UNIVERZA NA PRIMORSKEM
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN
INFORMACIJSKE TEHNOLOGIJE

Zaključna naloga
(Final project paper)

Detekcija epileptičnih napadov s pomočjo topografskih preslikav in globokega učenja

(Epileptic Seizure Detection Using Topographic Maps and Deep Machine Learning)

Ime in priimek: Patrik Kojanec
Študijski program: Bioinformatika
Mentor: doc. dr. Branko Kavšek
Somentor: dr. César A. Teixeira

Koper, julij 2019

Ključna dokumentacijska informacija

Ime in PRIIMEK: Patrik KOJANEC

Naslov zaključne naloge: Detekcija epileptičnih napadov s pomočjo topografskih preslikav in globokega učenja

Kraj: Koper

Leto: 2019

Število listov: 60 Število slik: 22 Število tabel: 10

Število prilog: 2 Število strani prilog: 14 Število referenc: 61

Mentor: doc. dr. Branko Kavšek

Somentor: dr. César A. Teixeira

Ključne besede: detekcija epileptičnih napadov, globoko strojno učenje, topografske mape

Izvleček:
Proti-epileptična zdravila so neučinkovita za kar tretjino epileptičnih bolnikov. Gradnja aparatov, ki samodejno pravočasno zaznajo epileptični napad na podlagi EEG signalov, predstavlja alternativo za izboljšanje kvalitete vsakdanjega življenja takih pacijentov. Sledeča teza je poskus metode automatične detekcije, ki sloni na topogrfskih mapah, izvlečenih iz signalov različnih frekvenčnih pasov, in globokega učenja. Model je bil sestavljen iz petih nevronskih mrež, ena za vsaki frekvenčni pas. Končno oznako za vsak vzorec model izbere glede na večino oznak iz posameznih mrež. Na tak način je model dosegel klasifikacijsko točnost 99.20%, občutljivost 96.48% in specifičnost 99.27% pri klasifikaciji 21 napadov enega pacijenta. Poleg tega, smo naučili nevronske mreže na vzorcih, ki so bili izbrani naključno iz intervalov, kjer se ni pojavljal epileptični napad. V 18 primerih je prišlo do napačne zaznave epileptičnega napada, kjer le-tega ni bilo, a so se te napačne zaznave vedno pojavljale v neposredni bližini samega epileptičnega napada. Take napačne klasifikacije se niso pojavile, ko so bili vzorci vzeti v času zadnjih pet minut pred napadom.

Key words documentation

Name and SURNAME: Patrik KOJANEC

Title of final project paper: Epileptic Seizure Detection Using Tpographic Maps and Deep Learning

Place: Koper

Year: 2019

Number of pages: 60 Number of figures: 22 Number of tables: 10

Number of appendices:2 Number of appendix pages: 14 Number of references: 61

Mentor: Assist. Prof. Branko Kavšek, PhD

Co-Mentor: Prof. César A. Teixeira, PhD

Keywords: epileptic seizure detection, deep learning, topographic maps

Abstract: One third of all epileptic patients is resistant to medical treatment. The construction of machines, that would detect an imminent epileptic attack based on EEG signals, represents an efficient alternative, that would help to increase their quality of life. In this, thesis we tried to implement an automatic detection method, based on the signal of different frequency sub-bands, using topographic maps and deep learning techniques. We constructed an ensemble of five convolutional neural networks, to classify samples of each sub-band and choose the final decision by a majority voting. The ensemble obtained 99.20% accuracy, 96.48% sensitivity and 99.27% specificity when detecting seizures of one patient. Moreover, in this thesis we identified on 18 of 21 seizures some false positive classifications really close to the seizure onset, when the networks were trained with samples taken randomly from the inter-ictal intervals. Such misclassifications did not occur when training was performed with samples taken within five minutes of the seizure onset.

Acknowledgement

I would like to thank my mentor and my co-mentor for the help and availability given during the progress of the project. I would like to thank the institutions that allowed me to do my internship: the University of Coimbra, the Department of Informatic Engeneering and the University of Primorska. I would also like to thank the personnel of the Clinical Informatics Laboratory of CISUC for the extensive help and consultations throughout this project. Finally, a special thank to my friends and family, always supporting me during this experience.

Contents

List of Tables

List of Figures

Appendices

List of Abbreviations

i.e.	that is
e.g.	for example
AEDs	Anti-Epileptic Drugs
DRE	Drug-Resistant Epilepsy
EEG	Electroencephalogram
ILAE	International League Against Epilepsy
ICA	Idependent Component Analysis
fMRI	functional Magnetic Resonance Imaging
PET	Positron Emission Tomography
CT	Computerized Tomography
GA	Genetic Algorithm
PSO	Particle Swarm Optimization
BCI	Brain Computer Interface
SVM	Support Vector Machine
ML	Machine Learning
RBF	Radial Basis Function
HMM	Hidden Markov Model
NB	Naïve Bayes
k-NN	k-Nearest Neighbour
DWT	Discrete Wavelet Transform
FTT	Fast Fourier Transform
ANN	Artificial Neural Network
RBFN	Radial Basis Function Network
MLP	Multi-layer Perceptron
DL	Deep Learning
CNN	Convolutional Neural Network
LSTM	Long-Short Termed Memory Network
bi-LSTM	bidirectional Long-Short Termed Memory Network

1 Introduction

Epilepsy is a neurological disorder characterized by sudden seizure attacks, that may cause in patients loss of consciousness and motor control. It is esteemed that epilepsy affects about 50 million people world-wide and represents up to 1% of the global burden of disease[42]. Although in the last decades many anti-epileptic drugs (AEDs) have been introduced [27], to more than 30% of the patients these treatments are ineffective. Therefore, their daily life activities are very restricted because of the unpredictability of the attacks. The development of different approaches, that could timely inform drug-resistant epilepsy (DRE) patients of an imminent epileptic attack, could be an efficient way to increase their quality of life.

The most used tool to monitor brain's electrical activity is the electroencephalogram (EEG). However, due to the complexity of the EEG signals, visual detection of epileptic seizures from the signal often results misinterpreted or mistaken. Therefore, in the last decades much research has been oriented towards finding automated detection procedures, that would efficiently analyze large chunks of signals, timely give out warnings and help the medical staff to deliver treatment on time[47].

Patients undergo long-term monitoring, to identify the zone of the ictal (seizure) onset and to establish its dynamics. This information may be used to determine if a surgery is needed or for the development of portable devices, that monitor the brain activity and warn patients of imminent epileptic attacks. Generally, there are two kinds of transitions from non-ictal to ictal state[36]. The first consists of gradual changes in brain's activity (or a cascade of changes) and are normally found in focal epilepsy. In this kind of epilepsy, seizures might be predicted by identifying a specific past time interval, which is usually referred as the pre-ictal time. However, such time interval is not well defined and it varies from patient to patient, raising objections about its existence. The other type of transition is more sudden and generally undetectable and unpredictable.

Since the first studies of epilepsy seizures with EEG, it is known that an epileptic attack has a detectable electrical discharge in the brain (EEG onset), prior to the manifestation of convulsions, lost of consciousness and others symptoms (clinical onset)[44]. The time window between these events usually ranges between 0 to 30 seconds, sometimes reaching over 1 minute. Therefore, being able to detect early enough the EEG onset

of the seizure could give enough time to the patient to get the treatment or at least to reach a safe environment.

Based on these motivations, this thesis will represent a model of epilepsy seizure detection based on topographic maps generated from EEG signals and deep machine learning classifying techniques. It will be structured as follows: section two will present some background information needed for a better understanding of the topic, section three will introduce the most common state-of-the-art techniques, section four will describe the data, parameters and the methodologies used in the experiment, section five will present the results, which will be followed by a discussion. A the end a conclusion will be taken and some suggestions for future work will be given.

2 Background information

This chapter will present some background knowledge, needed for a better understanding of the topic. The first section will bring up information about brain structure and it will define epilepsy. It will be followed by the section presenting the electroencephalograph and explain the meaning of its recordings.

2.1 The Brain

The brain is one of the most complex biological systems to be known. In animals, it is the main part of the central nervous system, that elaborates and interprets all the information sensed from the environment. The brain can be divided in different brain regions, each of them responsible for some tasks.
These regions are nonetheless networks of interconnected cells - neurons - that communicate between them using electrical signals. Each neuron may receive a signal from more than a hundred other neurons and propagate it to other hundreds. Thus, they form a vast network, that allows animals to perform cognitive and memory functions, as well as controlling movements[59]. Nowadays is quite well understood how single neurons work, however, the complex ways they communicate and interact, allowing animals to perform different tasks, is not so well comprehended.
Clearly, we can observe the brain function at different spatial scales: micro-scopic, meso-scopic and macro-scopic. Micro-scopic refers to individual neurons. Here the focus is on the proteins that compose their membrane, e.i. ion channels, ion pumps, receptors, on the neurotransmitters and neuromodulators, on the different pathways that are generated inside the cell, etc. Meso-scopic refers to elementary structures of few neurons, that joint together can perform more complex tasks. Such example are the cortical columns in the cortex or projections of neurons, that connect different regions. Macro-scopic refers to brain structures and even to whole brain regions, which can perform even the hardest functions. Here can be mentioned the hippocampus, thalamus, or the occipital lobe, known to be the region where visual input is initially processed. It is important to know, that for EEG monitoring, only the macro-scopic aspect is taken in account, since the recording of every electrode is influenced by all

the charges present in that region.

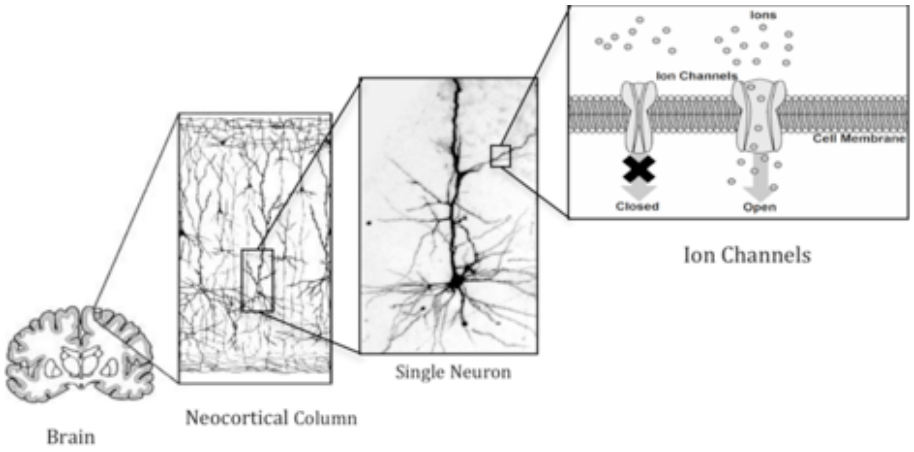


Figure 1: Picture representing the brain at all levels: micro, meso and macro-scopic. Source: [38]

2.2 Epilepsy

As previously mentioned, epilepsy is defined as a disorder of the brain characterized by enduring predisposition to generate epileptic seizures[15]. Seizures are nonetheless an abnormal activity of neuronal networks, that uncontrollably discharge and propagate the signal towards other regions of the brain. Neurons become hyper-active and synchronous and fire action potentials at a very high rate. The result of this event may be loss of consciousness and/or loss of motor control in the patient. This abnormal activity can continue for different minutes, until the oxygen supply ends or the chemical imbalances in the brain force its termination [59].
The dynamics of seizures can be very different between them, therefore the International League Against Epilepsy (ILAE) proposed a general classification method[32]. The classification of different kinds of epilepsy is operational (practical), thus is based on the manifestation of different symptoms during the attack, rather than on a scientific method. There are four main categories, regarding the onset of the seizure:

- focal epilepsy, whose seizures start from a certain brain region and are restricted to one hemisphere;
- general epilepsy, whose seizure's onset is identified in both hemispheres of the brain;
- combined focal and general epilepsy, that manifest both focal and general seizures;

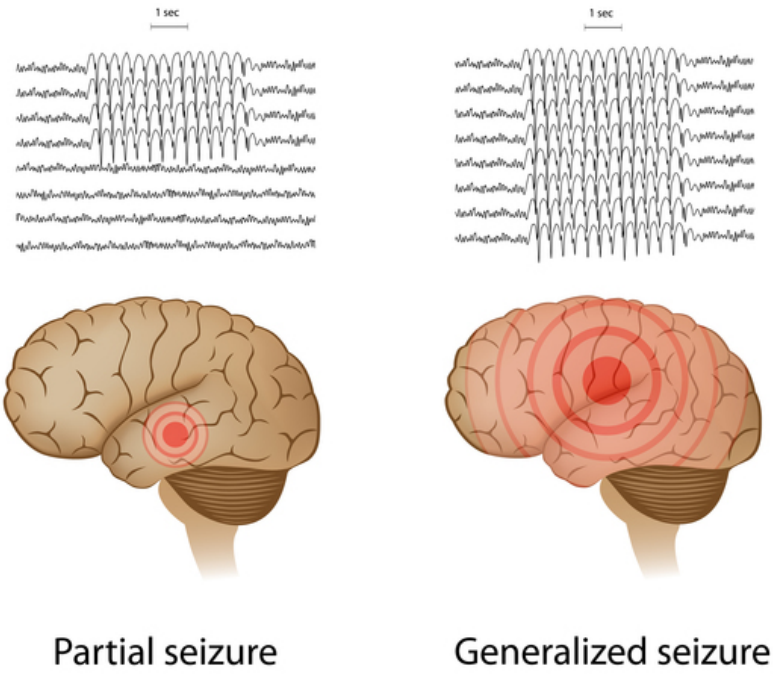


Figure 2: Image showing the differences in the signal between a focal and a general seizure. Source [21]

- unknown epilepsy, that is characterized by seizures, which don't have a defined locus of onset and can be later placed in one of the previous two categories.

Individual seizures can be furthermore organized regarding manifestations of motor and consciousness disabilities. Focal seizures, for instance, are divided in focal onset aware and focal onset impaired awareness seizures, based on the fact that the patient is aware of the surrounding environment or not. Moreover, a focal seizure can be characterized by motor-impairment manifestations, such as body stiffness and jerking movements, that will define the seizure as focal motor onset seizures. However, these two criteria of classification may not be always possible, since not all focal seizures manifest those events. Motor-nonmotor classification is also used for general and unknown onset seizures.

2.3 Electroencephalography

The most used technique to monitor brain electrical activity is called electroencephalography. It consists of recording electrical impulses from different brain structures using metal electrodes and conductive media. The recording is called electroencephalogram (EEG). Generally, there are two main types of EEG techniques: scalp EEG and in-

tracranial EEG[55]. Intracranial EEG consist of probes, that are implanted into the brain, therefore surgery is needed. It is a very invasive technique, but it gives high quality signals. However, in research it is not well accepted, because it lacks of "control" conditions in experiments. For example, intracranial EEG data is retrieved only from epileptic patients and not from healthy patients, because no healthy person will undergo the surgery needed to implant the device. Therefore, there will not be a set of data from a healthy population to compare the results with[59].
Scalp EEG, on the other hand, is a non-invasive technique, that can be performed repeatedly, with no side effects in children, adults and elderly people. Scalp EEG, as the word itself suggests, are recordings taken from the scalp, hence its recordings take in consideration mostly the activity of the cortex. Usually some gel is applied to the scalp's surface in order to receive a better signal. The signal recorded is an artefact of the whole brain activity near the electrode and is a product of both the charges, positive and negative. Therefore, the interpretation of the data can be ambiguous or unclear. E.g., there can be high activity in the recorded region, but since both charges neutralize each other, the recording shows otherwise. Thus, a greater amplitude of the signal is associated with high synchrony of the region. To record activity, other techniques are preferred, such as fMRI and PET[55][59].

10-20 Positioning System

In order to obtain useful and systematic information from the EEG recording, an appropriate positioning of the electrodes should be considered. Since the beginning of EEG research it was important to determine an international electrode positioning model, that would allow researchers to compare their studies. Such model should consider the following criteria: (1) locations of electrodes should be based on standard skull landmarks, (2) should provide a coverage of the whole head, (3) the nomenclature of the electrode should derive from brain regions other than only numbers, in order to avoid confusions and facilitate communication between interested people, (4) anatomical studies should be carried out to confirm the electrode positioning[33].
At the first International EEG Congress, held in London in 1947, Dr. Jasper was appointed to study this problem and to make recommendations on a general model for electrode positioning. In 1958, he proposed the 10-20 model[28], that still nowadays is recognized as general model for EEG probe positioning. The electrodes are named by the region where they are located:

- "Fp" - Frontal pole position of electrodes
- "F" - Frontal line of electrodes

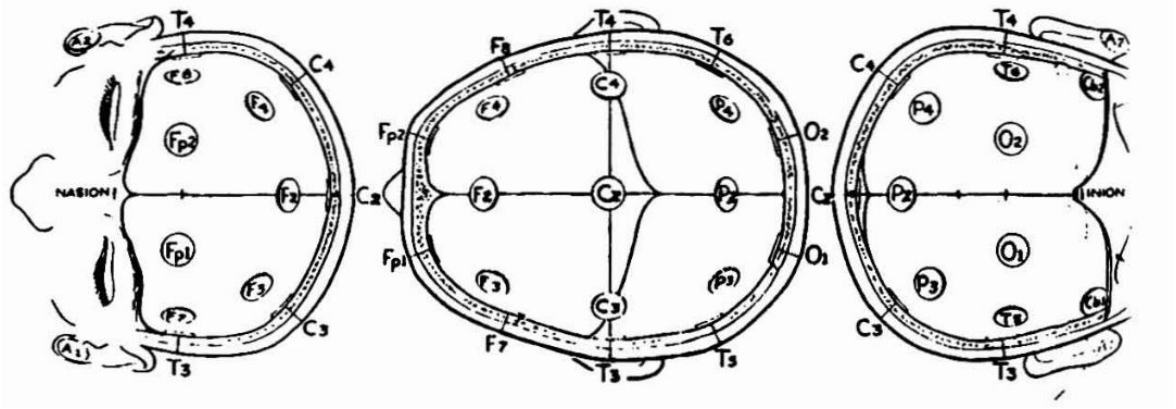


Figure 3: Illustration of the 10-20 model. Source: [56], page 18

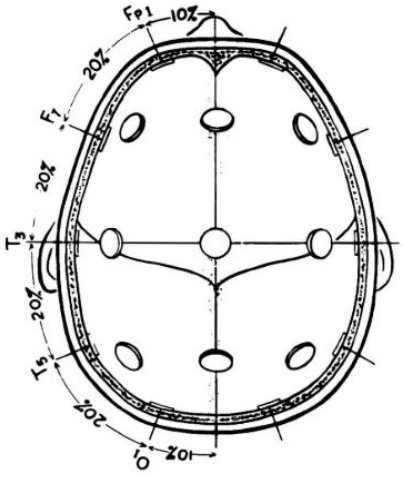


Figure 4: Spacing of the electrodes in the 10-20 model. Source: [56], page 16

.

- "C" - Central line of electrodes
- "P" - Parietal line of electrodes
- "O" - Occipital line of electrodes
- "T" - Temporal line of electrodes
- "Ph" - Pharyngeal position of electrodes
- "A" - Auricular position of electrodes
- "Cb" - Cerebellar position of electrodes

Odd numbers are located on the left hemisphere, while even numbers are located on the right. The electrodes placed among the nasion-inion axis are labeled with a "z", that denotes the number zero.

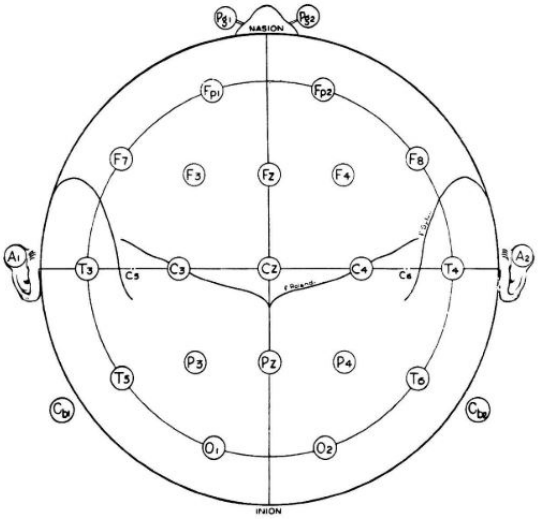


Figure 5: Schematic representation of the 10-20 model. FSource: [56], page 19

In the model's name, "10" denotes the distance in percentages, regarding the axis length in consideration, from standard landmarks, e.i. nasion, inion, left and right preauricular points, while "20" denote the distance in percentages, between electrodes, regarding the axis length in consideration. For example, as showed in figure 4, on the anterio-posterior (A-P) axis, the first electrode "Fp" is located at a distance of 10% of the total A-P length. The following electrode "F" is then located at 20% from the electrode "Fp". Following the same criteria for the other electrodes "C", "P" and "O", the "O" electrode will be located at 10% from the inion.

Anatomical studies were performed to validate this model. Homan et al. in 1987 [24] performed a non-invasive method, using computerized tomography (CT) and aluminum electrode markers to determine correlations between anatomical structures and electrode positioning. For the majority of subjects, even in the presence of skull asymmetry, as long as cerebral structures were not distorted, the 10-20 system provides scalp locations which correlate well with the expected cerebral structures. Other studies were performed using functional magnetic resonance imaging [fMRI] to define the best-fitting electrode position regarding the cortical anatomy[58], confirming that the electrodes in the 10-20 system cover the area of the cortical structures of major interest.

3 State of the art

With the raising of computational power and the advance of technology many studies were oriented towards finding new strategies for seizure automated detection using EEG data [47]. Generally, the methodology relies on machine learning algorithms, that learn to classify samples on behalf of some features extracted from the signal. A standard procedure for automated seizure detection follows a general framework with the following steps [7]:

- signal acquisition,
- signal processing,
- feature extraction,
- feature selection,
- classification,
- regularization,
- performance evaluation.

This chapter will present different techniques to approach the problem, that were already performed. The chapter will also be divided in sections, according the previously mentioned general pipeline, with the intention to present the topic in a more comprehensible way.

3.1 Signal acquisition

Early studies were mostly performed on data acquired from local health centers. Local databases contained a small number of patients and most of them with a small number of seizures. With such a limited amount of data, the specificity of prediction and detection algorithms could not be determined and such algorithms were prone to overfitting. The results of the studies often resulted biased.
To avoid this shortcomings nowadays studies rely on big amounts of data. The three

most used are the databases from epilepsy centers of Bonn and Freiburg, Germany, from the Children's Hospital in Boston, USA, and the EPILEPSIAE database, that is available since 2012 [50]. EPILEPSIAE was founded by the European Union and was started with six partners, from hospitals, universities and industry in France, Germany, Italy and Portugal. Compared to the previous three databases, which range from 5 to 23 total number of patients, 40 minutes to 142 hours of recording duration and from 59 to 189 total number of seizures, EPILEPSIAE contains recordings of 275 patients, averaging 165 hours of recording per patient and 9.68 seizures per patient. Besides EEG recordings, it also contains ECG recordings and extensive metadata on technical and clinical details of the recordings[32].

3.2 Signal Pre-Processing

Raw EEG data is a very noisy and non-stationary, thus has to undergo some pre-processing procedures, before being further analyzed. This section will present some techniques and approaches, that other researchers have already performed.
The analysis of a signal usually requires its decomposition in some chosen domain. The three main domains are: time, frequency and time-frequency. The decision about which will prevail in the study, should be taken carefully the beginning, allowing so the development of an efficient framework, avoiding losses of information and resolution.

3.2.1 Sliding Time Windows

The first step in pre-processing is defining the length of the sliding time window. A whole EEG recording is divided into small consecutive intervals, generally of equal length, for two main reasons: the first is that EEG is a non-stationary signal. Many mathematical tools for analysis assume the stationarity of the signal. One way to enforce an "artificial" stationarity is by segmenting the signal and making the analysis of the segments globally valid [59]. The other reason is more pragmatic; the ultimate goal of seizure detection and prediction is to timely give out warnings about an imminent seizure attack. This would be hard or impossible to perform by analyzing a whole recording, because algorithms would learn to identify an attack regarding the changes in a whole recording, rather than detecting significant consecutive changes of the signal that lead to a seizure. Therefore, this step is of fundamental importance and its setting can strongly influence future steps.
Sliding windows in different studies were set differently, regarding their length or their

overlapping characteristic. Overlapping, as the word suggests, means that the next window contains a segment from the previous window, therefore overlaps it. This is an effective technique, that can increase the number of windows, without reducing their length. On the other hand, selecting the length of a sliding window is a delicate step of pre-processing and an optimal window size has not been defined yet. Direito et al. in 2012 [12], while performing a study on seizure prediction, opted for five second sliding windows with no overlap. They stated, that it represents a good compromise between long windows, that lose temporary resolution and the assumption of stationarity, and short windows, that lose frequency resolution. Alotaiby et al. in 2014 [2], wrote a survey on seizure detection and prediction algorithms, where they reviewed 44 studies. Seizure detection studies (25) opted for sliding windows of a length ranging from 70 milliseconds to whole recordings, with most of them fitting in the range of 1 to 60 seconds. Similarly, prediction studies' (19) sliding windows sized between 2 seconds and whole recordings, with the majority in the range from 2 seconds to 10 minutes. Detection and prediction scores seemed to not be affected by the different sliding windows' length, however it has to be taken in consideration that authors have also used different classification methods and algorithms. Some different sliding window length opiton are shown in table 1, taken from the survey by Alotaiby et al., 2014. EMD stays for empirical mode decomposition, which transforms the signal to a group of intrinsic mode functions. SVD, which derives from singular-value decomposition, is an other kind of signal decomposition. The table also shows from which databases the data was taken.

In this thesis, since it aims to early detect seizures, sliding windows should be of such length that could raise an alarm in a range of 3 to 6 seconds, so that the patient could receive the warning before the clinical onset of the seizure. Therefore, an appropriate length could be anywhere between 1 and 5 seconds. However, too small frames will cut off smaller frequencies. Hence, using five seconds sliding window may result as an optimal option, since it will maintain a good resolution, even for frequencies under 1 Hz.

Overlapping could also be beneficial, since the sliding windows are relatively big when compared to the difference between clinical and EEG onset. Thus, using an overlap will increase the probability of early-detecting a seizure and it would augment the data, which could be beneficial for deep learning algorithms [10]. However, the final decisions should be based on the chosen methods and the overall framework.

Method for Detection	Domain	Database	Frame length
Tafreshi et al.	EMD	Five patients of Freiburg database (scalp and iEEG)	4 to 6 s
Orosco et al.	EMD	90 EEG segments acquired for nine patients	1 h
Guarnizo and Delgado	EMD	Five groups with 100 single-channel registers sampled at 173.61 Hz	Whole signal
Alam and Bhuiyan	EMD	Bonn University database (scalp EEG)	23.6 s
Bajaj and Pachori	EMD	Freiburg database for 21 patients (scalp and iEEG)	15 s
Rana et al.	Frequency	EEG and ECoG time series- BW: 0.5 to 100 Hz	10 to 60 s
Khamis et al.	Frequency	12 patients with six data records (R1 to R6)	32 s
Acharya et al.	Frequency	Self-recorded data	23.6 s
Vanrumste et al.	SVD	Simulated EEG signals	10 s
Shahid et al.	SVD	Recordings of four pediatric patients with 20 seizures	1 s
Runarsson and Sigurdsson	Time	Self-recorded data	Variable
Zhou et al.	Wavelet	Freiburg database of 21 patients	4 s w/o overlap.
Yoo et al.	Time	MIT database (scalp EEG)	2 s
Dalton et al.	Time	Dataset of 21 seizures	12 to 25 s
Liu et al.	Wavelet	Dataset of 509 h from 21 epileptic patients	4 s
Panda et al.	Wavelet	500 epochs of EEG data from 5 different brain activities	0.5 s
Khan et al.	Wavelet	Self-recorded data for five patients	25 s
Wang et al.	Wavelet	Changhai Hospital database (scalp EEG)	70 to 200 ms
Zainuddin et al.	Wavelet	University of Bonn database (scalp EEG)	23.6 s
Niknazar et al.	Wavelet	100 single-channel segments from Bonn University Database	23.6 s
Daou and Labeau	Wavelet	9 patients from MIT database- and 9 patients from MNI database	1 s
Mehta et al.	Wavelet	Self-recorded	10 s
Shoaib et al.	Wavelet	MIT database (scalp EEG)	2 s
Zandi et al.	Wavelet	EEG recordings from 14 patients (75.8 h with 63 seizures)	10 to 40 s

Table 1: Different authors used different frame lengths. Source [2].

3.2.2 Filtering, Normalizing and Noise Reduction

This step consists in removing artefacts that my alter further analysis results and normalizing recordings to make them comparable with others. EEG recordings can get altered by any electric field. Therefore, artefacts in EEG records can be produced by various sources, such as muscular contractions, eye blinks, or by the electrical field of the power line [59]. Next, there are presented some methods that might be used to further pre-process EEG signals.

Filtering frequency bands is a highly used step in EEG signal pre-processing. A 50Hz notch filter is commonly used to remove distortions caused by the power line. In addition, a band-pass filter is also very common, to filter frequencies outside of the 0.1 - 120 Hz range. The upper bound 120Hz is set regarding the Nyquist criterion, which states, that sampling rate of F_s Hz can only correctly represent signals that contain frequencies up to $F_s/2$. Since most machines have a sampling rate of 256Hz or higher, 120Hz presents a good upper bound candidate.

Normalizing data of a seizure is necessary to be able to compare it among data of other patients or among other seizures. Typical procedures are detrending the signal, which is obtained by removing its mean, and scaling to unit variance, which is obtained by dividing a detrended signal by its standard deviation [59].

The last step in signal pre-processing is dealing with noisy artefacts. Artefacts can be managed in four different ways: they can be ignored, if it is assumed that only minimally affect further feature extraction. Highly contaminated windows can be also rejected and not taken in account in further analysis. However, with this method a significant amount of data may be lost and this may affect future results. To reduce the data loss, artefacts can be removed by the EEG with separation methods, i.e. independent component analysis (ICA) or wavelet filtering. The last, more elegant solution, is to train classifying systems to properly deal with artefacts in the signal [59].

3.3 Feature extraction and selection

Raw EEG data itself contains a variety of information. Usually, recordings off EEG data are from the computational point of view unsuitable to use in machine learning algorithms. Therefore, raw EEG data normally undergoes processes of dimensionality reduction, called feature extraction and feature selection. First of all, a set of features of interest is extracted from pre-processed data, often with a transformation of the original data. Although some level of dimensionality reduction may already happen at this point, the features may still contain redundant or irrelevant information. A good

feature selection removes such features and, hence, can reduce learning time, improve learning accuracy and simplify learning results [9].

EEG extracted features can be univariate and multivariate, regarding the number of channels taken in consideration, and linear or non-linear. Linear features derive from frequency and amplitude information, while non-linear features use non-linear measures to describe the current state of the system or its dynamics [4]. The computational complexity of features increases as we advance from univariate linear to univariate non-linear and from multivariate linear to multivariate non-linear. No specific type of features has been proposed as the best, therefore the selection of features floats between the criteria of using the lower level of computational complexity and the major quantity of information, so it will achieve good classification results [7].

From previous studies it is known, that a combination of univariate and multivariate

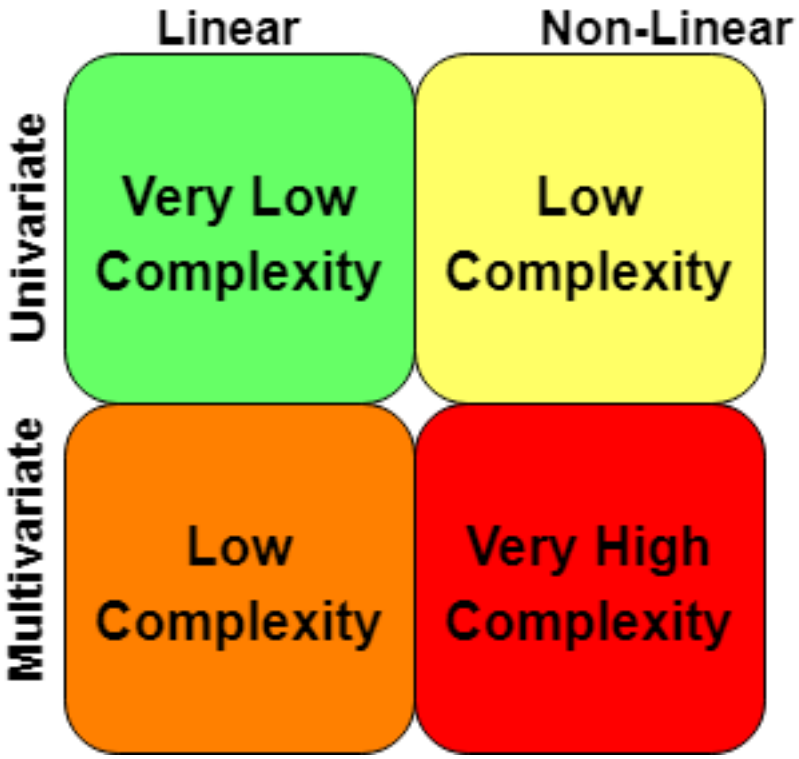


Figure 6: Representation of feature's complexity

features results in better predictive systems [40]. McSharry et al. 2003, revealed that non-linear measures did not perform better than the linear ones, even though EEG signals are of non-linear nature. In other studies with real-time algorithms [43] the computational complexity of non linear features was not suitable for such a dynamic model. Consequently it was advised that the use of non-linear features should be taken in consideration only when they outperform the linear ones.

In this thesis, since the study is oriented towards early detection, the computational

time is very limited. Perhaps, a selection of linear features and a restricted use of non-linear features could be more appropriate. However, the loss of information in such case could affect classification results. Thus, we will rely on the power of deep learning algorithms, that may eventually capture the same information that is held in the more sophisticated features.

3.3.1 Feature Selection Methods

The variety of feature selection methods and their combinations offer researchers and engineers many handholds to increase performances of their work, by reducing the computational overhead and performing classifying tasks better. They can be categorized according to different principles [9]:

- Concerning the labelling of the training data, selection methods can be supervised, unsupervised or partially supervised.
- Concerning the interaction with the model, selection methods can be divided into filter, wrapper or embedded methods.
- Concerning the search strategies, feature selection methods can be grouped into forward increase, backward deletion, and hybrid methods.
- Concerning the output, feature selection methods can be divided into subset selection and feature rank methods.

In the following sections it will present some feature selection methods, which will be divided into supervised and unsupervised. In these categories will be presented both filter and wrapper techniques of feature selection, with some state-of-art examples. A third section, named "Embedded Methods", will focus exclusively on embedded methods, since they represent a more recent approach.

Supervised Methods

Under supervised methods we identify all the methods for feature selection, that include labelled data for the training session. This methods select the most relevant features that will help classify future instances into two or more classes. Therefore, the correlation or relevance between the feature and the class label is the fundamental principle in this approach. Supervised filter models can be grouped in (1) relevance and redundancy based analysis, which could be also translated into two optimization

problems: max(Relevance) and min(Redundancy), or in (2) relevance and diversity analysis, which could be similarly translated into max(Relevance) and max(Diversity) optimization problems [9].

A different approach can be provided by supervised wrapper models, that often produce feature selection results simultaneously with the learning process, since the learning method is included in feature selection. Although wrapper models result in smaller feature subset size and better classification accuracy compared to filter models, they run with high time complexity and often display low learning scores. Common wrapper methods include genetic algorithm (GA)[1] and particle swarm optimization (PSO)[2] search strategies. Often wrapping methods are combined with filter methods, thus giving birth to hybrid methods. They consist in applying filters to remove noisy and irrelevant features and using wrapper model to find the optimal feature subset[9].

Unsupervised Methods

Unsupervised models can be as well divided into filter and wrapper methods, regarding the fact if they are based on cluster algorithms. Filter methods use an algorithm - selector - prior the learning algorithm, that will down-sample the feature set. It takes in account the statistical performance of the data as evaluation metric and decide which features better represent the dataset. Such approach is suitable for large data-sets, especially because of its low computational complexity, however, there is no feedback received from the training section. Hence, the selected features, which are believed to perform well individually, they may not be so efficient when used together in the training algorithm [52].

In the unsupervised wrapper models, features subsets are selected regarding to best clustering score. Hence, each feature subset has to undergo the process of evaluation through a pre-selected classifier, which requires time and high computational complexity, especially for large scale data. However, comparing to filter methods, there

[1]GA is a population based stochastic optimization algorithm, inspired by the process of natural selection, is an optimization algorithm, that relies on three genetic operators: mutations, crossover and selection. The first has the role to maintain a diversity in the new generation of possible solutions, the second to combine characteristics of the previous generation to create a new offspring and the third to select the most-fitting samples. Every new generation has a handful of new possible solutions, that may eventually find the optimal solution to the problem [39]

[2]PSO is also a population based stochastic optimization algorithm. Compared to the GA is more simple, in the way that not include genetic operators in every new generation. A population (i.e. swarm) of possible solutions, named particles, follows the current optimal solution in the "search-space". Each time a better position is discovered, it starts to guide the swarm. The process is repeated many times, hoping that an optimal solution will be found [31]

are higher chances to identify a well performing feature subset, since the selection and evaluation algorithms exchanges feedback. If the unsupervised feature selection is conducted in a single cluster, it is called a local wrapper method, otherwise, if it is conducted in all clusters, it is called global wrapper method [9].

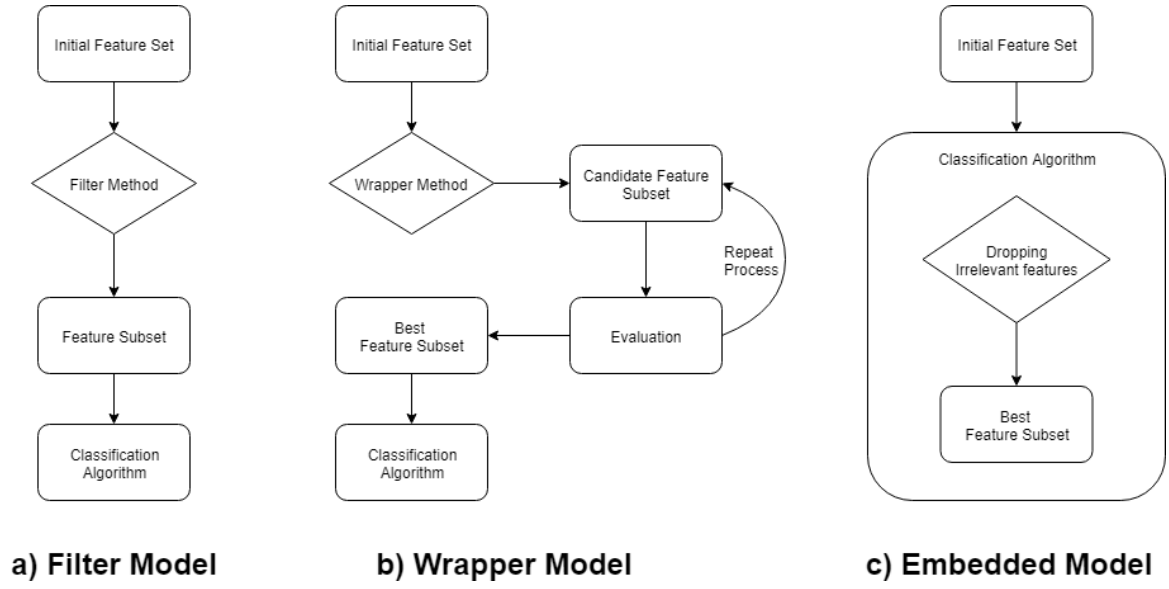


Figure 7: Schematic representation of feature selection models.

Embedded methods

Embedded methods differ from other feature selection methods in the way the interact with the learning algorithm [34]. The leading idea behind them is to merge the good characteristics of both filter and wrapper models: they aim to interact with the classifier, as wrapper methods, and to maintain a low time complexity, as filter methods.
While wrapper methods use learning algorithms to measure the quality of a feature subset, embedded methods select features during the learning phase and automatically output the most suitable ones after the learning procedure has ended. In other words, they do not separate feature selection from learning, but they incorporate (embed) it into the learning process. A known embedded method is based on least absolute shrinkage and selection operator (LASSO).
Compared to wrapper methods they may result more efficient in some aspects: there is no need to split the training data into training and validation sets and there is no need to restart training for each feature subset [18]. Embedded methods can be expected to be faster than wrapper methods, if the function that measures the quality of a scaling factor is evaluated faster than a cross-validation error estimation procedure. However, they are prone to over-fitting problems with limited training data available. In this

conditions filter methods would most likely perform better [9].

3.4 Topographic maps

An increased attention to spatial representations of brain electrical signals was first noticeable in the 60s-70s of the previous century. In 1969, Estrin and Uzgalis and Harris et al. already described some complex topographic systems, that were based on computational methods, while by 1987 at least nine systems were commercially available in the US [30]. Such a spread of the use of topographic maps was mostly due to the fact, that they simply represent various measures of brain electrical activity and so allow even personnel with limited amount of training to operate with such systems.

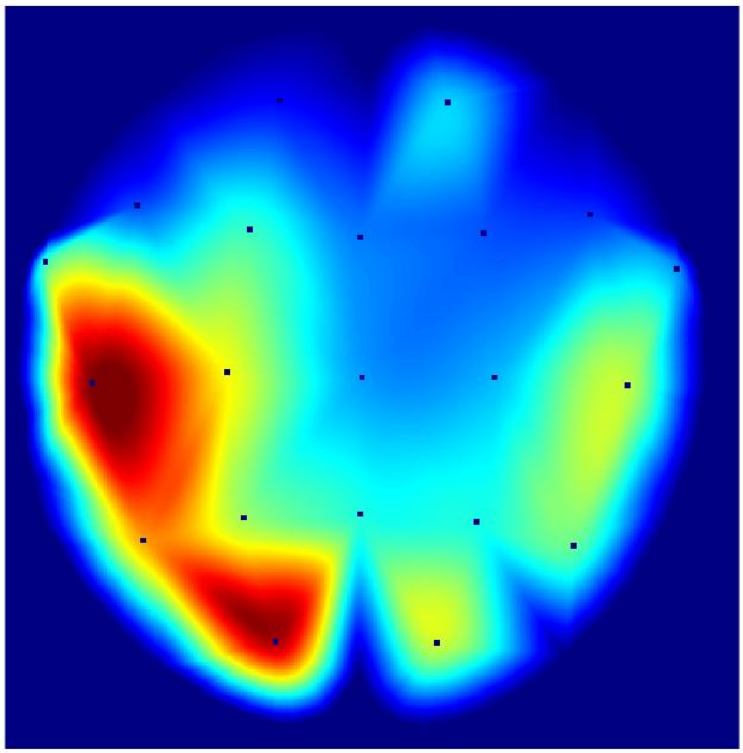


Figure 8: Example of a topographic map used in this thesis.

In its native formulation, a topographic map of a brain represents the values of the electric fields of the scalp, encoded in a colour scale. Commonly, red and yellow indicate higher synchronous activity or values, while blue indicates lower. A 2-D topographic map is constructed in two steps: first, the values retrieved from each electrode are

spatially distributed on a two dimensional grid, according to the original distribution on the scalp. Second, an interpolation function is performed, which assigns values to all the cells in between them. Typical interpolation functions are linear, cubic and spline functions. An example of 2-D topographic map is shown in figure 8.
Recently, with the arise of more powerful classifiers, the use of 3-D topographic maps, in the fields related to Brain Computer Interface(BCI) systems, gained much interest. A 3-D topographic map does not necessary denote the 3-D spatial distribution of the electrodes. Compared to conventional 2-D topographic maps, they have one more dimension, which can represent information from other domains, such as time or frequency domains, rather than spatial. Zhang et al. in 2019 [62], in an experiment on cross-task mental workload assessment, where they tried to learn new spatial-spectral-temporal EEG features with a recurrent 3-D convolutional neural network, they used 3-D topographic maps constructed as a stack of 40 2-D topographic maps. The topographic maps at each layer contained information at a different frequency of the original signal, raging from 1-40Hz. A representation of their method, taken from the original paper, is shown in figure 9.

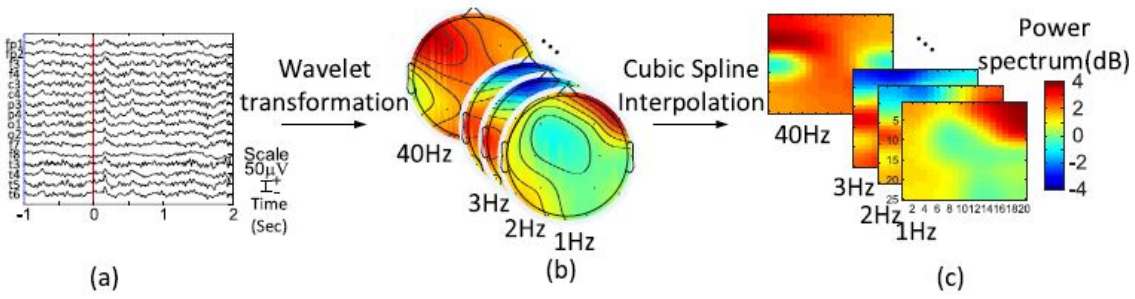


Figure 9: Example of 3-D topographic maps. Source: [62].

Since topographic maps are a graphic visualisation of the data, useful information can be retrieved from them using deep learning algorithms. In this thesis topographic maps will be fed to a convolutional neural network, that will be trained to distinguish seizure topographic maps from non-seizure topographic maps.

3.5 Classification

In machine learning classification models refer to models that allocate a sample to a class of similar samples. On the other hand regression models try to assign a value to the sample. In automated seizure detection, a classifier's role is to decide whether a

sample is representing a seizure or not, therefore binary classification models are used for this task. Classifiers consist mainly of machine learning algorithms, thus, their application is divided in two parts: learning phase and operational phase. The learning phase consists of presenting to the model some training samples, from which the model will adjust or train itself to recognize them and classify them with the best accuracy. Regarding the fact if the training sample is labelled or not, the learning phase can be supervised or unsupervised. Different classifiers obtain better results for each kind training [37].
In the following sections will be presented some of the most used classifiers in EEG based seizure detection. Since seizure detection is a binary classification problem, regression models will not be taken in consideration. Classifiers will be divided into linear and non-linear methods. A particular attention will be given to neural networks, that will be presented separately, although they could also be divided in linear and non-linear. The last section will focus especially on deep learning algorithms, since they present a powerful classification tool and they will be the main classification method used further on in the thesis.

3.5.1 Linear Classifiers

Linear classifiers use linear decision functions to distinguish samples between classes. They are widely used, especially in real-time and online seizure detection or prediction, because of their low computational requirements. They are generally simple and easy to implement, however their drawback is that they assume that data is linearly separable. This issue can be often bypassed by cnsidering non-linear decision functions in the algorithm. For this reason, linear machine learning algorithms, such as support vector machines (SVM), are one of the most common classifiers in seizure detection and prediction.

Support Vector Machine

A SVM in its native formulation is a linear classifier, that uses discriminant hyperplanes to classify samples in a feature space. Simple, yet efficient algorithm, selects the hyperplanes by optimizing the decision boundary such way, that maximises the separation margin. The margin is the minimum distance between the discriminant hyperplane and the samples closest to it. The intent of this approach is to maximise the chances of a correct classification for future samples [22]. Furthermore, SVM uses a regularization parameter, that allows a certain amount of errors during training.

SVM assumes that the samples are linearly separable. Taking in consideration the

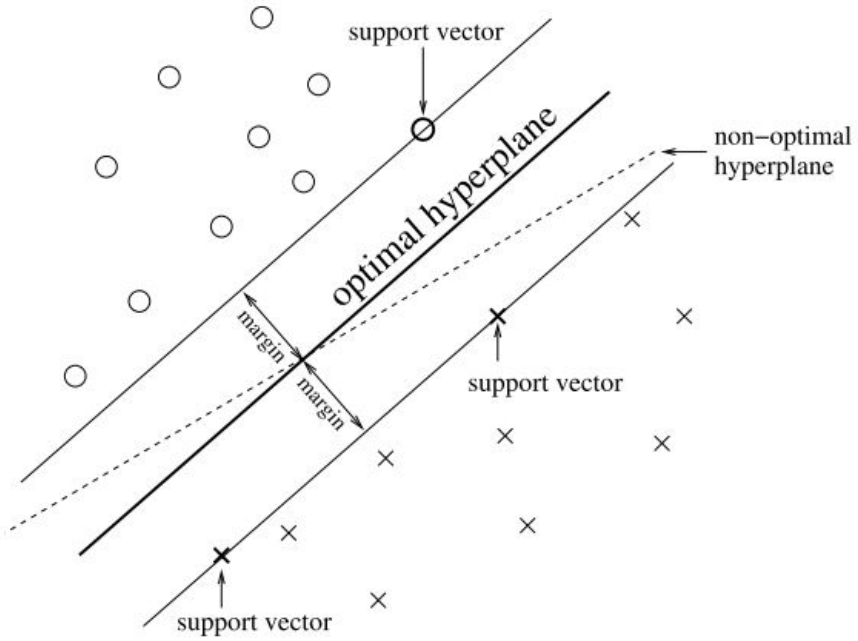


Figure 10: Finding the optimal hyperplane in the SVM. Source:[37]

figure 12, we can easily notice that a linear decision boundary can not efficiently discriminate the two labels on the plot B, as it does in the plot A. In such situations, a linear SVM would not be very accurate. However, a good characteristic about SVM is that it can easily learn to classify non-linear samples, by "pluging in" a non-linear kernel [29], [37]. In a recent paper on seizure detection, Harikumar et al., 2016 [20], compared a linear SVM, SVMs with different kernels and particle swarm optimization (PSO) algorithm. The results showed that RBF[3] SVM with gamma value of 10 outperformed all the rest of SVMs and PSO algorithms, by holding the highest accuracy of 99.20%. Results of this paper are presented in table 2.

Table 2: Comparison of seizure detection classification results with different SVMs and PSO algorithms. Source: [20]

Parameters	Linear SVM	$PolynomialSVMorder-2$	$RBFSVMwithGamma-5$	$RBFSVMwithGamma-10$	PSO
$Accuracy(\%)$	96.03	98.025	96.385	99.20	97.73
$Sensitivity(\%)$	100	100	100	100	97.68
$Specificity(\%)$	92.06	96.05	92.77	98.40	97.78

Radial Basis Function Kernel The Radial Basis function is often used in SVMs as a generalization kernel. A RBF is such function ϕ, so $\phi(x, c) = \phi(|x - c|)$ holds,

[3]Radial Basis Function (RBF) is explained in the following paragraph.

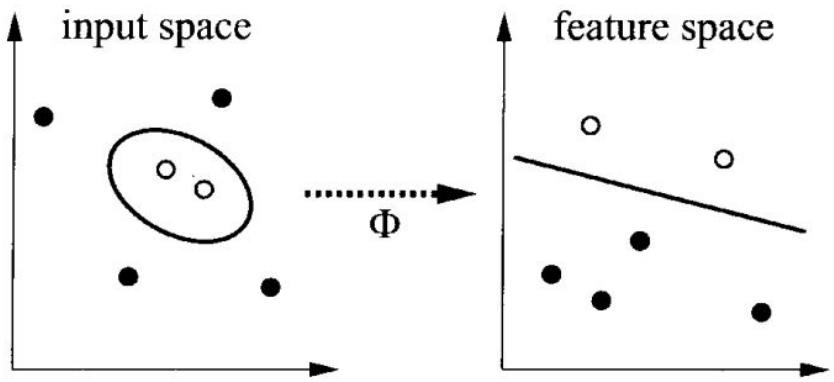


Figure 11: RBF-SVM solves non-linearity, by a mapping samples (Φ) in a higher linear feature space. Source:[49]

where c is the center. An example commonly used is the spherical Gaussian RBF:

$$\phi(x, c) = \exp^{\frac{|x-c|^2}{2\sigma}}$$

RBF is often used as activation function in a neural network, where the network discriminates samples using spheres or hyperspheres [6]. Usually it is constructed by first defining the number of radial functions and locating their centres, randomly or with with an algorithm, e.g k-means clustering. Then with training, the variance of the clusters is calculated and the weights and bias tuned [14].
Applying the RBF kernel to a SVM, does not alter the original algorithm. The idea is to use the kernel to map the non-linear samples into a higher feature-space, where the algorithm can classify data using linear decision functions. Moreover, the problems of selecting the number of functions and locating their centres is automatically solved by identifying the optimal hyperplanes and the support vectors in the higher linear feature space. Adding a kernel to the SVM does not change its characteristics. The model remains robust and still less prone to overfitting, even with high dimensional data, and not so computationally expensive. The simplicity, robustness and plasticity to generate other machine learning structures, made the SVMs a very popular classification tool, also in the fields of seizure detection and prediction.

3.5.2 Nonlinear Classifiers

As seen previously, sometimes linear classifiers are not efficient for a particular dataset, simply because the data cannot be divided with a linear function. Therefore other approaches are required to obtain better classification results. The next sections will

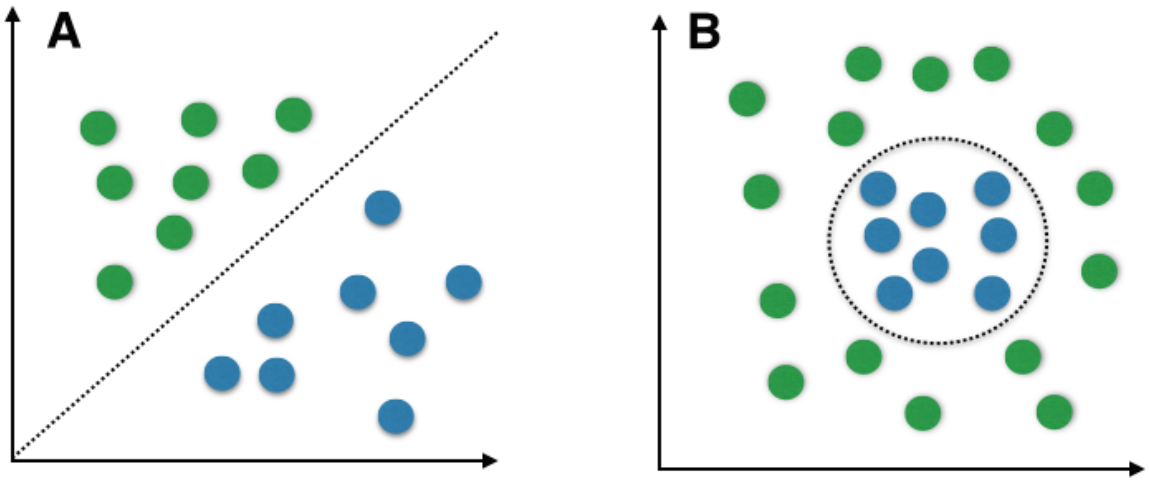


Figure 12: Graphic representation, showing that non-linearity is sometimes needed for better discrimination. Source: [46]

present some of the most common non-linear classifier used in seizure detection.

k-Nearest Neighbour

k-nearest neighbour is a non-parametric classifier, that works on both classification and regression. During training the algorithm locates the feature vectors in a multidimensional space and labels them with the appropriate class label. The next step consists in selecting a metric, normally Euclidean distance, and the 'k' parameter, which refers to the number of closest neighbours taken in consideration for a sample classification. Next, when an unknown sample is located in the feature space, its class is decided upon its k-nearest neighbours. Particular attention should be given to the tuning of the 'k' parameter, since could noticeable impact the results. High 'k' values tend to minimize the disturbances of noise, but also tend to make the boundaries between classes less distinct [8]. A good 'k' value selection can be based on heuristic hyperparameter optimization techniques or by identifying the empirical optimal value with the bootstrap method [19].

Bayesian Classifiers

Bayesian classifiers are probabilistic classifiers, that classify feature vectors regarding the probability of belonging to the class. Different Bayesian classifiers have been adopted in seizure detection and prediction studies, especially the Naïve Bayes classifier and the Hidden Markov Model.
Naïve Bayes classifier assumes strong independence among features of the feature vec-

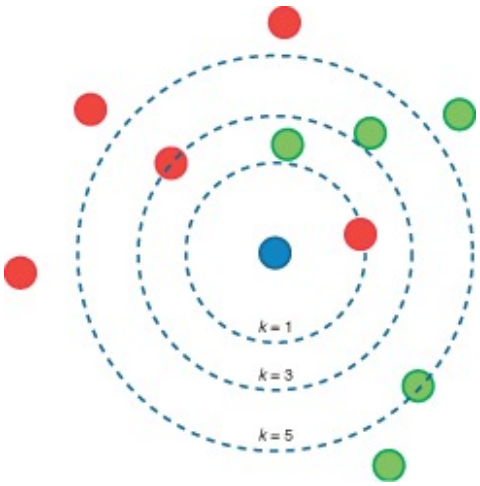


Figure 13: k-NN Classification.
Different settings of the "k" parameters may classify samples differently. Source: [49]

tor, which in real situations is difficult to encounter. Thus, the name naive. However, even when these conditions are not met, this method provides good results[14]. An instance is placed in the correct class when the probability of belonging to that class will be higher compared to the probability of other classes. Furthermore, the Bayesian inference can be replaced with maximum likelihood techniques, that make it computationally more feasible[14].
Hidden Markov Models (HMM) represent a popular branch of dynamic classifiers in speech recognition and particularly suitable for time series classification [45]. According to Wong et al.[61], HMM is especially well-suited for epilepsy studies since it is assumed that acquired EEG signals arise from underlying dynamical brain states. It consists of a probabilistic automaton, where the states are characterized with a probability distribution and transitions between states are determined by a set of probabilities. In each state an outcome is generated, regarding the probability distribution. During the process, the states are not visible from the outside, hence the name "hidden". Figure 14 shows an example of HMM. The image was taken from an experiment on seizure prediction. Therefore, four hidden states are taken in consideration: inter-ictal, pre-ictal, ictal and post-ictal. The visible part of the classifier are the clusters of point of interest, that were obtained with the segmentation of topographic maps.

In an experiment on epileptic seizure prediction, Direito et al. [12] used a hidden Markov model that would classify topographic maps, regarding their points of interest. On a study of 10 refractory epilepsy patients they averaged 89.31% accuracy, 94.59% sensitivity and 92.33% specificity.
Sharmila et al. [3] used patter recognition strategies to detect EEG signal abnormalities. They performed a discrete wavelet transform (DWT) to obtain features and used

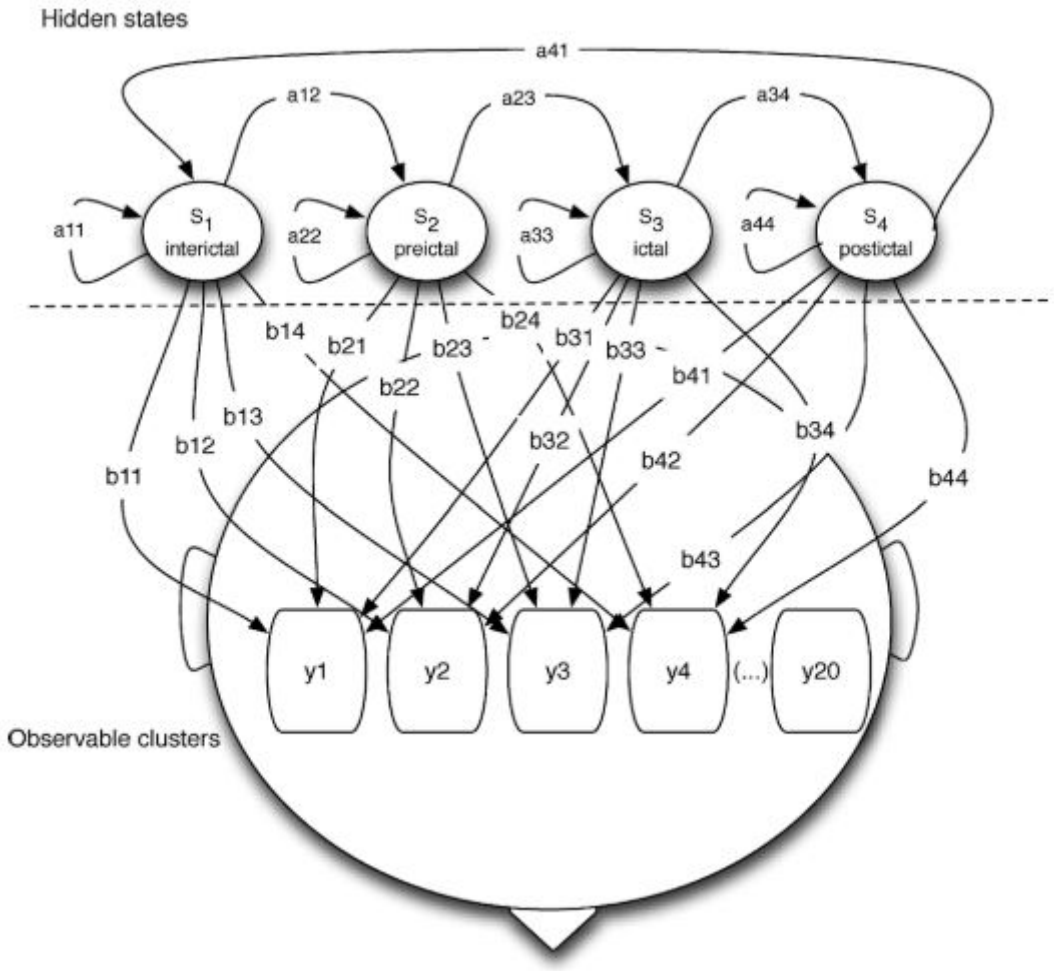


Figure 14: Example of a Hidden Markov Machine. Source:[12]

the naïve Bayes (NB) and the k-NN algorithms to classify them. With the NB they obtained 100% accuracy with 10 input features and 99.87% for k-NN, with only two features on a limited dataset.

3.5.3 Artificial Neural Networks

Among with linear classifiers, neural networks are the most used classifiers in the fields of seizure detection and prediction. The aim of these structures is to reproduce the real neural networks in the brain and their learning capability by a set of different interconnected layers of independent processing nodes, called neurons. In the typical neural networks, i.e. Artificial neural networks, the connections between neurons are weighted. The node sums the input signal and converts it to a value between 0 and 1. If the value reaches a certain threshold, an output signal is emitted from the neuron. In order for the network to give the correct result, the weights and other parameters have

to be tuned properly. Therefore, the network undergoes the "training" phase, where it learns to identify instances of a class, by an iterative adjustment of the parameters, until the point when the classification error drops under a certain level. Then, the network is supposed to be trained and able to classify correctly new instances[48].
Different artificial neural networks have been invented. One of the most representative ANNs is the Multi-layer Perceptron network (MLP). Similar to the previously seen radial basis function kernel Radial Basis Function network (RBFN).

Multi-Layer Perceptron and Radial Basis Function Network

The name perceptron is given by the homonym binary linear classifier - Perceptron, that classifies the feature vectors regarding hyperplanes in the feature space. A multi-layer perceptron is a class of feedforward artificial neural networks, composed by a stack of perceptrons. It comprehends an input and output layer, and several hidden layers in between them.
The layers are fully-connected, which means that every neuron is connected to every neuron from the previous layer. In the first layer a signal is inputted, then it passes through the nodes of the network, where is processed. Every node in the network, apart from the nodes of the first layer, have an activation function, usually nonlinear, such as

$$y(x) = tanh(x)$$

or

$$y(x) = (1 - e^{-x})^{-1},$$

that will allow only signals above a certain threshold to pass through the neuron. The neurons of the last (output) layer determine the class to which the instance appertains [37].
MLPs are usually used trained with backpropagation, in a supervised learning approach. Instances are paired with the correct class label and once they pass through the whole network, a classification error is estimated. At that point, a correction procedure starts from the last layer towards the first, where weights and bias values are modified, in order to obtain a higher classification score.
MLP are also know as universal approximators, meaning that with enough neurons and layers they can approximate any continuous function [37]. However, this characteristic makes them very sensitive to overtraining , especially with noisy and non-stationary data, such as EEG.
RBF network is another kind of neural network, that uses a radial basis function as activation function. Compared to MLPs, RBF networks can model any nonlinear function using a single hidden layer and require less training, since standard linear modelling

techniques can be employed. Moreover, they are not heuristic, hence do not suffer from problems such as local minima, which occurs in MLPs. However, RBF networks are known to give poor results, if the new data is located at the margins of the training sets, since the kernels do not represent efficiently the data far away from their centers[14].
In 2004, Srinivasan et al. [60] compared the classification performance of 4 different neural networks: MLP, Elman[4], probabilistic neural network and learning vector quantization. They discovered, that the Elman network outperformed the other three and yielded an accuracy, a specificity and a sensitivity of 98%. One year later, Subasi et al. [1] used a MLP-AAN to compare the classification scores of autoregression coefficients and features obtained with a fast Fourier transform of the signal of epileptic patients. The accuracy, specificity and sensitivity were 92.3%, 96.2% and 90.3% for the AR coefficients, while 91.6%, 93.6% and 89.8% for the FFT signals.

3.5.4 Deep Learning

More implicit knowledge is required to perform better decisions. Hence, more depth is required to represent more information from the dataset. Based on this premise, deep learning (DL) algorithms were designed [13]. Deep machine learning represents a branch of machine learning methods, that is based on representation learning. The network learns and extracts features from each of the numerous hidden layers of the structure [35]. The name "deep" is referred to those numerous hidden layers in the Artificial Neural Network-like structure. In the following sections are listed some of the most common DL networks.

Convolutional Neural Network (CNN)

One of the issues of the ANN is the susceptibility to translation and shift deviation, meaning that they fail to recognize an object if it is shifted or tilted, compared to the training samples. To overcome these limitations convolutional neural networks were designed. A CNN is a feedforward network, composed by an input and an output layer and multiple convolutional, pooling and fully-connected layers. Each layer can include different functions, e.g. activation functions, regularization functions and, for the output layer, a classifying function. Functions are usually also implemented as layers of the network.
CNN input layers can accept different kind of array data, as series, 2-D series, im-

[4]Elman network is a recurrent network with three layers. It is characterized by "context" units in the second layer, that mantain a sort of state in the network. This allows it to perform sequence-prediction tasks [11].

ages and 3-D images. In the convolutional layers, samples are organized in a feature map, regarding on a kernel - a set of weights, called filter bank. This allows to recognize different local patterns present among the samples. In image arrays such layers efficiently detect particular objects and shapes, independently from the position in pictures. Hence, the translation and shift problem, present in conventional neural networks, is overcome [35].
The next set of layers are pooling layers, that have the main task of down-sampling the outputs of the convolutional layers, by merging similar features into a single one. The down-sampling functions are average, max or sum [13]. In pictures, pooling upholds representing features from not varying too much, even though elements from which these are derived may be spatially distant or altered in appearance.
An other kind of layers are the fully-connected layers, where a neuron in the current layer is connected to all neurons from the previous one with weighted connections. The last fully-connected layer determines the number of output classes. During training with back-propagation, the algorithm corrects all the weights in the network, starting all the way from the last fully-connected layer to the first convolutional layer.
The last layer is the output layer, that is preceded by a classification function. The most common function is the "Softmax" function, that assigns to each possible class a number ranging from 0 to 1, obtained by a conversion the output of the last fully connected layer. The sum of these numbers is 1, hence, the function is similar to a distribution of the classification probability [16].
The role of the regularization functions is to reduce the training time and increase the learning accuracy. The activation functions are necessary to ensure the non-linearity, since the activation of a single neuron will lead to the activation of others. The most common activation functions in CNNs are presented in figure 15. An other regularization function is the "Dropout" function, that randomly removes a percentage of neurons from a layer in order to prevent the occurrence of phenomena such as overfitting. Beside these functions, a "Batch Normalization" function is often used with the intent to accelerate training, by normalizing the values in the hidden layers.

Long-Short Term Memory (LSTM) Network

An opposing idea to feed-forward networks are recurrent neural networks, characterised by their recursive approach. A product of the network is processed on behalf of previously learned information, that is stored in each node in form of inbuilt memory. Although RNNs are very powerful and dynamic systems, their training is very complicated since the back-propagated gradients tend to explode or vanish over many repetitions [5], [23]. RNNs resulted to perform really well, when operating with speech and text, or other sequential inputs [35].

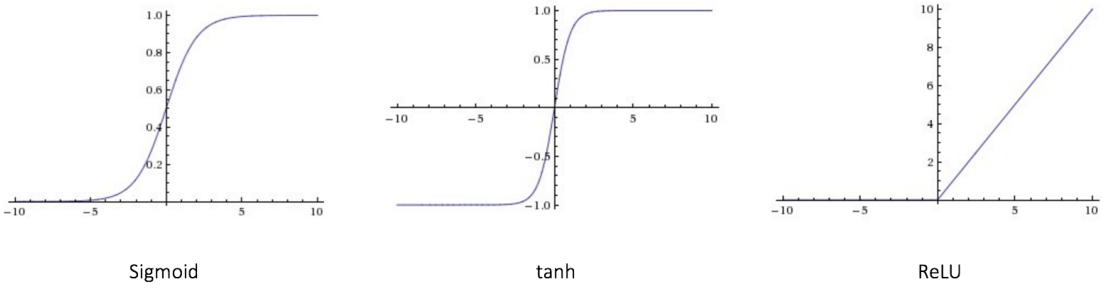


Figure 15: Common activation functions for CNNs.Source: [41]

A well established RNN is the Long Short-Termed Memory (LSTM) network, which main characteristic is the ability to learn long-term dependencies [23]. It uses a memory block within each neuron, that can modulated the cell state through three gates: (1) input gate, that decides if an information has to be updated and stores; (2) output gate, that decides if the information is used on the cell state; (3) forget gate, that deletes redundant or irrelevant information from the cell state. LSTM networks showed great results, especially when they include several depth layers per time step, enabling detection of either acoustics and character sequences [17].
An extension of common LSTM networks are bidirectional LSTM networks. A bidirectional approach presents both forward and backward sequences to two separated nets, that are connected to the same output layer. Their invention was mainly motivated by the fact, that the sequential readings tend to generate outputs that rely on previous context. Therefore, since bidirectional analysis would process every sequence in both ways, it would provide additional information, that may be profitable in terms of time efficiency and recognition performance. Basically, for every point in a given sequence, the such network has complete, sequential information about all points before and after it [17]. Furthermore, the issue of selecting an appropriate task-specific time window is avoided. It was found, that bi-LSTM is faster to train than other RNN and slightly more accurate, while solving speech recognition tasks.
Deep learning techniques are very recent. In 2016, Thodoroff et al. [57], used a recurrent convolutional neural network with LSTM memory nodes to build a mode for seizure detection. They have reached a 95% accuracy and an 85% sensitivity. Moreover, the model resulted very robust to missing channels, as the performance didn't decrease significantly, even with 16 channels missing. Hussain et al. [26] in 2018, created a seizure detection model, that is based on an ensemble of 1-D CNNs. 1-D CNNs take as input time series, which in this case was the EEG signal. In almost all the cases concerning epilepsy detection, the proposed system gives an accuracy of $99.1 \pm 0.9\%$. An other recent experiment by Hussain et al. [25], that also used CNNs, obtained

remarkable results. They input raw signal to the network, which had 4 convolutional layers. The results obtained were an average 90% sensitivity, 91% specificity and a 98.05% accuracy.

Deep learning represents a good candidate in physiological signal analysis, especially for its ability of dealing with multidimensional data and for its great representation learning performances. There is no need to perform an extensive feature extraction and selection for their training, which, in case of very complex biological signals as EEG signals, could potentially save much time and find the best feature subset by itself. However, the down part of DL is its training. This networks, because of its depth need many training samples, many training epochs and a good parameter tuning.

Convolutional networks are very effective for image classification, while LSTMs perform well on sequential data. Furthermore, bi-LSTM's abilities to learn important patterns in both directions could be applied to the seizure detection problem. Therefore, in this thesis we will use deep learning networks that will learn to detect an occurring seizure attack, on behalf of topographic maps that were generated from the power signal of five frequency sub-bands.

3.6 Regularization

False positive classifications could be really annoying to patients who rely on the system to early detect their seizures. One way to remove them, or at least reduce them, is to perform a post-processing task.

Generally speaking, this step is more common in seizure prediction systems, where they aim to detect the upcoming seizure minutes before the actual onset. The idea behind this process is to make the system ignore the single false positive classifications that occur sparsely and trigger an alarm when their concentration reaches a certain threshold. The two most common post-processing algorithms for seizure prediction are the firing power, that triggers an alarm when a certain threshold of false positive predictions is reached, and the Kalman Filter, generally applied to SVMs, that reduces the false positives using the classification variable [54].

In seizure detection the regularization methods are simpler. Classifiers do not classify all the samples correctly. Therefore, since in a real-time test there are much more non-seizure samples than seizure one, some false predictions might occur. To solve this issue generally an average moving filter (MAF) is applied. The filter, of a chosen size "k", attributes to the sample in the center of the filter the average value of all the samples in the filter. Thus, with a correct tuning of the parameter k, sparse misclasifications are removed.

3.7 Evaluation of the Performance

Evaluation of the results with some metrics, that is generally used in the field, allows to some extent the comparison of results. In seizure detection the most used evaluation metrics are sensitivity (SS), specificity (SP) and accuracy (AC). Often, some researchers also use the precision (PR). Formulas for their calculation are shown in table 4 and are calculated with the confusion matrix results the testing set. A general confusion matrix is shown in the table 3.

All these measures hold different information regarding the model's classification abil-

Table 3: Confusion Matrix Example. S = Seizure and N = Non-Seizure.

		Prediction Outcome	
Actual Value		**S**	**N**
	S	True Positive (TP)	False Negative (FN)
	N	False Positive (FP)	True Negative (TN)

ity: accuracy shows the percentage of correctly classified samples in the test. However, there is no information given about individual labels. On the other hand, sensitivity indicates the proportion, that the model will correctly classify a positive sample. The same measure for the opposite label is called specificity. The precision of a model, or sometimes mentioned as positive predictive value, is that measure representing the percentage that an instance, classified as positive in the test, is actually really positive. Similarly, although is less used, the same measure for the negative label is called negative predictive value.

While testing a seizure detection model, there is a much bigger proportions of non-

Table 4: Table showing different evaluation metrics and their formulas.

Accuracy (AC):	$AC = \frac{TP+TN}{TP+TN+FP+FN}$
Sensitivity (SS):	$SS = \frac{TP}{TP+FN}$
Specificity (SP):	$SP = \frac{TN}{TN+FP}$
Precision (PR):	$PR = \frac{TP}{TP+FP}$

ictal samples. The chances of having an elevated number of FP predictions are very high, even with a very accurate model. Therefore, a better understanding of the performance of the model is given by its specificity and precision.

4 Experimental Data and Setting

The study was conducted in *MATLAB*® environment. It follows a general pipeline, that comprises the following steps:

- patient selection,
- pre-processing,
- topographic map generation,
- classification,
- regularization,
- performance evaluation.

A graphic representation of the pipeline is also shown in figure 16.

This pipeline is similar to the one previously mentioned in chapter 3, although differring in some points. In this thesis, since we will use convolutional neural networks, we will not perform an extensive feature extraction and selection. The network is capable of recognizing important features of the image by itself and perform an efficient classification. The signal transformations, will be presented in the section "Pre-Processing". The aim of this chapter is to present detailed information about each step of the pipeline. The first section contains a description of the data analyzed in the thesis. It will be followed by a section explaining the pre-processing procedures and another one explaining how the topographic maps were generated. The next section describes the

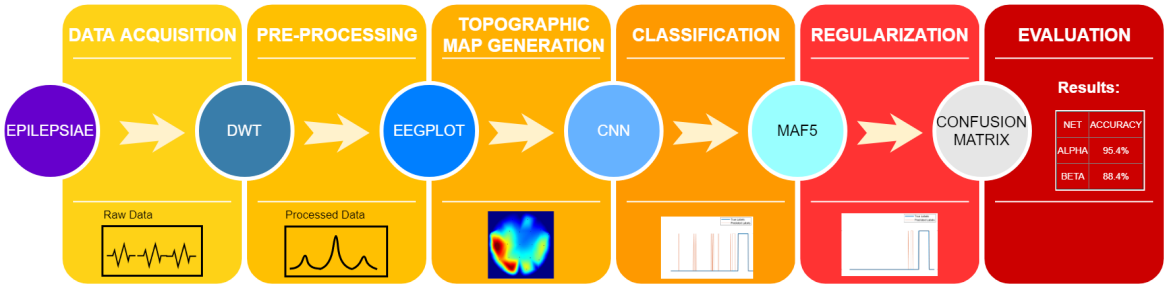


Figure 16: Graphic representation of the pipeline of the thesis.

construction and training of the CNN, and, finally, the regularization and evaluation methods will be presented.
The study can be divided into two parts. In the first one, the model was constructed with a limited amount of samples. However, given the interesting results from the first part, a second attempt, with a higher number of samples took place. The first part will be further on referred as “First Testing”, while the second as “Second Testing”.

4.1 Patient Selection from the EEG Database

Data was retrieved from the EPILEPSIAE database. The patients, that were selected from the database needed to meet some specifications. First, we focused only on scalp EEG, not taking in account intracranial EEG recordings. Then, we selected only patients with focal epilepsy, with a well defined focal region in the temporal lobe. Next, the patients were filtered by taking into account the recording sampling rate, which was set at 256Hz.
In total, five patients were identified for this study, two males and three females, from 22 to 54 years of age, averaging 39.8 years of age. Recordings duration range from 94.4 to 164.3 hours of length, averaging 125.46 hours and patients on average were struck by 2.65 seizures per day. Finally, we at first we carefully analyzed the patient 58602, who held the highest number of seizures.
Recordings were accounted for 24 channels, however 3 channels - ”RS”,”SP1”,”SP2” - were discarded before generating topographic maps. Table 5 shows which are the five patients with most seizures per day.

Table 5: Table showing the patients used in the study and their characteristics.

patient	age	gender	num. seiz.	sampling rate	duration(h)	num.seiz./day	focus localisation
58602	32	m	22	256	161.1	3.28	t
81102	41	m	13	256	96.9	3.22	t
85202	54	f	10	256	94.4	2.54	t
80702	22	f	10	256	110.6	2.17	t
95202	50	f	14	256	164.3	2.05	t
Average	39.8		13.8		125.46	2.65	t = temporal

4.2 Pre-Processing and Feature Extraction

Raw data comes in the form of binary files. Therefore, it needs to undergo few pre-processing steps, before it can be used to generate topographic maps. First of all, data has to be translated to .mat files, so it can be easily managed in further steps. Here, functions from the *EPILAB*® [53] package were used to perform the task. Next, a high-pass filter is used to remove the frequencies, between 0.1Hz and the Nyquist frequency, which in this case consists of 12 Hz, and then a 50 Hz notch filter is applied to remove possible power line artifacts.

Due to non-stationarity of the signal, data has to be analysed in short sliding time win-

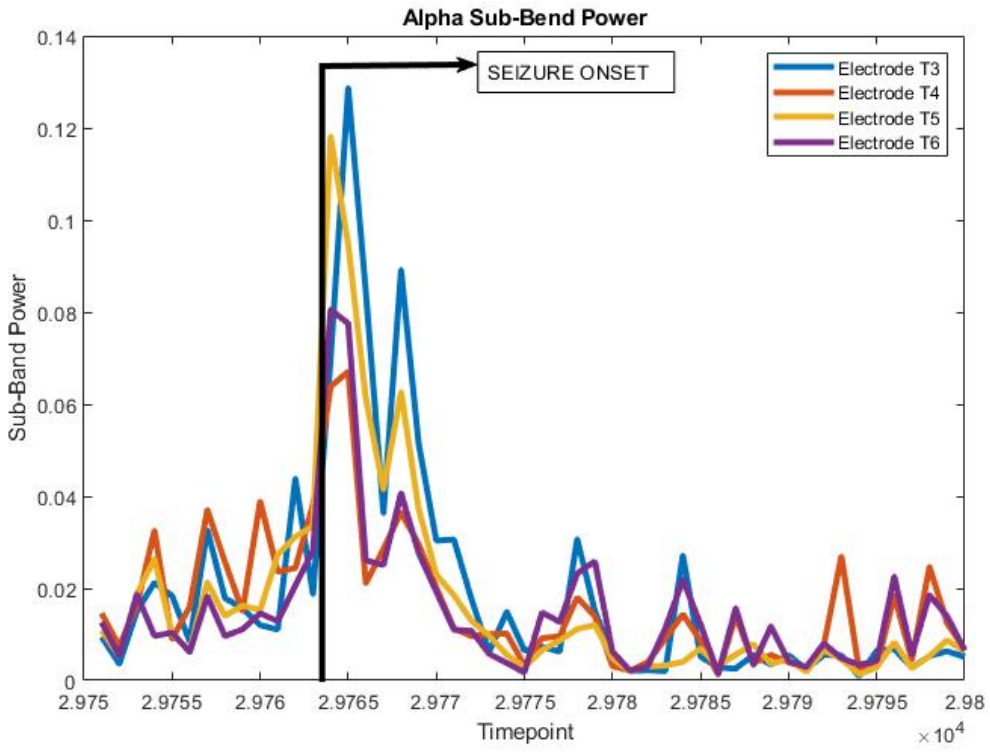


Figure 17: Example of the alpha frequency sub-band power signal, close to the seizure onset.

dows. Initially, the chosen length was of 5 seconds, with an overlap of 80%. This allows to assume stationarity in these five second windows and preserve frequency resolution. Furthermore, by overlapping by 80% we obtain four time more samples than we would have obtain without it and we might detect additional information that could not be captured between the end of a window and the beginning of another. Five seconds window's length represent a good compromise to keep sufficient time and frequency resolutions and is often used in EEG analysis.

While pre-processing the data in time windows, 5 basic frequency features were ex-

tracted. Features corresponded to the relative powers of different frequency sub-bands obtained with the Discrete Fourier Transform. The bands included the following frequencies: $\delta(0.5-3.5Hz), \theta(4-7.5Hz), \alpha(8-12Hz), \beta(13-35Hz), \gamma(35-60Hz)$. All vectors were normalized, by dividing their values by the total power of the signal's spectrum. In figure 17 there is an example of the alpha sub-band power calculated at the electrodes T3, T4, T5 and T6 at the time of seizure onset. The patient has focal epilepsy, with seizures starting from the left temporal lobe (channels T3 and T5).
After pre-processing, patient 58602 resulted having 974 seizure samples for each feature. However, one seizure was removed (originally seizure 13), since in the annotation file was described as not reliable. Therefore, the total number of seizure samples reduced to 929. We proceeded to generate topographic maps and later on to train the networks. However, the data, especially the seizure samples, was not sufficient for the testing phase. Thus, for the second round of training, we re-extracted data, with a higher overlapping percentage. The sliding windows now overlapped by 98%, meaning that the new window starts 0.1 s after the previous one. By doing this we augmented the data 10 times. We obtained 9290 seizure samples for each feature of the patient, which provided enough data for a qualitative training and testing.

4.3 Topographic Map Generation

The simplest way to implement a topographic map is to use a grid or matrix, where the values for each recording are spatially distributed on the grid according to their distribution on the scalp and then with an interpolation function fill the blank cells between them. In this thesis we used a pre-made function, eegplot().m by Ikaro Silva [51], freely available on MATLAB®file exchange. The interpolation function used was the cubic interpolation function. Maps were saved as coloured pictures in .jpg format, with a width of 884 pixels and a height of 766 pixels and a 24 bit depth.
For the first study, topographic maps were generated from the pre-processed data, using all of the seizure timepoints and the same number of non-seizure timpoints. Non-seizure samples were randomly selected from the non-seizure intervals. For the patient 58602 were generated 929 ictal topographic maps and 929 non-ictal (set A).
Further in the thesis, for a better representation of the performance, we generated topographic maps for the last five minutes before the seizure. We wanted to inspect and interpret the misclassification of some samples, especially of those very close to the seizure onset. Therefore, we had 6775 non-seizure samples, 300 (equal to 5 minutes) before each seizure of the patient and some samples after the seizure, corresponding to the half of the seizure timepoints (set B). The seizure samples where the same as

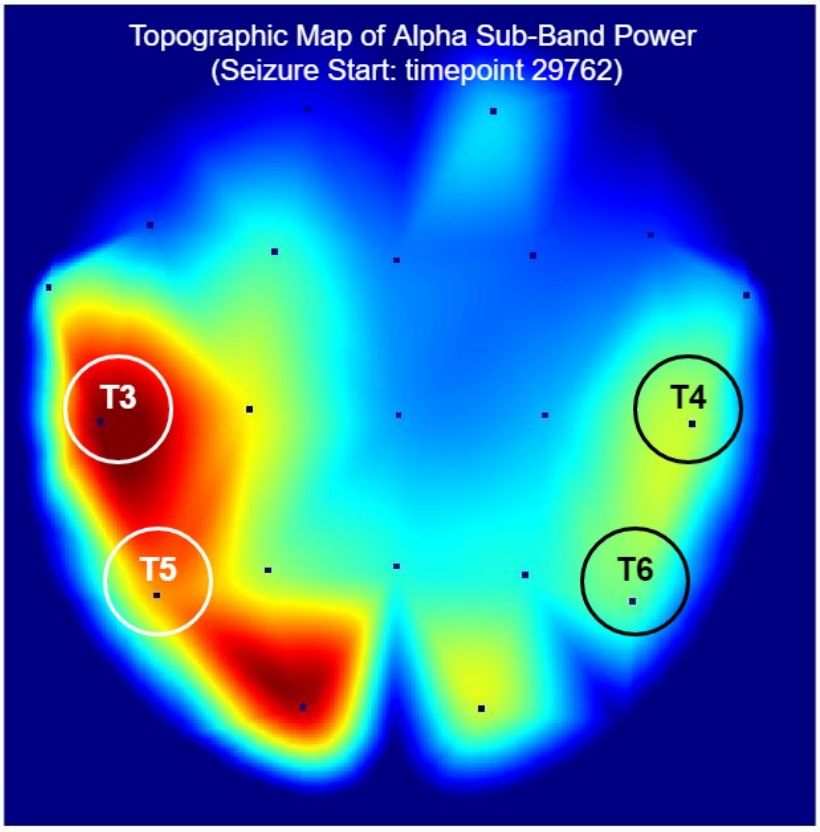


Figure 18: A topographic map of the seizure onset from the plot from figure 17, where the same electrodes are highlighted.

previously used for training (set A).
Once again we generated new topographic maps for the second study. In this case, the samples used for training were generated from each second in the range of 5 minutes to one minute prior the seizure. We intentionally left the last minute prior the seizure out of the training process, to see how will the model behave without that interval. Our intention was to obtain again some sequences of false alarms close to the seizure onset, as in some seizures of the first results (see figure **??**). Also samples taken randomly from the inter-ictal intervals were used balance the datasets. Finally, the dataset for each class had 8342 samples (set C).
For the second testing, we used intervals of 1 hour prior the seizure onset, where samples were taken every 2 seconds. Therefore, 37800 non-seizure and 950 seizure samples were generated (set D). A better representation of the sets of topographic maps used in this thesis is offered in table 6.

Table 6: Sets of topographic maps generated in the thesis.

Data Set	Objective	Description:
A (1st Study)	Training + Testing	- 929 ictal samples - 929 non-ictal samples taken randomly Separation of data: 80% train., 10% valid., 10% test
B (1st Study)	Testing	- 929 ictal samples (same as data set A) - 6775 non-ictal samples (5 min before the seizure + half the number of seizure samples, after the seizure)
C (2nd Study)	Training	- 8342 ictal samples - 5280 non-ictal samples (5-1 min before the seizure) - 3062 non-ictal samples (taken randomly) 20% of the samples were used for validation
D (2nd Study)	Testing	- 950 ictal samples - 37800 non-ictal samples (1 hour before the seizure, samples taken every two seconds)

4.4 Classification with CNN

In this thesis we used a Convolutional Neural Network, which is particularly suitable for image classification. The network was constructed with the $MATLAB$® Deep Learning Toolbox. In this section we will describe the chosen structure and the training procedures.

4.5 Network Structure

In this thesis, each feature has its own network, because each one needs different parameters to perform a better classification. However, all the networks have the same layer structure with the same 19 layers. They are described in table 7. The parameter selection for each network was performed with a grid search approach, presented in the next section.

The input layer takes images of size 766x884x3. Then the images enter the first convolutional layer, where 16 kernels of either size 3x3 or 5x5 pass through the image and generate a first feature map. Next, this feature maps are normalized in a batch normalization function, with the intent to speed up the learning process. A ReLU function is used as activation function, to rectify the values, before they undergo the pooling operation. This process is repeated three times, with three different convolutional layers. Next the feature maps pass through two fully connected layers, with the

Table 7: Example of a convolutional neural network used in the thesis.

L.	Name	Type	Description	Output	Learnables
1	InputLayer	Input	SIZE: 766x884x3	766x884x3	0
2	Conv1	Convolutional Layer	KERNEL:3x3x3 STRIDE:2x2 BIAS:1x1x3 NUM.KERNELS:16	383x442x16	448
3	BatchNorm1	Batch Normalization Function	CHANNELS: 16	383x442x16	32
4	ReLu1	ReLu Activation Function		383x442x16	0
5	MaxPool1	Max Pooling Layer	SIZE: 4x4 STRIDE 1x1	383x442x8	0
6	Conv2	Convolutional Layer	KERNEL:2x2x16 STRIDE:1x1 BIAS:1x1x3 NUM.KERNELS:8	383x442x8	520
7	BatchNorm2	Batch Normalization Function	CHANNELS: 8	192x221x4	8
8	ReLu2	ReLu Activation Function		383x442x8	0
10	MaxPool2	MaxPool Activation Function	SIZE: 3x3 STRIDE 1x1	192x221x8	0
11	Conv3	Convolutional Layer	KERNEL:2x2x8 STRIDE:1x1 BIAS:1x1x3 NUM.KERNELS:4	192x221x4	132
12	BatchNorm3	Batch Normalization Function	CHANNELS: 4	192x221x4	8
13	ReLu3	ReLu Activation Function		192x221x4	0
15	FullCon1	Fully Connected Layer	NODES:32	1x1x32	5431328
16	ReLu4	ReLu Activation Function		1x1x32	0
17	FullCon2	Fully Connected Layer	NODES:2	1x1x2	66
18	SoftMax	SoftMax Function		1x1x2	0
19	ClassOutput	Classification Layer			0

outputs 32 and 2 relatively and the Softmax function calculates the output.

4.6 First Study

Deep learning requires a big amount of samples for its training. The patient had 929 seizure samples, which are not enough for an efficient learning. Therefore, we decided to train CNNs separately for a subset of the previously extracted features and assemble them into one singular classifier. The final classifier would make its decision based on a majority voting. Hence, an odd number of features is suggested. We decided to use the relative powers of the frequency sub-bands δ,θ,α,β and γ.

4.6.1 Parameters Tuning: Grid Search Approach

During the first training of the CNN some parameters were fixed, while others were left to be tuned with a grid search. The variable parameters were the size of filters of the convolutional layers and the dropout percentage for both dropout layers. There

were two possible combination for the sizes for the filters at each layer and two possible dropout percentages for each dropout layer. Totally, each grid search had 32 possible combinations.
Grid search was performed for all five frequency sub-bands, delta, theta, alpha, beta, and gamma. Of all the images, 80% were used for training, 10% for validation during training and 10% for testing after training. For all the features, each of these subsets included the images with the same indexes. Totally, for each feature there were 1486 training samples, 186 validation samples and 186 testing samples.
Since the images were of big sizes, training required the use of mini-batches. We set the mini-batch size of 16 samples. Samples are randomly shuffled at the beginning of every training epoch. We opted for 32 training epochs, because represents a good compromise between training time duration and classification results. More epochs would require more training time for each network, which in a grid search approach is very restrictive. The chosen optimizer for training was the "Root Mean Square Propagation" (RMSprop) optimizer, since it is particularly suitable when using mini-batches. The training was validated every 30 iterations.
After training, every candidate network was tested with the test set. The final network for each feature was selected with two criteria: first, the true positive and true negative scores shall not differ for more than 10% between each other, second, the network with the highest accuracy is selected. The first criteria is needed to filter all the results, that obtained very high scores for only one class, but poorly classified the other. In table 8 are presented the confusion matrices of the best performing networks for each feature. Next, the nets were assembled and tested together.

4.6.2 Testing with a Pre-ictal Interval of 5 Minutes (B)

Next, we tried to test the performance of the ensemble on a interval from five minutes prior the seizure onset until the seizure end. Such performance is in fact more similar to the real life situation: the ultimate goal of the classifier is to warn the patient about an imminent attack. Hence, we are more interested on the performance of the model in this five minute window interval than from the classification of randomly sparse samples. With this premise, we generated new topographic maps, for every second of the five minute interval, preceding each seizure. The results are presented in the appendix A.

4.7 Second Study

Due to the few seizure samples, a new extraction of features, with a higher overlap (98%) was performed. Consequently, we opted to also re-train the networks, since more data would probably give better classification results. However, in this second round of training we did not perform a grid search of the parameters. Instead, we used the parameters that resulted as best in the previous grid search. Furthermore, compared to the first training, the samples used for the second training were taken from specific pre-ictal intervals, rather then from random locations in the recording, as explained in section 4.3.
For this training we had 8342 seizure samples for each feature, making 16684 total training samples. 20% of these were used for validation. Differently from the first training, networks were trained for 16 epochs, with a validation frequency of 500 iterations. This changes were made because of the highly increased number of samples.

4.7.1 Testing with a Pre-ictal Interval of 1 Hour (D)

Similarly as performed in the first study, we tested the performance of the ensemble on an interval prior to the seizure onset. However, this time the interval was much larger, covering one hour before the seizure onset. To reduce the time of the generation of topographic maps, we selected samples every two seconds (0.5 Hz) instead of one per second (1 Hz), as in the first testing. Figures for each seizure are presented in the appendix B.

4.8 Regularization

When testing on a five minute interval, we noticed that the number of false-positive predictions increased significantly. Therefore, since the data was in the form of a timeseries, we post-processed the results, with moving average filter (MAF) of size 5. The same procedure was also performed in the second testing, while testing on one hour pre-ictal intervals. An example of the effect of the MAF5 filter is shown in figure 19.

4.9 Evaluation

Almost all the ensembles were evaluated with unseen testing sets, with the exception of the B testing set, that included ictal data already used for training. We used accuracy, sensitivity and specificity to measure the performance of the model. The results are shown in the next section.

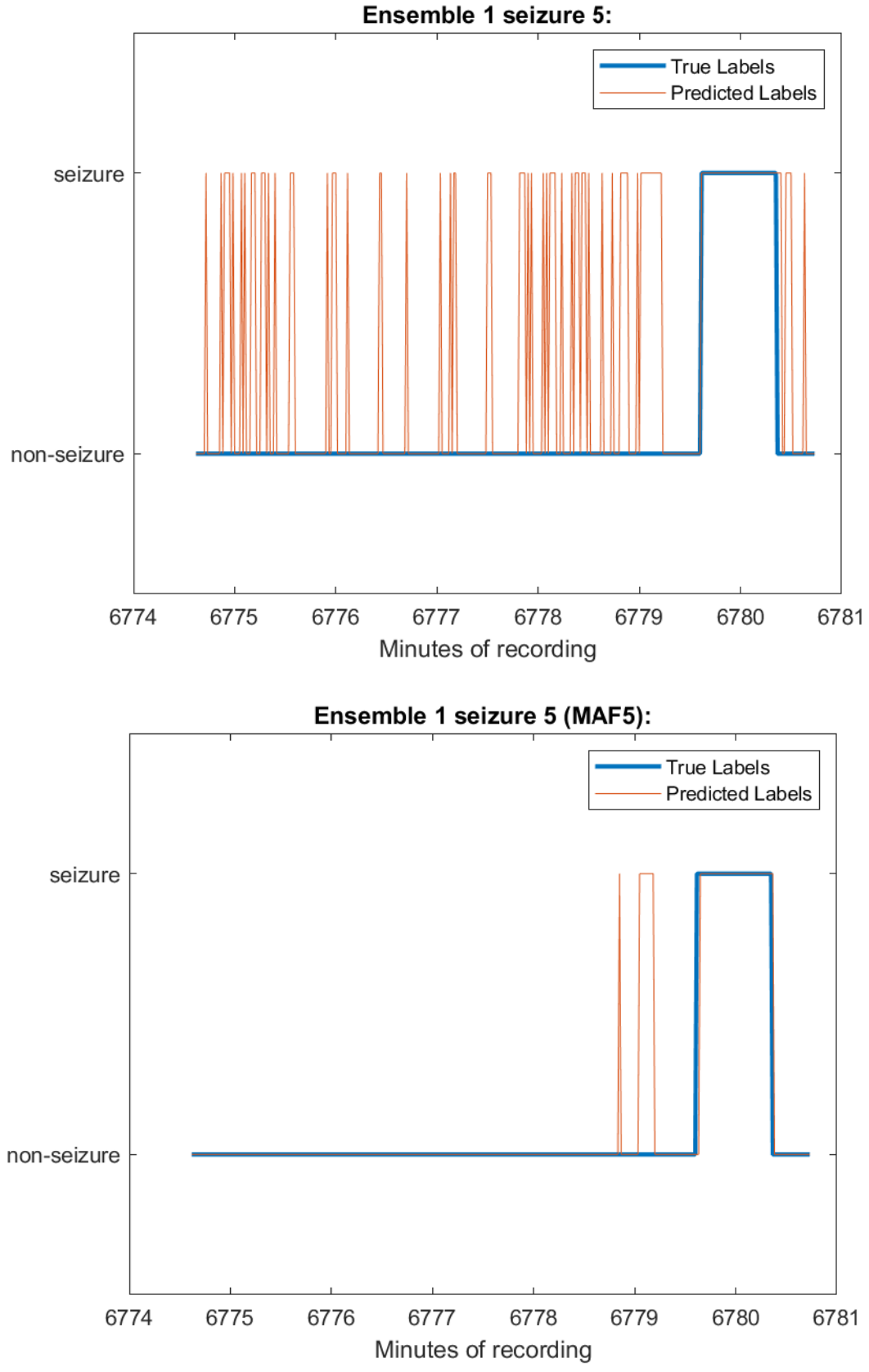


Figure 19: Figure showing the differences before and after applying the MAF5 filter.

5 Results

5.1 First Study Results

Grid Search Results: After only 32 epochs of training, good classification scores were obtained for the individual networks, even with the limited amount of samples. The most accurate network was the network of the theta sub-band relative power, with an accuracy of 95.16% and the least accurate was the network for the delta sub-band power, with an accuracy of 86.02%. More detailed results are shown in the table 8. Although some networks could be further trained, we decided to test the ensemble with the current networks. We used the same testing set as for testing the individual networks. The ensemble improved the accuracy and the specificity, however lost some sensitivity compared to the network of the Theta sub-band.

Table 8: Classification results of the A test set.

Sub-Band	TP	TN	FP	FN	AC	SS	SP
Alpha	85	77	16	8	87.1%	91.4%	82.8%
Beta	86	79	14	7	88.71%	92.47%	84.95%
Gamma	80	86	7	13	89.25%	86.02%	92.47%
Delta	80	80	13	13	86.02%	86.02%	86.02%
Theta	89	88	5	4	95.16%	95.7%	94.62%
Ensemble 1	88	91	2	5	96.23%	94.62%	97.85%

Five Minute Pre-Ictal Interval Testing: Testing the model on the five minute pre-ictal interval is believed to give more insights on the model's behavior close to the seizure onset. The model yield a high positive prediction value, misclassifying only 3 seizure samples out of 929, as can be seen in figure 20. However, careful interpretation of this results should be considered, due to the biased nature of the ictal testing data (see Discussion section).

This test also shows a noticeable increase of false positive classifications, that we managed to overcome with a moving average filter of size five. Considering the limited training data and the fact, that the networks were trained on randomly picked samples, the ensemble, after applying the MAF5 filter, still obtained 96.98% accuracy,

92.56% sensitivity and 97.23% specificity. Seizure-specific information of the test is presented in table 9.

Table 9: Seizure-wise results of the second ensemble testing the set B.

	Normal			MAF5		
Seizure	AC	SS	SP	AC	SS	SP
1	81.79%	100.00%	79.26%	98.37%	95.56%	98.76%
2	89.40%	100.00%	87.93%	93.75%	95.56%	93.50%
3	91.58%	100.00%	90.40%	94.02%	95.56%	93.81%
4	92.39%	100.00%	91.33%	96.47%	91.11%	97.21%
5	76.90%	100.00%	73.68%	96.47%	95.56%	96.59%
6	94.29%	97.78%	93.81%	97.83%	86.67%	99.38%
7	95.38%	100.00%	94.74%	98.91%	95.56%	99.38%
8	98.91%	100.00%	98.76%	98.91%	91.11%	100.00%
9	93.75%	100.00%	92.88%	95.92%	91.11%	96.59%
10	91.30%	100.00%	90.09%	98.10%	95.56%	98.45%
11	89.13%	100.00%	87.62%	93.75%	91.11%	94.12%
12	93.21%	100.00%	92.26%	98.10%	91.11%	99.07%
13	92.66%	97.78%	91.95%	96.20%	86.67%	97.52%
14	91.03%	100.00%	89.78%	98.37%	91.11%	99.38%
15	91.03%	100.00%	89.78%	95.38%	91.11%	95.98%
16	93.48%	100.00%	92.57%	98.91%	91.11%	100.00%
17	91.03%	100.00%	89.78%	93.48%	95.56%	93.19%
18	88.95%	100.00%	87.94%	95.06%	93.10%	95.24%
19	95.38%	97.78%	95.05%	97.01%	88.89%	98.14%
20	92.12%	100.00%	91.02%	94.84%	91.11%	95.36%
21	92.66%	100.00%	91.64%	97.83%	95.56%	98.14%
Average:	**90.99%**	**99.65%**	**89.88%**	**96.68%**	**92.56%**	**97.23%**

5.2 Second Traning Results

One Hour Pre-Ictal Interval Testing Similarly as in the first part, we tested the ensemble on a time series of samples. The pre-ictal interval had the length of one hour and the non-ictal samples were selected every two seconds (set D). After applying the MAF5 filter, the ensemble obtained 99.20% accuracy, 96.48% senesitivity

and 99.27% specificity. Seizure-specific results and results without using the MAF5 filter are presented in the table 10, while the confusion matrices of the D dataset classification are shown in figure 21.

Table 10: Seizure-wise results of the second ensemble testing the set D.

	Normal			MAF5		
Seizure	AC	SS	SP	AC	SS	SP
1	81.69%	100.00%	81.22%	98.81%	100.00%	98.78%
2	73.35%	100.00%	72.67%	97.78%	97.83%	97.78%
3	79.09%	97.83%	78.61%	98.43%	89.13%	98.67%
4	89.27%	100.00%	89.00%	99.89%	97.83%	99.94%
5	88.08%	95.65%	87.89%	99.13%	91.30%	99.33%
6	74.70%	100.00%	74.06%	98.27%	97.83%	98.28%
7	90.25%	100.00%	90.00%	99.67%	97.83%	99.72%
8	87.27%	100.00%	86.94%	99.40%	97.83%	99.44%
9	91.98%	100.00%	91.78%	99.51%	100.00%	99.50%
10	77.19%	100.00%	76.61%	98.16%	97.83%	98.17%
11	82.99%	100.00%	82.56%	99.78%	100.00%	99.78%
12	91.01%	100.00%	90.78%	99.67%	100.00%	99.67%
13	83.42%	100.00%	83.00%	99.08%	97.83%	99.11%
14	83.91%	100.00%	83.50%	99.30%	100.00%	99.28%
15	85.92%	95.65%	85.67%	99.35%	91.30%	99.56%
16	89.17%	100.00%	88.89%	99.46%	97.83%	99.50%
17	86.94%	100.00%	86.61%	99.51%	95.65%	99.61%
18	89.45%	100.00%	89.28%	99.95%	100.00%	99.94%
19	93.82%	97.83%	93.72%	99.78%	93.48%	99.94%
20	85.86%	95.65%	85.61%	99.13%	91.30%	99.33%
21	86.67%	95.65%	86.44%	99.19%	91.30%	99.39%
Average:	**85.34%**	**98.96%**	**84.99%**	**99.20%**	**96.48%**	**99.27%**

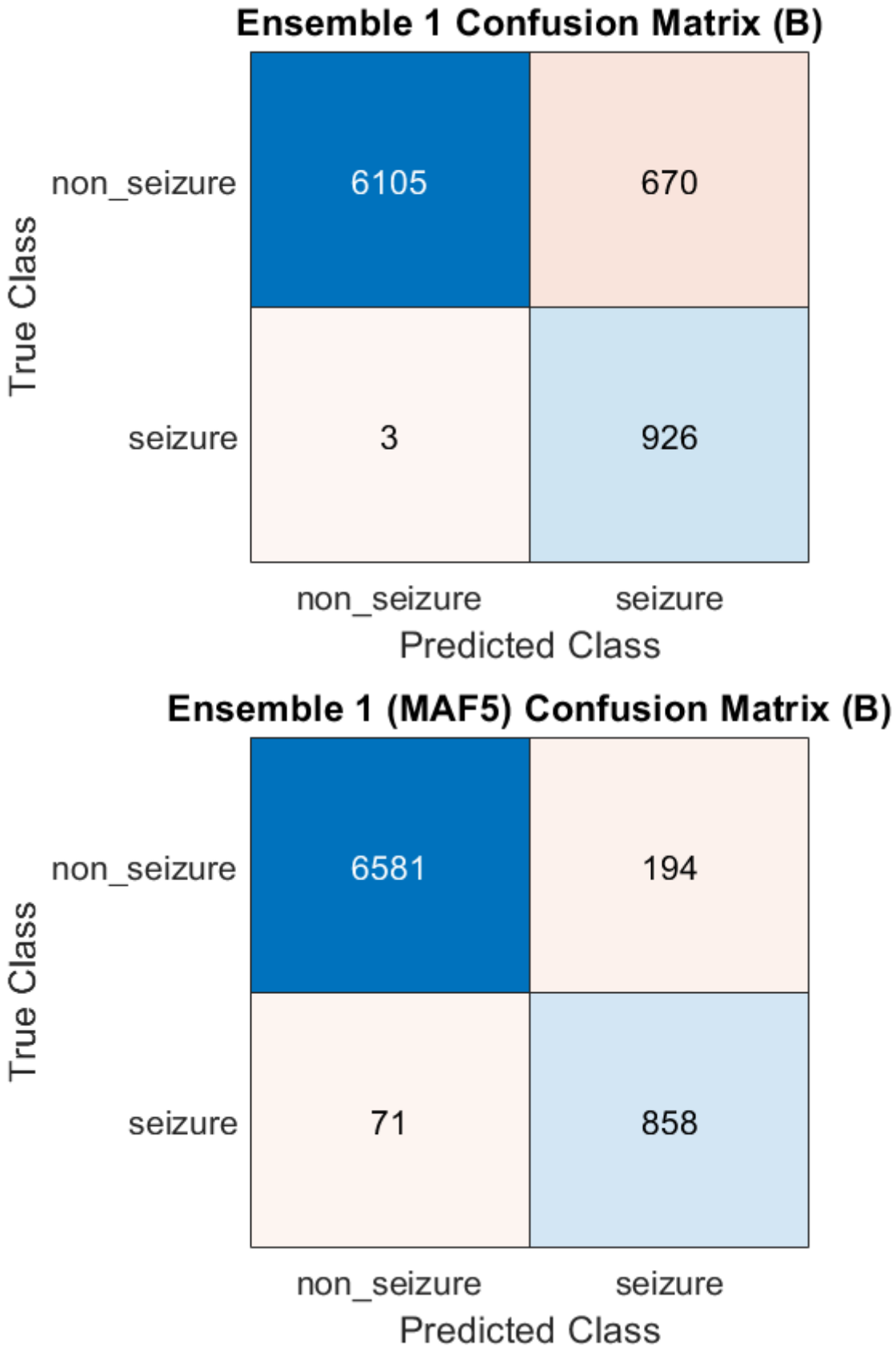


Figure 20: Confusion Matrices for the classification of the first ensemble (set B).

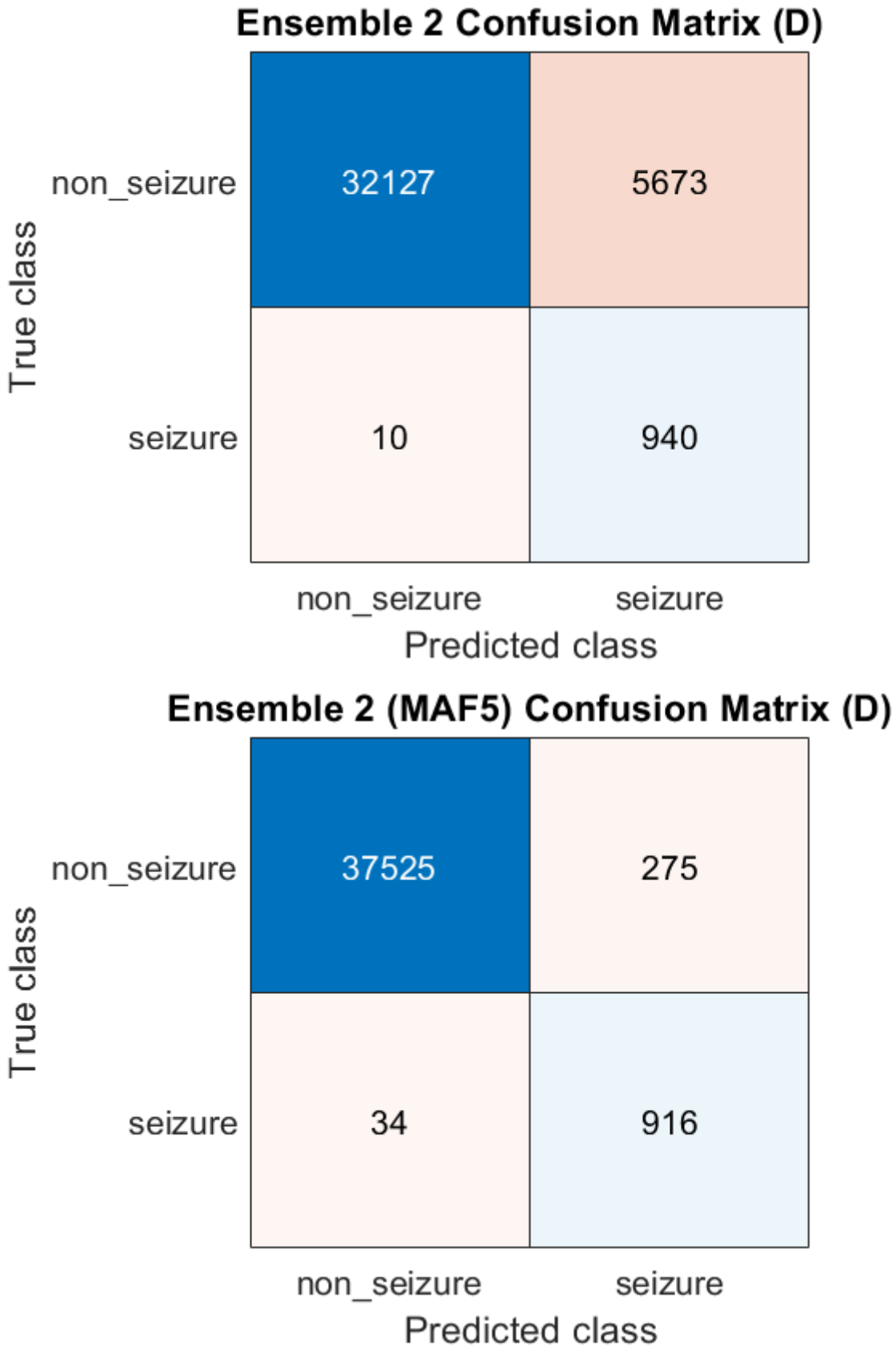


Figure 21: Confusion Matrices for the classification of the second ensemble (set D).

6 Discussion

As seen in table 8 the ensemble classified the testing samples with an accuracy of 0.974 and specificity of 0.969 and sensitivity of 0.979. However, it has to be taken in consideration the limited data, used for both training and testing and how it was extracted. Moreover, the non-ictal samples were taken randomly from the inter-ictal intervals, instead from an interval close to the seizure onset, which is believed to have more similarities with the ictal data. This might have made the two sets of samples more distinguishable and easier to train. Therefore, we proceeded with the testing of the dataset B.
We tried to see how this ensemble performs the classification samples close to the seizure onset. We extracted new samples for five minutes before the seizure onset. As we can easily notice from the first matrix in figure 20, there is a high number of false positives, while the seizures remain classified with a high score. We believe that this high classification score of the seizure set is due the fact that 80% of the seizure samples used for this testing, were also the same used for training the network. Therefore it is believed, that the network "remembered" those samples and classified them correctly.
The high increase of false positive classifications with the new test set are believed to be due the fact that the samples used for training, which are taken randomly from all the non-ictal intervals, were distinct from the ones really close the seizure onset. This leads to believe, that there is a noticeable change in the signal while approaching the transition from the non-ictal to the ictal phase. However, by applying the MAF5 filter, the score increased. The MAF5 filter could potentially suit the real-time detection, since it only requires a delay of few samples. In the case of 5 second sliding windows, with 80% overlap, this delay would be of 2 seconds with the filter size 5, which fits early detection necessities.

After applying a moving average filter, with a stride of five, to reduce the number of false positive classifications, we noticed that for 18 out of all 22 seizures, within a rage of 1 minute, some false positive classifications persisted. An example is shown in figure 22. This is curious, since this false predictions could be very useful in a real time model, because they can be used as alarms, warning the patient of an imminent attack. Thus, we retrained the networks with augmented data, to see if this false predictions persist.

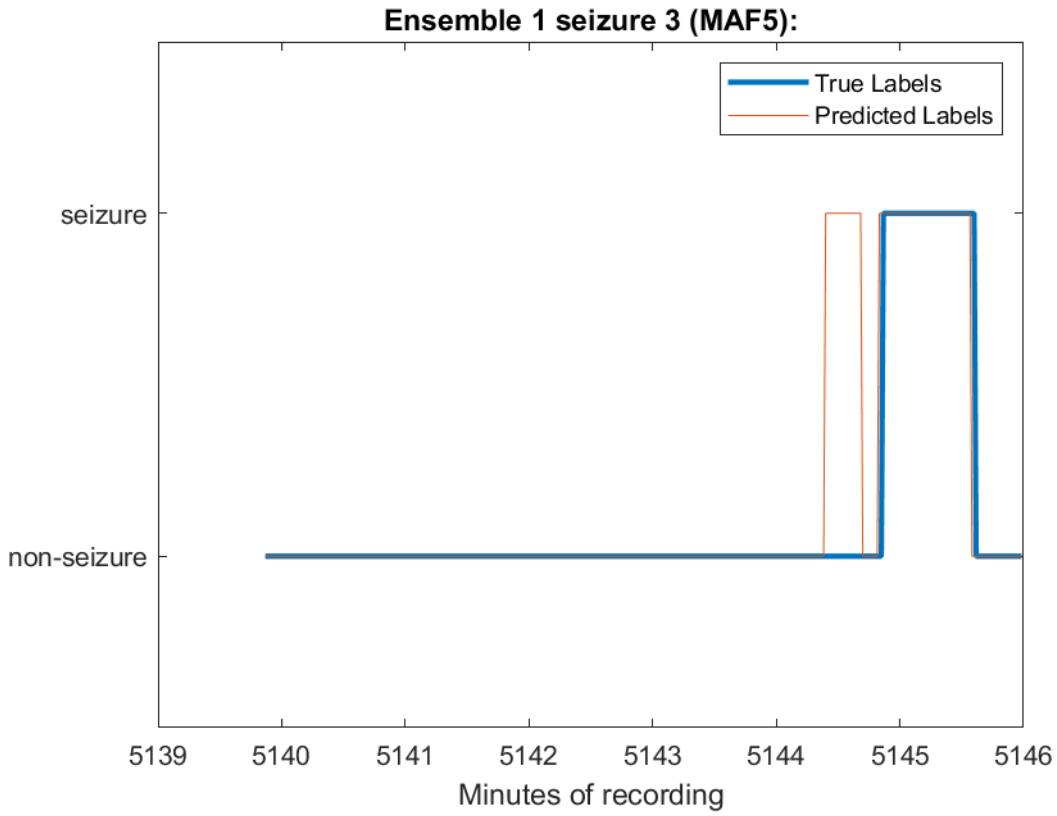


Figure 22: Figure showing the false predictions close to the seizure onset.

After testing the second ensemble on the dataset D, we could not identify the false alarms close to the seizure onset as in the previous testing. Moreover, only four seizures had a false alarm within five minutes before the seizure. We believe that this is a consequence of the training set C. Although samples from one minute before the onset were left out of training, with the intention of producing some false alarms, they were classified correctly in the test. An other evidence supporting our claim are the false predictions far from the seizure, that appear in most of the seizures. Hence, to get rid of them and produce the false predictions close to the seizure onset, the training set should include more samples that are far away from the seizure. However, this results confirm our initial statement, that is easier to discriminate non-ictal samples far from the seizure and seizure samples, than seizure samples and non-ictal samples close to them.

7 Future Perspective

The main drawback of this thesis is that it is performed with only one patient, therefore there are no other comparable results. As mentioned earlier, to reproduce the method on more patients could give more detailed explanation of the results.
An interesting aspect are the 3-D topographic maps. Nowadays deep learning methods are very powerful and efficiently deal with very complex data. By feeding the classifier with simple 2-D images of different sub-band, it will extract features, that only hold information about the current state of the brain. However, no information about the dynamics of the brain will be used in the decision process. Since the transition from non-ictal to ictal is generally gradual, the information about the dynamics may be useful for a more efficient classification. Therefore, using 3-D topographic maps, constructed as a time sequence of topographic maps, may give even more information than conventional 2-D maps and thus produce a higher classification score.
An alternative use of 3-D topographic maps could be on the frequency domain. Similarly as performed by Zgang et al. [62], topographic maps of different frequency sub-bands could be stacked in an array, before being fed to the CNN. At this point, the ensemble will not be further needed. Such approach would significantly reduce the training time, since there would only be one network used instead of five. However, the classification score might decrease, because the classification will depend only on the score of a single network rather than the majority of five different networks. Perhaps it could be interesting to compare of the two models.
An other interesting improvement could be limiting the number of electrodes. In this study 21 electrodes were used. Ideally, if the model would be so effective to be used at the clinical stage, patients would wear EEG caps to monitor their brain activity throughout the day. Therefore, limiting the number of electrodes will increase patient's comfort. Moreover, having less electrodes also means less data that need to be processed. Hence, it would be worth trying.

8 Conclusion

In this thesis we present a model for seizure detection, using topographic maps of the relative powers of different frequency sub-bands. To classify them we used an ensemble of five convolutional neural networks, one for each sub-band. The model, trained with samples randomly picked from the non-seizure interval, obtained good results on the initial testing set, however it produced a high number of false positive classifications when tested on an interval close to the seizure onset. We managed to drop such errors by applying a moving average filter, obtaining 96.68% accuracy, 92.56% sensitivity and 97.23% specificity. Furthermore, after applying the filter, we noticed that the false positive classifications were located close to the seizure onset.

We tried to obtain again similar results, by training the networks with considerably more samples taken between five and one minute before the seizure. Although the classification scores increased, we lost the false positive predictions close to the seizure onset. Perhaps, to keep them, the ensemble should have been trained with samples, that precede the seizure for more than five minutes. However, here limitations might occur due to the computational overhead that so many samples can produce.

Overall, the model obtained scores that are close to the state of the art results. Although this model does not have an early prediction performance, it still yields good detection scores. Furthermore, this thesis gives some insights on the early detection, that might be possible to perform, due to the diversity of the non-ictal samples, located far and close to the seizure.

9 Povzetek naloge v slovenskem jeziku

Proti-epileptična zdravila so neučinkovita na kar tretjini epileptičnih bolnikov. Gradnja aparatov, ki samodejno detektirajo nemuden epileptičen napad na podlagi EEG signalov, predstavlja alternativo za izboljšanje kvalitete vsakodnevnega življenja takih pacijentov. Sledeča teza je poskus metode automatične detekcije, ki sloni na topogrfskih mapah, izvlečenih iz signalov različnih frekvenčnih pasov, in globokega učenja.
Postopek detekcije epileptičnih napadov lahko strnemo na sedem glavnih korakov: pridobitev podatkov, obdelava podatkov, pridobivanje atributov, selekcija atributov, klasifikacija, urejanje in ocenitev. Vsekakor so v tem poteku možne spremembe, saj različne klasifikacijske metode potrebujejo različno obdelavo podatkov in različen izbor atributov. V zaključni nalogi smo uporabili klasifikacijske metode globokega učenja, ki so znane za to, da dobro delujejo s surovimi podatki in z osnovnimi atributi. Torej smo te korake izpustili in se osredotočili na sestavo topografskih preslikav.
Podatke smo pridobili iz base podatkov EPILEPSIAE. Signale smo morali obdelati, saj so EEG podatki nestacionarne narave. En postopek, da se jim umetno vsili stacionarnost, je da se jih spremeni v kratke zaporedne časovne intervale, imenovane drseča okna. V nalogi smo v dveh različnih poskusih uporabili okna dolžine petih sekund, enkrat z 80% prikrivanjem, drugič z 98%. Signale smo nato s diskretno Fourierjevo transformacijo spremenili iz čaovne domene v frekvenčno. Odstranjene so bile frekvence pod 0.1 Hz in nad 128 Hz (Nyquistjeva frekvenca) ter frekvenca 50 Hz, ker je občutljiva na motnje električne napeljave. Nato smo celoten signal razdelili na frekvenčne pasove $\delta(0.5-3.5Hz), \theta(4-7.5Hz), \alpha(8-12Hz), \beta(13-35Hz), \gamma(35-60Hz)$, ter izračunali njihovo relativno moč. Iz teh smo nato sestavili topografske preslikave oz. mape.
Klasifikcijski model je bil sestavljen iz petih globokih konvolucijskih nevronskih mrež, ena za vsak frekvenčni pas. Končno oznako za vsak vzorec model izbere glede na večino oznak iz posameznih mrež. Zaključno nalogo smo razdelili na dva dela: prvo in drugo testiranje.
Najprej smo iz surovih podatkov izvlekli vzorce v obliki drsečih oken dolžine petih sekund in z 80% prikrivanjem. Na tak način smo pridobili 929 vzorcev za vsako oznako. Vzorci, ki niso pripadali napadom, so bili izbrani naključno. Te smo razdeli v

podmnožice, in sicer 80% smo uporabili za uÄenje, 10% za preverjenje med uÄenjem (validacijo) ter 10% za testiranje. Prvo uÄenje je potekalo sočasno z iskanjem najboljših parametrov za posamezne mreže. Rezultati testiranja so prikazani v tabeli 8. Isti skupek mrež smo nato poskusili testirati na intervalu zadnjih pet minut pred vsakim napadom. Zaradi povečanja števila lažno pozitivnih napovedi smo vrsto vzorcev uredili s filtrom drsečne sredine. Primejrava obeh rezultatov je prikazana v sliki 20 ter v tabeli 9. Upoštevati je treba dejstvo, da je večji del uÄnih vzorcev za napade enak tistim, ki smo jih uporabili za testiranje, saj to močno vpliva na njihovo pravilno klasifikacijo. Poleg tega smo opazili, da se na 18 izmed 21 napadov, tik pred začetkom napada pojavi zaporedje napačnih klasifikacij. Take klasifikacije bi bile zelo uporabne, saj bi pomenilo, da je model zmožen krajših napovedi epileptičnih napadov in ne le njihovih detekcij.

Zaradi pomankanja vzorcev smo ponovno izvlekli nove podatke, toda tokrat so se drseča okna prekrivala za 98%, tako da smo dobili 10-krat več podatkov. Mreže smo tokrat uÄili z vzorci, ki so se nahajali med petimi in eno minuto pred napadom ter z nekim deležem nakjučno izbranim, da bi uravnotežili moč množic obeh oznak. Skupek mrež smo nato testirali na intervalu ene ure pred napadom. Rezultati so prikazani v sliki 21 in tabeli 10. Toda, napačnih klafikacij tik pred začetkom napada nismo zaznali. Vejretno so si bili vzorci zadnje minute in vzorci, uporabljeni za uÄenje, preveč podobni. Grafi posameznih napadov za oba testa so prikazani v prilogah A in B.

Če povzamemo, model je dosegel 99.20% klasifikacijsko točnost, 96.48% občutljivost in 99.27% specifičnost pri klasifikaciji 21 napadov enega pacijenta, kar se približuje rezultatom ostalih raziskav v stroki. Poleg tega, ko smo trenirali mreže z vzorci, ki so bili izbrani naključno iz intervalov, ki niso bili pod epileptičnim napadom, smo se pri 18 napadih spopadli z napačno pozitivnimi oznaki v neposredni bližini epileptičnega napada. Take napačne klasifikacije predstavljajo morebitno napovedovalno moč modela, ki pa se niso pojavile, ko so bili vzorci vzeti v času zadnjih pet minut pred napadom. Nadaljni cilj bi bil skušati obdržati take opomine za vsak napad, kar pa zajema veliko časovne in računske zahtevnosti.

10 Bibliography

[1] A. A.E. K. Abdulhamit Subasi M. Kemal Kiymika, "Neural Network Classification of EEG Signals by Using AR with MLE Preprocessing for Epileptic Seizure Detection," *Computational Applications*, vol. 10, no. 1, pp. 57–70, 2005.

[2] S. A. Alshebeili, F. E. Abd El-Samie, T. Alshawi, T. N. Alotaiby, and I. Ahmad, "EEG seizure detection and prediction algorithms: a survey," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, 2015. DOI: 10.1186/1687-6180-2014-183.

[3] S. Ashok and G. Purushothaman, "Detection of epileptic seizure from electroencephalogram signals based on feature ranking and best feature subset using mutual information estimation," *Journal of Medical Imaging and Health Informatics*, vol. 6, pp. 1850–1864, Dec. 2016. DOI: 10.1166/jmihi.2016.1938.

[4] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, "Epileptic seizure prediction using relative spectral power features," *Clinical Neurophysiology*, vol. 126, no. 2, pp. 237–248, 2015, ISSN: 18728952. DOI: 10.1016/j.clinph.2014.05.022. [Online]. Available: http://dx.doi.org/10.1016/j.clinph.2014.05.022.

[5] 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.

[6] K. P. Bennett, "Support Vector Machines: Hype or Hallelujah?" *SIGKDD Explorations*, vol. 2, no. 2, Dec. 2000.

[7] E. Bou, D. K. Nguyen, S. Rihana, and M. Sawan, "Biomedical Signal Processing and Control Towards accurate prediction of epileptic seizures : A review," *Biomedical Signal Processing and Control*, vol. 34, pp. 144–157, 2017, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2017.02.001. [Online]. Available: http://dx.doi.org/10.1016/j.bspc.2017.02.001.

[8] M. L. Brian S. Everitt Sabine Landau and D. Stahl, "Miscellaneous clustering methods," in *Cluster Analysis*, L. John Wiley Sons, Ed., 2011, ch. 8, pp. 215–255, ISBN: 9780470749913.

[9] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018, ISSN: 18728286. DOI: 10.1016/j.neucom.2017.11.077.

[10] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *CoRR*, vol. abs/1405.3531, 2014. arXiv: 1405.3531. [Online]. Available: http://arxiv.org/abs/1405.3531.

[11] H. Cruse, "Neural networks as cybernetic systems - part i," *Brains, Minds Media*, Jan. 2006.

[12] B. Direito, C. Teixeira, B. Ribeiro, M. Castelo-branco, and F. Sales, "Modeling epileptic brain states using EEG spectral analysis and topographic mapping," *Journal of Neuroscience Methods*, vol. 210, no. 2, pp. 220–229, 2012, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2012.07.006. [Online]. Available: http://dx.doi.org/10.1016/j.jneumeth.2012.07.006.

[13] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Computer Methods and Programs in Biomedicine*, vol. 161, no. April, pp. 1–13, 2018, ISSN: 18727565. DOI: 10.1016/j.cmpb.2018.04.005.

[14] A. H. Fielding, *Cluster and classification techniques for the biosciences*. Cambridge University Press, 2006. DOI: 10.1017/CBO9780511607493.

[15] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. E. Jr, L. Forsgren, J. A. French, D. C. Hesdorffer, B. I. Lee, G. W. Mathern, S. L. Mosh, M. Watanabe, E. Perucca, I. E. Scheffer, and S. Wiebe, "A practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475–482, 2014. DOI: 10.1111/epi.12550.

[16] T. Gorach, "Deep Convolutional Neural Networks- a Review," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 7, pp. 439–452, Jul. 2018.

[17] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm networks," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, vol. 4, 2005, 2047–2052 vol. 4. DOI: 10.1109/IJCNN.2005.1556215.

[18] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, no. March, pp. 1157–1182, 2011, ISSN: 00032670. DOI: 10.1016/j.aca.2011.07.027. arXiv: 1111.6189v1.

[19] P. Hall and B. U. Park, "Choice of neighbor order in nearest-neighbor classification," *The Annals of Statistics*, vol. 36, no. 5, I. of Mathematical Statistics, Ed., pp. 2135–2152, 2008. DOI: 10.1214/07-AOS537. arXiv: arXiv:0810.5276v1.

[20] R. Harikumar and M. Sunil Kumar, Prabhakar Manjusha, "Comparison of Support Vector Machine With Particle Swarm Optimization Technique for Eepilepsy Classification from EEG," *International Journal*, vol. 2, no. 1, pp. 421–428, 2016, ISSN: 0157-6895. DOI: 10.1007/s00249-003-0278-y. [Online]. Available: http://www.doaj.org/doaj?func=abstract{\&}id=946163.

[21] N. I. of Health, *Malignant migrating partial seizures of infancy*, https://ghr.nlm.nih.gov/art/large/brain-activity-partial-generalized-seizures.jpeg, Jun. 2019.

[22] M. Heyden, *Classification of EEG data using machine learning techniques*, Department of Automatic Control, Lund University, Box 118, SE-221 00, LUND, Sweden, 2016.

[23] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1–32, 1997.

[24] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10-20 system electrode placement," *Electroencephalography and Clinical Neurophysiology*, vol. 66, no. 4, pp. 376–382, 1987, ISSN: 00134694. DOI: 10.1016/0013-4694(87)90206-9.

[25] M. S. Hossain, S. U. Amin, M. Alsulaiman, and G. Muhammad, "Applying Deep Learning for Epilepsy Seizure Detection and Brain Mapping Visualization," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 15, no. 1s, pp. 1–17, 2019, ISSN: 15516857. DOI: 10.1145/3241056.

[26] L. Hussain, W. Aziz, S. Saeed, A. Idris, I. A. Awan, S. A. Shah, M. S. A. Nadeem, and S. Rathore, "Spatial Wavelet-Based coherence and coupling in EEG signals with eye open and closed during resting state," *IEEE Access*, vol. 6, pp. 37 003–37 022, 2018, ISSN: 21693536. DOI: 10.1109/ACCESS.2018.2844303.

[27] S. Jacob and A. B. Nair, "An Updated Overview on Therapeutic Drug Monitoring of Recent Antiepileptic Drugs.," eng, *Drugs in R&D*, vol. 16, no. 4, pp. 303–316, 2016, ISSN: 1179-6901 (Electronic). DOI: 10.1007/s40268-016-0148-6.

[28] H. H. Jasper, "International federation of societies for electroencephalography and clinical neurophysiology," *Electroencephalography and Clinical Neurophysiology*, vol. 10, no. 2, 367â373, 1958. DOI: 10.1016/0013-4694(58)90051-8.

[29] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Machine Learning: ECML-98*, C. Nédellec and C. Rouveirol, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 137–142, ISBN: 978-3-540-69781-7.

[30] E. M. Kahn, R. D. Weiner, R. P. Brenner, and R. Coppola, "Topographic maps of brain electrical activity-pitfalls and precautions," *Biological Psychiatry*, vol. 23, no. 6, pp. 628–636, 1988, ISSN: 00063223. DOI: 10.1016/0006-3223(88)90009-1.

[31] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," *IEEE*, pp. 1942–1948, 1995.

[32] J. Klatt, H. Feldwisch-drentrup, M. Ihle, V. Navarro, M. Neufang, C. Teixeira, C. Adam, M. Valderrama, A. Witon, M. L. V. Quyen, F. Sales, A. Dourado, J. Timmer, A. Schulze-bonhage, and B. Schelter, "The EPILEPSIAE database : An extensive electroencephalography database of epilepsy patients," *Epilepsia*, vol. 53, no. 9, pp. 1669–1676, 2012. DOI: 10.1111/j.1528-1167.2012.03564.x.

[33] G. H. Klem, H. O. Luders, H. Jasper, and C. Elger, "The ten-twenty electrode system of the International Federation," *Electroencephalography and Clinical Neurophysiology*, vol. 10, no. 2, pp. 371–375, 1999, ISSN: 00134694. DOI: 10.1016/0013-4694(58)90053-1. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10590970.

[34] T. N. Lal, O. Chapelle, and J. Weston, "Embedded Methods," in *Feature Extraction: Foundations and Applications Studies in Fuzziness and Soft Computing*. Springer, 2006, vol. 165, ch. 5, pp. 137–165.

[35] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, ISSN: 14764687. DOI: 10.1038/nature14539.

[36] F. Lopes, W. Blanes, S. N. Kalitzin, J. Parra, P. Suffczynski, and D. N. Velis, "Epilepsies as Dynamical Diseases of Brain Systems : Basic Models of the Transition Between Normal and Epileptic Activity," *Epilepsia*, vol. 44, pp. 72–83, 2003.

[37] F Lotte, M Congedo, A Lecuyer, F Lamarche, and B Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces.," eng, *Journal of neural engineering*, vol. 4, no. 2, R1–R13, 2007, ISSN: 1741-2560 (Print). DOI: 10.1088/1741-2560/4/2/R01.

[38] S. Lou, *Modeling the brain with computers*, http://sitn.hms.harvard.edu/flash/2012/brain-modeling/, Jul. 2012.

[39] M. Mitchell, *Genetic Algorithms : An Overview*. MIT Press, 1995, pp. 1–17.

[40] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: The long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2007, ISSN: 14602156. DOI: 10.1093/brain/awl241.

[41] A. Moujahid, *A practical introduction to deep learning with caffe and python*, http://adilmoujahid.com/posts/2016/06/introduction-deep-learning-python-caffe, Jun. 2016.

[42] W. H. Organization, *Atlas : Epilepsy care in the world*. World Health Organization, 1970. [Online]. Available: http://www.who.int/iris/handle/10665/43298.

[43] Y. Park, L. Luo, K. K. Parhi, and T. Netoff, "Seizure prediction with spectral power of EEG using cost-sensitive support vector machines," *Epilepsia*, vol. 52, no. 10, pp. 1761–1770, 2011. DOI: 10.1111/j.1528-1167.2011.03138.x.

[44] W Penfield and T. C. Erickson, *Epilepsy and cerebral localization*. Oxford, England: Charles C. Thomas, 1941, p. 623.

[45] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989, ISSN: 0018-9219. DOI: 10.1109/5.18626.

[46] S. Raschka, *Naive bayes and text classification*, https://sebastianraschka.com/Articles/2014_naive_bayes_1.html, Oct. 2014.

[47] J. Saini and M. Dutta, "An extensive review on development of EEG-based computer-aided diagnosis systems for epilepsy detection," *Network: Computation in Neural Systems*, vol. 28, no. 1, pp. 1–27, 2017, ISSN: 13616536. DOI: 10.1080/0954898X.2017.1325527. [Online]. Available: https://doi.org/10.1080/0954898X.2017.1325527.

[48] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015, ISSN: 18792782. DOI: 10.1016/j.neunet.2014.09.003. arXiv: arXiv:1404.7828. [Online]. Available: http://dx.doi.org/10.1016/j.neunet.2014.09.003.

[49] B. Schölkopf, K.-k. Sung, C. J. C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2758–2765, 1997.

[50] A. Schulze-Bonhage, A. Witon, M. Ihle, C. A. Teixeira, B. Schelter, H. Feldwisch-Drentrup, and J. Timmer, "EPILEPSIAE A European epilepsy database," *Computer Methods and Programs in Biomedicine*, vol. 106, no. 3, pp. 127–138, 2010, ISSN: 01692607. DOI: 10.1016/j.cmpb.2010.08.011. [Online]. Available: http://dx.doi.org/10.1016/j.cmpb.2010.08.011.

[51] I. Silva, *Eegplot()*, https://www.mathworks.com/matlabcentral/fileexchange/3279-eegplot, May 2013.

[52] S. Solorio-fernández and J. A.C.-o.J. F. Martínez-trinidad, "A review of unsupervised feature selection methods," *Artificial Intelligence Review*, 2019, ISSN: 1573-7462. DOI: 10.1007/s10462-019-09682-y. [Online]. Available: https://doi.org/10.1007/s10462-019-09682-y.

[53] C. A. Teixeira, B Direito, H Feldwisch-drentrup, M Valderrama, and R. P. Costa, "EPILAB : A software package for studies on the prediction of epileptic seizures," *Journal of Neuroscience Methods*, vol. 200, no. 2, pp. 257–271, 2011, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2011.07.002. [Online]. Available: http://dx.doi.org/10.1016/j.jneumeth.2011.07.002.

[54] C. A. Teixeira, B. Direito, M. Bandarabadi, and A. Dourado, "Output regularization of SVM seizure predictors : Kalman Filter versus the â Firing Power â method," in *34th Annual International Conference of the IEEE EMBS*, San Diego, California USA: IEEE, 2012, pp. 6530–6533, ISBN: 9781457717871.

[55] M. Teplan, "Teplan_review2002_Fundamentals of EEG measurement," *Measurment Science Review*, vol. 2, no. 2, pp. 1–11, 2002, ISSN: 15353893. DOI: 10.1021/pr0703501.

[56] "The ten twenty electrode system: International federation of societies for electroencephalography and clinical neurophysiology," *American Journal of EEG Technology*, vol. 1, no. 1, 13â19, 1961. DOI: 10.1080/00029238.1961.11080571.

[57] P. Thodoroff, J. Pineau, and A. Lim, "Learning Robust Features using Deep Learning for Automatic Seizure Detection," *Proceedings of Machine Learning and Healthcare*, vol. 56, 2016.

[58] V. L. Towle, J. Bolafios, D. Suarez, K. Tan, R. Grzeszczuk, D. N. Levin, R. Cakmur, S. A. Frank, and J.-p. Spire, "The spatial location of EEG electrodes : locating the best-fitting sphere relative to cortical anatomy," *Electroencephalography and Clinical Neurophysiology*, vol. 86, no. 1, pp. 1–6, 1993.

[59] A. Varsavsky, I. Mareels, and M. Cook, *Epileptic seizures and the eeg*. Jan. 2011, ISBN: 978-1-4398-1200-6-.

[60] N. S. V.Srinivasan C.Eswaran, "Epileptic detection using artificial neural," *International Conference on Signal Processing and Communication*, pp. 340–343, 2004.

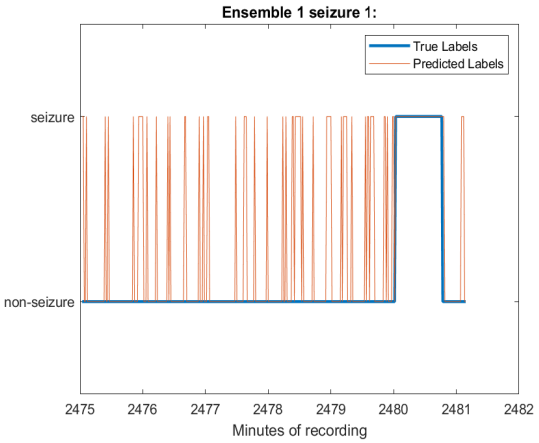
[61] S. Wong, A. B. Gardner, A. M. Krieger, and B. Litt, "A Stochastic Framework for Evaluating Seizure Prediction Algorithms Using Hidden Markov Models," *J Neurophysiol.*, vol. 97, no. 3, pp. 2525–2532, 2008.

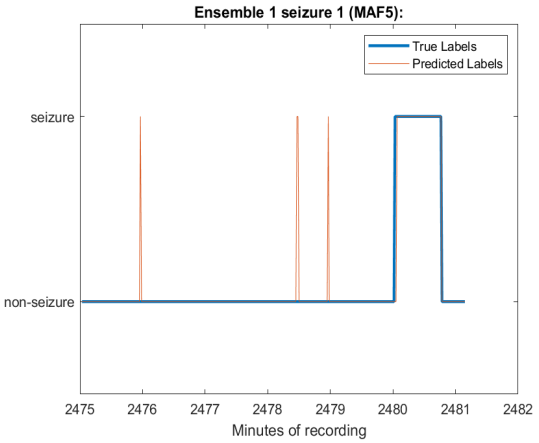
[62] P. Zhang, X. Wang, W. Zhang, and J. Chen, "Learning Spatial-Spectral-Temporal EEG Features With Recurrent 3D Convolutional Neural Networks for Cross-Task Mental Workload Assessment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 1, pp. 31–42, 2019, ISSN: 15344320. DOI: 10.1109/TNSRE.2018.2884641.

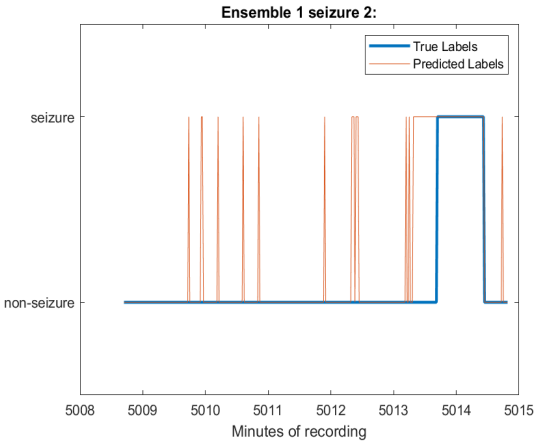
Appendices

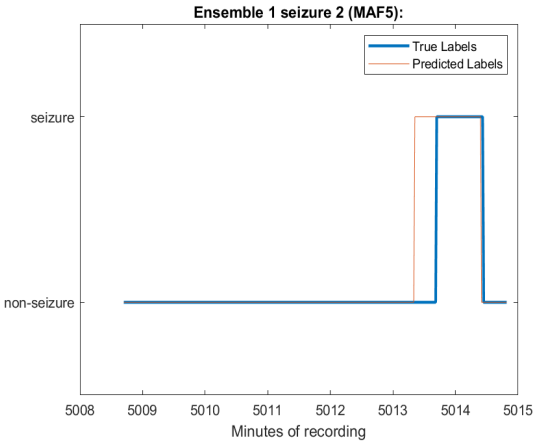
A Plots of the Seizures: dataset B

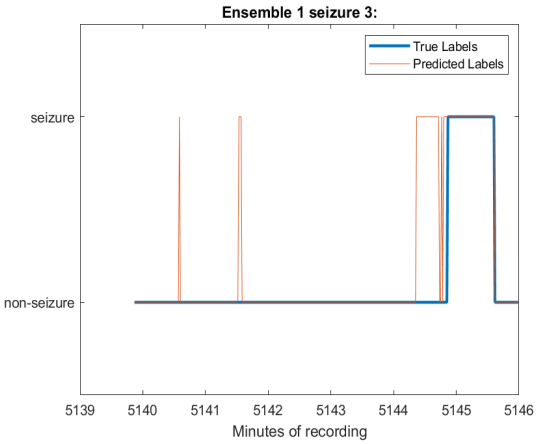
Here are presented the classifications for all the seizures in the B dataset.

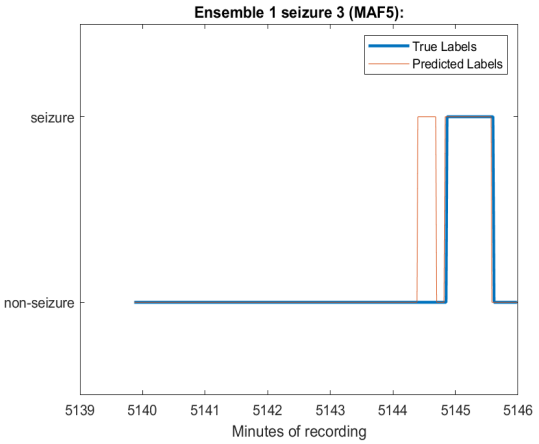

Ensemble 1 seizure 1:
True Labels
Predicted Labels
seizure
non-seizure
2475
2476
2477
2478
2479
2480
2481
2482
Minutes of recording

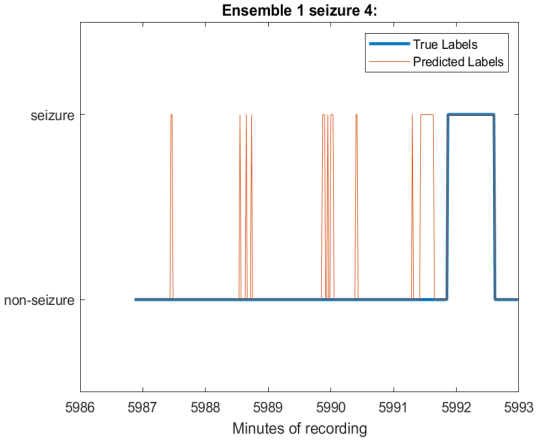

Ensemble 1 seizure 1 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
2475
2476
2477
2478
2479
2480
2481
2482
Minutes of recording

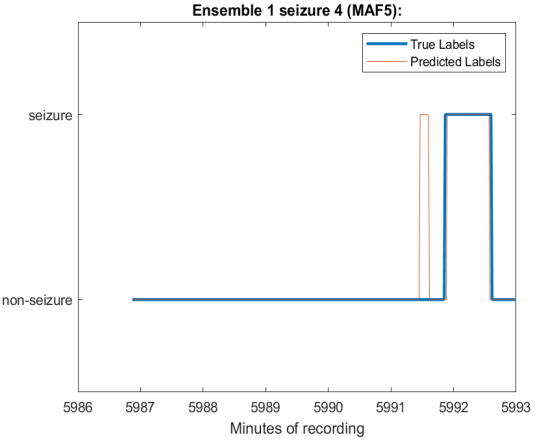

Ensemble 1 seizure 2:
True Labels
Predicted Labels
seizure
non-seizure
5008
5009
5010
5011
5012
5013
5014
5015
Minutes of recording

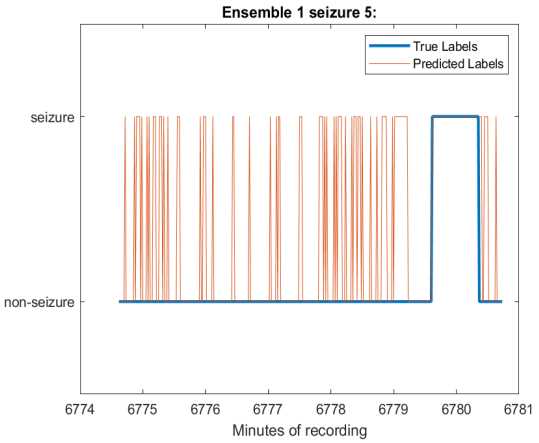

Ensemble 1 seizure 2 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
5008
5009
5010
5011
5012
5013
5014
5015
Minutes of recording

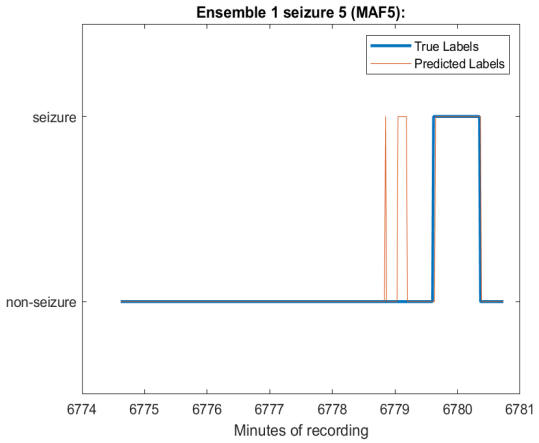

Ensemble 1 seizure 3:
True Labels
Predicted Labels
seizure
non-seizure
5139
5140
5141
5142
5143
5144
5145
5146
Minutes of recording

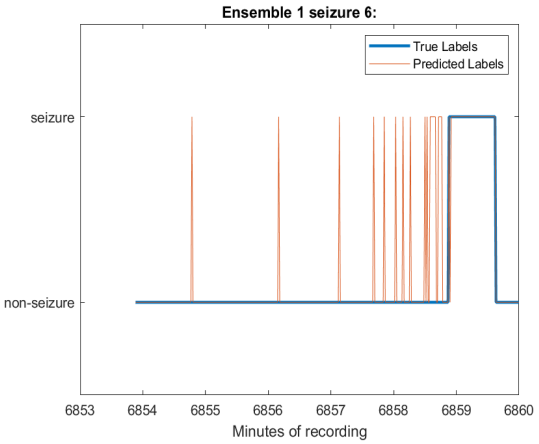

Ensemble 1 seizure 3 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
5139
5140
5141
5142
5143
5144
5145
5146
Minutes of recording

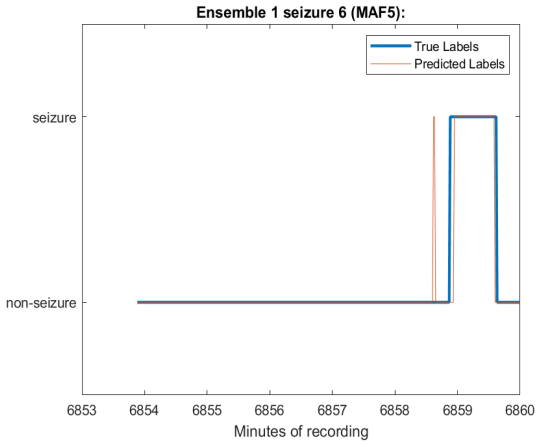

Ensemble 1 seizure 4:
True Labels
Predicted Labels
seizure
non-seizure
5986
5987
5988
5989
5990
5991
5992
5993
Minutes of recording

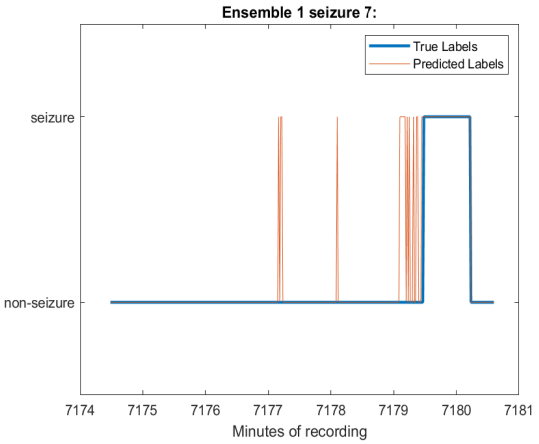

Ensemble 1 seizure 4 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
5986
5987
5988
5989
5990
5991
5992
5993
Minutes of recording

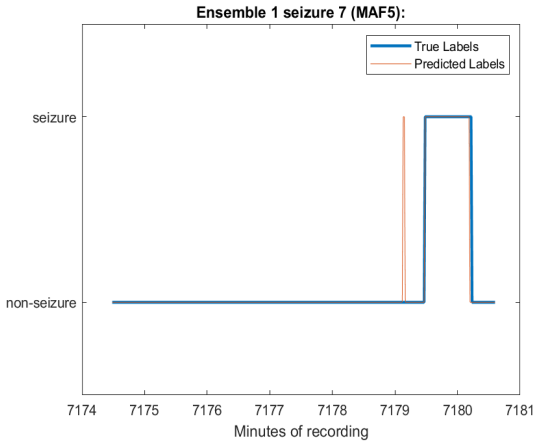

Ensemble 1 seizure 5:
True Labels
Predicted Labels
seizure
non-seizure
6774
6775
6776
6777
6778
6779
6780
6781
Minutes of recording

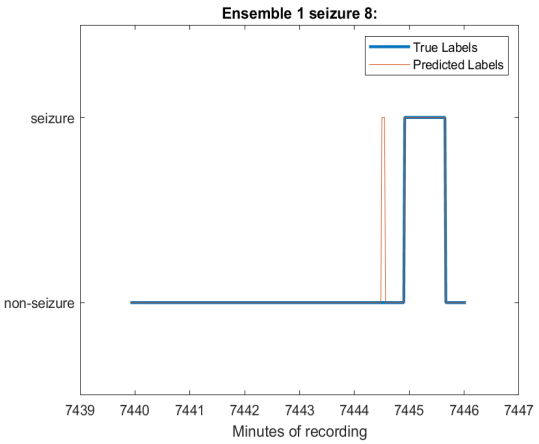

Ensemble 1 seizure 5 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
6774
6775
6776
6777
6778
6779
6780
6781
Minutes of recording

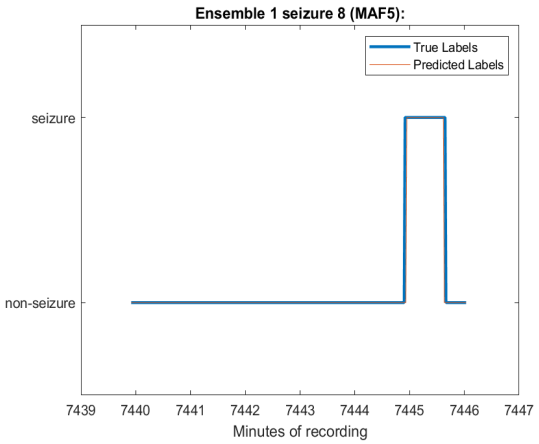

Ensemble 1 seizure 6:
True Labels
Predicted Labels
seizure
non-seizure
6853
6854
6855
6856
6857
6858
6859
6860
Minutes of recording

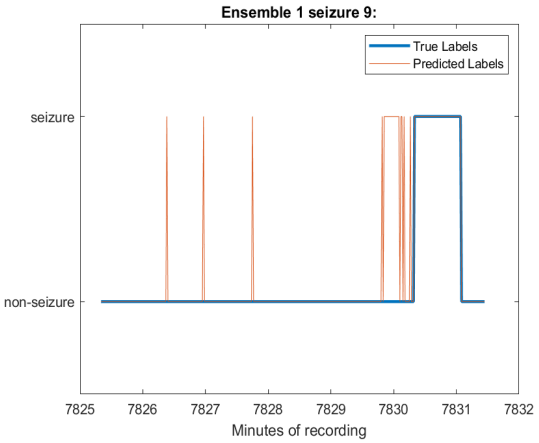

Ensemble 1 seizure 6 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
6853
6854
6855
6856
6857
6858
6859
6860
Minutes of recording

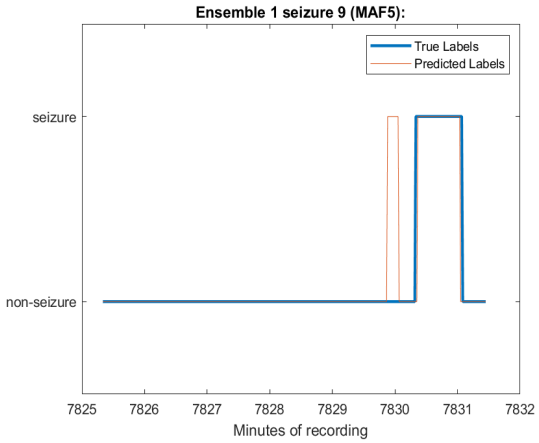

Ensemble 1 seizure 7:
True Labels
Predicted Labels
seizure
non-seizure
7174
7175
7176
7177
7178
7179
7180
7181
Minutes of recording

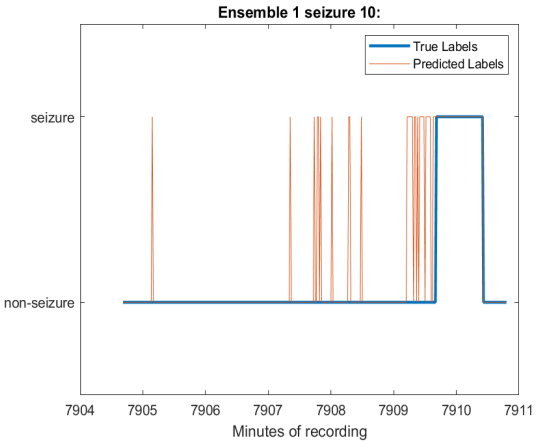

Ensemble 1 seizure 7 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
7174
7175
7176
7177
7178
7179
7180
7181
Minutes of recording

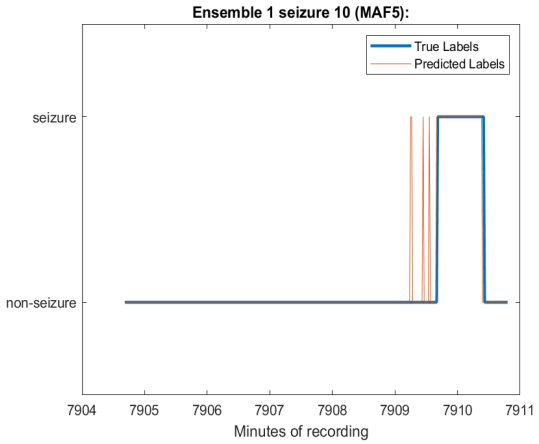

Ensemble 1 seizure 8:
True Labels
Predicted Labels
seizure
non-seizure
7439
7440
7441
7442
7443
7444
7445
7446
7447
Minutes of recording

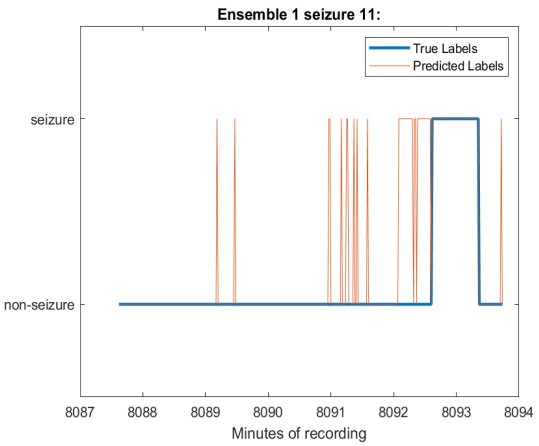

Ensemble 1 seizure 8 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
7439
7440
7441
7442
7443
7444
7445
7446
7447
Minutes of recording

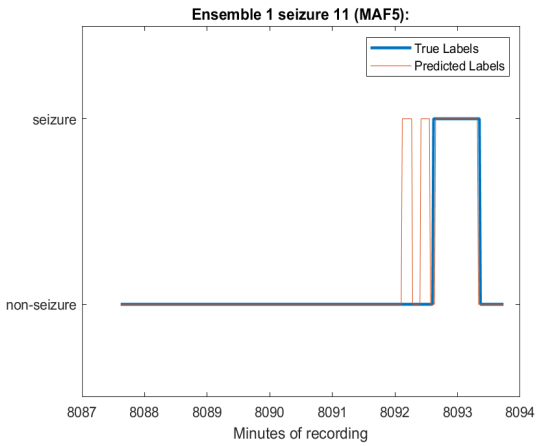

Ensemble 1 seizure 9:
True Labels
Predicted Labels
seizure
non-seizure
7825 7826 7827 7828 7829 7830 7831 7832
Minutes of recording

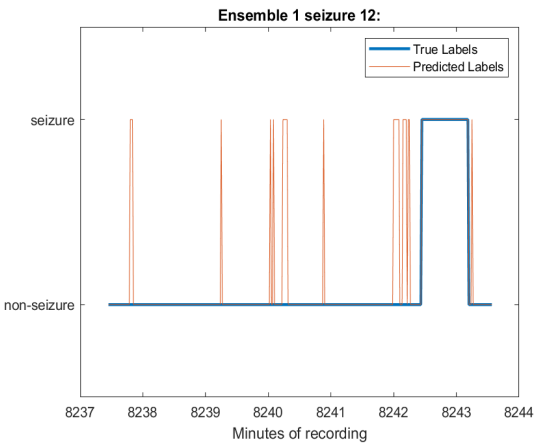

Ensemble 1 seizure 9 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
7825 7826 7827 7828 7829 7830 7831 7832
Minutes of recording

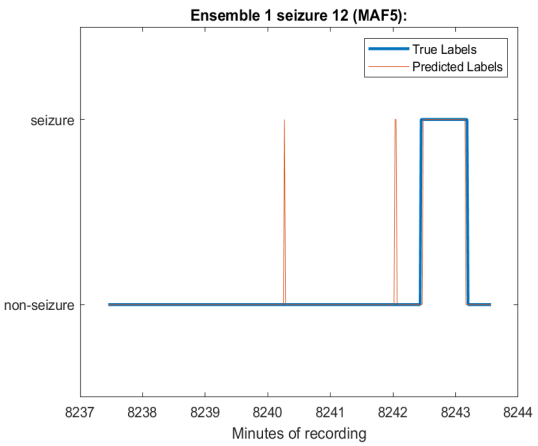

Ensemble 1 seizure 10:
True Labels
Predicted Labels
seizure
non-seizure
7904 7905 7906 7907 7908 7909 7910 7911
Minutes of recording

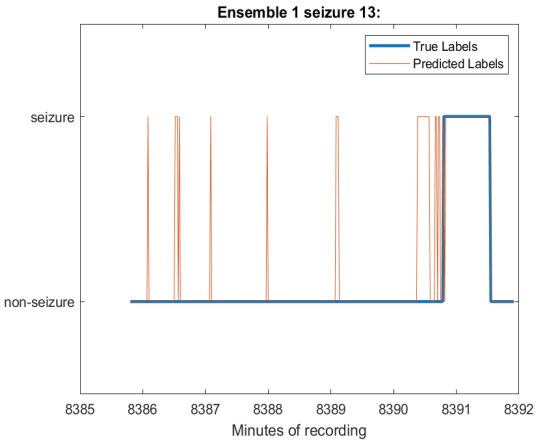

Ensemble 1 seizure 10 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
7904 7905 7906 7907 7908 7909 7910 7911
Minutes of recording

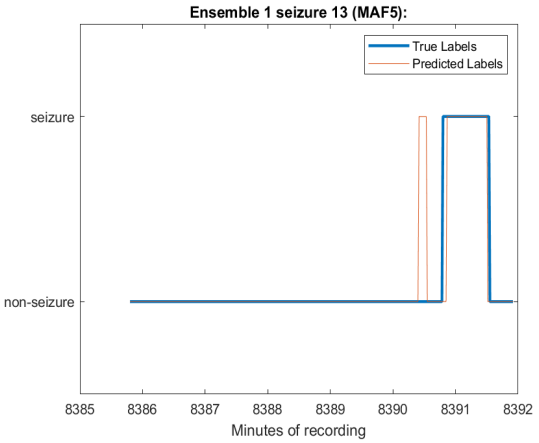

Ensemble 1 seizure 11:
True Labels
Predicted Labels
seizure
non-seizure
8087 8088 8089 8090 8091 8092 8093 8094
Minutes of recording

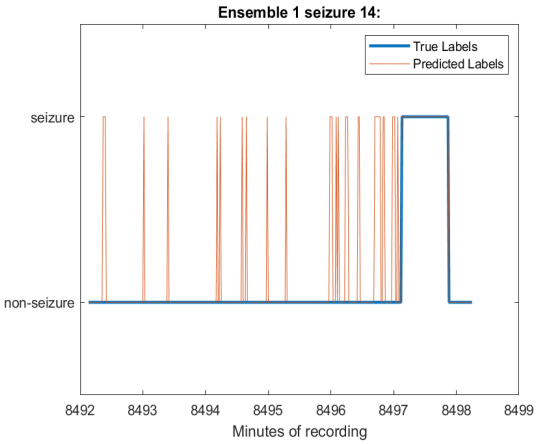

Ensemble 1 seizure 11 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
8087 8088 8089 8090 8091 8092 8093 8094
Minutes of recording

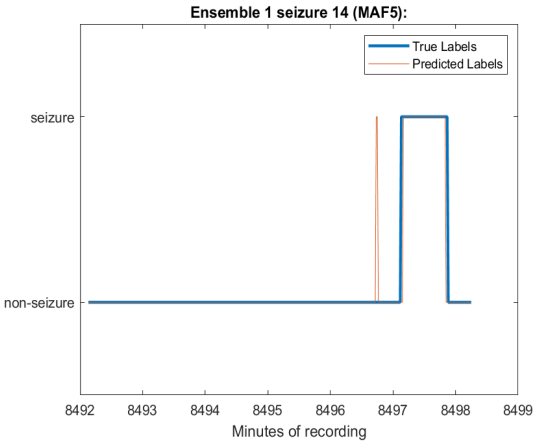

Ensemble 1 seizure 12:
True Labels
Predicted Labels
seizure
non-seizure
8237 8238 8239 8240 8241 8242 8243 8244
Minutes of recording

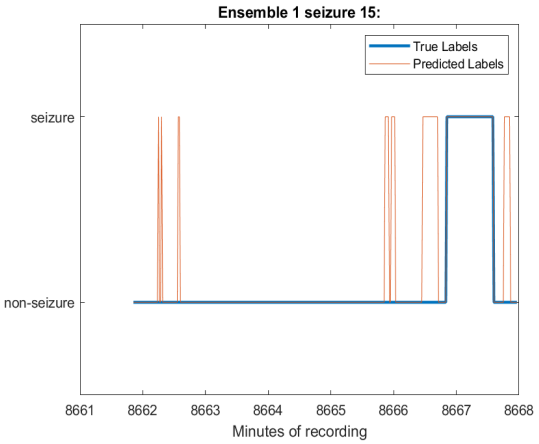

Ensemble 1 seizure 12 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
8237 8238 8239 8240 8241 8242 8243 8244
Minutes of recording

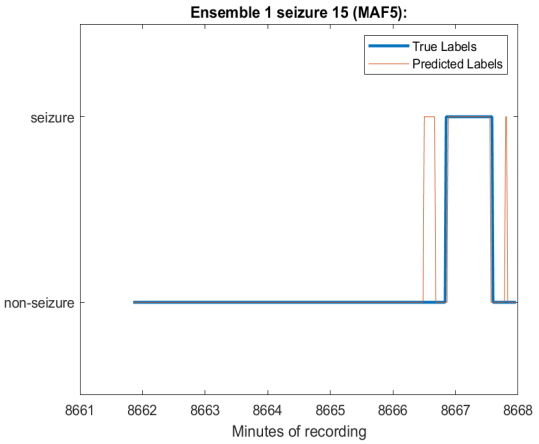

Ensemble 1 seizure 13:
True Labels
Predicted Labels
seizure
non-seizure
8385 8386 8387 8388 8389 8390 8391 8392
Minutes of recording

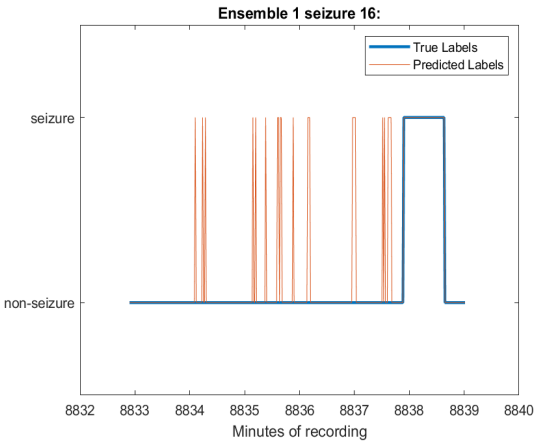

Ensemble 1 seizure 13 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
8385 8386 8387 8388 8389 8390 8391 8392
Minutes of recording

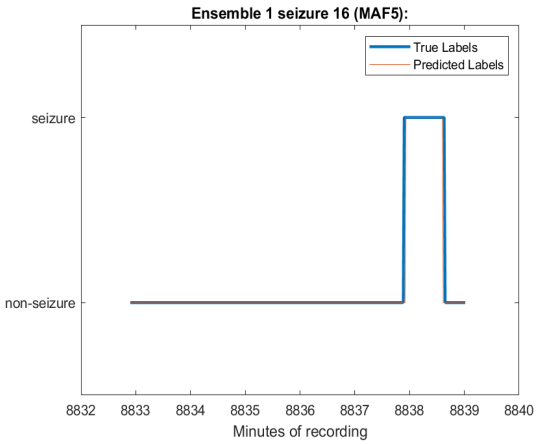

Ensemble 1 seizure 14:
True Labels
Predicted Labels
seizure
non-seizure
8492 8493 8494 8495 8496 8497 8498 8499
Minutes of recording

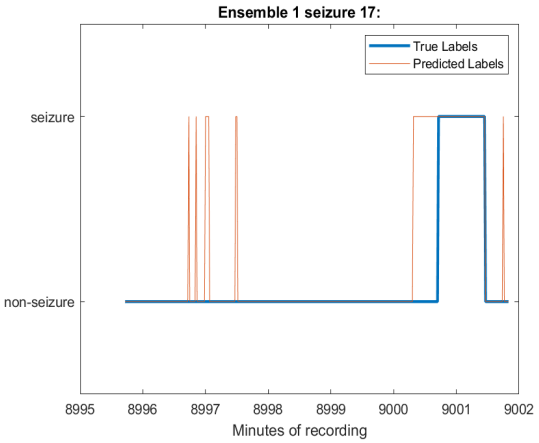

Ensemble 1 seizure 14 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
8492 8493 8494 8495 8496 8497 8498 8499
Minutes of recording

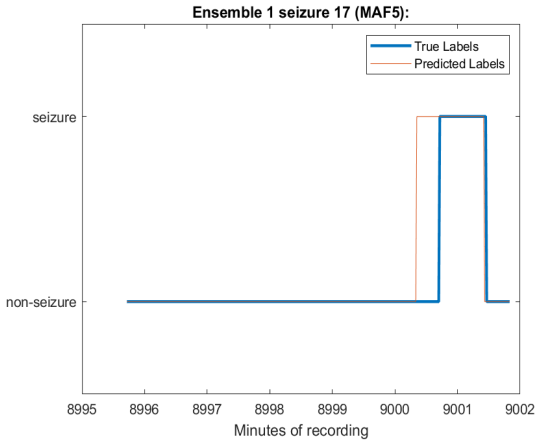

Ensemble 1 seizure 15:
True Labels
Predicted Labels
seizure
non-seizure
8661 8662 8663 8664 8665 8666 8667 8668
Minutes of recording

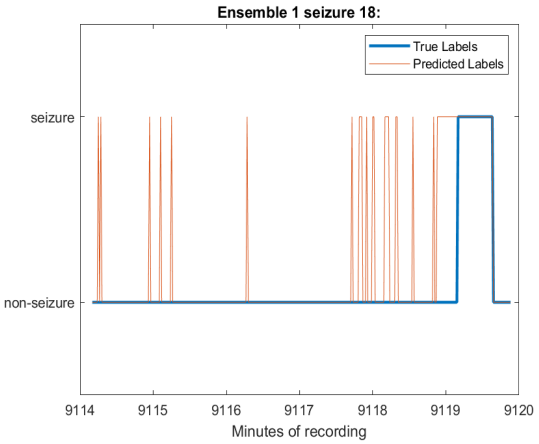

Ensemble 1 seizure 15 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
8661 8662 8663 8664 8665 8666 8667 8668
Minutes of recording

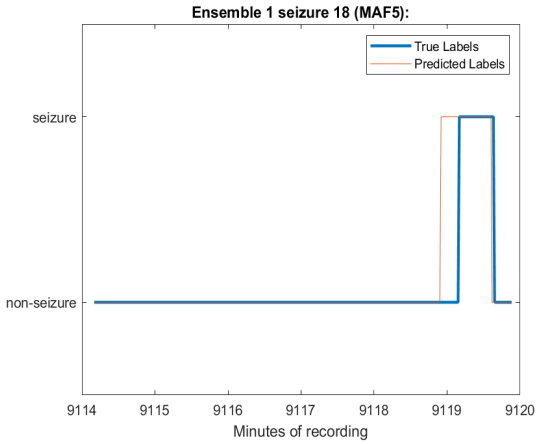

Ensemble 1 seizure 16:
True Labels
Predicted Labels
seizure
non-seizure
8832 8833 8834 8835 8836 8837 8838 8839 8840
Minutes of recording

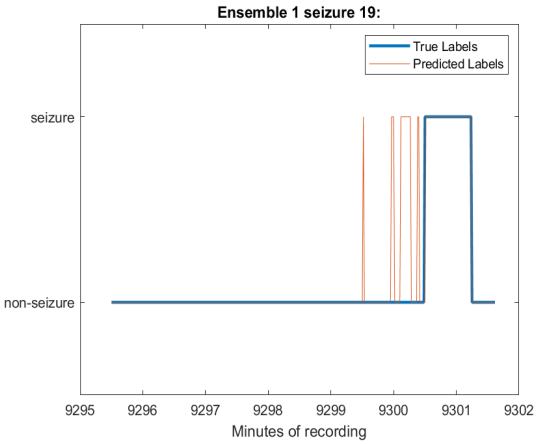

Ensemble 1 seizure 16 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
8832 8833 8834 8835 8836 8837 8838 8839 8840
Minutes of recording

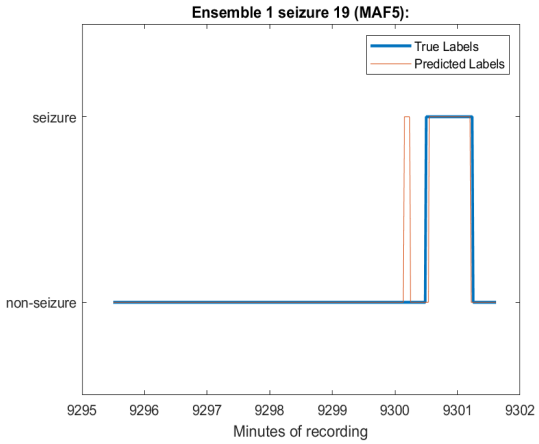

Ensemble 1 seizure 17:
True Labels
Predicted Labels
seizure
non-seizure
8995
8996
8997
8998
8999
9000
9001
9002
Minutes of recording

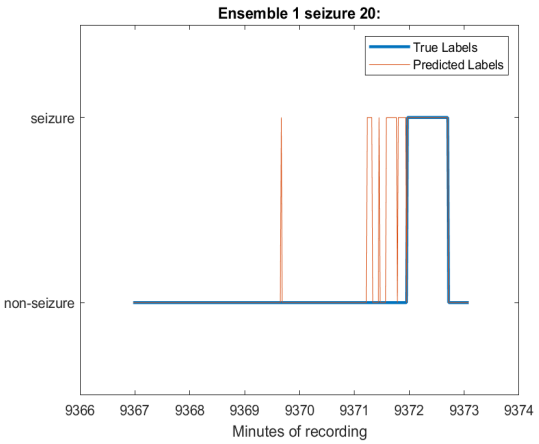

Ensemble 1 seizure 17 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
8995
8996
8997
8998
8999
9000
9001
9002
Minutes of recording

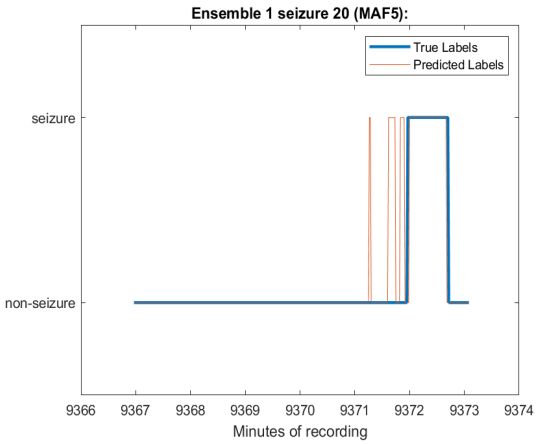

Ensemble 1 seizure 18:
True Labels
Predicted Labels
seizure
non-seizure
9114
9115
9116
9117
9118
9119
9120
Minutes of recording

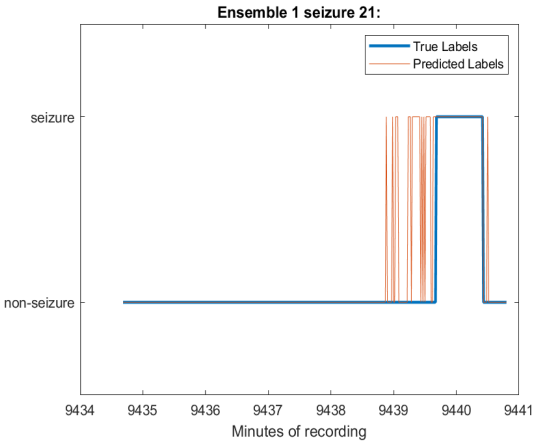

Ensemble 1 seizure 18 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
9114
9115
9116
9117
9118
9119
9120
Minutes of recording

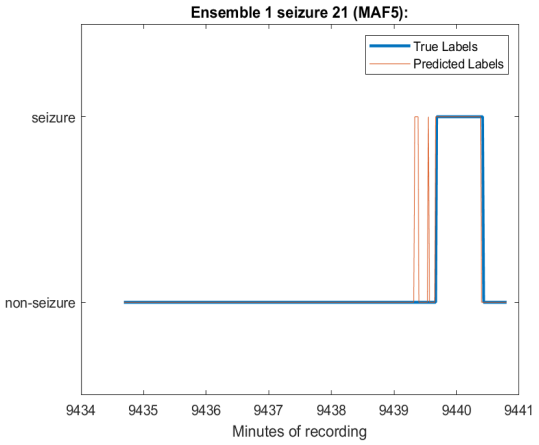

Ensemble 1 seizure 19:
True Labels
Predicted Labels
seizure
non-seizure
9295
9296
9297
9298
9299
9300
9301
9302
Minutes of recording


Ensemble 1 seizure 19 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
9295
9296
9297
9298
9299
9300
9301
9302
Minutes of recording


Ensemble 1 seizure 20:
True Labels
Predicted Labels
seizure
non-seizure
9366
9367
9368
9369
9370
9371
9372
9373
9374
Minutes of recording

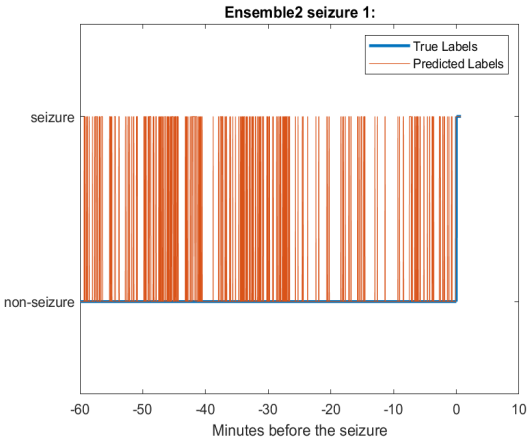

Ensemble 1 seizure 20 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
9366
9367
9368
9369
9370
9371
9372
9373
9374
Minutes of recording

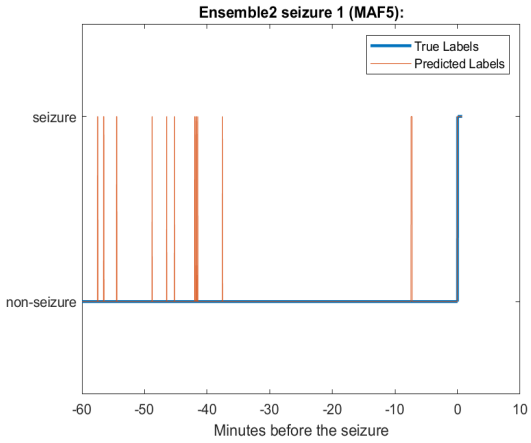

Ensemble 1 seizure 21:
True Labels
Predicted Labels
seizure
non-seizure
9434
9435
9436
9437
9438
9439
9440
9441
Minutes of recording

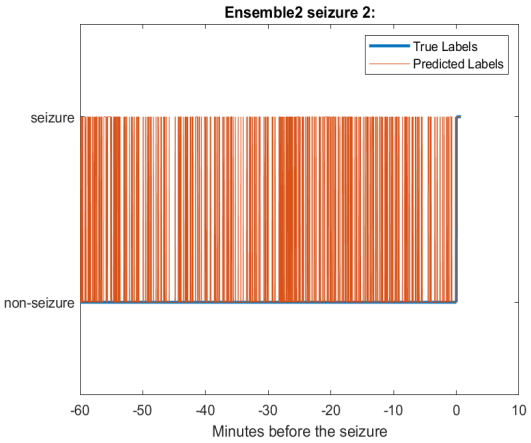

Ensemble 1 seizure 21 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
9434
9435
9436
9437
9438
9439
9440
9441
Minutes of recording

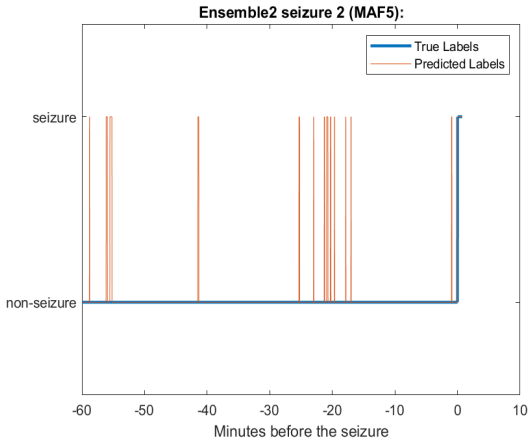
B Plots of the Seizures: dataset D

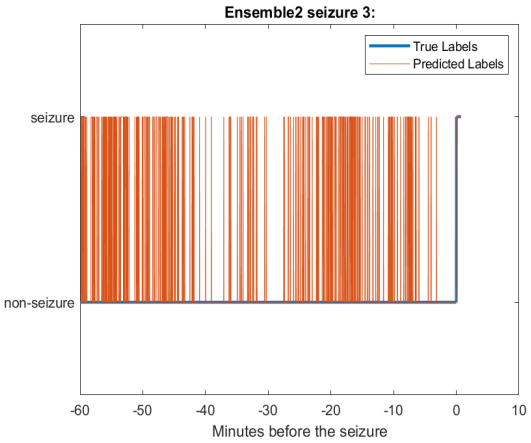
Here we present the results of the D testing dataset for every seizure.

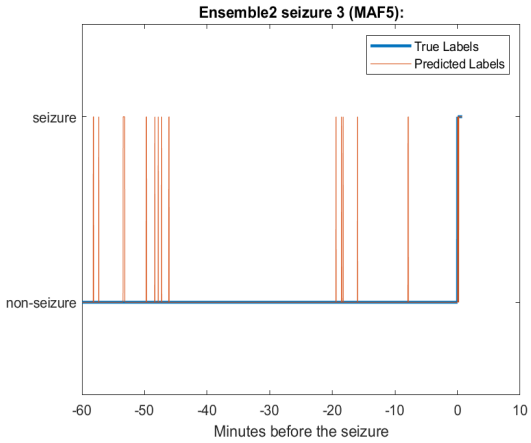

Ensemble2 seizure 1:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

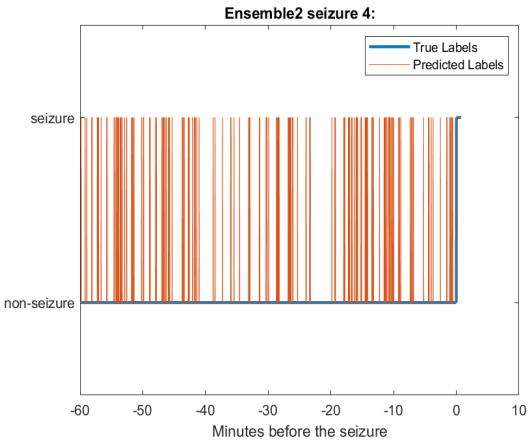

Ensemble2 seizure 1 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

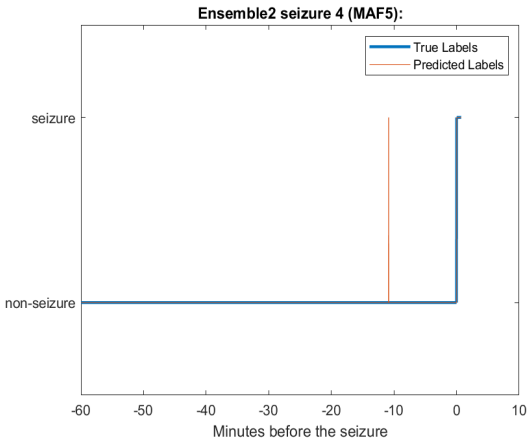

Ensemble2 seizure 2:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

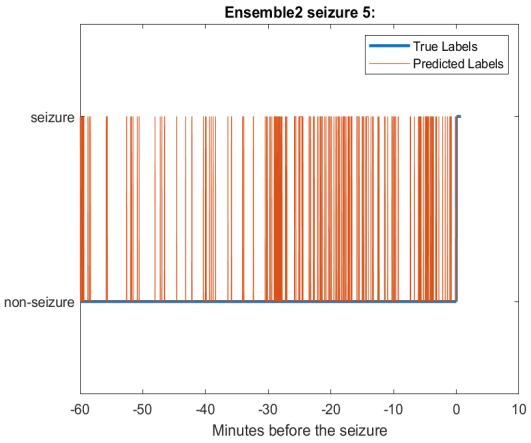

Ensemble2 seizure 2 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

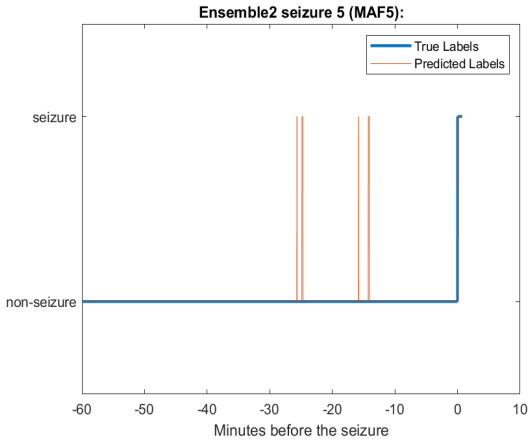

Ensemble2 seizure 3:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

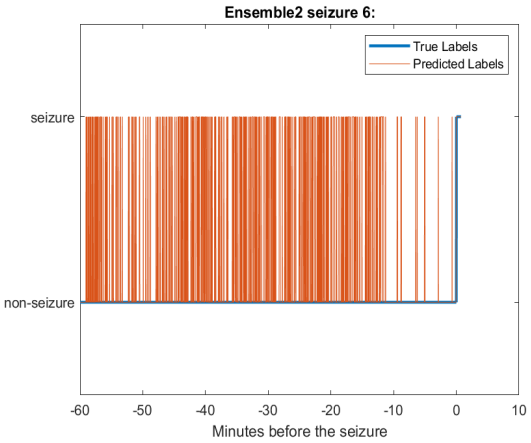

Ensemble2 seizure 3 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

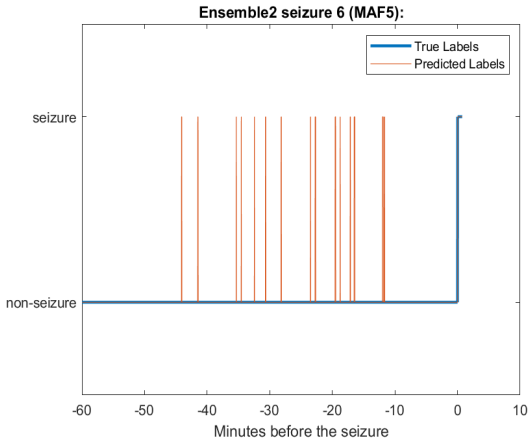

Ensemble2 seizure 4:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

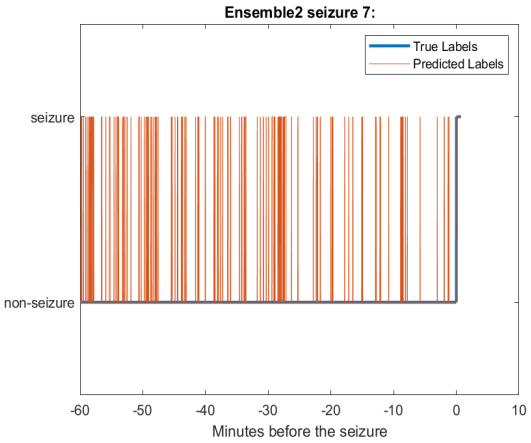

Ensemble2 seizure 4 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

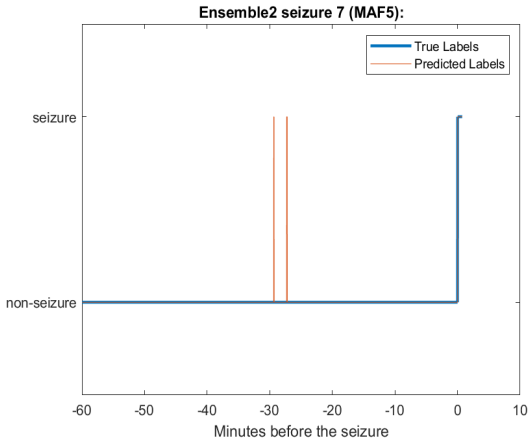

Ensemble2 seizure 5:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

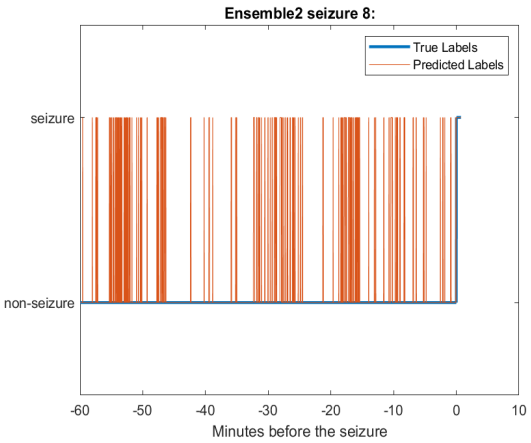

Ensemble2 seizure 5 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

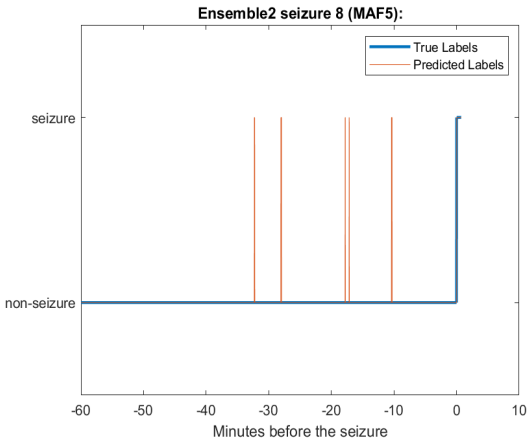

Ensemble2 seizure 6:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

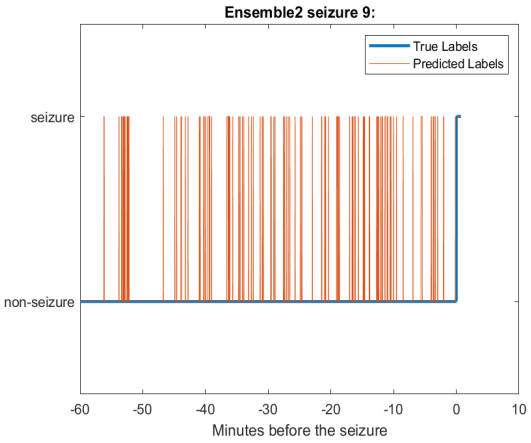

Ensemble2 seizure 6 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

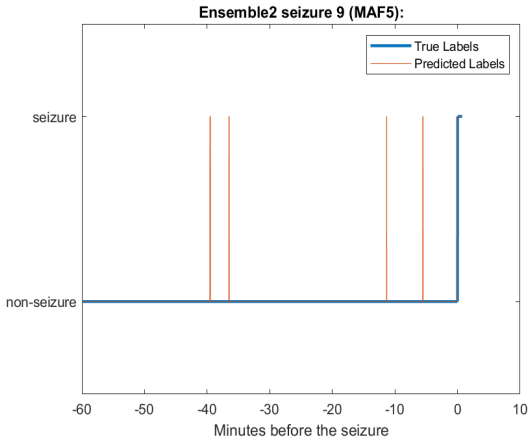

Ensemble2 seizure 7:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

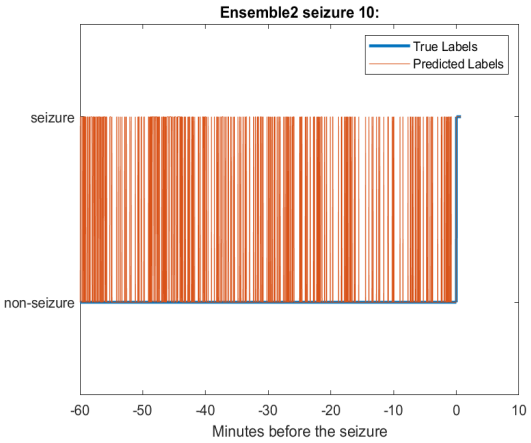

Ensemble2 seizure 7 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

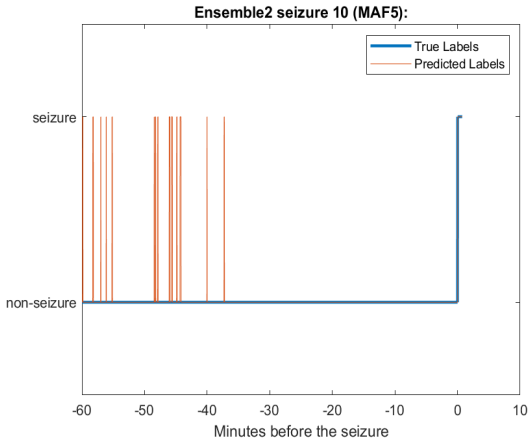

Ensemble2 seizure 8:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

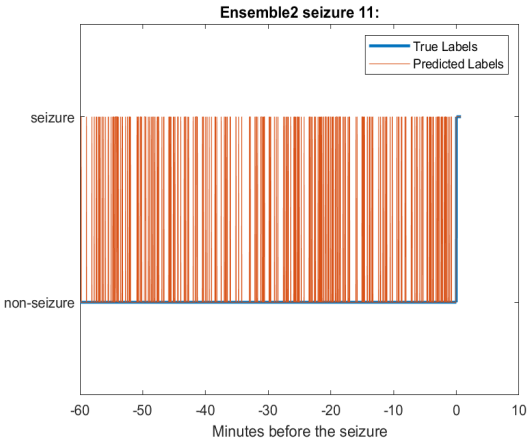

Ensemble2 seizure 8 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

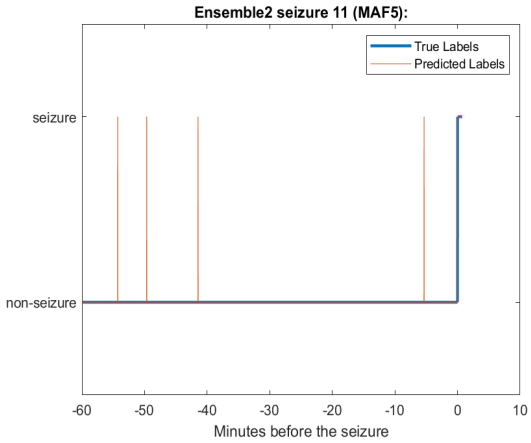

Ensemble2 seizure 9:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

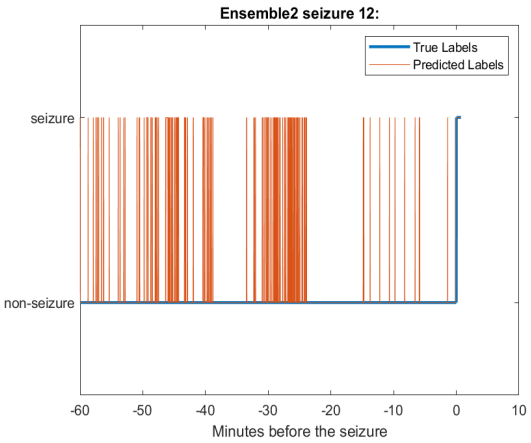

Ensemble2 seizure 9 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

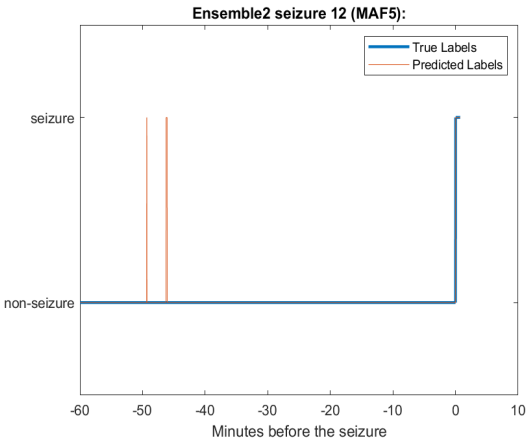

Ensemble2 seizure 10:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

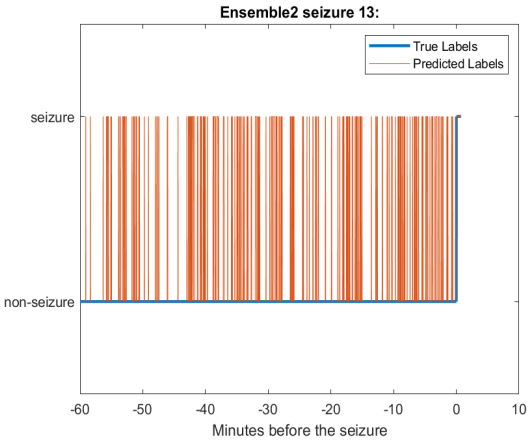

Ensemble2 seizure 10 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

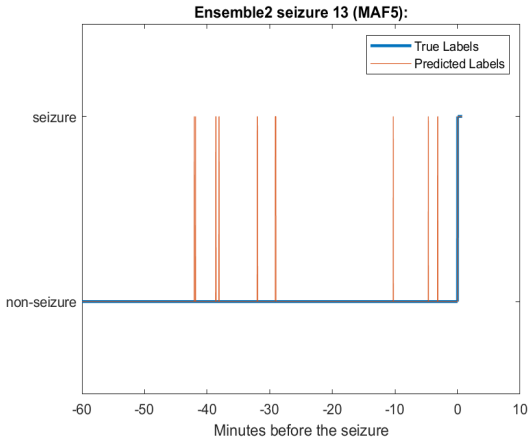

Ensemble2 seizure 11:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

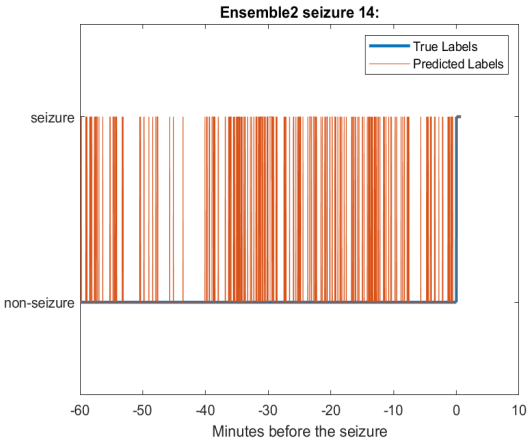

Ensemble2 seizure 11 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

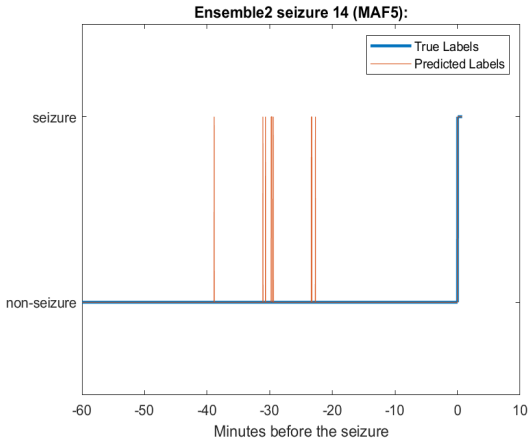

Ensemble2 seizure 12:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

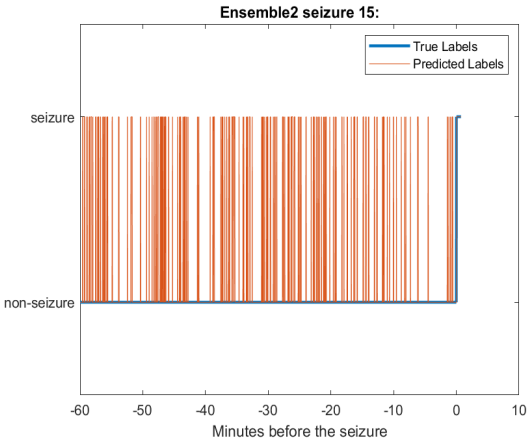

Ensemble2 seizure 12 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

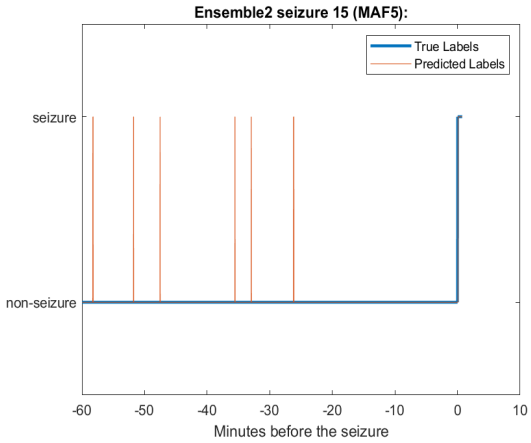

Ensemble2 seizure 13:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

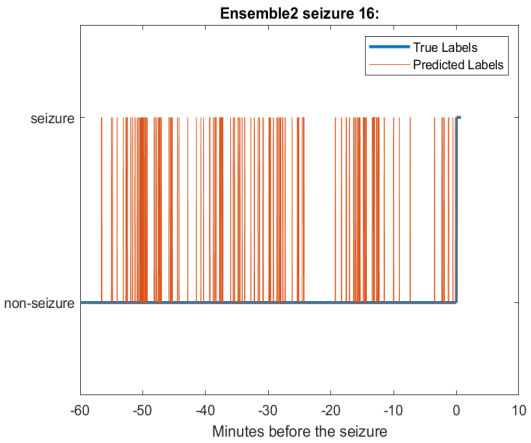

Ensemble2 seizure 13 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

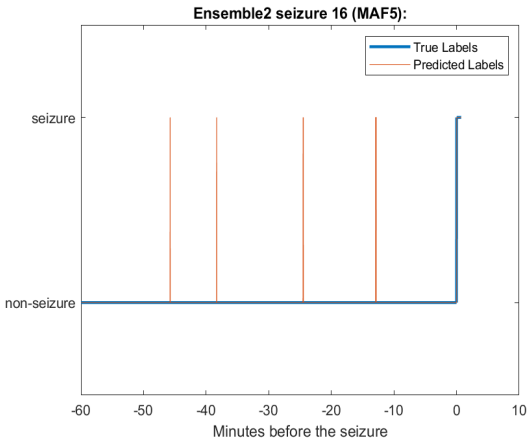

Ensemble2 seizure 14:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

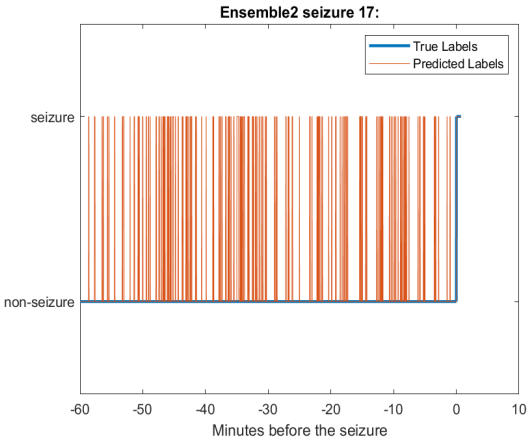

Ensemble2 seizure 14 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

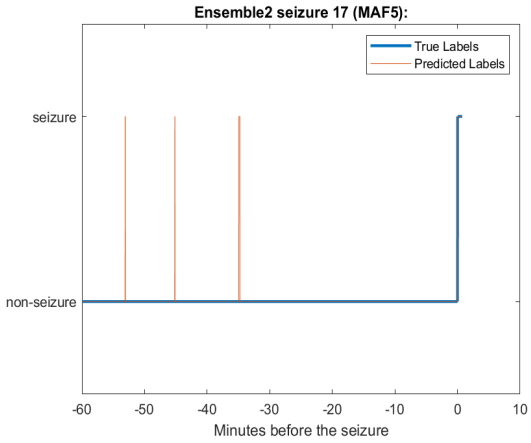

Ensemble2 seizure 15:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

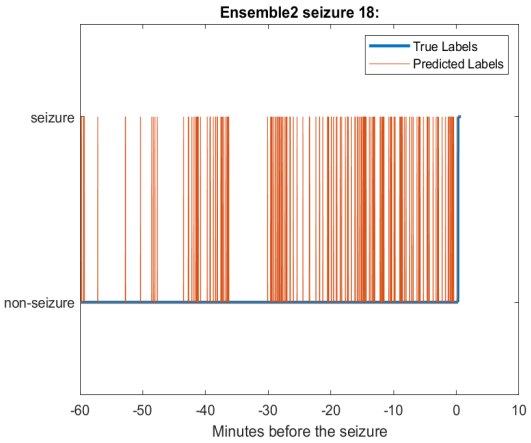

Ensemble2 seizure 15 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

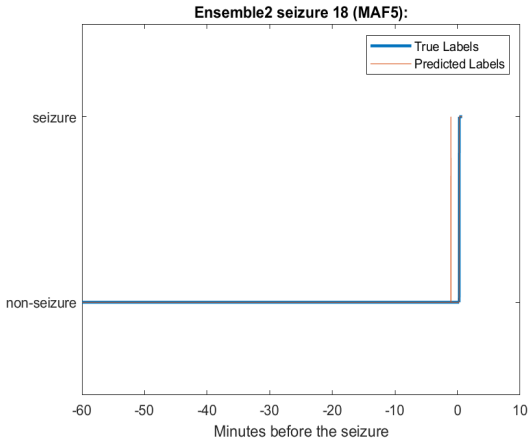

Ensemble2 seizure 16:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

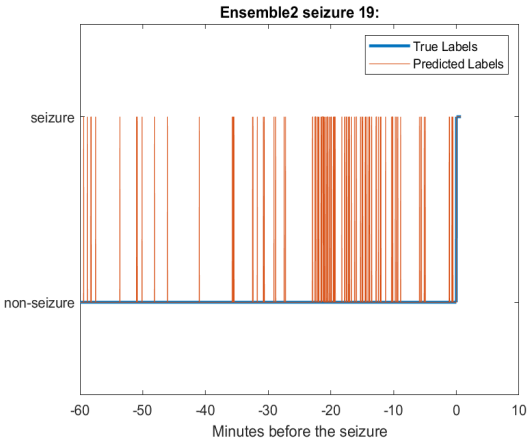

Ensemble2 seizure 16 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

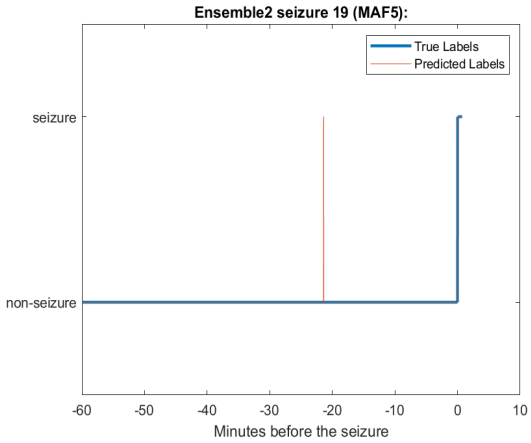

Ensemble2 seizure 17:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

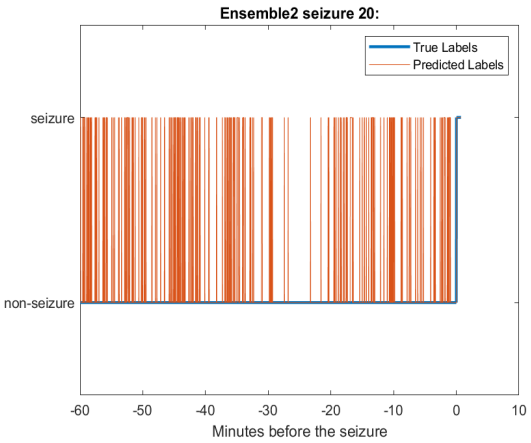

Ensemble2 seizure 17 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

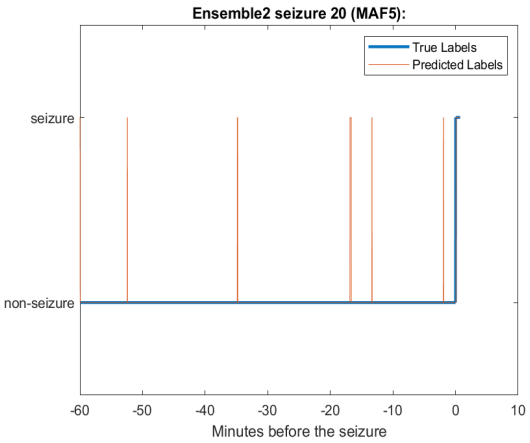

Ensemble2 seizure 18:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

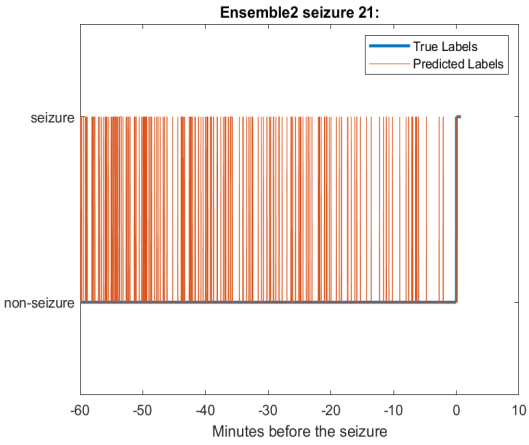

Ensemble2 seizure 18 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure

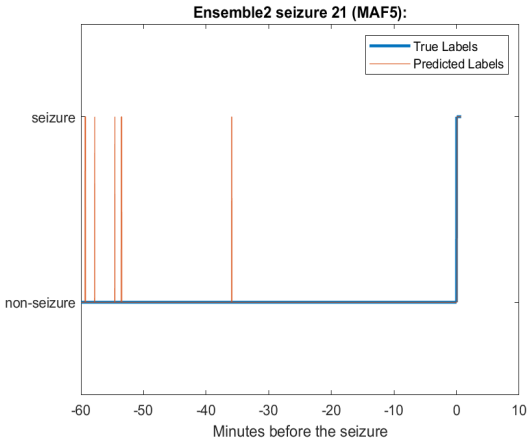

Ensemble2 seizure 19:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure


Ensemble2 seizure 19 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure


Ensemble2 seizure 20:
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure


Ensemble2 seizure 20 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60 -50 -40 -30 -20 -10 0 10
Minutes before the seizure


Ensemble2 seizure 21:
True Labels
Predicted Labels
seizure
non-seizure
-60
-50
-40
-30
-20
-10
0
10
Minutes before the seizure


Ensemble2 seizure 21 (MAF5):
True Labels
Predicted Labels
seizure
non-seizure
-60
-50
-40
-30
-20
-10
0
10
Minutes before the seizure