent Neural Network

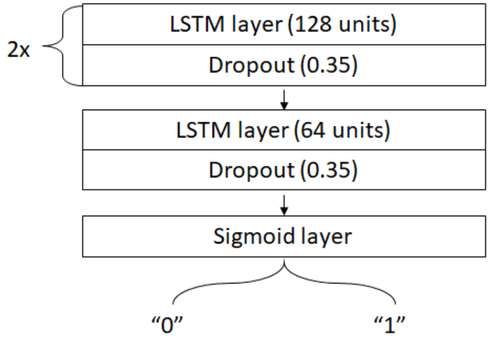
Several Recurrent Neural Networks were designed and explored during this dissertation, modifying the hyperparameters and extent in the complexity of all networks to encounter the best architecture to correctly detect IED samples. The architecture of the RNN consisted of stacked Long Short-term Memory layers with a final sigmoid layer for the classification task. Concerning the number of layers, experiments were performed from 1 to 7 LSTM layers, where the number of units composing each layer varied between 64, 128 and 256 units. Associated to each layer, Dropout regularization and Batch Normalization were applied, simultaneously and separately. The dropout rate changed from 0 to 0.35. Various experiments were done, combining all these variables. Not all parameters were combined, with exploration having been done iteratively according to the results that were obtained through the experiments performed. The experiments are explained in more detail later. A scheme of the Recurrent Neural Network architectures can be seen in Figures 4.1a, 4.1b, 4.1c, 4.2a, 4.2b and 4.2c.



(a) LSTM1 architecture

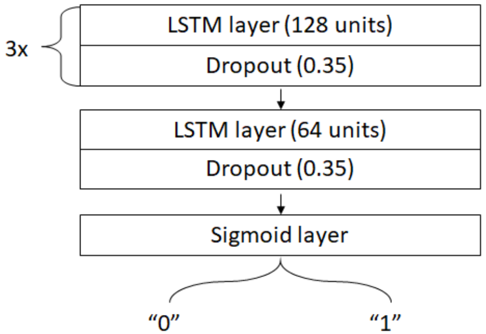


(b) LSTM2 architecture

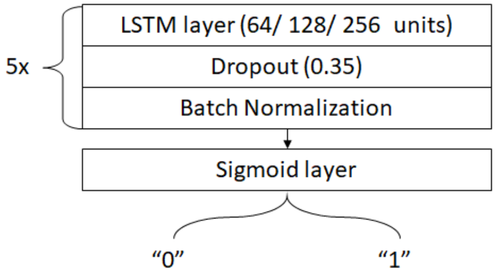


(c) LSTM3 architecture

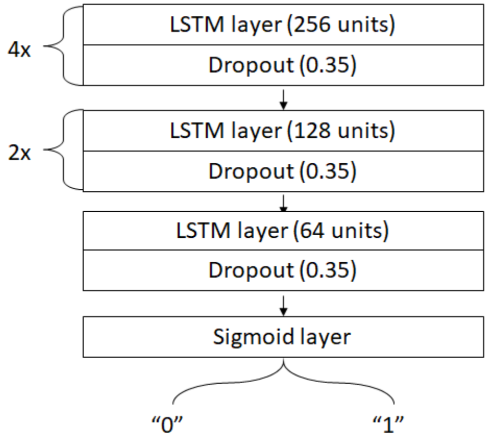
Figure 4.1: LSTM1, LSTM2 and LSTM3 models’ architecture



(a) LSTM4 architecture



(b) LSTM5 architecture

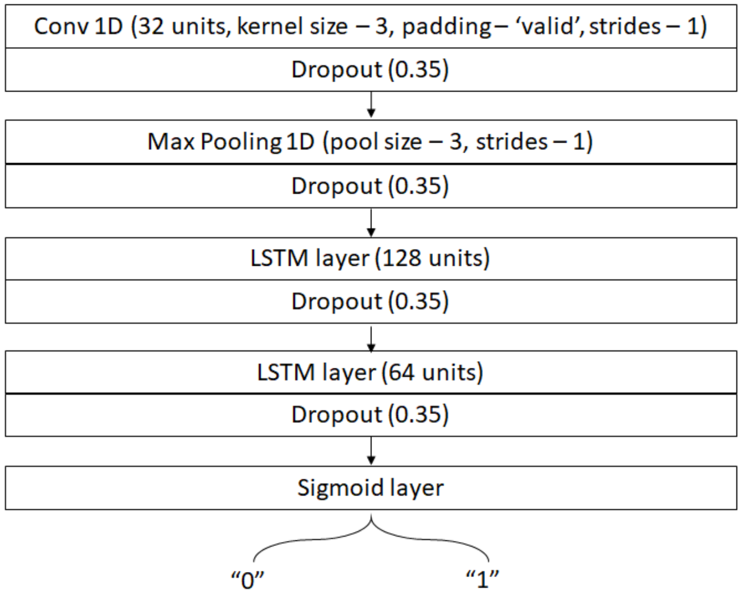


(c) LSTM7 architecture

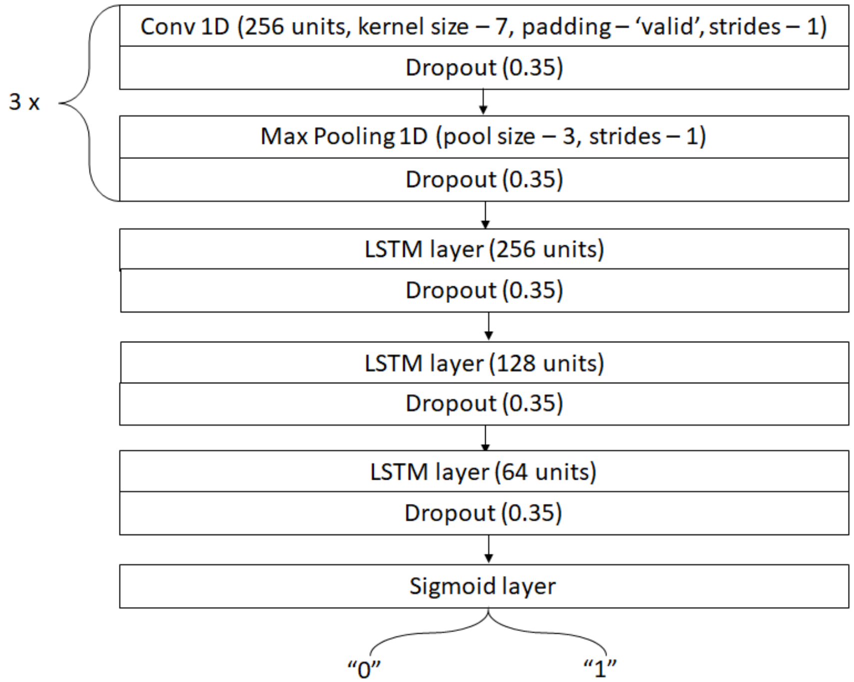
Figure 4.2: LSTM4, LSTM5 and LSTM7 models’ architecture

4.2.2 Hybrid Recurrent Convolutional Neural Network

The second model type chosen for experiments was a hybrid Recurrent Convolutional Neural Network. In total, four CNN+LSTM architectures were explored, two as a continuation of RNN models referenced before, and the other two inspired by [101], which achieved remarkable results for epileptic seizure detection. These models consist of the combination of convolutional and long short-term memory layers to overcome the performance of those two types of models isolated. Regarding the two first hybrid networks, they were constituted by 1D Convolutional and 1D Max-pooling layers, followed by LSTM layers. The first model, called CNN+LSTM1, had one 1D Convolutional and one 1D Max-pooling layers and, after that, two LSTM layers and one Sigmoid layer, with Dropout function applied to all the layers. The second one, named CNN+LSTM2, was structured with three 1D Convolutional, alternating with three 1D Max-pooling layers, followed by three LSTM and one Sigmoid layers, with also Dropout function applied to all layers. Concerning the models motivated by [101] the first one, called CNN+LSTM3, had a front-end that was constituted by four 1D convolutional layers, alternating with three 1D Max-pooling layers. In turn, a bi-directional LSTM layer was connected as the back-end of the deep neural network. For this network, Batch Normalization was firstly applied after each layer, as done in the aforementioned article. For CNN+LSTM4, the main difference with CNN+LSTM3 was the application of 2D Convolutional and Max-pooling layers instead of 1-dimensional functions. This was done due to the way data was structured, as a 2D matrix. As done with the RNN models, some alterations to the architecture were performed in order to try to reach the maximum performance of the model for the detection of IEDs. For these architectures, there were no modifications made to the length of the architecture, i.e., the number of layers remained equal throughout the whole dissertation. On the other hand, modifications in the hyperparameters and optimization functions like Dropout and Batch Normalization were made to assure the most successful design. In Figures 4.3a, 4.3b, 4.4a and 4.4b, all the architectures of the four CNN+LSTM models are shown.

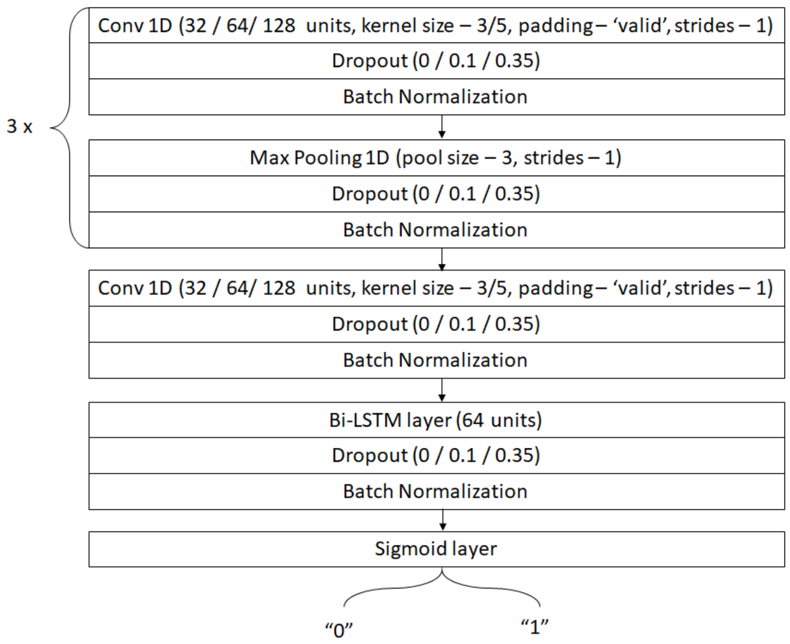


(a) CNN+LSTM1 model’s architecture

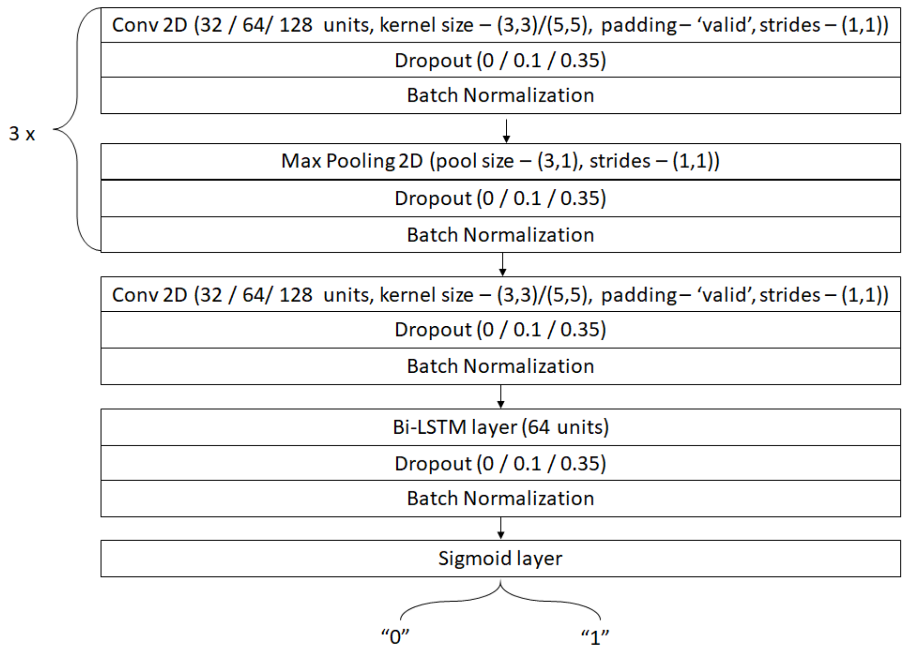


(b) CNN+LSTM2 model’s architecture

Figure 4.3: CNN+LSTM1 and CNN+LSTM2 models’ architecture



(a) CNN+LSTM3 model's architecture



(b) CNN+LSTM4 model's architecture

Figure 4.4: CNN+LSTM3 and CNN+LSTM4 models' architecture

4.2.3 Transformer Neural Network

The Transformer model is a recent type of deep network, first published in 2017, that works with different mechanisms used by other deep models since it has an attention-based process. There are few studies made with the application of these deep structures, and even fewer dealing with physiological signals processing [70]. Taking into consideration the remarkable results achieved in several areas like natural language processing, image processing, among others, the exploration of these models in signal processing, more specifically IED detection, was considered promising to study. Regarding the architectures designed to test these attention-based networks, these designs were motivated by two different architectures published by [67; 66]. Looking at the model published by *Vasawi et al* [66], the architecture designed included six Attention Modules, working as an encoder and, after that, a 1D Convolutional layer, a Flatten layer and Multi-Layer Perceptron was used to perform two-class classification. The architecture is explained in Section 2.3.3 and in the article referenced before.

Regarding the second architecture, called Vision Transformer, this presents some differences in the design, when compared to the first network. The network was composed of six Attention Modules, and the back-end equal to the first model. The main difference between these two architectures is in the design of the Attention Modules, more specifically with the application of the Layer Normalization function. In the Vision Transformer, the Layer Normalization is performed before the two sublayer components of the Attention Module (Multi-Head Attention and Feed Forward Network), instead of after these two sublayer components, as it is done in the Transformer model. The other difference is related to the preprocessing step of the input data, more concretely its position encoding. For the first model, the positional encoding is performed with the sinusoidal function and, for the Vision Transformer, the function applied is the 1D standard learnable positional embedding. Regarding the design of the input data for the Vision Transformer, the original matrix (250x18) was treated as an image, like it is done in the article referenced. The process performed over the data divides the matrix into patches, with the same dimension, and, then, each patch is flattened. After that, the flattened patches are reunited in a matrix, each row with a patch, forming a new shape to the matrix (150x30). Each patch has a 6x5 dimension. In Figures 4.5 and 4.6 , two schemes of aforementioned attention-based models can be seen.

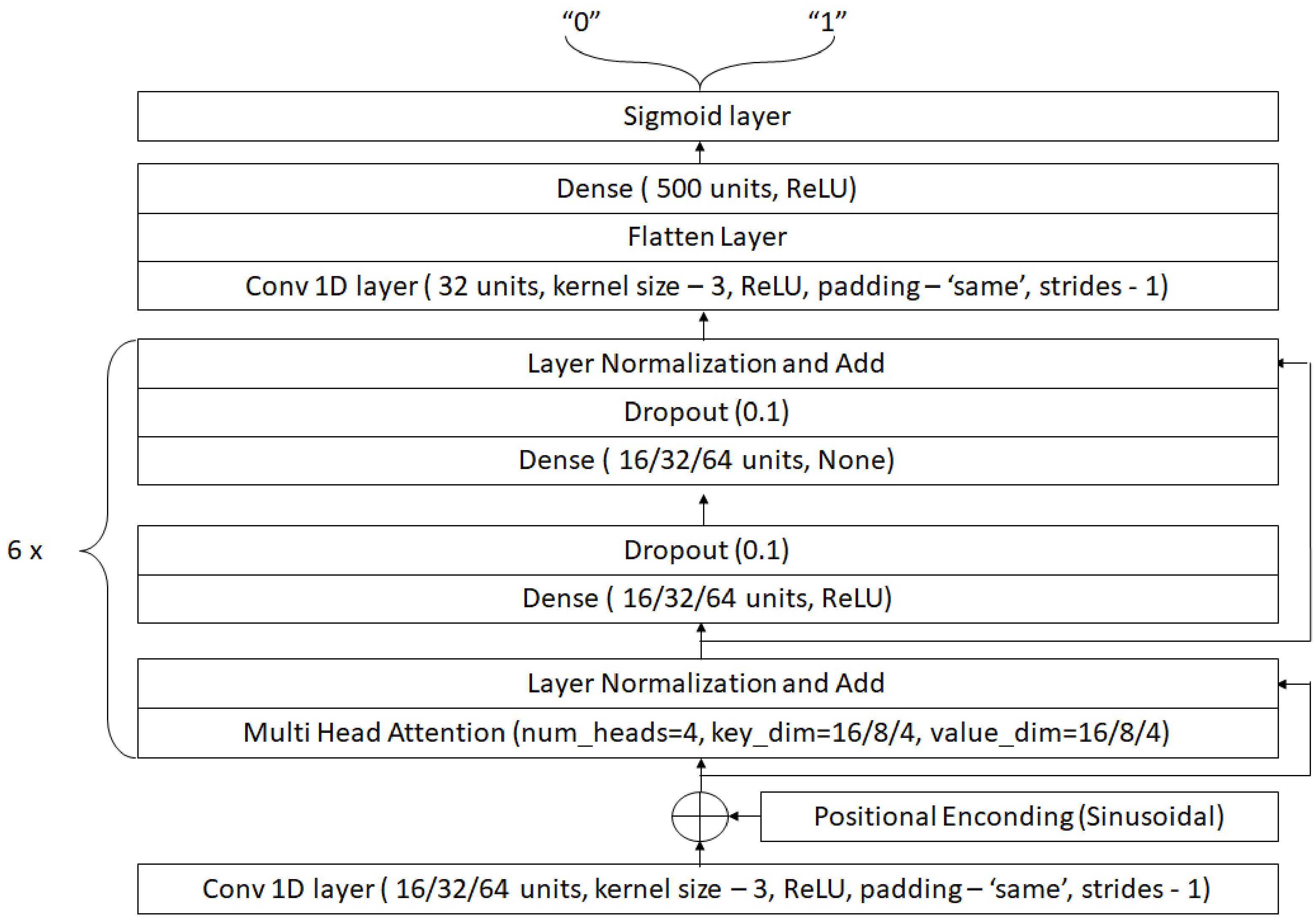


Figure 4.5: Transformer model architecture

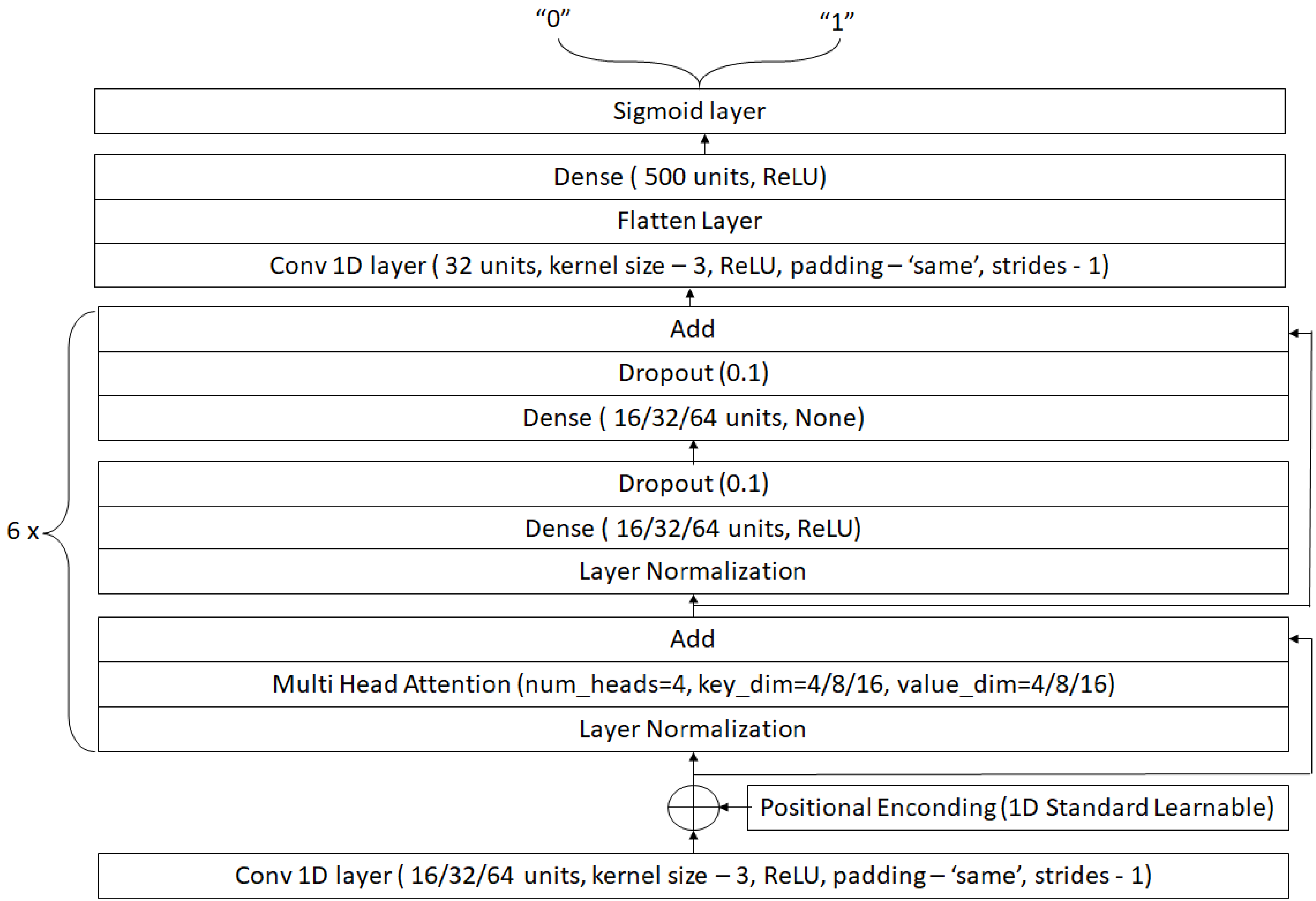


Figure 4.6: Vision Transformer architecture

4.2.4 Baseline model - VGG Neural Network

The VGG network was created in 2014[151], and it was comprised of several blocks of Zero Padding and Convolutional layers followed by Max Pooling layer. After these, Flattening preceded a block of two Fully Connected layers interspersed by Dropout. Finally, a Fully Connected layer with a Softmax activation function produced the output.

The first two blocks of layers were comprised of two Padding layers interspersed with Convolutional layers with, ending with Max Pooling. The following three blocks included one more Padding and Convolutional layers, before Pooling. The last block was constituted with a Flatten layer, followed by two Fully Connected layers, associated with two Dropout layers. In the end, a Softmax layer, with two units, was added to perform the two-class classification. In Figure 4.7, a scheme of the baseline model is shown.

4.2.5 Performance Assessment

Regarding the binary classification problem, several evaluation metrics were calculated to assess the performance obtained by the deep networks. Test Sensitivity, Specificity, AUC, and the number of True Positive, True Negative, False Positive and False Negative, as well as the TP and FP rate per hour were calculated, using three different thresholds: decision threshold of 0.5, a threshold to acquire a Specificity of 99% and similar values for Sensitivity and Specificity. All these evaluation criteria were used at different stages. Besides that, an average ROC curve was obtained for each five cross-validation loops.

4.3 Optimization Specifications

All the deep models were implemented with Python 3.7, using Keras 2.4.3 and a CUDA-enabled NVIDIA GPU. For the deep sequence models (Recurrent and hybrid Recurrent Convolutional Neural Networks), Stochastic Optimization was calculated with Adam optimizer [77], with a learning rate of 10^{-3}, β_1 of 0.9, β_2 of 0,99 and ε equal to 10^{-7}. On the other hand, for the attention-based networks, the optimization was performed using also Adam optimizer [77], but with different specifications (learning rate of 10^{-3}, β_1 of 0.9, β_2 of 0,98 and ε equal to 10^{-9}). The specifications for optimization of Attention models were based on the parameters used on [66].

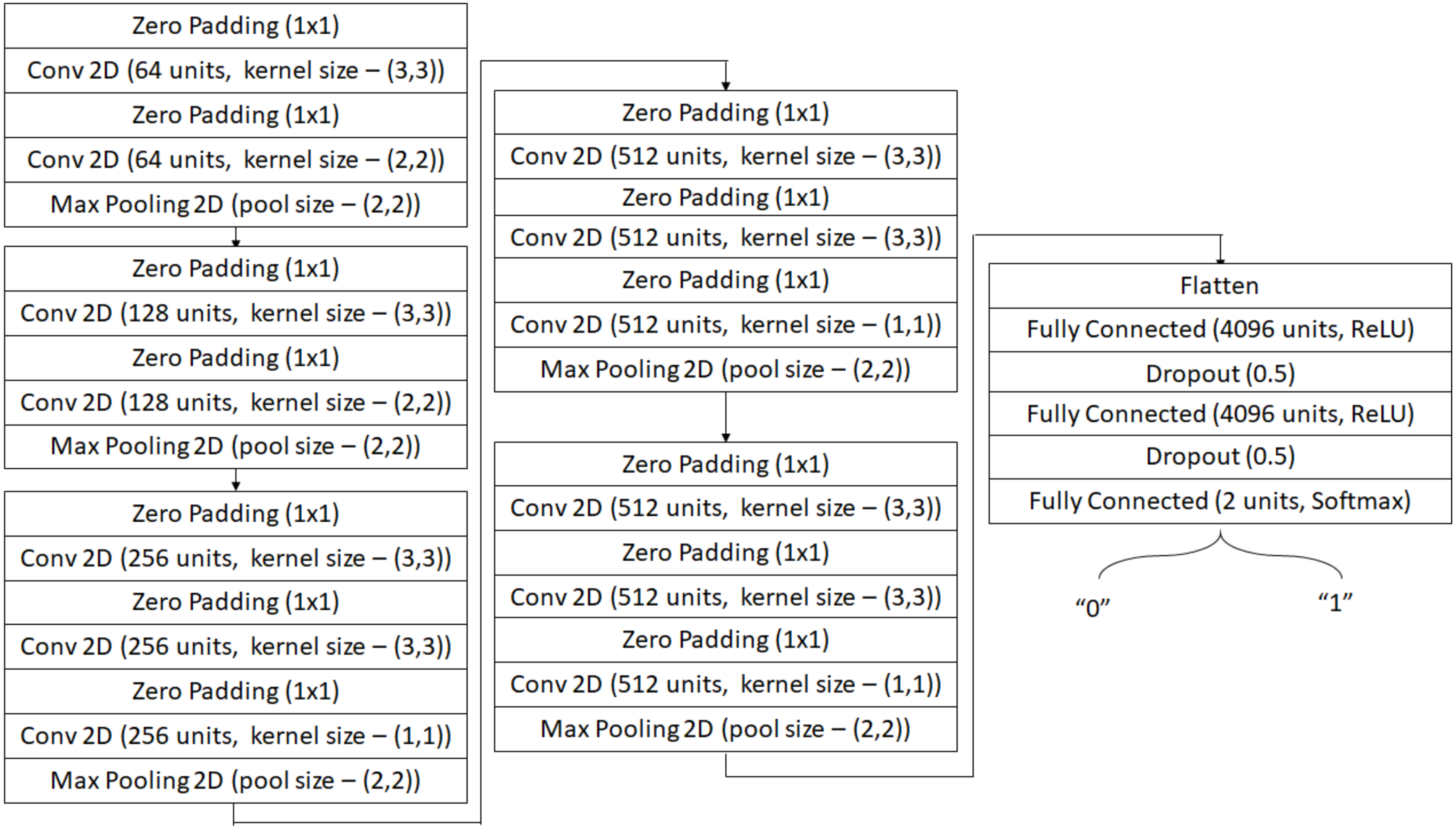


Figure 4.7: VGG baseline model’s architecture

Chapter 5

Results

Several deep architectures were experimented in order to assess the designs that more accurately could detect IED samples, using four different EEG databases. All the results obtained from these experiments are shown in the sections below.

5.1 Recurrent Neural Networks

For the exploration of deep sequence models, several neural networks were designed, motivated by the architectures from [97].

5.1.1 First round of experiments LSTM models

For the first round of trials, the goal was to replicate the work developed by [97], but using as database the Set A. Concerning the deep framework developed, two different architectures were created, which designs can be seen in Figures 4.1a and 4.1b. The first model, named LSTM1, is a one-to-one network, which means that from one LSTM layer, a single output is estimated. This model has 64 hidden units in the LSTM layer. On the other hand, the second model, named LSTM2, is a many-to-one network, i.e., from multiple stacks of layers, a single output is estimated. In this case, two LSTM layers were used, with 128 and 64 hidden units, respectively. The evaluation criteria for this phase was a decision threshold of 0.5 and a total of 10 training epochs. In Table 5.1, the results for this phase are shown.

Table 5.1: Results for LSTM1 and LSTM2 models for a decision threshold of 0.5

Model	Test Sensitivity	Test Specificity	TP / hour	FP / hour	AUC
LSTM1	9.8%	95.3%	1.97	64.23	0.52
LSTM2	29.5%	88.7%	11.61	225.36	0.59

5.1.2 Second round of experiments LSTM models

The second group of experiments that were implemented took into consideration the influence of the network's depth, maintaining the rest of the hyperparameters unchanged. In this way, 4 LSTM networks were experimented, from 2 to 7 LSTM layers, maintaining the units of each layer constant and with the use of dropout function with a rate of 0.35 through all the layers. The name attributed to each deep model was related to the number of deep layers (i.e., LSTM2, LSTM3, LSTM4 and LSTM7).

In the first approach, LSTM2, LSTM3, LSTM4 and LSTM7 were trained several times with the Set A, described in Section 4.1.3 , altering the class weight attributed to the positive class. The best performance was achieved with a class weight of 1:4 for most of the models. For a decision threshold of 0.5 and 20 epochs per training iteration, the results obtained were similar between the different architectures, with an increase in depth having a positive effect on the performance, as it can be seen in Table 5.2

Table 5.2: Results for Recurrent Neural Networks, with a threshold of 0.5

Model	Class Weight	Sensitivity	Specificity	AUC	TP / hour	FP / hour
LSTM2	1:2	78.8%	70.6%	0.80	27.26	519.78
LSTM3	1:4	83.7%	72.6%	0.83	28.94	483.63
LSTM4	**1:4**	**84.3%**	**72.5%**	**0.84**	**29.17**	**484.81**
LSTM7	1:4	85.7%	72.9%	0.85	29.63	478.92

From the results obtained, LSTM4 was the model that was chosen to continue to experiment. The reasons for this choice are explained in Section 6.2.2.

5.1.3 Third round of experiments LSTM models

For the third round of experiments, another recurrent model was added for exploration, named LSTM5, a five-layer LSTM deep network. As shown in the former phase, the increase of depth has a beneficial influence on the predictive ability of the networks. In this way, LSTM5 was designed to check if this improvement would be verified in a model that did not require much more computational power than the power needed for LSTM4. At this phase, the performance assessment and the number of epochs were changed, calculating the same evaluation metrics but now to criteria of test specificity equal to 99.0% and a condition of intersection between test specificity and test sensitivity, with a total of 60 epochs per each training iteration. This number of epochs and evaluation criteria was maintained throughout the rest of the experiments. At this phase, different class weight combinations were used (Automatic weight class attribution, 1:4, 1:10, 1:25 and 1:60), to assess which LSTM could detect IEDs more effectively. For both models, the best class weight combination was 1:10. For a test specificity of 99.0%, LSTM4 and LSTM5 achieved a sensitivity value of 4.7% and 5.6%, respectively, and, at the same time, reached equal values of specificity and sensitivity around 80.0% and 81.3%. These performances corresponded to an AUC value of 0.87 for LSTM4 and 0.88 for LSTM5. Taking into account the results obtained, the LSTM5 model was chosen to proceed to the third phase of experiments.

5.1.4 Fourth round of experiments LSTM models

For the fourth round, the objective was to infer the effect that a change to the hyperparameters and the layers that constitute the network, such as Dropout and Batch Normalization layers, can influence the performance achieved by the model. The first step was to experiment with the same LSTM5 model's design, changing, at each time, the implementation of Dropout, Batch Normalization or both at the same time. So, four different experiments were performed, with the use of Dropout with a dropout rate of 0.1, the use of Batch Normalization and, then, the combination of the Dropout and the Batch Normalization (with dropout rates of 0.1 and 0.35). It is important to note that, at the two former stages of experiments, Dropout with a rate of 0.35 was implemented in the model's architecture, so, at this phase, the tests with that layer implemented were already concluded. For all these experiments, the same class weight combination of 1:10 was used. The best result was obtained with Dropout (0.35 of dropout rate), reaching an AUC value of 0.88. The use of Dropout (0.1 of dropout rate) achieved an AUC value of 0.87, while the rest of the trials achieved the same value of AUC of 0.75. All the results can be seen in Table 5.3.

Table 5.3: LSTM5 third round of experiments results, with different layer combinations

	99.0% Specificity				Specificity = Sensitivity				
Layer combination	Sensitivity	Specificty	TP / hour	FP / hour	Sensitivity	Specificty	TP / hour	FP / hour	AUC
Dropout (0.35)	**5.6%**	**99.0%**	**1.92**	**17.64**	**81.3%**	**81.3%**	**28.38**	**337.26**	**0.88**
Dropout (0.1)	4.3%	99.0%	1.48	17.64	79.4%	79.4%	27.72	361.94	0.87
Batch Normalization	4.2%	99.0%	1.44	17.64	69.7%	69.7%	24.07	539.20	0.75
Dropout (0.35) + Batch Norm	4.1%	99.0%	1.41	17.64	69.2%	69.2%	23.83	550.80	0.75
Dropout (0.1) + Batch Norm	5.0%	99.0%	1.72	17.64	69.7%	69.7%	23.94	539.10	0.75

Afterwards, the next round of trials was performed, altering the number of LSTM units between 64 and 256 at each layer, maintaining the Dropout with dropout rate of 0.35. For this step, four different combinations of class weight were tested, 1:10, 1:20 and 1:30. Between all these trials, the best result was achieved for the LSTM5 model with 256 units per layer and with the use of the Dropout regularizer with a 0.35 dropout rate. In the Table 5.4, the final results are shown.

Table 5.4: LSTM5 third round of experiments results, with different layer units

		Specificity=99.0%				Specificity=Sensitivity				
Layer Units	Class Weight	Sensitivity	Specificity	TP / hour	FP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
64	1:20	4.2%	99.0%	1.44	17.64	79.4%	79.4%	27.50	368.15	0.85
128	1:10	5.6%	99.0%	1.92	17.64	81.3%	81.3%	28.38	337.26	0.88
256	**1:30**	**11.4%**	**99.0%**	**3.96**	**17.64**	**80.5%**	**80.5%**	**27.95**	**347.92**	**0.89**

5.1.5 Fifth round of experiments LSTM models

For the last stage of trials, the best LSTM5 architecture, with 256 units per layer and dropout of 0.35 associated at each, was tested with Sets B, C and D. The main focus on this stage was to infer if the implementation of different input data could improve the deep model's performance and to assess if data imbalance could be surpassed more efficiently using these techniques. For all the datasets, and as done in the other stages, different class weight combinations were used (1:4, 1:10, 1:20 and 1:30). Regarding the tests made with Set B, the best performance was achieved for the 1:10 class weight combination, reaching an average AUC of 0.94, with a sensitivity of 12.1% for 99.0% specificity and convergence of 88.7% between sensitivity and specificity. Looking at results with Set C, the design acquired the best predictive ability with the use of a 1:10 class weights, with a convergence of sensitivity and specificity around 78.0%, 1.9% of sensitivity for 99.0% specificity and an average AUC of 0.84. In the end, the application of Set D as input data reached the worst performance among all the datasets, with 3.4% sensitivity for 99.0% specificity and 73.4% specificity and sensitivity, with an overall AUC value of 0.76. The best results for the final LSTM5 model are shown in Table 5.5. In the appendix B, Figures B.1a and B.1b show the ROC curves corresponding to the best LSTM5 performances implementing sets A and B, respectively.

Table 5.5: Result for best LSTM5 design with Sets A and B

		Specficity=99.0%				Specificity=Sensitivity				
Dataset	Class Weight	Sensitivity	Specificity	TP / hour	FP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
Set A	1:30	11.4%	99.0%	3.96	17.64	80.5%	80.5%	27.95	347.92	0.89
Set B	**1:10**	**12.1%**	**99.0%**	**14.11**	**16.74**	**88.7%**	**88.7%**	**104.23**	**195.39**	**0.94**

5.2 Hybrid Recurrent Convolutional Neural Networks

Another objective was to study recurrence in other configurations besides pure Recurrent Neural Networks, like in combination with convolutional layers, in order to synergize the advantages and strengths of the two types of deep frameworks and surpass their limitations, to improve overall performance in IED detection.

5.2.1 First round of experiments CNN+LSTM models

Four neural networks were designed (has mentioned in Section 4.2), named CNN+LSTM1, CNN +LSTM2, CNN+LSTM3 and CNN+LSTM4, motivated by state of the art. Similar to the procedure described in the previous section, a series of experiments were done to optimize the various hybrid architectures. For the first group of trials, as done to the Recurrent models tested, the decision threshold defined was 0.5 and 20 epochs at each training iteration. At this phase, three different deep models were tested, CNN+LSTM1, CNN+LSTM2 and CNN+LSTM3. For each model, several tests were performed on various class weight combinations (Automatic, 1:4, 1:10, 1:15). In Table 5.6 the best performance results are shown.

Looking at the overall results, the other two architectures were discarded and CNN+LSTM3 was submitted to more trials. This decision is discussed in more detail in Section 6.3.1.

Table 5.6: Results for Hybrid Recurrent Convolutional Neural Networks, with a threshold of 0.5

Model	Class Weight	Sensitivity	Specificity	TP/hour	FP/hour	AUC
CNN+LSTM1	Automatic	81.6%	81.4%	28.22	328.48	0.82
CNN+LSTM2	1:15	64.7%	89.6%	22.38	183.30	0.77
CNN+LSTM3	**1:10**	**88.1%**	**77.0%**	**30.47**	**406.57**	**0.83**

5.2.2 Second round of experiments CNN+LSTM models

At this second phase of trials, another deep architecture was added, named CNN+LSTM4, aforementioned in Section 4.2 . This architecture allows us to study if 2D Convolution could work better with the image-shaped structure that database EEG epochs were organized in. CNN+LSTM3 and CNN+LSTM4 maintained the same design throughout this phase, with a total of 32 units in the 2D convolutional layer and 64 in the bidirectional LSTM layer. These models were submitted to several trials, with a change in the evaluation parameters (specificity of 99.0% and intersection between specificity and sensitivity) and increasing the training epochs to 60, implementing a set of class weight combinations (Automatic, 1:4, 1:10, 1:15, 1:25, 1:30, 1:60). The number With these trials, it was possible to understand, under those new evaluation conditions, which class weight combination could surpass with more efficiency the imbalance present in the database and achieve better performance. Throughout this phase, both model designs remained the same. The best results were achieved for 1:4 class weights for both models. An overall AUC of 0.90 and 0.93 was achieved, with test sensitivity of 13.5% and 16.9% for 99.0% specificity and 83.0% and 87.1% of specificity equals sensitivity for CNN+LSTM3 and CNN+LSTM4, respectively.

5.2.3 Third round of experiments CNN+LSTM models

Afterwards, as done with RNN models, the third stage of experiments was performed for both networks, making changes in the design of the deep frameworks, on hyperparameters and train's optimization functions (Dropout and Batch Normalization). At the second round of trials, Batch Normalization was already applied, so, in the third round of tests, the use of this function was not explored. In an initial phase, to a class weight combination of 1:10, Dropout function, with a dropout rate of 0.1 and 0.3, solo and in association with Batch Normalization, were added to both CNN+LSTM architecture, to infer which regularizer or combination worked better. For CNN+LSTM3, the best performances were obtained with the implementation of Batch Normalization only (the results obtained in the former round of trials and aforementioned) and with Batch Normalization plus Dropout at a rate of 0.1, with overall AUC of 0.90 and 0.89, respectively. Due to the close performance obtained for these two designs, it was decided to explore more extensively the use of Batch Normalization with Dropout together. So, as a next step, for the Batch Norm + Dropout(0.1) design, various class weight combinations were implemented. This study allowed to reach a better performance with a different class weight combination, in this case, 1:20. The results encountered had an average AUC of 0.92, test sensitivity of 29.6% for 99.0% specificity and 85.2% for specificity and sensitivity convergence.

Regarding CNN+LSTM4 results, the same exploration is done to CNN+LSTM3 was submitted to this deep neural network and a similar behavior was found. Besides the good results obtained with the application of Batch Normalization, which is mentioned above, also Batch Normalization + Dropout(0.1) combination generated an efficient model's performance, with 0.92 of AUC, test sensitivity of 21.9% for 99.0% specificity and 85.5% of convergence between sensitivity and specificity. Looking at the results obtained at this phase, for CNN+LSTM3, the utilization of Batch Normalization and Dropout (at 1:20 class weight) allowed an improvement of the model's execution. So, for the next round of experiments, this design was maintained. On the other hand, for the CNN+LSTM4 neural network, the use of Batch Normalization only or Batch Normalization and Dropout achieved similar performance. Despite that, only the CNN+LSTM4 design with the conjunction of Batch Normalization and Dropout continued to be explored, mostly related to the better sensitivity value achieved for 99.0% specificity criteria.

Afterwards, as the next round of trials, the number of units at each convolutional layer was explored. All the trials made before this stage were done with the same hyperparameters, i.e., 32 units in the convolutional layer and 64 in the bidirectional LSTM layer. Hence, tests were performed using 64 and 128 units for the 1D and 2D convolutional layers, maintaining the same number of units to Bi-LSTM. For these trials, 1:4, 1:10 and 1:20 weight class combinations were used. Regarding the CNN+LSTM3 outcome, no better result than the one achieved in an earlier stage was obtained. On the other hand, for the CNN+LSTM4 model, the best result was achieved for the architecture with 128 units and with Batch Normalization + Dropout(0.1). The best results are shown in Tables 5.7 and 5.8.

Table 5.7: CNN+LSTM3 third round of experiments results

			Specificity=99.0%				Specificity=Sensitivity				
Layer combination	Number of units	Class Weight	Sensitivity	Specificity	TP / hour	FP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
Batch Normalization + Dropout (0.1)	**32 + 64**	**1:20**	**29.6%**	**99.0%**	**10.25**	**17.64**	**85.2%**	**85.2%**	**29.63**	**269.03**	**0.92**
Batch Normalization + Dropout (0.1)	64 + 64	1:20	23.7%	99.0%	8.18	17.64	85.5%	85.5%	29.70	257.80	0.92
Batch Normalization + Dropout (0.1)	128 + 64	1:35	24.0%	99.0%	8.29	17.64	85.4%	85.4%	29.66	261.29	0.92

Table 5.8: CNN+LSTM4 third round of experiments results

			Specificity=99.0%				Specificity=Sensitivity				
Layer Combination	Number of units	Class Weight	Sensitivity	Specificity	TP / hour	FP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
Batch Normalization + Dropout (0.1)	32 + 64	1:4	21.9%	99.0%	7.58	17.64	85.5%	85.5%	29.59	261.83	0.92
Batch Normalization + Dropout (0.1)	64 + 64	1:4	19.8%	99.0%	6.83	17.64	86.2%	86.2%	29.89	247.66	0.93
Batch Normalization + Dropout (0.1)	**128 + 64**	**1:4**	**27.3%**	**99.0%**	**9.42**	**17.64**	**87.9%**	**87.9%**	**30.49**	**218.31**	**0.94**

5.2.4 Fourth round of experiments CNN+LSTM models

For a last round of experiments, the final CNN+LSTM architectures were tested with the other three datasets generated, i.e., Set B, C and D (described in Section 4.1.3). No modifications to the architecture were done during this phase, just a change in the weight class combination (1:1 for Set C only, 1:4, 1:10, 1:20 and 1:30 for all the sets). For both models, the results were similar. While the implementation of Sets C and D conducted worse performance from the deep neural networks, IED detection's ability improved with the implementation of Set B as input data for training and validation.

For the trials performed over the CNN+LSTM3 model, the best performance with the application of Set B was achieved for AUC of 0.95, with 21.2% of sensitivity for 99.0% specificity and the intersection of 90.4% between sensitivity and specificity, using 1:20 weight class combination. On the other hand, for Set C, the best performance was obtained for class weight classification 1:1, reaching, for 99.0% specificity criteria, the sensitivity of 9.2% and equal value between sensitivity and specificity of 82.9%, for an average AUC of 0.90. Ultimately, for Set D, the best results reached an average AUC of 0.75, with a convergence of sensitivity and specificity of 71.0% and 2.9% test sensitivity for 99.0% specificity, for 1:10 weight class conjunction.

Looking at the implementation of the three datasets on the CNN+LSTM4 architecture, the best performance was obtained with the implementation of Set B, for a class weight combination of 1:10, reaching an overall AUC of 0.96, the sensitivity of 26.1% to 99.0% specificity and convergence of specificity and sensitivity around 90.6%. Regarding the application of Set C, the best result was also obtained to a 1:10 weight class combination, with an average AUC of 0.90, reaching 14.7% of sensitivity to 99.0% specificity and 82.6% for specificity equal to sensitivity. Finally, the results for Set D's application are 6.0% for 99.0% specificity, 74.2% for sensitivity and specificity equality criteria and AUC of 0.76. All the evaluation criteria for the best performances from CNN+LSTM3 and CNN+LSTM4 models are presented in Table 5.9. In addition to this, the corresponding ROC curves are present in appendix B (B.2a, B.2b, B.3a and B.3b).

Table 5.9: Results for best CNN+LSTM3 and CNN+LSTM4 designs with Sets A and B

			Specificity=99.0%				Specificity=Sensitivity				
Model	Class Weight	Dataset	Sensitivity	Specificity	TP / hour	FP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
CNN + LSTM3	1:20	Set A	29.6%	99.0%	10.25	17.64	85.2%	85.2%	29.63	269.03	0.92
CNN + LSTM3	**1:20**	**Set B**	**21.2%**	**99.0%**	**24.80**	**16.74**	**90.4%**	**90.4%**	**106.50**	**165.10**	**0.95**
CNN + LSTM4	1:4	Set A	27.3%	99.0%	9.42	17.64	87.9%	87.9%	30.49	218.31	0.94
CNN + LSTM4	**1:10**	**Set B**	**26.4%**	**99.0%**	**30.95**	**16.74**	**90.6%**	**90.6%**	**106.52**	**157.85**	**0.96**

5.3 Transformer / Attention-based Neural Networks

The experiments done to explore the attention-based models were different. This was mostly related to the differences in the model's design, comparatively to the architecture of the deep sequence models.

5.3.1 First round of experiments for Transformer models

As the first round of experiments, the number of hyperparameters' influence in the model's performance was assessed, maintaining the same architecture through all the trials. Six Attention modules were used, changing the number of units in the Dense layers, the dimensions of key and value matrices and the number of heads in the Multi-Head function, under a correlation rule used in [66], which the dimensions of key and value matrices are equal to the number of units of Dense layers divided by the number of heads used in the Multi-Head Function. So, maintaining the same number of four heads, the units of Feed Forward layers and dimensions of the Multi-Head function changed between 64, 32, 16 and 16, 8, 4, respectively. This study was made for both the Vanilla Transformer and ViT models. For the architecture with more parameters, the weight class combinations were set between 1:200 and 1:600. On the other hand, for the two lighter designs, the weight class combination interval was defined between 1:200 and 1:300. These intervals of class weights were established because the values used for the deep sequence models did not reach positive results when applied to the attention-based architectures and, so, a different range of values was used. After evaluating all the results obtained, the best performance was achieved by the Transformer model, for the first and last architectures. In the Table 5.10, the best results for the six architectures can bee seen.

Table 5.10: Results for best performances from Attention-Based Neural Networks

		Specificity=99.0%				Specificity=Sensitivity				
Model	Class Weight	Sensitivity	Specificity	TP / hour	FP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
Transformer1	1:300	3.2%	99.0%	1.11	17.64	75.7%	75.7%	26.22	432.77	0.83
Transformer2	1:200	1.9%	99.0%	0.66	17.64	75.5%	75.5%	26.13	435.16	0.83
Transformer3	**1:275**	**1.3%**	**99.0%**	**0.43**	**17.64**	**75.6%**	**75.6%**	**26.17**	**433.73**	**0.84**
ViT1	1:500	2.7%	99.0%	0.93	17.64	70.2%	70.2%	24.32	528.40	0.77
ViT2	1:275	2.3%	99.0%	0.79	17.64	70.6%	70.6%	24.44	521.85	0.77
ViT3	1:225	1.5%	99.0%	0.51	17.64	71.8%	71.8%	24.81	502.21	0.79

5.3.2 Second round of experiments for Transformer models

After this round of experiments, for both of the best architectures and the best class weight combination, Set B was used to train and test both models to infer if the model's performance may improve, similar to what happened for deep sequence models. Contrarily to the results obtained for the other architectures, the performance of the attention-based models was worse when Set B was implemented, with an average AUC of 0.79 for both Transformer model 1 and Transformer model 3. In appendix B, Figures B.4a, B.4b, B.5a and B.5b show the ROC curves for Transformer1 and Transformer3 models, with implementation of Sets A and B, respectively.

Chapter 6

Discussion

The main objective of this dissertation was to explore deep architectures capable of effectively detecting IED samples in epileptic EEG records.

6.1 Class weight combinations

One of the principal obstacles to overtake was the imbalance of the first database, Set A. As can be seen in section 4.1.3, Set A, the first database used as input data, had a total of 2522 two-second epoch EEG samples that represent the positive class, around 4.2% of the total samples that constitute the database. So, having this in mind, some decisions were made to decrease the difference of class samples, to achieve better performance by the networks. The first decision taken was relative to the organization of the train and validation samples in five different parts. The particularity about the five parts was that, while the negative samples were distributed throughout these parts, the positive samples were all repeated at each one of them. This distribution was made to undersample the negative class samples at each training phase. Besides this decision, other techniques were implemented to overcome the imbalance issue. One example of this was the attribution of different weights to the class samples and the creation of other datasets with a more total number of IED samples. Class imbalance is a common issue present in databases for deep learning applications. This difficulty can be surpassed by several techniques, being one of them the attribution of different weights to the samples of the various classes. By default, deep learning algorithms assume that all the classes are balanced. So, in the cases where this does not happen, the deep network will be more biased towards the class with more samples, since the model does not have enough data from the minority class to properly learn and understand the specific patterns and features from this class [152]. To penalize the misclassification of the minority class, different weights can be given to the classes, which allows the algorithm to focus more on the minority class and reduce its classification errors. Regarding the attribution of the weights, this can be done automatically by the algorithm or weights can be attributed manually. During all the experiments, both ways were used to find the pair of class weights that could be more beneficial to an improvement of the deep neural network's performance. Due to the differences in design between all the deep models studied, a group of weight class combinations was implemented at each experiment phase and, as can be seen in the Results chapter, the class weight pair that matched with the best performance was always different, ranging from 1:2 until 1:30. In this way, no standardized or common behavior between the weight of the positive class and the algorithm's ability to detect correctly IEDs was confirmed, i.e., not always the highest weight attributed to the positive class leads to better algorithm performance.

6.2 Recurrent Neural Networks performance

6.2.1 First round of experiments for LSTM models

A priority of the dissertation was to design a deep sequence model capable of detecting IED samples correctly. So, as an initial step, the intent was to replicate the work developed by [97], as mentioned before. This first approach highlighted the main issue of Set A, the data class imbalance. The results obtained to test sensitivity and specificity show that there is an imbalance of classes in the database. As results for specificity are very high, but sensitivity is low, it is possible to assess that the model is very well tuned to correctly classify negative samples, but it did not generalize well for positive samples. In the absence of a relevant number of positive samples, i.e., samples classified with 1, the network was not able to fully capture the features of these samples and got too specific to detect negative samples.

6.2.2 Second round of experiments for LSTM models

At the second round of trials, two, three, four and seven-layer LSTM models were tested with Set A. This study allowed us to understand how the depth of the network could influence its performance. As can be seen in Table 5.2, it is noticeable that the increase in depth is translated into an improvement of performance, with the increase of AUC value, true positive per hour rate and a decrease of false positive per hour rate, which reveals a better predictive ability of the deepest model. In addition to this, Figure C.1, which contains the four confusion matrices, shows that LSTM4 and LSTM7 models (four and seven layers, respectively) can differentiate more accurately between normal EEGs and IEDs samples. After this phase of experiments, the LSTM4 model was the architecture chosen to proceed with experiments. Even though it was not the model which achieved the best performance, the trade-off between predictive ability and computation load appeared more advantageous, being three times faster to train and with less than one-fifth of the total parameters of the LSTM7 model.

6.2.3 Third round of experiments for LSTM models

With the correlation between the depth and the performance of the RNN, we decided to add another model (LSTM5) with an extra LSTM layer, in order to check if a better performance could be achieved by this deep structure with a model that did not have a much bigger computational load than the LSTM4 has. At this phase, the number of training epochs was also increased as an attempt to improve the deep model's training. Looking at the overall results at this phase, it was possible to conclude that the LSTM5 model reached a better classification performance. For both evaluation criteria, the LSTM5 model reached better test sensitivity, specificity and AUC values than the four-layer LSTM model. In addition to this, Figures C.2 and C.3 shows that, for 99% specificity and specificity equal sensitivity conditions, LSTM5 model's performance surpassed LSTM4, with a better ability to distinguish normal EEG samples from IEDs. This difference in performance associated with a similar computational load apart from both deep architectures was significant to discard the LSTM4 model and to continue exploring the five-layer recurrent neural network.

6.2.4 Fourth round of experiments for LSTM models

Regarding the next phase of trials, the number of hyperparameters and the design of the LSTM5 neural network was explored with the intent of achieving a better performance. This phase of trials was conducted to check how an increase or decrease in the number of units of the LSTM layers and how Dropout or Batch Normalization functions could influence the generalization ability from the recurrent neural network. Between all the architectures tested, the use of Dropout (with a rate of 0.35) and a total of 256 units at each LSTM layer achieved the best performance, as can be seen in Table 5.4. Taking into consideration the confusion matrices obtained for this deep model, presented in Figure C.4, it is possible to claim that for criteria of 99% specificity, that was a relevant improvement in performance, with an increase of test sensitivity of more than double. On the other hand, evaluation metrics for sensitivity equals specificity condition remained similar. With this exploration of hyperparameters and model's design, was possible to assess how small modifications in the network, maintaining the core of the design, can have a significant impact at the performance of the networks. Firstly, for the LSTM5 network, the incorporation of Batch Normalization layers to the network worsened its performance, with a decrease in average AUC values. Regarding the number of units at each layer, the increase of the model's width was proportional to the improvement of the predictive ability of the network.

6.2.5 Fifth round of experiments for LSTM models

After the experiments with Set A were completed, the three remaining sets were explored. Looking at the overall results, the implementation of Set B was the only dataset that improved the performance of the LSTM5 model, with an increase of evaluation metrics for both criteria. Set B is a dataset constituted with the same signals that are present in Set A but from records made with three different electrode montages (longitudinal bipolar, common average and source derivation). Even though the percentage of positive class samples did not increase significantly, the presence of more data samples allowed the network to learn more effectively specific patterns from the IEDs and normal EEG samples, which improved the ability of the model to distinguish these two distinct physiological patterns. On the other hand, the application of sets C and D did not develop the same success as the application of set B. These two sets were generated with the intent of increasing the total number of positive class samples through an artificial technique (SMOTE and data overlap with sliding window). For both datasets, the performance achieved was worse than the one reached with the implementation of sets A and B. Neither the creation of artificial positive samples from the original epochs nor the creation of new segments through the copy of parts of the original positive samples enabled the neural network to attain a better predictive ability. In Figures C.5 and C.6, confusion matrices obtained from the implementation of sets B, C and D are shown.

6.3 Hybrid Recurrent Convolutional Neural Networks performance

As mentioned in Section 5.2, in order to explore sequence-based networks not only by itself, but also in aggregation with other type of neural networks, the combination of convolutional and sequence-based layers was performed.

6.3.1 First round of experiments for CNN+LSTM models

At the first round of trials, three distinct hybrid convolutional recurrent neural networks were designed to survey how the two kinds of functions combined could influence the ability of the model to correctly differentiate normal EEG samples from interictal epileptiform discharges. The three designs contained a different number of LSTM and Convolutional layers to check how the presence or absence of each type of layer can influence the overall classification performance. Taking into consideration the results shown in Table 5.6, the CNN+LSTM3 design achieved the best performance. The combination of Convolutional and Max Pooling layers with a bi-directional LSTM layer dealt better with the EEG signals than the two other architectures. In addition to this, the computational load related to this architecture is lower, when compared with the two others models (CNN+LSTM1 and CNN+LSTM2 have a total of 279105 and 1389411 hyperparameters, respectively, while CNN+LSTM3 has a total of 27777 hyperparameters), and taking 1 second per training epoch, while CNN+LSTM1 and CNN+LSTM2 take 2 and 5 seconds per epoch, respectively. Consequently, the CNN+LSTM3 model was chosen to proceed with the experiments.

6.3.2 Second round of experiments for CNN LSTM models

At the second round of experiments, as mentioned in Section 5.2, a fourth deep model was added, named CNN+LSTM4. This model has the same design as CNN+LSTM3, but instead of 1D convolutional and max pooling layers, it has 2D layers. This decision was made taking into consideration the image-based structure, in which the EEG epochs were organized. At this phase, the design of both models was maintained, just changing the criteria established for performance evaluation and increasing the total number of training epochs. In the end, the overall performance was better than the performance achieved in the former phase, with a higher AUC value for the two neural networks. Furthermore, between the two models explored, CNN+LSTM4 was the one that achieved the best predictive ability, with better results at the various evaluation metrics, as can be seen in

Section 5.2. In addition to this, confusion matrices for both evaluation conditions can be seen in Figure C.7 and Figure C.8.

6.3.3 Third round of experiments for CNN+LSTM models

Since the two architectures achieved reasonable results, both of them continued to be explored in the third phase of trials, by altering the width of the network and the type of layers that constituted the architecture. With the experiments at this phase, it was possible to refine the architecture of both models, in order to improve the predictive ability of both. Regarding the alterations of the layers of the models, it was possible to assess that, contrarily to the results obtained for the LSTM5 model, the incorporation of Batch Normalization had a positive effect on the performance, moreover in combination with Dropout regularization. Furthermore, the exploration of the models' width brought up different outcomes for each model. Concerning the CNN+LSTM3 network, its performance maintained similar, independently of the number of units used in the Convolutional and Max Pooling layers. On the other hand, for the 2D hybrid architecture, the increase in the model's width provided an improvement in the performance. CNN+LSTM4 benefited from the presence of more units to more efficiently deal with the input data. All these inferences can be seen in Tables 5.7 and 5.8.

Concerning the comparison of the two models, even though CNN+LSTM4 reached a higher AUC value, it did not achieve better results for all the evaluation metrics, like in the case of the evaluation metrics for 99% specificity, which is also corroborated by Figure C.9. Furthermore, as aforementioned, the complexity of the network was something taken into account for an evaluation of the model's advantages. And, in this field, CNN+LSTM3 achieved good performance with a simpler model, as it has less than 2% of total parameters of CNN+LSTM4 architecture and it is six times faster to train. Since none of the models achieved a significantly higher predictive ability, both models were maintained to the next phase of experiments.

6.3.4 Fourth round of experiments for CNN+LSTM models

At the last round of trials, sets B, C and D were used as input data. The reasons pointed out in the last section are the same for the exploration of these datasets with this hybrid recurrent convolutional neural network. Looking at the overall results obtained, the same conclusions can be drawn. Regarding Set B implementation, this dataset allowed both deep models to surpass the performance achieved with set A. This dataset lets the hybrid models learn more efficiently the characteristics of the various EEG patterns and to achieve a better predictive ability to distinguish normal EEG samples from IED ones. On the other hand, the implementation of artificial positive class samples, through sets C and D, complicated the ability of the models to capture the specific patterns from the two types of EEG samples present. The artificial generation of positive class samples did not allow the deep models to learn effectively the attributes from this class. Moreover, comparing the results between the two sets, the utilization of the SMOTE technique led to better performance when compared to the use of positive data overlap. In Figures C.11, C.12, C.13 and C.14 all confusion matrices from the aforementioned results can be seen.

6.4 Transformers Neural Networks performance

After all the experiments with deep sequence models were done, it was decided to explore a recent type of deep architecture, named Transformers. These attention-based models showed promising results in various areas, as mentioned in Section 4.2.3. The path that the experiments took for this type of network was different from the one taken for the deep sequence models.

6.4.1 First round of experiments for Transformer models

For the first round of experiments, two different designs were tested, with three changes in the hyperparameters of each design. This modification of parameters was made under a correlation rule, explained in Section 5.3. The best results were all achieved by the Vanilla Transformer architectures, surpassing the performance obtained by ViT, as can be seen in Table 5.10. The ViT model was inspired by a deep architecture developed for image recognition, with a preprocessing of input data and the design of the architecture more specific to that area of interest. So, the more specificity of the ViT architecture for image recognition led to a worse predictive ability when applied in a signal processing task. Looking at the results obtained for the Vanilla Transformer designs, no significant alterations in the models' performance were noticed with the decrease of the hyperparameters of the network. Thus, at the next phase, just Vanilla Transformer models 1 and 3 continued to be explored. Vanilla Transformer 1 was chosen because it achieved the best results for both evaluation criteria. On the other hand, Vanilla Transformer 3 was not discarded due to two reasons: it was the simpler model, i.e., with the less number of parameters reached a similar performance in comparison with the other two Vanilla Transformer models and, in addition to this, had reached the higher AUC value between the three networks.

6.4.2 Second round of experiments for Transformer models

At the second round of experiments, Set B was implemented with the intent to potentiate the performance of the attention-based models that performed better in the last phase. The implementation of only set B was done taking into consideration the results obtained for other models, which set B was the only one that improved the predictive ability of the deep models when compared with set A results. In this case, the increase of diversity in positive and negative class samples was not helpful for the upgrade of the model's predictive ability. One reason for these worse results is the fact that the model's architecture was not capable of dealing with the increase of complexity of the signals used as input data. More diversity means more information to learn and properly understand and the attention-based model was not able to capture the specific features from the negative and positive class samples. In addition to this, the increase of complexity was not accompanied by the increase of the total samples present in the dataset, i.e., there was much information to learn and not enough samples as a source for learning.

6.5 Comparison of best performances

The refinement of the models explored to distinguish positive and negative class samples was one of the main goals of this dissertation. All this work was made associated with [1; 2], in which a VGG neural network was implemented to explore Set A and Set B, with the same objective of effectively detecting IED samples. The comparison of the best results was only done with the work developed by [1; 2] since the datasets used in this dissertation were different from the general datasets used on automation of epilepsy detection issue and, in addition to this, there is a lack of state of the art related to the implementation of deep sequence models for the IED detection tasks.

Regarding the best performance achieved between the different types of neural networks, this fell back over the hybrid recurrent convolutional neural networks, more concretely the CNN+LSTM4 model. The combination between convolutional and recurrent functions learned more successfully the main features of EEG samples present in the database when compared with the implementation of recurrence and attention-based models. Beyond this, comparing the two CNN+LSTM models explored, the one that implemented the 2D convolution and max pooling layers was the model that achieved the best performance. The CNN+LSTM4 architecture benefited from the grid structure in which the EEG epochs were organized, acquiring more efficiently the ability to distinguish the two types of EEG patterns.

Despite the good performance achieved for both sets A and B, none of the models designed achieved a better performance than the baseline model, aforementioned, with higher evaluation metrics for both 99.0% specificity condition and equal value of specificity and sensitivity criteria, as can be seen in Tables 6.1 and 6.2. The VGG model was capable of learning the main features of both EEG patterns and distinguish more correctly the IED samples from the normal EEG samples than the proposed networks tested at this work.

Even though the CNN+LSTM4 model achieved the best results, other models also obtained a remarkable performance. Taking this into consideration, all the work developed during this dissertation allows the establishment of a map with the various architectures that can deal effectively with the EEG datasets. These models have distinct features, with more or fewer parameters, time training and, obviously, classification performance, which can be significant to decide if a model can fit better the needs of a certain situation. For example, even though CNN+LSTM4 reached the best performance, the design CNN+LSTM3 also achieved relevant results, with a simpler architecture, possible to be used in a situation in which the computational power and resources are limited. Beyond this, the application of set B at the CNN+LSTM4 model led to a predictive performance similar to the performance achieved by the baseline model with the same dataset, with about 3% of the total parameters of the VGG network. For a better comprehension of this diversity, Table 6.3 retains the number of parameters and training time of the best models and Figures 6.1 and 6.2 shows a plot with the different ROC curves acquired from the neural networks that achieved the best performances under the implementation of Sets A and B, respectively. Looking at the plots with the different ROC curves, the comparative conclusions mentioned about the different models are corroborated, giving a clear idea of the predictive ability of the various networks and the difference in performance between them and, in combination with the results and design information, the advantages and drawbacks of each model can be encountered.

Table 6.1: Best performance results obtained for Set A, including baseline model from [1]

		99% Specificity			Specificity = Sensitivity				
Model	Set	Sensitivity	Specificity	TP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
Baseline VGG	A	79.0%	99.0%	40.27	93.0%	93.%	47.72	122.41	0.96
CNN+LSTM4	A	27.3%	99.0%	9.42	87.9%	87.9%	30.49	218.31	0.94
CNN+LSTM3	A	29.6%	99.0%	10.25	85.2%	85.2%	29.63	269.03	0.92
LSTM5	A	11.4%	99.0%	3.96	80.5%	80.5%	27.95	347.92	0.89
Transformer1	A	3.2%	99.0%	1.11	75.7%	75.7%	26.22	432.77	0.83
Transformer3	A	1.3%	99.0%	0.43	75.6%	75.6%	26.17	433.73	0.84

Table 6.2: Best performance results obtained for Set B, including baseline model from [2]

		99.0% Specificity			Specificity = Sensitivity				
Model	Set	Sensitivity	Specificity	TP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
Baseline VGG	B	74.2%	99.0%	38.05	89.5%	93.5%	105.30	114.00	0.95
CNN+LSTM4	B	26.4%	99.0%	30.95	90.6%	90.6%	106.52	157.85	0.96
CNN+LSTM3	B	21.2%	99.0%	24.80	90.4%	90.4%	106.50	165.10	0.95
LSTM5	B	12.1%	99.0%	14.11	88.7%	88.7%	104.23	195.39	0.94
Transformer1	B	2.2%	99.0%	2.57	71.7%	71.7%	84.43	479.59	0.79
Transformer3	B	0.8%	99.0%	0.93	71.3%	71.3%	83.70	487.93	0.79

Table 6.3: Training time and total number of parameters, with Sets A and B, for the deep models with the best performances

Model	Training time per epoch		Total number of Parameters
	Set A	Set B	
Baseline VGG	4 s	13 s	45570306
LSTM5	9 s	28 s	2383105
CNN+LSTM3	1 s	4 s	27777
CNN+LSTM4	6 s	19 s	1389411
Transformer 1	5 s	17 s	4018361
Transformer 3	8 s	25 s	4162937

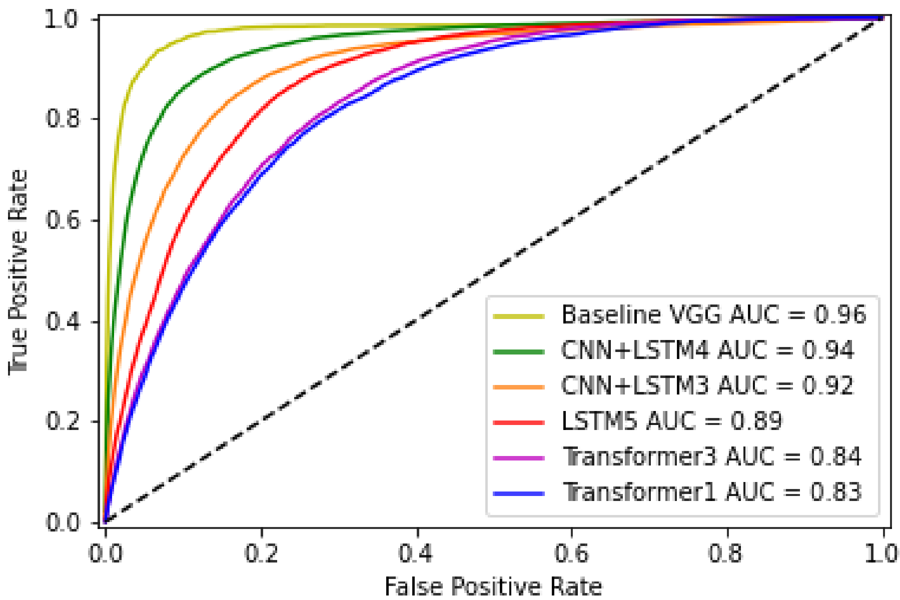


Figure 6.1: Plot of the ROC curve correspondent to the models with best performance, with Set A

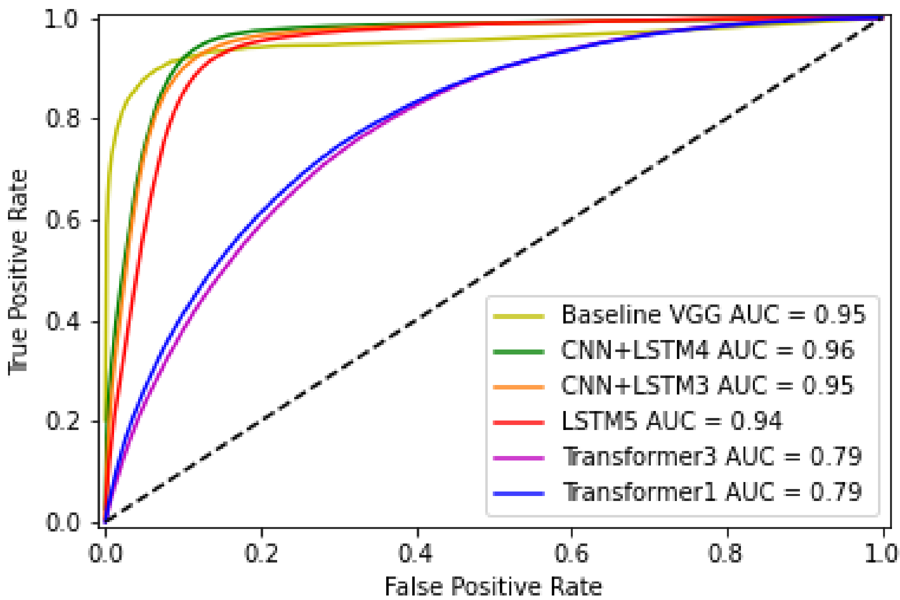


Figure 6.2: Plot of the ROC curve correspondent to the models with best performance, with Set B

Chapter 7

Conclusions and Future Work

7.1 Conclusions

The main conclusions that can be considered are that it is feasible to implement deep sequence models to correctly classify interictal epileptiform discharges between normal EEG samples and that the influence of class data imbalance in the overall performance of the models can be handled and the consequent errors can be diminished.

Class data imbalance is a common issue related to the EEG signals database and is also present in the original dataset of this dissertation. As an attempt to handle this problem, different class weights were attributed during all the experiments. This decision contributed to the improvement of the performance of the models. This improvement was first noticed for recurrent models referenced in section 5.1.2, comparing the results obtained at this phase with the results from the first approach, mentioned in Section 5.1.1, where was noticeable the inability to learn the IED features correctly. Looking at the rest of the experiments performed, the attribution of class weights had a positive effect on the predictive capacity of the networks and, in complement with other techniques, making it possible to surpass the class imbalance issue.

Regarding the performance of the deep models, the CNN+LSTM4 achieved the best predictive performance for both datasets A and B, with an overall AUC of 0.94 and 0.96, respectively. This network reached, for set A, 87.9% of test sensitivity and specificity, with a true positive per hour and false positive per hour rates of 30.49 and 218.31, respectively. For 99% specificity, a test sensitivity value of 27.3% was reached, with a decrease of true positive and false positive per hour rates to 9.42 and 17.64. Regarding the implementation of set B, the performance of this deep network was also successful, with a total value for sensitivity and specificity of 90.6%, with true and false positive per hour rates of 106.52 and 157.85, respectively. At 99% specificity criteria, test sensitivity decreased to 26.4%. Aside from these results, other models also achieved relevant results. Taking this into consideration, the work developed allowed the design of a map with a variety of neural networks that can perform effectively with IED databases. Each deep model presents different features, with advantages and drawbacks, being more useful in distinct situations.

With all the experiments with the deep sequence models concluded, we decided to explore a recent type of network: attention-based models. This decision was made because this kind of network reached remarkable results in various deep learning application areas, like speech recognition, natural language processing or image recognition. Moreover, there is a lack of studies surveyed in IED detection with this type of model. The results obtained with this type of network did not reach the results obtained by the sequential models. This worse performance is understandable, due to the limited time to properly explore these networks and to the lack of information related to the application of Transformer models in EEG signals, having a greater probability of not making the best decisions, while for sequence models there is already a lot of documented information so it is easier to know what is better or worse for performance.

This dissertation showed the potential of deep sequential neural networks applied in the automation of epilepsy through the analysis of EEG databases. The implementation of these algorithms could improve the diagnosis of epilepsy, reducing the amount of work and time that clinicians had to spend, bringing more reliable diagnostic outcomes. Another important point was the implementation of these models in IED detection, an application with limited state of the art currently published.

7.2 Future Work

This dissertation was developed with the intent to explore an alternative approach of deep learning in health, more specifically the IED detection for diagnosis of epilepsy. With this work, it was possible to exhibit not only the potential of these algorithms but also the current limitations and difficulties that are faced, to prepare strategies for the development of more robust and complete models for future implementations.

There are still several obstacles that need to be surpassed in order to reach more reliable algorithms. First, an important step is the increase of IED samples, which will aid the networks to learn more significant features from a higher variety of IED records. Aside from the exploration of Data Augmentation techniques already made, other methods could be experimented, since there are many alternatives. Moreover, the implementation of feature extraction techniques was not surveyed and could represent an alternative method to diminish the data imbalance and to produce a simpler and more relevant dataset, with less non-essential information. In addition to this, some additional experiments could be made with set B data. This dataset is composed of EEG signals from three different montages and a promising survey would be to use the signals from the two remaining montages that were not tested to check which montage is the best to extract and comprehend more efficiently the significant features inherent to each EEG pattern.

Secondly, future decisions could be to continue to explore the deep sequence models experimented, since there are still some other variables and hyperparameters that can improve the predictive performance of the networks. Regarding the attention-based networks, these models were less experimented and can sure bring very promising results. Having this said, work can be done at the exploration of the network's design in a deeper fashion.

Appendix A

Deep Learning applications in Epilepsy Detection: State of the Art

Table A.1: State of the Art of Deep Learning in Epilepsy Detection

Ref	Author	Year	Dataset	Preprocessing Method	Features Extraction	Deep Architecture	Results
[128]	Weng	1996	Database form Montreal Neurological Institute, with EEGs from five patients with seizure and non-seizure activity	Data Segmentation in 2.56 second epochs	Five features extracted - Two frequency domain features (dominant frequency and average power spectrum) through FFT and three time domain features (average amplitude, average duration, coefficient of variation)	Multilayer backpropagation network	Seizure detection and classifications are satisfactory.

[133]	Petrosian	2000	EEG signals acquired in this study were from two epileptic patients. Four preictal single-channel recordings (I1,I2,I3,I4) and three non-seizure recordings (N1,N2,N3)	No preprocessing method was applied	Wavelet Transform to obtain low and high-pass subands	Discrete-time recurrent multilayer perceptrons	Results for intracranial - Performance was poor with raw data, but the results were better using wavelet decomposed signals. Results for extracranial - Better network performance with using both high pass and low pass subband signals.
[134]	Bates	2000	Dataset created with 75 EEG channels	The pre-seizure and seizure data is segmented, filtered with a HP filter and normalized to [-1,1]	Raw data	Fully Connected architecture	It was found that the highest R-values were obtained from the seizure component of each electrode record. Also, high R-values were obtained at various points of the pre-seizure components, suggesting that there was a development of seizure activity within these components.
[95]	Liu	2002	The system was evaluated using 8-channel clinical EEG records from 81 epileptic subjects. The records ranged from 4 to 24 hours in length, totaling more than 800 hours of records	No preprocessing method was applied	Continuous Wavelet Transform used to obtain features like Relative Amplitude, Duration, Slope and Sharpness	Three similar MLP neural networks	The detection of epileptic events was 90% correct. The detection rate of sharp transients was 98% and overall false-detection rate was 6.1%.
[114]	Bates	2003	EEG one minute records were obtained by a patient with intractable epilepsy	The data was low-pass filtered, demeaned and normalized	Raw data	Three standard Recurrent Neural Network	In the analysis of the forty-nine sets of sEEG data, it was found that certain channels in a number of the data sets consistently showed relatively strong seizure-like activity, despite which of the three architectures was used. These channels represented specific regions in the cortical montage.

[127]	Davidson	2005	EEG data recorded from 5 healthy male volunteers	Data segmentation was performed, producing 1 second windows	Through Power Spectral Density, the logarithm of EEG frequency subbands is calculated and then converted in z-scores	6 LSTM layers with 3 units per layer	—–.
[126]	Srinivasan	2005	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz and data segmentation in 1s epoch	Three frequency domain features through EEG epoch power spectrum and two time domain features	Elman Neural Network	Accuracy of 99.6% with a single input feature, which was better than the results obtained by using the alternative types of deep models with two or more input features.
[100]	Guler	2005	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz and data segmentation in 256 data point segments	Lyapunov exponents	Elman Neural Network	Results from three class classification were: accuracy 96.79%, specificity (healthy segments) - 97.38%, sensitivity (interictal segments) - 96.88%, sensitivity (ictal segments) - 96.13%.
[137]	Kumar	2008	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz	Wavelet and Spectral Entropy	Elman Neural Network	Accuracy using Wavelet Entropy features as input - 98% , Accuracy using Spectral Entropy features as input - 96.7%.
[104]	Ubeyli	2009	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz and data segmentation in 256 data point segments	Feature extraction through Pisarenko PSD, MUSIC PSD and Minimum-Norm PSD algorithms	Elman Neural Network	Average Specificity of 99.5%, Sensitivity of 98.15% and Accuracy of 98.15%.
[131]	Gayatri	2010	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz	The proposed feature for the epileptic detection is named ApEn	Elman Neural Network	No results were published.

[106]	Orhan	2011	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz	Discrete wavelet transform to divide the raw EEG signals in its frequency subbands. From each subband some features are extracted like average power, mean, entropy and standard deviation of the wavelet coefficients	K-means as clustering method and MLPNN classifier	Average accuracy 98.13%, sensitivity 96.90% and specificity 98.66%.
[105]	Iscan	2011	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz	From Cross Correlation output to the time-domain samples, some features were extracted, like peak value of the cross correlogram, instant value corresponding to the peak value, centroid, equivalent width, mean square abcissa. For frequency domain features, Discrete Wavelet Transform is applied to obtain Power Spectral Density from each subband	Various Machine Learning algorithms	—.
[98]	Huang	2016	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz, data segmentation of each sample in 4 segments and data normalization	Raw data	Stacked Sparse Autoencoder	Two class classification with average accuracy of 96.00%, sensitivity of 98.67% and specificity of 93.83%.

[117]	Thodoroff	2016	CHB-MIT EEG database	No preprocessing method	Fast Fourier Transform to extract subbands and application of cubic interpolation with the 2D projection of electrodes coordinates	Recurrent Convolutional Neural Network with Bidirectional LSTM	For patient-specific detection, average sentivity of 95-100% and 0.3 h^{-1} of false positive rate. For Cross patient detection it was obtained an average sensitivity of 85% and a false positive rate of 0.8 per hour.
[144]	Antoniades	2016	Intracranial data from King's College London Hospital	Data segmentation in 160ms segments	Time-frequency features by using a Hanning-tapered window	Convolutional Neural Network	—–.
[118]	Vidyaratne	2016	CHB-MIT EEG database	Band-pass filter between 3-30 Hz and data segmentation in non-overlapping 1 second epochs	Raw Data	Bidirectional LSTM concatenated with Multilayer Perceptron	Two class classification, with overall sensitivity of 100%, specificity(false detections/hour) of $0.08h^{1}$ and detection delay of 7±27 seconds.
[111]	Shah	2017	TUH EEG seizure corpus database	Data Segmentation in 1 second epochs	Raw data	Recurrent Convolutional Neural Network varying the number of 2DCNN layers	The reduction of channels between 20-8 achieved sensitivity between 33-37% with false alarms in the range of 38-50 per 24 hours. False alarms increased (e.g., over 300 per 24 hours) when the number of channels was further reduced. Baseline performance of a system that used all 22 channels was 39% sensitivity with 23 false alarms.
[110]	Golmohammadi	2017	TUH EEG seizure corpus database	Data segmentation in 1 second epochs	Application of standard linear frequency cepstral coefficient-based feature extraction approach, with the extraction of the first and second derivatives of the LFCC's output	Hidden Markov Models with Stacked Denoising Autoencoder, with LSTM and a Recurrent Convolutional Neural Network	Average sensitivity around 30-35%, specificity 75-96% and false positive rate from 7 to 77 h^{-1}

[122]	Nussinov	2017	13959 IEDS from 46 epilepsy patients' records	Recordings were down-sampled from 1024Hz to 256Hz, then band-pass filtered from 0.5-70Hz with a Butter-worth third order filter, and notch filtered at 60 Hz with a Butterworth fourth-order filter	Raw Data	Convolutional Neural Network	—–.
[103]	Ullah	2018	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz and data segmentation using fixed size overlapping windows	Raw data	pyramidal one-dimensional convolutional neural network (P-1D-CNN) models	—–.
[101]	Abdelhameed	2018	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz	Raw data	Deep Convolutional Bidirectional LSTM network	100% accuracy in classification of normal vs ictal events. The three class classification results between the normal, interictal and ictal events accomplished a 98.89% average overall accuracy.
[121]	Tjepkema - Cloostermans	2018	Database constituted with 50 normal EEGs and 50 EEGs that contained focal interictal epileptiform discharges	Data segmentation in 2s epochs and band pass filtered in the 0.5–35 Hz frequency and downsampling to 125Hz	Raw data	Uni and Bidimensional CNN and RNN models	Test AUC achieved was 0.94. Validation results for IEDS was 47.40% for sensitivity, 98.00% for specificity and FPR of $0.6 min^{-1}$. For the normal EEGs, validation specificity was 99.90% and FPR $0.03 min^{-1}$.

[109]	Von Weltin	2018	TUH EEG seizure corpus database	No preprocessing method was applied	Feature extraction is performed using a method called standard linear frequency cepstral coefficient-based (LFCCs) and also the first and second derivative of LFCC's output	Convolutional Recurrent Neural Network (comparison between LSTM and GRU layers)	For CNN/GRU, the specificity value obtained was 30.83%, sensitivity of 91.49% and false positive alarm rate (per 24hours) of 21. For CNN/LSTM model, the specificity obtained was 30.83%, sensitivity of 97.10% and false positive alarm rate (per 24hours) of 6.
[116]	Tsiouris	2018	CHB-MIT EEG database	Signal segmentation in 5s samples	Extraction of time domain, frequency domain, cross correlation and graph theory features	Three different LSTM network architectures	High rates sensitivity and false prediction rates (FPR) of 0.11–0.02 per hour, depending on the duration of the preictal window.
[97]	Ahmedt - Aristizabal	2018	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz	Raw data	Two LSTM network architectures	On a public dataset, an average validation accuracy of 95.54% and an average AUC of 0.958 of the ROC curve between all sets.
[107]	Chowdhury	2019	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz. The noise of EEG signal has been removed by 6th order Butterworth filter. The cutoff frequency of the low pass filter has established at 60 Hz	Raw data	1-D convolutional neural network architecture	Overall results for accuracy, sensitivity and specificity above 99%.

[120]	Daoud	2019	CHB-MIT EEG database	The EEG signals were divided to non-overlapping 5 seconds segments	Raw data	A Deep Convolutional Neural Networks concatenated with Bidirectional LSTM and A Deep Convolutional Autoencoder concatenated with Bidirectional LSTM	Accuracy of 99.60% and false positive rate of 0.004 h^{-1}.
[119]	Tian	2019	CBH-MIT EEG database	No preprocessing method was applied	Fast Fourier Transform was applied to obtain frequency subbands. In time domain features, it was used the raw data and, in time-frequency domains, Wavelet Packet Decomposition was applied to obtain data in the mentioned domain	Convolutional Neural Network	—.
[102]	Dutta	2019	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz and data segmentation in 178 data point vectors that were used as input sequences	Raw data	Three different architectures: stacked simple RNN, LSTM and GRU layers	Overall accuracy obtained was 45,3% for simple RNN, for LSTM 69,48% and for GRU model 73.1%.

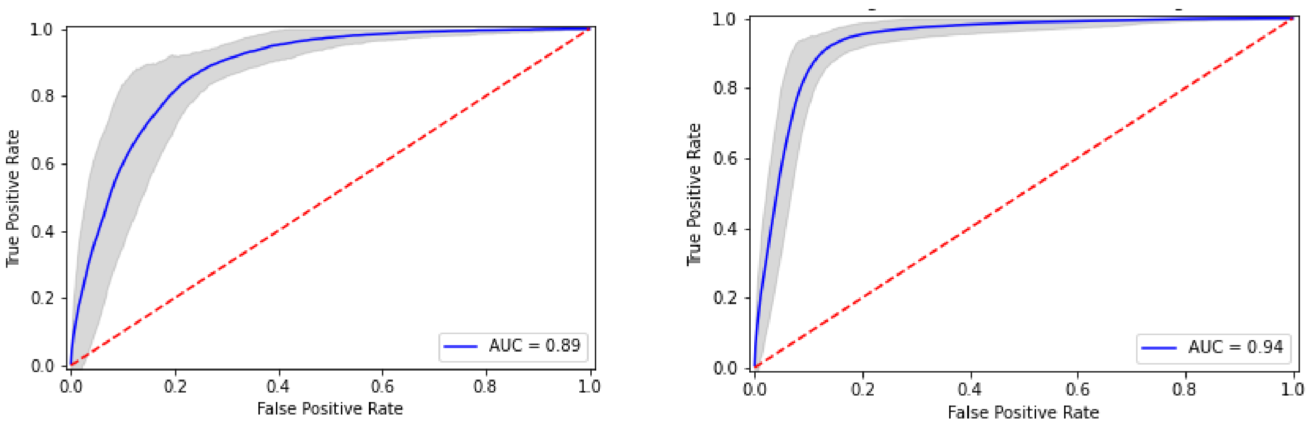
[96]	Zuo	2019	19 patients diagnosed with intractable epilepsy	A zero-phase finite impulse response filter was used to perform bandpass filtering for the data. The cutoff frequencies were 80–200 and 200–500Hz for ripples and fast ripples, respectively. After this, the data as segmented in 1s time series	Raw Data	Convolutional Neural Network	—–.
[94]	Michielli	2019	Sleep-EDF database	Band pass filter 0.3-45Hz, PCA for dimensionality reduction and 30 second data segmented in 1 second time series	A total of 55 frequency and time domain features were extracted from the EEG signals	LSTM + FC network	—–.
[113]	Roy	2019	TUH Abnormal EEG corpus database	No preprocessing method was applied	Raw data	Variations of Deep Convolutional Recurrent Neural Network, with the implementation of dense connections in the recurrent stage and filters of exponentially varying length in 1DCNN layers	Testing Accuracy of 86.57% for achitecture with two factors implmented.

[124]	Thomas	2019	EEG recordings from 156 subjects (93 epileptic patients with annotated IEDs and 63 healthy patients)	The data was downsampled to 128 Hz after applying an anti-aliasing FIR low pass filter. It was also applied two filters: an IIR notch filter of 60 Hz and a 1 Hz high pass-filter	Raw Data	Convolutional Neural Network and Support Vector Machines	——.
[125]	Fukumori	2019	EEG records of six patients with benign epilepsy with centro-temporal spikes	Waveforms were normalized at every channel before all processing. Then, a 1-s epochs were extracted, including 300 ms before and 700 ms after every detected peak	Two feature extraction methods were applied: A band pass filter (1-30Hz) with Discrete Wavelet Transform and the implementation of an 1DCNN layer	Two Recurrent Neural Network with Stacked LSTM and GRU layers	Average AUC around 0.83, with best result for strategy BP filter + DWT + LSTM (0.846).
[108]	Ahmedt - Aristizabal	2019	TUH EEG seizure corpus database	Data segmentation in overlapping windows	Fast Fourier Transform to extract signal frequency subbands	Two architectures: an hybrid LSTM network with an external memory controller and a standard Recurrent Convolutional Neural Network	Weighted F1-score was higher for first architecture (0.945) than for RCNN (0.824).
[99]	Hussein	2019	Dataset from the Department of Epileptology, University of Bonn	Band-pass filter at 0.53–40 Hz and data segmentation of each EEG sample by half	Raw data	Deep Recurrent Neural network concatenated with a Fully Connected layer	Average sensitivity, specificity and accuracy of 100% for two, three and five class classification.

[153]	Li	2020	The database included five brain activities from 109 subjects (eyes closed, imagining opening/closing left fist, imagining opening/closing right fist, imagining opening/closing both fist, imagining opening/closing both feet)	Data segmentation through sliding windows	Raw data	Recurrent Convolutional Neural Networks with different number of layers and with the implementation of LSTM or GRU layers	—–.
[112]	Liu	2020	TUH EEG seizure corpus database	Resampling of signals to 250Hz	Short Time Fourier Transform to obtain frequency domain representation. After that, intensity values for each sample was calculated, with log10	Three Deep bilinear models with CNN, LSTM and both as feature extractors	Average F1-score of 96.70% for bilinear-LSTM, 96.90% for bilinear-CNN and 97.40% for hybrid bilinear.
[123]	Jing	2020	A total of 9571 scalp EEG records with and without interictal epileptiform discharges	The EEGs were filtered (with a 60Hz notch filter and 0.5Hz high pass filter) and resampled to 128 Hz	Raw Data	Convolutional Neural Network	—–.
[115]	Bongiorni	2020	CHB-MIT EEG Database	Data Normalization and Data Segmentation in 1s batches	Time domain (Root Mean Squared Error, Standard Deviation, number of peaks above certain threshold) and Frequency domain (energy density of a spectral component, through FFT and PSD functions) features	Two LSTM architectures with different number of layers	Average accuracy of 97.40%, sensitivity of 19,67% and specificity of 99,70%.

Appendix B

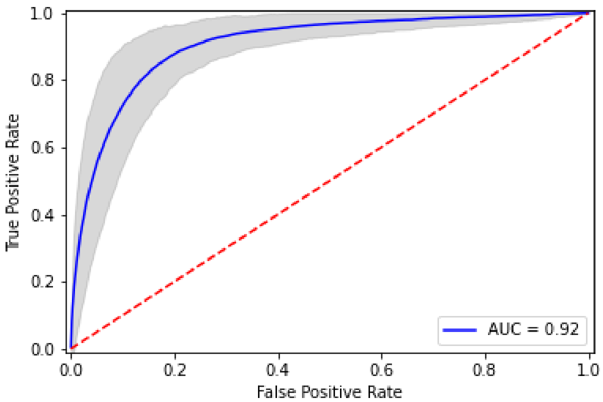
Additional Figures to chapter 5



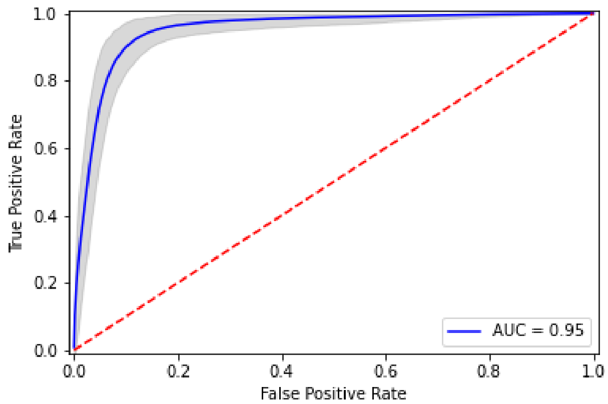
(a) ROC curve LSTM5 model Weight 1:30 Set A

(b) ROC curve LSTM5 model Weight 1:10 Set B

Figure B.1: LSTM5 ROC curves for Sets A and B

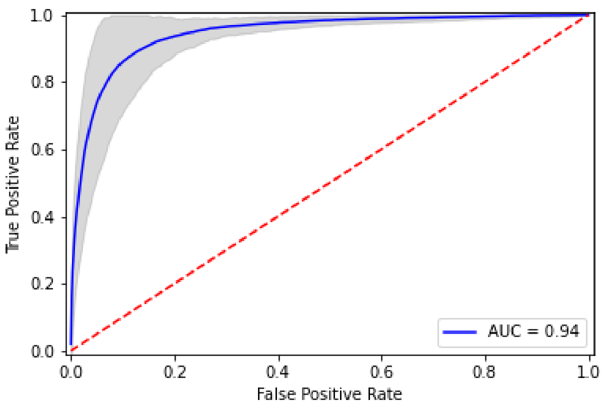


(a) ROC curve CNN+LSTM3 model Weight 1:20 Set A

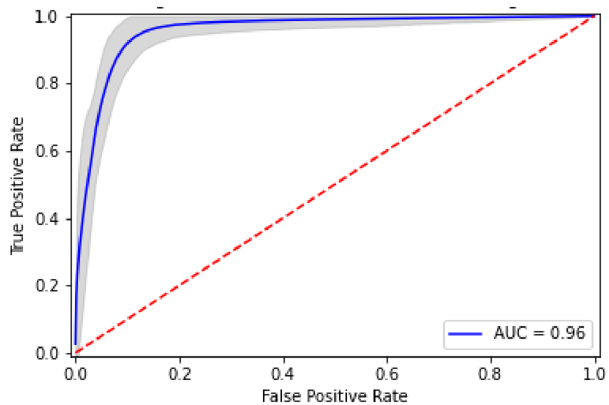


(b) ROC curve CNN+LSTM3 model Weight 1:20 Set B

Figure B.2: CNN+LSTM3 ROC curves for Sets A and B

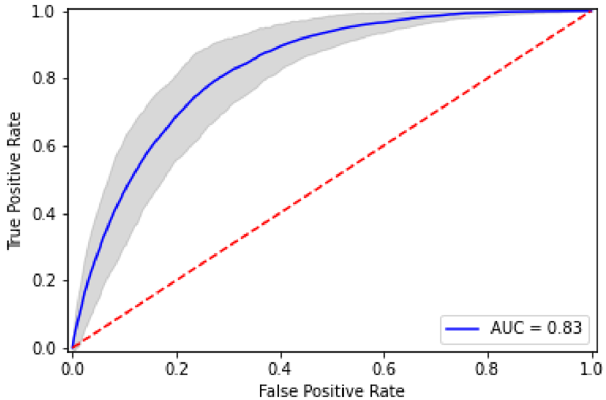


(a) ROC curve CNN+LSTM4 model Weight 1:4 Set A

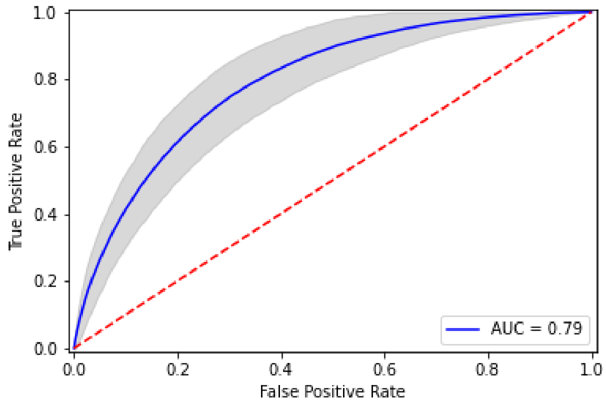


(b) ROC curve CNN+LSTM4 model Weight 1:10 Set B

Figure B.3: CNN+LSTM4 ROC curves for Sets A and B

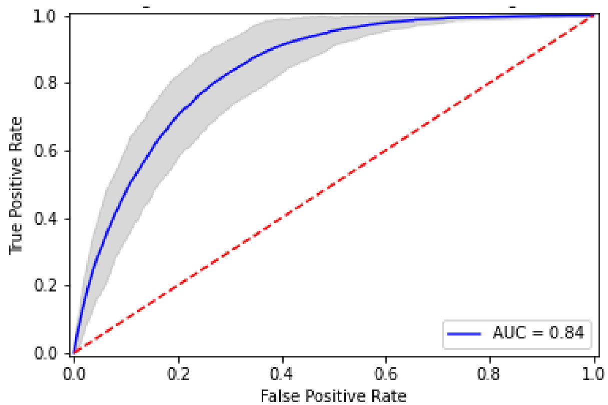


(a) ROC curve Transformer 1 model Weight 1:300 Set A

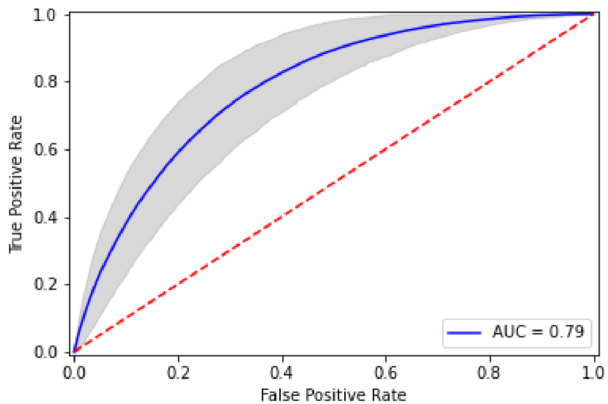


(b) ROC curve Transformer 1 model Weight 1:300 Set B

Figure B.4: Transformer 1 ROC curves for Sets A and B



(a) ROC curve Transformer 3 model Weight 1:275 Set A



(b) ROC curve Transformer 3 model Weight 1:275 Set B

Figure B.5: Transformer 3 ROC curves for Sets A and B

Appendix C

Additional Figures to chapter 6

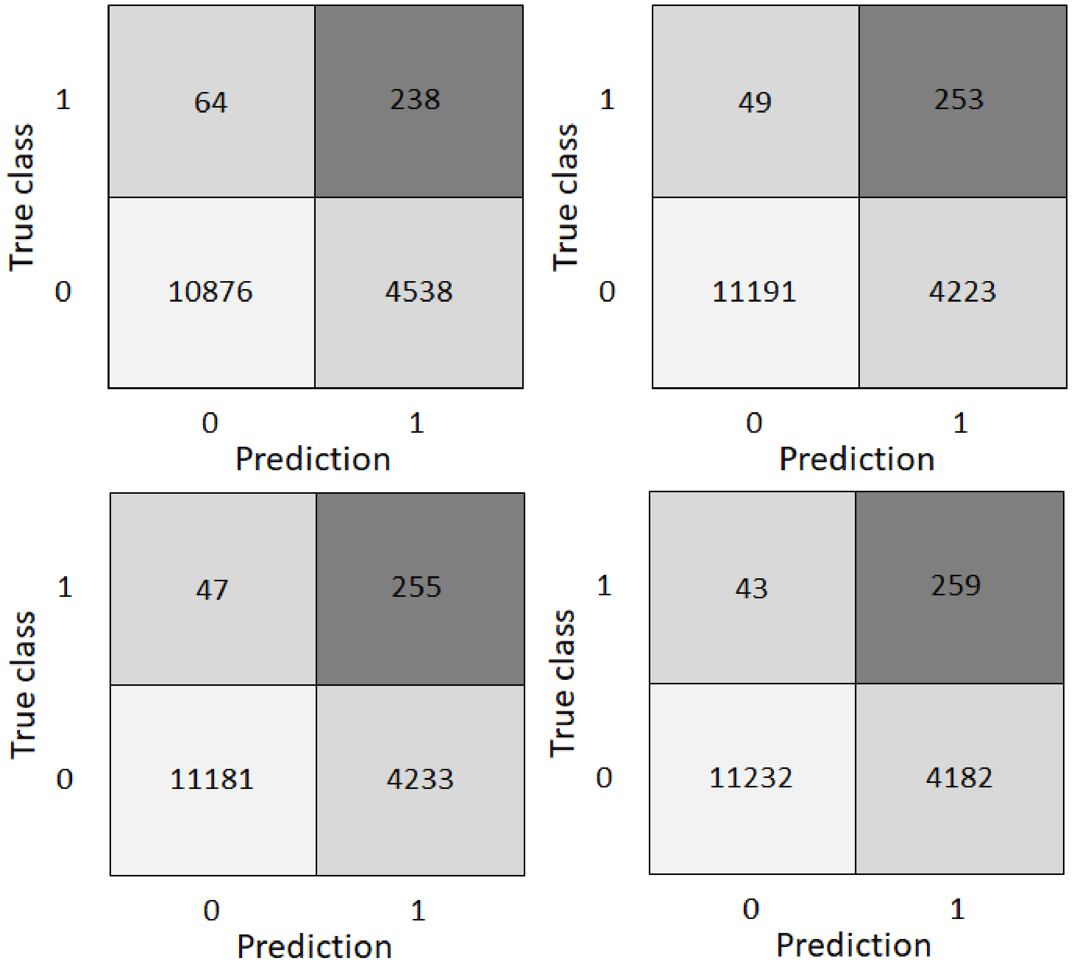


Figure C.1: Non normalized confusion matrices of: upper left) LSTM2; upper right) LSTM3; lower left) LSTM4; lower right) LSTM7 at a threshold decision of 0.5

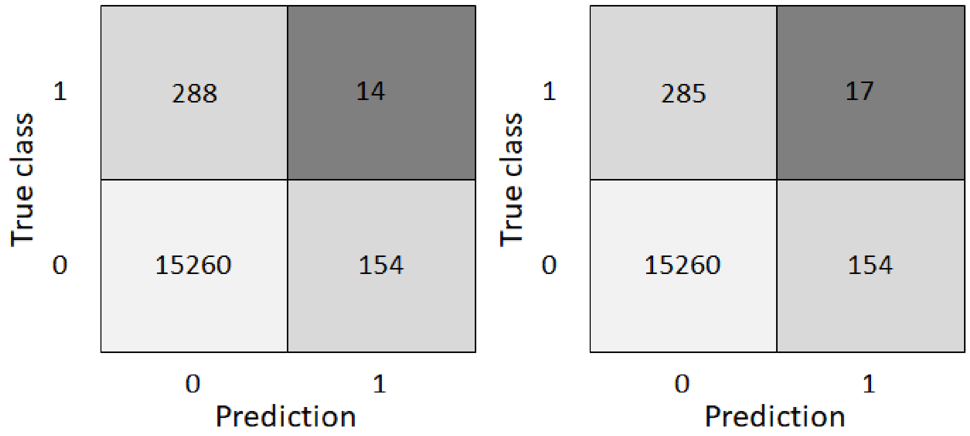


Figure C.2: Non normalized confusion matrices of: left) LSTM4; right) LSTM5 at a evaluation criteria of 99% specificity

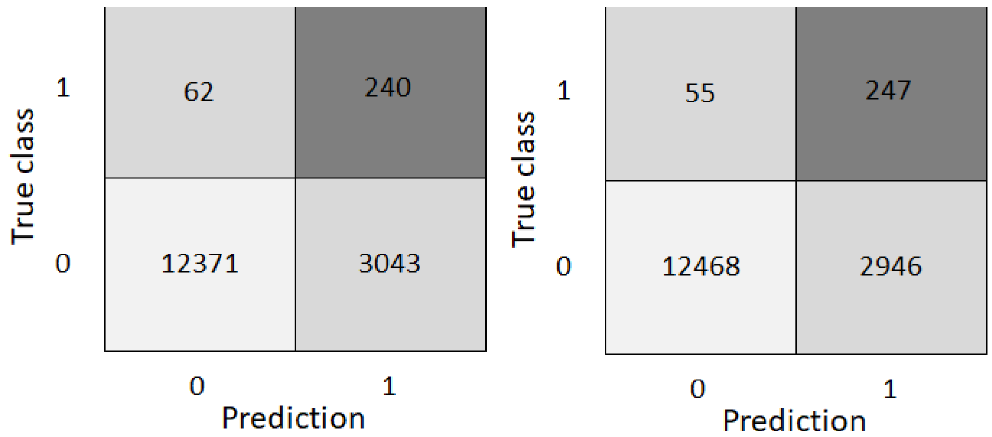


Figure C.3: Non normalized confusion matrices of: left) LSTM4; right) LSTM5 at a evaluation criteria of specificity equal to sensitivity

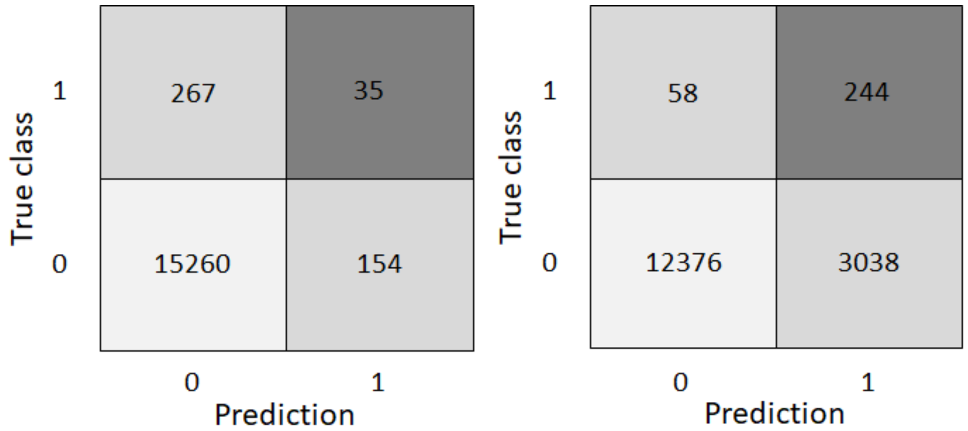


Figure C.4: Non normalized confusion matrices of LSTM5 best model's design: left) for an evaluation criteria of 99% specificity; right) for an evaluation criteria of specificity equal to sensitivity, Set A

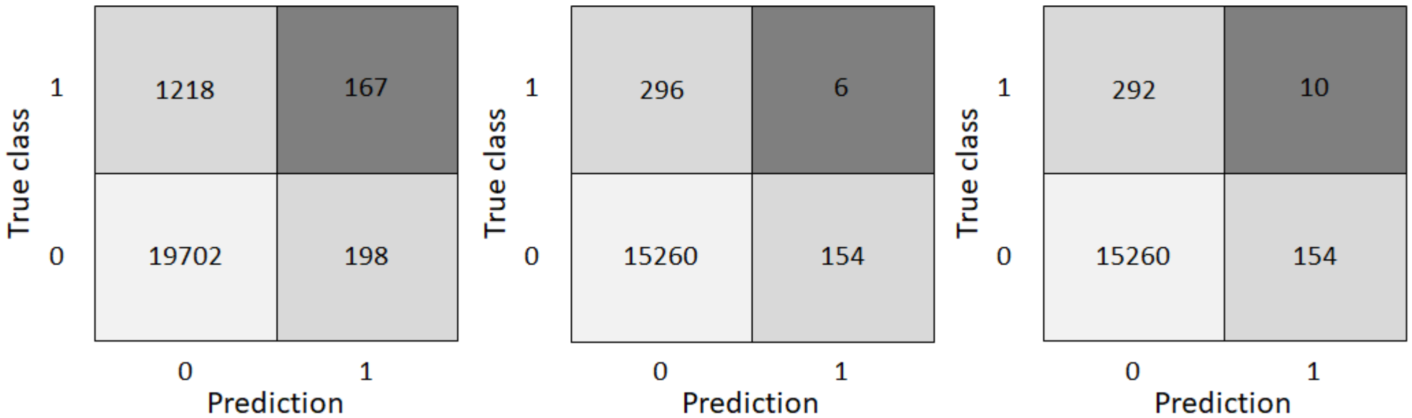


Figure C.5: Non normalized confusion matrices of LSTM5 best model's design for an evaluation criteria of 99% specificity: left) with Set B; center) with Set C; right) with Set D

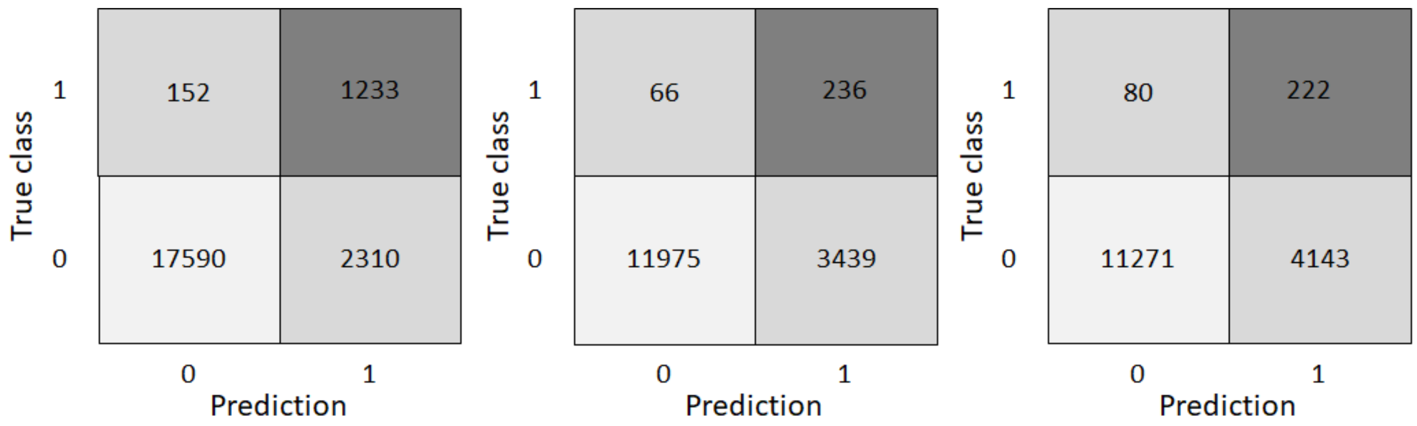


Figure C.6: Non normalized confusion matrices of LSTM5 best model's design for an evaluation criteria of equal value for sensitivity and specificity: left) with Set B; center) with Set C; right) with Set D

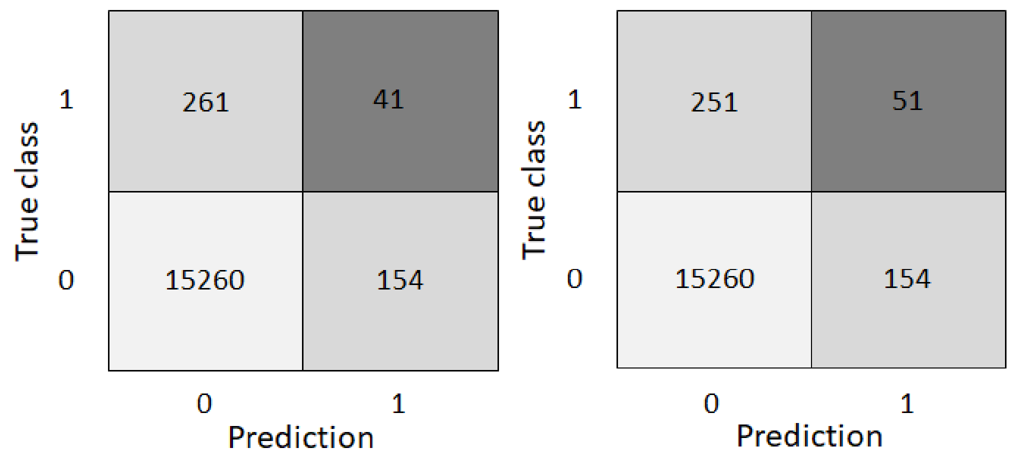


Figure C.7: Non normalized confusion matrices for an evaluation criteria of 99% specificity: left) CNN+LSTM3 model; right) CNN+LSTM4 model

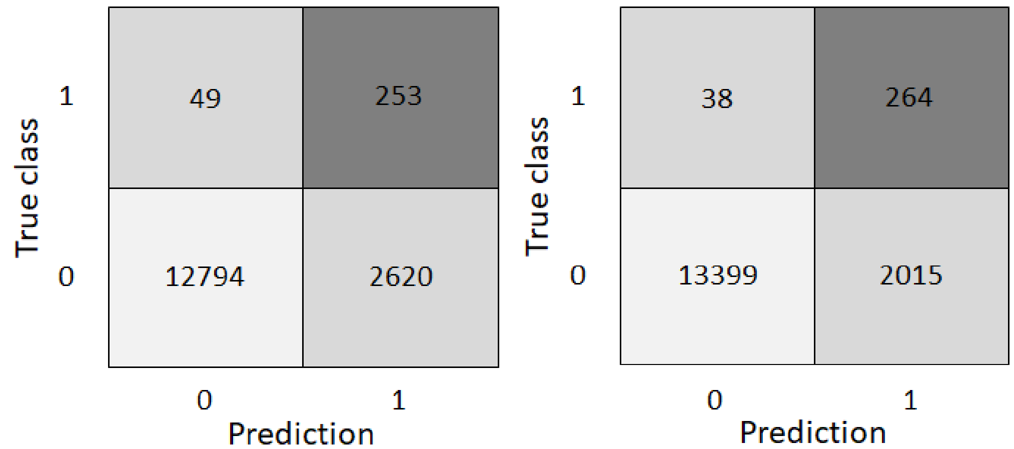


Figure C.8: Non normalized confusion matrices for an evaluation criteria of equal value for sensitivity and specificity: left)CNN+LSTM3 model ; right) CNN+LSTM4 model

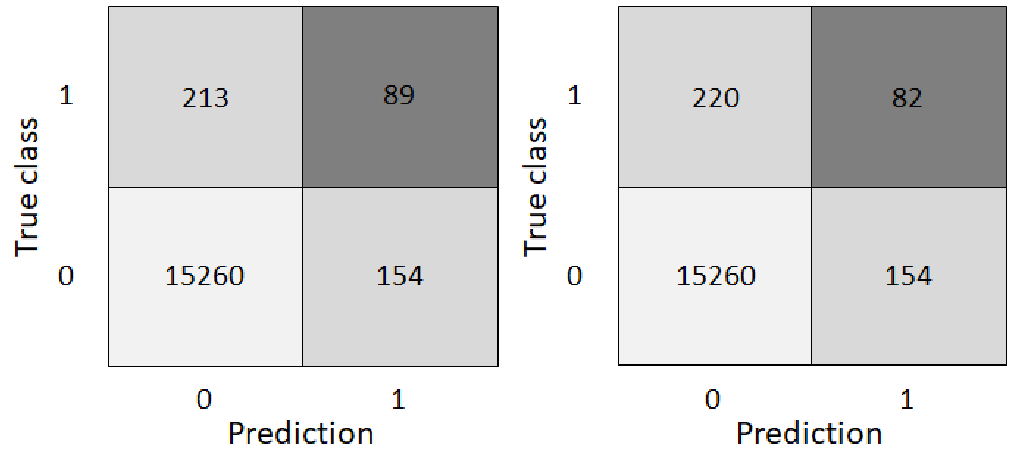


Figure C.9: Non normalized confusion matrices for an evaluation criteria of 99% specificity with the best design of: left) CNN+LSTM3 model; right) CNN+LSTM4 model, with Set A

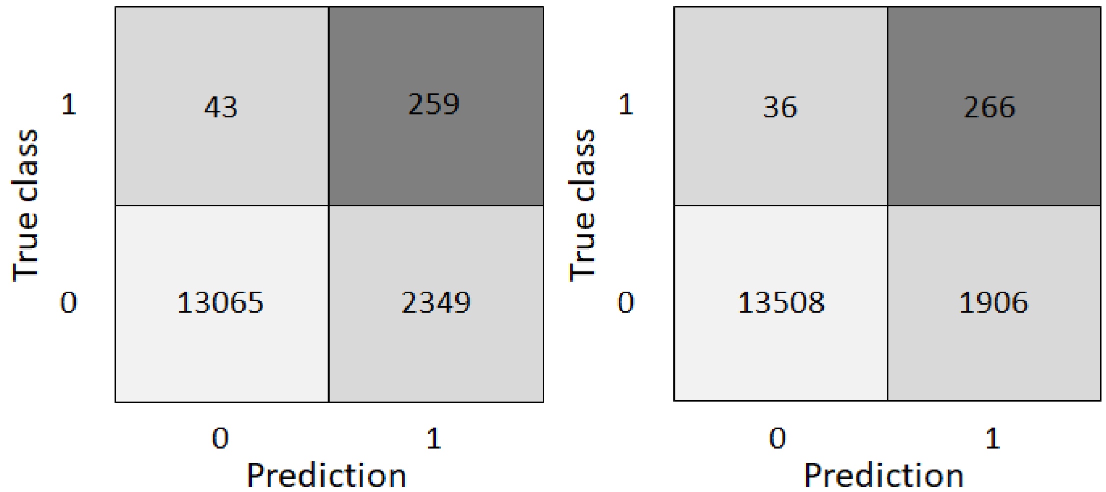


Figure C.10: Non normalized confusion matrices for an evaluation criteria of equal value for sensitivity and specificity with the best design of: left)CNN+LSTM3 model ; right) CNN+LSTM4 model, with Set A

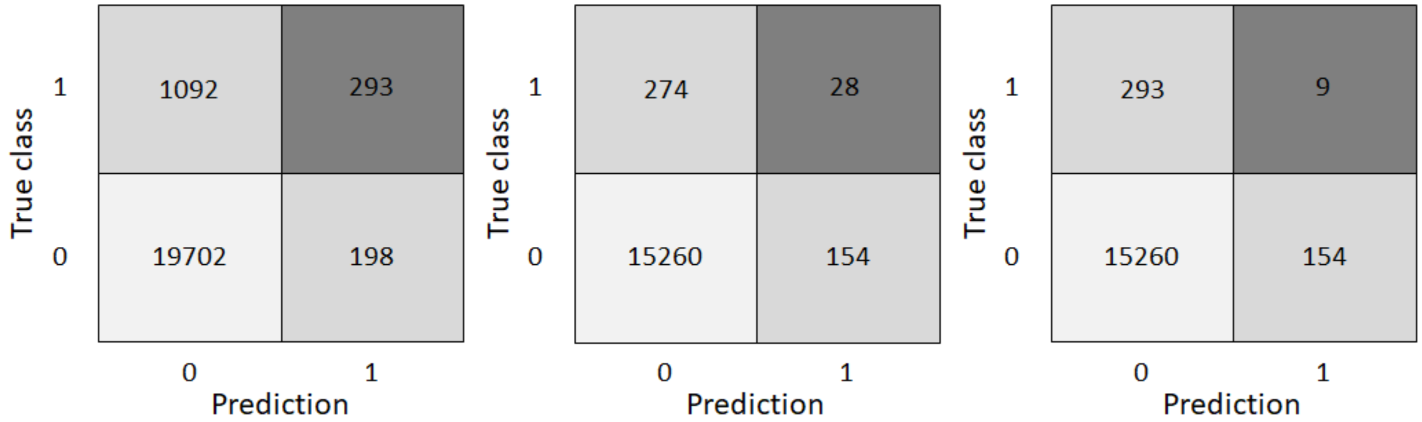


Figure C.11: Non normalized confusion matrices of CNN+LSTM3 best model's design for an evaluation criteria of 99% specificity: left) with Set B; center) with Set C; right) with Set D

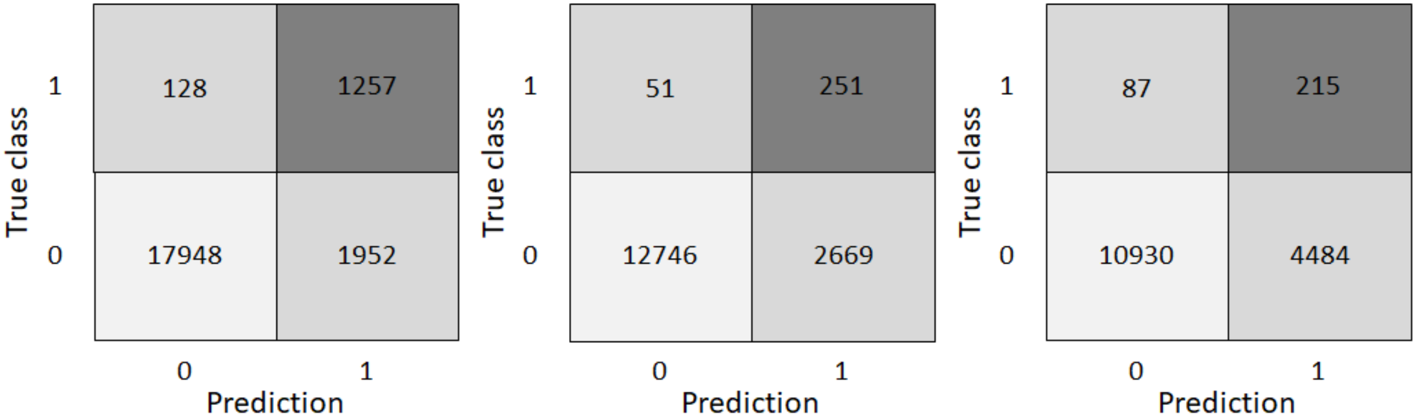


Figure C.12: Non normalized confusion matrices of CNN+LSTM3 best model's design for an evaluation criteria of equal value for sensitivity and specificity: left) with Set B; center) with Set C; right) with Set D

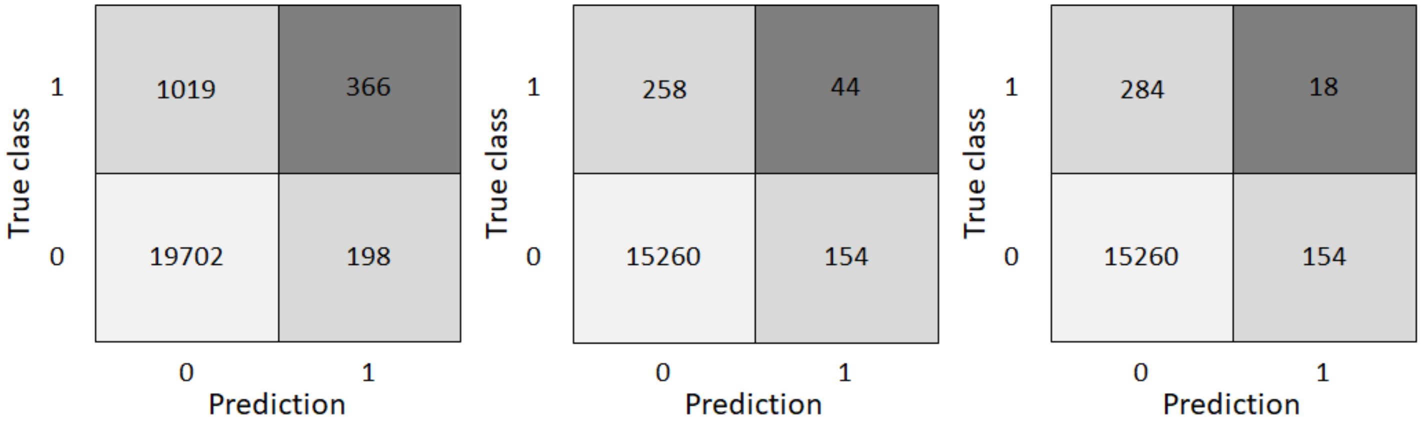


Figure C.13: Non normalized confusion matrices of CNN+LSTM4 best model's design for an evaluation criteria of 99% specificity: left) with Set B; center) with Set C; right) with Set D

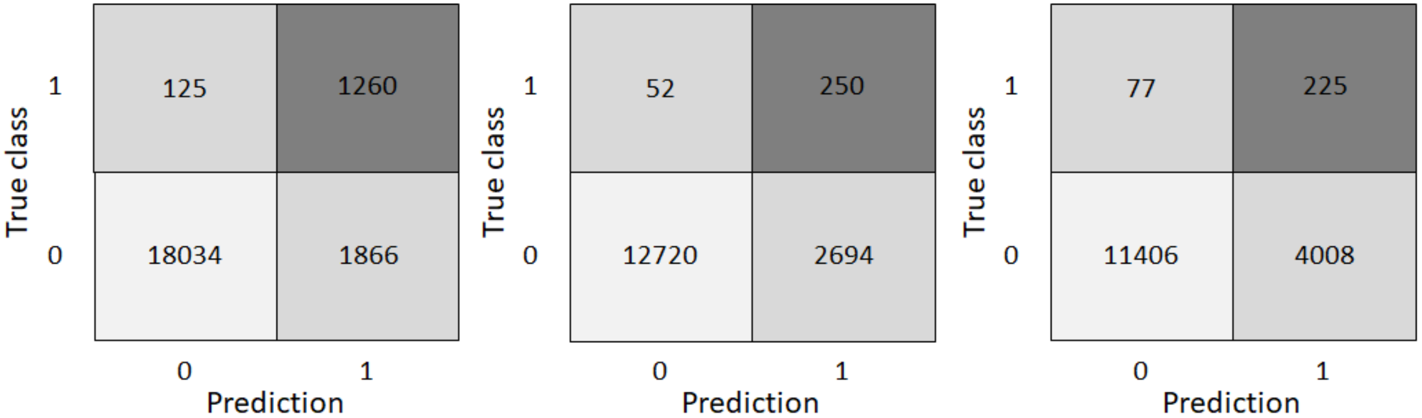


Figure C.14: Non normalized confusion matrices of CNN+LSTM4 best model's design for an evaluation criteria of equal value for sensitivity and specificity: left) with Set B; center) with Set C; right) with Set D

Appendix D

Example of article for publication

DETECTION OF EPILEPSY IN EEGS USING SEQUENCE MODELS

Miguel Marques ⋆ *Luís Teixeira* ⋆ *Catarina da Silva Lourenço* †

⋆ Faculty of Engineering of University of Porto
† Department of Clinical Neurophysiology, University of Twente, the Netherlands

ABSTRACT

The automation of interictal epileptiform discharges through deep learning models can increase assertiveness and reducing time spent on epilepsy diagnosis, making the process faster and more reliable. It was demonstrated that deep sequence networks can be a feasible algorithm to effectively detect IED samples, under different configurations. From three different deep networks, it was reached average AUC values of 0.96, 0.95 and 0.94, with convergence of test specificity and sensitivity values around 90%, which is traduced in a good ability to detect IED samples in EEG records.

***Index Terms*—** Deep Learning, Interictal Epileptiform Discharges, Recurrent Neural Networks, Electroencephalogram

1. INTRODUCTION

Epilepsy is a neurologic disorder characterized by an enduring predisposition to generate epileptic seizures and by the neurobiologic, cognitive, psychological, and social consequences of this condition. An epileptic seizure is a "transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain." Fisher et al. (2005). This event can be expressed by various body reactions, depending on which part of the brain gets involved (such as loss of awareness with body shaking, confusion, and difficulty responding; visual or other sensory symptoms; isolated posturing of a single limb) Johnson (2019).

Nowadays, the diagnosis of epilepsy is done by visual assessment of Electroencephalography recordings Rosenow et al. (2015). Regarding the analysis of EEG recording, interictal epileptiform discharges (IED) detection is one of the most prominent indicatives for the presence of epilepsy disorder. Between seizures, the brain of some patients with epilepsy generates pathological patterns of activity that are clearly distinguished from the activity observed during the seizure itself and normal brain activity. The detection of these patterns requires the visual analysis of EEG recordings, which corresponds to a prolonged and demanding visual assessment. In addition to this, subjectivity and intra/inter-variability are also factors that affect the diagnosis Lodder et al. (2011). All these reasons often leads to difficulties in a precise diagnosis, making, unfortunately, misdiagnosis common, with 20%-30% of epileptic patients misdiagnosed Benbadis (2009); Beach and Reading (2005).

To improve the detection of interictal epileptiform discharges in EEG, diminishing time and increasing the accuracy of diagnosis, deep learning algorithms might be the solution needed. These algorithms have been employed in various subjects and, their application in the health field is becoming more usual and is gaining importance.

The implementation of deep sequence models for the IED detection was surveyed several times. Guler et al. (2005) implemented an Elman Neural Network to test a three-class classification (normal vs interictal vs ictal events), reaching sensitivity for IED segments equals to 96.9%. For a pre-ictal vs interictal classification, Tsiouris et al. (2018) implemented three LSTM networks concatenated with Fully Connected and a final Dense layers. For an event-based evaluation, sensitivity and false positive rate per hour for 15 min, 30 min, 1 hour and 2 hours of pre-ictal window segments was 100.0% for all window segments and 0.107, 0.063, 0.032 and 0.2 h^{-1} for each segment, respectively. On 2018, Tjepkema-Cloostermans et al. (2018) developed a CNN and hybrid CNN+LSTM architectures, with a different number of layers, to properly identify interictal epileptiform discharges. The 2D CNN and 2D CNN+LSTM architectures, which reached the better results, obtained an average AUC of 0.94, with the corresponding value of sensitivity for the IED samples, for all architectures, around 47.4% (range 18.8–78.4%). Regarding the work developed by Daoud and Bayoumi (2019), a Dense Convolutional Neural Network concatenated with a Bidirectional LSTM layer and a Deep Convolutional Stacked Autoencoder (DCAE) concatenated with a Bidirectional LSTM network were implemented, achieved a sensitivity of 99.7% and false positive rate per hour of 0.004 h^{-1} for both methods, under patient-specific detection. Lastly, Bongiorni and Balbinot (2020) implemented a Recurrent Neural Network composed by LSTM layers for the classification between ictal and interictal segments, with overall accuracy of 97.4% , 19.7% for sensitivity and specificity of 99.7%.

A common issue with EEG databases is the lack of IED samples present, which can affect negatively a correct detection. Deep learning algorithms, such as the models mentioned above, require a great amount of data to be trained correctly and to obtain a good predictive performance. So, in order to provide a larger number of IED samples to the deep networks, several data augmentation techniques can be performed. Data overlap with a sliding window or the application of SMOTE technique are examples of techniques commonly used. In addition to this, other alternative that might be explored is the attribution of different weights to the classes present in the database, so the network can give more importance to the minority class and decrease the predictive errors related to the lack of examples of that class.

The aim of this study was to design and test different deep sequence architectures to automatize the detection of IEDs in EEG recordings, improving the efficiency of the EEG patterns classification.

2. METHODS

2.1. EEG data and pre-processing

The database is comprised of two second EEG epochs with two types of EEG patterns. Normal EGGs, from 67 healthy subjects without any epileptic episodes and interictal epileptiform discharges from focal (50 patients) and generalized (49 patients) epilepsy. As pre-processing steps, the data was filtered in the 0.5-35 Hz range to reduce artefacts and downsampled to 125 Hz to reduce input size (and consequently computational load). The acquisition of the EEG signals was performed under three different electrode montages: longitudinal bipolar, common average and source derivation. After that, each recording was split into two second non-overlapping epochs, yielding a 18x250 (channels x time) matrix for each epoch.

2.2. Database Structuring

For the development of this work, four different datasets were created. First, Set A is constituted by all the normal EEG samples at the negative class and focal and generalized IED samples at the positive class. Secondly, Set B was created with the same signals presented in Set A, but with the records from the three different montages rather than just from one, as in Set A. After that, other two sets were created, with the intent to increase artificially positive class samples. Set C was created with the IED samples increased by the implementation of SMOTE technique and Set D with the application of a data overlap around 70% with a sliding window. In the Table 1 a description of all datasets can be seen.

The data was randomly split 80% in train and validation and 20% in test. After, in order to decrease the difference between positive and negative samples, the negative samples present in the set for train and validation were divided in 5 different parts. All the positive samples were saved in the 5 parts and, on the other hand, negative class samples were distributed similarly between the five parts, as a manner to under sampling the total number of negative class samples. In the end, five-cross validation was implemented to each part and the evaluation results were calculated as an average of the training to the five parts.

Table 1: Databases description (number of epochs for train, validation and test)

<table>
<tr><th rowspan="3">Dataset</th><th colspan="3">Train & Validation</th><th colspan="3">Test</th></tr>
<tr><th colspan="2">Epochs</th><th rowspan="2">Duration (h)</th><th colspan="2">Epochs</th><th rowspan="2">Duration (h)</th></tr>
<tr><th>Negative</th><th>Positive</th><th>Negative</th><th>Positive</th></tr>
<tr><td rowspan="2">Set A</td><td colspan="2">43589</td><td rowspan="2">24.2 h</td><td colspan="2">15716</td><td rowspan="2">8.7 h</td></tr>
<tr><td>41369</td><td>2220</td><td>15414</td><td>302</td></tr>
<tr><td rowspan="2">Set B</td><td colspan="2">145709</td><td rowspan="2">80.9 h</td><td colspan="2">21285</td><td rowspan="2">11,8</td></tr>
<tr><td>139373</td><td>6336</td><td>19900</td><td>1385</td></tr>
<tr><td rowspan="2">Set C</td><td colspan="2">88320</td><td rowspan="2">49.1 h</td><td colspan="2">15716</td><td rowspan="2">8.7 h</td></tr>
<tr><td>44160</td><td>44160</td><td>15414</td><td>302</td></tr>
<tr><td rowspan="2">Set D</td><td colspan="2">50249</td><td rowspan="2">27.9 h</td><td colspan="2">15716</td><td rowspan="2">8.7 h</td></tr>
<tr><td>41369</td><td>8880</td><td>15414</td><td>302</td></tr>
</table>

2.3. Deep Learning models

All the deep models were implemented at python 3.7, using Keras 2.4.3 and a CUDA-enabled NVIDIA GPU. For the deep sequence models (Recurrent and hybrid Recurrent Convolutional Neural Networks), Stochastic Optimization was calculated with Adam optimizer, with a learning rate of 10^{-3}, β_1 of 0.9, β_2 of 0,99 and ϵ equal to 10^{-7}. Five different neural network architectures were explored, three deep sequence models and two attention-based networks. Regarding the deep sequence models, the first networks designed was a five-layer Long Short-term Memory neural network Goodfellow et al. (2016), named LSTM1, with a total of 256 units per layer and, in addition, a Dropout layer with 0.35 of dropout rate associated with each LSTM layer.

The other two models are hybrid Recurrent Convolutional Neural Network, with a similar design, inspired by Abdelhameed et al. (2018). The first one, called CNN+LSTM1, had a front-end that was constituted by four 1D convolutional layers, alternating with three 1D Max-pooling layers. In turn, a bi-directional LSTM layer was connected as the back-end of the deep neural network. For this network, Batch Normalization and Dropout with a dropout rate of 0.1 layers were applied after each layer. For the last model, called CNN+LSTM2, the main difference with CNN+LSTM1 was the application of 2D Convolutional and Max-pooling layers instead of 1 dimensional functions.

Concerning the attention-based networks, also called Transformer, it is a recent type of deep network, first published in 2017, that works with different mechanisms used by other deep models, since it has an attention-based process. It achieved remarkable results in several areas like natural language processing, image processing, among others. On the other hand, there are few studies made with the application of these deep structures dealing with physiological signals processing and, consequently, IED detection Dosovitskiy et al. (2020); Song et al. (2018); Jiang et al. (2020); Yan et al. (2019). Regarding the architectures designed to test these attention-based networks, these designs were motivated by Vaswani et al. (2017). The architecture designed included six Attention Modules, working as an encoder and, after that, an 1D Convolutional layer, a Flatten layer and Multi Layer Perceptron was used to perform two class classification. The modifications on the architecture focused on the number of units in the Dense layers, the dimensions of key and value matrices and the number of heads in the Multi Head function, under a correlation rule used in Vaswani et al. (2017), where the dimensions of key and value matrices are equal to the number of units of Dense layers divided by the number of heads used in the Multi Head Function. So, maintaining the same number of four heads, the units of Feed Forward layers and dimensions of the Multi Head function changed between 64 and 16 and between 16 and 4, respectively.

At the experiments performed, it was implemented a batch size of 512 and 60 training epochs.

2.4. Performance Assessment

Regarding the binary classification problem, several evaluation metrics were calculated to assess the performance obtained by the deep networks. Test Sensitivity, Specificity, AUC, the True Positive and False Positive rate per hour were calculated, using two different decision thresholds: a threshold to acquire a specificity of 99% and intersection between sensitivity and specificity. In addition to this, an average ROC curve was obtained for several five-fold cross-validation loops.

3. RESULTS

Regarding the results obtained for the LSTM1 model, several experiments were made, under the implementation of Set A, in order to design the more efficient architecture to distinguish properly IED samples from normal EEG epochs. This came up with a LSTM network containing five layers, as mentioned in section 2.3. For Set A, the test sensitivity value reached 11.4% for 99.0% specificity, and intersection between sensitivity and specificity of 80.5%. After the experiments made with Set A, LSTM1 was tested with the three remaining datasets, to study if the increase of total number of positive class samples could have a positive effect in the network's performance. Regarding Set B results, the sensitivity value reached 12.1% for 99.0% specificity and the convergence value for sensitivity and specificity reached 88.7%. With the implementation of Set B, the average AUC value improved from 0.89 to 0.94. Afterwards, Sets C and D were also implemented. The performance achieved with these sets obtained sensitivity values of 1.9% and 3.4% for 99.0% specificity and equal value for sensitivity and specificity of 78.0% and 73.4%, with average AUC values of 0.84 and 0.76 for Sets C and D, respectively. In Table 2 , all the aforementioned results are shown.

Secondly, similar experiments were performed over CNN+LSTM1 and CNN+LSTM2, reaching to different designs. The 1D CNN+LSTM network reached its best performance with 32 units on the Convolutional and Max Pooling layers, with the association of Batch Normalization and Dropout with a dropout rate of 0.1. On the other hand, CNN+LSTM2 reached the best performance with 128 units per layer. For the trials performed over CNN+LSTM3 model, the performance, under the implementation of Set A, reached 29.6% of sensitivity for 99.0% specificity and intersection of sensitivity and specificity of 85.2%. Secondly, the best performance was achieved with the application of Set B, with AUC of 0.95, 21.2% of sensitivity for 99.0% specificity and the intersection of 90.4% between sensitivity and specificity.For the implementation of Set C, it was reached, for 99.0% specificity, sensitivity of 9.2% and equal value between sensitivity and specificity of 82.9%, for an average AUC of 0.90. Ultimately, for Set D, the best results reached an average AUC of 0.75, with a convergence of sensitivity and specificity of 71.0% and 2.9% test sensitivity for 99.0% specificity. In Table 3, all the results can be seen.

Concerning the implementation of Set A on CNN+LSTM2 model, at a 99.0% specificity criteria, sensitivity reached 27.9%, with a convergence of sensitivity and specificity of 87.9%, corresponding to an average AUC value of 0.94. Looking at the implementation of the other three datasets, the best performance was obtained with the implementation of Set B, reaching an overall AUC of 0.96, sensitivity of 26.1% to 99.0% specificity and convergence of specificity and sensitivity around 90.6%. Regarding the application of Set C, the best performance achieved an average AUC of 0.90, reaching 14.7% of sensitivity to 99.0% specificity and 82.6% for specificity equal to sensitivity. At last, the results for Set D's application are 6.0% for 99.0% specificity, 74.2% for sensitivity and specificity equality criteria and AUC of 0.76. In Table 4, all the results for CNN+LSTM2 model can be seen

At last, the Transformer models did not reach such a good performance as done with the other models. Concerning Transformer 1 exploration with Set A, this model achieved, for 99% specificity, a sensitivity of 3.2%, with an intersection between sensitivity and specificity of 75.7%, corresponding to an average AUC value of 0.83. In addition to this, Transformer 2 achieved, with the application of Set A, a convergence value of sensitivity and specificity of 75.6% and sensitivity of 1.3% for 99% specificity, for an average AUC value of 0.84. Regarding the implementation of the other sets, only set B was the dataset explored, due to the improvement of performance achieved by the deep sequence models when this set was applied. Regarding the results obtained by the two Transformer models, they achieved, for a 99% specificity, 2.2% and 0.8% of sensitivity, intersection between sensitivity and specificity of 71.7% and 71.3% and an average AUC value of 0.79, respectively for Transformer 1 and 2 models.

Table 2: Final Results for the LSTM1 model, with the implementation of Sets A, B, C and D

			Specificity=99.0%				Specificity=Sensitivity				
Model	Class Weight	Dataset	Sensitivity	Specificity	TP / hour	FP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
LSTM1	1:30	Set A	11.4%	99.0%	3.96	17.64	80.5%	80.5%	27.95	347.92	0.89
LSTM1	1:10	Set B	12.1%	99.0%	14.11	16.74	88.7%	88.7%	104.23	195.39	0.94
LSTM1	1:10	Set C	1.9%	99.0%	0.64	17.64	78.0%	78.0%	27.06	393.89	0.84
LSTM1	1:20	Set D	3.4%	99.0%	1.19	17.64	73.4%	73.4%	25.42	474.45	0.76

Table 3: Results for best performances of CNN+LSTM1 model , with the implementation of Sets A, B, C and D

			Specificity=99.0%				Specificity=Sensitivity				
Model	Class Weight	Dataset	Sensitivity	Specificity	TP / hour	FP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
CNN + LSTM1	1:20	Set A	29.6%	99.0%	10.25	17.64	85.2%	85.2%	29.63	269.03	0.92
CNN + LSTM1	1:20	Set B	21.2%	99.0%	24.80	16.74	90.4%	90.4%	106.50	165.10	0.95
CNN + LSTM1	1:1	Set C	9.2%	99.0%	3.18	17.64	82.9%	82.9%	28.77	305.61	0.90
CNN + LSTM1	1:10	Set D	2.9%	99.0%	1.01	17.64	71.0%	71.0%	24.60	513.59	0.75

In Table 5, all the results for Transformer 1 and 2 models can be seen.

4. DISCUSSION

As mentioned before, the main objective of this dissertation was to explore deep sequence architectures capable of detect effectively IED samples in epileptic EEG records. For a reliable accomplishment of this objective, a main obstacle to overtake was the imbalance of the first dataset, Set A. The total presence of positive class samples correspond to 4.2% of total EEG samples. So, having this in mind, some decisions were made to decrease the difference of class samples, to achieve a better performance by the networks. The first decision was relative to the organization of the datasets in six different parts, one for test and the rest for train and validation. The particularity about the latter five parts was that, while the negative samples were similarly distributed throughout these parts, the positive samples were all repeated at each one of them. This distribution was made to under sampling the negative class samples at each training phase. Besides this decision, other techniques were implemented to overcome the imbalance issue. One example of this was the attribution of different weights to the class samples and the creation of other datasets with a increase of IED samples total number.

Between all the techniques implemented to reduce the influence of the databases imbalance, different class weights were attributed at every experiment performed, in order to survey the influence of these several class weights in the ability of the models to learn feasibly the principal features of the EEG samples. As can be seen in the Results chapter, the class weight pair that matched with the best performance was always different, between 1:10, 1:20 and 1:30. In this way, no standardized or common behavior between the weight of the positive class and the algorithm's ability to detect correctly IEDs was confirmed, i.e., not always the highest weight attributed to the positive class is traduced in a better algorithm performance.

Regarding the comparison between the five deep sequence models, CNN+LSTM2 achieved the best performances for most of the datasets, with the highest AUC values reached for Set A, B and D around 0.94, 0.96 and 0.76 respectively. The 2D hybrid Recurrent Convolutional Neural Network was the model that achieved, for most of the experiments, the best predictive ability to properly distinguish normal EEG from IED samples. The CNN+LSTM2 architecture benefited from the grid structure which the EEG epochs were organized, acquiring more efficiently the ability to distinguish the two types of EEG patterns. Even though CNN+LSTM2 model achieved the best results, other models also obtained a remarkable performance. Taking this into consideration, all the work developed during this dissertation permit the establishment of a map with the various architectures that can deal effectively with the EEG datasets. These models have distinct features, with more or less parameters, time training and, obviously, classification performance, which can be significant to decide if a model can fit better the needs of a certain situation. For example, even though CNN+LSTM2 reached the best performance, the design CNN+LSTM1 also achieved relevant results, with a more simpler architecture, possible to be used in a situation which the computational power and resources are limited. For a better comprehension of this diversity, Table 6 retains the number of parameters and training time of the best models and Figures 1 and 2 shows a plot with the different ROC curves acquired from the neural networks that achieved the best performances under implementation of Sets A and B, respectively. This work was developed in continuity with the studies made by Lourenço et al. (2020) and da Silva Lourenço

Table 4: Results for best performances of CNN+LSTM2 model , with the implementation of Sets A, B, C and D

			Specificity=99.0%				Specificity=Sensitivity				
Model	Class Weight	Dataset	Sensitivity	Specificity	TP / hour	FP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
CNN + LSTM2	1:4	Set A	27.3%	99.0%	9.42	17.64	87.9%	87.9%	30.49	218.31	0.94
CNN + LSTM2	1:10	Set B	26.4%	99.0%	30.95	16.74	90.6%	90.6%	106.52	157.85	0.96
CNN + LSTM2	1:10	Set C	14.7%	99.0%	5.07	17.64	82.6%	82.6%	28.67	308.56	0.90
CNN + LSTM2	1:20	Set D	6.0%	99.0%	2.08	17.64	74.2%	74.2%	25.71	459.04	0.76

Table 5: Final Results for the Transformer 1 and 3 models, with the implementation of Sets A and B

			Specificity=99.0%				Specificity=Sensitivity				
Model	Class Weight	Dataset	Sensitivity	Specificity	TP / hour	FP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
Transformer1	1:300	Set A	3.2%	99.0%	1.11	17.64	75.7%	75.7%	26.22	432.77	0.83
Transformer1	1:300	Set B	2.2%	99.0%	2.57	16.74	71.7%	71.7%	84.43	479.59	0.79
Transformer2	1:275	Set A	1.3%	99.0%	0.43	17.64	75.6%	75.6%	26.17	433.73	0.84
Transformer2	1:275	Set B	0.8%	99.0%	0.93	16.74	71.3%	71.3%	83.70	487.93	0.79

et al. (2021), which performed similar experiments on a VGG model using the same Sets A and B. Regarding Set A implementation, the baseline model achieved better performance than the deep sequence models, with a sensitivity of 79.0% for 99.0% specificity and a convergence between sensitivity and specificity of 93.0%, corresponding to an average AUC value of 0.96, the highest in comparison with the other AUC values obtained for Set A, as can be seen in Table 7. Similar results were obtained for Set B implementation on the baseline model. This network reached 74.2% of sensitivity for 99% specificity. For the intersection between sensitivity and specificity, that convergence did not occur. Instead, close values of sensitivity and specificity were obtained, which were 89.5% and 93.5%, respectively. The VGG model was capable of learn the main features of both EEG patterns and distinguish more correctly the IED samples from the normal EEG samples than the proposed networks tested at this work.

Table 6: Training time, for Set A and B, and total number of parameters for the deep models with the best performances

Model	Training time per epoch		Total number of Parameters
	Set A	Set B	
Baseline VGG	4 s	13 s	45570306
LSTM1	9 s	28 s	2383105
CNN+LSTM1	1 s	4 s	27777
CNN+LSTM2	6 s	19 s	1389411
Transformer 1	5 s	17 s	4018361
Transformer 2	8 s	25 s	4162937

Looking at the utilization of the datasets, the implementation of Set B was the only dataset that improved the performance of the deep sequence models, when compared with Set A's results. The presence of data samples from the three montages allowed the LSTM1, CNN+LSTM1 and CNN+LSTM2 networks to learn more effectively specific patterns from the IEDs and normal EEG samples, which improved the ability of the models to distinguish these two distinct physiological patterns. However, In the case of attention-based models, the increase of diversity in positive and negative class samples was not helpful for the upgrade of model's predictive ability. One reason for these worse results is the fact that the model's architecture was not capable of dealing with the increase of complexity of the signals used as input data. The more diversity means the more information to learn and properly understand and the attention-base modes were not able to capture the specific features from the negative and positive class samples. Regarding the application of Sets C and D did not develop the same success as the application of Set B. These two sets were generated with the intent of increase the total number of positive class samples through an artificial technique (SMOTE and data overlap with sliding window). For both datasets, the performance achieved was worse than the one reached with the implementation of Sets A and B. Neither the creation of artificial positive samples from the original epochs nor the creation of new segments through the copy of parts of the original positive samples enabled the neural network to attain a better predictive ability. Concerning the influence of these two datasets, Set C reached superior results compared to Set D, showing that SMOTE technique is a more efficient method to produce artificially more positive class samples rather than overlap the information already existent in the original IED samples.

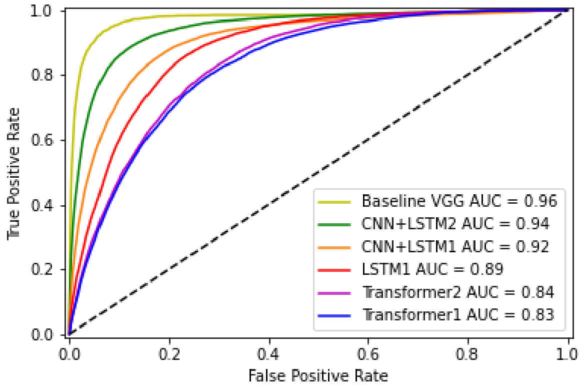


Fig. 1: Plot of the ROC curve correspondent to the models with best performance, with Set A

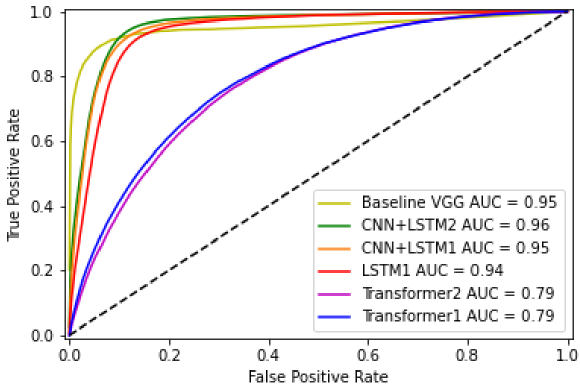


Fig. 2: Plot of the ROC curve correspondent to the models with best performance, with Set B

5. CONCLUSION

The principal conclusion is that is feasible to implement deep sequence models to correctly classify interictal epileptiform discharges between normal EEG samples, in order to assess the presence or absence of epilepsy. Furthermore,the work allowed the investigation of a common issue related to EEG signals database, which is the imbalance of the various EEG patterns present in the database, and to explore several techniques to overcome this issue and potentiate the performance of the networks, in order to achieves a reliable predictive ability.

References

Abdelhameed, A. M., Daoud, H. G., and Bayoumi, M. (2018). Deep Convolutional Bidirectional LSTM Recurrent Neural Network for Epileptic Seizure Detection. *2018 16th IEEE International New Circuits and Systems Conference, NEWCAS 2018*, pages 139–143.

Beach, R. and Reading, R. (2005). The importance of acknowledging clinical uncertainty in the diagnosis of epilepsy and non-epileptic events. *Archives of Disease in Childhood*, 90(12):1219–1222.

Benbadis, S. (2009). The differential diagnosis of epilepsy: A critical review. *Epilepsy and Behavior*, 15(1):15–21.

Bongiorni, L. and Balbinot, A. (2020). Evaluation of recurrent neural networks as epileptic seizure predictor. *Array*, 8:100038.

da Silva Lourenço, C., Tjepkema-Cloostermans, M. C., and van Putten, M. J. (2021). Efficient use of clinical EEG data for deep learning in epilepsy. *Clinical Neurophysiology*, 132(6):1234–1240.

Daoud, H. and Bayoumi, M. A. (2019). Efficient Epileptic Seizure Prediction Based on Deep Learning. *IEEE Transactions on Biomedical Circuits and Systems*, 13(5):804–813.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2020). An Image

Table 7: Best performance results obtained for Set A, including baseline model from Lourenço et al. (2020)

		99% Specificity			Specificity = Sensitivity				
Model	Set	Sensitivity	Specificity	TP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
Baseline VGG	A	79.0%	99.0%	40.27	93.0%	93.%	47.72	122.41	0.96
CNN+LSTM2	A	27.3%	99.0%	9.42	87.9%	87.9%	30.49	218.31	0.94
CNN+LSTM1	A	29.6%	99.0%	10.25	85.2%	85.2%	29.63	269.03	0.92
LSTM1	A	11.4%	99.0%	3.96	80.5%	80.5%	27.95	347.92	0.89
Transformer1	A	3.2%	99.0%	1.11	75.7%	75.7%	26.22	432.77	0.83
Transformer2	A	1.3%	99.0%	0.43	75.6%	75.6%	26.17	433.73	0.84

Table 8: Best performance results obtained for Set B, including baseline model from da Silva Lourenço et al. (2021)

		99.0% Specificity			Specificity = Sensitivity				
Model	Set	Sensitivity	Specificity	TP / hour	Sensitivity	Specificity	TP / hour	FP / hour	AUC
Baseline VGG	B	74.2%	99.0%	38.05	89.5%	93.5%	105.30	114.00	0.95
CNN+LSTM2	B	26.4%	99.0%	30.95	90.6%	90.6%	106.52	157.85	0.96
CNN+LSTM1	B	21.2%	99.0%	24.80	90.4%	90.4%	106.50	165.10	0.95
LSTM1	B	12.1%	99.0%	14.11	88.7%	88.7%	104.23	195.39	0.94
Transformer1	B	2.2%	99.0%	2.57	71.7%	71.7%	84.43	479.59	0.79
Transformer2	B	0.8%	99.0%	0.93	71.3%	71.3%	83.70	487.93	0.79

is Worth 16x16 Words: Transformers for Image Recognition at Scale. *CoRR*, pages 1–21.

Fisher, R. S., Van Emde Boas, W., Blume, W., Elger, C., Genton, P., Lee, P., and Engel, J. (2005). Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia*, 46(10):1701–1702.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*, volume 29. MIT Press. http://www.deeplearningbook.org.

Guler, N. F., Ubeyli, E. D., and Guler, I. (2005). Recurrent neural networks employing Lyapunov exponents for EEG signals classification. *Expert Systems with Applications*, 29(3):506–514.

Jiang, K., Liang, S., Meng, L., Zhang, Y., Wang, P., and Wang, W. (2020). A two-level attention-based sequence-to-sequence model for accurate inter-patient arrhythmia detection. *Proceedings - 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020*, (September 2020):1029–1033.

Johnson, E. L. (2019). Seizures and Epilepsy. *Medical Clinics of North America*, 103(2):309–324.

Lodder, S., Putten, V., and Antonius, M. J. (2011). Automated eeg analysis: characterizing the posterior dominant rhythm. *Journal of neuroscience methods*, 200(1):86–93.

Lourenço, C., Tjepkema-Cloostermans, M. C., Teixeira, L. F., and van Putten, M. J. (2020). Deep Learning for Interictal Epileptiform Discharge Detection from Scalp EEG Recordings. *IFMBE Proceedings*, 76:1984–1997.

Rosenow, F., Klein, K. M., and Hamer, H. M. (2015). Non-invasive EEG evaluation in epilepsy diagnosis. *Expert Review of Neurotherapeutics*, 15(4):425–444.

Song, H., Rajan, D., Thiagarajan, J. J., and Spanias, A. (2018). Attend and diagnose: Clinical time series analysis using attention models. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, pages 4091–4098.

Tjepkema-Cloostermans, M. C., de Carvalho, R. C., and van Putten, M. J. (2018). Deep learning for detection of focal epileptiform discharges from scalp EEG recordings. *Clinical Neurophysiology*, 129(10):2191–2196.

Tsiouris, , Pezoulas, V. C., Zervakis, M., Konitsiotis, S., Koutsouris, D. D., and Fotiadis, D. I. (2018). 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.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 2017-December(Nips):5999–6009.

Yan, G., Liang, S., Zhang, Y., and Liu, F. (2019). Fusing Transformer Model with Temporal Features for ECG Heartbeat Classification. *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, pages 898–905.

References

[1] C. Lourenço, M. C. Tjepkema-Cloostermans, L. F. Teixeira, and M. J. van Putten, "Deep Learning for Interictal Epileptiform Discharge Detection from Scalp EEG Recordings," *IFMBE Proceedings*, vol. 76, 2020.

[2] C. da Silva Lourenço, M. C. Tjepkema-Cloostermans, and M. J. van Putten, "Efficient use of clinical EEG data for deep learning in epilepsy," *Clinical Neurophysiology*, vol. 132, no. 6, pp. 1234–1240, 2021.

[3] R. S. Fisher, W. Van Emde Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. Engel, "Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia*, vol. 46, no. 10, pp. 1701–1702, 2005.

[4] E. L. Johnson, "Seizures and Epilepsy," *Medical Clinics of North America*, vol. 103, no. 2, pp. 309–324, 2019.

[5] L. G. Kim, T. L. Johnson, A. G. Marson, and D. W. Chadwick, "Prediction of risk of seizure recurrence after a single seizure and early epilepsy: Further results from the mess trial," *Lancet Neurology*, vol. 5, no. 4, pp. 317–322, 2006.

[6] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. Engel, L. Forsgren, J. A. French, M. Glynn, D. C. Hesdorffer, B. I. Lee, G. W. Mathern, S. L. Moshe, E. Perucca, I. E. Scheffer, T. Tomson, M. Watanabe, and S. Wiebe, "ILAE Official Report: A practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475–482, 2014.

[7] F. Rosenow, K. M. Klein, and H. M. Hamer, "Non-invasive EEG evaluation in epilepsy diagnosis," *Expert Review of Neurotherapeutics*, vol. 15, no. 4, pp. 425–444, 2015.

[8] S. Lodder, V. Putten, and M. J. Antonius, "Automated eeg analysis: characterizing the posterior dominant rhythm," *Journal of neuroscience methods*, vol. 200, no. 1, pp. 86–93, 2011.

[9] S. L. Moshe, E. Perucca, P. Ryvlin, and T. Tomson, "Epilepsy: New advances," *The Lancet*, vol. 385, no. 9971, pp. 884–898, 2015.

[10] E. Beghi, "The Epidemiology of Epilepsy," *Neuroepidemiology*, vol. 54, no. 2, pp. 185–191, 2020.

[11] H. E. Scharfman, "The Neurobiology of Epilepsy," *Current Neurology and Neuroscience Reports*, pp. 348–354, 2007.

[12] S. Weinstein, "Seizures and epilepsy: An overview," *Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Engineering, and Physics*, pp. 65–77, 2016.

[13] R. S. Fisher, J. H. Cross, J. A. French, N. Higurashi, E. Hirsch, F. E. Jansen, L. Lagae, S. L. Moshe, J. Peltola, E. Roulet Perez, I. E. Scheffer, and S. M. Zuberi, "Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology," *Epilepsia*, vol. 58, no. 4, pp. 522–530, 2017.

[14] W. G. Lennox, "The heredity of epilepsy as told by relatives and twins," *Journal of the American Medical Association*, vol. 146, no. 6, pp. 529–536, 1951.

[15] J. Engel, "A proposed diagnostic scheme for people with epileptic seizures and with epilepsy: Report of the ILAE task force on classification and terminology," *Epilepsia*, vol. 42, no. 6, pp. 796–803, 2001.

[16] J. F. Annegers and W. A. Rocca, "Causes of epilepsy: Contributions of the rochester epidemiology project," *Mayo Clinic Proceedings*, vol. 71, no. 6, pp. 570–575, 1996.

[17] K. M. Fiest, K. M. Sauro, S. Wiebe, S. B. Patten, J. Dykeman, T. Pringsheim, and D. L. Lorenzetti, "Prevalence and incidence of epilepsy," *Neurology*, no. 88, pp. 296–303, 2017.

[18] S. J. Smith, "EEG in neurological conditions other than epilepsy: When does it help, what does it add?," *Neurology in Practice*, vol. 76, no. 2, pp. 8–12, 2005.

[19] R. P. Gregory, T. Oates, and R. T. G. Merry, "Electroencephalogram epileptiform abnormalities in candidates for aircrew training," *Electroencephalogr Clin Neurophysiol.*, vol. 86, pp. 75–77, 1993.

[20] L. Zivin and C. A. Marsan, "Incidence and prognostic significance of "epileptiform" activity in the EEG of non-epileptic subjects," *Brain*, vol. 91, pp. 751–778, 12 1968.

[21] N. Kane, J. Acharya, S. Beniczky, L. Caboclo, S. Finnigan, P. W. Kaplan, H. Shibasaki, R. Pressler, and M. J. A. M. V. Putten, "A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings," *Clinical Neurophysiology Practice*, vol. 2, pp. 170–185, 2017.

[22] M. D. Curtis, J. G. R. Jefferys, and M. Avoli, "Interictal epileptiform discharges in partial epilepsy different ied patterns in epileptic patients," *Jasper's Basic Mechanisms of the Epilepsies*, pp. 1–24, 2012.

[23] S. Benbadis, "The differential diagnosis of epilepsy: A critical review," *Epilepsy and Behavior*, vol. 15, no. 1, pp. 15–21, 2009.

[24] R. Beach and R. Reading, "The importance of acknowledging clinical uncertainty in the diagnosis of epilepsy and non-epileptic events," *Archives of Disease in Childhood*, vol. 90, no. 12, pp. 1219–1222, 2005.

[25] R. Roberts, "National Clinical Guideline Centre for Acute Chronic Conditions (Great Britian). Epilepsies: the Diagnosis and Management of the Epilepsies in Adults and Children in Primary and Secondary Care: Pharmacological Update of Clinical Guideline," *London: Royal College of Physicians (UK)*, 2012.

[26] A. T. Berg, S. F. Berkovic, M. J. Brodie, J. Buchhalter, J. H. Cross, W. Van Emde Boas, J. Engel, J. French, T. A. Glauser, G. W. Mathern, S. L. Moshe, D. Nordli, P. Plouin, and I. E. Scheffer, "Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on Classification and Terminology," *Epilepsia*, vol. 51, no. 4, pp. 676–685, 2010.

[27] T. Glauser, E. Ben-Menachem, B. Bourgeois, A. Cnaan, C. Guerreiro, R. Kalviainen, R. Mattson, J. A. French, E. Perucca, and T. Tomson, "Updated ILAE evidence review of antiepileptic drug efficacy and effectiveness as initial monotherapy for epileptic seizures and syndromes," *Epilepsia*, vol. 54, no. 3, pp. 551–563, 2013.

[28] D. Hirtz, A. Berg, D. Bettis, C. Camfield, P. Camfield, P. Crumrine, W. D. Gaillard, S. Schneider, and S. Shinnar, "Practice parameter: Treatment of the child with a first unprovoked seizure: Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Practice Committee of the Child Neurology Society," *Neurology*, vol. 60, no. 2, pp. 166–175, 2003.

[29] V. Novak, "Epilepsy Treatment Introduction," *Clinical Therapeutics*, vol. 39, no. 8, pp. 105–106, 2017.

[30] H. Berger, "Uber das Elektrenkephalogramm des Menschen - Dritte Mitteilung," *Archiv fur Psychiatrie und Nervenkrankheiten*, vol. 94, no. 1, pp. 16–60, 1931.

[31] C. Rizzo, "EEG Signal Acquisition," *Clinical Electroencephalography*, pp. 53–73, 2019.

[32] J. S. Kumar and P. Bhuvaneswari, "Analysis of electroencephalography (EEG) signals and its categorization - A study," *Procedia Engineering*, vol. 38, pp. 2525–2536, 2012.

[33] S. S. Spencer, P. Guimaraes, A. Katz, J. Kim, and D. Spencer, "Morphological Patterns of Seizures Recorded Intracranially," *Epilepsia*, vol. 33, no. 3, pp. 537–545, 1992.

[34] S. S. Spencer, J. Kim, and D. D. Spencer, "Ictal spikes: a marker of specific hippocampal cell loss," *Electroencephalography and Clinical Neurophysiology*, vol. 83, pp. 104–111, 1992.

[35] R. Kennett, "Modern electroencephalography," *Journal of Neurology*, vol. 259, no. 4, pp. 783–789, 2012.

[36] A. Tandle, N. Jog, P. D'cunha, and M. Chheta, "Classification of Artefacts in EEG Signal Recordings and EOG Artefact Removal using EOG Subtraction," *Communications on Applied Electronics*, vol. 4, no. 1, pp. 12–19, 2016.

[37] M. Teplan, "Fundamentals of EEG measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.

[38] G. Klem, H. Luders, H. Jasper, and C. Elger, "The ten-twenty electrode system of the international federation. the international federation of clinical neurophysiology," *Electroencephalography and clinical neurophysiology. Supplement*, vol. 52, pp. 3–6, 1999.

[39] R. C. Burgess, *Filtering of neurophysiologic signals*, vol. 160. Elsevier B.V., 2019.

[40] N. Houmani, F. Vialatte, E. Gallego-Jutgla, G. Dreyfus, V. H. Nguyen-Michel, J. Mariani, and K. Kinugawa, "Diagnosis of Alzheimer's disease with electroencephalography in a differential framework," *PLoS ONE*, vol. 13, no. 3, pp. 1–19, 2018.

[41] N. K. Al-qazzaz, S. Hamid, B. Ali, S. A. Ahmad, K. Chellappan, S. Islam, and J. Escudero, "Role of EEG as biomarker in the early detection and classification of dementia," *Scientific World Journal*, vol. 2014, 2014.

[42] S. J. M. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," *Journal of Neurology, Neurosurgery and Psychiatry*, 2005.

[43] I. G. Campbell, "EEG recording and analysis for sleep research," *Current Protocols in Neuroscience*, no. SUPPL.49, 2009.

[44] Y. Sun, C. Wei, V. Cui, M. Xiu, and A. Wu, "Electroencephalography: Clinical Applications During the Perioperative Period," *Frontiers in Medicine*, vol. 7, no. June, pp. 1–10, 2020.

[45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, vol. 29. MIT Press, 2016. http://www.deeplearningbook.org.

[46] W. J. Zhang, G. Yang, Y. Lin, C. Ji, and M. M. Gupta, "On definition of deep learning," in *2018 World Automation Congress (WAC)*, pp. 1–5, 2018.

[47] Y. Lecun, *Generalization and network design strategies*. Elsevier, 1989.

[48] S. Patel and A. Patel, "Deep Leaning Architectures and its Applications A Survey," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 6, pp. 1177–1183, 2018.

[49] S. C. Chelgani, B. Shahbazi, and E. Hadavandi, "Learning representations by back-propagating errors," *Measurement: Journal of the International Measurement Confederation*, vol. 114, no. 2, pp. 102–108, 2018.

[50] B. Iung, "Bidirectional recurrent neural network," *EMC - Traite de medecine AKOS*, vol. 8, no. 2, pp. 1–4, 2013.

[51] K. Cho, B. Van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1724–1734, 2014.

[52] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in Neural Information Processing Systems*, vol. 4, pp. 3104–3112, 2014.

[53] J. B. Pollack, "Recursive distributed representations," *Artificial Intelligence*, vol. 46, no. 1-2, pp. 77–105, 1990.

[54] R. Socher, C. C. Y. Lin, A. Y. Ng, and C. D. Manning, "Parsing natural scenes and natural language with recursive neural networks," *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, pp. 129–136, 2011.

[55] R. Socher, E. H. Huang, J. Pennington, A. Y. Ng, and C. D. Manning, "Dynamic pooling and unfolding recursive autoencoders for paraphrase detection," *Proceedings of NIPS 2011*, pp. 1–9, 2011.

[56] J. F. Kolen and S. C. Kremer, "Gradient Flow in Recurrent Nets: The Difficulty of Learning Long Term Dependencies," *A Field Guide to Dynamical Recurrent Networks*, 2010.

[57] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, vol. 6, no. 2, pp. 107–116, 1998.

[58] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[59] A. M. Schaefer, S. Udluft, and H. G. Zimmermann, "Learning long-term dependencies with recurrent neural networks," *Neurocomputing*, vol. 71, pp. 2481–2488, 2008.

[60] G. Singhal, "Introduction to lstm units in rnn." https://www.pluralsight.com/guides/introduction-to-lstm-units-in-rnn, 9 2020. (Accessed on 12/20/2020).

[61] K. Smagulova and A. P. James, *Overview of long short-term memory neural networks*, vol. 14. Springer International Publishing, 2020.

[62] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," *31st International Conference on Machine Learning, ICML 2014*, vol. 5, pp. 3771–3779, 2014.

[63] R. Kiros, R. Salakhutdinov, and R. S. Zemel, "Unifying visual-semantic embeddings with multimodal neural language models," 2014.

[64] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, no. 3, pp. 6645–6649, 2013.

[65] G. Singhal, "Lstm versus gru units in rnn." https://www.pluralsight.com/guides/lstm-versus-gru-units-in-rnn, 9 2020. (Accessed on 12/21/2020).

[66] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 2017-December, no. Nips, pp. 5999–6009, 2017.

[67] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *CoRR*, pp. 1–21, 2020.

[68] H. Song, D. Rajan, J. J. Thiagarajan, and A. Spanias, "Attend and diagnose: Clinical time series analysis using attention models," *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, pp. 4091–4098, 2018.

[69] K. Jiang, S. Liang, L. Meng, Y. Zhang, P. Wang, and W. Wang, "A Two-level Attention-based Sequence-to-Sequence Model for Accurate Inter-patient Arrhythmia Detection," *Proceedings - 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020*, no. September 2020, pp. 1029–1033, 2020.

[70] G. Yan, S. Liang, Y. Zhang, and F. Liu, "Fusing Transformer Model with Temporal Features for ECG Heartbeat Classification," *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, pp. 898–905, 2019.

[71] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, 2016.

[72] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer Normalization," 2016.

[73] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," *34th International Conference on Machine Learning, ICML 2017*, vol. 3, pp. 2029–2042, 2017.

[74] S. Ruder, "An overview of gradient descent optimization algorithms," 2017.

[75] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Networks*, vol. 12, no. 1, pp. 145–151, 1999.

[76] V. Bushaev, "Understanding rmsprop — faster neural network learning." `https://towardsdatascience.com/understanding-rmsprop-faster-neural-network-learning-62e116fcf29a`, 9 2018. (Accessed on 01/05/2021).

[77] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, 2015.

[78] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2017.

[79] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," *2014 IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014*, pp. 1015–1018, 2014.

[80] J. Z. Cheng, D. Ni, Y. H. Chou, J. Qin, C. M. Tiu, Y. C. Chang, C. S. Huang, D. Shen, and C. M. Chen, "Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans," *Scientific Reports*, vol. 6, no. March, pp. 1–13, 2016.

[81] HealthIT.gov, "What is an electronic health record (ehr)." `https://www.healthit.gov/faq/what-electronic-health-record-ehr`, 9 2019. (Accessed on 01/17/2021).

[82] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, "Learning to diagnose with LSTM recurrent neural networks," *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, pp. 1–18, 2016.

[83] T. A. Lasko, J. C. Denny, and M. A. Levy, "Computational Phenotype Discovery Using Unsupervised Feature Learning over Noisy, Sparse, and Irregular Clinical Data," *PLoS ONE*, vol. 8, no. 6, 2013.

[84] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting Clinical Events via Recurrent Neural Networks.," *JMLR workshop and conference proceedings*, vol. 56, pp. 301–318, 2016.

[85] X. Ying, "An Overview of Overfitting and its Solutions," *Journal of Physics: Conference Series*, vol. 1168, no. 2, 2019.

[86] G. Raskutti, M. J. Wainwright, and B. Yu, "Early stopping and non-parametric regression: An optimal data-dependent stopping rule," *Journal of Machine Learning Research*, vol. 15, pp. 335–366, 2014.

[87] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[88] B. Scheepers, P. Clough, and C. Pickles, "The misdiagnosis of epilepsy: Findings of a population study," *Seizure*, vol. 7, no. 5, pp. 403–406, 1998.

[89] "10 / 20 System Positioning - Manual," *Gloucester Road*, no. 2410, pp. 77–79, 2012.

[90] M. R. Nuwer, G. Comi, R. Emerson, A. Fuglsang-Frederiksen, M. Guerit, H. Hinrichs, A. Ikeda, F. J. Luccas, and P. Rappelsburger, "IFCN standards for digital recording of clinical EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 106, no. 3, pp. 259–261, 1998.

[91] M. R. Nuwer, "10-10 electrode system for EEG recording," *Clinical Neurophysiology*, vol. 129, no. 5, p. 1103, 2018.

[92] H. J. Landau, "Sampling, Data Transmission, and the Nyquist Rate," *Proceedings of the IEEE*, vol. 55, no. 10, pp. 1701–1706, 1967.

[93] Mingui Sun, M. S. Scher, R. E. Dahl, N. D. Ryan, S. Iyengar, B. Kosanovic, and R. J. Sclabassi, "Analysis of aliasing and quantization problems in eeg data acquisition," in *[1993] Proceedings of the Twelfth Southern Biomedical Engineering Conference*, pp. 280–282, 1993.

[94] N. Michielli, U. R. Acharya, and F. Molinari, "Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals," *Computers in Biology and Medicine*, vol. 106, no. January, pp. 71–81, 2019.

[95] H. S. Liu, T. Zhang, and F. S. Yang, "A multistage, multimethod approach for automatic detection and classification of epileptiform EEG," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 12 I, pp. 1557–1566, 2002.

[96] R. Zuo, J. Wei, X. Li, C. Li, C. Zhao, Z. Ren, Y. Liang, X. Geng, C. Jiang, X. Yang, and X. Zhang, "Automated detection of high-frequency oscillations in epilepsy based on a convolutional neural network," *Frontiers in Computational Neuroscience*, vol. 13, no. February, pp. 1–11, 2019.

[97] D. Ahmedt-Aristizabal, C. Fookes, K. Nguyen, and S. Sridharan, "Deep classification of epileptic signals," *arXiv*, pp. 332–335, 2018.

[98] D. S. Huang, K. Han, and A. Hussain, "Classification of Epileptic EEG Signals with Stacked Sparse Autoencoder based on Deep Learning," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9773, 2016.

[99] R. Hussein, H. Palangi, R. K. Ward, and Z. J. Wang, "Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals," *Clinical Neurophysiology*, vol. 130, no. 1, pp. 25–37, 2019.

[100] N. F. Guler, E. D. Ubeyli, and I. Guler, "Recurrent neural networks employing Lyapunov exponents for EEG signals classification," *Expert Systems with Applications*, vol. 29, no. 3, pp. 506–514, 2005.

[101] A. M. Abdelhameed, H. G. Daoud, and M. Bayoumi, "Deep Convolutional Bidirectional LSTM Recurrent Neural Network for Epileptic Seizure Detection," *2018 16th IEEE International New Circuits and Systems Conference, NEWCAS 2018*, pp. 139–143, 2018.

[102] K. K. Dutta, "Multi-class time series classification of EEG signals with recurrent neural networks," *Proceedings of the 9th International Conference On Cloud Computing, Data Science and Engineering, Confluence 2019*, pp. 337–341, 2019.

[103] I. Ullah, M. Hussain, E. u. H. Qazi, and H. Aboalsamh, "An automated system for epilepsy detection using EEG brain signals based on deep learning approach," *Expert Systems with Applications*, vol. 107, pp. 61–71, 2018.

[104] E. D. Ubeyli, "Analysis of EEG signals by implementing eigenvector methods/recurrent neural networks," *Digital Signal Processing: A Review Journal*, vol. 19, no. 1, pp. 134–143, 2009.

[105] Z. Iscan, Z. Dokur, and T. Demiralp, "Classification of electroencephalogram signals with combined time and frequency features," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10499–10505, 2011.

[106] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13475–13481, 2011.

[107] T. T. Chowdhury, A. Hossain, S. A. Fattah, and C. Shahnaz, "Seizure and Non-Seizure EEG Signals Detection Using 1-D Convolutional Neural Network Architecture of Deep Learning Algorithm," *1st International Conference on Advances in Science, Engineering and Robotics Technology 2019, ICASERT 2019*, vol. 2019, pp. 1–4, 2019.

[108] D. Ahmedt-Aristizabal, T. Fernando, S. Denman, L. Petersson, M. J. Aburn, and C. Fookes, "Neural memory networks for seizure type classification," *arXiv*, 2019.

[109] E. Von Weltin, T. Ahsan, V. Shah, D. Jamshed, M. Golmohammadi, I. Obeid, and J. Picone, "Gated recurrent networks for seizure detection," *arXiv*, 2018.

[110] M. Golmohammadi, S. Ziyabari, V. Shah, D. Silvia Lopez, I. Obeid, and J. Picone, "Deep architectures for automated seizure detection in scalp EEGs," *arXiv*, 2017.

[111] V. Shah, M. Golmohammadi, S. Ziyabari, E. Von Weltin, I. Obeid, and J. Picone, "Optimizing channel selection for seizure detection," *2017 IEEE Signal Processing in Medicine and Biology Symposium, SPMB 2017 - Proceedings*, vol. 2018-January, pp. 1–5, 2017.

[112] T. Liu, N. D. Truong, S. Member, and A. Nikpour, "Epileptic Seizure Classification With Symmetric and Hybrid Bilinear Models," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2844–2851, 2020.

[113] S. Roy, I. Kiral-Kornek, and S. Harrer, "Chrononet: A deep recurrent neural network for abnormal EEG identification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11526 LNAI, pp. 47–56, 2019.

[114] R. R. Bates, M. Sun, M. Scheuer, and R. Sclabassi, "Seizure detection by recurrent backpropagation neural network analysis," *4th International Symposium on Uncertainty Modeling and Analysis, ISUMA 2003*, pp. 312–317, 2003.

[115] L. Bongiorni and A. Balbinot, "Evaluation of recurrent neural networks as epileptic seizure predictor," *Array*, vol. 8, p. 100038, 2020.

[116] Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals," *Computers in Biology and Medicine*, vol. 99, pp. 24–37, 2018.

[117] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," 2016.

[118] L. Vidyaratne, A. Glandon, M. Alam, and K. M. Iftekharuddin, "Deep recurrent neural network for seizure detection," in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 1202–1207, 2016.

[119] X. Tian, Z. Deng, W. Ying, K. S. Choi, D. Wu, B. Qin, J. Wang, H. Shen, and S. Wang, "Deep Multi-View Feature Learning for EEG-Based Epileptic Seizure Detection," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 10, pp. 1962–1972, 2019.

[120] H. Daoud and M. A. Bayoumi, "Efficient Epileptic Seizure Prediction Based on Deep Learning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 5, pp. 804–813, 2019.

[121] M. C. Tjepkema-Cloostermans, R. C. de Carvalho, and M. J. van Putten, "Deep learning for detection of focal epileptiform discharges from scalp EEG recordings," *Clinical Neurophysiology*, vol. 129, no. 10, pp. 2191–2196, 2018.

[122] M. Abou Jaoude, J. Jing, H. Sun, C. Jacobs, K. Pellerin, B. Westover, S. Cash, and A. Lam, "Detection of mesial temporal lobe epileptiform discharges on intracranial electrodes using deep learning," *Physiology & behavior*, vol. 176, no. 10, pp. 139–148, 2017.

[123] J. Jing, H. Sun, J. A. Kim, A. Herlopian, I. Karakis, M. Ng, J. J. Halford, D. Maus, F. Chan, M. Dolatshahi, C. Muniz, C. Chu, V. Sacca, J. Pathmanathan, W. Ge, J. Dauwels, A. Lam, A. J. Cole, S. S. Cash, and M. B. Westover, "Development of Expert-Level Automated Detection of Epileptiform Discharges during Electroencephalogram Interpretation," *JAMA Neurology*, vol. 77, no. 1, pp. 103–108, 2020.

[124] M. B. W. John Thomas, Luca Comoretto, Jing Jin, Justin Dauwels, Sydney S. Cash, "EEG Classification via Convolutional Neural Network-Based Interictal Epileptiform Event Detection," *Physiology & behavior*, vol. 176, no. 3, pp. 139–148, 2019.

[125] K. Fukumori, H. T. Thu Nguyen, N. Yoshida, and T. Tanaka, "Fully Data-driven Convolutional Filters with Deep Learning Models for Epileptic Spike Detection," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2019-May, pp. 2772–2776, 2019.

[126] V. Srinivasan, C. Eswaran, and N. Sriraam, "Artificial neural network based epileptic detection using time-domain and frequency-domain features," *Journal of Medical Systems*, vol. 29, no. 6, pp. 647–660, 2005.

[127] P. R. Davidson, R. D. Jones, and M. T. R. Peiris, "Detecting behavioral microsleeps using eeg and lstm recurrent neural networks," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 5754–5757, 2005.

[128] W. W. Khorasani, "An Adaptive Structure Neural Networks with Application to EEG Automatic Seizure Detection," *Neural Networks*, vol. 9, no. 7, pp. 1223–1240, 1996.

[129] A. V. Oppenheim and R. W. Schafer, "Discrete Time Signal Processing 2nd Edition," 1998.

[130] L. Tan, "Upsampling and downsampling." https://www.eetimes.com/multirate-dsp-part-1-upsampling-and-downsampling/, 04 2008. (Accessed on 01/25/2021).

[131] M. Gayatri, A. Kumar, M. Janghu, M. Kaur, and T. V. Prasad, "Implementation of Epileptic EEG using Recurrent Neural Network," *IJCSNS International Journal of Computer Science and Network Security*, vol. 10, no. 3, p. 290, 2010.

[132] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, "A Regularity Statistic for medical data analysis," *Journal of Clinical Monitoring and Computing*, pp. 335–345, 1991.

[133] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch, "Recurrent neural network based prediction of epileptic seizures in intra and extracranial EEG," *Neurocomputing*, vol. 30, no. 1-4, pp. 201–218, 2000.

[134] R. R. Bates, M. Sun, M. Scheuer, and R. Sclabassi, "Detection of seizure foci by recurrent neural networks," *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, vol. 2, pp. 1377–1379, 2000.

[135] A. Harati, M. Golmohammadi, S. Lopez, I. Obeid, J. Picone, N. Engineering, and D. Consortium, "Improved EEG Event Classification Using Differential Energy," *IEEE Signal Process Med Biol Symp.*, pp. 1–13, 2016.

[136] S. Lopez, A. Gross, S. Yang, M. Golmohammadi, I. Obeid, and J. Picone, "An analysis of two common reference points for EEGs," *2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–5, 2016.

[137] S. P. Kumar, N. Sriraam, and P. G. Benakop, "Automated detection of epileptic seizures using wavelet entropy feature with recurrent neural network classifier," *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, pp. 17–21, 2008.

[138] O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schürmann, and E. Ba ar, "Wavelet entropy: A new tool for analysis of short duration brain electrical signals," *Journal of Neuroscience Methods*, vol. 105, no. 1, pp. 65–75, 2001.

[139] N. Kannathal, M. L. Choo, U. R. Acharya, and P. K. Sadasivan, "Entropies for detection of epilepsy in EEG," *Computer Methods and Programs in Biomedicine*, vol. 80, no. 3, pp. 187–194, 2005.

[140] A. Shoeb, "Chb-mit scalp eeg database v1.0.0." https://physionet.org/content/chbmit/1.0.0/, 6 2010. (Accessed on 01/06/2021).

[141] A. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," *Diss. Massachusetts Institute of Technology*, pp. 157–162, 2009.

[142] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 64, no. 6, p. 8, 2001.

[143] V. Shah, E. von Weltin, S. Lopez, J. R. McHugh, L. Veloso, M. Golmohammadi, I. Obeid, and J. Picone, "The Temple University Hospital seizure detection corpus," *Frontiers in Neuroinformatics*, vol. 12, no. November, pp. 1–6, 2018.

[144] A. Antoniades, L. Spyrou, C. C. Took, and S. Sanei, "Deep Learning for Epileptic intracranial EEG data," *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, 2016.

[145] T. Yang, L. B. Yang, C. W. Wu, and L. O. Chua, "Fuzzy cellular neural networks: theory," *Proceedings of the IEEE International Workshop on Cellular Neural Networks and their Applications*, vol. 35, no. 10, pp. 181–186, 1996.

[146] J. L. Elman, "Finding Structure in Time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.

[147] J. Picone, "Continuous Speech Recognition Using Hidden Markov Models," *IEEE ASSP Magazine*, vol. 7, no. 3, pp. 26–41, 1990.

[148] C. B. Swisher, C. R. White, B. E. Mace, K. E. Dombrowski, A. M. Husain, B. J. Kolls, R. R. Radtke, T. T. Tran, and S. R. Sinha, "Diagnostic Accuracy of Electrographic Seizure Detection by Neurophysiologists and Non-Neurophysiologists in the Adult ICU Using a Panel of Quantitative EEG Trends," *Journal of Clinical Neurophysiology*, vol. 32, no. 4, pp. 324–330, 2015.

[149] T. Miconi, J. Clune, and K. O. Stanley, "Differentiable plasticity: Training plastic neural networks with backpropagation," *35th International Conference on Machine Learning, ICML 2018*, vol. 8, pp. 5728–5739, 2018.

[150] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. February 2017, pp. 321–357, 2002.

[151] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–14, 2015.

[152] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, 2019.

[153] Y. Li, H. Yang, J. Li, D. Chen, and M. Du, "EEG-based intention recognition with deep recurrent-convolution neural network: Performance and channel selection by Grad-CAM," *Neurocomputing*, vol. 415, pp. 225–233, 2020.