UNIVERSIDADE D
COIMBRA

Tiago André Cruz Coelho

EEG EPILEPSY SEIZURE PREDICTION: A MULTI-OBJECTIVE EVOLUTIONARY APPROACH

Thesis submitted to the Faculty of Science and Technology of the University of Coimbra for the degree of Master in Biomedical Engineering with specialization in Clinical Informatics and Bioinformatics, supervised by Prof. Dr. César Teixeira and MSc Mauro Pinto.

November 2020

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D COIMBRA

Tiago André Cruz Coelho

EEG Epilepsy Seizure Prediction: A Multi-Objective Evolutionary Approach

Thesis submitted to the
University of Coimbra for the degree of
Master in Biomedical Engineering

Supervisors:
Prof. Dr. César Alexandre Domingues Teixeira (CISUC)
Mestre Mauro Filipe da Silva Pinto (CISUC)

Coimbra, 2020

This work was developed in collaboration with:

Center for Informatics and Systems of the University of Coimbra

Agradecimentos

Gostaria de agradecer, em primeiro lugar, a ambos os meus orientadores por todo o apoio que me prestaram durante este ano letivo. Ao Professor Doutor César Teixeira, cujos anos de experiência académica e científica foram indispensáveis para a concretização deste projeto. Ao Mauro Pinto, aluno de doutoramento, para além de todas as horas despendidas em discussões e esclarecimentos, fico-lhe grato pelas aprendizagens que me proporcionou tanto a nível científico como pessoal.

Aos colegas e amigos que passaram pela minha vida ao longo destes cinco anos, um grande obrigado. O meu percurso académico não teria sido o mesmo sem todos os momentos que partilhámos, tanto de euforia como de desespero.

Exprimo a minha profunda gratidão à minha família, que permitiu que tudo isto fosse possível. O que alcancei foi graças ao vosso esforço, aos vossos ensinamentos e à confiança que sempre depositaram em mim. Sem vocês como molde, não seria metade da pessoa que sou hoje. Travámos esta batalha juntos e, do mesmo modo, enfrentaremos todas as que o futuro nos trouxer.

À Andreia, agradeço por caminhar ao meu lado desde há vários anos. Entre bons e maus momentos, estiveste sempre presente e pronta a ser o meu porto seguro. Obrigado pelo teu apoio e amor incondicional e por me fazeres sorrir e chorar como mais ninguém consegue.

Por fim, obrigado a Coimbra, a cidade que me viu nascer e crescer. Tal como a vida, feita de subidas e descidas, será sempre o canto do mundo a que tenho o prazer de chamar "casa".

"For long you live and high you fly
And smiles you'll give and tears you'll cry
And all you touch and all you see
Is all your life will ever be."

ROGER WATERS

Resumo

Cerca de 30% dos doentes epiléticos não conseguem obter uma supressão eficaz das crises a longo prazo após administração de fármacos anticonvulsivantes. Para estes indivíduos, a previsão de crises foi identificada como uma abordagem promisora, que utiliza algoritmos de previsão para alertar o paciente atempadamente e permitir que sejam tomadas medidas preventivas. A determinação do período pré-ictal, no qual o cérebro transita de um estado normal para um de hiperexcitabilidade, é considerado um passo crítico.

Apesar da abundância de abordagens baseadas no Eletroencefalograma (EEG), muitas não atingem a aplicabilidade clínica à custa de uma capacidade de generalização insuficiente. Além do mais, vários estudos recentes têm adotado métodos baseados em *Deep Learning* que, apesar de atingirem um desempenho significativo em relação aos classificadores tradicionais, resultam na perda de interpretabilidade clínica, dificultando a sua implementação em dispositivos médicos.

O presente trabalho consistiu no desenvolvimento de um esquema de previsão baseado em dados de monitorização pré-cirúrgica da base de dados EPILEPSIAE, aplicando Algoritmos Evolucionários Multi-Objetivo (MOEA) para otimização de modelos. Nomeadamente, foram experimentadas três implementações diferentes: MOGA, NSGA-II e um SMS-EMOA modificado. Cada um otimizou um conjunto de modelos de previsão específicos para cada paciente, procurando gerar um conjunto ótimo de *features* computacionalmente leves para treinar um classificador baseado em Máquinas de Vetores de Suporte (SVM).

Considerando um grupo de 36 doentes com Epilepsia do Lobo Temporal (TLE), a metodologia proposta alcançou um desempenho estatisticamente significativo para 19%, 30% e 14% considerando o previsor aleatório e 89%, 83% e 86% usando análise *surrogate*, respetivamente, para cada MOEA. Foram também obtidos valores geralmente baixos para a Taxa de Falsos Positivos por Hora (FPR/h), sendo que um terço dos doentes apresentou valores abaixo de 0.15, que é um requisito para aplicabili-

dade em contexto real. Estes resultados demonstram a possibilidade de identificar o período pré-ictal e, simultaneamente, manter a interpretabilidade dos modelos, o que poderá contribuir para aprofundar a compreensão das dinâmicas cerebrais e melhorar a capacidade preditiva.

Palavras-chave: Epilepsia, Processamento de EEG, Previsão de crises, Algoritmos evolucionários

Abstract

An estimated 30% of epileptic patients cannot achieve effective long-term seizure suppression after antiepileptic drug administration. For these individuals, seizure prediction has been identified as a promising approach, making use of prediction algorithms to alert the patient in a timely manner and allow for preventive measures to be taken. The determination of the pre-ictal period, in which the brain transitions from a normal state to a seizure, is considered a critical step.

Despite the abundance of approaches based on the Electroencephalogram (EEG) signal, many fail to achieve real-world applicability on account of insufficient generalization capability. Moreover, recent studies have began to adopt Deep Learning methods, which despite attaining remarkable performance when compared to traditional classifiers, result in a loss of clinical interpretability, hindering their implementation on medical devices.

This work concerns the development of a prediction scheme based on pre-surgical monitoring data from the EPILEPSIAE database, employing Multi-Objective Evolutionary Algorithms (MOEAs) for model optimization. Namely, we experimented with three different implementations: MOGA, NSGA-II and a modified SMS-EMOA. Each one optimized a set of patient-specific prediction models, aiming to generate an optimal set of computationally light features to train a Support Vector Machine (SVM) classifier.

Considering a group of 36 patients suffering from Temporal Lobe Epilepsy (TLE), the proposed methodology performed above chance for 19%, 30% and 14% considering the random predictor and 89%, 83% and 86% using surrogate analysis, respectively, for each MOEA. Generally low False Positive Rate per Hour (FPR/h) scores were also obtained, with a third of all patients presenting average values lower than 0.15, which is a requirement for real-life applicability. These results demonstrate the possibility of identifying the pre-ictal period while maintaining model interpretability, which may contribute to a better understanding of brain dynamics

and enhance predictive performance.

Keywords: Epilepsy, EEG processing, Seizure prediction, Evolutionary Algorithms

Contents

List of Figures

List of Tables

List of Abbreviations

AED Anti-epileptic drug. 6, 9, 19, 21
AR Autoregressive. 41, 115, 116

CNN Convolutional Neural Network. 2, 44, 45, 46

DM Decision Maker. xxiv, 71, 72, 76, 78, 93, 121, 122
DRE Drug-Resistant Epilepsy. 1, 9, 27, 34, 53, 54
DWT Discrete Wavelet Transform. 41, 114

EA Evolutionary Algorithm. xx, 3, 22, 23, 24, 26, 29, 43
EEG Electroencephalogram. xii, xiii, xiv, xv, 1, 2, 3, 5, 8, 9, 10, 11, 12, 13, 14, 17, 28, 31, 32, 34, 35, 37, 38, 41, 42, 43, 44, 45, 50, 53, 54, 56, 64, 73, 82, 90, 95, 113, 114, 115, 116, 117, 118
EU European Union. 3

FBTC Focal to Bilateral Tonic-Clonic. 7, 55
FFT Fast Fourier Transform. 45, 114
FIR Finite Impulse Response. 37
FOA Focal Onset Aware. 7, 8, 55, 85, 86, 93
FOIA Focal Onset Impaired Awareness. 7, 8, 14, 55, 85, 86, 93
FPR/h False Positive Rate per Hour. xiii, xix, xxiv, 3, 16, 17, 18, 19, 28, 32, 37, 45, 48, 50, 58, 64, 71, 72, 74, 78, 82, 83, 84, 85, 88, 93, 95, 123

iEEG intracranial EEG. 11, 12, 28, 34
IIR Infinite Impulse Response. 37
ILAE International League Against Epilepsy. 5, 6, 8, 9
IT Intervention Time. 15

LSTM Long Short-Term Memory. 2, 44, 46

mDAD maximum Difference Amplitude Distribution histograms. 43

MOEA Multi-Objective Evolutionary Algorithm. xiii, xxiii, xxiv, 26, 27, 29, 57, 71, 74, 75, 76, 78, 82, 83, 84, 86, 88, 90, 93, 95, 96, 122
mRMR minimum Redundance Maximum Relevance. 43
MTLE Mesial Temporal Lobe Epilepsy. 9

PCA Principal Component Analysis. 43
PSD Power Spectral Density. 41, 114

RNN Recurrent Neural Network. 46

SOP Seizure Occurrence Period. xix, xxiv, 2, 3, 14, 15, 16, 19, 28, 35, 37, 38, 44, 48, 51, 59, 64, 72, 73, 83, 86, 87, 88, 93, 123
SPH Seizure Prediction Horizon. xix, 3, 14, 15, 16, 28, 35, 37, 38, 44, 48, 51, 53, 63, 64, 73, 83, 84
SUS Stochastic Universal Sampling. 66
SVM Support Vector Machine. xiii, 44, 45, 57, 63, 73, 83, 95

TLE Temporal Lobe Epilepsy. xiii, 9, 12, 27, 53, 54, 73, 83, 84

1

Introduction

This chapter contains a short overview of the motivation behind this study in section 1.1, as well as the context surrounding it in section 1.2. The main goals and expected contributions are discussed in section 1.3. Lastly, the outline of this document can be found in section 1.4.

1.1 Motivation

Epilepsy is a chronic neurological condition characterized by recurrent seizures. With a prevalence of more than 3 million patients in Europe alone (and about 1% of the global population), it is one of the most common diseases of the human brain. Due to the apparent unpredictable nature of seizure occurrence, patients suffer not only physical (e.g. risk of trauma), but also social and psychological (e.g. stigma and discrimination) consequences [52, 60]. Additionally, it is associated with significant economic implications, including health-care needs and loss of productivity [114].

Although seizure control through antiepileptic drug administration or surgical intervention has a success rate of 70%, new treatment concepts must be developed for the remaining group of patients, such as seizure prediction. By laying the basis for future warning systems or closed-loop intervention devices, it may allow to, respectively, decrease anxiety levels regarding their unpredictability and reduce/suppress seizure effects, effectively improving the patient's quality of life [50, 52].

1.2 Context

Since the 1970's, seizure prediction has been largely based on the Electroencephalogram (EEG) signal dynamics, with increasing advances throughout the years. The signal has been used to continuously monitor epileptic patients and successfully evaluate and diagnose Drug-Resistant Epilepsy (DRE). The main goal is to develop tools which acquire on-line data and raise alarms/warnings some time before each

seizure, providing time for intervention. After starting with simple thresholding, modern approaches have adopted machine learning with promising results. Even so, several questions arise concerning the reliability and validity of previously developed algorithms [75].

The current approaches face several factors which hamper their real-life implementation. For instance, several of the used databases are comprised of discontinuous and/or short-time EEG recordings. Additionally, these require manual annotation of each seizure as well as contextual information (e.g. sleep, patient movement), a demanding and complex task which should be performed by experts. The signal's morphology and characteristics are also not fully understood.

A particularly significant subject is the determination of the pre-ictal period. It is described as the time interval preceding the seizure onset where the brain transitions from a normal state to a seizure, and it defines a critical step in seizure prediction algorithms. Authors have not reached a consensus regarding an optimal value for this time interval, with different studies reporting that it not only varies between patients but also between seizures within each patient [42, 62, 81].

Other issues include the clinical heterogeneity in epilepsy diagnosis, data imbalance (normal brain activity is prevalent), as well as the presence of concept drifts: considering that most of the data derives from patients in pre-surgical monitoring, where anticonvulsant medication is suppressed, neurophysiological conditions will be different and seizures will be more frequent than in everyday life. Other physiological phenomena such as circadian rhythm or stress also contribute to alterations in brain dynamics [56].

Another core aspect in the current framework is feature selection, not only to ensure a reliable performance for the prediction algorithm but also to meet the needs for interpretable results with possible clinical knowledge extrapolation. Once again, no consensus was reached concerning the optimal group of features or feature types. Furthermore, many studies employ post-processing methods in order to improve the prediction performance. These, however, are simply ad hoc solutions with the purpose of reducing the number of false alarms [11, 102].

Recent machine learning studies have adopted Deep Learning techniques such as Long Short-Term Memory (LSTMs) and Convolutional Neural Networks (CNNs), which outperform most traditional classifiers by handling the time property directly. Despite their ability to perform automatic feature engineering, there is an inevitable loss of interpretability that severely affects their applicability in a clinical context.

In order to attain a high degree of confidence in new methodologies, these should be evaluated considering several factors: the duration of the Seizure Occurrence

Period (SOP) (during which the seizure will occur) and Seizure Prediction Horizon (SPH) (time for intervention), False Positive Rate per Hour (FPR/h) and sensitivity. Their direct impact on the patient's well-being must be taken into account (e.g., overly long SOP and SPH may induce overwhelming amounts of stress) and tests on long-term recordings should be carried out to evaluate real-life applicability.

1.3 Research goals

Regarding the main objective, which is to advance the current EEG seizure prediction framework, the expected contributions of this thesis are the following:

- Exploitation of the European Epilepsy Database, with the development of a new prediction approach based on its data. It is the largest epilepsy database, created as part of the EU-funded EPILEPSIAE project, containing long-term recordings of 275 patients with extensive clinical metadata.
- Exploitation of several aspects of Evolutionary Algorithms (EAs), such as fitness function design and multi-objective optimization, and how these might contribute to seizure prediction as well as understanding the underlying brain dynamics.
- Contribution towards a new feature construction scheme, aiming to achieve both satisfactory performance and interpretability.

1.4 Outline

This document is divided into five chapters beyond the introduction.

Chapter 2 presents background information regarding epilepsy, the EEG signal, seizure prediction concepts and evolutionary computing terminology that will be referred to throughout this document.

Chapter 3 showcases the state of art concerning EEG-based seizure prediction.

Chapter 4 describes the various methods employed throughout the experimental work, including signal processing, machine learning and EAs.

Chapter 5 reports not only the results obtained from the machine learning algorithms, but also their statistical validation. A thorough discussion on the obtained results and other aspects of this thesis is also provided.

Chapter 6 presents a conclusion and addresses future perspectives in this field of study.

2

Background Concepts

This chapter introduces the main concepts needed to follow this document. Firstly, a brief overview of epilepsy, its clinical definitions and classification is presented in section 2.1, followed by the characterization of the Electroencephalogram (EEG) signal and its application in neurophysiology in section 2.2. Section 2.3 introduces all the necessary notions within the field of seizure prediction, while section 2.4 provides insight on the main concepts of evolutionary computing. Lastly, a short summary is included in section 2.5.

2.1 Epilepsy

2.1.1 Clinical definitions

Epilepsy is defined, according to the International League Against Epilepsy (ILAE) in 2005, as "*a disorder of the brain characterized by an enduring predisposition to generate epileptic seizures and by the neurobiological, cognitive, psychological, and social consequences of this condition*". Additionally, its definition "*requires the occurrence of at least one epileptic seizure*" [38].

An epileptic seizure is defined as "*a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain*" [38].

These conceptual definitions, however, do not cover all the clinical circumstances. Hence, the ILAE Task Force proposed an operational definition of the disease in 2014 [37], which states that epilepsy is a disease of the brain defined by any of the following conditions:

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

In this new, practical definition, the risk of recurrence after a single unprovoked seizure is given greater consideration, allowing clinicians to initiate treatment after some initial unprovoked seizures. The term "unprovoked" indicates absence of a temporary/reversible factor lowering the threshold and causing a seizure at that point in time.

It is worth noting that epilepsy is not necessarily a life-long condition, and is considered to be "resolved" if a person has been seizure-free for the last 10 years, with at least the last 5 years off Anti-epileptic drugs (AEDs), or when that person has passed the age of an age-dependent epilepsy syndrome. However, when epilepsy is resolved it does not guarantee that it will not return.

2.1.2 Classification

The ILAE Task Force, in addition to its clinical definition, also updated the classification of epilepsy diagnosis. As seen in figure 2.1, it is presented in three levels: firstly, the seizure type, then the epilepsy type and thirdly, the epilepsy syndrome.

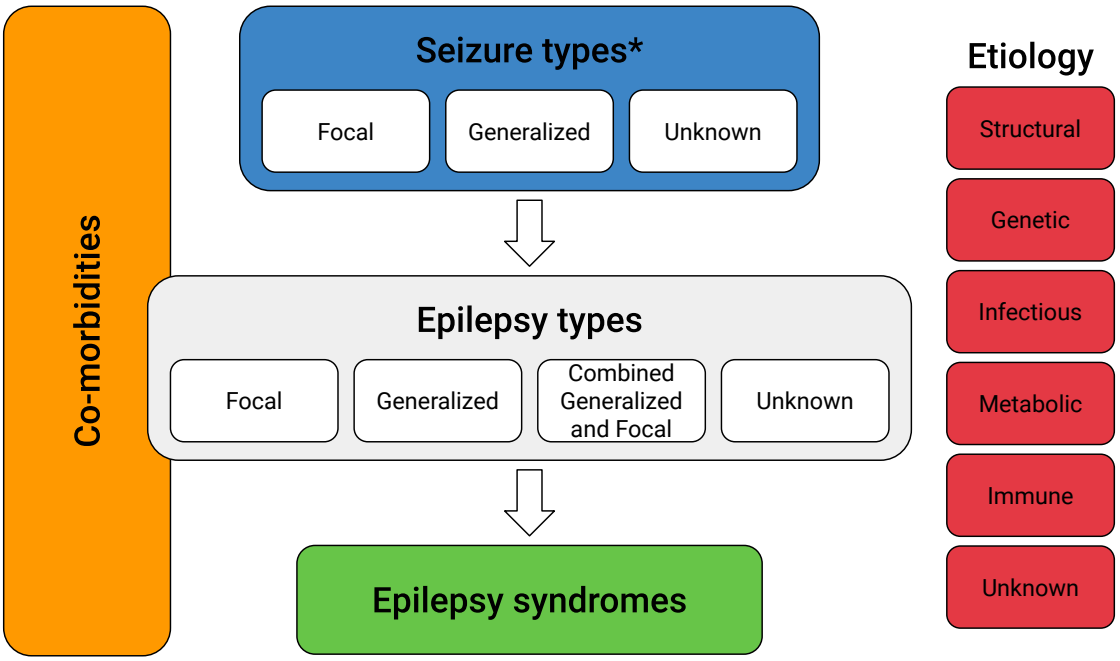


Figure 2.1: ILAE 2017 framework for classification of epilepsy. *Denotes onset of seizure. Adapted from: Sheffer et al. [90]

The new classification also incorporates etiology along each stage, as it often carries significant treatment implications at each step of the diagnosis. It is divided into different subgroups according to their potential therapeutic consequences but, for the sake of simplicity, its influence will not be discussed in this document [90].

Seizure type

The clinician's first task is to ascertain if an event has the characteristics of a seizure. When it is proven to be a seizure, its classification begins with the determination of whether the initial manifestations (onset) of the seizure are focal, generalized or unknown.

Focal seizures correspond to seizures with clinical or EEG onsets originating within networks limited to one hemisphere, whereas generalized ones involve bilaterally distributed networks, that is, both hemispheres.

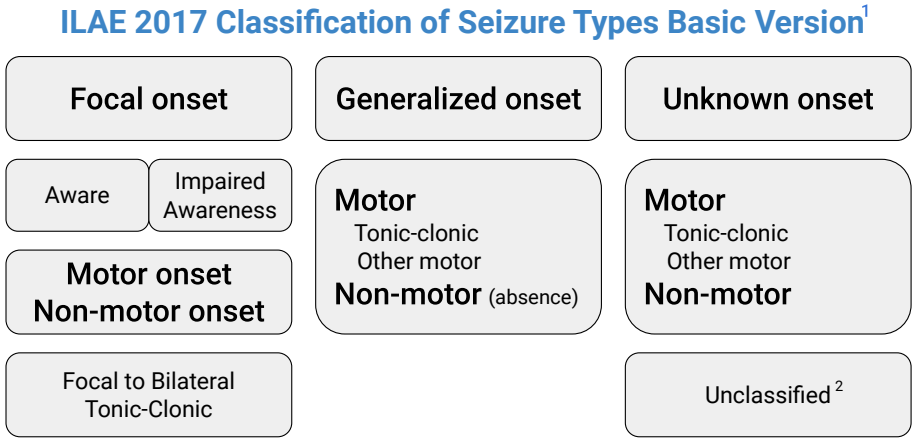


Figure 2.2: The basic ILAE 2017 operational classification of seizure types. [1]Definitions, other seizure types and descriptors are listed in the accompanying paper and glossary of terms. [2] Due to inadequate information or inability to place in other categories. Adapted from: R.S. Fisher et al. [39]

As shown in the figure above, seizure types can be subgrouped as motor or non-motor, according to behavioral manifestations. This is also valid for seizures of unknown onset, where the clinician lacks sufficient information to determine if it is focal or generalized.

Focal seizures can also be classified based on the person's level of awareness of self and environment during the seizure, even if immobile. If awareness is retained, it is classified as Focal Onset Aware (FOA), otherwise, the seizure is categorized as Focal Onset Impaired Awareness (FOIA). A particular case that should be noted are Focal to Bilateral Tonic-Clonic (FBTC) seizures [34]: the onset is limited to one

hemisphere, but rapidly propagates to the other, often resulting in tonic (muscle stiffness) and clonic (jerking movements) symptoms.

It is worth noting that, depending on the available resources, classification according to seizure type may be the maximum level possible for diagnosis as there may be no access to EEG, video and imaging studies [39, 90].

Despite the lack of guidelines provided by the ILAE Task Force, several other aspects are often registered and analyzed by clinicians. Hence, in practice, it is common to annotate the patient's state of vigilance at the time of the seizure (awake, non-REM stages I-IV, REM) as well as the onset's localization in terms of hemispheres (left, right, bilateral) and brain lobes (frontal, parietal, occipital, temporal) [60].

Epilepsy type

The second level of the classification framework identifies the type of epilepsy that affects the patient and assumes a diagnosis of epilepsy based on the 2014 definition. As illustrated in figure 2.1, epilepsy types can be grouped in four different categories:

- Focal: includes unifocal and multifocal disorders and seizures that involve one hemisphere; the most common seizure types include FOA, FOIA, focal motor, focal non-motor and focal to bilateral tonic-clonic ones, with the interictal EEG characterized by focal epileptiform discharges.
- Generalized: characterized by generalized spike-wave activity on the EEG, with seizure types including absence, myoclonic, atonic, tonic and tonic-clonic ones.
- Combined Generalized and Focal: characterized by having generalized and focal seizures with the interictal EEG showing generalized spike-wave and focal epileptiform discharges.
- Unknown: the patient is confirmed to have epilepsy, but the clinician is unable to conclude whether the epilepsy type is focal or generalized from the information available.

As shown, each epilepsy type is associated with a high degree of complexity, given that each category contains multiple types of seizures [90].

Epilepsy syndrome

The third level of classification is an epilepsy syndrome diagnosis. An epilepsy syndrome refers to a group of characteristics including seizure types, EEG and imag-

ing features which tend to occur together. Additionally, others such as age at onset and remission (where applicable), seizure triggers and diurnal variation are often considered. However, it should be noted that there is no formal classification of syndromes by the ILAE [90].

The identification of an epilepsy syndrome is helpful as it provides information on the underlying causes and which AEDs might be suitable. This is especially relevant given that several syndromes display seizure aggravation with particular medications, which can be prevented through appropriate early diagnosis [53].

Temporal Lobe Epilepsy (TLE) represents the most common form of focal epilepsy, and is divided in two types: Mesial Temporal Lobe Epilepsy (MTLE), which accounts for over 80% of all TLE cases and neocortical/lateral temporal lobe epilepsy. This syndrome is characterized by seizures originating in temporal structures such as the hippocampus or surrounding areas, typically diagnosed in early to late adolescence. Anti-seizure medication is often ineffective, thus making surgery a frequent recommendation [35, 95]. It is worth noting that various seizure prediction studies have focused on TLE patients [11, 16, 30].

Drug-Resistant Epilepsy (DRE)

The ILAE Task Force, in 2009, defined DRE as "*failure of adequate trials of two tolerated, appropriately chosen and used AED schedules (whether as monotherapies or in combination) to achieve sustained seizure freedom*". Seizure freedom was also defined, in the same report, as "*freedom from all types of seizures for 12 months or three times the preintervention interseizure interval, whichever is longer*" [63].

This group of patients is of serious concern, bearing in mind that they are subjected to the unpredictable nature of seizure occurrence and its physical and social consequences. They are the focus of seizure prediction approaches (as well as those with ineffective surgical intervention), which can significantly improve their quality of life, preventing severe accidents and injuries, which are one of the main causes of death related to epilepsy [114].

2.2 EEG

2.2.1 Overview

The Electroencephalogram (EEG) measures and records the electrical activity of the brain in the form of a time series, representing the potential voltage fluctuations in space and time. These electrical potentials arise from summated excitatory

and inhibitory postsynaptic potentials which are generated primarily by cortical pyramidal cells [2, 81].

Considering that epileptic seizures occur due to a malfunction in the electrophysiological system of the brain, causing a sudden, synchronous electrical discharge in a group of neurons, it is naturally understood why the EEG became the primary clinical tool used when handling patients suffering from epilepsy [2, 50].

Generally, the potentials captured by the EEG can be categorized into two types of phenomena: oscillations and transients. Oscillations, in the brain, refer to rhythmic fluctuations of neurons or populations of neurons. These can be produced by excitatory/inhibitory networks or interactions between them [23]. They can be characterized in terms of frequency band activity: delta (0-4Hz), theta (4-8Hz), alpha (8-13Hz), beta (13-30Hz) and gamma (>30Hz) [11]. It should be noted that there is no general consensus among authors regarding these bandwidths.

Concerning transients, they can be considered normal or abnormal. Normal transients include various sleep potentials as well as non-cerebral electrical potentials (e.g. eye blinks, cardiac impulses, muscle activity). Abnormal transients are subgrouped into epileptiform (e.g. spike and wave complexes, sharp or slow wave discharges) and non-epileptiform. Epileptiform discharges can be identified during the inter-ictal stage and often form the basis for diagnosis of epilepsy [81].

Furthermore, besides the aforementioned activity, there can be other, non-physiological artifacts present in the EEG. These are generated by electromagnetic fields outside the patient's body, as is the case with power-line interference (50 or 60Hz), changes in the electrodes' impedance, and interference from the environment (movement and medical instrumentation) [112].

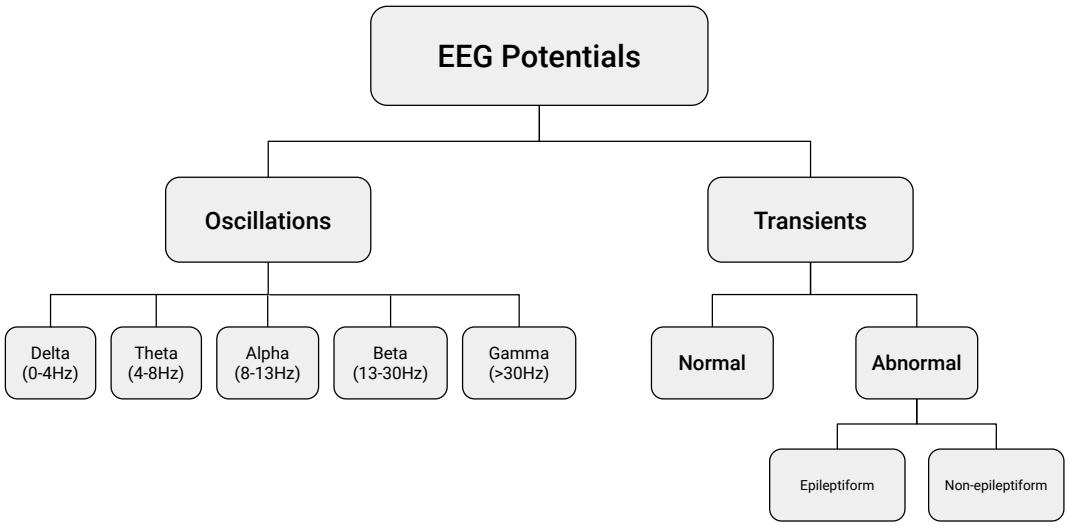


Figure 2.3: Categorization of EEG activity.

2.2.2 Signal acquisition

The EEG signal is acquired by placing several electrodes either on the scalp (scalp EEG) or inside the patient's skull (intracranial EEG (iEEG)). The signal's spatial resolution is determined by the number of electrodes used and their respective localization, while time resolution depends on the sampling frequency.

Scalp EEG

Scalp EEG represents the most common acquisition method, as it is non-invasive. Electrodes are placed on the patient's scalp according to the international 10-20 system (as depicted in figure 2.4) and, through the use of an electroconductive gel, the existing impedance is reduced [6, 23]. The standard set of electrodes in adults is comprised of 21 electrodes for recording electrical activity and an additional one for ground reference, when using a referential montage. Bipolar montages are also commonly used, as explained below [81, 89].

When acquiring the signal, each electrode measures the voltage difference between the electrode itself and a reference. After acquisition, one can choose to maintain the original signals or create what is referred to as bipolar montages, which typically involve the difference between the voltages measured from adjacent electrodes [81].

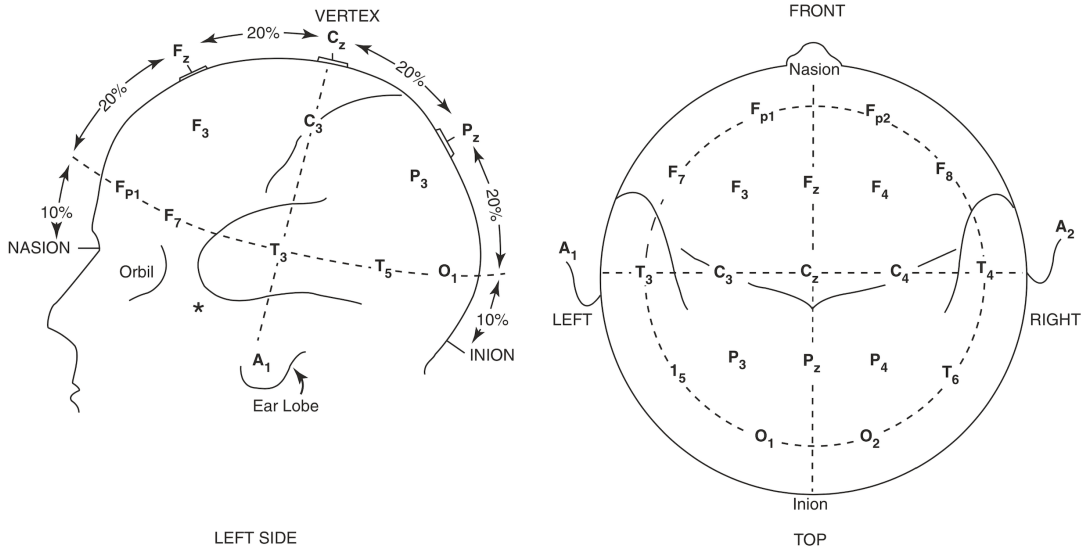


Figure 2.4: The international 10-20 system used for electrode placement. Source: M. Sazgar and M. G. Young, *Overview of EEG, Electrode Placement, and Montages* [89]

iEEG

iEEG is an invasive method, in which electrodes are placed directly on the exposed surface of the patient's brain during surgery. A craniotomy (surgical incision into the skull) is performed and subdural grids or strips with numbered electrodes are placed on a small region of the cortex, as illustrated by figure 2.5 [6, 81].

For intracranial recordings, there are also different options for a reference electrode: another intracranial electrode, taking the average of all electrodes on the grid/strip, or an extracranial electrode. Depth electrodes can also be placed in deep brain structures and have been proven useful in detecting the spread of epileptiform activity in patients with TLE. However, major risks when placing these should be considered, such as hemorrhage or infection [23, 36, 81].

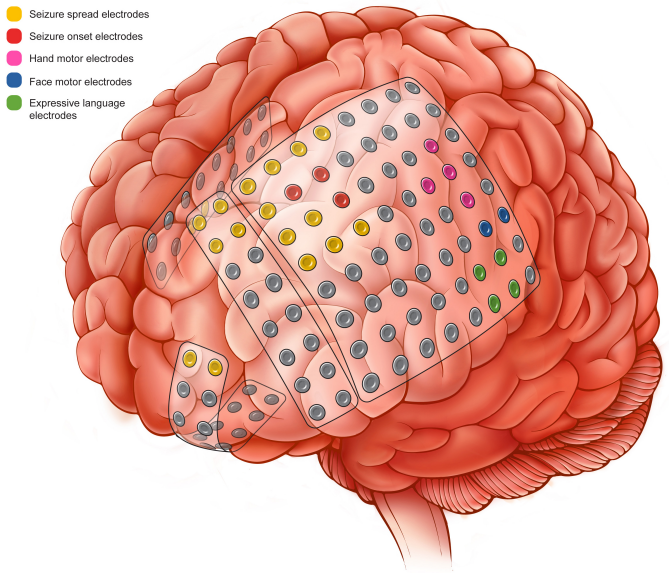


Figure 2.5: Illustration of intracranial electrode placement. Grids/strips with numbered electrodes are placed directly on the surface of the patient's brain. Source: J. Watson, *Epilepsy. June 2011.* [55]

Scalp EEG vs. iEEG

Non-invasive, scalp EEG recordings have a number of limitations when compared to invasive ones, considering that muscle artifacts are more abundant and they cannot accurately capture part of the activity in the beta and gamma bands. Moreover, iEEG recordings have a higher signal to noise ratio, since electrodes are placed closer to the brain structures. However, considering the inherent ethical reasons, increased risk (up to 4%) of infection/hemorrhage and since the method does not cover wider regions of the brain, invasive recordings bear some shortcomings

as well [61, 81]. Epilepsy patients in pre-surgical monitoring are only subjected to intracranial recordings when the seizure localization is inconclusive [36].

Each acquisition technique can be associated with different goals in terms of its application in seizure prediction. Invasive recordings are preferred for implantable, closed-loop brain stimulation devices (e.g. the RNS system [98]), despite the aforementioned risk and maintenance requirements such as device migration. For simpler warning systems, however, scalp recordings, whilst noisy, may be desirable for their reduced cost and ease of access to the general population [42, 62, 75, 94].

In either case, it is worth noting that the EEG is a complex signal which aims to approximate the true underlying complexity of the linear and non-linear nature of the interactions between neurons [75].

2.2.3 Epileptic epoch segmentation

The EEG signal of an epileptic patient can be divided into different periods in time, relative to the seizure, as shown in figure 2.6. Hence, each seizure episode consists of a pre-ictal period (preceding the seizure), an ictal period (corresponding to the seizure), a post-ictal period (following the seizure) and, finally, an inter-ictal period (between the post- and pre-ictal periods of consecutive seizures).

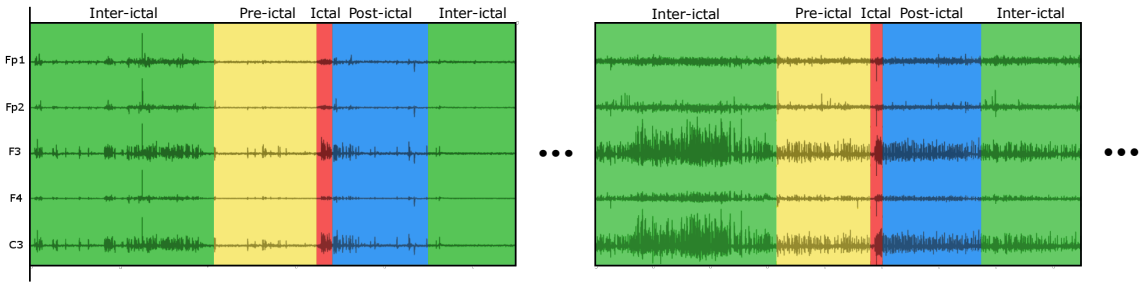


Figure 2.6: Different periods of a seizure episode annotated on an EEG signal: inter-ictal (green), pre-ictal (yellow), ictal (red) and post-ictal (blue). Data selected from patient 16202 from the EPILEPSIAE database.

It should be stressed that the pre-ictal period is the most difficult one to determine and is not even annotated by neurophysiologists, as it is associated with a considerable degree of heterogeneity: its characteristics can vary not only between patients, but also between seizures [42, 62]. Furthermore, abnormal non-epileptiform activity contributes to the difficulty of the task and, in some cases, seizures are not preceded by any specific patterns captured in the EEG [2]. Understanding the underlying dynamics and detecting this complex transition between the inter-ictal and ictal stages is the fundamental challenge in the field of seizure prediction.

2.3 Seizure Prediction

2.3.1 Main concepts

The main goal of seizure prediction is to develop tools able to forecast/anticipate the occurrence of seizures, based on on-line data, and notify the patient in a timely manner. Thus, in addition to providing an occurrence period, in which the seizure takes place, the algorithm should allow enough time for the patient to take preventive measures and avoid harmful consequences. In order to reduce stress levels, the number of false alarms must be strictly limited [11].

Seizure prediction vs. detection

A parallel, although different research field is seizure detection, whose algorithms aim at an early detection of the beginning (onset) of the seizure. These, in contrast to seizure prediction algorithms, do not provide time for the patient to take action, but can aid clinicians in identifying the epileptic focus and selecting appropriate medication. Both can also be implemented in closed-loop intervention systems, although prediction is preferable for more timely responses [81, 85].

Seizure onset

Two types of onset can be considered: clinical and electrographic. The former refers to the moment when the first clinical manifestations of the seizure become visible, while the latter relates to the first observable changes in the EEG. Given that symptoms typically become apparent later and may prove difficult to identify (as is the case of FOIA and non-motor seizures), it is reasonable to annotate the beginning of the seizure based on the EEG onset [81].

Seizure prediction characteristic

Since its early beginnings in the 1970s, seizure prediction research grew in interest and started to show promising results. However, by the turn of the century, evaluation of the developed algorithms did not achieve such optimistic findings [81].

Accordingly, in 2003, a "seizure prediction characteristic" was suggested by Winterhalder et al. [113] to assess and compare different seizure prediction methodologies, based on statistical, clinical and behavioral considerations. As such, two fundamental concepts were introduced: Seizure Occurrence Period (SOP) and Seizure Prediction Horizon (SPH), illustrated in figure 2.7.

While the ideal scenario would be to predict the exact moment of the onset, some uncertainty is to be expected. Hence, the SOP is defined as the period during which the seizure is presumed to occur. Moreover, to make therapeutic or behavioral intervention possible, it is essential do define a minimum time interval between the alarm generated by the prediction method and the beginning of SOP. This time window indicates the SPH, also known as Intervention Time (IT).

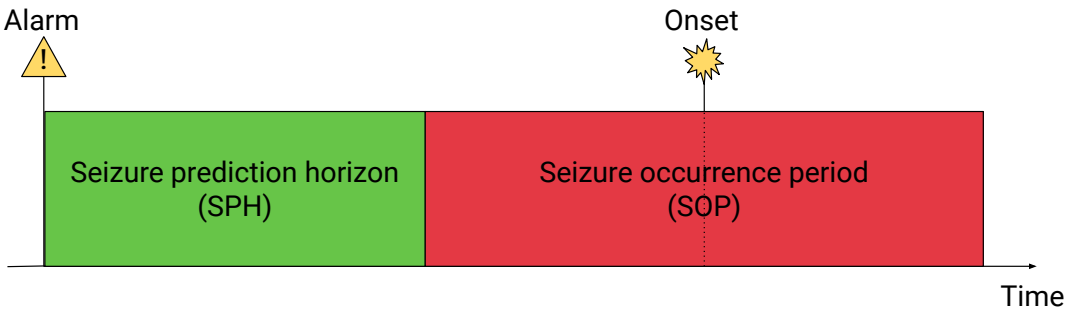


Figure 2.7: Visual representation of SOP and SPH. Adapted from: Winterhalder et al. [113]

Both periods must be accounted for when analyzing the prediction: for an alarm to be considered as true, the seizure onset must appear sometime during the defined SOP. In other words, if the onset is located outside of the SOP period, even if within the SPH, it is classified as a false alarm (i.e. false positive).

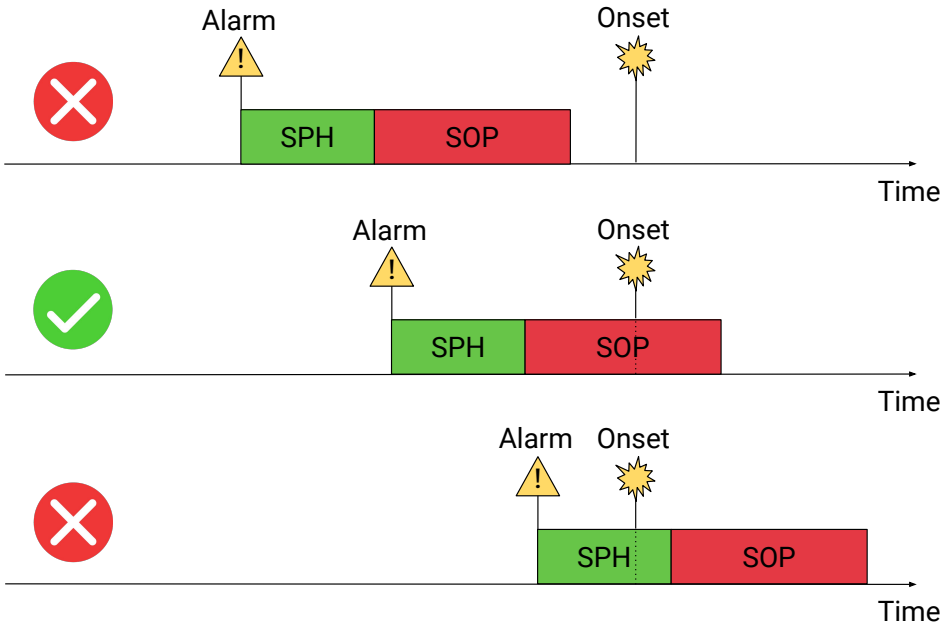


Figure 2.8: Visual representation of true and false alarms in seizure prediction.

In literature, SOP values have been found ranging from minutes to hours [81]. This can not only impact the performance of the algorithm but also the patient's life, as large time intervals may induce an overwhelming amount of stress/anxiety. Thus, both SOP and SPH values must be chosen according to their impact on the patient. For instance, let us consider the following example: if the chosen SOP value spans 6 hours, and a given patient has 4 seizures within a single day, then, technically, no prediction is being made despite the fact that the algorithm accurately anticipates all of the seizures.

Regarding lower/upper bounds for SPH and SOP, there are no standard values. Nonetheless, considering that intervention systems only require a few seconds to act, this could indicate a suitable value for SPH_{min} (albeit, for warning systems, a longer interval is required). Furthermore, future implanted devices (with drug delivery or electrical stimulation) may cause adverse side effects over prolonged intervention times, which could help to determine an appropriate value for SOP_{max}, in addition to the aforementioned effect on stress levels [81, 113]. It is worth mentioning that authors typically choose, in supervised learning approaches, an equal duration for $SOP + SPH$ as the assumed pre-ictal period, as depicted in figure 2.9.

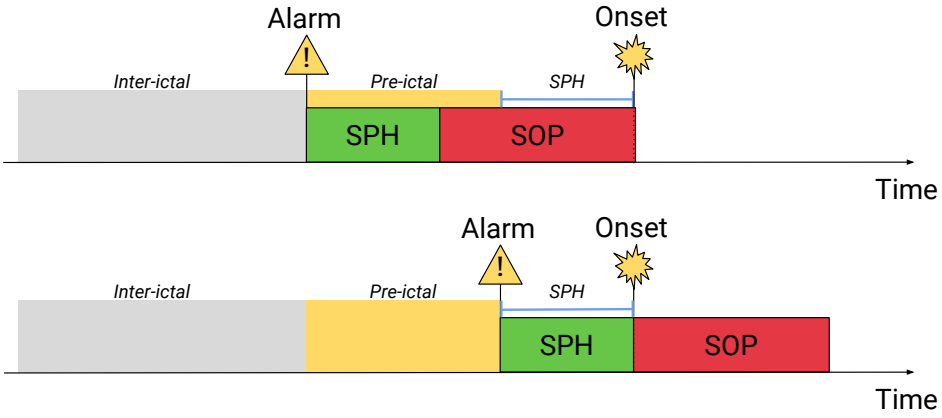


Figure 2.9: Visual representation of the relationship between SOP, SPH and the assumed pre-ictal period. Both alarms are correct, and are raised in the first and last pre-ictal sample, respectively, representing the two extreme cases.

2.3.2 Performance evaluation

In addition to the aforementioned concepts, two performance metrics were defined for seizure prediction: sensitivity and False Positive Rate per Hour (FPR/h). These follow the notion of true/false alarms as explained previously.

Sensitivity is, therefore, defined as the fraction of correctly predicted seizures, as described by equation (2.1):

$$Sensitivity = \frac{Predicted\ seizures}{All\ seizures}. \tag{2.1}$$

By using this definition, sensitivity can be analyzed in a more effective manner, whereas with the typical definition (equation (2.4), presented below), two accurately predicted seizures could display different values depending on how many instants of the time series were labeled correctly.

Regarding FPR/h, it is defined as the ratio between the number of false alarms and the total duration of the inter-ictal period:

$$FPR/h = \frac{N_{alarms}}{Interictal_{duration}}, \tag{2.2}$$

where N_{alarms} represents the number of false alarms and the $Interictal_{duration}$ can be computed by equation (2.3):

$$Interictal_{duration} = Recording_{duration} - N_{seizures} \times (SOP + SPH). \tag{2.3}$$

This definition, therefore, avoids the ambiguity of measuring the specificity of the prediction algorithm, which would be the portion of time, during the inter-ictal period, where the patient is not under a false alarm (false waiting time) [81, 111].

After analyzing the output of the algorithm and comparing it to clinical annotations on the EEG signal, one may define a confusion matrix as follows:

Table 2.1: Confusion matrix for evaluation of the seizure prediction performance.

		Clinical label	
		Pre-ictal	Inter-ictal
Output	Pre-ictal	TP	FP
	Inter-ictal	FN	TN

It should be stressed, however, that the previous concepts must not be envisioned as common machine learning metrics, such as the ones expressed below:

$$Sample\ Sensitivity = \frac{TP}{TP + FN} \tag{2.4}$$

$$False\ Positive\ Rate = 1 - Sample\ Specificity = 1 - \frac{TN}{TN + FP}. \tag{2.5}$$

In summary, seizure prediction can be seen as a binary problem, where the pre-ictal state corresponds to the positive class. The performance of the algorithm is assessed using sensitivity (equation (2.1)), to evaluate if seizures are effectively predicted, and FPR/h (equation (2.2)), which better reflects the impact of the number of false alarms, as exemplified in figure 2.10.

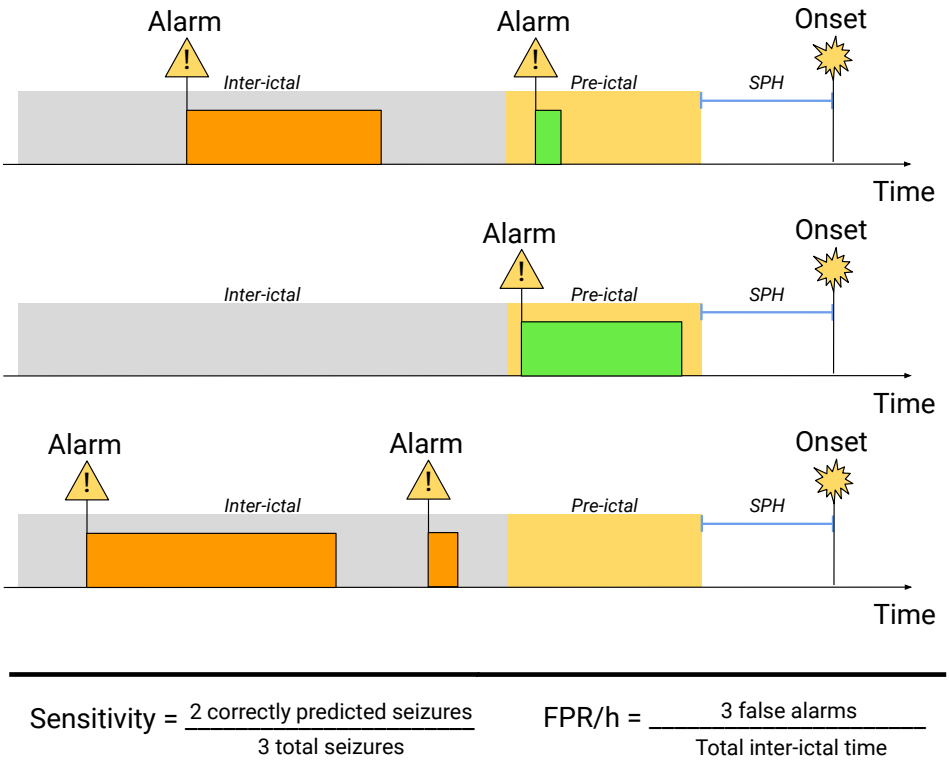


Figure 2.10: An example of performance assessment, using sensitivity and FPR/h.

It is worth mentioning that, generally, there is a trade-off relationship between both performance metrics: an increase in sensitivity leads to an increase in FPR/h. Setting a lower threshold to trigger an alarm increases the number of crossings during the pre-ictal stage (increasing sensitivity), but the same happens for the inter-ictal stage (increasing false positives and, thus, FPR/h) [113]. Hence, both metrics should be evaluated simultaneously, since an excessive number of false alarms may cause the patient to ignore warnings or it could result in side effects from unnecessary intervention.

Optimal metric values

Despite the lack of a standard metric value for FPR/h, a maximum value must be defined to establish a level of acceptable performance from the clinical point of view. Therefore, the average seizure incidence could indicate an appropriate range.

Thus, Winterhalder et al. [113] proposed a value of 0.15 for FPR/h_{max}, which corresponds to an average incidence of 3.6 seizures per day. This is the case for patients in pre-surgical monitoring, whose AED intake is decreased resulting in artificially frequent seizures. These also constituted, at the time, the majority of available databases of epilepsy recordings.

It should be noted that, under normal circumstances, patients with pharmacorefractory local epilepsy experience an average of three seizures per month, which corresponds to 0.0042 seizures per hour. As such, an FPR/h_{max} of 0.15 would mean that, even if all seizures can be predicted correctly, 50% would be false alarms during pre-surgical monitoring and, outside these conditions, the number would escalate to 97%. Hence, this maximum value can only be considered reasonable for patients in pre-surgical monitoring [81, 113].

2.3.3 Statistical validation

In addition to the previous aspects, statistical validation is another important issue to consider in the field of seizure prediction. For an algorithm to be considered valid, it must meet the minimum requirement of performing above chance [11, 81].

Random predictor

Schelter et al. [91, 92] proposed a random predictor based on a homogeneous Poisson process for the false predictions. In this approach, the probability of raising an alarm for any single sampling point of a feature extracted from a time series is:

$$P_{Poiss} = \frac{FP}{N}, \tag{2.6}$$

where FP and N are the number of false alarms and the number of samples, respectively. Considering a time period of duration equal to SOP and that the product of FPR/h_{max} and SOP is substantially lower than one (valid when assuming the patient is not under continuous warning), the probability of raising at least one alarm within SOP is given by equation (2.7):

$$P \approx 1 - e^{-FPR/h_{max}W} \approx FPR/h_{max} \times SOP. \tag{2.7}$$

This probability P forms the basis for a significance level for assessing whether the sensitivity $S(FPR/h_{max}, SOP, SPH)$ is greater than that of a random predictor.

Moreover, this significance level should take into account the fact that more than one seizure is being analyzed and that the probability of prediction by chance increases with the number of electrodes used and the degrees of freedom d (e.g. number of prediction models). Hence, in order to reflect these aspects, the probability of randomly predicting at least k out of K seizures follows a binomial distribution:

$$P_{binom,d}(k,K,P) = 1 - \left[\sum_{1}^{j\leq k} \binom{K}{j} P^j (1-P)^{K-j}\right]^d. \tag{2.8}$$

Finally, the critical value σ_{low} taken into account to test for statistical significance is given by (2.9) for a significance level α:

$$\sigma_{low} = \frac{argmax_k\{P_{binom,d}(k,K,P) > \alpha\}}{K} \times 100\%. \tag{2.9}$$

Interestingly, this assessment methodology and its analytic expressions provide information regarding the minimum number of seizures that the data must contain to guarantee that a performance above chance level can be demonstrated. Nevertheless, it operates assuming a homogeneous Poisson distribution for the false predictions, which might not be the case.

Surrogate data testing

Surrogate time series analysis is another proposed method to evaluate predictive performance. In this approach, the original seizure onset times are randomly shuffled in order to generate artificial ones. Afterwards, if the predictive performance of the algorithm is higher with statistical significance for the original onset times, then it can be said to outperform a random predictor [9, 11, 81].

Schelter et al. [91] have suggested techniques based on bootstrapping, including seizure-time and measure-profile surrogates: the former keeps the inter-seizure-interval distribution, while the latter keeps the original seizure times unchanged. This approach offers considerable flexibility in terms of testing different null hypotheses by composing suitable sets of assumptions and constraints, although great care must be taken during its implementation to avoid hidden bias. On this matter, Andrzejak et al. [8] stress the importance of defining an appropriate null hypothesis for the conjectures in the chosen methodology.

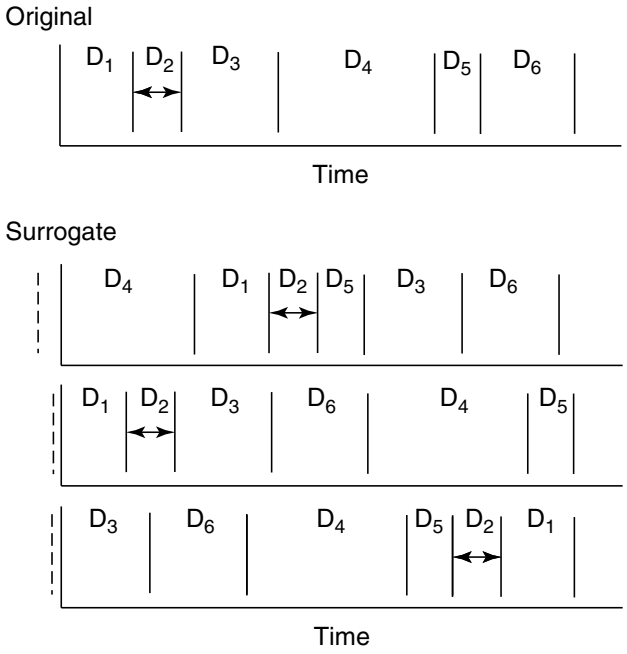


Figure 2.11: Original seizure times and the surrogate times bootstrapped from the inter-seizure intervals. The arbitrary onset times for the surrogates are obtained from a uniform distribution and are indicated by the dashed vertical lines. Source: Schelter et al. [91]

2.3.4 Concept drift and data imbalance

Seizure prediction algorithms face a problem found in several machine learning applications: when training a model from data, the concept of interest may depend on some hidden context that is not given explicitly in the form of predictive features. Changes in the hidden context can induce modifications in the target concept, which is commonly referred to as a concept drift. These changes may also cause shifts in the underlying data distribution (virtual concept drift), making it necessary to revise the current model [109].

In the context of epilepsy, this drift is a consequence of changes in brain dynamics due to the patient's daily routine, stressful situations or medication. The latter is especially relevant, bearing in mind that, as previously stated, numerous databases (e.g. EPILEPSIAE [60]) are mostly comprised of patients in pre-surgical monitoring, whose drug intake is reduced. Consequently, as the average seizure incidence is considerably lower outside this period [113], it may induce significant changes in the data distribution and impact the performance of the trained models. In particular, seizures are often absent during the first few hours in these recordings, while the effect of the AEDs is still wearing off, and become progressively more frequent thereafter.

Furthermore, class imbalance is another significant issue in seizure prediction, given the fact that the inter-ictal period is considerably longer than the pre-ictal one. As a result of this, the pre-ictal class is severely underrepresented, which introduces bias in the learning algorithm towards the majority class. Additionally, when learning from on-line data, the model may not receive positive class instances for long periods of time, impairing the learner's ability to adequately learn the pre-ictal class boundary [47].

Hence, it is desirable to develop robust classifiers that are able to remain stable and unchanged to irrelevant events (outliers) while being capable of handling changes (drifts) in concepts over time. Ensemble techniques and the inclusion of exogenous variables have been found to aid in this regard [47, 109].

2.4 Evolutionary Computing

2.4.1 Overview

Evolutionary computing is a research area within the field of computer science, concerned with a class of search algorithms based on the Darwinian principles of natural selection, drawing inspiration from molecular genetics [32]. These algorithms are designated as Evolutionary Algorithms (EAs), encompassing multiple variants: evolutionary programming, evolution strategies, genetic algorithms and genetic programming, which are all population-based metaheuristic methods [13].

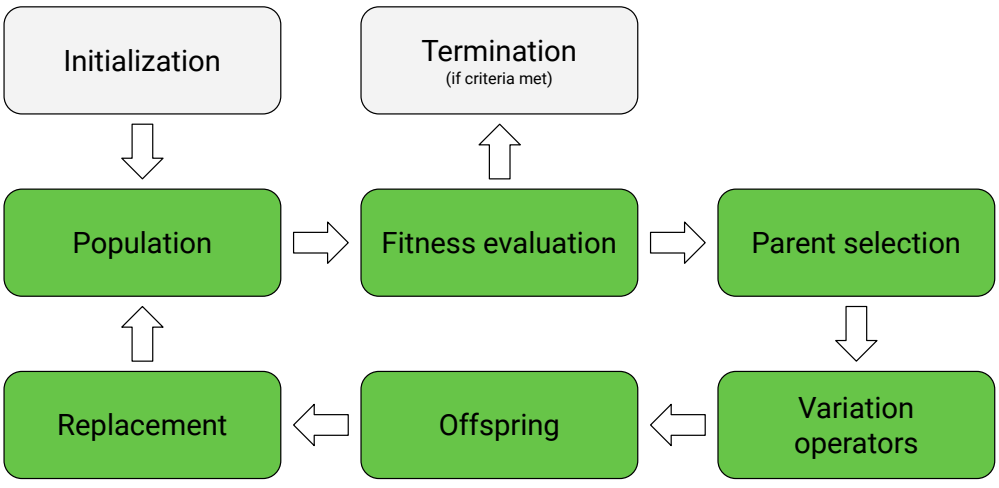


Figure 2.12: Flowchart of the typical evolutionary framework. Adapted from: Towards Data Science [105]

This set of algorithms is based on a common framework (figure 2.12), which can be summarized in the following steps:

- Initialization: an initial population of individuals is created, comprised of a diverse set of candidate solutions to a given problem.
- Parent selection: each individual is evaluated according to one or more fitness functions, measuring the quality of the solution (the higher, the better).
- Variation operators: after selecting a group of individuals based on their fitness scores (parents), these will be used as the basis for the next generation; new individuals (offspring) are created by altering or combining individuals from the previous generation.
- Termination: new generations are created until a candidate solution with sufficient fitness is found or a previously set computational limit is reached.

The previous steps describe the evolutionary process, which results in a population that should become progressively better adapted to the environment with each generation. In other words, the algorithm searches for increasingly better solutions for the problem at hand, navigating through the search space over each iteration.

Essentially, the main goal of EAs is to be applicable to a wide range of problems and deliver good (not necessarily optimal) solutions within acceptable time. Namely, instead of iterating through every possible solution, these algorithms explore a set of solutions which are generated and tested according to one or more fitness functions tailored to the given problem [12, 32, 33].

2.4.2 Main concepts

Representation

The first step in defining an EA is to connect the context of the problem at hand and the problem-solving space where evolution will take place. Thus, an object or solution in the original problem is referred to as phenotype, while its encoding within the EA is called genotype. The mapping from the phenotypes onto a set of corresponding genotypes is defined as the representation [33, 88].

Similarly to its biological counterpart, the mapping between the genotype space and phenotype space should be such that for each genotype there is at most one corresponding phenotype. In other terms, the same phenotype may be coded by different genotypes, but a genotype will only correspond to a single phenotype [32].

In order to make it possible to specify, store and evaluate candidate solutions in a way that can be manipulated by a computer, these can be represented in different

formats: in binary (as bit-strings), integers, real values, graphs, etc. When choosing a representation, it is important to choose the one that is best suited for the problem.

Fitness function

After mapping the space of all possible candidate solutions (phenotype space) to the space where the evolutionary search takes place (genotype space), it is necessary to define the requirements the population should adapt to meet. This is accomplished by using what is commonly referred to as the fitness function [13, 32].

The fitness function is a function that assigns an heuristic measure of quality to genotypes. Each solution computed in the genotype space is converted to its corresponding phenotype, evaluated according to the function and given a fitness score. This score will allow to distinguish between high-quality and low-quality individuals, which is essential to manage the population and improve the performance of later generations [12, 33].

Population

In EAs, the role of the population is to hold the representation of possible solutions to the problem. It is a multiset of genotypes (where copies of an individual are possible), which forms the unit of evolution. Each individual is a static object, that is, it does not change or adapt, it is the population that does [32].

In most applications, the population size is constant, remaining unchanged throughout the evolutionary search, which produces the limited resources needed to create competition. The fitness of the population is improved using two types of operators: selection operators aim to increase the mean quality of solutions in the population, while variation operators help to create diversity within the population, promoting novelty [13, 33].

Selection operators

The selection operators work at the population level: parent selection chooses the individuals which will form the basis for the next generation, while survivor selection (commonly referred to as replacement) decides which individuals will be allowed in to the following generation.

Parent selection is typically probabilistic, giving to high-quality individuals a greater chance of becoming parents than those with lower fitness scores. However, low-quality individuals have a small but positive probability, to prevent the search from becoming excessively "greedy", that is, to get stuck in a local optimum. Pop-

ular methods include: fitness proportional selection, in which the probability that an individual is selected depends on its relative fitness score (i.e. compared to the rest of the population); ranking, where the population is sorted based on fitness score and probabilities are assigned according to their rank; tournament selection, where individuals and their fitness score are compared in groups of randomly chosen individuals [12, 32, 44].

Survivor selection, on the other hand, is often deterministic and used in a different stage of the evolutionary cycle, that is, after the offspring is created from the selected parents. This selection can be made based on age (each individual only exists for a number of generations) or the individual's fitness score, for which there are different approaches, such as replacing the worst individuals or selecting survivors based on tournaments, among others [32, 96].

Variation operators

The role of variation operators is to create new individuals from old ones, generating new candidate solutions for the problem. There are two types of operators, classified according to the number of individuals involved: mutation (single individual) and recombination (two or more individuals).

Mutation is a unary variation operator, applying a small, random change to a genotype. It represents the smallest step in the search space of the problem, creating one child from a single parent. The intensity or magnitude of the mutation is regulated by parameters such as mutation rate or mutation probability (depending on the implementation) [32].

Recombination (also known as crossover) creates a new solution by merging the information contained within two (or more) parent genotypes. It is applied probabilistically, according to a crossover rate p_c: the offspring are created from two selected parents with probability p_c, or by simply copying one of the parents with probability $1 - p_c$ [13, 32].

Both types of variation operators depend on the chosen representation. For instance, when considering a binary representation, the mutation operator allows each bit to flip with a small probability, while in integer or real-valued representations, each gene has a probability of changing to another value from its domain or a small positive or negative value is added to its current value. Crossover operators, usually, either combine different parts of the parent genotypes (such as n-point crossover, depicted in figure 2.13) or create a new allele (gene value) that lies between those of the parents [13, 32, 33].

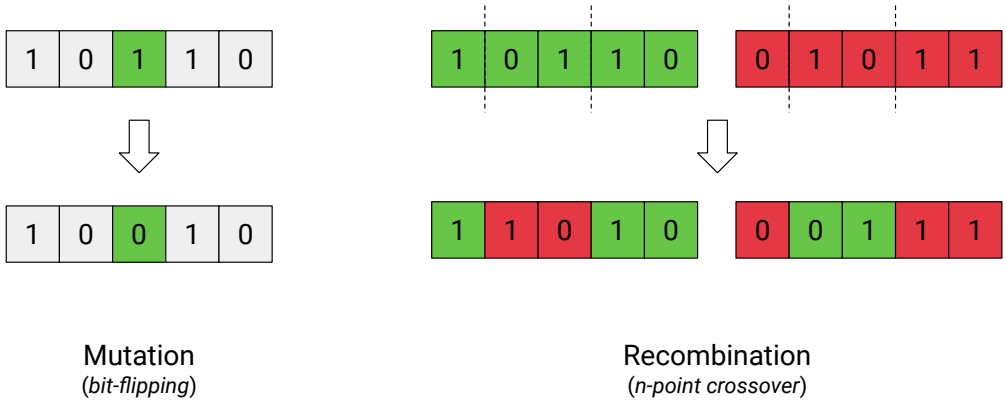


Figure 2.13: An example of mutation and recombination operators in EAs, using a binary representation.

Multi-Objective Evolutionary Algorithms (MOEAs)

MOEAs differ from their single-objective counterpart by the fact that the quality of each solution is defined by its fitness in relation to several, possibly conflicting objectives. Selection is then based on dominance (relative fitness) rather than on an absolute score. Given two solutions (A and B), with fitness scores (a_i and b_i) according to some set of n objective values, one solution is said to dominate the other if its score is at least as high for all objectives, and is strictly higher for at least one [32], as expressed in the following equation:

$$A \succeq B \iff \forall i \in \{1, ..., n\}, a_i \geq b_i \wedge \exists i \in \{1, ..., n\}, a_i > b_i. \tag{2.10}$$

This leads to the concept of the Pareto front (illustrated in figure 2.14), which is the set of all non-dominated solutions. The quality of each one of these solutions cannot be increased with respect to any of the objective functions without negatively affecting one of the others [13, 32].

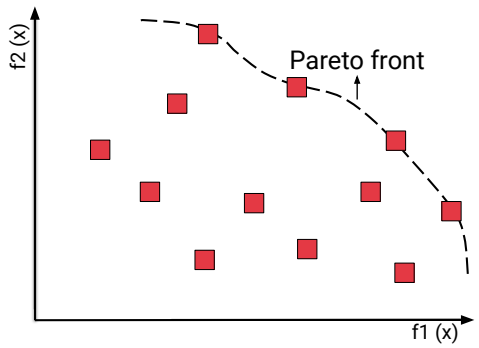


Figure 2.14: Illustration of the Pareto front (dashed line), assuming maximization for two objective functions ($f_1(x)$, $f_2(x)$).

The aim of MOEAs is, therefore, to distribute the population evenly along this front, in order to find a diverse set of high-quality solutions [22, 32].

Several approaches were suggested to solve multi-objective problems, such as Fonseca and Fleming's multi-objective genetic algorithm (MOGA) [41], Srinivas and Deb's non-dominated sorting algorithm (NSGA) [97] and Horn et al.'s niched Pareto genetic algorithm (NPGA) [48]. These have two common characteristics: assigning fitness based on non-dominated sorting and preserving diversity among solutions of the same non-dominated front.

These earlier algorithms, despite achieving good performance for a large range of problems, are heavily dependent on parameter tuning. In order to reduce this dependency and enhance the convergence properties of MOEAs, elitism was introduced [32]. Notably, the revised NGSA-II algorithm by Deb et al. [28] and SMS-EMOA by Beume et al. [18] accomplish this in different ways. The former ranks the non-dominated solutions in different classes (extreme solutions are given the highest rank) and sorts these within each class according to a distance measure, while the latter uses a concept called hypervolume, which measures the volume of the dominated portion of the objective space bounded by a reference point [17].

It is worth mentioning that the search ability of these algorithms is severely affected when the number of objectives is increased. In particular, various difficulties may arise: decrease of selection pressure (probability that the most fit individual is selected as a parent, controlling the rate of convergence) towards the Pareto front, exponential increase of the number of solutions required to approximate the entire Pareto front and difficulty of the visualization of solutions (which, in turn, makes the choice of a final solution more complicated) [54].

2.5 Summary

Epilepsy

Epilepsy is characterized by a considerable heterogeneity regarding types of seizure, types of epilepsy and epilepsy syndromes. In particular, a seizure can be described by its initial manifestations/symptoms, awareness and epileptic focus localization concerning lobes and/or hemispheres. TLE represents the most common form of focal epilepsy, defined by seizures with a temporal lobe focus. DRE patients, which do not achieve sustained seizure freedom through medication, represent the focus of seizure prediction given that they are exposed to the physical and social implications of the disease. This group of patients is often subjected to monitoring for

weeks/months to evaluate their condition before undertaking surgical interventions, hence why most databases are comprised of data acquired during this period.

EEG

The EEG measures and records the electrical activity of the brain, representing the primary physician tool to assess patients suffering from epilepsy. Its potentials can be categorized in two main types: oscillations and transients. The former refer to rhythmic patterns (with different frequency bandwidths) while the latter concern sharp transitions, which may be normal (e.g. eye blinks) or abnormal (which may be related to epilepsy).

Two acquisition techniques may be used: scalp EEG or iEEG. Despite the fact that iEEG recordings have a higher signal to noise ratio than the non-invasive approach, scalp EEG recordings can capture low-frequency activity more accurately. It is worth noting that iEEG, as an invasive method, presents a considerable risk (up to 4%) of infection/hemorrhage. In either case, the EEG is an approximation of the true complexity of brain activity.

Considering that not all of the captured epileptiform activity can anticipate seizures, accurate predictions require a thorough analysis of all types of EEG activity.

Seizure prediction

For supervised learning approaches, it is essential to divide each seizure episode into pre-ictal, ictal, post-ictal and inter-ictal. The primary goal of seizure prediction is to detect the pre-ictal period and correctly anticipate a seizure, providing an occurrence period (SOP) and intervention time (SPH). However, the fact that this transitional stage differs between patients and between seizure episodes presents a major challenge.

The gold standard metrics are sensitivity (percentage of correctly predicted seizures) and FPR/h. Suitable methodology must concern an adequate SPH duration (long enough for patients to take preventive measures) and evaluate performance for a range of SOP values (which should not be overly long as to not induce stress/anxiety). Furthermore, statistical validation should be conducted where performing above chance is the minimum requirement. Finally, proposed approaches should handle the presence of concept drifts and data imbalance.

Evolutionary computing

Evolutionary approaches are population-based metaheuristic search algorithms, drawing inspiration from the Darwinian principles of natural selection. A population of candidate solutions is generated and becomes progressively better adapted to the environment, that is, the algorithm navigates through the search space finding increasingly better solutions for the given problem. Instead of iterating through every possible solution, EAs operate in a more flexible way, exploring sets of solutions based on fitness functions tailored to the problem.

Each solution in the problem space is referred to as phenotype, while its encoding within the EA is called genotype. The mapping between these spaces is called representation, which can be made using different formats (bit-strings, real values, etc.). This should be chosen according to the suitability to the problem.

Selection and variation operators are used to improve the fitness (heuristic measure of quality) of the population: the former aim to increase the mean quality of the solutions, while the latter help to create diversity. Variation operators include mutation, that introduces small changes to a single individual, and recombination, which creates new solutions from two individuals (the parents).

Multi-objective approaches have been proposed with the aim of optimising solutions for problems with conflicting objectives. Instead of measuring quality based on a single fitness function, solutions are compared on the basis of dominance, that is, relative fitness to a set of objective functions.

The benefits of adopting EAs become clear when one envisions machine learning as a search problem. Considering that the main goal is to discover a model which can approximate an unknown mapping function from inputs to outputs, many decisions throughout its design narrow down the space of all possible models: choosing which data to train on, the algorithm's configuration, etc. By using search algorithms, one can refine how this space of possible mappings is explored as the model is fit, in order to take full advantage of the chosen training data. Particularly, adopting MOEAs enables one to explore different trade-offs and how they influence the classification model in a more flexible way than standard machine learning training.

3

State of the Art

This chapter provides an overview of the current state of the art in seizure prediction based on the Electroencephalogram (EEG) signal and machine learning. Firstly, the general framework is presented in section 3.1. Sections 3.2 to 3.5 review conventionally used techniques and common features, followed by sections 3.6 to 3.8 which examine classification, regularization and performance evaluation methods. Finally, a summary of the main aspects and a small discussion are provided in section 3.9.

3.1 Pipeline overview

Current algorithmic seizure prediction methodologies follow, in general, a set of steps primarily concerning signal processing and machine learning, as illustrated in the figure below:

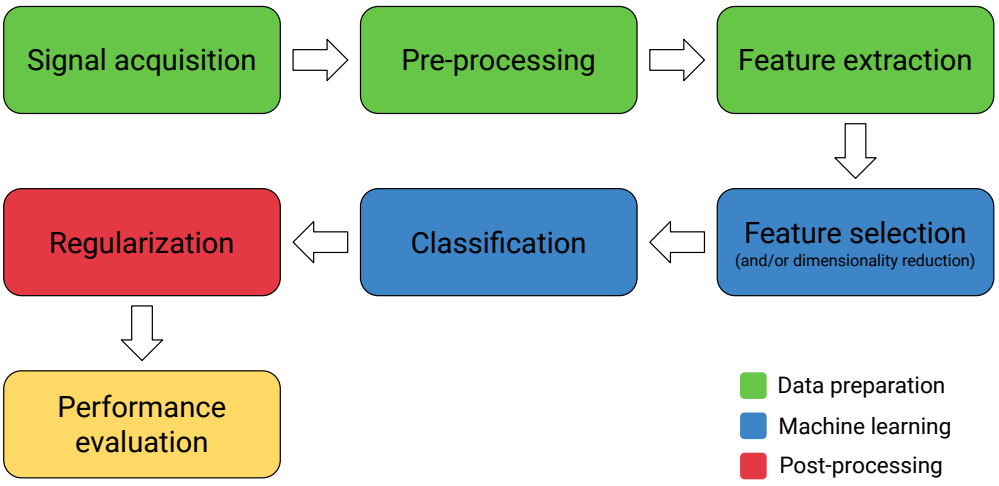


Figure 3.1: Color-coded flowchart of the typical seizure prediction framework. Adapted from: Bou Assi et al. [11].

Concisely, after collecting the EEG recordings, each step can be described as follows:

- Pre-processing: enhancing the signal quality, usually through filtering, followed by segmentation using sliding window analysis.
- Feature extraction and selection: collecting features from the time series data and selecting the ones which better discriminate each epileptic state.
- Classification: training machine learning models using the previously selected set of features.
- Regularization: smoothing the classifier output with post-processing methods.

After the development of the algorithm, its performance is then assessed using sensitivity, False Positive Rate per Hour (FPR/h) and statistical validation (against a random predictor or surrogate time series analysis) [11].

Differences in Deep Learning approaches

Recently, with the increase in computational power and available data, more advanced machine learning models have surfaced, referred to as Deep Learning. These have become the state of the art in many fields, and seizure prediction is no exception.

These models are representation learning methods, which when given raw data can automatically determine the representations needed for detection or classification tasks. With multiple layers of representation, complex functions can be learned and, in the case of classification, certain aspects of the input that are important for discrimination can be amplified, in addition to suppressing irrelevant variations [69].

In the context of the seizure prediction pipeline, the main modifications introduced by these methods are present in the feature extraction, feature selection/dimensionality reduction and classification stages. These will be analyzed and discussed throughout each of the following sections.

3.2 Signal acquisition

Considering that the EEG is a significantly complex signal, the chosen type(s) of recordings can greatly influence the results. Furthermore, the data set used to train and evaluate the proposed methodologies can play a major role. Table 3.1 presents an overview of these aspects concerning studies from the past 20 years of research in EEG seizure prediction.

Table 3.1: Overview of the signal acquisition characteristics from EEG seizure prediction studies of the past 20 years.

Study	Database	Patients	Electrodes	Signals
Daoud et al. (2019) [27]	CHB-MIT	8 epileptic	23 (reduced after selection)	Scalp EEG
Khan et al. (2018) [57]	CHB-MIT, MSSM	28 epileptic	22	Scalp EEG
Truong et al. (2018) [107]	Freiburg, CHB-MIT, Kaggle (American Epilepsy Society)	28 humans, 2 dogs	6, 22, 16 (respectively)	Scalp EEG, iEEG
Tsiouris et al. (2018) [108]	CHB-MIT	23 epileptic	18	Scalp EEG
Agarwal et al. (2018) [4]	Kaggle (UPenn and Mayo Clinic)	8 epileptic	All electrodes	iEEG
Chamseddine et al. (2018) [20]	Kaggle (American Epilepsy Society)	1 epileptic dog	16	iEEG
Sun et al. (2018) [99]	Kaggle (American Epilepsy Society)	2 humans, 5 dogs	16	iEEG
Aggarwal et al. (2017) [5]	Personal	10 epileptic, 2 healthy	18	Scalp EEG
Kiral-Korkek et al. (2017) [59]	NeuroVista	10 epileptic	16	iEEG
Direito et al. (2017) [30]	EPILEPSIAE	216 epileptic	F7, FZ, F8, T5, PZ, T6 6 random 6 in focal region	Scalp EEG, iEEG
Bandarabadi et al. (2015) [16]	EPILEPSIAE	24 epileptic	3 in focal region and 3 far from local region	Scalp EEG, iEEG
Bandarabadi et al. (2015) [14]	EPILEPSIAE	18 epileptic	2 in focal region	Scalp EEG, iEEG
Bou Assi et al. (2015) [10]	Kaggle (American Epilepsy Society)	5 epileptic dogs	16	iEEG
Rasekhi et al. (2015) [87]	EPILEPSIAE	10 epileptic	3 in focal region and 3 far from local region	Scalp EEG, iEEG
Alvarado-Rojas et al. (2014) [7]	EPILEPSIAE	53 epileptic	All electrodes	iEEG
Teixeira et al. (2014) [104]	EPILEPSIAE	278 epileptic	F7, FZ, F8, T5, PZ, T6 6 random 6 in focal region	Scalp EEG, iEEG
Moghim et al. (2014) [74]	Freiburg	21 epileptic	6	iEEG
Rabbi et al. (2013) [84]	EPILEPSIAE	1 epileptic	2 channels in epileptic region	iEEG
Rasekhi et al. (2013) [86]	EPILEPSIAE	10 epileptic	3 in focal region and 3 far from local region	Scalp EEG, iEEG
Bandarabadi et al. (2012) [15]	EPILEPSIAE	12 epileptic	3 in focal region and 3 far from local region	Scalp EEG, iEEG
Valderrama et al. (2012) [110]	EPILEPSIAE	12 epileptic	All electrodes	Scalp EEG, iEEG, ECG
Teixeira et al. (2012) [103]	EPILEPSIAE	10 epileptic	3 in focal region and 3 far from local region	Scalp EEG
Acharya et al. (2012) [1]	Bonn	5 epileptic, 5 healthy	N.A.	iEEG
Direito et al. (2012) [29]	EPILEPSIAE	10 epileptic	3 in focal region and 3 far from local region	Scalp EEG
Direito et al. (2011) [31]	EPILEPSIAE	3 epileptic	Post-selection based on feature discriminative power	Scalp EEG
Park et al. (2011) [82]	Freiburg	18 epileptic	6	iEEG
Chisci et al. (2010) [21]	Freiburg	9 epileptic	3 in focal region and 3 far from local region	iEEG
Mirowski et al. (2009) [73]	Freiburg	21 epileptic	3 in focal region and 3 far from local region	iEEG
Adeli et al. (2007) [3]	Bonn	5 epileptic, 5 healthy	N.A.	iEEG
Mormann et al. (2005) [78]	Bonn	5 epileptic	All electrodes	iEEG
Le Van Quyen et al. (2005) [68]	Bonn	5 epileptic	All electrodes	iEEG
D'Alessandro et al. (2003) [26]	Personal	4 epileptic	All electrodes	iEEG
Mormann et al. (2003) [77]	Bonn	18 epileptic	N.A.	iEEG
Mormann et al. (2003) [76]	Bonn	10 epileptic	N.A.	iEEG
Le Van Quyen et al. (2001) [67]	Personal	23 epileptic	21 or 27	Scalp EEG, iEEG
Geva et al. (1998) [43]	Personal	25 epileptic rats	N.A.	Scalp EEG

Databases

Several databases have become available, where the ones from the University of Freiburg [21, 73, 82], the University of Bonn [1, 3, 76–78], the European Database on Epilepsy (EPILEPSIAE) [7, 14–16, 29–31, 52, 60, 84, 86, 87, 103, 104, 110] and, more recently, the Children's Hospital Boston (CHB-MIT) [27, 57, 107, 108] and the Kaggle American Epilepsy Society [10, 20, 99] databases are the most widely used.

Notably, the EPILEPSIAE database [60] is significantly larger than the remaining, containing recordings lasting 165 hours (on average) from 275 Drug-Resistant Epilepsy (DRE) patients in pre-surgical monitoring, in addition to extensive metadata which enables stratification. The NeuroVista database curated by Cook et al. [24] is also worth noting for containing long-term recordings (up to two years) from 15 patients outside of monitoring units, therefore representing real-life data which should benefit research concerning concept drift and facilitate clinical translation.

With the exception of Teixeira et al. [104], Direito et al. [30] and Alvarado-Rojas et al. [7], no other studies included more than 30 patients. Additionally, some studies have included non-epileptic humans [1, 3, 5] or epileptic animals such as dogs [10, 20, 99, 107] or rats [43]. The vast majority worked on pre-surgical monitoring data which, as mentioned in section 2.3.4, restricts applicability [42].

Electrode selection

Different approaches have been made regarding electrode selection. While many studies choose to work with all available electrodes, others choose a number of electrodes from the focal region and others far from it, or even select only some from the focal region.

This step leads to different assumptions regarding the seizure generation process. By choosing random electrodes, for instance, one assumes that this process can be captured at any brain location. On the other hand, when choosing electrodes only from the focal region it is presumed that the activity in this region is enough to capture the process. When electrodes are chosen from the focal region and others far from it, the assumption is that it is necessary to relate the information from the focal region to other regions of the brain. None of these assumptions, thus far, have proven to be more correct. Hence, choosing all the available electrodes can be intuitive, considering it contains more information, despite the increase in computational cost and patient discomfort [75].

Recording type

Both types of recording (scalp EEG and intracranial EEG (iEEG)) have been considered. Several studies have compared their performance, but no conclusions could be reached given that results were either statistically non-significant, limited by the reduced number of subjects or lacked validation.

While iEEG can be considered more suitable for chronic intervention devices due to its proximity to the brain and reduced number of artifacts, scalp EEG is able

to provide information regarding general brain state rather than localized data [11].

3.3 Signal pre-processing

Given that the main goal is to develop a tool which is able to receive on-line data and process it in real-time, the proposed methods must take into account its real-life feasibility. Therefore, the first step is, typically, data segmentation by sliding window analysis. Afterwards, in order to enhance signal quality, other steps may be employed such as filtering and/or artifact removal. Lastly, the pre-ictal period, Seizure Occurrence Period (SOP) and Seizure Prediction Horizon (SPH) are defined (although they may be handled later during classification, their range should be specified before the machine learning stage for the sake of not being influenced by performance results).

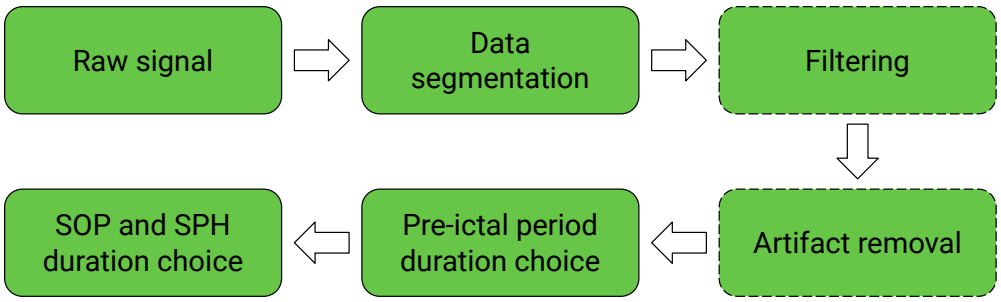


Figure 3.2: Flowchart of the typical signal processing pipeline in seizure prediction. Optional steps are represented with a dashed line. Defining the pre-ictal period is required if a supervised learning approach is used.

In table 3.2 an overview is given concerning signal processing steps in different studies. Generally, no significant efforts are made in filtering and artifact removal operations, considering that the EEG signal is complex and not fully understood. Sliding window analysis is characterized by time length and percentage of overlap between consecutive windows, with considerable uniformity among studies. Pre-ictal duration, required for supervised learning approaches (as illustrated in figure 3.2), is an aspect where authors tend to vary. Chosen SOP duration (assumed as equal to the pre-ictal period duration, as explained next) is heterogeneous and most studies do not present performance evaluation for the entire range of selected values. SPH duration is often omitted in most studies, as the vast majority of authors only consider a minimal time interval of a few seconds which, unless the goal is to implement it in a closed-loop intervention system, may be unrealistic in terms of applicability, given that patients may require more reasonable intervention time to take effective preventive measures before the seizure onset [11, 62].

Table 3.2: Overview of the signal processing characteristics from EEG seizure prediction studies of the past 20 years.

Study	Sliding Window	Filtering	Pre-ictal period duration (minutes)	SPH duration (minutes)[#]
Daoud et al. (2019) [27]	N.A.	None	60	N.A.
Khan et al. (2018) [57]	1 second 0% overlap	0 - 128Hz band-pass	10	N.A.
Truong et al. (2018) [107]	30 seconds Overlap N.A.	50 and 60Hz notch DC (0Hz) removed	30	5
Tsiouris et al. (2018) [108]	5 seconds 0% overlap	None	15, 30, 60, 120	N.A.
Agarwal et al. (2018) [4]	1 second 0% overlap	1 - 47Hz band-pass	N.A.	N.A.
Chamseddine et al. (2018) [20]	5 seconds 0% overlap	60Hz notch 0 - 190Hz band-pass	60	N.A.
Sun et al. (2018) [99]	30 seconds 0% overlap	Resampling 0.1 - 180Hz band-pass	60	5
Aggarwal et al. (2017) [5]	5 seconds 0% overlap	50Hz notch 0.5 - 50Hz band-pass	60	4 - 10
Kiral-Korkek et al. (2017) [59]	N.A.	N.A.	15 (before real-time tunability)	Minutes to hours
Direito et al. (2017) [30]	5 seconds 0% overlap	50Hz notch	10, 20, 30, 40	1/6
Bandarabadi et al. (2015) [16]	5 seconds 0% overlap	50Hz notch	10, 20, 30, 40	N.A.
Bandarabadi et al. (2015) [14]	8 seconds 50% overlap	50Hz notch	10, 30, 50, 70	N.A.
Bou Assi et al. (2015) [10]	5 seconds 0% overlap	50Hz notch 0.5 - 180Hz band-pass	60	5
Rasekhi et al. (2015) [87]	5 seconds 0% overlap	50Hz notch	10, 20, 30, 40	N.A.
Alvarado-Rojas et al. (2014) [7]	5 seconds 0% overlap	Bands of interest from 0.5Hz to 140Hz	10, 30, 60	N.A.
Teixeira et al. (2014) [104]	5 seconds 0% overlap	50Hz notch	10, 20, 30, 40	1/6
Moghim et al. (2014) [74]	5 and 9 seconds Overlap N.A.	Artifact removal with EEGLAB	5	N.A.
Rabbi et al. (2013) [84]	10 seconds 50% overlap	60Hz notch 0.5 - 100Hz band-pass	15, 30, 45	N.A.
Rasekhi et al. (2013) [86]	5 seconds 0% overlap	50Hz notch	10, 20, 30, 40	N.A.
Bandarabadi et al. (2012) [15]	5 seconds 0% overlap	None	10, 20, 30, 40	N.A.
Valderrama et al. (2012) [110]	5 seconds 0% overlap	N.A.	5, 10, 20, 30, 45, 60	N.A.
Teixeira et al. (2012) [103]	5 seconds 0% overlap	50hz notch	10, 20, 30, 40	N.A.
Acharya et al. (2012) [1]	23.6 seconds 20% overlap	0.5 - 85Hz band-pass*	N.A.	N.A.
Direito et al. (2012) [29]	5 seconds 0% overlap	50Hz notch	Statistically selected for each patient (2 - 60)	2
Direito et al. (2011) [31]	5 seconds 0% overlap	None	30, 40	N.A.
Park et al. (2011) [82]	20 seconds 50% overlap	Artifact removal 50Hz and 100Hz notch	30	N.A.
Chisci et al. (2010) [21]	2 seconds 0% overlap	50Hz notch	15	N.A.
Mirowski et al. (2009) [73]	5 seconds Overlap N.A.	50Hz notch 0.5 - 120Hz band-pass	50	N.A.
Adeli et al. (2007) [3]	N.A.	50Hz notch 0.5 - 60Hz band-pass*	N.A.	N.A.
Mormann et al. (2005) [78]	17 - 20.5 seconds 0% overlap	0.5 - 85Hz band-pass*	5, 30, 120, 240	N.A.
Le Van Quyen et al. (2005) [68]	5 seconds 0% overlap	Bands of interest from 0.5Hz to 30Hz*	30	N.A.
D'Alessandro et al. (2003) [26]	10 seconds 25% overlap	60Hz notch 0.1 - 100Hz band-pass	10	10
Mormann et al. (2003) [77]	23.6 seconds 20% overlap	0.5 - 85Hz band-pass*	N.A.	N.A.
Mormann et al. (2003) [76]	23.6 seconds 20% overlap	0.5 - 85Hz band-pass*	N.A.	N.A.
Le Van Quyen et al. (2001) [67]	30 seconds 0% overlap	0.5 - 99Hz band-pass 0.1 - 70Hz band-pass	60	N.A.
Geva et al. (1998) [43]	1 second 0% overlap	50Hz notch 1 - 30Hz band-pass	5	N.A.

* Pre-processed data from the University of Bonn.

N.A. assumes the duration of 1 or 2 sliding windows.

Data segmentation

With the purpose of simulating an on-line time series scenario, the EEG signal is segmented and analyzed in short windows to extract features chronologically. Window length varies between 1 to 30 seconds among the selected studies, while overlap percentage is commonly 0% or 50%.

The choice of window length and amount of overlap is primarily influenced by the trade-off between computational cost and execution speed. Considering the number of electrodes, sampling frequency and recording duration, many authors have chosen to use a 5-second window with no overlap, which is considered a compromise between capturing specific patterns and signal stationarity assumptions [11].

Filtering and artifact removal

In general, this step includes the removal of powerline interference (50Hz or 60Hz), band-pass filtering and removal of abnormal transients (considered artifacts). Signal decomposition into frequency bands of interest or wavelet coefficients may also be included, since all EEG activity except oscillations is removed.

Both Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) time-domain filters have been widely used. While FIR ones induce a linear phase response and allow zero-phase distortion, IIR filters have shown to not cause significant ripple within EEG frequencies of interest [11].

Cut-off frequencies for the band-pass filter differ between authors. Generally, frequency components below 0.5Hz are eliminated, which are considered breathing artifacts. Regarding the high-frequency cut-off, after which it can be considered as noise, some studies have investigated the discriminative ability of low- to high-gamma activity (as high as 500Hz), such as the one by Alvarado-Rojas et al. [7].

Pre-ictal period, SOP and SPH duration

No optimal or standard pre-ictal period duration has been established so far. Thus, authors have adopted fixed periods ranging from 5 minutes to 4 hours, or experimented with different periods [7, 14–16, 29–31, 78, 84, 86, 87, 103, 104, 108, 110]. In particular, Teixeira et al. [104] reported a decrease in FPR/h with no significant differences in sensitivity for longer periods, although the regularization method may have contributed to this.

This step significantly contributes to the difficulty in evaluating and comparing different seizure prediction methodologies. Several authors only state the chosen

pre-ictal period P used for supervised learning which, as illustrated in figure 2.9, can be related to the duration of SOP and SPH:

$$P = SOP + SPH. \tag{3.1}$$

Since SPH duration is commonly not mentioned, one must assume a minimal value corresponding to the sliding window length (particularly, the last one before the seizure onset), as it theoretically maximizes sensitivity. Equation (3.1) can then be approximated to:

$$P \approx SOP \; for \; SPH \approx 0. \tag{3.2}$$

Additionally, given that most studies only report the optimal values for the optimal P, fair comparisons cannot be made between methods when the corresponding values of P, SOP and SPH differ.

3.4 Feature extraction

Feature extraction constitutes the most heterogeneous step as researchers have proposed numerous approaches. Nevertheless, no specific type of features has been determined as optimal.

In general, the extracted features aim to capture three characteristics representative of seizure activity: an increase in energy (caused by electrical discharge in the brain), a shift in spectral content from low to high frequencies and an increase of synchronization in neuronal activity.

Features from the EEG may be extracted on a single or multi-channel basis. Single-channel analysis is carried out by selecting a given electrode and it is based primarily in local activity measures, while a multi-channel approach incorporates information from two or more electrodes. These analyses may be performed, respectively, through univariate and multivariate features [6, 11]. Some authors have reported increased predictive performance when combining both types [78].

Additionally, features may also be classified as linear or non-linear. Performance differences between these types have been addressed by several studies [45, 72, 78], but no conclusion can be drawn due to conflicting results. Nonetheless, non-linear features may be limited in terms of real-life applicability considering their increased computational cost [11, 82].

The various features extracted from the EEG can then be grouped according to their linearity and whether they were computed from one or more channels, as depicted in figure 3.3.

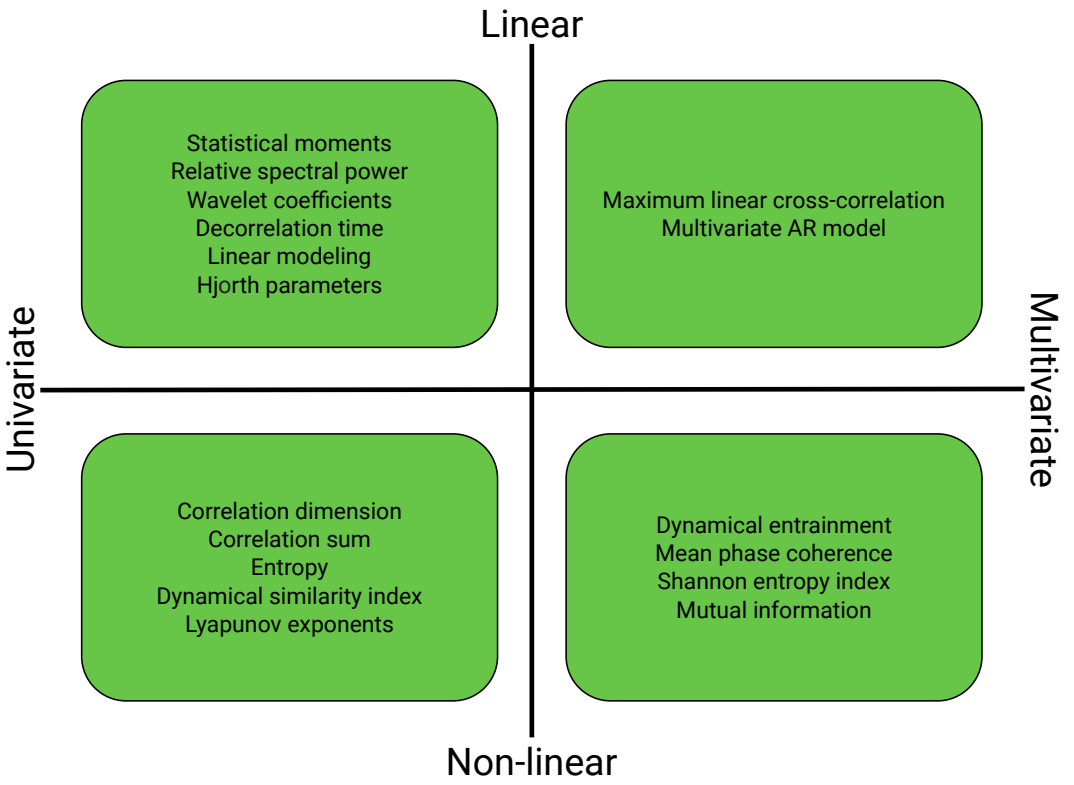


Figure 3.3: Categorization of common features used in seizure prediction, according to their linearity and whether they are univariate/multivariate.

Table 3.3 presents a summary of linear and non-linear features used in studies from the past 20 years. In general, linear univariate features are the most commonly used in the selected set of studies. Regarding multivariate features, authors tend to mainly choose non-linear measures.

The preference for linear univariate features results from being computationally lighter and simpler to interpret clinically. Given that multivariate features are, in general, heavier than univariate ones, they require additional processing power. As such, authors with enough computational resources for multivariate features are, usually, able to handle the additional requirements of non-linear multivariate ones, as the increase in computational cost is relatively small.

It is worth noting that various studies from recent years, which adopt Deep Learning approaches, have chosen to employ automatic feature engineering methods using time series data [4, 27, 59]. However, several still perform traditional feature engineering, primarily with features based on frequency band or wavelet decomposition, and use these as input for the classification models [4, 20, 57, 73, 99, 107, 108]. In terms of interpretability, traditional features are preferable when compared to the automatically extracted ones.

Table 3.3: Overview of univariate and multivariate feature extraction from EEG seizure prediction studies of the past 20 years.

	Linear univariate features							Non-linear univariate features				Linear multivariate features		Non-linear multivariate features			
Study	**Statistical moments**	**Spectral band related**	**Wavelets**	**Linear modelling**	**Energy**	**Hjorth parameters**	**Decorrelation time**	**Correlation dimension/sum**	**Lyapunov exponent**	**Dynamic Similarity Index**	**Entropy**	**Ratio**	**Correlation**	**Dynamical entrainment**	**Mean Phase Coherence**	**Entropy**	**Synchrony**
Daoud et al. (2019) [27]																	
Khan et al. (2018) [57]			•														
Truong et al. (2018) [107]																	
Tsiouris et al. (2018) [108]	•	•	•		•		•						•				
Agarwal et al. (2018) [4]		•											•				
Chamseddine et al. (2018) [20]		•															
Sun et al. (2018) [99]		•															
Aggarwal et al. (2017) [5]															•		
Kiral-Korkek et al. (2017) [59]																	
Direito et al. (2017) [30]	•	•	•	•	•	•	•										
Bandarabadi et al. (2015) [16]		•										•					
Bandarabadi et al. (2015) [14]		•															
Bou Assi et al. (2015) [10]		•				•	•										
Rasekhi et al. (2015) [87]	•	•	•	•	•	•	•					•					
Alvarado-Rojas et al. (2014) [7]		•															
Teixeira et al. (2014) [104]	•	•	•	•	•	•	•										
Moghim et al. (2014) [74]	•	•	•		•			•	•								
Rabbi et al. (2013) [84]										•					•		•
Rasekhi et al. (2013) [86]	•	•	•	•	•	•	•										
Bandarabadi et al. (2012) [15]		•										•					
Valderrama et al. (2012) [110]	•	•	•		•	•	•										
Teixeira et al. (2012) [103]	•	•	•	•	•	•	•				•						
Acharya et al. (2012) [1]											•						
Direito et al. (2012) [29]		•															
Direito et al. (2011) [31]	•	•	•	•	•	•	•										
Park et al. (2011) [82]		•															
Chisci et al. (2010) [21]				•													
Mirowski et al. (2009) [73]													•	•			•
Adeli et al. (2007) [3]			•					•	•								
Mormann et al. (2005) [78]	•	•				•	•	•	•				•		•	•	
Le Van Quyen et al. (2005) [68]																	•
D'Alessandro et al. (2003) [26]			•		•						•					•	
Mormann et al. (2003) [77]															•		
Mormann et al. (2003) [76]													•		•		
Le Van Quyen et al. (2001) [67]										•							
Geva et al. (1998) [43]	•		•										•				

In the following sections, a general description of the most common features of each type is given. For a more in-depth review, please refer to Appendix A.

Linear univariate features

These features correspond to mathematical measures computed from phase/ frequency and amplitude information of the signal, that comply with the linearity property. Linear feature extraction assumes the quasi-stationarity of the EEG within each sliding window.

The most widely used features in terms of time-domain analysis of the EEG are the first four statistical moments (mean, variance, skewness, kurtosis) due to their simplicity and linearity properties [15, 30, 31, 43, 74, 78, 86, 87, 103, 104, 108, 110]. However, other measures concerning the signal's dynamics have also been adopted by several studies. Accumulated energy, for instance, operates under the assumption that seizure-generation processes lead to an increase in brain activity, as reported by Rasekhi et al. [86, 87], Valderrama et al. [110], Teixeira et al. [103, 104], Moghim et al. [74] and D'Alessandro et al. [26]. The three Hjorth parameters (activity, mobility and complexity) have also been used to capture differences in signal dynamics between the pre-ictal and inter-ictal stages, as shown in various studies [10, 30, 78, 86, 87, 103, 104, 110].

Moreover, to evaluate neuronal synchronization, other features have been employed. Decorrelation time, which corresponds to the first zero-crossing of the autocorrelation function (see A.8) of the signal, can serve as an indicator of periodicity in neural activity [10, 30, 31, 78, 86, 87, 103, 104, 108, 110, 110]. Autoregressive (AR) models have also been used to model the EEG, where their predicted output is a weighted sum of previous signal values. Authors have used either the modelling error [30, 31, 86, 87, 103, 104], which has been found to increase due to pre-ictal changes in brain activity, or the values of modeling coefficients as features [21].

Concerning the spectral dynamics of the EEG, different methods have been adopted. In particular, under the assumption that seizure-generation processes lead to changes in the normal rhythmic brain activities [78, 82], several authors have decomposed the signal into different frequency bands (delta, theta, alpha, beta, gamma) and computed their relative spectral power [7, 10, 15, 16, 29–31, 74, 78, 82, 86, 87, 103, 104, 108]. This is done by dividing the power in the sub-band and the total power in the signal, based on the signal's Power Spectral Density (PSD) (see A.5). Another common form of analyzing the EEG's spectral content is to compute its Discrete Wavelet Transform (DWT), which is able to reflect both frequency and temporal location properties of the signal [74]. The resulting wavelet coefficients

have been used by several studies to compute other measures [3, 26, 30, 57, 74, 86, 87, 104, 108, 110] or, in the case of Deep Learning approaches by Khan et al. [57] and Tsiouris et al. [108], wavelet-transformed EEG signals were used as input for their models. Additionally, the spectral edge power has been adopted by various authors which reported a power transfer from low to high frequencies during the pre-ictal stage [10, 30, 31, 78, 86, 87, 103, 104, 110].

Non-linear univariate features

New features were introduced by dynamical system theory, such as correlation dimension [3, 74], Lyapunov exponents [3, 74, 78], the dynamical similarity index [67, 78, 84] and various measures of entropy [1, 26, 103]. Considering that the EEG is a noisy and non-stationary time series, chaotic measures can aid in interpreting brain dynamics. Moreover, the sporadic nature of seizures can be inexplicable using linear concepts as sometimes seizures may result from external inputs which might not always be present beforehand [50].

Generally, all of these measures aim to capture a significant characteristic of seizures: an increase of synchronization in neuronal activity. In other words, by envisioning the brain as a chaotic system, its behavior becomes progressively more predictable the closer it is to the seizure. Despite their use in multiple studies, it is worth noting that the predictive power of the correlation dimension [45] and the Lyapunov exponents [64, 65] has been contested by some authors.

Linear multivariate features

Multivariate features are able to analyze the interactions between different brain regions in terms of synchronization, by incorporating information from different electrodes/channels. Brainwave synchronization patterns have been demonstrated to help differentiate between the inter-ictal and pre-ictal stages [11, 73].

Several multivariate measures have been proposed, such as Rasekhi et al. [87] who experimented with differences and ratios between different univariate linear measures for seizure prediction. Bandarabadi et al. [15, 16] have also proposed multivariate features based on spectral power. However, the most widely used linear multivariate feature is maximum linear cross-correlation, which quantifies the lag synchronization between two EEG channels or, in other words, it measures how identical two signals are in terms of phase and amplitude but shifted in time [75, 76]. Numerous studies have reported drops in synchronization during the pre-ictal stage, followed by hyper-synchronization during seizures [73, 76, 78, 104, 108].

Non-linear multivariate features

In addition to linear measures, multivariate non-linear features have been employed in the seizure prediction field. These also aim to capture synchrony changes in brain activity with measures based on mutual information and similarity between electrode channels in the EEG [11]. The two most prominent are mean phase coherence [5, 16, 78, 84], proposed by Mormann et al. [79] and often compared to maximum linear cross-correlation, and dynamical entrainment [73], proposed by Iasemidis et al. [49], which builds upon the concept of the Lyapunov exponents but applied to multiple channels.

3.5 Feature selection

After extracting a large number of features from the EEG, a high-dimensional space is created. In order to lighten the computational cost, improve classification performance and avoid overfitting, various feature selection or reduction methods have been employed [11].

Generally, feature selection methods operate under the same principle, which is to maximize relevance by preferring features with high discriminative power, while minimizing similarity by eliminating redundant measures. Authors have adopted methods such as maximum Difference Amplitude Distribution histograms (mDAD) [15, 16], minimum Redundance Maximum Relevance (mRMR) [10, 15, 16, 87] or ReliefF [74].

An alternate approach to feature selection relies on the use of Evolutionary Algorithms (EAs). This class of algorithms, as described in section 2.4, is a population-based method for finding sub-optimal solutions within acceptable time [32]. In the present context, candidate solutions are different sets of features and/or classifier hyperparameters, which are then selected according to operators based on the principles of natural selection, such as mutation and recombination [11]. Studies such as the ones by Direito et al. [31], Bou Assi et al. [10] and D'Alessandro et al. [26] have followed this methodology, with different genetic structures and fitness functions.

Additionally, dimensionality reduction methods such as Principal Component Analysis (PCA) may also be employed [2, 73]. This method transforms high-dimensional data into a low-dimensional orthogonal feature space, where each new feature is a principal component perpendicular to each other and ranked by variance. In Deep Learning approaches, reduction is performed either by convolutional layers [4, 20, 57, 99, 107] or through autoencoders [27].

3.6 Classification

Based on the remaining set of features, a chosen classification model can be trained and applied to detect the pre-ictal period. Numerous models have been adopted by authors, where Support Vector Machines (SVMs) [1, 4, 10, 14, 15, 21, 30, 31, 73, 74, 82, 87, 103, 104, 110], Convolutional Neural Networks (CNNs) [20, 57, 59, 73, 99, 107] and Long Short-Term Memory (LSTMs) [20, 27, 108] are the most commonly used.

As stated in section 2.3.4, data imbalance represents a serious issue in seizure prediction, as the inter-ictal period is significantly longer than the pre-ictal one. While many authors have addressed this by undersampling (dropping inter-ictal samples) [10, 16, 30, 86, 104, 110], others have adopted cost-sensitive classification approaches [82, 107].

It should be noted that, as mentioned in section 3.3, the duration of SOP and SPH can be handled at the classification stage. In practice, this is done when labelling the EEG epochs for training/testing the model. As such, the ones labelled as pre-ictal correspond to the SOP, while the epochs preceding the onset which are removed for training concern the SPH (in other words, the classifier is not trained with data belonging to the chosen time interval prior to the seizure).

Partitioning

There is considerable heterogeneity in the chosen partitioning methods, with no standard procedure. As such, different approaches lead to different assumptions regarding the seizure-generation process.

For instance, studies which choose a determined number of seizures from the entire set of patients for training and test on the remaining seizures assume that seizure-generation is similar between different individuals [14, 15, 26, 76, 77, 110]. On the other hand, most studies assume a patient-specific approach, in which the model is trained and tested for each subject, assuming that seizures are generated differently for each patient [7, 30, 74, 84, 87, 104]. Additionally, most studies ignore the order in which seizures occur, assuming nonexistence of a concept drift, while studies such as the ones by Teixeira et al. [103, 104], Alvarado-Rojas et al. [7] and, notably, Kiral-Kornek et al. [59] assume there is a time dependence by using earlier seizures to train and later data to test their models.

At any rate, partitioning should not be performed by selecting random epochs from the EEG, as it is not a common machine learning problem. Instead, methods

must be suitable for time series, as each seizure episode represents a chronology of events. Furthermore, data from the same ictal event should not be used for both training and testing, since performance will be biased by the fact that the model was trained with epochs close to the ones used for testing.

Support Vector Machines (SVMs)

SVMs are widely adopted classifiers which are capable of producing non-linear decision boundaries by employing non-linear kernel functions. They are characterized by a good generalization capability and low number of hyperparameters [11, 30].

The algorithm determines an optimal hyperplane (decision boundary) by maximizing the margin of separation between two classes or, in other words, maximizing the shortest distance between the decision function and the closest data points of each class [30, 46]. In cases where data is not linearly separable, the feature space is implicitly transformed into a higher dimensional space with non-linear kernels [100].

This possibility of linearizing the feature space is, from the perspective of interpretability, the main appeal of adopting SVMs. Despite this, it should be noted that a high number of features may lead to loss of interpretability. In the context of seizure prediction, SVMs have been demonstrated to outperform other classification models in terms of FPR/h [11].

Convolutional Neural Networks (CNNs)

CNNs are Deep Learning models designed to process data that comes in the form of multiple arrays (e.g. images) and are capable of learning optimal time-invariant local features from the input. In other words, these networks are able to build high-level representations and automatically learn features that handle the time property directly [69, 73].

In terms of architecture, CNNs typically stack various convolutional layers, which build feature maps with filtering operations using kernels. These are then followed by pooling layers, that learn features from the resulting maps of the previous layers, which can then be used by classification layers. Lastly, dropout layers are often employed to prevent overfitting by setting the output of random units to zero during training [57].

In the case of seizure prediction, the time series data is converted to a compatible format (either raw or using Fast Fourier Transform (FFT)/wavelet decomposition) to serve as input. These networks can then capture short-term temporal dependencies in the EEG and automate feature engineering [57, 107].

Long Short-Term Memory (LSTMs)

LSTMs are a different type of Deep Learning models, based on Recurrent Neural Networks (RNNs). They introduce the concept of gates, which control what information needs to be stored in memory and what must be discarded by the algorithm. In other words, it regulates the learning rate so that the network can better adjust to large sequences of data [108].

For seizure prediction, LSTM networks have an advantage over CNNs, since they are capable of learning temporal features of brain activity during different states while maintaining long-time dependencies, which can greatly benefit predictive performance [27, 108].

It should be noted that, despite their advantages such as learning from raw data and robustness to noise, Deep Learning approaches require large amounts of data, are prone to overfitting and can be notoriously difficult to interpret [93].

3.7 Regularization

In order to attenuate the number of false alarms raised by the classification model, regularization methods are employed. Generally, these consist of functions that account for the signal's temporal dynamics and smooth the output of the classifier accordingly [11]. Two methods have been used by several studies: Kalman filtering [7, 20, 21, 82, 107] and the Firing Power method [15, 16, 20, 30, 86, 87, 104].

Kalman filter

The Kalman filter is based on the state estimation of a linear dynamic system:

$$\begin{cases} s_{k+1} = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix} s_n + w_n \\ y_n = \begin{bmatrix} 1 & 0 \end{bmatrix} s_n + z_n \end{cases}, \tag{3.3}$$

where s_k is the state of the system, y_k the classifier's predicted output, τ the prediction interval and w_k and z_k are white noise vectors. An alarm is only raised when the filter output crosses a given threshold [20, 82].

Firing Power

The Firing Power method was proposed by Teixeira et al. [102] and employed in numerous studies. Considering the binary classifier output O_k ($O_k = 1$ for pre-ictal, $O_k = 0$ for inter-ictal) and τ the number of samples in a sliding window (of equal size to the pre-ictal period), the firing power of sample k is given by (3.4):

$$FP[n] = \frac{\sum_{k=n-\tau}^{n} O[k]}{\tau}. \tag{3.4}$$

Simply put, this method quantifies the relative number of samples classified as pre-ictal and raises an alarm based on a threshold. Despite the number of studies using Firing Power, no optimal threshold has been identified [11]. It is worth noting that since the window is the same duration as the pre-ictal period, only a single alarm may be raised for each seizure and an equally long refractory period is introduced.

In comparison to the previous regularization approach, the Firing Power method (illustrated below in figure 3.4) demonstrated to be more conservative in raising alarms, due to a longer memory of classification dynamics and the additional time constraints [103]. It should be noted that both methods require parameter tuning, even though it is more straightforward in the case of Firing Power.

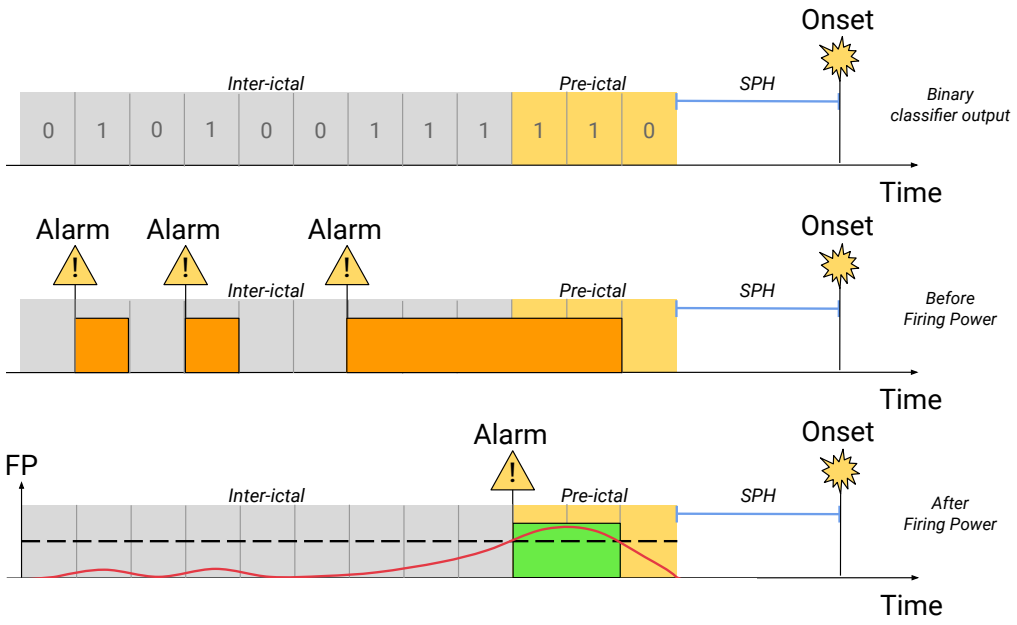


Figure 3.4: Visual representation of the effect of regularization on the classification output and number of false alarms. Given a certain threshold (dashed line), an alarm is only raised when enough consecutive epochs are classified as pre-ictal. The continuous regularized output is represented with a red line.

3.8 Performance evaluation

Finally, as with any classification problem, the developed methodology is evaluated according to a given set of metrics. As presented in section 2.3.2, the gold standard metrics for seizure prediction algorithms are sensitivity and FPR/h, which should be reported for a range of SOP and SPH values [113].

While most studies present results in terms of sensitivity, this is not the case for FPR/h. The latter is presented as specificity by various studies which, as discussed previously, does not accurately reflect the impact of false alarms. Some have reported values of FPR/h below 0.15 [15, 16, 21, 57, 73, 87, 104, 108] which, for patients in pre-surgical monitoring, has been considered reasonable [113]. Rabbi et al. [84] displayed an FPR/h of 0.46, which was the highest among studies that included the metric.

It should be noted that, even for studies that evaluate performance with both sensitivity and FPR/h metrics, a fair comparison cannot be made, considering that these are rarely presented for different values of SOP and SPH, along with the differences in patient type and recording duration. Furthermore, most studies do not conduct statistical validation of their results, either by comparison to a random predictor or surrogate data testing.

Table 3.4 presents an overview of classification, regularization and performance evaluation procedures undertaken by studies from the past 20 years, where the aspects discussed in the previous sections can be seen clearly.

Table 3.4: Overview of classification, regularization, performance and other related aspects from EEG seizure prediction studies of the past 20 years.

Study	Partitioning	Classification	Regularization	FPR/h	Sensitivity (%)	Statistical validation
Daoud et al. (2019) [27]	Leave-One-Out with seizures	CNN, Bi-LSTM	N.A.	0.00	99.72	No
Khan et al. (2018) [57]	K-fold with recordings	CNN	N.A.	0.14	87.80	Yes
Truong et al. (2018) [107]	Leave-One-Out with seizures	CNN	Kalman filter	0.06 - 0.21	75.00 - 81.40	Yes
Tsiouris et al. (2018) [108]	K-fold with recordings	LSTM	N.A.	0.02 - 0.11	99.28 - 99.84	No
Agarwal et al. (2018) [4]	Training: 50% seizures Testing: remaining seizures	SVM, CNN	N.A.	N.A.	96.47	No
Chamseddine et al. (2018) [20]	Training: 80% samples Testing: remaining samples	LSTM, CNN, GRU	Firing Power, Kalman filter	N.A.	88.70	No
Sun et al. (2018) [99]	N.A.	LDA, CNN, RNN, Linear regression	N.A.	N.A.	N.A.	No
Aggarwal et al. (2017) [5]	N.A.	Thresholding	N.A.	N.A.	80.00	No
Kiral-Kornek et al. (2017) [59]	Training: first 2 months Testing: remaining duration	CNN	N.A.	N.A.	68.60 (mean)	Yes
Direito et al. (2017) [30]	Training: 2 - 3 seizures / patient Testing: remaining seizures	SVM	Firing Power	0.20	38.47	Yes
Bandarabadi et al. (2015) [16]	Training: first 3 seizures / patient Testing: remaining seizures	SVM	Firing Power	0.10	75.80	Yes
Bandarabadi et al. (2015) [14]	N.A.	N.A.	N.A.	N.A.	N.A.	Yes
Bou Assi et al. (2015) [10]	Training: 80% segments Testing: remaining segments	SVM, ANFIS	N.A.	N.A.	85.49	No
Rasekhi et al. (2015) [87]	Training: first 3 seizures / patient Testing: remaining seizures	SVM	Firing Power	0.11	60.90	Yes
Alvarado-Rojas et al. (2014) [7]	Training: first 4 seizures / patient and at least 10 hours Testing: remaining seizures	Thresholding	Kalman filter	0.33	68.00	Yes
Teixeira et al. (2014) [104]	Training: 2 - 3 seizures / patient Testing: remaining seizures	SVM, ANN	Firing Power	$\leq$0.35	>50.00	Yes
Moghim et al. (2014) [74]	Training: 70% samples / patient Testing: 30% samples / patient	SVM	N.A.	N.A.	91.14	No
Rabbi et al. (2013) [84]	Training: 1 seizure / patient Testing: 5 seizures / patient	ANFIS	N.A.	0.46	80.00	No
Rasekhi et al. (2013) [86]	Training: first 3 seizures / patient Testing: remaining seizures	SVM	Firing Power	0.15	73.90	No
Bandarabadi et al. (2012) [15]	Training: 3 seizures Testing: remaining seizures	SVM	Firing Power	0.15	76.09	No
Valderrama et al. (2012) [110]	Training: 50% seizures Testing: 50% seizures	SVM	N.A.	N.A.	94.50	Yes
Teixeira et al. (2012) [103]	Training: first 3 seizures / patient Testing: remaining seizures	SVM	Firing Power, Kalman filter	0.20	77.00	No
Acharya et al. (2012) [1]	Training: 67% segments Testing: 33% segments	SVM, KNN, Decision tree, Probabilistic NN, GMM, Naive Bayes, Fuzzy Sugeno	N.A.	N.A.	97.60 (mean)	No
Direito et al. (2012) [29]	N.A.	Hidden Markov model	N.A.	N.A.	94.59	No
Direito et al. (2011) [31]	Training: 3 seizures / patient Testing: remaining seizures	SVM	N.A.	N.A.	89.54 (best model)	No
Park et al. (2011) [82]	Leave-One-Out with 1 inter-ictal and 1 pre-ictal segments	SVM	Kalman filter	0.27	97.50	No
Chisci et al. (2010) [21]	Training: 15 mins before first 2 or 3 seizures, inter-ictal random Testing: remaining seizures	SVM	Kalman filter	0.00 (some patients)	100.00	No
Mirowski et al. (2009) [73]	Testing: last 1 or 2 seizures and 33% inter-ictal samples Training: remaining seizures	CNN, SVM, Logistic regression	N.A.	0.00 (some patients)	100.00 (some patients)	Yes
Adeli et al. (2007) [3]	N.A.	N.A.	N.A.	N.A.	N.A.	Yes
Mormann et al. (2005) [78]	N.A.	N.A.	N.A.	N.A.	N.A.	Yes
Le Van Quyen et al. (2005) [68]	N.A.	KNN, Mahalanobis distance	N.A.	N.A.	84.00	No
D'Alessandro et al. (2003) [26]	Training: 70% seizures Testing: 30% seizures	Probabilistic NN	N.A.	0.28	62.50	No
Mormann et al. (2003) [77]	Leave-One-Out with 1 patient for testing	Thresholding	N.A.	N.A.	N.A.	No
Mormann et al. (2003) [76]	Leave-One-Out with 1 patient for testing	N.A.	Parameter optimization	N.A.	N.A.	Yes
Le Van Quyen et al. (2001) [67]	N.A.	N.A.	N.A.	N.A.	N.A.	No
Geva et al. (1998) [43]	N.A.	Unsupervised fuzzy clustering	N.A.	N.A.	N.A.	No

3.9 Summary

Pipeline overview

Generally, after collecting the EEG recordings, the seizure prediction framework is comprised of pre-processing, feature extraction through sliding window analysis and feature selection/reduction steps, followed by classification, regularization and performance evaluation methods.

The main changes introduced by recent studies result from the adoption of Deep Learning models, which are capable of performing automatic feature engineering from raw data and handling the time property directly as another dimension of the feature space. Despite this, these models require larger amounts of data and can be notoriously difficult to interpret when compared to traditional machine learning models. It is worth noting that, in both approaches, a higher number of used features as well as their complexity may severely affect interpretability.

Identified shortcomings

Overall, current seizure prediction methodologies, despite achieving favorable results, present several issues which should not be overlooked.

One of the clear limitations in the current methodology resides in the fact that most databases only contain recordings of patients in pre-surgical monitoring, spanning over a few days. In order to truly understand the behavior of the proposed prediction tools and make them clinically viable, long-term recordings (comprising months/years) of data from everyday life are required. Moreover, studies should not be based on discontinuous data segments or selected epochs [62].

Feature extraction/engineering constitutes the step with the most heterogeneity in the current state of the art. While many studies present superior performance using various features with a high degree of complexity, it is achieved at the expense of interpretability. To make these methodologies a feasible solution, results must be clinically explainable.

Several studies also disregard the existence of concept drifts, namely in terms of time dependence between seizures and other aspects which may alter the data distribution over time (e.g. medication). Additionally, given that the pre-ictal period is not annotated by physicians, most authors label it by assuming one or more time lengths. Towards these issues, new prediction models should be adaptable and extensive search must be carried out regarding the duration of the pre-ictal period.

Another concerning issue is the frequent omission of the FPR/h metric, which

does not allow for a proper evaluation of the prediction effectiveness and how the proposed system may impact the patient in terms of stress/anxiety. Furthermore, given that most studies lack either form of statistical validation and do not present results for a range of SOP or SPH (the latter of which is often omitted altogether), a fair comparison between them cannot be made.

4

Methodology

This chapter describes the various steps concerning the development of the proposed methodology for seizure prediction. Section 4.1 presents the problem and the underlying assumptions, followed by a brief description of the used data in section 4.2. Sections 4.3 to 4.5 characterize, respectively, the pre-processing pipeline, algorithm design and the different experimental scenarios. Finally, a brief summary is given in section 4.6.

4.1 Problem statement

The current study concerns the development of a patient-specific seizure prediction scheme, based on Electroencephalogram (EEG) data. It is targeted at Drug-Resistant Epilepsy (DRE) patients in pre-surgical monitoring and, in particular, ones suffering from Temporal Lobe Epilepsy (TLE), which is the most common form of focal epilepsy [35, 95].

Considering that the used data (described next) concerns patients in pre-surgical monitoring units, this study may only present a proof of concept for the proposed prediction system. In order to truly assess its real-life applicability, long-term data from everyday life must be used [24, 62].

Several decisions regarding data processing and algorithm design throughout the conducted research were made under the following assumptions:

- seizure-generation processes can be captured by patterns found in the EEG signal, where the transition to an ictal state is assumed to be gradual and not influenced by external stimuli (e.g. flickering lights) [25, 78];
- these patterns exist up to 90 minutes before each seizure (maximum pre-ictal period) and can be identified before the last 10 minutes (SPH);
- seizures are time dependent and concept drifts are present [42, 62, 75];
- seizures separated by less than 4 and a half hours (270 minutes) belong to the same, dependent cluster of seizures [42, 62, 75];

- a set of 5 linear univariate features exists which is able to handle concept drift and detect the seizure-generation patterns [62, 75]
- the pre-ictal period varies between patients [42, 62, 81];

It should be noted that the interseizure interval used to define seizure clusters was chosen for two main reasons: it must provide enough time to train the prediction algorithm and, when used as a selection criterion for the data, it should allow to keep an adequate number of seizures from the available recordings.

4.2 Data description

The EEG data of 36 DRE patients (21 males and 15 females, with ages ranging from 13 to 67 years), along with its respective metadata, were selected from the European Database on Epilepsy (EPILEPSIAE) [52, 60]. Overall, 1988.8 hours of recordings were used, comprising a total of 208 seizure episodes, as presented in table 4.1 along with their respective clinical classification, pattern description and vigilance state.

The selection requirements for the group of patients described above were the following:

- EEG scalp recordings, with a sampling frequency of 256Hz;
- only contain seizures with a focus located on the temporal lobe (i.e. only patients suffering from TLE);
- only seizures preceded by, at least, 4 and a half hours (270 minutes) of recorded signal data not containing any ictal events;
- a minimum of 4 seizures per patient which meet the previous criteria, to ensure that enough are available to train and test the algorithm.

The clinical annotations including seizure type/activity pattern classification and vigilance state at the time of the seizure were used for patient stratification.

Table 4.1: Patient information including age, gender and number of seizures which met the selection criteria.

Patient ID	Age	Gender	Number of seizures	Seizure classification	Seizure activity pattern	Vigilance at seizure onset	Recording duration (h)
402	55	F	5	FOIA, FBTC, FOIA, FBTC, FOIA	t, t, t, t, t	A, A, A, A, A	41.6
8902	67	F	5	UC, FOIA, FOIA, FOIA, FOIA	a, b, a, m, a	A, A, A, A, A	34.5
11002	41	M	4	UC, FOIA, FOIA, FOIA	?, s, a, t	A, R, A, A	23.7
16202	46	F	7	UC, FBTC, UC, FOIA, FOIA, FOIA, FOIA	r, ?, r, r, r, ?, r	A, A, A, A, A, A, A	46.4
23902	36	M	5	FOA, FOA, FOA, FOA, FOA	t, t, t, d, t	A, A, A, A, A	45.9
30802	28	M	8	FOA, FOA, FOA, FOA, FOA, FOA, FOA	t, t, t, t, t, t, t	R, A, 2, A, A, R, 2, 2	73.6
32702	62	F	5	FOIA, FOIA, FOIA, FOIA, FOIA	t, t, t, r, a	A, A, A, A, A	34.2
46702	15	F	5	FOA, FOIA, FOIA, FBTC, FOIA	a, a, t, b, t	A, 2, A, 2, A	24.9
50802	43	M	5	FOIA, UC, UC, FOIA, FBTC	t, t, t, t, t	A, 2, 2, 2, A	47.6
53402	39	M	5	FOA, FOA, FOA, FOA, FOIA	?, ?, ?, ?, t	A, A, 2, A, A	61.6
55202	17	F	8	FOIA, FOIA, FOA, UC, UC, FOA, UC, FOIA	t, d, t, t, t, r, r	A, A, A, A, A, A, A	77.3
56402	47	M	4	UC, UC, UC, FBTC	t, ?, ?, a	A, A, A, A	32.2
58602	32	M	6	FOIA, FOIA, FOIA, FOIA, FOIA, FOIA	r, t, t, r, r, t	A, R, A, A, A, 2	35.3
59102	47	M	5	FOA, FOIA, FOIA, FOIA, FOA	?, t, t, t, t	A, A, A, A, A	94.2
60002	55	M	6	FOIA, FOIA, FOIA, UC, FOIA, FOIA	d, c, t, t, d, d	1, A, A, R, R, 1	164.3
64702	51	M	5	FOA, FBTC, FBTC, FBTC, FBTC	?, m, t, t, t	A, A, A, A, 2	43.6
75202	13	M	7	FOA, FOA, UC, FOA, FOA, FOA, FOA	t, t, t, t, t, ?, t	2, 2, A, A, A, A, A	64.6
80702	22	F	6	FOIA, FOIA, UC, FOIA, FBTC, FOIA	b, b, ?, c, c, c	A, A, A, A, A, A	41.5
85202	54	F	5	FOIA, FOIA, UC, UC, UC	m, c, m, m, m	2, A, A, A, A	32.4
93402	67	M	5	FBTC, FOIA, FOIA, UC, UC	t, t, t, t, t	2, 2, 2, 2, 2	66.0
93902	50	M	6	FOA, FOIA, FBTC, FOIA, FOIA, UC	t, t, d, d, d, d	A, A, 2, A, 2, A	32.2
94402	37	F	7	FOA, UC, FOIA, UC, FOA, UC, FOA	?, d, b, t, ?, b, ?	A, A, A, 2, A, 2, A	42.3
95202	50	F	7	FBTC, FOIA, FOIA, FOIA, UC, FOIA, UC	b, b, b, m, b, b, t	2, 2, 2, 2, 2, 2, 2	102.2
96002	58	M	7	FOIA, FOIA, FOIA, FOIA, UC, FOIA, FOIA	t, t, t, d, a, t, a	A, A, A, A, A, A, A	94.1
98102	36	M	5	FOA, UC, UC, UC, FBTC	?, ?, ?, ?, ?	A, A, A, A, A	57.7
98202	39	M	7	FOIA, FOIA, FOIA, FBTC, FOIA, FOIA, FOIA	t, a, t, t, t, t, t	A, A, A, A, A, A, A	47.9
101702	52	M	5	FOIA, FOIA, FOIA, FOIA, FOIA	t, t, t, r, r	A, A, A, 2, A	35.8
102202	17	M	7	FOA, UC, FOIA, UC, FOA, FOIA, UC	b, ?, t, ?, t, t, t	2, A, 2, A, A, 2, A	63.4
104602	17	F	5	FOIA, FBTC, FBTC, FBTC, UC	t, a, t, t, d	A, 2, 2, 2, 2	27.2
109502	50	M	4	FOIA, FOIA, UC, UC	t, t, t, t	A, A, A, A	53.9
110602	56	M	5	FOIA, FOIA, FOIA, FOIA, FOA	t, t, t, t, t	A, A, A, A, A	37.9
112802	52	M	6	UC, FOIA, UC, FOIA, FOIA, UC	t, t, t, t, t, t	A, A, A, A, A, A	123.4
113902	29	F	6	UC, FOIA, FOIA, FOIA, UC, FOIA	t, d, t, t, t, t	A, A, 2, A, 2, A	34.7
114702	22	F	8	FOIA, FOIA, UC, FOIA, FOIA, FOIA, FOIA, FOIA	t, t, t, t, d, t, d, t	A, A, A, A, A, A, A, A	46.0
114902	16	F	7	FOA, FOIA, FOIA, FBTC, UC, FOIA, FOIA	s, b, s, t, r, a, t	A, A, A, 2, A, A, A	62.6
123902	25	F	5	FBTC, FBTC, FOIA, FOIA, FOA	t, t, t, t, t	2, 2, R, A, A	42.1

Seizure classification: unclassified (UC), Focal Onset Aware (FOA), Focal Onset Impaired Awareness (FOIA), Focal to Bilateral Tonic-Clonic (FBTC); Seizure pattern: unclear (?), rhythmic sharp waves (s), rhythmic alpha waves (a), rhythmic delta waves (d), rhythmic theta waves (t), rhythmic beta waves (b), repetitive spiking (r), cessation of inter-ictal activity (c), amplitude depression (m); Vigilance state at the time of the seizure: awake (A), Non-REM sleep stage I (1), Non-REM sleep stage II (2), REM sleep stage (R)

4.3 Pre-processing

After loading the raw signal data, which was originally split into 1 hour segments, the first step was to find any possible gaps in the recording. Consequently, all instances of missing data were identified and handled by value imputation, taking the average between the amplitude values of the samples immediately before and after each gap. The segments were then compiled differently for each set of seizures: the 4 hours preceding each onset were extracted for the first three chronological seizures, on which the algorithm is trained; for the remaining seizures, used for testing, all available recordings preceding it were retrieved, starting from 30 minutes after the previous ictal event (to account for post-ictal activity) until each onset.

Subsequently, the multi-channel signals were segmented in 5-second non-overlapping windows and filtered using a 50Hz notch filter (to remove power-line interference) followed by a fourth-order Butterworth high-pass filter with a cut-off frequency at 0.5Hz (to remove the DC component and minimize motion artifacts).

Finally, 24 linear univariate features were extracted from each of the 19 EEG channels (see figure 4.4) within each window. In the time domain, the first four statistical moments (mean, variance, skewness and kurtosis) and the three Hjorth parameters (activity, mobility and complexity) were computed, while in the frequency domain the relative spectral power of the delta (0.5-4Hz), theta (4-8Hz), beta (8-12Hz), alpha (13-30Hz), low gamma (30-79Hz) and high gamma (79-128Hz) bands was extracted, along with the spectral edge frequency at three different cut-off percentages (50%, 75% and 90%) and the energy of each wavelet coefficient (D1 to D8, using the Daubechies 4 mother wavelet). These steps are summarized in figure 4.1 below.

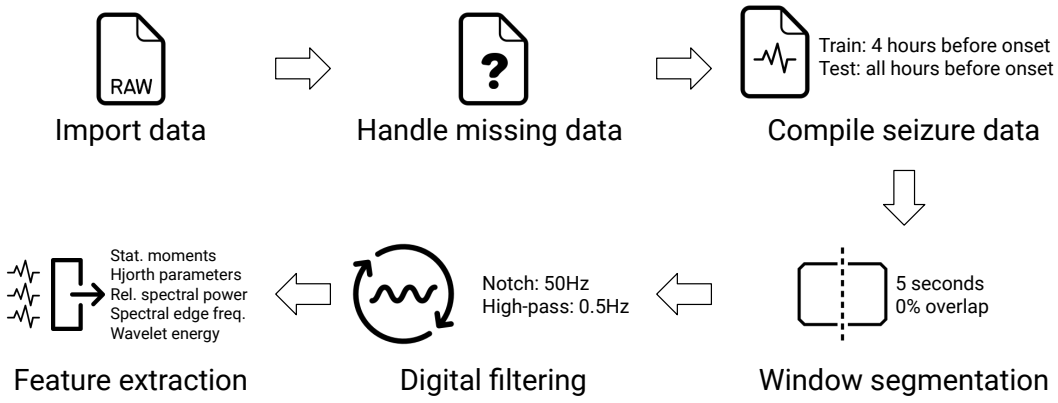


Figure 4.1: Flowchart illustrating the pre-processing pipeline of this study.

4.4 Algorithm design

4.4.1 Overview

The proposed approach for seizure prediction consists in the application of a Multi-Objective Evolutionary Algorithm (MOEA), where each individual within the population corresponds to a set of high-level features (constructed from the ones extracted in the previous stage [106]) and a minimum pre-ictal period (P_{min}). In other words, the algorithm searches for different configurations to build a machine learning model to anticipate seizures. As described in section 2.4, based on the principles of evolution, increasingly better solutions are found through each generation.

Algorithm 1 Pseudocode illustrating the evolutionary algorithm's common steps.

```
population = randomInitialization(number_individuals, number_features)
population.evaluate()
generation_count = 1
while generation_count ≤ max_generations do
  parents = selectParents(population)
  offspring = recombine(parents, crossover_rate)
  offspring = mutate(offspring, mutation_rate)
  offspring.evaluate()
  population = selectSurvivors(population, offspring)
  generation_count = generation_count + 1
end while
```

After initializing the population with a set of randomly generated individuals (as indicated above in algorithm 1), their fitness scores are evaluated. Namely, for each individual, its high-level features are extracted using a sliding window and the respective labels (inter-ictal and pre-ictal) are assigned. Then, a linear Support Vector Machine (SVM) is trained and tested and, from the resulting output, two performance metrics are computed: sample sensitivity (given by (2.4)) and time under false alarm (given by (2.5)). These metrics correspond to the first two fitness functions, in addition to a third objective which takes into account the number of different electrodes (and their corresponding lobes) within the feature set.

Once every individual's fitness has been evaluated, parent selection is performed. The selected parents are then recombined or cloned, with a given probability, producing offspring. Next, the generated offspring is subjected to mutation, also with a given probability. Finally, a replacement strategy is applied to select which individuals take part in the following generation.

Three different implementations were used: Fonseca and Fleming's MOGA [41],

Deb et al.'s NSGA-II [28] and a modified version of Beume et al.'s SMS-EMOA [18]. Both parent selection and replacement are accomplished in different ways in each of these three methods. The evolution process is repeated in each generation until the designated stopping criteria is met, which was set to a maximum of 50 generations. The population size was fixed at 100 individuals, with each one holding a set of 5 high-level features. These parameters were tuned with the aim of keeping runtime shorter than the considered time length between independent seizures: approximately 2 hours on a machine equipped with an Intel Core i5-3230M 2.6GHz processor, 8GB of RAM running on macOS Mojave 10.14.6 and using Python 3.7 on Spyder 4.0.1.

The evolutionary search is run and trained on a set of training seizures (which, in this study, were the first three seizures of each patient). Once the search is completed, the Pareto optimal solutions are tested on a different set of seizures, unknown to the machine learning model, as shown in figure 4.2. Their performance is then evaluated according to the gold standard metrics for seizure prediction: seizure sensitivity and False Positive Rate per Hour (FPR/h). Statistical validation is also performed against a random predictor and by surrogate time series analysis.

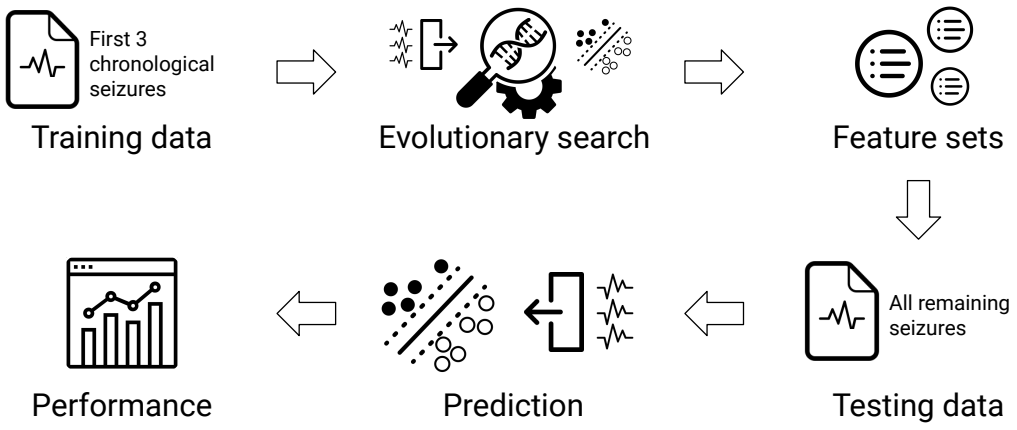


Figure 4.2: Flowchart illustrating the main steps in the algorithm.

4.4.2 Representation

Genotype

Each individual represents a set of 5 features as well as a minimum pre-ictal period (P_{min}), as illustrated in figure 4.3. Each feature is composed of 13 different genes, whose possible values are represented in figure 4.4 and described below:

- Active feature group: a binary gene that determines which type of feature is active: concerning the time domain or the frequency domain;
- Active time feature: a binary gene that dictates which time domain feature is active: a statistical moment or a Hjorth parameter;
- Active frequency feature: a binary gene that specifies which type of frequency domain feature is active: one related to band division or the spectral edge frequency;
- Active frequency band feature: a binary gene that determines which type of feature based on frequency band division is active: the relative spectral power or wavelet coefficient energy;
- Statistical moment: indicates which of the four statistical moments is used (mean, variance, skewness or kurtosis);
- Hjorth parameter: indicates which of the three Hjorth parameters is used (activity, mobility or complexity);
- Frequency band: indicates which band's relative spectral power is used (delta, theta, beta, alpha, low gamma or high gamma);
- Wavelet coefficient: indicates which detail coefficient's energy is used (ranging from D1 to D8);
- Spectral edge frequency cut-off: indicates which cut-off is used to compute the spectral edge frequency (50%, 75% or 90%);
- Electrode: dictates which electrode the information is extracted from;
- Window length: dictates the duration of the time-window from which information is extracted, where the possible values were 1, 5, 10, 15 or 20 minutes;
- Delay: specifies the duration of the delay (relative to the minimum pre-ictal period) applied to the aforementioned time-window, ranging from 0 to 30 minutes in 5-minute increments;
- Mathematical operator: specifies which operation (mean, variance or integral) is applied to the data within the indicated window.

Additionally, there is a single gene for each individual which dictates P_{min}, ranging from 30 to 60 minutes in 5-minute increments. Its value relates to the Seizure Occurrence Period (SOP) duration as given by (3.1). It is worth noting that, while figure 4.4 illustrates the genotype using a graph representation, in practice, each gene is encoded using an integer assigned to the corresponding node.

The design of each graph in the genotype was informed by the order relations between the values of each gene. The electrode graph follows the 10-20 system, where neighboring electrodes are neighboring nodes. In the mathematical operator

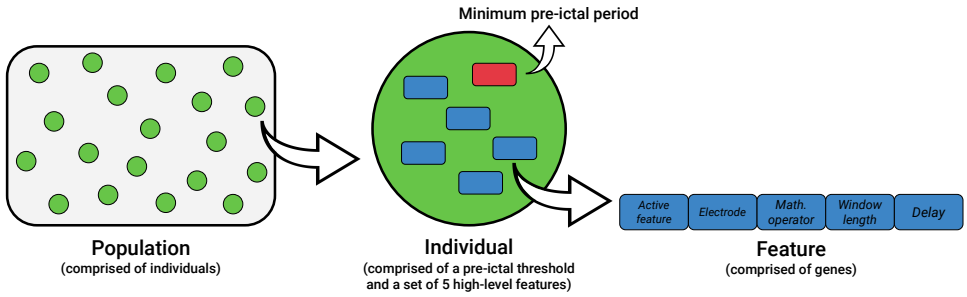


Figure 4.3: Representation of the evolutionary algorithm's population. Note that each feature actually contains 13 genes, unlike the current simplified illustration.

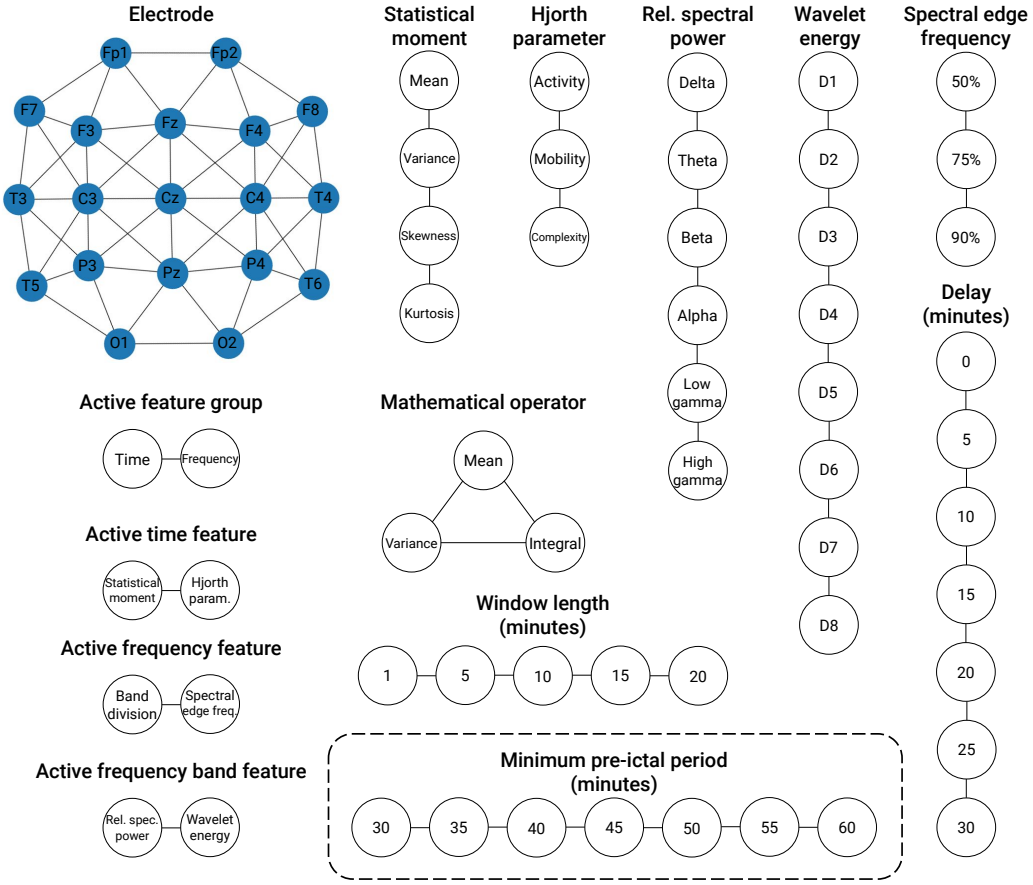


Figure 4.4: Graph representation of each individual's genotype. Every gene with the exception of the minimum pre-ictal period (highlighted with a dashed line) concerns each feature.

gene, since no order relation could be established, all nodes are neighbors of each other. The statistical moment gene is sorted by ascending order of the moment order. The Hjorth parameters are sorted by the order of the derivative used to compute each one (as given by (A.10), (A.11) and (A.12)). The frequency bands in the relative spectral power gene are sorted by ascending order of frequency. The

wavelet energy gene is sequenced by increasing level of decomposition. Every other gene with the exception of the binary genes follow a sequence of increasing numbers (e.g. the spectral edge frequency gene is sorted by the cut-off percentage).

Phenotype

The genotype-phenotype mapping process is split into two stages: firstly, the construction of a high-level feature extractor, followed by the determination of the pre-ictal period.

The decoding process is straightforward with the exception of the genes that dictate which feature is selected for extraction, as shown by the tree diagram in figure 4.5. Then, as depicted in figure 4.6, each decoded feature in the individual's genotype is organized chronologically according to its delay gene value, such that the ones with the higher delay duration are placed further back in the timeline.

By using a set of features with time windows that not only span different lengths but also comprise distinct moments in time, one may capture the chronology of events leading up to a seizure, effectively handling the time property without resorting to more complex machine learning models [71].

Once the "active" features are selected and placed accordingly in the timeline, the mathematical operators are employed to perform sliding window analysis on the low-level features extracted in the pre-processing stage. By performing feature construction, the resulting high-level features may improve the performance and discriminative ability in comparison to the original set, particularly in high-dimensional classification problems [106, 115]. Finally, the pre-ictal period duration is determined by the sum of the individual's minimum pre-ictal period and the shortest delay found within its set of features.

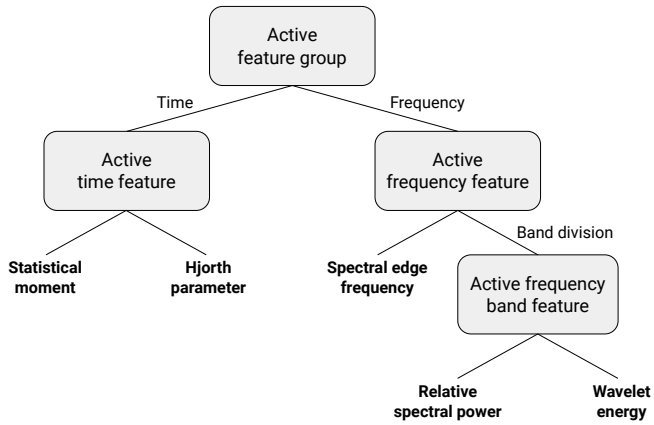


Figure 4.5: Tree diagram illustrating how the "active" feature is decoded from the values of 4 binary genes.

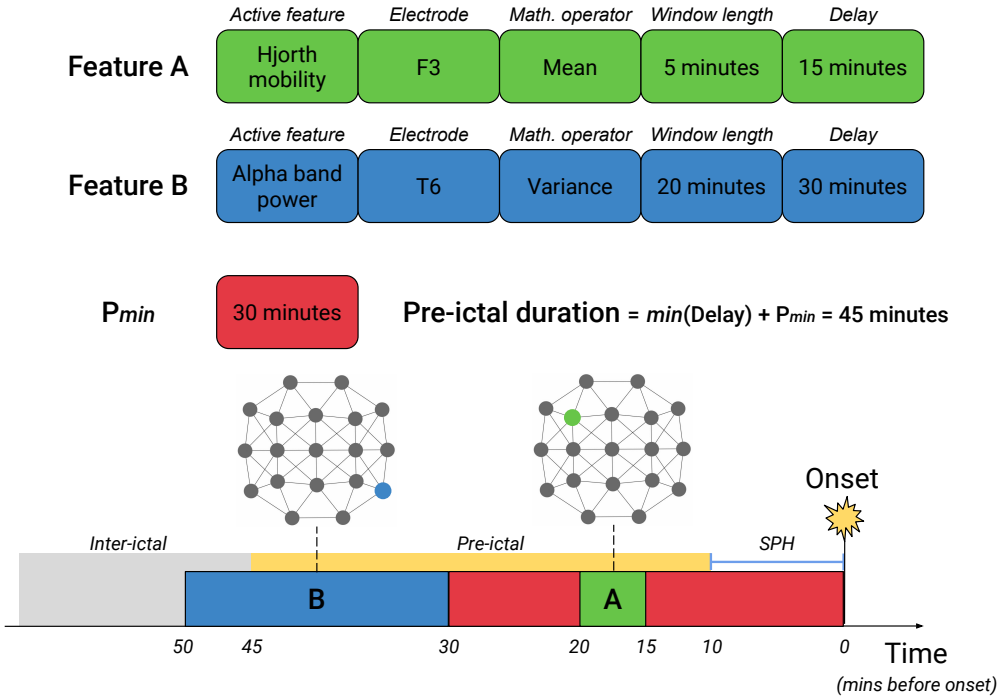


Figure 4.6: Visual representation of the genotype decoding process. In this example, two features are to be extracted from different brain regions and arranged chronologically. The individual's minimum pre-ictal period is also presented in red.

4.4.3 Fitness evaluation

The evaluation process of each individual consists of three main steps: high-level feature extraction, classification model training and performance evaluation.

As described previously, sliding window analysis is performed using a non-overlapping 30-second window on the decoded set of features. Each window is also labeled as inter-ictal (0) or pre-ictal (1) according to the determined pre-ictal duration.

Before constructing the machine learning model, the extracted features require some processing steps. First, the features are scaled using z-score standardization, as given by (4.1):

$$z = \frac{x - \mu}{\sigma}, \tag{4.1}$$

where μ and σ correspond to the mean and standard deviation, respectively. Next, in order to handle the class imbalance discussed in section 2.3.4, a balanced weight

was computed for each class i according to (4.2) and assigned to each sample:

$$C_{w_i} = \frac{N_S}{N_C N_{C_i}}, \tag{4.2}$$

where N_S is the number of training samples, N_C is as the number of classes and N_{C_i} is the number of samples from class C_i. Finally, a Seizure Prediction Horizon (SPH) was implemented by assigning a null training weight to samples within the last 10 minutes before the onset.

Subsequently, the processed features are used to train a machine learning classification model. For this study, a linear SVM classifier (implemented using the scikit-learn library [83]) was chosen for its relatively fast convergence and ease of interpretability. Its training phase is carried out iteratively, such that for a set of N_{tr} training seizures (a subset of all available N seizures), the classifier begins by training on the first seizure and validating on the second one and, thereafter, proceeds to train on all previous seizures and validates on the following one, as indicated in figure 4.7 below.

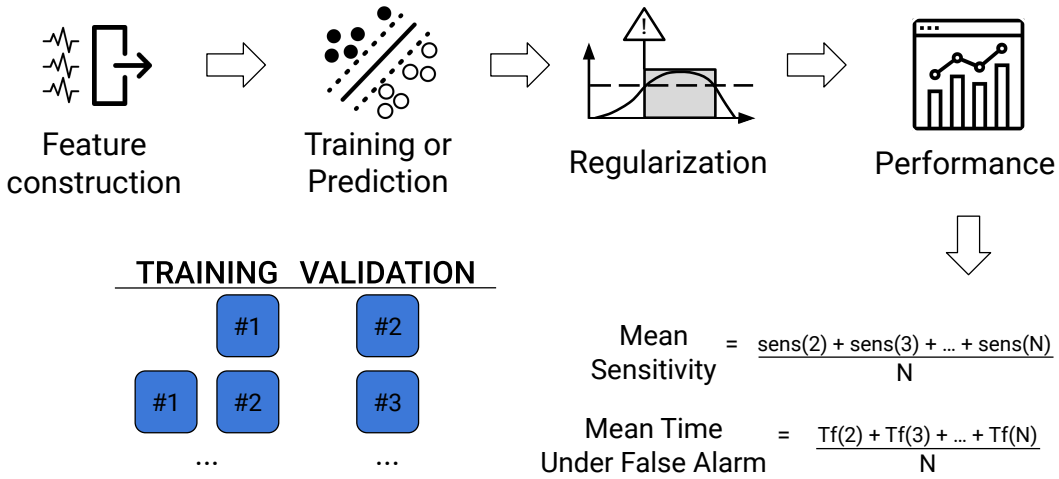


Figure 4.7: Flowchart illustrating the fitness evaluation scheme. For each individual, its high-level features are constructed, a classifier is trained iteratively (as shown above), and its performance is assessed after output regularization (i.e. after implementing the Firing Power method and a refractory behaviour).

In each iteration of the training phase, the classifier's output is processed in two stages. Firstly, the Firing Power method is used to attenuate the number of false alarms and smooth out the output, using the procedure described in section 3.7. A reasonable threshold of 0.7 was set, such that an alarm is raised when $FP[n] > 0.7$. Afterwards, a refractory behavior is also implemented, such that no

alarms can be raised after SOP+SPH minutes. This is done to approximate the alarm firing mechanism to a real one: for instance, 2 near-consecutive alarms, for the same seizure, should only correspond to one alarm, in order to avoid causing any confusion to the patient regarding SOP and SPH.

It is worth noting that the introduction of a refractory behavior in the classifier requires a modification of the FPR/h formula (equation (2.2)) when assessing performance in the testing phase. In order to compensate for the number of samples after each raised alarm which are not analyzed (i.e. during the refractory period), the total refractory period $T_{refractory}$ is subtracted from the $Interictal_{duration}$:

$$FPR/h = \frac{N_{alarms}}{Interictal_{duration} - T_{refractory}}. \tag{4.3}$$

The classifier's performance is evaluated in each iteration of the training phase, concerning sample sensitivity (given by (2.4)) and time under false alarm T_f (given by (2.5)). These two metrics, despite not being part of the standard framework, were used in the fitness evaluation scheme given that seizure sensitivity, due to a low number of training seizures, can only take a few different values, which does not allow to properly differentiate between individuals, and both FPR/h and seizure sensitivity do not adequately reflect the number of correctly classified samples. Each metric value is averaged after all the iterations have been run.

Each of the two aforementioned metrics correspond to an objective, where each of the fitness functions is valued between 0 and 1. Maximization is assumed for each objective, such that the higher the fitness score the better. For the sample sensitivity objective, the metric score corresponds directly to the fitness value. In the case of time under false alarm, its fitness is computed by subtracting the metric score from 1, since it is meant to minimized (simply put, this equates to maximizing specificity). Lastly, there is a third objective which takes into account the number of different electrodes $N_{Electrodes}$ (and corresponding lobes N_{Lobes}), whose fitness score is computed according to (4.4):

$$Fitness_{Electrodes} = 1 - \frac{N_{Electrodes} \times N_{Lobes}}{N_{Features} \times 5}, \tag{4.4}$$

where $N_{Features}$ is the number of features of the individual (in this case, five). This objective aims to promote solutions which do not require a large number of electrodes and focus on a particular region of the brain, potentially sparing the patient from having to wear a full EEG electrode cap and improving comfort.

4.4.4 Selection configurations

Once every individual has been evaluated, parents are selected to reproduce and generate offspring. In this study, considering a population of N individuals, a group of $\frac{N}{2}$ parents are selected and recombine until N offspring are produced. The offspring is then subjected to mutation (with a given probability, just as before for recombination) and a replacement strategy is put in place to select the N individuals that will make up the following generation.

As stated previously, several parent and survivor selection schemes exist with different implications on selection pressure, which individuals are prioritized, etc. The three multi-objective implementations used in this study are presented next.

MOGA

Firstly, the population is ranked based on dominance: for each individual, its rank is equal to the number of individuals dominating it plus one (e.g. all non-dominated individuals are assigned rank 1). Afterwards, a fitness value is assigned by interpolating from the best individual (rank 1) to the worst (rank $n \leq N$), using a linear function (although others can be used). This fitness value is then averaged within the individuals of each rank, so that all of them have an equal probability of being sampled [41].

Furthermore, fitness sharing is implemented within each rank, with the aim of promoting diversity. The fitness f of each individual is adjusted depending on how many individuals fall within a certain distance d:

$$f'_i = \frac{f_i}{\sum_j sh(d(i,j))}, \tag{4.5}$$

where $sh(d)$ is a sharing function given by (4.6):

$$sh(d) = \begin{cases} 1 - (\frac{d}{\sigma_{share}})^{\alpha} & if \quad d \leq \sigma_{share}, \\ 0 & otherwise. \end{cases} \tag{4.6}$$

The parameter α controls the shape of the sharing function (where $\alpha = 1$ is linear) while σ_{share} indicates the distance required for two individuals to belong to the same niche. A suitable value for σ_{share} can be found by solving (4.7) considering q objectives:

$$N\sigma_{share}^{q-1} - \frac{\prod_{i=1}^{q}(M_i - m_i + \sigma_{share}) - \prod_{i=1}^{q}(M_i - m_i)}{\sigma_{share}} = 0, \tag{4.7}$$

where M and m are, respectively, vectors containing the maximum and minimum

fitness values of each objective found within the non-dominated front.

Considering the previously assigned fitness, parents are selected using Stochastic Universal Sampling (SUS). It functions in a similar fashion to fitness proportionate selection, where highly fit individuals have a bigger chance of being sampled. However, instead of generating a random number between 0 and 1 multiple times, it only does so once and then generates equally spaced numbers within that interval, so that less fit individuals have a more reasonable probability of being chosen [32].

Regarding the replacement strategy, MOGA uses a generational approach combined with the introduction of a small percentage of random immigrants. As such, $(1-\delta)N$ parents are recombined and mutated to produce $(1-\delta)N$ offspring, and the remaining individuals are randomly generated and added to the new generation. In this study, a percentage of 10% ($\delta = 0.1$) was chosen as found in the literature [40, 101].

NSGA-II

In this implementation, the ranking process is referred to as non-dominated sorting. The algorithm iteratively searches for all the non-dominated solutions in the population that have not been labelled as belonging to a previous front. After labelling the new front, a front counter is incremented and the process is repeated until all solutions have been ranked.

Afterwards, the individuals within each rank are sorted according to a measure, the crowding distance, which corresponds to the average side length of a cuboid defined by its nearest neighbours in the same front [28]. In essence, the larger the crowding distance is, the fewer solutions occupy the vicinity of a given individual. It is computed, for each objective m, in the following manner: sort individuals according to their fitness score f, assign an infinite value to boundary solutions (so that they are always selected) and compute the distance measure for the remaining solutions as given by (4.8). The overall crowding distance is calculated as the sum of the distance values concerning each objective.

$$F(i) = F(i) + \frac{f(i+1)_m - f(i-1)_m}{f_m^{max} - f_m^{min}}. \tag{4.8}$$

For parent selection, the population is ranked using non-dominated sorting and then, within each rank, individuals are sorted by crowding distance (the higher, the better). Parents are then chosen using binary tournaments and reproduce until N offspring are generated.

With respect to the replacement strategy, NSGA-II uses an elitist approach.

Firstly, after evaluating the newly created offspring, the $2N$ individuals (current generation and offspring) are ranked with non-dominated sorting. Then, entire fronts are added into the new generation, starting from the rank 1 individuals, followed by rank 2, and so on, until a new set can no longer be accommodated. This last set of solutions is then sorted according to crowding distance, and the better ones are chosen to fill out the rest of the new population.

SMS-EMOA (modified)

In the original implementation of the algorithm [18], a single individual is generated by selecting random parents and applying variation operators. Then, the population is ranked using non-dominated sorting, just as in NSGA-II. Finally, the last rank is sorted according to the hypervolume contribution $\Delta_{\mathscr{S}}$ of each solution. This is accomplished by first computing the hypervolume indicator $\mathscr{S}$ considering all the individuals belonging to that rank, computing the same measure but without accounting for the current solution s and then subtracting them, as given by (4.9):

$$\Delta_{\mathscr{S}}(s,R_v) = \mathscr{S}(R_v) - \mathscr{S}(R_v \setminus \{s\}), \tag{4.9}$$

where R_v denotes the worst ranked front. In the current study, the hypervolume indicator is computed using the implementation found in the PyGMO library [19], which uses dimension-dependent methods described in [80].

For this study, however, a number of modifications were made in order to better suit the intended application of the algorithm and adequately explore the search space within the same number of generations. Hence, after non-dominated sorting, individuals within each rank are sorted based on the hypervolume contribution $\Delta_{\mathscr{S}}$ metric described before. Parents are selected through binary tournaments and reproduce until N offspring are generated (as opposed to creating just one individual from two random parents).

After evaluating the $2N$ individuals in the combined population, these are ranked with non-dominated sorting and entire fronts are placed into the new generation until they no longer fit. The last front which cannot fit in its entirety is then sorted using the $\Delta_{\mathscr{S}}$ metric and the best individuals are selected to complete the population. Essentially, this modified version of SMS-EMOA closely resembles NSGA-II with the exception of the crowding distance, which is replaced by the $\Delta_{\mathscr{S}}$ metric, in which solutions located in high curvature regions of the front are valued relatively higher [18].

4.4.5 Variation operators

Between parent selection and the replacement strategy, the variation operators are applied in order to create new individuals, that is, new candidate solutions for the problem. Recombination merges the information contained within the genotypes of two parents probabilistically, according to a crossover rate ρ_c, copying one of the parents otherwise. Mutation is then applied with a probability ρ_m (mutation rate) to the generated offspring, introducing a small, random change to one of the genes.

Mutation

An individual can be mutated in one of two ways: one of its features has one of its properties altered, or the individual's minimum pre-ictal period is changed. Weighted probabilities are employed to ensure that the search space is explored equally across all genes. First, a random choice must be made between mutating P_{min} or one of the features. The probability of choosing either depends on how many features the individual has and the number of values each gene can take (both in terms of period time lengths and feature properties), respectively given by (4.10) and (4.11):

$$P_m(period) = \frac{N_{periods}}{N_{periods} + N_{features} \times N_{properties}}, \tag{4.10}$$

$$P_m(features) = \frac{N_{features} \times N_{properties}}{N_{periods} + N_{features} \times N_{properties}}. \tag{4.11}$$

If the minimum pre-ictal period is chosen, its value is changed to the one immediately above or below it, as illustrated in figure 4.4. Otherwise, one of the individual's features is randomly chosen (with equal probability), and a weighted probability decides which of its properties is mutated according to the number of values each one can take, as expressed by (4.12):

$$P_m(property_i) = \frac{N_{values-1}(property_i)}{\sum_{i=1}^{N_{properties}} N_{values}(property_i)}. \tag{4.12}$$

Each property can then be changed depending on the connected nodes in the respective graph:

- All binary genes (active feature group, active time feature, active frequency feature, active frequency band feature) simply switch to the other possible value;
- Electrode: switch to one of the neighboring electrodes (e.g. the FP1 electrode can mutate to FP2, FZ, F3 or F7);

- Mathematical operator: switch to any of the other two possible values, since all the nodes are connected;
- Every other gene changes to the value immediately above or below it, since each of those graphs represent a sequence.

Essentially, once the target gene is chosen, the mutation operator is applied by changing the current value to one of the neighboring nodes.

Recombination

When two individuals are recombined, they merge genes from each one of their features and their minimum pre-ictal period. However, prior to recombining each feature pair, a distance measure is computed between each feature from each parent, in order to only merge features that are conceptually close to one another. For instance, it is more sensible to recombine two time domain features than it is to merge a time domain feature with another from the frequency domain.

The distance between two properties (one from each feature) is obtained by computing the shortest distance between them (in its respective graph) and dividing it by the number of possible values that property can take. Then, the distance $D(f_i,f_j)$ between two features (one from each parent) is the sum of all the distances between properties, as expressed by (4.13):

$$D(f_i,f_j) = \sum_{k=1}^{N_{properties}} \frac{d(f_i(property_k), f_j(property_k))}{N_{values}(property_i)}. \tag{4.13}$$

The features of each parent are then sorted according to $D(f_i,f_j)$. The sorting is done iteratively, meaning that for each feature from parent 1, the feature with the lowest distance value from parent 2 is assigned and matched to it. A potential drawback from this is that after matching that feature to one from parent 1, it can no longer be reassigned even if a lower $D(f_i,f_j)$ is found.

Once all the features from both parents have been matched according to the distance measure, their properties are recombined in the following manner:

- If both parents have the same gene value, the offspring will take that value;
- If the parents have values in neighboring nodes, one of them is randomly chosen;
- Electrode: first, a random path from all the shortest paths in the electrode graph starting from one of the parents' electrodes to the other is selected; then, a random electrode within that path is chosen (e.g. for parents with electrodes F3 and F4, the offspring may copy one of them or become CZ or FZ);

- In every other gene, a value between those of the parents is chosen.

Additionally, the same conditions (except the one specific to the electrode gene) apply when recombining minimum pre-ictal period (P_{min}) genes from two parents.

4.5 Experimental setup

Using the methodology described in the previous section as well as the data presented in section 4.2, various scenarios will be tested in order to evaluate different possible configurations of the seizure prediction scheme. Namely, three different multi-objective implementations (MOGA, NSGA-II and a modified SMS-EMOA), considering three objectives (sample sensitivity, time under false alarm and the number of different electrodes/lobes within the feature set), as per figure 4.8 below.

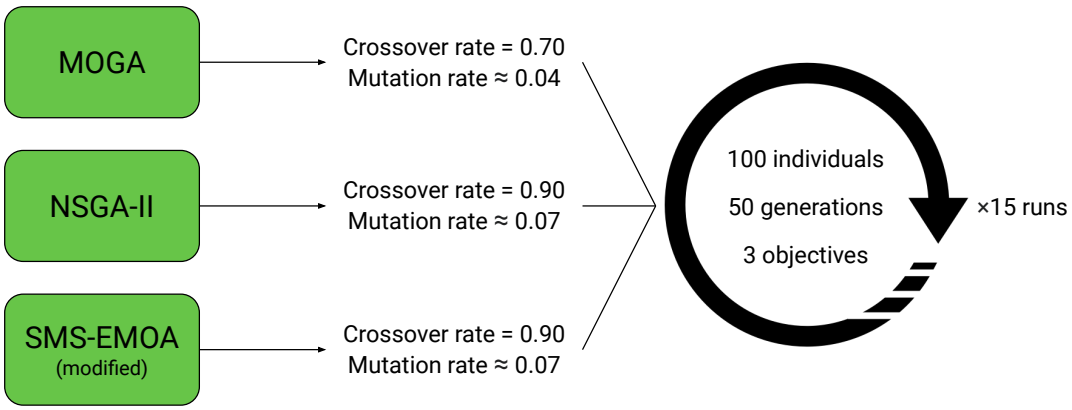


Figure 4.8: Flowchart illustrating the experimental setup of this study for the training phase.

Concerning parameter tuning for each implementation, their mutation and crossover rates were chosen according to the ones suggested by the original authors [28, 40]: $\rho_c = 0.90$ and $\rho_m \approx 0.07$ for NSGA-II and the modified SMS-EMOA; $\rho_c = 0.70$ and $\rho_m \approx 0.04$ for MOGA. Considering $\ell = 14$, given that the used genotype comprises 14 genes, the mutation rates were computed using (4.14) for both NSGA-II and the modified SMS-EMOA and (4.15) in the case of MOGA:

$$\rho_m = \frac{1}{\ell}, \tag{4.14}$$

$$\rho_m = 1 - (\alpha\mu)^{-\frac{1}{\ell}}, \tag{4.15}$$

where $\alpha = 0.9$ and $\mu = 2$, as indicated in [40]. It is worth noting that the modified

SMS-EMOA has the same parameterization as NSGA-II since, as stated in section 4.4.4, similar selection and replacement strategies are used.

In a real-life context, the algorithm would be executed once on a set of training seizures and a single minimum pre-ictal period as well as a single set of features would be chosen to predict future seizures. However, given the academic context and exploratory nature of this study, 15 runs of the evolutionary search are performed for each patient and each MOEA variant. In order evaluate the performance range of each algorithm, multiple solutions within each run will be tested on unseen data and the mean seizure sensitivity and FPR/h will be presented.

Moreover, a Decision Maker (DM) is implemented to select which individuals from the set of Pareto-optimal solutions are used in the testing phase. Hence, only individuals with a sufficiently high fitness score in the first two objectives (sample sensitivity and time under false alarm) are tested. This restriction is put in place due to the inclusion of the third objective, which despite finding solutions that extract information from a low number of electrodes, results in some Pareto-optimal solutions with inadequate classification performance within the training set.

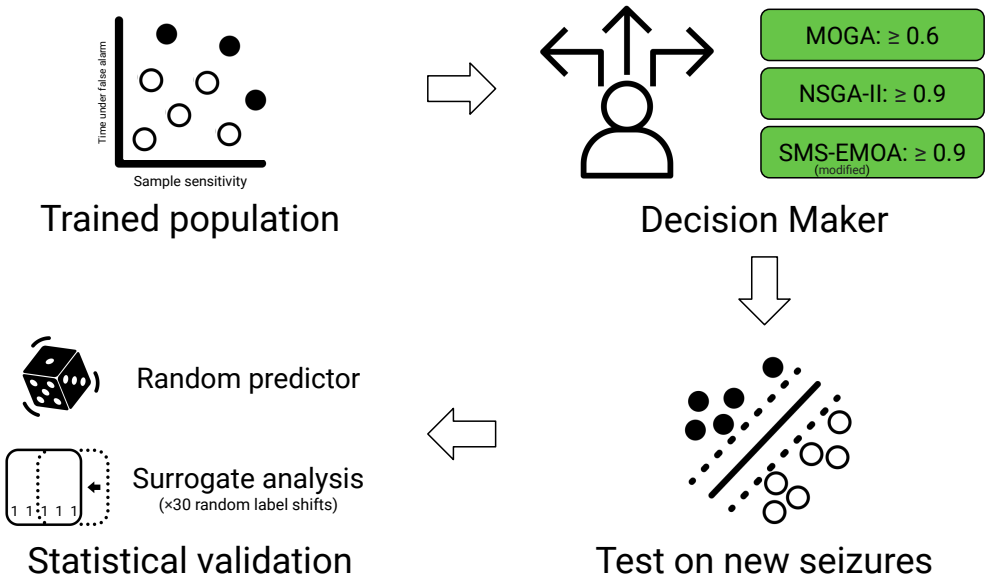


Figure 4.9: Flowchart illustrating the experimental setup of this study for the testing phase.

As shown in figure 4.9, for both elitist variants (NSGA-II and the modified SMS-EMOA), a minimum fitness threshold for the DM was set to 0.9 considering that, for every execution, a considerable number of solutions that met the criteria was always obtained. However, in the case of MOGA, a lower limit of 0.6 was chosen given that fitness scores were relatively lower than the former. In the event where

no solutions could be selected within a run, the threshold was decreased further to 0.5 to guarantee that the DM selected at least one individual from that execution.

Concerning statistical validation, the algorithm's performance will be evaluated by comparing it against the random predictor (based on a Poisson distribution) as well as through surrogate time series analysis. The former's sensitivity was computed using (2.8) considering $d = 1$ to validate each model individually and $d = N_m$ for the average random predictor (to validate the set of N_m patient-specific models). The SOP and FPR/h of each individual is used for (2.7) when comparing each individual to a random predictor, and the average of each one is used when validating the average sensitivity from all the solutions. Furthermore, the product between the number of solutions and the number of testing seizures is used for the k parameter in (2.8) in the case of the average random predictor.

Regarding the surrogate predictor, the same prediction pipeline is used for each of the selected individuals. However, for each tested seizure, the correspondent preictal labels were randomly shifted in time within the duration of the recording. After performing random label shifts 30 times, the average sensitivity is computed and compared against the one obtained for each individual. The average sensitivity from all selected solutions is also validated against the average of the previous surrogate sensitivity values.

As stated in section 2.3.3, in order to consider each solution's performance as above chance level, it must meet the requirement of beating the random/surrogate predictor, that is, achieve a higher sensitivity with statistical significance. In this study, a significance level of $\alpha = 0.05$ is considered.

Moreover, the ratio of solutions that outperform each of the random or surrogate predictors is also statistically validated. This validation is performed using a procedure inspired by Alvarado-Rojas et al. [7], who used a binomial distribution to verify if the number of validated patients was statistically significant for the whole group. Thus, the probability of at least i of I solutions beating the random or surrogate predictors is given by the following expression:

$$P_{binom}(i,I,\alpha) = \sum_{j=1}^{I} \binom{I}{j} \alpha^j (1-\alpha)^{(I-j)}. \tag{4.16}$$

In this study, the ratio of statistically valid solutions is considered significant if higher than the ratio obtained by (4.16), considering $\alpha = 0.05$.

Finally, a phenotype study is conducted at two different levels: for the whole group and for a specific patient. In the former case, SOP duration will be analyzed in order to check for any correlation to the observed prediction performance. Con-

cerning the latter, the phenotypes of the solution set will be examined in terms of gene prevalence to showcase the possibility of extrapolating clinical knowledge using the proposed methodology.

4.6 Summary

The proposed methodology for seizure prediction is based on scalp EEG data from 36 patients undergoing pre-surgical monitoring and suffering from TLE, retrieved from the EPILEPSIAE database. In addition to signal data concerning 19 electrodes (following the 10-20 system), extensive metadata including seizure type/pattern classification and vigilance state was also used for stratification.

The time-series data was segmented using 5-second windows with no overlap. 4 hours preceding each seizure onset are used for training, while the testing set of seizures includes all available recordings starting from 30 minutes after the previous ictal event. Signals were filtered using a notch filter at 50Hz as well as a high-pass filter with a cut-off frequency at 0.5Hz. 24 linear univariate features were extracted from each channel: four statistical moments, three Hjorth parameters, relative spectral power (delta, theta, beta, alpha, low and high gamma), spectral edge frequency (with three different cut-offs) and wavelet coefficient energy (from D1 to D8, considering the Daubechies 4 mother wavelet).

A multi-objective evolutionary approach is used to perform high-level feature extraction as well as feature and parameter selection (namely, the chosen minimum pre-ictal period, which will dictate the duration of SOP). Each individual in the population represents a possible configuration for a machine learning prediction algorithm. The evolutionary search is run and trained on a set of training seizures (first 3 chronological seizures), while the chosen solutions are applied on a set of testing seizures, unknown to the prediction algorithm.

In each generation, a population of 100 individuals, each representing a set of 5 features and a minimum pre-ictal period, are evaluated. The high-level features are constructed from the linear univariate features using a 30-second sliding window with no overlap. The SOP duration is equal to the sum of the individual's minimum pre-ictal period and the shortest delay found within its feature set, meaning it can range from 20 to 80 minutes. Class weights are applied to each sample, and an SPH is introduced by assigning a null weight to the samples from the last 10 minutes before the onset. A linear SVM classifier is trained iteratively, accumulating data from each seizure and testing on the following one. The classification output is processed using the Firing Power method with a threshold of 0.7 and a refractory

behaviour is also implemented to produce a more realistic alarm firing mechanism. Each of the first two evaluated objectives concern a performance metric: sample sensitivity and time under false alarm. A third objective is introduced to account for how many different electrodes and lobes are found within the feature set. For each fitness function, maximization is assumed.

Three different MOEA implementations were used: Fonseca and Fleming's MOGA [41], Deb et al.'s NSGA-II [28] and a modified version of Beume et al.'s SMS-EMOA [18]. After evaluating the population, parents are selected depending on the chosen method and recombined to produce offspring. The resulting offspring is then mutated (with a given probability) and a replacement strategy is employed, also depending on the selected implementation. Once the stopping criteria are met (a maximum of 50 generations), the Pareto optimal solutions are selected and applied on the testing set, with their performance evaluated using the gold standard metrics for seizure prediction: seizure sensitivity and FPR/h. Statistical validation is also performed against the random predictor and by surrogate time series analysis. Besides patient stratification, a phenotype study was also conducted.

5

Results and Discussion

This chapter presents the results obtained in the training and testing phases (sections 5.1 and 5.2, respectively) along with their comprehensive analysis. Section 5.3 contains a phenotype study, followed by a brief summary provided in section 5.4.

5.1 Training phase

As mentioned in section 4.5, for each patient, 15 executions of the evolutionary search were run for each of the three considered Multi-Objective Evolutionary Algorithms (MOEAs). Table 5.1 shows the average fitness scores of the entire set of Pareto-optimal individuals found within those runs.

Overall, for each of the three objectives, that is, sample sensitivity ($S_{samples}$), time under false alarm (T_f) and number of different electrodes/lobes within the feature set (N_E), the following fitness values were obtained: 0.62 ± 0.29, 0.82 ± 0.16 and 0.78 ± 0.19 for MOGA; 0.88 ± 0.16, 0.92 ± 0.10 and 0.86 ± 0.14 for NSGA-II; 0.88 ± 0.16, 0.93 ± 0.09 and 0.83 ± 0.16 for the modified SMS-EMOA. Statistically significant differences between them were found for every MOEA except between NSGA-II and the modified SMS-EMOA in the first objective (see table B.2), considering a two-tailed t-test with a significance level of $\alpha = 0.05$.

Focusing on the fitness values of the first two objectives, it can be seen that prediction performance is generally satisfactory, displaying averages above 0.80 with the exception of the populations trained with MOGA. This is to be expected considering that both NSGA-II and the modified SMS-EMOA employ elitist replacement strategies, which guarantee that the individuals from the best rank are kept for the next generation. In contrast, MOGA uses a generational replacement method, in which non-dominated solutions can potentially be lost.

However, the standard deviations are considerably large indicating that, despite the abundance of high-performing solutions, others which perform worse are also present. Additionally, it is worth noting the effect of the third objective (N_E), since

the number of different electrodes/lobes in the feature set is not indicative of the prediction performance, meaning that a given solution may be Pareto-optimal due to a high fitness score in the last objective at the expense of the first two objectives. Furthermore, as the number of objectives increases, more individuals tend to lie in the first non-dominated front, resulting in a decrease of selection pressure [58].

Table 5.1: Fitness scores obtained in the evolutionary algorithm training phase.

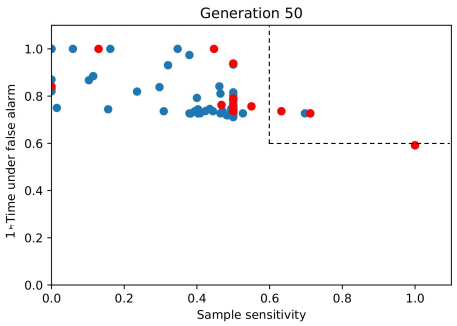
	MOGA			NSGA-II			SMS-EMOA (modified)		
Patient ID	$S_{samples}$	T_f	N_E	$S_{samples}$	T_f	N_E	$S_{samples}$	T_f	N_E
402	0.59 ± 0.28	0.78 ± 0.18	0.76 ± 0.20	0.89 ± 0.17	0.84 ± 0.14	0.89 ± 0.12	0.90 ± 0.17	0.90 ± 0.13	0.88 ± 0.14
8902	0.67 ± 0.21	0.75 ± 0.07	0.75 ± 0.17	0.81 ± 0.16	0.82 ± 0.09	0.83 ± 0.16	0.80 ± 0.17	0.83 ± 0.10	0.81 ± 0.17
11002	0.69 ± 0.23	0.76 ± 0.22	0.80 ± 0.20	0.95 ± 0.07	0.94 ± 0.13	0.88 ± 0.10	0.95 ± 0.07	0.96 ± 0.09	0.80 ± 0.19
16202	0.49 ± 0.32	0.77 ± 0.19	0.74 ± 0.19	0.85 ± 0.19	0.89 ± 0.13	0.86 ± 0.13	0.83 ± 0.15	0.89 ± 0.10	0.85 ± 0.14
23902	0.71 ± 0.26	0.84 ± 0.15	0.78 ± 0.20	0.92 ± 0.11	0.94 ± 0.06	0.82 ± 0.16	0.94 ± 0.08	0.95 ± 0.06	0.75 ± 0.17
30802	0.42 ± 0.24	0.93 ± 0.11	0.82 ± 0.18	0.90 ± 0.14	0.98 ± 0.06	0.80 ± 0.18	0.83 ± 0.22	0.96 ± 0.07	0.87 ± 0.12
32702	0.67 ± 0.25	0.82 ± 0.14	0.73 ± 0.22	0.90 ± 0.11	0.94 ± 0.06	0.86 ± 0.14	0.90 ± 0.13	0.94 ± 0.08	0.82 ± 0.17
46702	0.57 ± 0.31	0.78 ± 0.17	0.78 ± 0.20	0.81 ± 0.20	0.89 ± 0.11	0.85 ± 0.13	0.74 ± 0.20	0.89 ± 0.12	0.84 ± 0.14
50802	0.88 ± 0.21	0.96 ± 0.06	0.91 ± 0.09	0.99 ± 0.01	0.99 ± 0.01	0.93 ± 0.06	0.99 ± 0.01	1.00 ± 0.00	0.88 ± 0.10
53402	0.75 ± 0.23	0.84 ± 0.17	0.89 ± 0.11	0.94 ± 0.12	0.97 ± 0.04	0.88 ± 0.12	0.94 ± 0.10	0.97 ± 0.05	0.83 ± 0.18
55202	0.72 ± 0.25	0.83 ± 0.12	0.85 ± 0.14	0.91 ± 0.10	0.97 ± 0.05	0.88 ± 0.11	0.89 ± 0.13	0.95 ± 0.07	0.81 ± 0.15
56402	0.69 ± 0.23	0.77 ± 0.16	0.77 ± 0.18	0.91 ± 0.11	0.86 ± 0.11	0.83 ± 0.15	0.88 ± 0.14	0.88 ± 0.10	0.78 ± 0.18
58602	0.55 ± 0.30	0.78 ± 0.18	0.76 ± 0.21	0.90 ± 0.12	0.90 ± 0.11	0.82 ± 0.16	0.91 ± 0.10	0.92 ± 0.10	0.81 ± 0.21
59102	0.74 ± 0.24	0.92 ± 0.12	0.80 ± 0.19	0.99 ± 0.01	1.00 ± 0.02	0.89 ± 0.14	0.99 ± 0.05	0.99 ± 0.01	0.91 ± 0.09
60002	0.65 ± 0.30	0.83 ± 0.15	0.82 ± 0.19	0.87 ± 0.13	0.94 ± 0.07	0.86 ± 0.12	0.92 ± 0.12	0.95 ± 0.05	0.76 ± 0.22
64702	0.75 ± 0.23	0.80 ± 0.15	0.81 ± 0.18	0.95 ± 0.09	0.96 ± 0.06	0.89 ± 0.12	0.93 ± 0.10	0.91 ± 0.08	0.81 ± 0.16
75202	0.52 ± 0.29	0.84 ± 0.16	0.75 ± 0.21	0.88 ± 0.14	0.94 ± 0.07	0.80 ± 0.18	0.90 ± 0.13	0.94 ± 0.08	0.81 ± 0.17
80702	0.63 ± 0.30	0.77 ± 0.15	0.72 ± 0.20	0.80 ± 0.20	0.91 ± 0.10	0.85 ± 0.13	0.86 ± 0.17	0.90 ± 0.11	0.81 ± 0.17
85202	0.41 ± 0.21	0.86 ± 0.12	0.76 ± 0.19	0.81 ± 0.22	0.89 ± 0.13	0.82 ± 0.14	0.83 ± 0.21	0.90 ± 0.12	0.82 ± 0.13
93402	0.68 ± 0.26	0.87 ± 0.14	0.81 ± 0.15	0.94 ± 0.10	0.96 ± 0.07	0.89 ± 0.13	0.96 ± 0.09	0.95 ± 0.08	0.87 ± 0.10
93902	0.57 ± 0.26	0.80 ± 0.18	0.74 ± 0.22	0.80 ± 0.19	0.91 ± 0.11	0.84 ± 0.14	0.82 ± 0.19	0.88 ± 0.13	0.81 ± 0.15
94402	0.69 ± 0.26	0.80 ± 0.18	0.79 ± 0.19	0.89 ± 0.14	0.94 ± 0.08	0.86 ± 0.16	0.90 ± 0.14	0.94 ± 0.07	0.89 ± 0.09
95202	0.62 ± 0.33	0.81 ± 0.15	0.78 ± 0.21	0.84 ± 0.19	0.91 ± 0.10	0.89 ± 0.11	0.85 ± 0.14	0.93 ± 0.06	0.82 ± 0.21
96002	0.73 ± 0.25	0.94 ± 0.06	0.84 ± 0.17	0.96 ± 0.07	0.99 ± 0.01	0.84 ± 0.16	0.98 ± 0.04	0.99 ± 0.01	0.92 ± 0.07
98102	0.55 ± 0.30	0.81 ± 0.15	0.73 ± 0.21	0.77 ± 0.19	0.86 ± 0.11	0.83 ± 0.17	0.82 ± 0.18	0.88 ± 0.11	0.80 ± 0.16
98202	0.30 ± 0.23	0.87 ± 0.10	0.78 ± 0.19	0.76 ± 0.22	0.90 ± 0.12	0.84 ± 0.13	0.77 ± 0.23	0.89 ± 0.12	0.77 ± 0.21
101702	0.62 ± 0.31	0.78 ± 0.16	0.81 ± 0.18	0.92 ± 0.11	0.92 ± 0.08	0.87 ± 0.13	0.94 ± 0.12	0.92 ± 0.10	0.84 ± 0.16
102202	0.56 ± 0.30	0.82 ± 0.15	0.76 ± 0.21	0.84 ± 0.17	0.89 ± 0.09	0.83 ± 0.17	0.83 ± 0.18	0.88 ± 0.09	0.83 ± 0.16
104602	0.67 ± 0.29	0.85 ± 0.14	0.81 ± 0.19	0.94 ± 0.08	0.94 ± 0.08	0.91 ± 0.09	0.91 ± 0.10	0.95 ± 0.06	0.90 ± 0.11
109502	0.69 ± 0.26	0.78 ± 0.18	0.80 ± 0.18	0.96 ± 0.09	0.94 ± 0.10	0.90 ± 0.09	0.9 ± 0.17	0.97 ± 0.06	0.91 ± 0.07
110602	0.56 ± 0.32	0.83 ± 0.16	0.80 ± 0.16	0.94 ± 0.10	0.93 ± 0.10	0.85 ± 0.12	0.93 ± 0.11	0.94 ± 0.08	0.80 ± 0.15
112802	0.61 ± 0.30	0.76 ± 0.17	0.78 ± 0.19	0.78 ± 0.20	0.91 ± 0.08	0.84 ± 0.16	0.84 ± 0.16	0.91 ± 0.08	0.83 ± 0.18
113902	0.67 ± 0.28	0.81 ± 0.17	0.80 ± 0.17	0.86 ± 0.16	0.92 ± 0.08	0.80 ± 0.18	0.85 ± 0.16	0.93 ± 0.08	0.84 ± 0.13
114702	0.72 ± 0.25	0.81 ± 0.16	0.81 ± 0.16	0.96 ± 0.05	0.94 ± 0.08	0.88 ± 0.13	0.95 ± 0.06	0.96 ± 0.07	0.89 ± 0.08
114902	0.60 ± 0.32	0.79 ± 0.14	0.80 ± 0.17	0.89 ± 0.15	0.97 ± 0.05	0.86 ± 0.14	0.91 ± 0.14	0.96 ± 0.05	0.85 ± 0.12
123902	0.63 ± 0.23	0.91 ± 0.13	0.80 ± 0.18	0.93 ± 0.11	0.96 ± 0.05	0.85 ± 0.13	0.90 ± 0.12	0.97 ± 0.06	0.77 ± 0.19

$S_{samples}$: sample sensitivity; T_f: time under false alarm; N_E: number of different electrodes/lobes within the feature set

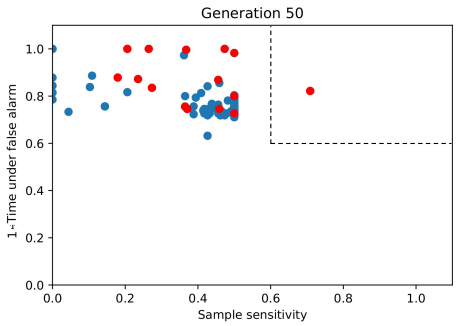
For the reasons given above, a Decision Maker (DM) was employed to select Pareto-optimal solutions with high fitness scores for the first two objectives ($S_{samples}$ and T_f), as described in section 4.5. The threshold values (figure 4.8) were chosen in accordance with the average fitness values observed for those objectives in each algorithm variant. Table 5.2 shows the fitness scores of the resulting set of individuals, while figure 5.1 illustrates the decision making process for each MOEA.

By analyzing the 2D scatter plots (selected from the training phase of patient 85202), it becomes even more evident why a DM was implemented. In the case of MOGA (figures 5.1a and 5.1b), most Pareto-optimal solutions have a low fitness score in the $S_{samples}$ objective, despite an overall high fitness in the T_f objective

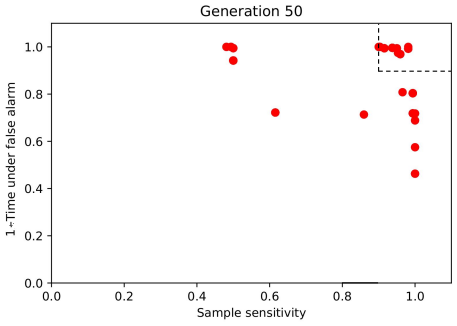
(i.e. solutions display a high specificity at the cost of a low sensitivity). As for NSGA-II (figures 5.1c and 5.1d) and the modified SMS-EMOA (figures 5.1e and 5.1f), there is a clear tendency for most solutions to belong in the non-dominated front, even when a low fitness score is obtained in the first two objectives. Hence, only high-performing solutions (within the training set) were kept for testing.



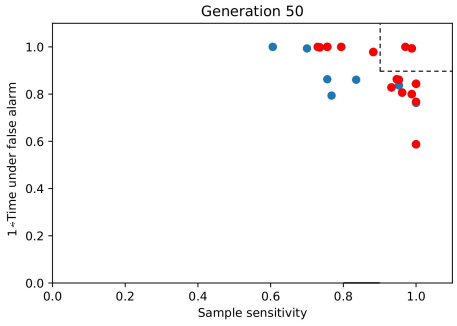
(a) Patient 85202: MOGA Run #8



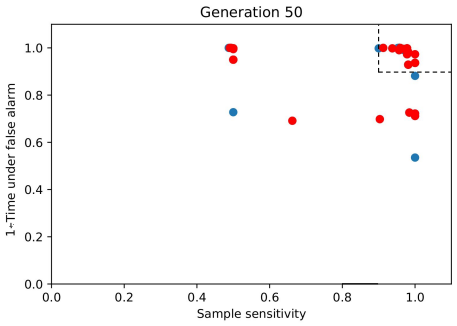
(b) Patient 85202: MOGA Run #12



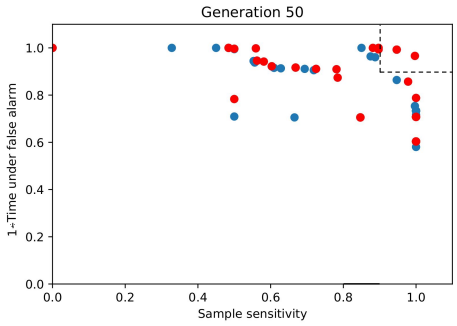
(c) Patient 85202: NSGA-II Run #3



(d) Patient 85202: NSGA-II Run #12



(e) Patient 85202: SMS-EMOA Run #3



(f) Patient 85202: SMS-EMOA Run #9

Figure 5.1: 2D scatter plots representing the fitness scores of two selected runs from each algorithm variant. Pareto-optimal solutions are represented in red. The Decision Maker thresholds are illustrated by dashed lines.

Table 5.2: Fitness scores of the solution set selected by the Decision Maker.

	MOGA			NSGA-II			SMS-EMOA (modified)		
Patient ID	$S_{samples}$	T_f	N_E	$S_{samples}$	T_f	N_E	$S_{samples}$	T_f	N_E
402	0.80 ± 0.11	0.72 ± 0.07	0.64 ± 0.24	0.97 ± 0.03	0.98 ± 0.02	0.84 ± 0.15	0.98 ± 0.02	0.99 ± 0.02	0.85 ± 0.15
8902	0.79 ± 0.11	0.74 ± 0.04	0.74 ± 0.17	0.85 ± 0.04	0.86 ± 0.04	0.83 ± 0.13	0.85 ± 0.06	0.85 ± 0.04	0.78 ± 0.18
11002	0.77 ± 0.15	0.83 ± 0.15	0.71 ± 0.25	0.98 ± 0.03	0.99 ± 0.01	0.88 ± 0.1	0.98 ± 0.02	0.98 ± 0.03	0.82 ± 0.15
16202	0.69 ± 0.14	0.71 ± 0.09	0.70 ± 0.20	0.96 ± 0.05	0.97 ± 0.04	0.87 ± 0.11	0.96 ± 0.04	0.94 ± 0.04	0.83 ± 0.15
23902	0.84 ± 0.13	0.82 ± 0.11	0.73 ± 0.21	0.97 ± 0.03	0.96 ± 0.03	0.78 ± 0.18	0.96 ± 0.03	0.97 ± 0.03	0.72 ± 0.17
30802	0.77 ± 0.16	0.88 ± 0.09	0.65 ± 0.24	0.98 ± 0.03	0.99 ± 0.01	0.77 ± 0.19	0.99 ± 0.02	0.98 ± 0.03	0.86 ± 0.15
32702	0.85 ± 0.12	0.76 ± 0.09	0.66 ± 0.25	0.97 ± 0.03	0.96 ± 0.03	0.77 ± 0.16	0.97 ± 0.03	0.98 ± 0.02	0.82 ± 0.13
46702	0.82 ± 0.14	0.75 ± 0.09	0.64 ± 0.23	0.93 ± 0.06	0.94 ± 0.05	0.82 ± 0.12	0.90 ± 0.06	0.91 ± 0.05	0.72 ± 0.14
50802	0.95 ± 0.05	0.96 ± 0.07	0.91 ± 0.09	0.99 ± 0.01	0.99 ± 0.01	0.93 ± 0.06	0.99 ± 0.01	1.00 ± 0.00	0.88 ± 0.10
53402	0.85 ± 0.12	0.85 ± 0.12	0.89 ± 0.09	0.98 ± 0.02	0.97 ± 0.02	0.87 ± 0.13	0.97 ± 0.03	0.98 ± 0.02	0.86 ± 0.12
55202	0.89 ± 0.11	0.77 ± 0.09	0.83 ± 0.15	0.96 ± 0.04	0.98 ± 0.02	0.89 ± 0.10	0.95 ± 0.05	0.97 ± 0.04	0.76 ± 0.15
56402	0.79 ± 0.15	0.76 ± 0.09	0.72 ± 0.20	0.93 ± 0.05	0.95 ± 0.04	0.79 ± 0.15	0.96 ± 0.04	0.93 ± 0.05	0.73 ± 0.15
58602	0.79 ± 0.11	0.78 ± 0.10	0.62 ± 0.19	0.96 ± 0.03	0.97 ± 0.03	0.75 ± 0.19	0.98 ± 0.03	0.96 ± 0.03	0.80 ± 0.17
59102	0.89 ± 0.11	0.92 ± 0.11	0.76 ± 0.21	0.99 ± 0.01	1.00 ± 0.01	0.89 ± 0.14	0.99 ± 0.01	0.99 ± 0.01	0.91 ± 0.09
60002	0.87 ± 0.12	0.76 ± 0.10	0.80 ± 0.20	0.96 ± 0.04	0.96 ± 0.03	0.82 ± 0.11	0.96 ± 0.04	0.96 ± 0.03	0.72 ± 0.19
64702	0.86 ± 0.12	0.79 ± 0.09	0.74 ± 0.20	0.98 ± 0.03	0.98 ± 0.03	0.87 ± 0.13	0.97 ± 0.03	0.94 ± 0.02	0.76 ± 0.18
75202	0.79 ± 0.13	0.78 ± 0.11	0.69 ± 0.22	0.97 ± 0.03	0.96 ± 0.03	0.71 ± 0.21	0.98 ± 0.03	0.97 ± 0.03	0.73 ± 0.17
80702	0.85 ± 0.13	0.74 ± 0.06	0.62 ± 0.20	0.96 ± 0.03	0.96 ± 0.04	0.77 ± 0.12	0.97 ± 0.04	0.94 ± 0.04	0.79 ± 0.14
85202	0.64 ± 0.15	0.76 ± 0.05	0.68 ± 0.21	0.95 ± 0.04	0.97 ± 0.05	0.73 ± 0.12	0.97 ± 0.04	0.97 ± 0.04	0.72 ± 0.11
93402	0.83 ± 0.12	0.85 ± 0.12	0.79 ± 0.15	0.98 ± 0.02	0.99 ± 0.02	0.85 ± 0.14	0.98 ± 0.02	0.99 ± 0.02	0.85 ± 0.09
93902	0.78 ± 0.18	0.74 ± 0.12	0.56 ± 0.24	0.96 ± 0.06	0.94 ± 0.05	0.80 ± 0.16	0.97 ± 0.04	0.93 ± 0.07	0.75 ± 0.19
94402	0.83 ± 0.13	0.78 ± 0.11	0.73 ± 0.19	0.96 ± 0.03	0.97 ± 0.03	0.81 ± 0.19	0.97 ± 0.03	0.97 ± 0.03	0.86 ± 0.11
95202	0.87 ± 0.12	0.74 ± 0.09	0.76 ± 0.20	0.97 ± 0.04	0.96 ± 0.05	0.85 ± 0.12	0.95 ± 0.05	0.92 ± 0.04	0.78 ± 0.22
96002	0.89 ± 0.11	0.92 ± 0.06	0.84 ± 0.18	0.98 ± 0.02	0.99 ± 0.01	0.84 ± 0.17	0.99 ± 0.02	0.99 ± 0.01	0.92 ± 0.07
98102	0.81 ± 0.12	0.75 ± 0.08	0.66 ± 0.20	0.89 ± 0.06	0.88 ± 0.06	0.70 ± 0.21	0.94 ± 0.07	0.94 ± 0.06	0.65 ± 0.16
98202	0.58 ± 0.12	0.78 ± 0.08	0.76 ± 0.20	0.97 ± 0.03	0.97 ± 0.03	0.70 ± 0.12	0.96 ± 0.03	0.96 ± 0.03	0.59 ± 0.21
101702	0.83 ± 0.13	0.75 ± 0.11	0.71 ± 0.21	0.95 ± 0.05	0.96 ± 0.05	0.88 ± 0.11	0.98 ± 0.03	0.97 ± 0.03	0.84 ± 0.15
102202	0.82 ± 0.14	0.76 ± 0.07	0.69 ± 0.21	0.95 ± 0.05	0.92 ± 0.04	0.69 ± 0.21	0.93 ± 0.07	0.89 ± 0.06	0.72 ± 0.17
104602	0.89 ± 0.12	0.80 ± 0.10	0.80 ± 0.18	0.97 ± 0.02	0.98 ± 0.02	0.89 ± 0.10	0.97 ± 0.03	0.96 ± 0.03	0.87 ± 0.14
109502	0.85 ± 0.10	0.79 ± 0.13	0.78 ± 0.19	0.98 ± 0.03	0.99 ± 0.01	0.88 ± 0.10	0.99 ± 0.01	0.98 ± 0.02	0.90 ± 0.07
110602	0.80 ± 0.13	0.80 ± 0.08	0.73 ± 0.17	0.97 ± 0.03	0.98 ± 0.02	0.82 ± 0.13	0.96 ± 0.03	0.98 ± 0.02	0.76 ± 0.15
112802	0.83 ± 0.14	0.73 ± 0.06	0.70 ± 0.19	0.94 ± 0.06	0.91 ± 0.06	0.77 ± 0.16	0.95 ± 0.06	0.93 ± 0.05	0.76 ± 0.23
113902	0.85 ± 0.12	0.78 ± 0.10	0.74 ± 0.18	0.96 ± 0.03	0.95 ± 0.03	0.66 ± 0.19	0.97 ± 0.03	0.95 ± 0.03	0.76 ± 0.13
114702	0.87 ± 0.13	0.77 ± 0.10	0.78 ± 0.17	0.97 ± 0.03	0.98 ± 0.02	0.87 ± 0.14	0.97 ± 0.03	0.98 ± 0.02	0.90 ± 0.06
114902	0.87 ± 0.12	0.74 ± 0.07	0.76 ± 0.19	0.98 ± 0.02	0.97 ± 0.02	0.86 ± 0.13	0.97 ± 0.02	0.98 ± 0.03	0.82 ± 0.11
123902	0.84 ± 0.12	0.83 ± 0.11	0.75 ± 0.21	0.98 ± 0.02	0.97 ± 0.03	0.82 ± 0.14	0.97 ± 0.03	0.99 ± 0.02	0.68 ± 0.21

$S_{samples}$: sample sensitivity; T_f: time under false alarm; N_E: number of different electrodes/lobes within the feature set

Overall, the solution set selected by the DM resulted in a substantial improvement of fitness scores across all patients, objectives and used algorithms, most prominently in the MOGA solution sets. Fitness scores average at 0.83 ± 0.14, 0.79 ± 0.11 and 0.74 ± 0.20 for MOGA, 0.97 ± 0.04, 0.97 ± 0.04 and 0.83 ± 0.16 for NSGA-II and, finally, 0.97 ± 0.04, 0.97 ± 0.04 and 0.81 ± 0.16 for the modified SMS-EMOA. Statistically significant differences between them were identified for every MOEA except between NSGA-II and the modified SMS-EMOA in the first and second objectives (see table B.3), using a two-tailed t-test with a significance level of $\alpha = 0.05$.

5.2 Testing phase

Each of the solutions selected by the DM was tested on new set of seizures, unknown to the machine learning models. Their performance was then assessed in terms of sensitivity and False Positive Rate per Hour (FPR/h), along with statistical validation against a random predictor and surrogate analysis, as per section 4.5.

5.2.1 MOGA

Table 5.3: Seizure prediction performance and statistical validation results obtained for the models optimized with MOGA.

Patient ID	Sensitivity	FPR/h	Random sensitivity	Solutions beat random (%)	Surrogate sensitivity	Solutions beat surrogate (%)
402	0.14 ± 0.22	0.14 ± 0.09	0.53 ± 0.12	0.00	0.15 ± 0.09	28.30#
8902	0.25 ± 0.28	0.16 ± 0.09	0.50 ± 0.04*	3.51	0.22 ± 0.07	44.44#
11002	0.16 ± 0.37	0.27 ± 0.18	0.90 ± 0.30*	0.00	0.19 ± 0.12	16.13#
16202	0.11 ± 0.15	0.20 ± 0.11	0.50 ± 0.00*	0.00	0.11 ± 0.06	33.33#
23902	0.08 ± 0.19	0.24 ± 0.10	0.50 ± 0.04*	0.71	0.22 ± 0.10	15.71#
30802	0.34 ± 0.26	0.20 ± 0.08	0.60 ± 0.00*	9.38	0.20 ± 0.06*	68.75#
32702	0.32 ± 0.32	0.29 ± 0.15	0.50 ± 0.06*	8.73#	0.22 ± 0.10*	52.38#
46702	0.27 ± 0.28	0.40 ± 0.17	0.50 ± 0.00	3.08	0.28 ± 0.10	35.38#
50802	0.22 ± 0.25	0.25 ± 0.06	0.50 ± 0.00*	0.00	0.21 ± 0.06	42.31#
53402	0.26 ± 0.30	0.23 ± 0.08	0.50 ± 0.00*	5.81	0.16 ± 0.08*	44.19#
55202	0.32 ± 0.17	0.19 ± 0.05	0.60 ± 0.00*	0.00	0.18 ± 0.04*	54.03#
56402	0.03 ± 0.18	0.26 ± 0.14	0.97 ± 0.16*	0.00	0.18 ± 0.12	3.39
58602	0.06 ± 0.13	0.20 ± 0.09	0.67 ± 0.00*	0.00	0.17 ± 0.07	16.00#
59102	0.13 ± 0.22	0.18 ± 0.06	0.50 ± 0.00	0.00	0.17 ± 0.06	25.74#
60002	0.21 ± 0.20	0.10 ± 0.02	0.67 ± 0.00*	0.00	0.15 ± 0.05*	55.96#
64702	0.09 ± 0.19	0.13 ± 0.09	0.50 ± 0.19*	0.00	0.11 ± 0.09	16.67#
75202	0.29 ± 0.20	0.15 ± 0.06	0.50 ± 0.00*	5.00	0.18 ± 0.07*	61.25#
80702	0.10 ± 0.15	0.19 ± 0.11	0.67 ± 0.00*	0.00	0.17 ± 0.07	24.18#
85202	0.52 ± 0.35	0.30 ± 0.16	0.50 ± 0.00	25.92#	0.20 ± 0.10*	77.78#
93402	0.03 ± 0.12	0.25 ± 0.08	0.50 ± 0.00*	0.00	0.23 ± 0.08	6.45
93902	0.30 ± 0.29	0.16 ± 0.09	0.65 ± 0.10*	7.14	0.16 ± 0.07*	57.14#
94402	0.09 ± 0.14	0.27 ± 0.07	0.50 ± 0.00*	0.00	0.22 ± 0.05	12.00#
95202	0.15 ± 0.17	0.17 ± 0.07	0.50 ± 0.00*	0.00	0.14 ± 0.06	38.20#
96002	0.14 ± 0.16	0.21 ± 0.06	0.50 ± 0.00*	0.00	0.15 ± 0.05	35.56#
98102	0.47 ± 0.30	0.12 ± 0.10	0.59 ± 0.21*	12.00#	0.12 ± 0.10*	77.33#
98202	0.07 ± 0.11	0.29 ± 0.10	0.50 ± 0.00	0.00	0.24 ± 0.08	6.12
101702	0.35 ± 0.30	0.23 ± 0.13	0.50 ± 0.00	7.69	0.19 ± 0.10*	58.24#
102202	0.18 ± 0.19	0.20 ± 0.07	0.50 ± 0.00*	1.39	0.15 ± 0.07	36.11#
104602	0.36 ± 0.36	0.30 ± 0.14	0.50 ± 0.00*	15.24#	0.27 ± 0.08*	51.43#
109502	0.46 ± 0.50	0.17 ± 0.09	1.00 ± 0.00*	0.00	0.14 ± 0.09*	45.74#
110602	0.25 ± 0.33	0.21 ± 0.09	0.49 ± 0.05	9.68#	0.17 ± 0.07*	40.86#
112802	0.06 ± 0.12	0.19 ± 0.05	0.67 ± 0.00*	0.00	0.12 ± 0.06	15.63#
113902	0.31 ± 0.14	0.15 ± 0.13	0.58 ± 0.23*	12.37#	0.13 ± 0.10*	80.41#
114702	0.15 ± 0.16	0.31 ± 0.07	0.60 ± 0.00*	0.00	0.25 ± 0.06	13.82#
114902	0.39 ± 0.22	0.15 ± 0.06	0.50 ± 0.00*	13.26#	0.12 ± 0.06*	84.69#
123902	0.04 ± 0.13	0.10 ± 0.06	0.54 ± 0.16	0.00	0.07 ± 0.06	7.79

* denotes that the average sensitivity obtained is higher than that of the average random predictor or the average surrogate sensitivity with statistical significance; # indicates that the ratio of individuals that beat the random or surrogate predictors is statistically significant.

Averaging across all 36 patients, a sensitivity of 0.21 ± 0.27 and an FPR/h of 0.21 ± 0.12 were obtained. The sensitivity of the random predictors averaged at 0.58 ± 0.16. These were outperformed, on average, by 3.91 ± 5.96% of solutions. Overall surrogate predictor sensitivity was 0.18 ± 0.09 and 38.43 ± 22.83% were validated. The average random predictor (whose sensitivity values are presented in table B.5) was beaten for 28 patients (78%), while the average surrogate predictor was outperformed for 15 (42%). Although only 7 patients (19%) had a significant

ratio of validated solutions for the random predictor, 32 out of 36 patients (89%) had a statistically significant above chance ratio for the surrogate predictor.

5.2.2 NSGA-II

Table 5.4: Seizure prediction performance and statistical validation results obtained for the models optimized with NSGA-II.

Patient ID	Sensitivity	FPR/h	Random sensitivity	Solutions beat random (%)	Surrogate sensitivity	Solutions beat surrogate (%)
402	0.03 ± 0.12	0.14 ± 0.06	0.50 ± 0.02*	0.00	0.17 ± 0.07	5.61
8902	0.16 ± 0.23	0.21 ± 0.11	0.50 ± 0.00*	0.00	0.22 ± 0.08	30.32#
11002	0.02 ± 0.13	0.39 ± 0.19	0.94 ± 0.24*	0.00	0.25 ± 0.12	1.83
16202	0.07 ± 0.11	0.12 ± 0.07	0.53 ± 0.14*	0.21	0.10 ± 0.05	28.48#
23902	0.12 ± 0.23	0.22 ± 0.10	0.50 ± 0.06*	1.72	0.24 ± 0.12	18.39#
30802	0.27 ± 0.26	0.21 ± 0.08	0.60 ± 0.00*	1.51	0.20 ± 0.05*	52.85#
32702	0.24 ± 0.31	0.23 ± 0.15	0.50 ± 0.00*	7.07#	0.17 ± 0.09*	40.98#
46702	0.26 ± 0.26	0.31 ± 0.16	0.50 ± 0.00*	1.47	0.23 ± 0.09*	49.56#
50802	0.18 ± 0.24	0.26 ± 0.07	0.50 ± 0.00*	0.00	0.23 ± 0.08	35.44#
53402	0.37 ± 0.30	0.26 ± 0.07	0.50 ± 0.00*	8.78#	0.18 ± 0.07*	64.86#
55202	0.28 ± 0.15	0.24 ± 0.06	0.60 ± 0.00*	0.00	0.20 ± 0.04*	45.30#
56402	0.04 ± 0.18	0.22 ± 0.14	1.00 ± 0.00*	0.00	0.17 ± 0.11	3.57
58602	0.08 ± 0.15	0.21 ± 0.07	0.67 ± 0.00*	0.00	0.19 ± 0.08	6.68#
59102	0.12 ± 0.22	0.19 ± 0.06	0.50 ± 0.00*	0.21	0.18 ± 0.07	24.26#
60002	0.21 ± 0.18	0.10 ± 0.02	0.67 ± 0.00*	0.00	0.16 ± 0.06*	57.62#
64702	0.05 ± 0.15	0.12 ± 0.08	0.56 ± 0.21*	0.00	0.12 ± 0.08	10.13#
75202	0.32 ± 0.22	0.18 ± 0.07	0.50 ± 0.00*	12.28#	0.19 ± 0.08*	58.95#
80702	0.08 ± 0.15	0.16 ± 0.07	0.67 ± 0.00*	0.00	0.14 ± 0.06	19.00#
85202	0.59 ± 0.36	0.30 ± 0.10	0.50 ± 0.00*	37.72#	0.20 ± 0.08*	80.25#
93402	0.07 ± 0.21	0.25 ± 0.12	0.50 ± 0.03*	2.36	0.23 ± 0.09	11.45#
93902	0.34 ± 0.23	0.20 ± 0.13	0.58 ± 0.23*	18.37#	0.16 ± 0.11*	61.68#
94402	0.03 ± 0.09	0.31 ± 0.09	0.50 ± 0.00*	0.00	0.20 ± 0.06	0.96
95202	0.04 ± 0.10	0.14 ± 0.04	0.50 ± 0.00*	0.00	0.13 ± 0.05	8.62#
96002	0.15 ± 0.14	0.21 ± 0.06	0.50 ± 0.00*	0.00	0.15 ± 0.05	40.82#
98102	0.38 ± 0.34	0.14 ± 0.12	0.58 ± 0.19*	11.11#	0.11 ± 0.08*	61.30#
98202	0.02 ± 0.06	0.28 ± 0.12	0.50 ± 0.02*	0.00	0.20 ± 0.06	1.11
101702	0.46 ± 0.28	0.28 ± 0.15	0.50 ± 0.00*	11.66#	0.23 ± 0.10*	71.90#
102202	0.08 ± 0.14	0.16 ± 0.07	0.51 ± 0.06*	0.00	0.12 ± 0.06	21.36#
104602	0.28 ± 0.32	0.22 ± 0.11	0.50 ± 0.04*	8.71#	0.20 ± 0.07*	47.33#
109502	0.16 ± 0.36	0.13 ± 0.08	1.00 ± 0.07*	0.00	0.10 ± 0.08*	15.90#
110602	0.25 ± 0.34	0.18 ± 0.08	0.49 ± 0.08*	13.66#	0.16 ± 0.07*	39.70#
112802	0.05 ± 0.12	0.19 ± 0.06	0.67 ± 0.00*	0.00	0.12 ± 0.06	12.16#
113902	0.28 ± 0.12	0.11 ± 0.16	0.34 ± 0.33*	40.40#	0.10 ± 0.10*	71.88#
114702	0.18 ± 0.18	0.30 ± 0.08	0.60 ± 0.00*	0.25	0.24 ± 0.07	19.47#
114902	0.47 ± 0.16	0.15 ± 0.05	0.50 ± 0.05*	13.13#	0.12 ± 0.05*	98.12#
123902	0.01 ± 0.08	0.10 ± 0.07	0.56 ± 0.16*	0.00	0.08 ± 0.07	2.97

* denotes that the average sensitivity obtained is higher than that of the average random predictor or the average surrogate sensitivity with statistical significance; # indicates that the ratio of individuals that beat the random or surrogate predictors is statistically significant.

Overall sensitivity was 0.18 ± 0.26, while FPR/h averaged at 0.21 ± 0.12. The sensitivity of the random predictors averaged at 0.57 ± 0.17 and, on average, these were outperformed by 5.29 ± 9.67% of solutions. Mean surrogate predictor sensitivity was 0.18 ± 0.09, with 33.91 ± 25.79% of solutions validated. The average

random predictor (see table B.5) was outperformed for every patient (100%) and the average surrogate predictor was beaten for 16 (44%) of them. 11 patients (30%) had a significant ratio of validated solutions for the random predictor, while 30 (83%) had a statistically significant ratio for the surrogate predictor.

5.2.3 SMS-EMOA (modified)

Table 5.5: Seizure prediction performance and statistical validation results obtained for the models optimized with the modified SMS-EMOA.

Patient ID	Sensitivity	FPR/h	Random sensitivity	Solutions beat random (%)	Surrogate sensitivity	Solutions beat surrogate (%)
402	0.04 ± 0.13	0.14 ± 0.04	0.50 ± 0.00*	0.00	0.16 ± 0.06	7.47#
8902	0.37 ± 0.26	0.17 ± 0.12	0.50 ± 0.00*	4.19	0.22 ± 0.06*	70.06#
11002	0.01 ± 0.10	0.38 ± 0.21	0.88 ± 0.32*	0.00	0.22 ± 0.14	0.95
16202	0.04 ± 0.12	0.16 ± 0.12	0.56 ± 0.11*	0.32	0.10 ± 0.09	9.29#
23902	0.03 ± 0.13	0.22 ± 0.09	0.50 ± 0.02*	0.00	0.18 ± 0.10	6.62#
30802	0.32 ± 0.22	0.21 ± 0.08	0.60 ± 0.00*	0.86	0.23 ± 0.05*	55.68#
32702	0.20 ± 0.25	0.38 ± 0.22	0.50 ± 0.00*	0.65	0.21 ± 0.12	33.40#
46702	0.24 ± 0.25	0.33 ± 0.16	0.50 ± 0.00*	0.00	0.24 ± 0.09	45.41#
50802	0.14 ± 0.22	0.25 ± 0.08	0.50 ± 0.00*	0.00	0.21 ± 0.10	28.09#
53402	0.19 ± 0.25	0.25 ± 0.08	0.50 ± 0.00*	0.61	0.18 ± 0.08	36.50#
55202	0.29 ± 0.11	0.23 ± 0.04	0.60 ± 0.00*	0.00	0.21 ± 0.05*	49.84#
56402	0.06 ± 0.24	0.27 ± 0.10	1.00 ± 0.00*	0.00	0.19 ± 0.09	6.24
58602	0.03 ± 0.09	0.30 ± 0.11	0.64 ± 0.13*	0.00	0.17 ± 0.07	7.91#
59102	0.07 ± 0.17	0.18 ± 0.04	0.50 ± 0.00*	0.00	0.17 ± 0.06	13.32#
60002	0.27 ± 0.24	0.09 ± 0.01	0.67 ± 0.00*	0.00	0.14 ± 0.05*	59.40#
64702	0.25 ± 0.25	0.14 ± 0.06	0.50 ± 0.05*	0.20	0.14 ± 0.10*	49.38#
75202	0.36 ± 0.22	0.15 ± 0.05	0.50 ± 0.00*	14.65#	0.17 ± 0.06*	73.11#
80702	0.07 ± 0.14	0.14 ± 0.06	0.66 ± 0.03*	0.22	0.14 ± 0.06	18.74#
85202	0.42 ± 0.22	0.25 ± 0.10	0.50 ± 0.00*	3.10	0.22 ± 0.07*	78.57#
93402	0.17 ± 0.32	0.24 ± 0.05	0.50 ± 0.00*	9.83#	0.24 ± 0.07	22.61#
93902	0.19 ± 0.18	0.18 ± 0.09	0.66 ± 0.04*	0.59	0.16 ± 0.08*	47.20#
94402	0.14 ± 0.18	0.34 ± 0.14	0.50 ± 0.00*	0.00	0.21 ± 0.06	28.50#
95202	0.07 ± 0.11	0.15 ± 0.06	0.50 ± 0.00*	0.00	0.12 ± 0.06	20.88#
96002	0.11 ± 0.13	0.22 ± 0.07	0.50 ± 0.00*	0.00	0.15 ± 0.06	35.62#
98102	0.45 ± 0.20	0.13 ± 0.10	0.56 ± 0.17*	3.15	0.12 ± 0.08*	86.94#
98202	0.05 ± 0.10	0.32 ± 0.11	0.50 ± 0.01*	0.00	0.23 ± 0.06	1.28
101702	0.73 ± 0.28	0.22 ± 0.12	0.49 ± 0.06*	49.03#	0.19 ± 0.09*	94.88#
102202	0.20 ± 0.18	0.19 ± 0.06	0.50 ± 0.00*	0.00	0.12 ± 0.04*	59.45#
104602	0.35 ± 0.32	0.28 ± 0.14	0.50 ± 0.00*	9.74#	0.24 ± 0.09*	58.74#
109502	0.28 ± 0.45	0.12 ± 0.10	0.96 ± 0.18*	0.00	0.08 ± 0.07*	28.02#
110602	0.05 ± 0.18	0.25 ± 0.10	0.50 ± 0.02*	1.72	0.20 ± 0.08	8.32#
112802	0.00 ± 0.04	0.18 ± 0.05	0.67 ± 0.00*	0.00	0.10 ± 0.04	1.68
113902	0.28 ± 0.13	0.12 ± 0.10	0.52 ± 0.27*	19.15#	0.12 ± 0.10*	67.93#
114702	0.21 ± 0.22	0.36 ± 0.10	0.60 ± 0.00*	0.00	0.24 ± 0.06	36.12#
114902	0.32 ± 0.12	0.13 ± 0.04	0.51 ± 0.05*	1.61	0.11 ± 0.04*	98.24#
123902	0.02 ± 0.09	0.08 ± 0.05	0.56 ± 0.19*	0.00	0.06 ± 0.04	3.38

* denotes that the average sensitivity obtained is higher than that of the average random predictor or the average surrogate sensitivity with statistical significance; # indicates that the ratio of individuals that beat the random or surrogate predictors is statistically significant.

Sensitivity averaged at 0.18 ± 0.26 and FPR/h at 0.22 ± 0.12. Mean sensitivity for the random predictors was 0.58 ± 0.16, which were beaten, on average, by 3.32

± 8.86% of solutions. The surrogate predictor sensitivity averaged at 0.17 ± 0.09 and 37.49 ± 28.25% of solutions were validated. The average random predictor (see table B.5) was beaten for every patient (100%) while 15 (42%) outperformed the average surrogate predictor. As for the ratios of solutions above chance, only 5 (14%) were considered statistically significant for the random predictor, whereas 31 (86%) were validated for the surrogate predictor.

5.2.4 Comparative analysis

Concerning the gold standard metrics for seizure prediction, statistically significant differences between the three variants were found, except when comparing NSGA-II with the modified SMS-EMOA in terms of sensitivity and with MOGA in terms of FPR/h (see table B.4). Nevertheless, the evaluated models for all three of them generally display a low FPR/h and low sensitivity.

Focusing on FPR/h, it is worth noting the number of patients where, on average, FPR/h ≤ 0.15, which is considered reasonable for patients in pre-surgical monitoring [113]. Individually, this was observed for 8 patients using models optimized with MOGA, 10 for NSGA-II and 11 in the case of the modified SMS-EMOA. In total, the criteria was met for at least one of the MOEAs in 12 out of 36 patients.

With regard to sensitivity, the low scores may be attributed to the reduced number of seizures available for testing, which is also reflected in the high standard deviations. For instance, with 2 testing seizures, the prediction models can only achieve 3 possible sensitivity scores: 0.0, 0.5 or 1.0. It is worth mentioning that significant correlations were found between the number of testing seizures for each patient and the standard deviations observed for the sensitivity metric: Pearson correlation coefficients of $\rho = -0.53$, $\rho = -0.44$ and $\rho = -0.40$ were obtained for each algorithm variant, respectively.

Additionally, due to computational cost limitations, as mentioned in section 4.1, the training set of each patient-specific model comprises only 12 hours of recording (4 hours preceding each of the first 3 seizures), which may prove insufficient. Some connections were found between fitness scores (in training) and performance (in testing), namely, for the objective concerning time under false alarm: $\rho = -0.12$, $\rho = -0.10$ and $\rho = -0.11$, when compared to sensitivity; $\rho = 0.10$, $\rho = 0.17$ and $\rho = 0.15$, when compared to FPR/h.

Finally, the duration of the Electroencephalogram (EEG) recording displayed a considerable influence in the prediction performance. This is supported by the negative correlations found between the duration of the available recordings for

each patient and the sensitivity ($\rho = -0.12$, $\rho = -0.09$ and $\rho = -0.08$) and FPR/h ($\rho = -0.42$, $\rho = -0.36$ and $\rho = -0.40$) metrics. However, this is to be expected given that correctly identifying the pre-ictal period without raising false alarms is a more challenging task for a higher amount of data.

As for statistical validation, some differences were observed when comparing each model's performance to that of the random and surrogate predictors. For instance, the average random predictor was outperformed for a higher number of patients than the average surrogate one. This is, however, because the great majority of the sensitivity scores for the former (included in table B.5) were close to (or exactly) zero which, in turn, is a result of using the number of testing seizures multiplied by the number of prediction models for the k parameter in (2.8).

Moreover, in terms of the ratios of above chance models, the surrogate predictor achieved a significantly higher percentage of validated patients than the random predictor. The reason for this is that the sensitivity of each random predictor is frequently higher than that of each model, as its value depends only on the given Seizure Occurrence Period (SOP), FPR/h and number of testing seizures (given that the d and α parameters were fixed at 1 and 0.05, respectively). Hence, despite the increased computational complexity of the surrogate predictor, it is more flexible and adaptable to the used data (e.g. in terms of the number of seizures and recording duration), which may provide a more robust validation of the results [8].

Table 5.6: Comparison between the results obtained for each MOEA and those of other studies concerning TLE patients from the EPILEPSIAE database.

Study	Number of TLE patients	Sensitivity	FPR/h	Validated with random (%)	Validated with surrogate (%)
Direito et al. (2017)	130	0.38	0.23	N.A.	N.A.
Alvarado-Rojas et al. (2014)	34	0.66	0.33	8.82	N.A.
Teixeira et al. (2014)	190	0.75	0.32	N.A.	N.A.
This study (MOGA)	36	0.21	0.21	19.44	88.89
This study (NSGA-II)	36	0.18	0.21	30.56	83.33
This study (modified SMS-EMOA)	36	0.18	0.22	13.89	86.11

Lastly, the results achieved can be compared to previous studies which were also performed on the EPILEPSIAE database [7, 30, 104]. Table 5.6 summarizes the performance obtained in each study considering only patients diagnosed with Temporal Lobe Epilepsy (TLE). Both Direito et al. [30] and Teixeira et al. [104] use Support Vector Machines (SVMs) (although the latter also employs neural network classifiers), while Alvarado-Rojas et al. [7] chose a simpler threshold classifier. Direito et al. [30] and Teixeira et al. [104] included a Seizure Prediction Horizon (SPH), albeit considerably short (10 seconds, in both cases). None of the selected studies have validated their results against a surrogate predictor.

The results obtained by Direito et al. [30], despite achieving a superior sensitivity, are similar to the current study in terms of FPR/h. Regarding validation, the authors only presented the percentage of validated patients for the entire set (11%, out of 218 patients) and not for the TLE group. Alvarado-Rojas et al. [7] outperform the proposed methodology in terms of sensitivity, but report a higher FPR/h and a lower percentage of patient-specific models above chance level, regardless of the MOEA used to optimize the models developed in this study. Teixeira et al. [104] report a significantly higher sensitivity, despite a higher FPR/h. Since only statistically significant differences between both metrics are addressed, no comparisons concerning validation can be made.

Generally speaking, the methodology developed in the present study has achieved a lower FPR/h and a higher percentage of patient-specific models above chance level (concerning the random predictor). Nevertheless, it is clearly outperformed regarding sensitivity. In addition to the aforementioned reasons, it is worth noting that several parameters such as the sliding window step and the number of high-level features were chosen based on runtime and were not fine-tuned with regard to testing performance. As a matter of fact, most limiting factors in terms of sensitivity stem from training time restraints. Moreover, a notably longer SPH (10 minutes) was adopted, which makes the prediction task more challenging.

A possible improvement may be found, precisely, in the model optimization/ training phase. For instance, seizure prediction performance could be enhanced by executing an evolutionary search for every new seizure, using data from the previous N seizures. In other words, the feature set and model parameters could be re-selected periodically with the availability of new seizure data. Despite the increase in computational complexity and the need for more testing seizures to properly evaluate this approach, real-time applicability may be achievable, considering that MOEA execution time for a maximum of 50 generations was relatively fast (about 2 hours, as detailed in section 4.4.1). Hence, despite the lower sensitivity obtained, the proposed methodology takes into account the need for algorithmic solutions applicable in real-time, with minimal computational requirements and high power efficiency [62].

5.2.5 Patient stratification

With the purpose of assessing whether the overall percentage of prediction models above chance level could be increased when restricting analysis to a selected group of patients, various stratification criteria were set. Namely:

- only patients whose seizures were classified as either Focal Onset Aware (FOA) or Focal Onset Impaired Awareness (FOIA);
- only patients who displayed rhythmic activity patterns;
- only patients who were awake.

Assuming that seizures with an unknown classification ("UC") and unclear patterns ("?") from table 4.2 comply with the first two criteria (respectively), the overall values were calculated for each algorithm variant considering each patient group, as presented in tables 5.7, 5.8 and 5.9.

Table 5.7: Patient stratification results considering models optimized by MOGA.

Stratification	Sensitivity	FPR/h	Random sensitivity	Solutions beat random (%)	Surrogate sensitivity	Solutions beat surrogate (%)
FOA/FOIA	0.22 ± 0.28	0.21 ± 0.11	0.58 ± 0.15	4.30 ± 6.28	0.18 ± 0.08	41.46 ± 21.09
Rhythmic patterns	0.18 ± 0.27	0.23 ± 0.12	0.59 ± 0.18	3.03 ± 4.55	0.19 ± 0.10	32.62 ± 21.27
Awake patient	0.20 ± 0.28	0.21 ± 0.11	0.60 ± 0.18	2.16 ± 3.97	0.18 ± 0.09	32.29 ± 19.13

Table 5.8: Patient stratification results considering models optimized by NSGA-II.

Stratification	Sensitivity	FPR/h	Random sensitivity	Solutions beat random (%)	Surrogate sensitivity	Solutions beat surrogate (%)
FOA/FOIA	0.20 ± 0.26	0.22 ± 0.12	0.58 ± 0.18	6.44 ± 11.52	0.18 ± 0.09	36.97 ± 23.87
Rhythmic patterns	0.16 ± 0.25	0.22 ± 0.12	0.57 ± 0.19	4.77 ± 9.21	0.18 ± 0.09	28.79 ± 22.78
Awake patient	0.15 ± 0.24	0.20 ± 0.10	0.59 ± 0.17	2.14 ± 4.26	0.17 ± 0.08	25.40 ± 16.30

Table 5.9: Patient stratification results considering models optimized by the modified SMS-EMOA.

Stratification	Sensitivity	FPR/h	Random sensitivity	Solutions beat random (%)	Surrogate sensitivity	Solutions beat surrogate (%)
FOA/FOIA	0.19 ± 0.27	0.22 ± 0.13	0.59 ± 0.18	4.47 ± 11.12	0.18 ± 0.09	40.28 ± 27.25
Rhythmic patterns	0.14 ± 0.24	0.22 ± 0.12	0.58 ± 0.18	2.48 ± 5.26	0.18 ± 0.09	29.23 ± 22.74
Awake patient	0.12 ± 0.24	0.21 ± 0.12	0.60 ± 0.18	0.64 ± 1.24	0.16 ± 0.09	25.81 ± 24.49

Each group is comprised of 21, 23 and 16 patients, respectively, as shown in table B.6. Tables 5.10 and 5.11 report the percentage of patients whose models outperform the average random/surrogate predictors along with the percentage of patients with a significant ratio of models above chance level.

Overall, the only group that achieves an improvement, albeit minor, in mean sensitivity and/or FPR/h is the one which concerns seizure classification (only FOA or FOIA). No significant changes were found in the average sensitivity of the random or surrogate predictors. Conversely, the percentages of validated patients displayed a substantial increase.

Table 5.10: Percentage of patients validated considering the average random predictor and whose ratio of models above chance level is statistically significant.

Stratification	MOGA		NSGA-II		SMS-EMOA (modified)	
	Beat average random (%)	Significant above chance ratio (%)	Beat average random (%)	Significant above chance ratio (%)	Beat average random (%)	Significant above chance ratio (%)
FOA/FOIA	80.95	19.05	100.00	33.33	100.00	14.28
Rhythmic patterns	73.91	13.04	100.00	26.09	100.00	17.39
Awake patient	75.00	18.75	100.00	18.75	100.00	0.00

Table 5.11: Percentage of patients validated considering the average surrogate predictor and whose ratio of models above chance level is statistically significant.

Stratification	MOGA		NSGA-II		SMS-EMOA (modified)	
	Beat average surrogate (%)	Significant above chance ratio (%)	Beat average surrogate (%)	Significant above chance ratio (%)	Beat average surrogate (%)	Significant above chance ratio (%)
FOA/FOIA	52.38	100.00	52.38	90.48	47.62	90.48
Rhythmic patterns	34.78	82.60	39.13	73.91	30.43	78.26
Awake patient	31.25	87.50	31.25	81.25	25.00	81.25

Concerning the random predictor, only the seizure classification criterion managed to achieve a higher percentage of patients where the average random predictor was outperformed for models optimized with MOGA (every patient had already been validated for the other two algorithms, so no further conclusions could be drawn). Improvements in terms of the number of significant above chance level ratios were found for the first group in models optimized with NSGA-II and the modified SMS-EMOA, as well as in the latter for the second group (only seizures with rhythmic patterns). As for the surrogate predictor, both validation percentages were only improved for the first patient group.

The previous results suggest that, for patients who suffer from FOA and/or FOIA seizures, the current methodology may prove particularly viable. This is further supported by the positive correlations found between this patient group and the sensitivity metric: $\rho = 0.19$, $\rho = 0.19$ and $\rho = 0.17$, considering each MOEA variant, respectively.

5.3 Phenotype analysis

In order to demonstrate the possible benefits of developing interpretable prediction models for clinical knowledge extrapolation, the resulting phenotype was examined at the group level and, subsequently, for a selected patient. Table 5.12 presents the average SOP found for each patient, while figure 5.2 illustrates the relative prevalence of each one.

Table 5.12: Average SOP duration found for each patient in each algorithm variant.

	SOP duration (minutes)		
Patient ID	**MOGA**	**NSGA-II**	**SMS-EMOA (modified)**
402	41.32 ± 8.02	44.71 ± 6.75	40.78 ± 6.63
8902	44.06 ± 6.78	43.03 ± 6.78	45.45 ± 7.46
11002	43.47 ± 7.54	42.85 ± 7.55	37.40 ± 10.37
16202	40.28 ± 7.10	36.25 ± 6.42	39.54 ± 4.76
23902	44.39 ± 7.62	42.32 ± 8.42	43.70 ± 7.27
30802	44.53 ± 6.04	43.04 ± 6.94	43.04 ± 6.06
32702	42.02 ± 9.22	40.98 ± 7.76	39.92 ± 5.79
46702	43.08 ± 7.16	39.10 ± 7.14	42.24 ± 5.35
50802	48.08 ± 4.40	44.40 ± 6.85	44.83 ± 6.75
53402	47.32 ± 6.04	38.92 ± 6.39	43.33 ± 8.51
55202	47.06 ± 5.57	42.93 ± 4.76	44.52 ± 5.50
56402	42.03 ± 8.21	40.99 ± 9.86	38.45 ± 9.96
58602	42.20 ± 6.87	40.12 ± 5.85	42.56 ± 6.88
59102	43.71 ± 5.96	41.55 ± 7.15	40.33 ± 6.61
60002	43.81 ± 7.29	44.05 ± 6.28	41.21 ± 5.98
64702	43.29 ± 9.13	47.97 ± 6.64	47.05 ± 6.23
75202	43.94 ± 8.93	44.23 ± 8.18	44.34 ± 7.57
80702	42.36 ± 7.75	40.36 ± 7.06	42.17 ± 4.07
85202	40.92 ± 5.94	43.35 ± 6.33	43.52 ± 7.18
93402	43.31 ± 6.87	44.68 ± 7.92	44.48 ± 6.80
93902	40.60 ± 11.03	40.37 ± 5.78	36.55 ± 9.42
94402	42.85 ± 6.33	41.66 ± 7.83	40.46 ± 6.69
95202	42.47 ± 6.62	42.62 ± 7.59	44.19 ± 10.39
96002	43.72 ± 6.35	43.56 ± 6.55	38.24 ± 5.06
98102	42.80 ± 9.67	40.17 ± 9.24	32.61 ± 11.60
98202	46.02 ± 6.31	40.50 ± 6.44	41.49 ± 7.20
101702	43.85 ± 6.67	46.01 ± 7.63	38.67 ± 12.40
102202	41.25 ± 8.97	41.36 ± 6.55	38.78 ± 7.53
104602	45.14 ± 6.34	41.74 ± 7.09	41.53 ± 7.91
109502	45.42 ± 6.63	46.80 ± 6.09	46.18 ± 5.42
110602	44.35 ± 6.85	43.73 ± 7.82	37.99 ± 5.99
112802	42.76 ± 6.65	40.39 ± 9.27	43.41 ± 7.49
113902	43.86 ± 7.95	39.41 ± 9.64	40.38 ± 7.89
114702	45.04 ± 6.16	41.31 ± 7.70	41.10 ± 8.52
114902	42.40 ± 8.61	38.38 ± 5.98	40.62 ± 7.59
123902	44.68 ± 7.44	43.28 ± 8.59	45.30 ± 6.49

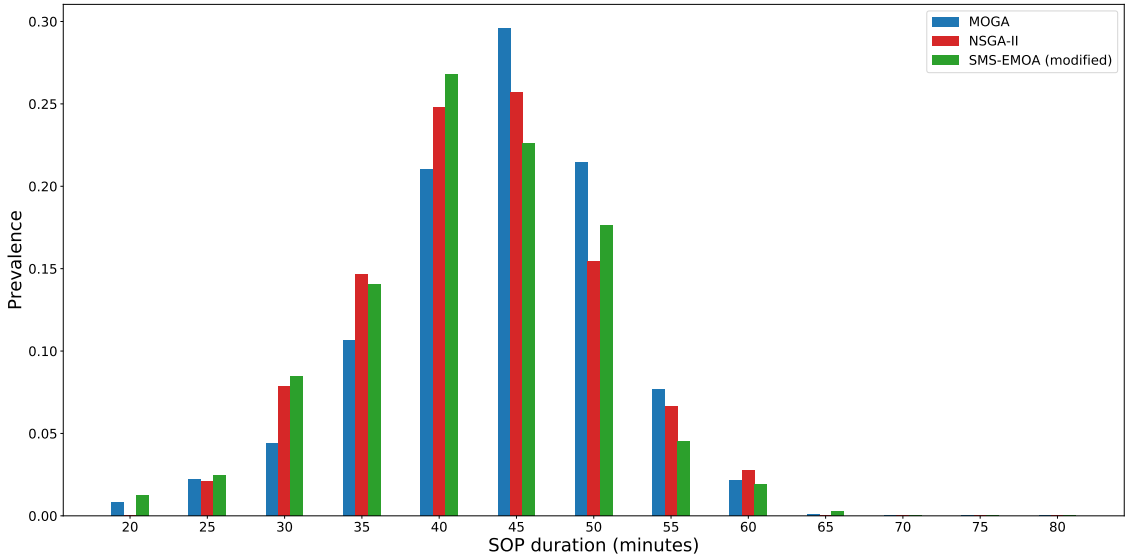


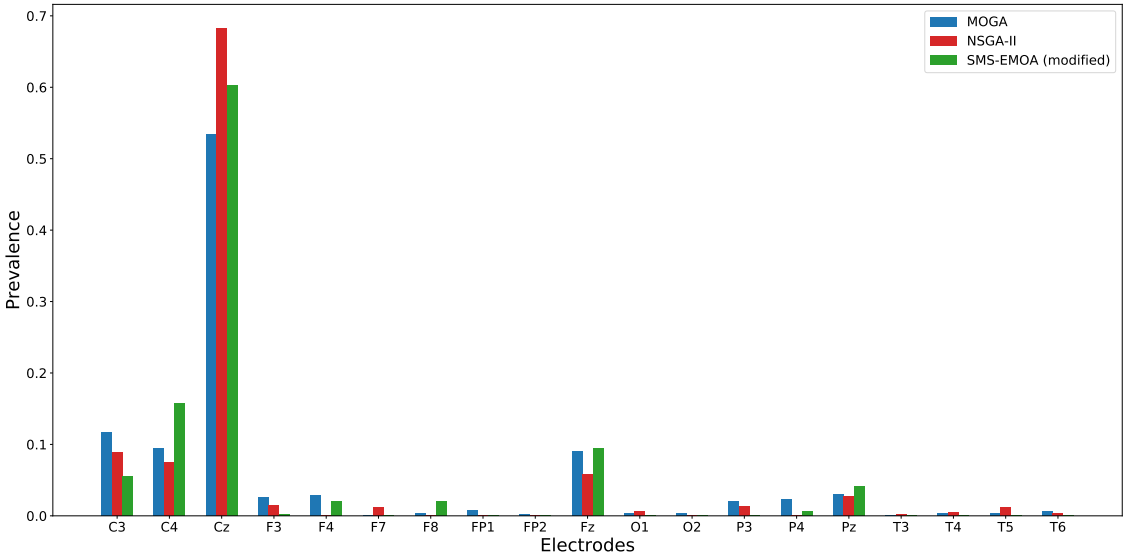
Figure 5.2: Histogram of the prevalence (valued between 0 and 1) of each SOP duration within models optimized by different algorithms.

Averaging across all 36 patients, the mean SOP duration in minutes for each MOEA was 43.74 ± 7.52, 42.54 ± 7.61 and 41.85 ± 7.86, respectively. Despite all algorithms presenting statistically significant differences between one another, there is a clear preference for SOPs between 35 and 50 minutes across all three variants.

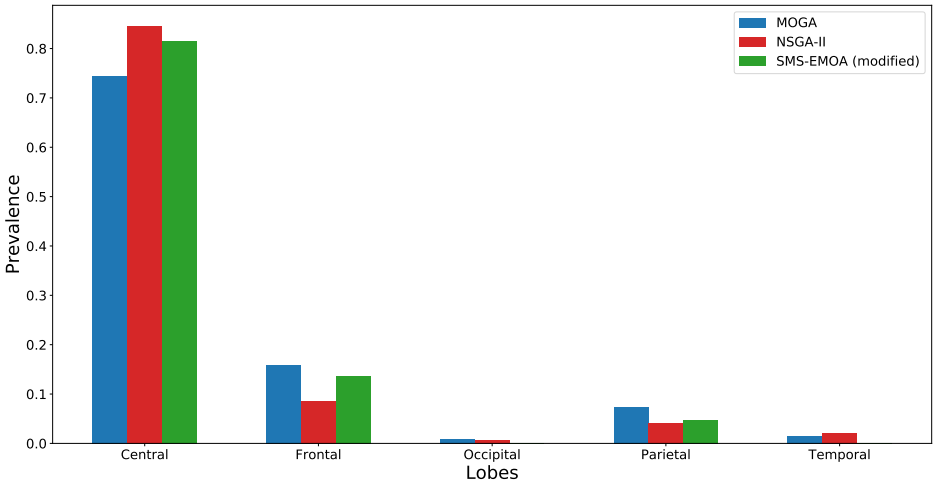
Moreover, it is worth noting that neither MOEA displays any preference for either extreme case (i.e. SOPs of 20 or 80 minutes). Another remarkable aspect lies in the fact that virtually no models used a SOP longer than 65 minutes, in spite of the fact that the prediction task becomes less challenging for longer SOPs (at the cost of patient stress levels, as explained in section 2.3.1).

Next, patient 114902 was selected since its patient-specific models achieved the highest percentage of models above chance level for all three MOEA variants (concerning the surrogate predictor) at a reasonable sensitivity and FPR/h ≤ 0.15. The following figures represent the presence of several genes in the resulting phenotypes, concerning the electrodes/lobes where information was extracted from, the types of features used as well as the delays and window lengths within the feature set.

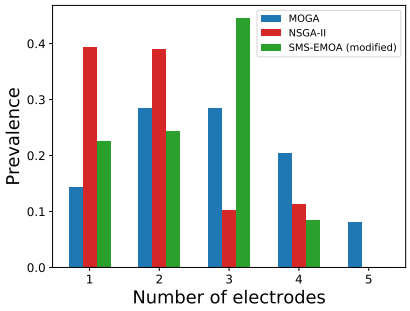
As indicated in figures 5.3a and 5.3b, the most common electrodes were Cz, C3, C4, Fz and Pz for all three variants and, accordingly, the central, frontal and parietal lobes were the most frequent. Notably, over 70% of all models optimized with either MOEA contained one or more electrodes placed in the central lobe. These results may support network theory, which proposes that even focal seizures stem from abnormal activity in a large-scale functional network extending over lobes and hemispheres [62]. Hence, although this patient's epileptic focus is situated in



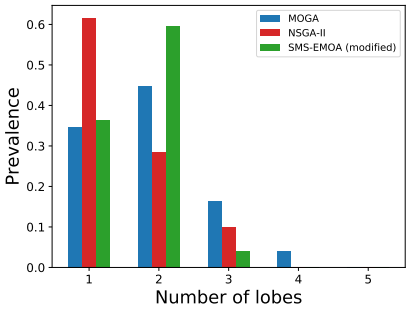
(a) Electrode prevalence



(b) Lobe prevalence



(c) Number of unique electrodes



(d) Number of unique lobes

Figure 5.3: Histogram of the prevalence (valued between 0 and 1) of different electrodes/lobes as well as their unique number within each model.

the temporal lobe, pre-ictal changes or patterns may be encountered among the remaining brain regions.

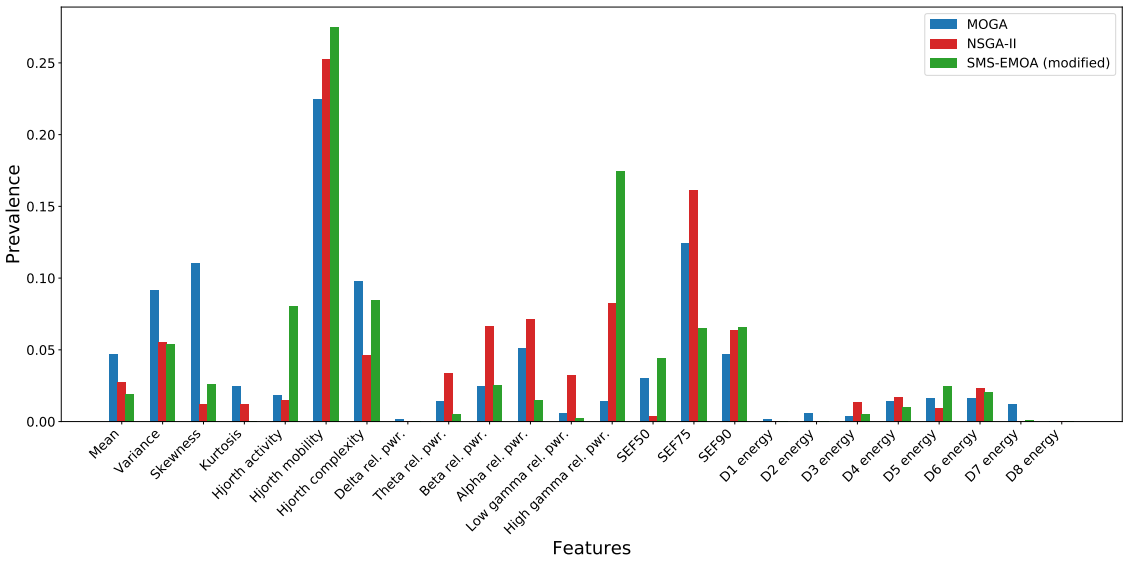
Moreover, a large proportion of the prediction models did not require more than 2 or 3 electrodes, where the majority was placed in either 1 or 2 lobes, as observed in figures 5.3c and 5.3d. A probable cause is the inclusion of the third objective in the evolutionary search, which benefits solutions with low spatial requirements. Hence, in addition to providing insight on brain phenomena, these models may aid in terms of signal acquisition and, in particular, patient comfort (e.g. 60% of the models optimized with NSGA-II only require electrodes placed in a single lobe).

By analyzing figures 5.4a and 5.4b, it can be seen that, with the exception of models optimized by NSGA-II, time domain features were preferred. In particular, the Hjorth mobility measure was present in over 20% of all models optimized with either MOEA. Among features from the frequency domain, the relative spectral power of the high gamma range and the spectral edge frequency with a cut-off at 75% were the most common.

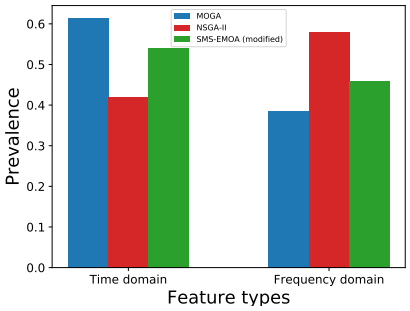
Despite the preference for time domain features, the vast majority of the models contained features from both types, as seen in figure 5.5b. Apart from minor exceptions, every prediction model contained at least three different features within its set, regardless of the MOEA used to optimize it. Concerning the mathematical operators (figure 5.5), no clear trends were identified.

Finally, some patterns can be found in figure 5.6 concerning the delay and window length genes. The duration of each one is commonly around 15 and 10 minutes, respectively. More importantly, however, is their heterogeneity within each feature set: most models present 3 different delays and 2 to 4 different window lengths. In other words, most prediction models do not focus on the feature values within the same time frame, but instead appear to search for seizure-related patterns in different time periods, hinting at the possibility of capturing a sequence of events leading up to a seizure.

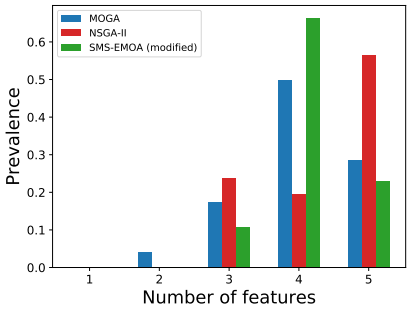
Overall, these findings allow for clinical interpretation of the prediction models by providing insight in terms of the seizure-generation processes in two different ways. First, when assessing the feature sets resulting from the evolutionary search, whose prevalence may be indicative of key properties found within the EEG. Then, when analyzing the electrodes from which those features are extracted, which may shed light on connectivity between brain regions during epileptogenesis.



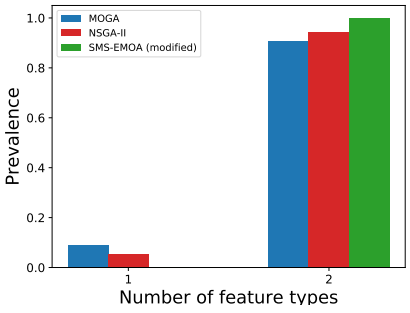
(a) Feature prevalence



(b) Feature type prevalence

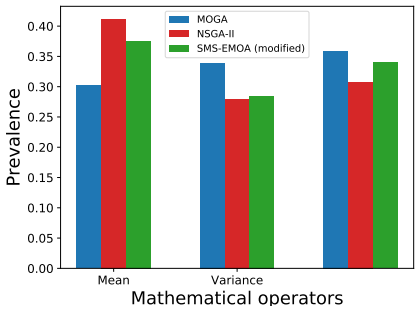


(c) Number of unique features

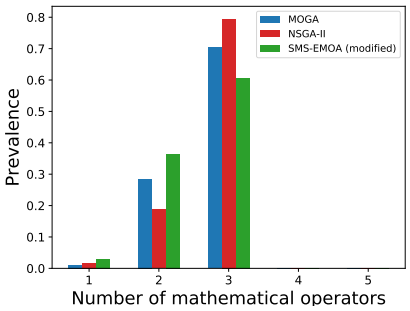


(d) Number of unique feature types

Figure 5.4: Histogram of the prevalence (valued between 0 and 1) of different features as well as their unique number within each model.

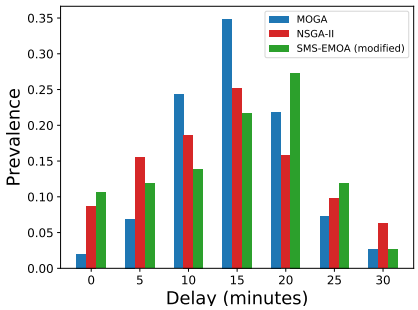


(a) Mathematical operator prevalence

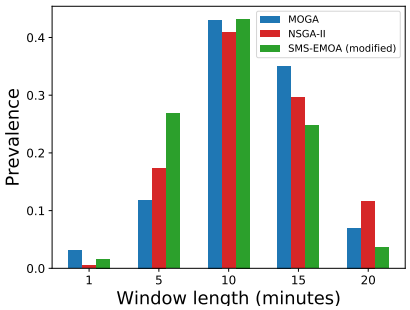


(b) Number of unique operators

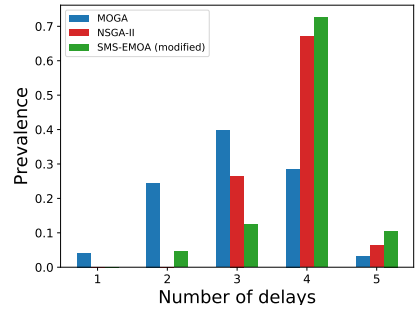
Figure 5.5: Histogram of the prevalence (valued between 0 and 1) of different mathematical operators as well as their unique number within each model.



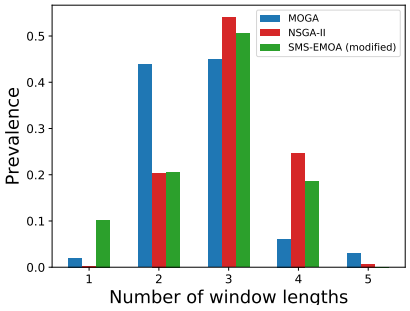
(a) Delay duration prevalence



(b) Window length prevalence



(c) Number of unique delays



(d) Number of unique window lengths

Figure 5.6: Histogram of the prevalence (valued between 0 and 1) of different delays/window lengths as well as their unique number within each model.

5.4 Summary

Training performance considering the solution set selected by the DM was generally satisfactory for each of the three objectives (sample sensitivity, time under false alarm and number of different electrodes/lobes within the feature set) across all three MOEAs. Fitness scores averaged at 0.83 ± 0.14, 0.79 ± 0.11 and 0.74 ± 0.20 for MOGA, 0.97 ± 0.04, 0.97 ± 0.04 and 0.83 ± 0.16 for NSGA-II and, finally, 0.97 ± 0.04, 0.97 ± 0.04 and 0.81 ± 0.16 for the modified SMS-EMOA.

In terms of testing performance, the following sensitivity and FPR/h were achieved: 0.21 ± 0.27 and 0.21 ± 0.12 for MOGA, 0.18 ± 0.26 and 0.21 ± 0.12 for NSGA-II and 0.18 ± 0.26 and 0.22 ± 0.12 for the modified SMS-EMOA. Concerning statistical validation, the random predictor was outperformed for 19%, 30% and 14%, respectively, and the surrogate predictor was beaten by a significant number of models for 89%, 83% and 86% of all 36 patients.

Patient stratification results indicated a reasonable increase in validation percentage when restricting analysis to patients whose seizures were either classified as either FOA or FOIA. The remaining groups (rhythmic activity patterns and patients who were awake prior to the seizure onset) did not achieve any improvement.

By analyzing the resulting phenotypes for each set of models, a clear preference for SOPs between 35 and 50 minutes was observed for all three MOEA variants. As for the selected patient, several trends were found concerning gene prevalence. Firstly, information is extracted mostly from electrodes placed on the central lobe, with most models requiring a low number of different electrodes and lobes. Then, the vast majority of the feature sets included 3 or more different features from both the time and frequency domains. Finally, features tend to analyze different time frames, possibly capturing the sequence of events preceding the seizure onset.

6

Conclusion

The present study concerned the development of machine learning models optimized by three different Multi-Objective Evolutionary Algorithms (MOEAs) for seizure prediction based on Electroencephalogram (EEG) data. It can be said that the established research goals were achieved, namely:

- A considerable percentage of the patient-specific prediction models was statistically validated considering both random and surrogate predictors, with generally low False Positive Rate per Hour (FPR/h) values;
- The several hours of available signal data were used to train, test and evaluate the models, in addition to patient stratification made possible by the extensive metadata present in the EPILEPSIAE database;
- The application of MOEAs for model optimization not only allowed to explore the vast search space of possible configurations but also provided the possibility for clinical knowledge extrapolation from the obtained phenotypes.

Moreover, despite the computational complexity of the training phase, which had to be restricted to 4 hours of data before each seizure onset, the proposed methodology is relatively light and straightforward: basic pre-processing steps, followed by high-level feature extraction and the application of a linear Support Vector Machine (SVM) for classification. Considering the relatively short runtime, it may be possible to build a real-time application adapted to the seizure rate of patients in pre-surgical monitoring. The majority of the developed models also presented a low number of required electrodes, a potential improvement regarding power efficiency and patient comfort.

Nevertheless, when compared to other studies using data from the same database, the present methodology is outperformed in terms of sensitivity, which may indicate the need for models with a higher degree of complexity. Furthermore, considering that data was retrieved from patients in pre-surgical monitoring units, this work may only serve as a proof of concept for the developed prediction system. In order to truly assess its applicability in a real-life scenario, long-term data from everyday

life conditions must be used [42].

Concerning future work, a stronger focus on the pre-processing stage of the current pipeline could potentially improve performance. Namely, refined data cleanup and automated artifact removal procedures may lead to more robust models with a lower chance of producing false alarms. Another possible modification, given the observed runtime for the training phase, could be to make the MOEA periodically search for new configurations (i.e. feature sets and model parameters) with the availability of incoming seizure data.

Furthermore, employing the same methodology in different datasets would allow for comparisons to other seizure prediction studies. In particular, long-term data containing naturally-occurring seizures such as the one curated by Cook et al. [24] would be highly valuable not only to evaluate recent approaches but also for the development and validation of novel prediction algorithms, pushing towards a wider clinical acceptance.

Bibliography

[1] U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K.-H. Ng, and J. S. Suri, "Automated diagnosis of epileptic EEG using entropies," *Biomedical Signal Processing and Control*, vol. 7, no. 4, pp. 401–408, 2012.

[2] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: a review," *Knowledge-Based Systems*, vol. 45, pp. 147–165, 2013.

[3] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 2, pp. 205–211, 2007.

[4] P. Agarwal, H.-C. Wang, and K. Srinivasan, "Epileptic seizure prediction over EEG data using hybrid CNN-SVM model with edge computing services," in *MATEC Web of Conferences*, vol. 210. EDP Sciences, 2018, p. 03016.

[5] G. Aggarwal and T. K. Gandhi, "Prediction of epileptic seizures based on mean phase coherence," *bioRxiv*, p. 212563, 2017.

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

[7] C. Alvarado-Rojas, M. Valderrama, A. Fouad-Ahmed, H. Feldwisch-Drentrup, M. Ihle, C. Teixeira, F. Sales, A. Schulze-Bonhage, C. Adam, A. Dourado *et al.*, "Slow modulations of high-frequency activity (40–140 hz) discriminate preictal changes in human focal epilepsy," *Scientific reports*, vol. 4, p. 4545, 2014.

[8] R. G. Andrzejak, D. Chicharro, C. E. Elger, and F. Mormann, "Seizure prediction: Any better than chance?" *Clinical Neurophysiology*, vol. 120, no. 8, pp. 1465–1478, 2009.

[9] R. G. Andrzejak, F. Mormann, T. Kreuz, C. Rieke, A. Kraskov, C. E. El-

ger, and K. Lehnertz, "Testing the null hypothesis of the nonexistence of a preseizure state," *Physical Review E*, vol. 67, no. 1, p. 010901, 2003.

[10] E. B. Assi, M. Sawan, D. Nguyen, and S. Rihana, "A hybrid mRMR-genetic based selection method for the prediction of epileptic seizures," in *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2015, pp. 1–4.

[11] E. B. Assi, D. K. Nguyen, S. Rihana, and M. Sawan, "Towards accurate prediction of epileptic seizures: A review," *Biomedical Signal Processing and Control*, vol. 34, pp. 144–157, 2017.

[12] T. Back, *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford university press, 1996.

[13] T. Bäck, D. B. Fogel, and Z. Michalewicz, *Evolutionary computation 1: Basic algorithms and operators*. CRC press, 2018.

[14] M. Bandarabadi, J. Rasekhi, C. A. Teixeira, M. R. Karami, and A. Dourado, "On the proper selection of preictal period for seizure prediction," *Epilepsy & Behavior*, vol. 46, pp. 158–166, 2015.

[15] M. Bandarabadi, C. A. Teixeira, B. Direito, and A. Dourado, "Epileptic seizure prediction based on a bivariate spectral power methodology," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 5943–5946.

[16] 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.

[17] T. Bartz-Beielstein, J. Branke, J. Mehnen, and O. Mersmann, "Evolutionary algorithms," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, no. 3, pp. 178–195, 2014.

[18] N. Beume, B. Naujoks, and M. Emmerich, "SMS-EMOA: Multiobjective selection based on dominated hypervolume," *European Journal of Operational Research*, vol. 181, no. 3, pp. 1653–1669, 2007.

[19] F. Biscani and D. Izzo, "esa/pagmo2: pagmo 2.15.0," Apr. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3738182

[20] A. Chamseddine and M. Sawan, "Deep learning based method for output regularization of the seizure prediction classifier," in *2018 IEEE Life Sciences Conference (LSC)*. IEEE, 2018, pp. 118–121.

[21] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio, and F. Fuggetta, "Real-time epileptic seizure prediction using AR models and support vector machines," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 5, pp. 1124–1132, 2010.

[22] C. A. C. Coello, G. B. Lamont, D. A. Van Veldhuizen *et al.*, *Evolutionary algorithms for solving multi-objective problems*. Springer, 2007, vol. 5.

[23] M. X. Cohen, *Analyzing neural time series data: theory and practice*. MIT press, 2014.

[24] M. J. Cook, T. J. O'Brien, S. F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, W. D'Souza, R. Yerra, J. Archer, L. Litewka *et al.*, "Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study," *The Lancet Neurology*, vol. 12, no. 6, pp. 563–571, 2013.

[25] F. H. L. da Silva, W. Blanes, S. N. Kalitzin, J. Parra, P. Suffczynski, and D. N. Velis, "Dynamical diseases of brain systems: different routes to epileptic seizures," *IEEE transactions on biomedical engineering*, vol. 50, no. 5, pp. 540–548, 2003.

[26] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, and B. Litt, "Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients," *IEEE transactions on biomedical engineering*, vol. 50, no. 5, pp. 603–615, 2003.

[27] H. Daoud and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE transactions on biomedical circuits and systems*, vol. 13, no. 5, pp. 804–813, 2019.

[28] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[29] B. Direito, C. Teixeira, B. Ribeiro, M. Castelo-Branco, F. Sales, and A. Dourado, "Modeling epileptic brain states using EEG spectral analysis and topographic mapping," *Journal of neuroscience methods*, vol. 210, no. 2, pp. 220–229, 2012.

[30] B. Direito, C. A. Teixeira, F. Sales, M. Castelo-Branco, and A. Dourado, "A realistic seizure prediction study based on multiclass SVM," *International journal of neural systems*, vol. 27, no. 03, p. 1750006, 2017.

[31] B. Direito, F. Ventura, C. Teixeira, and A. Dourado, "Optimized feature subsets for epileptic seizure prediction studies," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 1636–1639.

[32] A. E. Eiben, J. E. Smith *et al.*, *Introduction to Evolutionary Computing*. Springer, 2003, vol. 53.

[33] A. E. Eiben and J. Smith, "From evolutionary computation to the evolution of things," *Nature*, vol. 521, no. 7553, pp. 476–482, 2015.

[34] Epilepsy Foundation, "Focal to bilateral tonic-clonic seizures (secondarily generalized seizures)," 2019, last accessed: 11 November 2019. Available at https://www.epilepsy.com/learn/types-seizures/focal-bilateral-tonic-clonic-seizures-aka-secondarily-generalized-seizures.

[35] ——, "Temporal Lobe Epilepsy (TLE)," 2019, last accessed: 17 November 2019. Available at https://www.epilepsy.com/learn/types-epilepsy-syndromes/temporal-lobe-epilepsy-aka-tle.

[36] ——, "Video EEG Monitoring with Invasive Electrodes," 2019, last accessed: 25 November 2019. Available at https://www.epilepsy.com/learn/treating-seizures-and-epilepsy/surgery/tests-surgery/video-eeg-monitoring-invasive-electrodes.

[37] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. Engel Jr, L. Forsgren, J. A. French, M. Glynn *et al.*, "ILAE official report: a practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475–482, 2014.

[38] R. S. Fisher, W. V. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. Engel Jr, "Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia*, vol. 46, no. 4, pp. 470–472, 2005.

[39] R. S. Fisher, J. H. Cross, J. A. French, N. Higurashi, E. Hirsch, F. E. Jansen, L. Lagae, S. L. Moshé, J. Peltola, E. Roulet Perez *et al.*, "Operational classification of seizure types by the international league against epilepsy: Position paper of the ILAE commission for classification and terminology," *Epilepsia*, vol. 58, no. 4, pp. 522–530, 2017.

[40] C. M. Fonseca, "Multiobjective genetic algorithms with application to control engineering problems." Ph.D. dissertation, University of Sheffield, 1995.

[41] C. M. Fonseca, P. J. Fleming *et al.*, "Genetic algorithms for multiobjective optimization: Formulation discussion and generalization." in *Icga*, vol. 93, no. July. Citeseer, 1993, pp. 416–423.

[42] D. R. Freestone, P. J. Karoly, and M. J. Cook, "A forward-looking review of seizure prediction," *Current opinion in neurology*, vol. 30, no. 2, pp. 167–173, 2017.

[43] A. B. Geva and D. H. Kerem, "Forecasting generalized epileptic seizures from the EEG signal by wavelet analysis and dynamic unsupervised fuzzy clustering," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 10, pp. 1205–1216, 1998.

[44] P. J. Hancock, "An empirical comparison of selection methods in evolutionary algorithms," in *AISB Workshop on Evolutionary Computing*. Springer, 1994, pp. 80–94.

[45] M. A. F. Harrison, I. Osorio, M. G. Frei, S. Asuri, and Y.-C. Lai, "Correlation dimension and integral do not predict epileptic seizures," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 15, no. 3, p. 033106, 2005.

[46] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[47] T. R. Hoens, R. Polikar, and N. V. Chawla, "Learning from streaming data with concept drift and imbalance: an overview," *Progress in Artificial Intelligence*, vol. 1, no. 1, pp. 89–101, 2012.

[48] J. Horn, N. Nafpliotis, and D. E. Goldberg, "A niched pareto genetic algorithm for multiobjective optimization," in *Proceedings of the first IEEE conference on evolutionary computation. IEEE world congress on computational intelligence*. Ieee, 1994, pp. 82–87.

[49] L. Iasemidis, D.-S. Shiau, P. M. Pardalos, W. Chaovalitwongse, K. Narayanan, A. Prasad, K. Tsakalis, P. R. Carney, and J. C. Sackellares, "Long-term prospective on-line real-time seizure prediction," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 532–544, 2005.

[50] L. D. Iasemidis, "Epileptic Seizure Prediction and Control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 5, pp. 549–558, 2003.

[51] L. D. Iasemidis and J. C. Sackellares, "The evolution with time of the spatial

distribution of the largest lyapunov exponent on the human epileptic cortex," *Measuring chaos in the human brain*, pp. 49–82, 1991.

[52] M. Ihle, H. Feldwisch-Drentrup, C. A. Teixeira, A. Witon, B. Schelter, J. Timmer, and A. Schulze-Bonhage, "EPILEPSIAE - A European epilepsy database," *Comput. Methods Programs Biomed.*, vol. 106, no. 3, pp. 127–138, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.cmpb.2010.08.011

[53] International League Against Epilepsy (ILAE), "Epilepsy syndromes," 2019, last accessed: 11 November 2019. Available at https://www.epilepsydiagnosis.org/syndrome/epilepsy-syndrome-groupoverview.html.

[54] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimization: A short review," in *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*. IEEE, 2008, pp. 2419–2426.

[55] Jack Watson, "Epilepsy. June 2011." 2019, last accessed: 25 November 2019. Available at https://jacklynwatson.com/illustration/epilepsy/.

[56] P. J. Karoly, H. Ung, D. B. Grayden, L. Kuhlmann, K. Leyde, M. J. Cook, and D. R. Freestone, "The circadian profile of epilepsy improves seizure forecasting," *Brain*, vol. 140, no. 8, pp. 2169–2182, 2017.

[57] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal onset seizure prediction using convolutional networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 2109–2118, 2017.

[58] V. Khare, X. Yao, and K. Deb, "Performance scaling of multi-objective evolutionary algorithms," in *International conference on evolutionary multi-criterion optimization*. Springer, 2003, pp. 376–390.

[59] I. Kiral-Kornek, S. Roy, E. Nurse, B. Mashford, P. Karoly, T. Carroll, D. Payne, S. Saha, S. Baldassano, T. O'Brien *et al.*, "Epileptic seizure prediction using big data and deep learning: toward a mobile system," *EBioMedicine*, vol. 27, pp. 103–111, 2018.

[60] J. Klatt, H. Feldwisch-Drentrup, M. Ihle, V. Navarro, M. Neufang, C. Teixeira, C. Adam, M. Valderrama, C. Alvarado-Rojas, A. Witon, M. Le Van 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.

[61] S. Kovac, V. N. Vakharia, C. Scott, and B. Diehl, "Invasive epilepsy surgery evaluation," *Seizure*, vol. 44, pp. 125–136, 2017.

[62] L. Kuhlmann, K. Lehnertz, M. P. Richardson, B. Schelter, and H. P. Zaveri, "Seizure prediction—ready for a new era," *Nature Reviews Neurology*, vol. 14, no. 10, pp. 618–630, 2018.

[63] P. Kwan, A. Arzimanoglou, A. T. Berg, M. J. Brodie, W. Allen Hauser, G. Mathern, S. L. Moshé, E. Perucca, S. Wiebe, and J. French, "Definition of drug resistant epilepsy: consensus proposal by the ad hoc task force of the ILAE commission on therapeutic strategies," *Epilepsia*, vol. 51, no. 6, pp. 1069–1077, 2010.

[64] Y.-C. Lai, M. A. F. Harrison, M. G. Frei, and I. Osorio, "Inability of Lyapunov exponents to predict epileptic seizures," *Physical review letters*, vol. 91, no. 6, p. 068102, 2003.

[65] ——, "Controlled test for predictive power of Lyapunov exponents: their inability to predict epileptic seizures," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 14, no. 3, pp. 630–642, 2004.

[66] M. Le Van Quyen, J. Martinerie, M. Baulac, and F. Varela, "Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings," *Neuroreport*, vol. 10, no. 10, pp. 2149–2155, 1999.

[67] M. Le Van Quyen, J. Martinerie, V. Navarro, P. Boon, M. D'Havé, C. Adam, B. Renault, F. Varela, and M. Baulac, "Anticipation of epileptic seizures from standard EEG recordings," *The Lancet*, vol. 357, no. 9251, pp. 183–188, 2001.

[68] M. Le Van Quyen, J. Soss, V. Navarro, R. Robertson, M. Chavez, M. Baulac, and J. Martinerie, "Preictal state identification by synchronization changes in long-term intracranial EEG recordings," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 559–568, 2005.

[69] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[70] X. Li and G. Ouyang, "Nonlinear similarity analysis for epileptic seizures prediction," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 64, no. 8, pp. 1666–1678, 2006.

[71] R. C. Madeo, S. M. Peres, and C. A. Lima, "Overview on support vector

machines applied to temporal modeling," *Anais do IX Encontro Nacional de Inteligência Artificial*, pp. 1–6, 2012.

[72] P. E. McSharry, L. A. Smith, and L. Tarassenko, "Prediction of epileptic seizures: are nonlinear methods relevant?" *Nature medicine*, vol. 9, no. 3, pp. 241–242, 2003.

[73] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky, "Classification of patterns of EEG synchronization for seizure prediction," *Clinical neurophysiology*, vol. 120, no. 11, pp. 1927–1940, 2009.

[74] N. Moghim and D. W. Corne, "Predicting epileptic seizures in advance," *PloS one*, vol. 9, no. 6, 2014.

[75] 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, 2006.

[76] F. Mormann, R. G. Andrzejak, T. Kreuz, C. Rieke, P. David, C. E. Elger, and K. Lehnertz, "Automated detection of a preseizure state based on a decrease in synchronization in intracranial electroencephalogram recordings from epilepsy patients," *Physical Review E*, vol. 67, no. 2, p. 021912, 2003.

[77] F. Mormann, T. Kreuz, R. G. Andrzejak, P. David, K. Lehnertz, and C. E. Elger, "Epileptic seizures are preceded by a decrease in synchronization," *Epilepsy research*, vol. 53, no. 3, pp. 173–185, 2003.

[78] F. Mormann, T. Kreuz, C. Rieke, R. G. Andrzejak, A. Kraskov, P. David, C. E. Elger, and K. Lehnertz, "On the predictability of epileptic seizures," *Clinical neurophysiology*, vol. 116, no. 3, pp. 569–587, 2005.

[79] F. Mormann, K. Lehnertz, P. David, and C. E. Elger, "Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients," *Physica D: Nonlinear Phenomena*, vol. 144, no. 3-4, pp. 358–369, 2000.

[80] K. Nowak, M. Märtens, and D. Izzo, "Empirical performance of the approximation of the least hypervolume contributor," in *International Conference on Parallel Problem Solving From Nature*. Springer, 2014, pp. 662–671.

[81] I. Osorio, H. P. Zaveri, M. G. Frei, and S. Arthurs, *Epilepsy: the intersection of neurosciences, biology, mathematics, engineering, and physics*. CRC press, 2016.

[82] 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.

[83] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[84] A. F. Rabbi, L. Azinfar, and R. Fazel-Rezai, "Seizure prediction using adaptive neuro-fuzzy inference system," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 2100–2103.

[85] S. Ramgopal, S. Thome-Souza, M. Jackson, N. E. Kadish, I. S. Fernández, J. Klehm, W. Bosl, C. Reinsberger, S. Schachter, and T. Loddenkemper, "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy," *Epilepsy & behavior*, vol. 37, pp. 291–307, 2014.

[86] J. Rasekhi, M. R. K. Mollaei, M. Bandarabadi, C. A. Teixeira, and A. Dourado, "Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods," *Journal of neuroscience methods*, vol. 217, no. 1-2, pp. 9–16, 2013.

[87] ——, "Epileptic seizure prediction based on ratio and differential linear univariate features," *Journal of medical signals and sensors*, vol. 5, no. 1, p. 1, 2015.

[88] F. Rothlauf, "Representations for genetic and evolutionary algorithms," in *Representations for Genetic and Evolutionary Algorithms*. Springer, 2002, pp. 9–30.

[89] M. Sazgar and M. G. Young, *Overview of EEG, Electrode Placement, and Montages*. Cham: Springer International Publishing, 2019, pp. 117–125. [Online]. Available: https://doi.org/10.1007/978-3-030-03511-2_5

[90] I. E. Scheffer, S. Berkovic, G. Capovilla, M. B. Connolly, J. French, L. Guilhoto, E. Hirsch, S. Jain, G. W. Mathern, S. L. Moshé *et al.*, "ILAE classification of the epilepsies: position paper of the ILAE commission for classification and terminology," *Epilepsia*, vol. 58, no. 4, pp. 512–521, 2017.

[91] B. Schelter, R. G. Andrzejak, and F. Mormann, "Can your prediction algo-

rithm beat a random predictor," *Seizure prediction in epilepsy: from basic mechanisms to clinical applications. New York: Wiley*, pp. 237–48, 2008.

[92] B. Schelter, M. Winterhalder, T. Maiwald, A. Brandt, A. Schad, A. Schulze-Bonhage, and J. Timmer, "Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 16, no. 1, p. 013108, 2006.

[93] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[94] A. Schulze-Bonhage, F. Sales, K. Wagner, R. Teotonio, A. Carius, A. Schelle, and M. Ihle, "Views of patients with epilepsy on seizure prediction devices," *Epilepsy & behavior*, vol. 18, no. 4, pp. 388–396, 2010.

[95] J. Sheng, S. Liu, H. Qin, B. Li, and X. Zhang, "Drug-resistant epilepsy and surgery," *Current neuropharmacology*, vol. 16, no. 1, pp. 17–28, 2018.

[96] J. Smith, "On replacement strategies in steady state evolutionary algorithms," *Evolutionary Computation*, vol. 15, no. 1, pp. 29–59, 2007.

[97] N. Srinivas and K. Deb, "Muiltiobjective optimization using nondominated sorting in genetic algorithms," *Evolutionary computation*, vol. 2, no. 3, pp. 221–248, 1994.

[98] F. T. Sun and M. J. Morrell, "The RNS system: responsive cortical stimulation for the treatment of refractory partial epilepsy," *Expert review of medical devices*, vol. 11, no. 6, pp. 563–572, 2014.

[99] M. Sun, F. Wang, T. Min, T. Zang, and Y. Wang, "Prediction for high risk clinical symptoms of epilepsy based on deep learning algorithm," *IEEE Access*, vol. 6, pp. 77 596–77 605, 2018.

[100] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[101] C. Teixeira, "Soft-computing techniques applied to artificial tissue temperature estimation," 2008.

[102] C. Teixeira, B. Direito, H. Feldwisch-Drentrup, M. Valderrama, R. Costa, C. Alvarado-Rojas, S. Nikolopoulos, M. Le Van Quyen, J. Timmer, B. Schelter

et al., "EPILAB: A software package for studies on the prediction of epileptic seizures," *Journal of Neuroscience Methods*, vol. 200, no. 2, pp. 257–271, 2011.

[103] C. Teixeira, B. Direito, M. Bandarabadi, and A. Dourado, "Output regularization of SVM seizure predictors: Kalman filter versus the "firing power" method," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 6530–6533.

[104] C. A. Teixeira, B. Direito, M. Bandarabadi, M. Le Van Quyen, M. Valderrama, B. Schelter, A. Schulze-Bonhage, V. Navarro, F. Sales, and A. Dourado, "Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients," *Computer methods and programs in biomedicine*, vol. 114, no. 3, pp. 324–336, 2014.

[105] Towards Data Science, "Introduction to evolutionary algorithms," 2018, last accessed: 4 February 2020. Available at https://towardsdatascience.com/introduction-to-evolutionary-algorithms-a8594b484ac.

[106] B. Tran, B. Xue, and M. Zhang, "Genetic programming for feature construction and selection in classification on high-dimensional data," *Memetic Computing*, vol. 8, no. 1, pp. 3–15, 2016.

[107] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, S. Ippolito, and O. Kavehei, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Networks*, vol. 105, pp. 104–111, 2018.

[108] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Computers in biology and medicine*, vol. 99, pp. 24–37, 2018.

[109] A. Tsymbal, M. Pechenizkiy, P. Cunningham, and S. Puuronen, "Dynamic integration of classifiers for handling concept drift," *Information fusion*, vol. 9, no. 1, pp. 56–68, 2008.

[110] M. Valderrama, C. Alvarado, S. Nikolopoulos, J. Martinerie, C. Adam, V. Navarro, and M. Le Van Quyen, "Identifying an increased risk of epileptic seizures using a multi-feature EEG–ECG classification," *Biomedical Signal Processing and Control*, vol. 7, no. 3, pp. 237–244, 2012.

[111] S. Wang, W. A. Chaovalitwongse, and S. Wong, "Online seizure prediction

using an adaptive learning approach," *IEEE transactions on knowledge and data engineering*, vol. 25, no. 12, pp. 2854–2866, 2013.

[112] D. M. White and C. A. Van Cott, "EEG artifacts in the intensive care unit setting," *American journal of electroneurodiagnostic technology*, vol. 50, no. 1, pp. 8–25, 2010.

[113] M. Winterhalder, T. Maiwald, H. Voss, R. Aschenbrenner-Scheibe, J. Timmer, and A. Schulze-Bonhage, "The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods," *Epilepsy & Behavior*, vol. 4, no. 3, pp. 318–325, 2003.

[114] World Health Organization, "Epilepsy," 2019, last accessed: 21 October 2019. Available at https://www.who.int/news-room/fact-sheets/detail/epilepsy.

[115] H. Zhao, A. P. Sinha, and W. Ge, "Effects of feature construction on classification performance: An empirical study in bank failure prediction," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2633–2644, 2009.

Appendices

A

Detailed Description of Common Features

Linear univariate features

Statistical moments

Statistical moments can be used to describe the amplitude distribution of the signal. Due to their simplicity and linearity properties, they are the most widely used features in seizure prediction studies [15, 30, 31, 43, 74, 78, 86, 87, 103, 104, 108, 110].

Table A.1 describes the first four statistical moments, where x represents the input vector and N the number of samples within a sliding window.

Table A.1: Statistical moments used in feature extraction.

Order	Designation	Description	Formula
1^{st}	Mean	Measure of the average amplitude	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ (A.1)
2^{nd}	Variance	Measure of dispersion of amplitude around mean.	$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2$ (A.2)
3^{rd}	Skewness	Measure of asymmetry of amplitude distribution.	$\chi = \frac{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^3}{\sigma^3}$ (A.3)
4^{th}	Kurtosis	Measure of relative peakedness / flatness of amplitude distribution.	$\kappa = \frac{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^4}{(N-1)\sigma^4} - 3$ (A.4)

Relative spectral power

EEG activity is comprised of oscillations and transients (figure 2.3), where it is assumed that seizure-generation processes lead to changes in the normal rhythmic brain activities [78, 82]. Numerous studies have decomposed the EEG into different

frequency bands (delta, theta, alpha, beta, gamma) and computed their relative power [7, 10, 15, 16, 29–31, 74, 78, 82, 86, 87, 103, 104, 108].

The relative spectral power of each band can be computed by firstly computing the Power Spectral Density (PSD) of the signal within a window (assumed to be short enough to be considered quasi-stationary and long enough to capture low-frequency brain activity) [16]. In its simplest form, this can be performed by applying the Fast Fourier Transform (FFT) and averaging the squared coefficients belonging to the band of interest [30]. The normalized spectral power NP_i for a frequency band i is then computed by dividing the spectral power in the sub-band (P_i) and the total power in the signal (P_{tot}):

$$NP_i = \frac{P_i}{P_{tot}} = \frac{\sum_i PSD(x)}{\sum_{total} PSD(x)}. \tag{A.5}$$

Spectral edge frequency/power

Spectral power in the EEG is primarily concentrated in the lower frequency range, below 40Hz. The spectral edge frequency (f) corresponds to the frequency below which 50% of the total power of the signal is located. Thus, the spectral edge power is a measure of power existing below f [75].

Studies have used this measure and reported a power transfer from low to high frequencies during the pre-ictal stage [10, 30, 31, 78, 86, 87, 103, 104, 110].

Wavelet coefficients

The Discrete Wavelet Transform (DWT) is a time-frequency domain transform, alternative to the FFT, which is able to reflect both frequency and temporal location properties of the signal [74]. Simply put, wavelets provide a time-variant decomposition adapted to the signal, capable of capturing minor details and sudden changes by granting a higher frequency resolution for lower frequencies and higher time resolution for higher frequencies [3]. Wavelet coefficients measure the correlation between the signal and different translations and scales of a mother wavelet [30].

Several studies have used wavelet decomposition and computed other measures, such as signal energy, from its coefficients [3, 26, 30, 57, 74, 86, 87, 104, 108, 110]. In particular, Rasekhi et al. [86] highlighted the predictive power of low-frequency wavelet coefficients, while Bandarabadi et al. [15] developed patient-specific orthogonal mother wavelets through genetic algorithm optimization. Deep Learning approaches by Khan et al. [57] and Tsiouris et al. [108] have also used wavelet-transformed EEG signals as input for their models.

Accumulated energy

The first step to calculate accumulated energy is to average all successive values of energy within a time window of N samples:

$$E_a = \frac{1}{N} \sum_{i=1}^{N} x_i^2. \tag{A.6}$$

Afterwards, for a determined number of consecutive windows (W), it is possible to calculate the energy accumulation for epoch k, as indicated in (A.7):

$$E_{ac_k} = \frac{1}{W} \sum_{i=1}^{W} E_a + E_{ac_{k-1}}. \tag{A.7}$$

The underlying assumption of using this measure is that seizure-generation processes lead to the intensification of brain activity which results in accumulation of energy, as reported in studies by Rasekhi et al. [86, 87], Valderrama et al. [110], Teixeira et al. [103, 104], Moghim et al. [74] and D'Alessandro et al. [26].

Decorrelation time

The auto-correlation function of a signal is defined as the correlation between the signal $x(t)$ and versions time shifted by k samples:

$$c_{xx_k} = \frac{\sum_{i=1}^{N-k} x_i x_{i+k}}{(N-1)\sigma^2}. \tag{A.8}$$

Decorrelation time corresponds to the first zero-crossing of c_{xx_k} and is an indicator of signal periodicity [30]. Brain activity changes due to seizure-generation may be captured by analyzing variations in decorrelation time, as demonstrated in various studies [10, 30, 31, 78, 86, 87, 103, 104, 108, 110, 110].

Linear modeling

Autoregressive (AR) models can be used to model the EEG, assuming stationarity of the signal. In AR modeling, the predicted output in each instant t is a weighted sum of previous p values (p being the order of the model) plus a constant term c and white noise ε. By considering Y_t the predicted value for instant t and

ar_i the model parameters, an AR model can be expressed by (A.9):

$$Y_t = c + \sum_{i=1}^{P} ar_i Y_{t-1} + \varepsilon_t. \tag{A.9}$$

Seizure prediction studies have made use of AR modeling, either based on prediction error [30, 31, 86, 87, 103, 104], which has been found to increase due to pre-ictal changes in brain activity, or the values of modeling coefficients as features [21].

Hjorth parameters

Three parameters have been defined by Bo Hjorth for quantitative description of the EEG: activity, mobility and complexity. Activity is proportional to the variance of the signal ($\sigma^2(x)$), while mobility provides an estimate of the mean frequency and complexity assesses signal bandwidth [75, 86]. These can be defined, respectively, by the following expressions:

$$H_a = \sigma^2(x), \tag{A.10}$$

$$H_m = \sqrt{\frac{\sigma^2(x')}{\sigma^2(x)}}, \tag{A.11}$$

$$H_c = \sqrt{\frac{\sigma^2(x'')\sigma^2(x)}{(\sigma^2(x'))^2}}, \tag{A.12}$$

where x' and x" concern the first and second derivatives of the signal. Differences in mobility and complexity between the pre-ictal and inter-ictal stages have been reported [10, 30, 78, 86, 87, 103, 104, 110].

Non-linear univariate features

Correlation dimension/sum

The reconstruction of a state trajectory $\vec{s_t}$ in a phase plane for a scalar time series x can be computed by the method of delayed coordinates. Hence, in each instant t, for an embedded dimension m and a time delay τ:

$$\vec{s_t} = (x_t, x_{t-\tau}, ..., x_{t-(m-1)\tau}). \tag{A.13}$$

Considering a time series with N samples, each $N-(m-1)\tau$ reconstructed vectors depicts an instantaneous state of the system in an m-dimensional space [75]. The sequential plot of the states defines the state space trajectory, which can illustrate the system's dynamics [1].

The correlation sum C_s (or correlation integral) estimates the local probability density in state space, by counting the numbers of vector pairs in it that are closer than a given hypersphere radius ε. Considering Θ the Heaviside function, W the Theiler window and $M = N-(m-1)\tau$:

$$C_s(\varepsilon) = \frac{2}{(M-W)(M-W-1)} \sum_{i=1}^{M} \sum_{j=i+W}^{M} \Theta(\varepsilon - |\vec{s_j} - \vec{s_k}|) \tag{A.14}$$

From the local slope of the correlation sum, the correlation dimension CD may be obtained by (A.15) and corresponds to an estimate of the number of active degrees of freedom of random points within a state space. Studies by Adeli et al. [3] and Moghim et al. [74] have used it for chaos analysis of the EEG signal dynamics. However, its predictive power has been contested by Harrison et al. [45].

$$CD = \lim_{N\to\infty} \lim_{\varepsilon\to 0} \frac{ln(C_s(\varepsilon))}{ln(\varepsilon)}. \tag{A.15}$$

Entropy

Entropy can be used as a measure of regularity and predictability of the EEG signal fluctuations. Considering that seizures are characterized by an increase of synchronization in neuronal activity (thus, becoming more predictable) [75], various entropy measures have been proposed and exploited to detect changes from the inter-ictal to the pre-ictal stage, such as approximate, sample or spectral entropies [1, 26, 103].

Dynamical similarity index

The dynamical similarity index ζ_t was proposed by Le Van Quyen et al. [66] as a measure of similarity between segments of the EEG, namely, between a reference inter-ictal segment r and a sliding test window t. Mathematically, it is computed by (A.16), where C_{rt} is the cross-correlation sum and C_{rr} and C_{tt} are the auto-

correlation sums of the reference and test window, respectively.

$$\zeta_t = \frac{C_{rt}}{\sqrt{C_r C_t}}. \tag{A.16}$$

This index ranges from zero to one and it can be interpreted as the degree of stationarity, that is, $\zeta_t \approx 1$ when the two segments have the same underlying dynamics [70]. Shifts to lower values during pre-ictal changes have been reported and used in seizure prediction studies [67, 78, 84].

Lyapunov exponents

The largest Lyapunov exponent is a measure of a system's chaotic behavior, quantifying the exponential divergence of nearby trajectories in state space [75]. In other words, given that system predictability is sensitive regarding the initial conditions, chaos can be quantified by measuring the increase in distance between trajectories over time, as given by (A.17):

$$d_t = d_0 e^{Lt}, \tag{A.17}$$

where L_t represents the rate of divergence and an estimate of the largest Lyapunov exponent. This measure was shown to reduce significantly minutes before the pre-ictal stage, indicating that brain dynamics become more predictable preceding seizure occurrence [51]. Despite its use in some studies [3, 74, 78], others have contested its predictive power due to noise present in the EEG signal [64, 65].

Linear multivariate features

Maximum linear cross-correlation

Maximum linear cross-correlation is used to quantify the lag synchronization between two EEG channels. In other words, it can characterize the condition where two signals are identical in terms of phase and amplitude but shifted by a time lag τ [75, 76]. This is done by first computing the linear cross-correlation function C_{xy}:

$$C_{xy}(\tau) = \int_{-\infty}^{+\infty} s_x(t+\tau) s_b(t) dt. \tag{A.18}$$

Afterwards, C_{xy} is normalized to the interval [0,1] and its maximum value is

taken, as given by (A.19):

$$C_{max} = \max_{\tau} \left\{ \left| \frac{C_{xy}(\tau)}{\sqrt{C_{xx}(0)C_{yy}(0)}} \right| \right\}. \tag{A.19}$$

High values ($C_{max} \approx 1$) indicate lag synchronization while lower values represent unsynchronized signals. Significant drops were found during the pre-ictal stage, followed by hyper-synchronization during seizures, as shown by several studies [73, 76, 78, 104, 108].

Non-linear multivariate features

Mean phase coherence

Mean phase coherence was proposed by Mormann et al. [79] to quantify phase synchronization between signals from different channels. Considering the phases of two channels θ_x and θ_y and a time series of length N, the measure is defined by (A.20):

$$R = \left| \frac{1}{N} \sum_{j=1}^{N} e^{i(\theta_x(t) - \theta_y(t))} \right|. \tag{A.20}$$

R represents a normalized value between zero and one, where two channels are fully synchronized when $R = 1$ and completely unsynchronized for $R = 0$. It has been used in numerous studies [5, 16, 78, 84] and compared to maximum linear cross-correlation by Mormann et al. [76], which reported similar predictive performance for both measures.

Dynamical entrainment

Iasemidis et al. [49] proposed a measure to quantify the non-linear behavior of signals from two different channels. It is defined as the statistical difference between the largest Lyapunov exponents L_{max} over a number of l consecutive windows for two signals, by using the index T_{xy} retrieved from a paired t-test:

$$T_{xy} = \sqrt{l} \frac{|\langle L_{max,x} - L_{max,y} \rangle|}{\sigma_{xy}}. \tag{A.21}$$

A low value for T_{xy} indicates high entrainment or, in other words, synchronized activity between two channels [75]. Mirowski et al. [73] adopted this measure.

B

Supplementary Results

Table B.1: Number of Pareto-optimal individuals before and after employing the Decision Maker (DM)

	Pareto-optimal solutions			Selected by the Decision Maker		
Patient ID	**MOGA**	**NSGA-II**	**SMS-EMOA (modified)**	**MOGA**	**NSGA-II**	**SMS-EMOA (modified)**
402	239	1323	1101	53	392	522
8902	278	1406	1495	171	221	167
11002	174	1389	1051	62	928	737
16202	251	1232	1291	54	467	312
23902	235	1209	1444	140	696	1012
30802	179	1221	1317	32	859	695
32702	245	1164	1233	126	410	464
46702	279	1378	1446	65	341	196
50802	88	1202	1246	78	1202	1246
53402	151	1131	1078	86	888	822
55202	200	1447	1136	124	872	640
56402	232	1453	1224	118	308	433
58602	240	1379	1143	50	524	354
59102	149	1426	1401	101	1422	1381
60002	214	1398	1368	109	604	564
64702	205	1252	1352	108	908	488
75202	254	1449	1333	80	570	621
80702	270	1325	1217	91	321	459
85202	196	1446	1177	27	395	420
93402	200	1294	1057	124	847	712
93902	200	1237	1467	42	381	339
94402	227	1448	1363	100	628	628
95202	217	1301	1344	89	383	364
96002	134	1419	1111	90	1242	1033
98102	253	1444	1338	75	261	222
98202	213	1452	1348	49	360	392
101702	239	1267	1049	91	669	567
102202	265	1470	1216	72	323	291
104602	206	1448	1248	105	769	606
109502	205	1122	1394	94	805	1010
110602	290	1456	1271	93	864	697
112802	291	1417	1433	96	370	596
113902	225	1444	1426	97	448	449
114702	239	1188	973	123	796	598
114902	253	1124	1015	98	693	684
123902	189	1188	1349	77	809	680
TOTAL	7925	47949	45455	3190	22976	21401

Table B.2: p-values obtained for the t-test performed to assess whether the differences between fitness score means are statistically significant, considering all Pareto-optimal solutions.

	MOGA vs. NSGA-II	MOGA vs. SMS-EMOA	NSGA-II vs. SMS-EMOA
$S_{samples}$	0.00	0.00	1.00
T_f	0.00	0.00	0.00
N_E	0.00	0.00	0.00

Table B.3: p-values obtained for the t-test performed to assess whether the differences between fitness score means are statistically significant, considering the solutions selected by the DM.

	MOGA vs. NSGA-II	MOGA vs. SMS-EMOA	NSGA-II vs. SMS-EMOA
$S_{samples}$	0.00	0.00	1.00
T_f	0.00	0.00	1.00
N_E	0.00	0.00	0.00

Table B.4: p-values obtained for the t-test performed to assess whether the differences between test results for each MOEA were statistically significant.

	MOGA vs. NSGA-II	MOGA vs. SMS-EMOA	NSGA-II vs. SMS-EMOA
Sensitivity	0.00	0.00	1.00
FPR/h	1.00	0.00	0.00
Random sensitivity	0.00	1.00	0.00
Solutions beat random (%)	0.00	0.00	0.00
Surrogate sensitivity	1.00	0.00	0.00
Solutions beat surrogate (%)	0.00	0.07	0.00

Table B.5: Sensitivity values obtained for the average random predictor, that is, considering the average Seizure Occurrence Period (SOP), average FPR/h and the total number of models and testing seizures.

Patient ID	MOGA	NSGA-II	SMS-EMOA (modified)
402	0.58	0.00	0.00
8902	0.00	0.00	0.00
11002	0.00	0.00	0.00
16202	0.00	0.00	0.00
23902	0.00	0.00	0.00
30802	0.00	0.00	0.00
32702	0.00	0.00	0.00
46702	0.58	0.00	0.00
50802	0.00	0.00	0.00
53402	0.00	0.00	0.00
55202	0.00	0.00	0.00
56402	0.00	0.00	0.00
58602	0.00	0.00	0.00
59102	0.63	0.00	0.00
60002	0.15	0.00	0.00
64702	0.00	0.00	0.00
75202	0.00	0.00	0.00
80702	0.05	0.00	0.00
85202	0.50	0.00	0.00
93402	0.00	0.00	0.00
93902	0.00	0.00	0.00
94402	0.00	0.00	0.00
95202	0.00	0.00	0.00
96002	0.00	0.00	0.00
98102	0.40	0.00	0.00
98202	0.38	0.00	0.00
101702	0.38	0.00	0.00
102202	0.00	0.00	0.00
104602	0.15	0.00	0.00
109502	0.00	0.00	0.00
110602	0.38	0.00	0.00
112802	0.00	0.00	0.00
113902	0.06	0.00	0.00
114702	0.00	0.00	0.00
114902	0.00	0.00	0.00
123902	0.62	0.00	0.00

Table B.6: Patient groups for each stratification criterion: seizure classification, activity patterns and vigilance state.

Patient ID	FOA/FOIA seizures	Rhythmic patterns	Awake
402		•	•
8902	•		•
11002	•	•	
16202			•
23902	•	•	•
30802	•	•	
32702	•		•
46702		•	
50802		•	
53402	•	•	
55202	•		•
56402		•	•
58602	•		
59102	•	•	•
60002	•		
64702			
75202	•	•	
80702			•
85202	•		
93402		•	
93902		•	
94402	•	•	
95202			
96002	•	•	•
98102			•
98202		•	•
101702	•		
102202	•	•	
104602		•	
109502	•	•	•
110602	•	•	•
112802	•	•	•
113902	•	•	
114702	•	•	•
114902			
123902		•	