UNIVERSIDADE D
COIMBRA

Gonçalo Laranjeira Pires dos Santos Costa

Machine Learning Methods for Epileptic Seizure Risk Assessment

Thesis submitted to the Faculty of Sciences 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 Dr. Mauro Pinto.

September 2023

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D COIMBRA

Machine Learning Methods for Epileptic Seizure Risk Assessment

Gonçalo Laranjeira Pires dos Santos Costa

Thesis submitted to the Faculty of Sciences and Technology of the University of Coimbra for the degree of Master in Biomedical Engineering with specialization in Clinical Informatics and Bioinformatics.

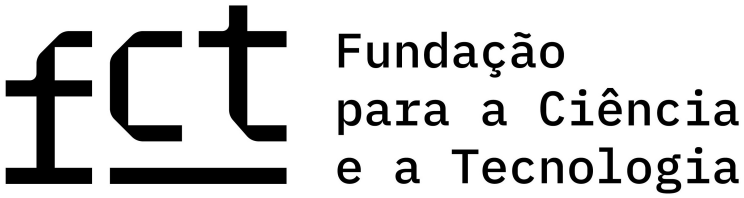
Supervisors:
Prof. Dr. César Alexandre Domingues Teixeira (CISUC)
Dr. Mauro Filipe da Silva Pinto (CISUC)

Coimbra, 2023

This work was developed in collaboration with:

Center for Informatics and Systems of the University of Coimbra

This work is funded by FCT- Foundation for Science and Technology, I.P., within the scope of the projects: CISUC - UID/CEC/00326/2020 with funds from the European Social Fund, through the Regional Operational Program Centro 2020; and project RECoD - PTDC/EEI-EEE/5788/2020 financed with national funds (PIDDAC) via the Portuguese State Budget.



Agradecimentos

Em primeiro lugar gostaria de agradecer ao Professor Doutor César Teixeira e ao Doutor Mauro Pinto por toda a ajuda e orientação no decorrer deste projeto. Ouço muitas pessoas a falar do quão complicado é fazer uma tese de mestrado, mas toda a vossa disponibilidade e atenção tornou este processo muito mais agradável e facilitado do que estava à espera.

À minha mãe, a primeira que me tirou do caminho da medicina para a biomédica. "Deves ser a única mãe do mundo que não quer o filho em medicina" foi uma frase que foi dita várias vezes, mas de facto as mães sabem mais, porque esse foi o melhor conselho que podia ter recebido. Obrigado por me aturares e apoiares, de mau-humor quando as coisas corriam mal e eufórico ao ponto de ser irritante quando corriam bem, a tua preocupação foi dos principais fatores que me permitiu fazer este mestrado da forma que o fiz.

Ao meu pai, que andou em processo de tese de mestrado ao mesmo tempo que eu. Não importa o resultado de qualquer um dos nossos mestrados, ter conseguido ensinar-te finalmente a diferença entre o Grammarly e o QuillBot foi das maiores vitórias dos últimos tempos. Obrigado pelas conversas que ajudavam a desbloquear certas partes do meu trabalho, por todos os conselhos no decorrer deste processo, por todas as vezes que ignoraste os murros na secretária quando o Python dava erro e por não me deserdares quando ia à sala pegar nos teus chinelos e atirá-los ao teto quando o código corria como era suposto.

Ao resto da minha família, ao Avô Pires, à Avó Nanda, ao Tio Toni, à Tia Íris, ao Avô Carlos, à Avó Lai, à Tia Guida e à Bisa Zélia, pelos valores e princípios que sempre me incutiram, por todo o apoio, preocupação e cuidado que tiveram não só durante a minha vida académica, mas também em toda a minha vida e que sei que sempre terão. E um obrigado em especial aos meus meninos, a Kika, o Afonso e a Margarida por me distraírem das responsabilidades e deixarem-me entrar no vosso mundo de brincadeira e maluquice.

À Mari, a maior peste que conheci neste curso. Tanto te adoro e te quero dar um abraço como te odeio e te quero arrancar a cabeça. És sem dúvida a irmã que nunca tive, obrigado por tudo o que fizeste por mim, especialmente nos momentos piores. Sei que posso contar contigo para tudo e em todo o lado (a não ser que esse lado seja frio, haja bicicletas, a cerveja custa 4 € e o café 3 €), e espero que saibas que te estou eternamente grato por estes últimos 5 anos. Estou a rezar para que tenhas conseguido entregar a tua tese com 0 erros amarelos no Overleaf!

À Tini, a minha companheira bioinformática, que não bastou ter de levar comigo durante as cadeiras de mestrado quase todas, ainda teve de me aturar durante o projeto. Obrigado por tudo, já disse várias vezes que sem ti a trabalhar comigo ou não tinha feito o curso, ou tinha-me mandado da janela da sala de mestrado porque o raio do código não corria. Companheira de programação e de sushi, só te desejo todo o sucesso do mundo que alguém ridiculamente trabalhadora como tu mereces.

Ao Mestre e ao Traçaço, os meus meninos mais velho e mais novo que sei que vou levar para a vida. "The battles comes and goes but my soldiers are eternals". A amizade que me deram ao longo destes anos é algo que nunca esperei e que vou sempre estimar. Desde ajudas com coisas relacionadas com esta tese a ajudas a acabar de comer a travessa de marisco ou canecas de cerveja preta, sempre lá estiveram para mim. Espero e sei que, não importa a que parte do mundo a nossa vida nos levar, esta ligação vai estar sempre presente e forte.

Aos restantes membros d'"Os Borgas", obrigado por terem marcado de uma forma extraordinária o meu percurso académico. Aos velhotes, o Dias, o Artur e o Jójó, obrigado por me terem aceite de braços abertos e me terem incutido os valores de Coimbra. Ao meu maninho Mendas, obrigado por teres estado ao meu lado durante todo este percurso, aprendi muito contigo. Aos mais novitos, o Esgaçu, o Tiaguinho, o Gui, o Ivo, o aldeão Arnaut, o Jaimito, o Morais e o Topê, obrigado por me fazerem ver todos os anos o que é ser um estudante desta cidade magnífica.

A todos os meus restantes amigos, o maior obrigado que consigo expressar! Em específico aos "Cruzados", obrigado por todos os anos de diversão e alegria que me proporcionaram. À Madri Botas, obrigado por seres quem és, uma jovem com espírito de velha professora, com quem posso falar de tudo (até mesmo escolher um tema para uma tese de mestrado), e que, já agora, ainda me está a dever uma sopa da pedra..... Às minhas afis, Sebasti, Sofia e Maria, e à minha Laurita Gentlemansss, obrigado por me terem deixado acompanhar o vosso percurso e por nunca deixarem que um momento convosco seja triste ou aborrecido. À Fipas, obrigado por teres tido toda a paciência do mundo quando voltei da Holanda e o meu cérebro ainda estava um pouco em papa, és das mais antigas e espero que recebamos os dois dinheiro suficiente para fazer uma jantarada semanal no sushi mais caro do país.

E por último, mas não menos importante, à minha Carol. Foste a adição mais recente, quando entraste na minha vida ainda estava a começar a descobrir o que era epilepsia, mas rapidamente chegaste a um nível destacado da maior parte. Estiveste comigo durante todo o processo de escrita desta tese e posso dizer que as minhas idas à Guarda, Viseu ou Seia, e as tuas a Coimbra ou Figueira deram-me o estado de espírito perfeito para conseguir fazer um projeto destes. Fazes-me querer ser melhor, obrigado por toda a atenção, preocupação, carinho e paciência para aturar um chato como eu. Não tenho palavras para ti, um gigante obrigado por tudo aquilo que és!

"We make our world significant by the courage of our questions and the depth of our answers"

Carl Sagan

Resumo

A epilepsia afeta cerca de 1% da população mundial. Os medicamentos antiepiléticos são uma excelente opção para controlar a ocorrência de crises, mas não funcionam em aproximadamente um terço dos pacientes. Os dispositivos de alerta que utilizam algoritmos de previsão ou avaliação do risco de crises podem trazer aos pacientes um novo conforto e qualidade de vida. Estes algoritmos tentam detetar o período pré-ictal de uma crise, um momento de transição entre a atividade cerebral normal e a crise, e transmitem essa informação ao utilizador.

Ao longo dos anos, foram desenvolvidos muitos estudos de previsão de crises utilizando metodologias baseadas no Eletroencefalograma (EEG), que disparam um alarme quando detetam fases precoces do período pré-ictal. No entanto, poucas destas metodologias têm sido clinicamente aplicáveis. Esta ineficácia deve-se principalmente a três razões: a falta de bases de dados apropriadas; o sinal EEG ainda não é totalmente compreendido; determinar e detetar o período pré-ictal é uma tarefa incrivelmente árdua.

Estudos recentes sugerem uma mudança de perspetiva. A avaliação do risco de crises adota uma abordagem probabilística do problema em questão, em contraste com a abordagem baseada em alarmes na previsão de crises. A avaliação do risco de crises substitui os alarmes acionados para simbolizar a deteção de um período pré-ictal na previsão de crises por uma análise contínua. Dependendo da probabilidade de ocorrência de uma crise, são definidos diferentes estados de risco que são constantemente apresentados ao paciente.

O presente trabalho tem como objetivo explorar metodologias capazes de avaliar o risco de crises e estabelecer uma comparação com os resultados da previsão de crises.

Usando dados de 40 pacientes com epilepsia da base de dados *EPILEPSIAE*, três algoritmos de previsão de crises e três de avaliação do risco de crises específicos para cada paciente foram implementados e comparados. Uma Regressão Logística, um conjunto de *Support Vector Machines* (SVMs) e um conjunto de *Shallow Neural Networks* (SNNs) foram utilizados como classificador para a previsão de crises e para a avaliação do risco.

Nenhuma das metodologias de previsão de crises obteve resultados suficientemente bons para definir a melhor metodologia proposta. Entre os modelos que utilizam a Regressão Logística e o conjunto de SNNs, não houve equilíbrio entre um valor de sensibilidade suficientemente elevado e um valor da taxa de falsos alarmes por hora (FPR/h) suficientemente baixo para ser clinicamente aceitável.

A melhor metodologia proposta de avaliação do risco de crises utilizou o classificador de Regressão Logística. Obteve uma sensibilidade de 0.28±0.37, um *Time in Warning* de 0.13±0.14, um *Brier Score* de 0.19±0.11 e um *Brier Skill Score* de 0.01±0.15, em que 50% dos

modelos dos pacientes obtiveram resultados estatisticamente significativos superiores ao acaso.

Comparando os resultados nas duas áreas de estudo, um aumento da sensibilidade é observado no modelo de avaliação de risco em relação ao de previsão. Houve uma melhoria de 115% para a Regressão Logística, de 146% para o modelo do conjunto SVMs e de 70% para o modelo de conjunto SNNs. Os resultados da validação estatística foram semelhantes. O número de pacientes que apresentaram uma melhoria em relação ao acaso aumentou na avaliação do risco de crises em 300%, 171% e 125%, respetivamente.

Portanto, com este estudo é possível concluir que a metodologia de avaliação do risco de crises supera a de previsão, tanto em termos de sensibilidade como de melhoria em relação ao acaso, e pode ser mais adequada para dispositivos de alerta de crises.

Palavras-Chave: Epilepsia, Avaliação do risco de crises, Previsão de crises, Aprendizagem Computacional, Eletroencefalograma

Abstract

Epilepsy affects around 1% of the population worldwide. Anti-Epileptic Drugs (AEDs) are an excellent option for controlling seizure occurrence but do not work for approximately one-third of patients. Warning devices employing seizure prediction or forecasting algorithms could bring patients new-found comfort and quality of life. These algorithms attempt to detect a seizure's preictal period, a transitional moment between regular brain activity and the seizure, and relay this information to the user.

Over the years, many seizure prediction studies using Electroencephalogram (EEG)-based methodologies have been developed, triggering an alarm when detecting early stages of the preictal period. However, few of these methodologies have been clinically applicable. This inefficacy is mainly for three reasons: the lack of proper databases; the EEG signal is still not fully understood; determining and detecting the preictal period is an incredibly arduous task.

Recent studies have suggested a shift in view. Seizure forecasting takes a probabilistic approach to the problem in question instead of the alarm-based approach in seizure prediction. Seizure forecasting replaces the triggered alarms that symbolize the detection of a preictal period in seizure prediction with a continuous analysis. Depending on the probability of a seizure, different risk states are defined and constantly displayed to the patient.

The present work aims to explore methodologies capable of seizure forecasting and establish a comparison with seizure prediction results.

Using data from 40 epilepsy patients from the EPILEPSIAE database, three patient-specific seizure prediction algorithms and three seizure forecasting ones were implemented and compared. A Logistic Regression, a Support Vector Machine (SVM) ensemble, and a Shallow Neural Network (SNN) ensemble classifier were employed for both prediction and forecasting.

None of the seizure prediction methodologies achieved satisfactory enough results to define the best proposed methodology. Between the models using the Logistic Regression and the SNN ensemble, there was no balance between a Sensitivity (SS) value sufficiently high and a False Positive Rate per Hour (FPR/h) value sufficiently low to be clinically acceptable.

The best proposed seizure forecasting methodology used the Logistic Regression classifier. It achieved a SS of 0.28±0.37, a Time In Warning (TiW) of 0.13±0.14, a Brier Score (BS) of 0.19±0.11, and a Brier Skill Score (BSS) of 0.01±0.15, where 50% of patient models achieved statistically significant results higher than chance.

Comparing the performance results in both fields of study, an increase of the SS is observed in forecasting relative to prediction. There was a 115% improvement in the Logistic Regression model, 146% in the SVM ensemble model, and 70% in the SNN ensemble model. Findings for

statistical validation were similar. The number of patients that displayed an improvement over chance increased in seizure forecasting by 300%, 171%, and 125%, respectively.

Therefore, with this study, it is possible to conclude that the seizure forecasting methodology outperforms the prediction one, both in SS and improvement over chance, and may be more suitable for seizure warning devices.

Keywords: Epilepsy, Seizure Forecasting, Seizure Prediction, Machine Learning, Electroencephalogram

Contents

List of Figures

List of Tables

List of Abbreviations

AED Anti-Epileptic Drug. xi, 1, 2, 14, 15, 17, 23, 28

BS Brier Score. xi, 24, 25, 29, 56, 57, 67, 68, 73, 75, 77
BSS Brier Skill Score. xi, 24, 25, 29, 45, 56, 67, 68, 69, 73, 75, 77

CD Concept Drift. 2, 28, 29, 78
CNN Convolutional Neural Network. 41, 44, 49

DL Deep Learning. xvii, 32, 41, 44, 45
DRE Drug-Resistant Epilepsy. 1, 15, 16, 28, 35, 48

EEG Electroencephalogram. xi, 2, 3, 5, 8, 9, 10, 12, 13, 15, 17, 18, 20, 22, 27, 28, 29, 31, 32, 35, 37, 38, 44, 47, 48, 49, 50, 74, 75, 97, 98, 99, 100

FBCT Focal to Bilateral Tonic-Clonic. 7, 8
FFT Fast Fourier Transform. 44, 98
FOA Focal Onset Aware. 7, 8
FOIA Focal Onset Impaired Awareness. 7, 8
FPR/h False Positive Rate per Hour. xi, 22, 23, 26, 29, 38, 44, 45, 57, 67, 73, 74, 77

ICA Independent Component Analysis. 38, 99
IED Interictal Epileptiform Discharge. 9, 13, 17, 29
iEEG Intracranial EEG. 10, 12, 13, 17, 20, 29, 35
ILAE International League Against Epilepsy. 5, 6, 7, 8, 15
IoC Improvement over Chance. 25, 57, 67, 68, 69, 74

LOOCV Leave-One-Out Cross-Validation. 54
LSTM Long Short-Term Memory Network. 41, 44

ML Machine Learning. 2, 3, 23, 31, 32, 43, 59, 74, 77

RC Reliability Curve. 24, 25, 29, 56, 68, 69
RNS Responsive Neurostimulation. 16, 35

SNN Shallow Neural Network. xi, xix, 47, 52, 53, 54, 55, 59, 62, 63, 66, 67, 68, 69, 71, 73, 74, 75, 77, 101, 103

SOP Seizure Occurrence Period. 20, 21, 22, 25, 26, 29, 36, 50, 51, 52, 53, 54, 55, 56, 59, 70, 73, 75
SP Specificity. 18, 23
SPH Seizure Prediction Horizon. xix, 20, 21, 22, 29, 36, 38, 51, 56, 70, 75
SS Sensitivity. xi, xii, 18, 22, 23, 24, 29, 38, 44, 45, 56, 57, 66, 67, 68, 69, 70, 73, 74, 75, 77
ssEEG Sub-scalp EEG. 12, 20, 35, 75
SUDEP Sudden Unexpected Death in Epilepsy. 9, 14, 17
SVM Support Vector Machine. xi, xix, 41, 43, 47, 50, 52, 53, 54, 55, 59, 61, 63, 67, 69, 70, 73, 77, 101

TiW Time In Warning. xi, 24, 29, 45, 56, 67, 69, 73, 75, 77
TLE Temporal Lobe Epilepsy. 8, 28, 48

VNS Vagus Nerve Stimulation. 16, 17

1

Introduction

This chapter begins with an overview of the motivation behind this project in Section 1.1 while also presenting the major differences between prediction and forecasting and current limitations. Section 1.2 presents the expected goals and contributions. Finally, Section 1.3 outlines the structure of this thesis.

1.1 Motivation

Epilepsy is one of the most common neurological diseases. It affects around 1% of the world's population and is characterized by recurrent seizures. One significant problem patients with epilepsy suffer from is the apparent unpredictable nature of seizures [1].

The first line of treatment is the use of Anti-Epileptic Drugs (AEDs). However, approximately one-third of patients with epilepsy suffer from Drug-Resistant Epilepsy (DRE), a condition where the use of medication is not enough to achieve seizure-free lives [2, 3]. The inability to control seizures can lead to physical problems, such as an increased risk of accidental injury, brain injury, cognitive decline, or even death, and psychological ones, such as neuropsychological deficits (memory loss and attention difficulties), depression, anxiety, or psychoses [4–6].

In cases where seizure control cannot be achieved through medication, surgery, or neurostimulation, the objective becomes informing the patients when a seizure will occur or an estimation of the seizure likelihood through warning devices [7].

1.1.1 Seizure Prediction and Forecasting

Researchers take two main approaches when developing an algorithm for a seizure warning device. One is seizure prediction, where an alarm is raised when the algorithm detects the preictal period. This alarm-based view means that whenever an alarm is raised, the information that is given by the algorithm is that a seizure is sure to occur in a very near period of time. The other is seizure forecasting, where the algorithm warns about the likelihood of a seizure but does not guarantee it.

From an optimal point of view, seizure prediction would be the ideal option. When a seizure is guaranteed to occur, it gives a warning, and actions can be taken accordingly. However, seizure generation is a very complex area, and there is still much to evolve, so as of today, seizure forecasting is considered more feasible and flexible than seizure prediction. Even though seizure forecasting takes a probabilistic approach, which is rarely fully confident of events, it avoids the

crisp approach of prediction that, despite sometimes being correct, is linked to significant rates of false alarms. This may happen in cases where the brain is highly susceptible to a seizure but does not fully develop it [8]. Access to the probabilistic likelihood of seizures allows patients to make an informed decision based on a certain degree of uncertainty [9, 10].

1.1.2 Seizure Prediction and Forecasting Limitations

Current prediction and forecasting approaches display several noteworthy limitations.

Firstly, the most effective way to diagnose and analyze epilepsy is through the Electroencephalogram (EEG) [11]. However, this signal is not yet entirely understood. Moreover, most public EEG databases' data comes from presurgical monitoring, where patients suffer from AEDs withdrawal and sleep deprivation. These conditions do not reflect everyday seizure activity. Significant advancements come with the transition to ultra-long-term databases containing months to years of daily-life data. However, there are still not many of these databases, and the existing ones do not have public access (some do, but only share a considerably smaller, discontinuous, and more limited subset). Furthermore, even then, the rare seizure occurrence creates a considerable class imbalance with the data received [12–14].

Brain dynamics are also a topic of great complexity. Patients with epilepsy have faulty brain seizure regulation mechanisms that allow seizures to occur [8]. Current state-of-the-art uses supervised learning techniques in mostly EEG data to classify information as interictal, where no seizure occurs, or preictal, where a seizure is about to occur. Thus, a critical part of the process is determining the preictal period. However, there is no consensus among authors on its length, and some evidence even suggests the preictal period may vary between patients and between the same patient's seizures [12, 13, 15, 16]. A way to approach this limitation is the development of patient-specific algorithms.

Another problem hindering seizure prediction algorithms' performances is the existence of Concept Drifts (CDs). An elevated number of real-world variables can affect and influence a patient's seizure propensity by altering brain dynamics. These variables include the circadian, diurnal, and ultradian cycles, the sleep-wake cycle, medication changes, and changes in behavior or mood [13–15, 17, 18].

1.2 Goals and Contributions

Epileptic seizure prediction aims to develop an algorithm to anticipate an epileptic seizure and warn patients or caregivers before it begins. Over the years, significant advancements have been made in this area with the use of EEG signals [19].

Considering the recent shift from seizure prediction to forecasting, which more aptly deals with seizure generation mechanisms, this thesis aims to evaluate this transition in the EPILEPSIAE database [20]. Therefore, this thesis aims to develop a Machine Learning (ML) patient-tailored seizure forecasting algorithm that uses information on past seizures to determine the constant probability of a seizure and compare its performance to a seizure prediction one.

Thus, using scalp EEG data from the previously mentioned database and several ML techniques, this thesis is expected to have the following contributions:

- Development of three patient-specific seizure forecasting algorithms using three different classification techniques;
- Development of three patient-specific seizure prediction algorithms using three different classification techniques;
- Evaluate the performance of the different classifiers' models in each field mentioned above;
- Establish a comparison between seizure prediction and forecasting by evaluating the performance of the developed methodologies in the same set of patients.

1.3 Thesis Structure

This document is organized into five chapters beyond the introduction:

Chapter 2 provides background information regarding epilepsy, the EEG signal, treatment and therapeutics options, and provides an introduction to seizure prediction and forecasting.

Chapter 3 presents the state-of-the-art concerning EEG-based seizure prediction and forecasting.

Chapter 4 describes the various methodology steps employed throughout the experimental work.

Chapter 5 reports not only the results obtained in this study but also a discussion and an interpretative analysis.

Chapter 6 presents a conclusion and addresses future perspectives.

1.4 Scientific Contributions

During this thesis, a scientific contribution to the field of epilepsy seizure forecasting and prediction was made. A paper entitled "*Machine Learning Methods in Seizure Prediction and Forecasting: What Is the Best Approach?*" was written and submitted to the International Conference on Data Science and Engineering in Healthcare, Medicine & Biology organized by the IEEE Engineering in Medicine and Biology Society. The full conference paper can be found in Appendix A.

2

Background Concepts

This chapter introduces the main concepts required to follow this thesis. Firstly, Section 2.1 presents some notions and definitions related to epilepsy and seizures, followed by an overview of the Electroencephalogram (EEG) signal in Section 2.2. Section 2.3 refers to the present-day treatment and therapeutic options. Section 2.4 provides the necessary concepts to understand seizure prediction and forecasting, and Section 2.5 details some relevant concept drifts regarding seizure prediction and forecasting. Finally, Section 2.6 brings forth a summary of the key concepts discussed during this chapter.

2.1 Epilepsy and Seizure Concepts

Epilepsy is a disease that can develop at any time, with the highest incidence age being in the first years and after 65 years old. It is considered one of the most common neurological diseases, affecting around 1% of the world's population. [1, 21].

However, despite its significance and prevalence, there was no agreed-upon general definition of either epilepsy or seizures until 2005. In that year, representatives of the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE) proposed the following definition of epilepsy as [22] "*a disorder of the brain characterized by an enduring predisposition to generate epileptic seizures and by the neurobiologic, cognitive, psychological, and social consequences of this condition.*" The ILAE and the IBE also add that the definition "*requires the occurrence of at least one epileptic seizure*" after defining an epileptic seizure as "*a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain.*"

In 2014, the ILAE Task Force proposed an operational, clinical definition of epilepsy that designated it as a brain disease (formerly a disorder) characterized by any of the following conditions [23]:

1. "*At least two unprovoked (or reflex) seizures occurring >24 h apart.*"
2. "*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.*"
3. "*Diagnosis of an epilepsy syndrome.*"

Furthermore, the 2005 definition did not consider that one can be off medication with no seizures for decades and still be deemed to have epilepsy. Accordingly, the Task Force also set a time limit for the disease that goes as follows [23]: "*Epilepsy is considered to be resolved for individuals who had an age-dependent epilepsy syndrome but are now past the applicable age or those who have remained seizure-free for the last 10 years, with no seizure medicines for the last 5 years*". The term "resolved" indicates that even though the disease may return, the patient at that moment no longer has it.

2.1.1 Classification

The classification of epilepsy and seizures is another aspect of great importance when dealing with and evaluating people suffering from the disease. When there are good and precise classification guidelines, it is easier to examine a patient. It also facilitates understanding the types of seizures they suffer from and any others they may suffer, providing insight into possible triggers and any risks of coexisting conditions (learning difficulties, intellectual disabilities, autism spectrum disorder, among others) [24].

Despite the previously discussed work by the ILAE to improve the definitions of epilepsy and seizures, its classification is a topic that had not been subject to change since a 1989 update by the same association (barring a revision made in 2010).

The endurance of this classification, though, does not mean that it was flawless. With the advances in neurobiology, genomic technologies, and molecular biology, a new classification that focuses more on the scientific side of the subject and therefore makes epilepsy much more understandable by clinicians, patients, and caretakers was long overdue [25, 26].

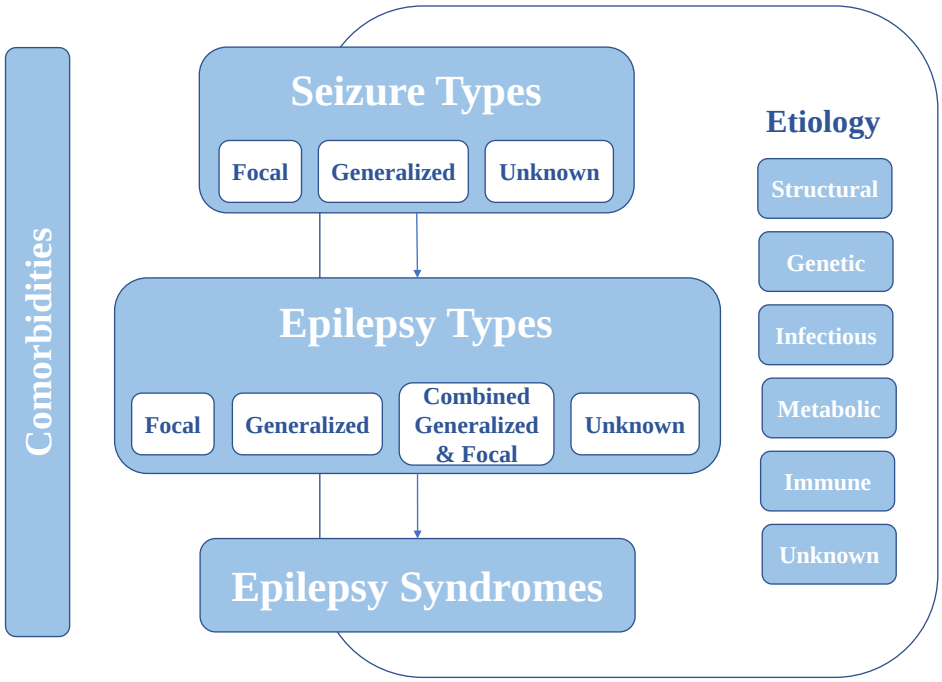


Figure 2.1: International League Against Epilepsy 2017 framework for the Classification of Epilepsies. Adapted: Scheffer et al. (2017) [24].

Accordingly, in 2017, the ILAE proposed an updated version of the Classification of Epilepsies composed of three different levels, as can be seen in Figure 2.1, starting with seizure types, then moving onto epilepsy types, and finally, epilepsy syndromes. The depth of the diagnosis can vary depending on the resources available to the practicing clinician. However, ideally, it should cover all three levels and consider the patient's etiology and comorbidities [24]. However, the importance of these two topics is beyond the scope of this thesis and will not be further discussed.

2.1.1.1 Seizure Types

After a seizure, the clinician's first step toward a diagnosis is to classify it into a seizure type, defined as [27] "*a useful grouping of seizure characteristics for purposes of communication in clinical care, teaching, and research.*" The type of seizure depends on where the initial manifestations begin in the brain. As depicted in Figure 2.2, the seizure can be branched into the following categories: focal, generalized, unknown (if the onset is obscured or missed), and subcategories: motor, non-motor.

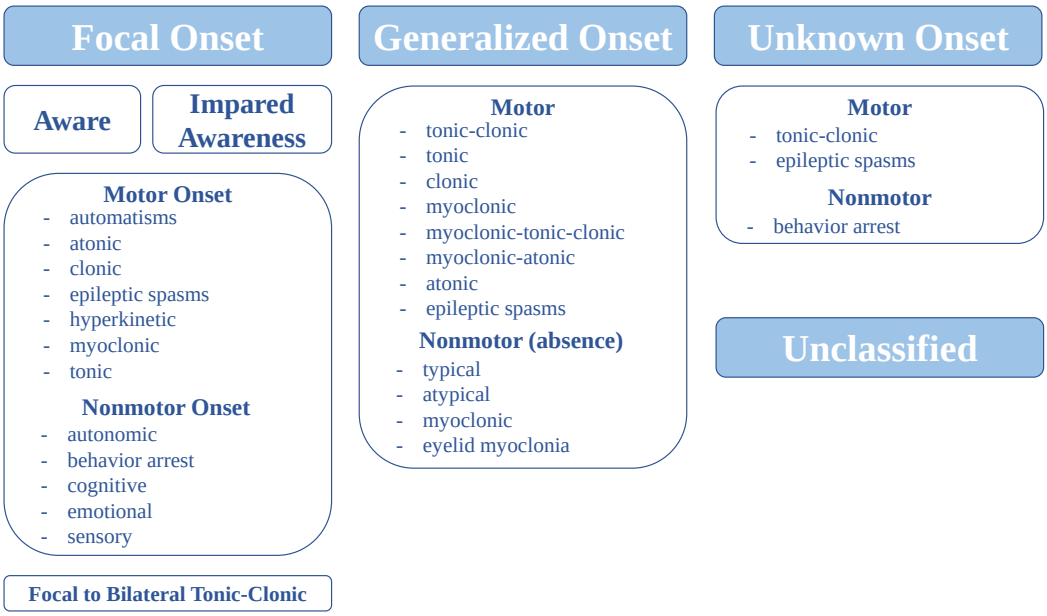


Figure 2.2: International League Against Epilepsy 2017 expanded classification of seizure types. Adapted: Fisher et al. (2017) [27].

A focal seizure is defined as [27] "*originating within networks limited to one hemisphere. They may be discretely localized or more widely distributed. Focal seizures may originate in subcortical structures.*" They can be further subdivided depending on the state of awareness of the patient. When awareness is retained (the patient is aware of himself and his surroundings, even if immobile), the seizure is designated as Focal Onset Aware (FOA); otherwise, it is defined as Focal Onset Impaired Awareness (FOIA). A unique type of seizure is the Focal to Bilateral Tonic-Clonic (FBCT), which has a focal onset but quickly propagates to another hemisphere. As the name indicates, this can result in tonic and clonic symptoms.

A generalized seizure is defined as [27] "*originating at some point within, and rapidly engaging, bilaterally distributed networks.*" Opposite to focal onset seizures, awareness is not a good classifier, as a significant part of generalized seizure results in individuals with either impaired awareness or loss of consciousness, despite the possibility of being partially retained.

2.1.1.2 Epilepsy Types

The next step in the classification is determining the epilepsy type, which presupposes that the patient was diagnosed based on the ILAE 2014 definition of epilepsy [23]. As shown in Figure 2.1, the types of epilepsy are divided into the following categories:

- Focal: encapsulates not only unifocal and multifocal disorders but also includes seizures involving one hemisphere. The seizure types in this group include FOA, FOIA, FBCT, focal motor, and focal non-motor. The interictal EEG tends to display focal epileptiform discharges;
- Generalized: encapsulates seizures in a wide range, such as absence, myoclonic, atonic, tonic, and non-tonic. The interictal EEG tends to show generalized spike-wave activity;
- Combined Generalized & Focal: encapsulates both focal and generalized seizures. The interictal EEG shows both focal epileptiform and generalized spike-wave discharges;
- Unknown: if there is not enough information for the clinician to assess if the epilepsy type is focal or generalized, but there is a diagnosis of epilepsy.

This level of classification has the possibility of being the end of the diagnosis as there are cases where a clinician cannot make an epilepsy syndrome diagnosis [24].

2.1.1.3 Epilepsy Syndrome

The following level of epilepsy classification comes in the diagnosis of an epilepsy syndrome, understood to be a cluster of features comprised of seizure types, EEG, and specific imaging characteristics that usually occur together. Additionally, it can incorporate, among others, the following features: age at onset and remission (where applicable), seizure triggers, and diurnal variation [24].

Even though there is a great variety of well-known syndromes, the ILAE has yet to perform a formal classification. The correct identification of an epilepsy syndrome helps assign the proper medication to the patients [3, 24].

Temporal Lobe Epilepsy (TLE) is the most common epilepsy syndrome. One characteristic of patients with TLE is that they have focal epilepsy. It is so frequent that 60% of people with focal seizures have TLE [28].

2.1.2 Seizure Clusters and Status Epilepticus

Many patients with epilepsy may suffer from seizure clusters [29, 30]. Seizure clusters are not among the ILAE epilepsy definitions, so there is no global agreement on its description. It is

generally defined as "*acute episodes of deterioration in seizure control.*" In other words, seizure clusters consist of a series of consecutive seizures with short interictal periods between them. Currently, there is no consensus on this interictal period's length [31].

Seizure clusters, if not quickly dealt with, have the potential to progress into Status Epilepticus (SE), and SE is not likely to stop occurring without external assistance. Moreover, this event will frequently develop treatment resistance in the patient, leading to severe neurological damage and even Sudden Unexpected Death in Epilepsy (SUDEP). So there needs to exist easy ways to deliver rescue treatment during everyday life, which Section 2.3 will cover [30, 32, 33].

2.2 EEG

The EEG is a high-temporal-resolution technique for medical imaging, and it is considered one of the most effective ways to diagnose brain disorders, examining and interpreting their characteristics.

The human brain works on electrical activity fired by neurons. Thus, the EEG registers the voltage fluctuations caused by the excitatory/inhibitory postsynaptic potentials created in the cortical pyramidal neurons [34–36].

This sort of high-temporal-resolution imaging tool is ideal for studying processes related to neurocognition. For one, it can record cognitive dynamics when cognition occurs since these processes are fast, and high-temporal-resolution techniques effectively capture events with these characteristics. Moreover, it is outstanding in recording neural activity, and its signal is multidimensional as it comprehends time, space, frequency, and power and phase.

However, it is not a global answer for all studies. It is ineffective at determining an accurate functional location when researching slower cognitive processes and testing theories concerning deep brain regions [35].

Generally, the EEG can capture two different phenomena (see Figure 2.3): oscillations and transients, subdivided into normal and subnormal [37].

Oscillations concern activity that starts in the cortex and is dependent on thalamocortical reciprocity to a certain degree. Furthermore, they are often described as rhythmic fluctuations in the excitation-inhibition dynamics of groups of neurons. Normal oscillations can be categorized by their waveform frequency and thus divided into the following groups: delta (0.5 - 4 Hz), theta (4 - 8 Hz), alpha (8 - 13 Hz), beta (13 - 30 Hz), and gamma ($>$ 30 Hz); abnormal oscillation consists of burst-suppression and seizures. It should also be noted that, despite being relatively well defined, there is no definitive consensus among authors on the frequency bands of normal oscillations and that the waves' characteristics are not constant as time goes by.

The transients are composed of sharp changes that distinguish themselves from any background activity. Normal transients incorporate sleep episodes and artifacts. The latter can be either physiological (caused mainly by body movement issues) or external (caused by cable defects, 50-60 Hz power supply interference, and electrical noise, for instance). Abnormal transients can not only be a foundation to diagnose epilepsy (Interictal Epileptiform Discharges (IEDs)) but also the grounds to find other brain diseases (non-epileptiform) [34, 35, 37, 38].

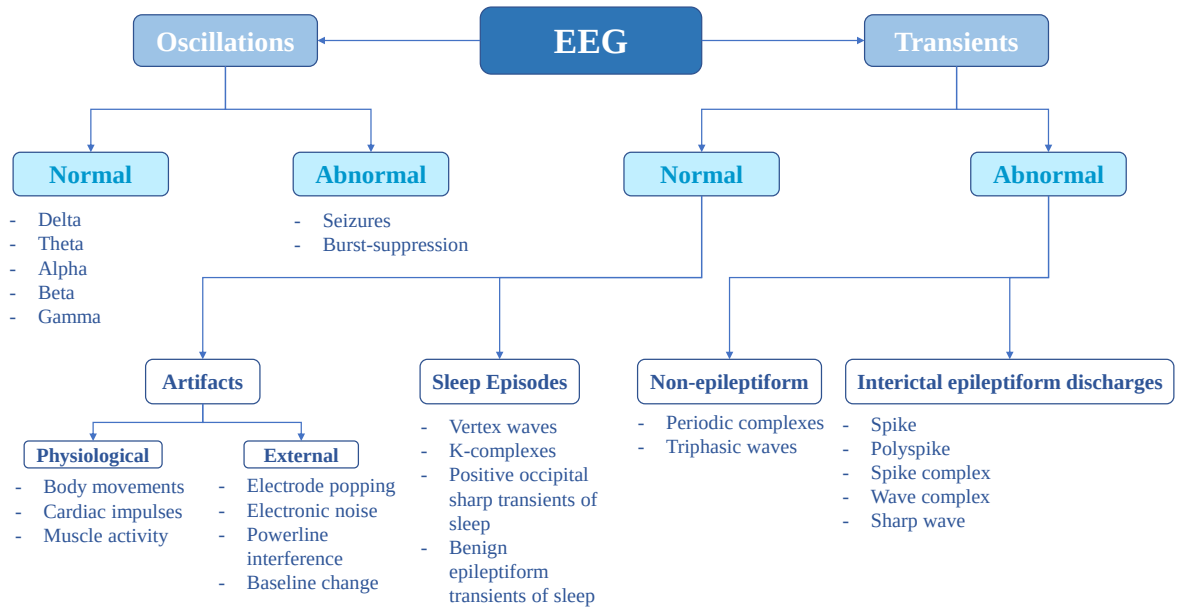


Figure 2.3: Categorization of Electroencephalogram activity. Adapted: Osorio et al. (2016) [37] and Sanei and Chambers (2007) [38].

2.2.1 Signal Acquisition

The acquisition of the EEG signal is executed by placing electrodes either in the scalp (scalp EEG) or inside the patient's skull (Intracranial EEG (iEEG)). The spatial resolution of the signal is determined by the number of electrodes and their corresponding localization, whereas the sampling frequency determines the time resolution.

2.2.1.1 Scalp EEG

The scalp EEG (or just EEG) captures the signal through the use of electrodes placed on the patient's scalp and, as a non-invasive technique, is considered the most common way for its acquisition [39, 40].

The electrodes are placed according to a long-standing international convention called the 10-20 system (see Figure 2.4 a)), consisting of 21 recording electrodes and one ground electrode. The electrodes on the left side of the skull are given odd numbers, the ones on the right even numbers, and the ones on the midline have the letter "z." They are also given different prefixes that signify their region: "Fp" (Fronto-polar); "F" (Frontal); "C" (Central); "T" (Temporal); "P" (Parietal); "O" (Occipital); "A" (Auricular). This system can bypass the need for placement on the eyes and set some fixed distances by looking at particular anatomic positions and considering the electrode interval of 10-20% of those. This facilitates comparisons between different patients. It even allows adding extra electrodes to improve spatial resolution and location (see Figure 2.4 b)) [37, 38, 41].

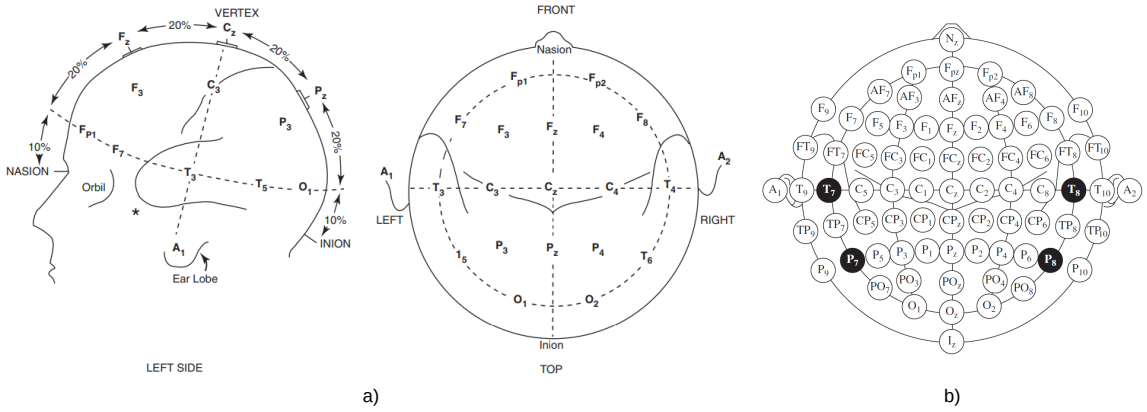


Figure 2.4: International 10-20 system for electrode placement: a) represents the default placements and b) represents the extended 75 electrodes 10-20 configuration. Source: Sanei and Chambers (2007) [38] and Sazgar and Young (2019) [41].

Some of the most used montages for scalp EEG include (see Figure 2.5) [34, 41]:

- Longitudinal Bipolar (Double Banana): for activity traveling between the front and back of the brain; suitable when the focus is highly localized waveforms of low-to-medium amplitude.

- Transverse Bipolar: similar to the previous montage, except the activity travels between the brain hemispheres.

- Referential: a reference electrode is chosen; it should be selected with caution so as not to contaminate any recorded activity; there is a more considerable interelectrode distance, leading to a waveform with higher amplitude.

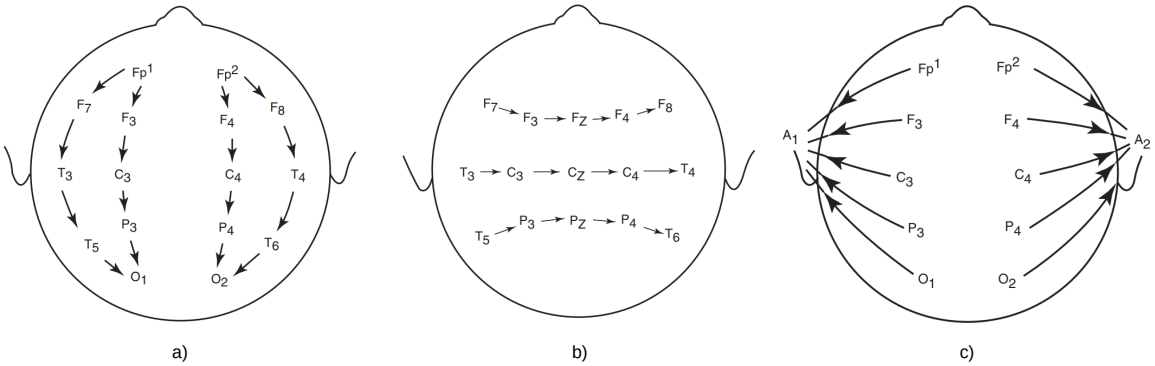


Figure 2.5: Three common scalp Electroencephalogram montages: in a) the longitudinal bipolar montage; in b) the transverse bipolar montage; in c) the referential montage. Source: Sazgar and Young (2019) [41].

Nevertheless, there are restraints attached to this method of acquisition. Despite being relatively cheap and not particularly hard to perform, it captures multiple external artifacts. Also, scalp EEG may be ineffective in recording specific frequencies in the beta and gamma bands and cannot reach the deep brain. Moreover, the equipment to connect the skin and the electrodes requires constant maintenance, and using the same electrodes is usually not recommended for over two weeks [34, 42].

2.2.1.2 iEEG

There are times when the scalp EEG does not provide enough information, requiring more in-depth analysis [34].

The iEEG is an invasive method of capturing the signal by placing the electrodes directly in the deep brain or on the brain surface. Unlike scalp EEG, there is no convention or standard established for electrode placement.

There are several types of iEEG (see Figure 2.6). The first one is Electrocorticography (ECoG), which uses cortical electrodes. These electrodes are placed on the surface of the cerebral cortex in grids or strips, usually on only one of the hemispheres. However, cortical electrodes cannot reach the deep brain, so the second method is Stereotaxic EEG (sEEG). The sEEG uses depth electrodes to record information from deeper structures like the hippocampus, temporal lobes, or insula [34, 43, 44].

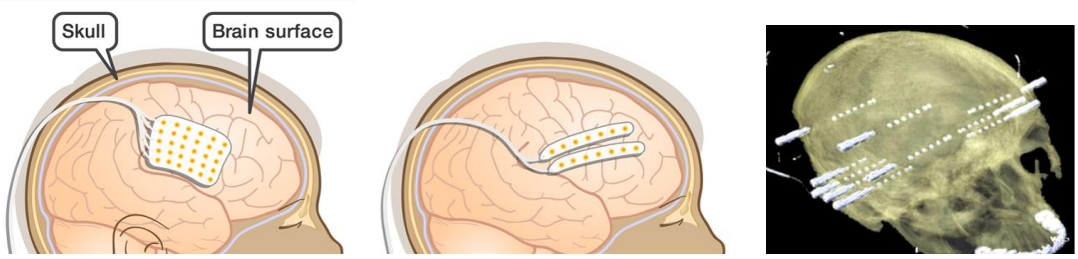


Figure 2.6: Electrode placement in different Intracranial Electroencephalogram methods: a) and b) for Electrocorticography; c) for Stereotaxic Electroencephalogram. Source: Parvizi and Kastner (2018) [43] and [45].

Consequently, the iEEG has many benefits compared to scalp EEG. It can achieve significant precision and area specificity and the ability to reach the deep brain while recording a cleaner signal with significantly fewer artifacts. However, it also has the downside of requiring surgery and thus increasing its associated risks [34, 37].

Additionally, Sub-scalp EEG (ssEEG) is a technology that has gained traction recently, where the electrodes are placed between the skull and the scalp. This technique shows signal quality improvements from scalp EEG and does not require the same electrode care as iEEG. Its minimal invasiveness allows for easy use outside of the hospital environment and facilitates ultra-long-term recordings [46].

2.2.2 EEG in Epilepsy

When it comes to epilepsy, the EEG is the most effective tool for diagnosis by outlining the epileptogenic part of the brain and analyzing the patient's seizures. Analysis of the recorded EEG signals can be used to learn more about the features of epileptic seizures and allow for the distinction between its different phases. EEG analysis can also discriminate between normal and epileptic data [11, 16].

2.2.2.1 IEDs

A characteristic aspect of epilepsy and EEG is that the brain with epilepsy produces distinctive field potentials called IEDs. These waveforms strongly contrast with background activity and comprise an essential component of the electroclinical definition of epilepsies. IEDs have previously contributed significantly to the diagnosis of epilepsy and continue to do so in the current classification of focal or generalized epilepsies. IEDs and characteristic patterns in abnormal EEGs also aid in identifying epilepsy syndromes.

One appealing aspect of IEDs is that they can be seen in up to 90% of patients in the scalp EEG when using conventional electrodes set per the international 10-20 system. However, the bulk of IEDs recorded directly at the cortical surface are not evident on scalp EEG or at least do not demonstrate their epileptic origin, according to simultaneous iEEG and scalp EEG recordings [11, 37, 39].

2.2.2.2 Epileptic Period Segmentation

The EEG of a patient with epilepsy who suffers from a seizure can be segmented into distinct periods in time (see Figure 2.7):

- Preictal: the period before a seizure;
- Ictal: the period during a seizure;
- Postictal: the period after a seizure;
- Interictal: the period between the postictal and preictal periods of two consecutive seizures.

To summarize, the EEG records the pre, post, and interictal phases when no seizure occurs and the ictal phase during a seizure.

Correctly identifying the preictal stage allows for the prediction of seizure onset. However, even though the brain's electrophysiological behavior alters when entering the preictal stage, it involves several intricate processes. The process of accurately annotating it, especially manually, is an arduous task, as the characteristics differ not only between patients but also between different seizures by the same patient [16].

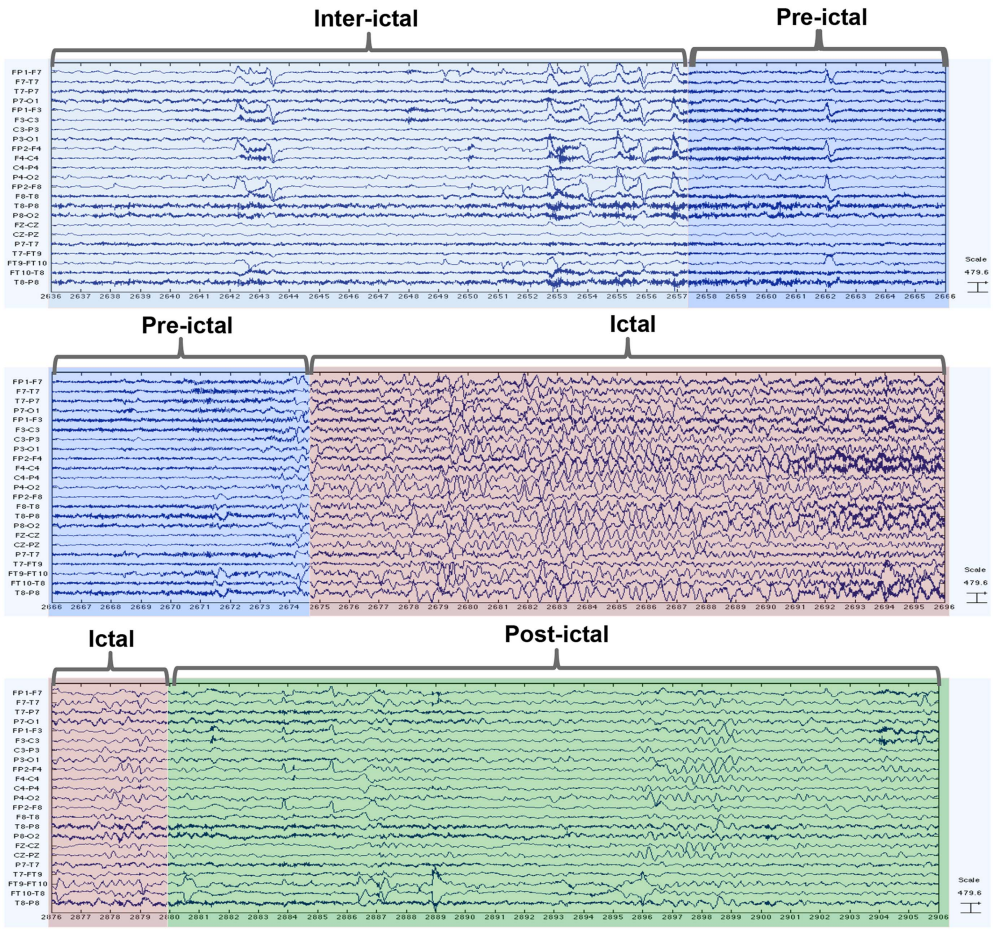


Figure 2.7: Electroencephalogram signal highlighting the stages of an epileptic seizure episode. Source: Moghim and Corne (2014) [47].

2.3 Treatment & Therapeutics

There is a substantial stigma concerning epilepsy, particularly epileptic seizures [48]. Patients with the disease may suffer from neuropsychological deficits, namely memory loss and attention difficulties. Moreover, the inability to control seizures has not only social and psychological consequences but also physical ones, as it increases accidental injury rates and the risk of brain injury, cognitive decline, and SUDEP [4–6]. To this effect, treatment for epilepsy that aims to eliminate seizures as quickly as possible, with no side effects, is of great importance [21].

2.3.1 Anti-Epileptic Drugs (AEDs)

Anti-Epileptic Drugs (AEDs) are some of the most prescribed centrally active agents, being the first option when looking at treatments for epilepsy. Since the present knowledge lies in the idea that the disease is the cause of improper balance of excitation-inhibition, several of the AEDs being used either block excitatory mechanisms or enhance inhibitory ones [2, 49]. Trials also show that if AEDs start being administered right after a patient's first epileptic seizure, it reduces the risk of a second seizure (or at least extends the interval to it) [50].

Drug-Resistant Epilepsy (DRE)

However, AEDs only help around two-thirds of patients, as the rest seem resistant to this type of medication, and in some cases, even seem to worsen the symptoms [2, 3].

Drug-Resistant Epilepsy (DRE) was defined by the ILAE Task Force in 2009 as [51] "*failure of adequate trials of two tolerated and appropriately chosen and used AED schedules (whether as monotherapies or in combination) to achieve sustained seizure freedom*" and seizure freedom as "*freedom from all types of seizures for 12 months or three times the preintervention interseizure interval, whichever is longer.*"

Patients suffering from this type of epilepsy are a subject of great concern. Starting with infants, DRE can lead to institutionalization from epileptic encephalopathies with developmental delay. In the older children to young adults age group, the disease brings challenges regarding social skills acquisition and proper integration into society. In general, the mortality rates of individuals with DRE are five to ten times higher than the general population, and there is a higher probability of developing depression, anxiety, and psychoses [6, 21]. So, there must be an alternative to deal with these types of cases.

2.3.2 Surgery

In cases where AEDs fail, surgery may be the best option to achieve seizure control. However, the effectiveness of the surgery is not universal. Certain types of epilepsies, underlying pathologies, and an inaccurate localization of the epileptogenic brain region can hinder its success. Nonetheless, particularly in cases with drug-resistant focal epilepsy, the surgery frequently leads to improved cognition, behavior changes, and a better quality of life [52].

To ensure the effectiveness of the surgery, there needs to occur presurgical monitoring to help identify the epileptogenic zone and assess the brain regions to be resected. This evaluation process rests on an approach dependent on the collaboration between different diagnostic modalities. The degree of presurgical evaluation varies from patient to patient, as the type of epilepsy and epilepsy syndromes vary. After acquiring the clinical history and performing a regular physical examination, the evaluation starts by applying video-EEG monitoring, neuroimaging, neuropsychological tests, and psychological analysis. The next step is a multimodal one, where several other exams are taken. Finally, a crucial portion of the evaluation lies in predicting and attenuating any postoperative functional impairments.

Due to the abundant availability of data during presurgical monitoring, most databases used in the study and development of prediction algorithms in epilepsy come from this context.

However, the surgery is heavily underused. There are many misconceptions regarding the procedure and an unwillingness of patients to undergo evaluation by a group of specialists at an epilepsy center, even though the mortality rate is nearly zero. This means that only a tiny percentage of suitable patients will undergo surgery, and even those who do are often too late, having already suffered irreversible psychological and social consequences [21, 52, 53].

2.3.3 Neurostimulation

Not all patients suffering from DRE are eligible for resective surgery. In these cases, neurostimulation methods can be an option. Studies on electrophysiology and the notion of epileptic networks have aided a better understanding of epilepsy.

Neurostimulation techniques can be categorized by looking at the stimulation target, which can be invasive or non-invasive. It can also vary in the method of stimulation, divided into chronic programmed (open loop) or responsive (closed loop).

Vagus Nerve Stimulation (VNS), approved in 1997 by the Food and Drug Administration (FDA), was the first approved technique for neurostimulation therapy. VNS is one of the invasive methods, requiring the implementation of a neurocybernetic prosthesis under the chest skin. Other invasive neurostimulation options include Deep Brain Stimulation (DBS) and Responsive Neurostimulation (RNS) System, the first closed-loop option. These methods and their efficacy are backed by numerous clinical trials and are good, safe options for treatment.

Regarding the non-invasive techniques, a similar option to the VNS exists called External Trigeminal Nerve Stimulation (eTNS), although not as effective. Repetitive Transcranial Magnetic Stimulation (rTMS) and Transcranial Direct Current Stimulation (tDCS) are among the non-invasive options. However, non-invasive methods still lack evidence of efficacy and clinical trials [54, 55].

2.3.4 Rescue Medication

Rescue medication is a fundamental tool for the control of epilepsy. It has been shown to lessen the likelihood of recurrent seizures and diminish the length of prolonged seizures [56]. During seizure clusters or prolonged seizures ($\geq$ 5 minutes), rescue medication consists of a benzodiazepine intake. As these episodes will likely occur outside of the hospital environment, there must be methods of administration that are not intravenous. Several studies show that non-intravenous delivery routes are quicker than intravenous ones [30, 32]. The high number of children suffering from these emergencies further supports the need for easy access to these procedures in schools [30].

The most frequent method is a rectal diazepam gel. However, there is the downside of requiring a partially undressed patient and the necessity of a private space to administer it [29]. An article published in 2021 showed some alternatives for the rectal diazepam gel that the US Food and Drug Administration has already approved. These methods include intranasal diazepam, intranasal midazolam and intramuscular midazolam. However, the previously mentioned are not without downsides. Intranasal methods have physiological dependencies on the bioavailability and absorption of the drug. Intramuscular options tend to hurt and require expertise for the injection. Other routes still in development involve buccal, intrapulmonary, and sublingual [32]. Nevertheless, recent evidence shows that intranasal diazepam and midazolam are as similarly effective and secure in treating seizure clusters as the rectal gel [57].

2.3.5 Warning Devices

For patients that AEDs and epilepsy surgery are not viable options for seizure control, there must be a way to try and improve the quality of life. Uncontrolled seizures can cause seizure--related injuries and social discomfort, not to mention an increased risk of SUDEP. So, a way to monitor biosignals and predict seizures can give the patient enough time to either minimize consequences or use rescue medication. Warning devices use algorithms for automatic seizure prediction or forecasting and could help eliminate the anxiety caused by the unpredictability of epilepsy in everyday life. These methods can use different biosignals. Examples include the analysis of IEDs on EEG, the detection of motor responses in seizures using accelerometry, and the evaluation of alterations in physiological parameters with Electrocardiography (ECG), respiratory monitors, or pulse oximetry [7].

A study from 2016 with 141 patients from epilepsy centers in Freiburg, Germany, and Coimbra, Portugal, showed that the demand for this sort of system is high. The study also showed that patients wanted a high-sensitivity prediction algorithm while discarding specificity and that short prediction windows were enough to prevent social mishaps. It also found that the current algorithms for this type of use do not yet meet these patients' requirements, and many were unwilling to use or participate in studies for ambulatory devices [58]. Furthermore, a study by Bruno et al. (2018) [59] surveyed 87 patients with epilepsy, caregivers, and healthcare professionals. The study aimed to assess the current use of digital technology and the willingness to use this option. Results show that 80% of patients were receptive to using a wearable device to track seizures. Caregivers believed that the use of wearable devices could be a way to reduce their workload, and healthcare professionals responded that these devices had the potential to help with patient management, specifically for diagnosis and treatment.

A promising warning device for seizure prediction has been developed. The NeuroVista Seizure Advisory System was implemented in 15 patients for two years as a phase I clinical trial of the device. The patients were selected based on seizure frequency and level of independence. Some patients had previously tried options such as resective surgery or neurostimulation, specifically VNS. Using iEEG, the system could register the signal and determine a certain probability of a seizure. This trial showed that 11 subjects performed well, generating a sensitivity higher than 65% and performance more significant than randomly predicting events. It is the first human study to successfully demonstrate prospective seizure prediction using ambulatory EEG recordings. However, due to the results not being sufficiently good and the belief that the technology was too invasive, the funding was cut [60, 61].

2.4 Seizure Forecasting and Prediction

The primary goal of seizure prediction is to create an algorithm capable of anticipating seizures and triggering a warning. It strays from seizure detection, which aims to detect a seizure as it occurs. An alarm should provide the patient with a sufficiently large and well--defined time window to allow for preventive measures. Additionally, the number of false alarms should be taken to an absolute minimum [5, 48].

Seizure forecasting has a different objective. Instead of alarms triggered when the preictal state is detected, the goal is a continuous assessment of the risk of an epileptic seizure. Even though it does not guarantee its occurrence, when the algorithm identifies a brain state as high-risk, a seizure is likely to follow in a reasonable and well-defined time window [10].

With ways to capture and use data from the patient, mainly EEG, the objective of seizure forecasting and prediction algorithms is to detect early signs of the preictal stage. This way, patients can take steps to prevent the seizure from happening, be prepared for when it comes, or at least reduce the stress and anxiety that comes from its unpredictability [12, 62].

2.4.1 Seizure Onset

The seizure onset marks the start of an epileptic seizure. There are two types of onsets: the EEG (or electrographic) onset, which refers to the moment when the first changes in the EEG signal are noticeable, and the clinical onset, which refers to the moment when the first symptoms appear. As the EEG onset precedes the clinical one, algorithms for seizure prediction tend to consider the EEG onset [13, 37].

2.4.2 Lead Seizure

A significant number of seizures is required for trustworthy results. This analysis also takes into account seizure clusters, discussed in Section 2.1. Prediction and forecasting models consider each seizure an independent event. So, there exists the consensus of using only the first seizure of the cluster. This seizure, preceded by a period of normality, is called a "lead seizure" [31]. Again, as there is no set agreement on the definition of seizure cluster and this minimum period to set them apart, these vary from study to study [63]. Studies consider that the minimum time separating lead seizures can range from 1 hour [64] up to 8 hours [60].

2.4.3 Detection vs Prediction

Seizure detection is separate from seizure prediction (see Figure 2.8). Seizure detection is a similar field of research, focusing on detecting the EEG onset before the patient develops symptoms (clinical onset). Additionally, it must be capable of detecting whether seizures are presently occurring or not [13, 65].

Returning to a study mentioned in Section 2.3 [58], over 90% of participants were greatly interested in a way to predict an epileptic seizure. However, prediction methods are less reliable than detection methods, with lower values of Sensitivity (SS) and Specificity (SP). On the other hand, unless there is an immediate intervention to prevent a seizure from happening after its EEG onset, seizure detection is of little practical use for the patient. Nevertheless, it can offer comprehensive seizure data to clinicians who manage epilepsy [12, 13, 65, 66].

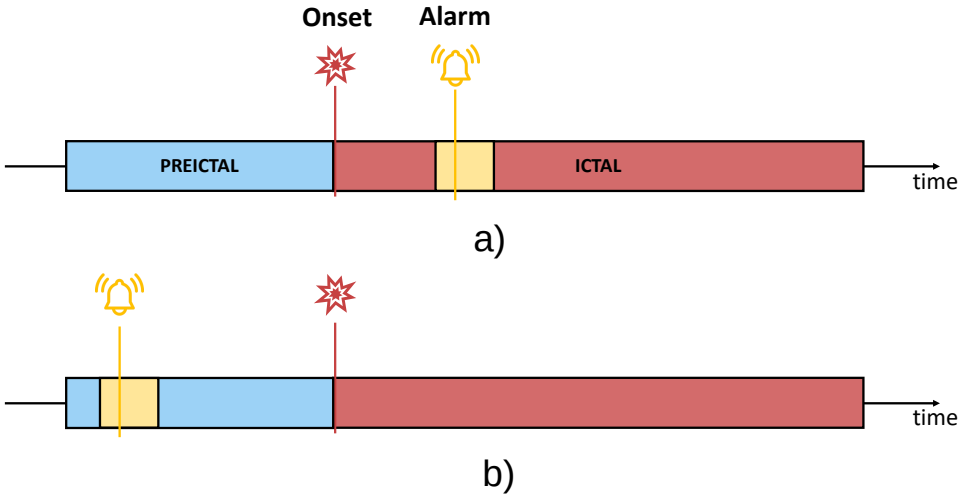


Figure 2.8: Comparison between seizure detection a) and seizure prediction b).

2.4.4 Forecasting vs Prediction

Seizure forecasting is another parallel area to seizure prediction that has been gaining traction. Seizure forecasting moves away from categorical seizure prediction assessments and instead focuses on identifying the brain state by computing the possibility of a seizure occurring. In other words, whereas in seizure prediction an alarm would be given before the ictal period, seizure forecasting shows the probability of one occurring (see Figure 2.9). This option can provide not only patients but also caregivers with more control over everyday life. It can also help with treatment, as an evaluation of the likelihood of a seizure can impact its type and potentially reduce side effects and costs [9, 67].

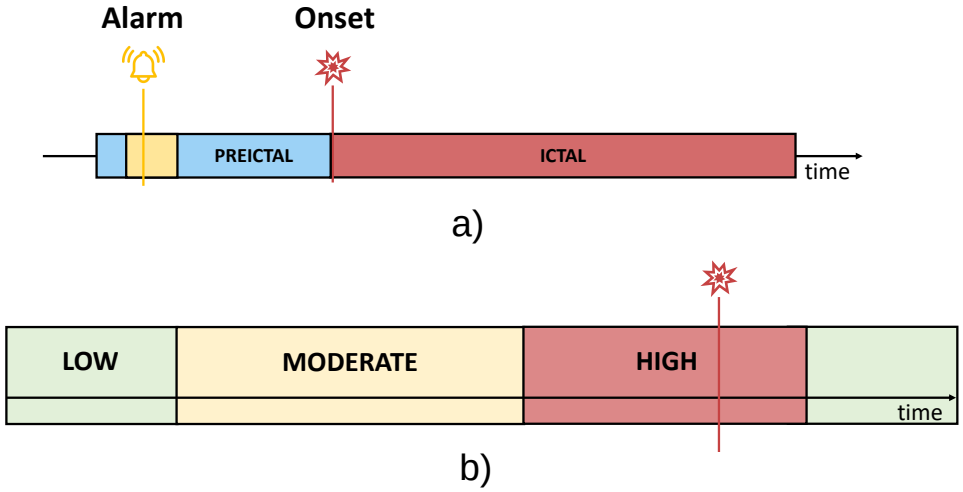


Figure 2.9: Comparison between seizure prediction a) and seizure forecasting b) as done in the NeuroVista [60] clinical trial.

There are times when a seizure may be about to occur, but the brain's natural self-regulation mechanisms may stop it before it fully develops. These mechanisms may be somewhat impaired in individuals with epilepsy, allowing for some seizures to occur. However, they still may work to a certain degree. Seizure self-termination may also occur in seizures that have not fully developed yet. A prediction algorithm could interpret these signals as a preictal state and raise an alarm, even though no seizure will follow [8].

Despite still being a seizure prediction study, the NeuroVista clinical trial [60] discussed in Section 2.3 as a warning device constituted the first transitional steps to seizure forecasting. Using the continuous EEG recordings obtained through iEEG, the system told the user if the seizure likelihood was Low, Moderate, High, or Uncertain. Ever since this trial, knowledge about seizure patterns has increased. Due to the possibility of acquiring the EEG signal for an extended period, the ability to provide tailored forecasts from patients' seizure cycles is now a topic of great interest. These seizure cycles may follow circadian, multiday, or seasonal patterns. Continuous EEG recordings using fast and slow brain activity cycles provide the best estimate of seizure likelihood as of today.

Providing the occurrence probability of a seizure is more achievable than predicting the following one, and user surveys show that this type of technology is desirable. However, most algorithms' dataset is relatively small. Additionally, this data is mainly collected through iEEG, so the invasive nature of the process might make it undesirable for the users [9, 68]. For this reason, studies using different methods for signal acquisition have started to gain traction, particularly ssEEG [46].

2.4.5 Seizure Prediction Characteristic

Up to the early 2000s, there was substantial growth in seizure prediction and the development of different methods. However, there needed to be a recognized way to evaluate the performance of these methods.

Consequently, in 2003, a publication by Winterhalder et al. [19] tried to solve this problem by taking clinical, behavioral, and statistical considerations to propose the "seizure prediction characteristic." They introduced two terms: Seizure Occurrence Period (SOP) and Seizure Prediction Horizon (SPH).

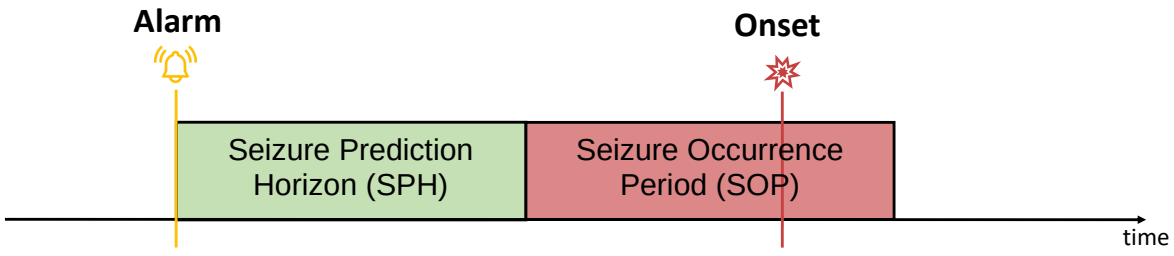


Figure 2.10: Visual representation of the definition of Seizure Prediction Horizon and Seizure Occurrence Period. Adapted: Winterhalder et al. (2003) [19].

As shown in Figure 2.10, the SOP is the period during which it is expected for the seizure to occur. However, if any therapeutic intervention or behavioral change is expected, there must be a minimum time window between the alarm raised and the time when the seizure will occur. This window defines the SPH, also known as Intervention Time (IT). For a prediction to be correct, the seizure has to occur during the SOP (see Figure 2.11).

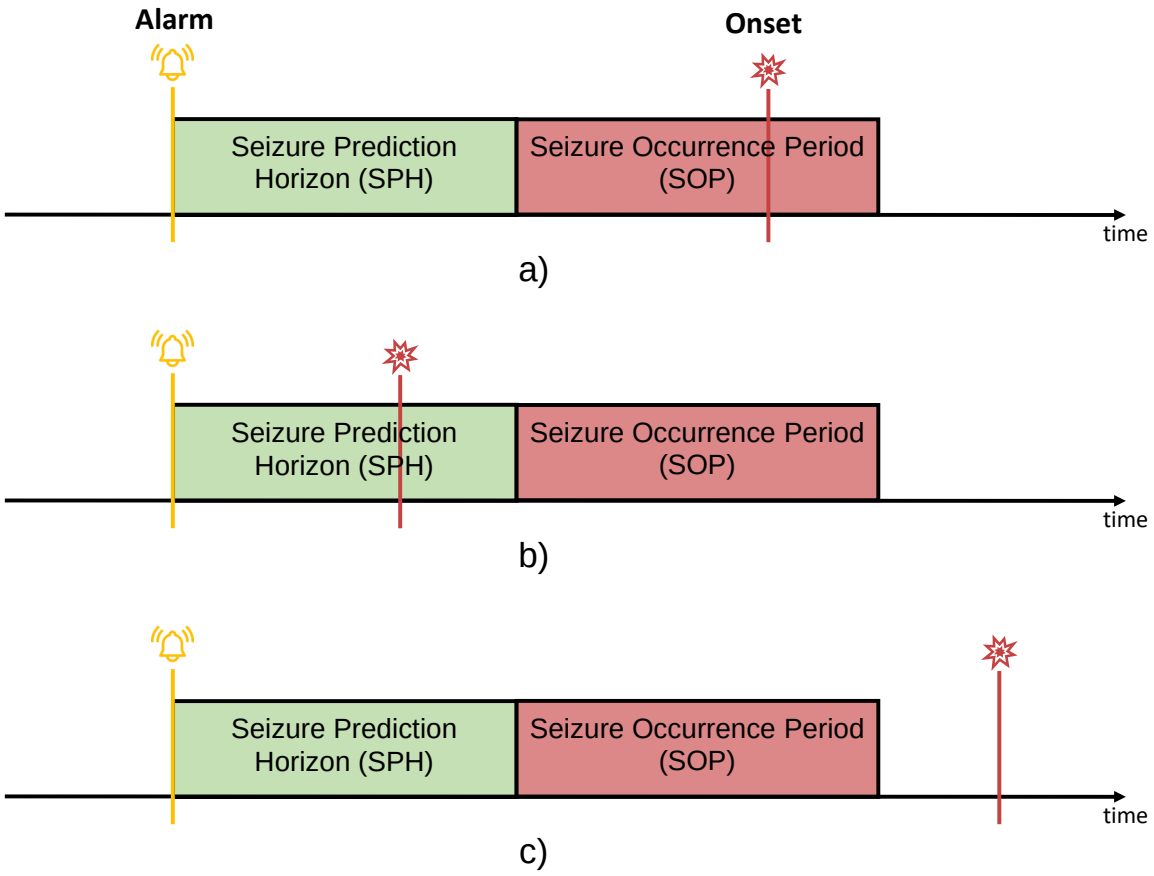


Figure 2.11: Comparison between a true alarm a) and false alarms b) and c) with the Seizure Prediction Horizon and Seizure Occurrence Period concepts in mind.

The current literature has yet to agree on the optimal SOP values. These can range anywhere from seconds to minutes to even hours. Nonetheless, the SPH must have a minimum time to allow for intervention. Additionally, as there is no exact time when a seizure may occur, a maximum limit for the SOP should also be set. SOPs that are too long may cause psychological stress for the patient.

Additionally, long SOPs may not be helpful for the patient. For example, in a patient with five seizures per day, having a SOP of six hours is essentially useless. Even though it anticipates seizures, it does not predict any of them. Summarily, a prediction method requires a SPH that is big enough to allow for intervention and a SOP with a duration that allows the correct prediction of seizures without causing additional stress and anxiety for the patient [13, 19, 69].

2.4.6 Postprocessing

The classifiers of seizure prediction or forecasting algorithms are trained to make classifications on independent EEG segments. So, a usual step after classification is postprocessing the classifiers' output to smooth it out, specifically a regularization function. Two methods that can be deployed include the Kalman Filter and the Firing Power [12].

The Firing Power is a technique that uses a sliding window analysis to quantify the rate of preictal samples. It is a way to smooth the classifier's output over time, sounding an alarm if it exceeds a normalized threshold. There is no optimal value for the threshold. The Kalman filtering achieves the estimation of the states of a linear dynamic system tending close to the actual measurements. If this method's output is classified as a preictal state, it raises an alarm [12]. A 2012 study by Teixeira et al. [70] found that despite the Kalman filtering achieving better SS values, the number of false alarms was large enough for the model to be considered clinically impractical. Therefore, the Firing Power is a more conservative and superior method due to the ability to create time constraints and maintain a more extended memory of the classification dynamics.

2.4.7 Performance Assessment

For the evaluation of the performance itself, several metrics are analyzed. These metrics vary between seizure prediction and seizure forecasting. In this subsection, performance metrics and options for statistical validation will be discussed.

2.4.7.1 Prediction

Returning to Winterhalder et al. (2003) [19], this study defined two performance metrics: SS and False Positive Rate per Hour (FPR/h).

The SS depicts the ratio between the correctly predicted seizures (true alarms) and the total number of seizures, as shown in Equation 2.1:

$$SS = \frac{Number\ of\ true\ alarms}{Number\ of\ seizures}. \tag{2.1}$$

The FPR/h represents the number of seizures incorrectly predicted in one hour. This metric illustrates the ratio between the number of false alarms and the interval during which the model can raise a false alarm, the interictal period. However, when an alarm is fired, it is also associated with a SPH and a SOP, during which no other alarms can be raised. So this period, called the refractory period and composed of the sum of the SPH and SOP, occurring every time an alarm is triggered, must be subtracted from the total interictal time, as Equation 2.2 shows:

$$FPR/h = \frac{N_F}{\triangle_I - N_F * (\triangle_{SPH} + \triangle_{SOP})}, \tag{2.2}$$

where N_F is the number of false alarms, $\triangle_I$ the interictal length, $\triangle_{SPH}$ the SPH length, and $\triangle_{SOP}$ the SOP length.

The goal for any prediction model is for all the seizures to be predicted and no false alarms to be raised (SS of one and a null FPR/h). However, achieving this goal is difficult. There may be a trade-off relationship between these two metrics. A model with a high FPR/h used in a warning device may cause patients not to take alarms seriously, as many of the ones triggered amount to no seizure. Patients who take every alarm seriously will be subjected to significant stress and anxiety levels.

Consequently, a maximum false prediction rate (FPR/h_{max}) should be set. Winterhalder et al. (2003) [19] used the average seizure incidence in the context of presurgical monitoring to propose an FPR/h_{max} with a value of 0.15 seizures per hour or 3.6 seizures per day. It should be noted that these values are much higher than usual, as these patients suffered a reduction of AEDs. Taking the example of patients with pharmacorefractory focal epilepsy, with around three seizures per month (0.0042 per hour), the FPR/h_{max} should be much lower.

It is also possible to consider these performance metrics in the scope of a standard binary Machine Learning (ML) problem. Here, the samples are classified into two classes: the interictal (0) and the preictal (1) period. When the training of the prediction algorithm finishes, the resulting model is applied to the test data. The confusion matrix in 2.12 shows the relationship between the model output and the actual data, according to each sample [19, 37].

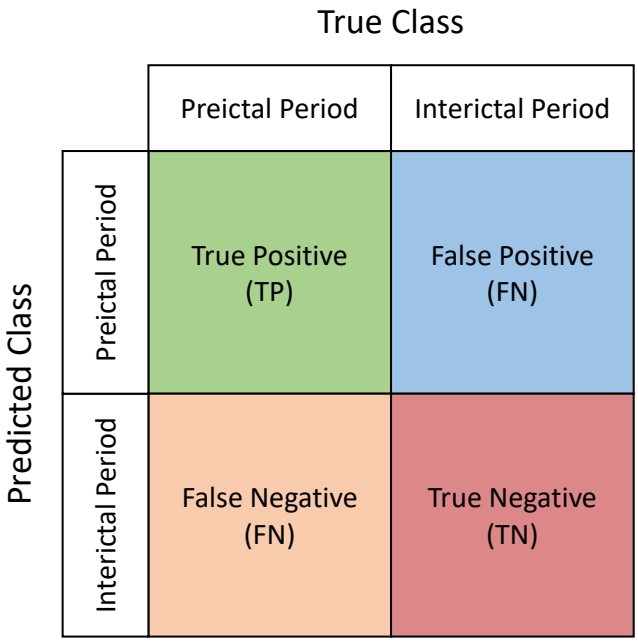


Figure 2.12: Confusion matrix for seizure prediction performance evaluation.

The values from the confusion matrix can be used to calculate the Sample SS (Equation 2.3) and the Sample SP (Equation 2.4).

$$SS_{Sample} = \frac{TP}{TP + FN} \tag{2.3}$$

$$SP_{Sample} = \frac{TN}{TN + FP} \tag{2.4}$$

2.4.7.2 Forecasting

Seizure forecasting has different performance metrics: the SS, the Time In Warning (TiW), the Brier Score (BS), the Brier Skill Score (BSS), and the Reliability Curves (RCs).

While in seizure prediction the SS accounts for the alarms correctly raised during the preictal period, in forecasting, it is merely necessary for the algorithm to be in a high warning state during part of this period. So, considering an algorithm with an output such as in Figure 2.9 b), the SS represents the proportion of seizures that occur during a high-risk state.

The TiW, also known as Time Under Warning (TUW), is what its name suggests: TiW is the time spent in a warning. More specifically, it is the time the model spends in a high-risk state divided by the total time [17, 71, 72].

In seizure forecasting, the mathematical models output a probabilistic value (seizure risk) between 0 and 1. The BS and BSS metrics consider this. Starting with the BS, this metric evaluates the performance of probabilistic forecasting methods by measuring the difference between these continuous forecasts and the observed seizure rates. It assesses the degree of success in matching different forecast probabilities to its observed probabilities of suffering from a seizure. Good forecasts tend to 0, while bad forecasts tend to 1. Essentially, as Equation 2.5 shows, the BS calculated by the mean squared error between the forecast and the observation:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2, \tag{2.5}$$

where f_i is the forecast probability, o_i is an observation (0 represents the absence and 1 the occurrence of a seizure), and N is the number of forecast time points [10, 17]. However, Karoly et al. (2017) [17] take a different approach by considering the uncertainty (baseline rate of seizures) and the resolution (average predictive power above the baseline rate). Equation 2.6 presents this new way of calculating the BS:

$$\begin{aligned} BS &= Reliability - Resolution + Uncertainty \\ &= \frac{1}{N}\sum_{i=1}^{N_f} n_i(f_i - b_i)^2 - \frac{1}{N}\sum_{i=1}^{N_f}(b_i - b)^2 + b(1-b), \end{aligned} \tag{2.6}$$

where N is the number of forecasts, N_f is the number of forecast bins, b is the seizure baseline rate, b_i is the rate of seizure occurrence when the forecast is in the bin in order i, f_i is the average forecast, and n_i is the number of forecasts per corresponding bin.

The BSS is defined as the improvement of the calculated BS over an uninformed reference forecast BS_{ref}, as shown in Equation 2.7:

$$BSS = 1 - \frac{BS}{BS_{ref}}. \tag{2.7}$$

For seizure forecasting, Baud et al. (2022) [10] suggest that an unskilled reference such as random chance, shuffled forecasts, or uninformative forecasts may be the option for the BS_{ref}. Certain studies [17, 73] utilize a randomly shuffled forecast. These surrogate forecasts randomly draw probabilities from the same distribution classifier's output, and a BS_{ref} is calculated by the mean of the BSs from all surrogates.

If the BSS approximates 1, the forecasts are successful and can truly show the probability of seizures. On the other hand, if the forecasts do not prove to be better than the uninformed reference, the BSS will tend to 0. If the forecasts are worse than the reference, the BSS will be negative.

The final metric is RCs. They serve as a compelling visualization method to represent the components of the BS (see Figure 2.13). It displays a graph comparing the forecast seizure rate to the actual seizure rate, which is determined based on the frequency of seizures occurring in the preictal stage after each forecast. This graph is made using the quantization in probability bins made in calculating the BS. In practical terms, for instance, it searches for all the samples that show a 0.8 likelihood of a seizure and compares them to the actual samples. An ideal classifier would match 80% of these to samples in the SOP. The optimal RC is a diagonal line indicating that the forecast probabilities match actual results [17, 37, 74].

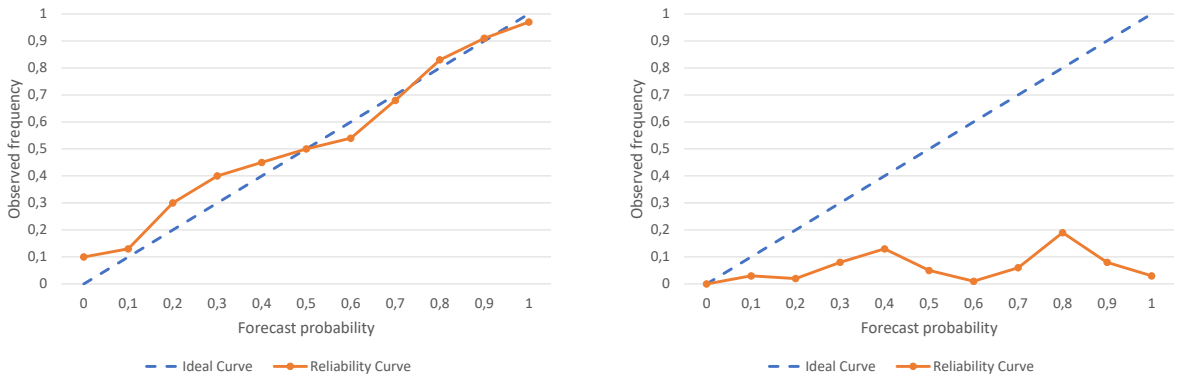


Figure 2.13: Illustration of Reliability Curves (RCs), where a) is a good RC and b) is a bad RC.

2.4.7.3 Statistical Validation

Another fundamental aspect to take into account when dealing with seizure prediction and forecasting is statistical validation. To ensure that any algorithm performs above chance-level, it must outperform one based on chance [13]. The goal is to quantify the percentage of patients that show Improvement over Chance (IoC) [73]. This is identical in seizure prediction and forecasting. This topic will address two of the most widely used techniques: unspecific predictors and surrogate analysis.

Unspecific Predictors

Winterhalder et al. (2003) [19] proposed the Random Prediction Method, where alarms are triggered randomly without the use of any information from the EEG signal. Equation 2.8 calculates the probability p of an alarm in a small interictal interval I:

$$p = FPR/h \cdot I. \tag{2.8}$$

Taking into account a more extensive time interval W, Equation 2.9 calculates the probability P of at least one alarm:

$$P = 1 - (1 - FPR/h \cdot I)^{W/I} \approx 1 - e^{-FPR/h \cdot W} for\ I \ll W. \tag{2.9}$$

For $W = SOP$, this depicts the sensitivity of a random prediction approach, as it signifies the chance of a single alarm happening within the SOP.

Later, Schelter et al. (2008) [69] proposed the analytic random predictor based on a homogeneous Poisson process for false predictions. So Equation 2.10 shows the probability of raising an alarm at any single sampling point of a feature extracted from a time series.

$$P_{Poiss} = \frac{FP}{N}, \tag{2.10}$$

where FP is the number of false predictions, and N is the number of samples. Considering a period of the same duration as the SOP and the product between the FPR/h and the SOP being much smaller than one (making sure the patient is not subjected to a continuous warning), Equation 2.11 can be used. This is a way to calculate the probability P that at least one alarm fires within the SOP:

$$P \approx 1 - e^{-FPR/h \cdot SOP} \approx FPR/h \cdot SOP. \tag{2.11}$$

This probability P sets the foundations for a significance level α to assess if the sensitivity S(FPR/h, SOP, SPH) outperforms that of a random predictor. Equation 2.12 calculates the probability of randomly predicting a minimum of k seizures out of K seizures following a binomial distribution:

$$P_{binom}(k,K,P) = 1 - \left[\sum_{j \leq k} \binom{K}{j} P^j (1-P)^{K-j} \right]. \tag{2.12}$$

Finally, Equation 2.13 calculates the critical value used for testing the statistical significance alpha for a certain significance level:

$$\sigma = \frac{argmax_k \{P_{binom}(k,K,P) > \alpha\}}{K} \cdot 100\%. \tag{2.13}$$

The essential advantage of the random predictor is its simple mathematical expression that does not require the EEG signal, making it computationally efficient. It also gives insight into the minimum number of seizures required in the data to guarantee that the performance surpasses the chance level. However, it assumes a uniform distribution of false alarms over time as it is based on a homogeneous Poisson process, which may not effectively address particular seizure dynamics.

Surrogate Analysis

Schelter et al. (2008) [69] also discussed seizure-predictor surrogates, an alternative to the analytic random predictor. Based on Monte Carlo simulations, these rest on constrained randomizations of the proposed prediction/forecasting model that shares specific properties with them. This method offers more versatility than analytical random predictors as it enables the testing of different null hypotheses by combining suitable sets of assumptions and limitations, as any particular assumption can be transformed into a relevant randomization constraint. Finally, the model's performance is compared to the surrogate one.

The algorithm is considered better than chance if it is higher with statistical significance, by rejecting the null hypothesis that the proposed model does not outperform the surrogate one. Figure 2.14 shows one possible strategy.

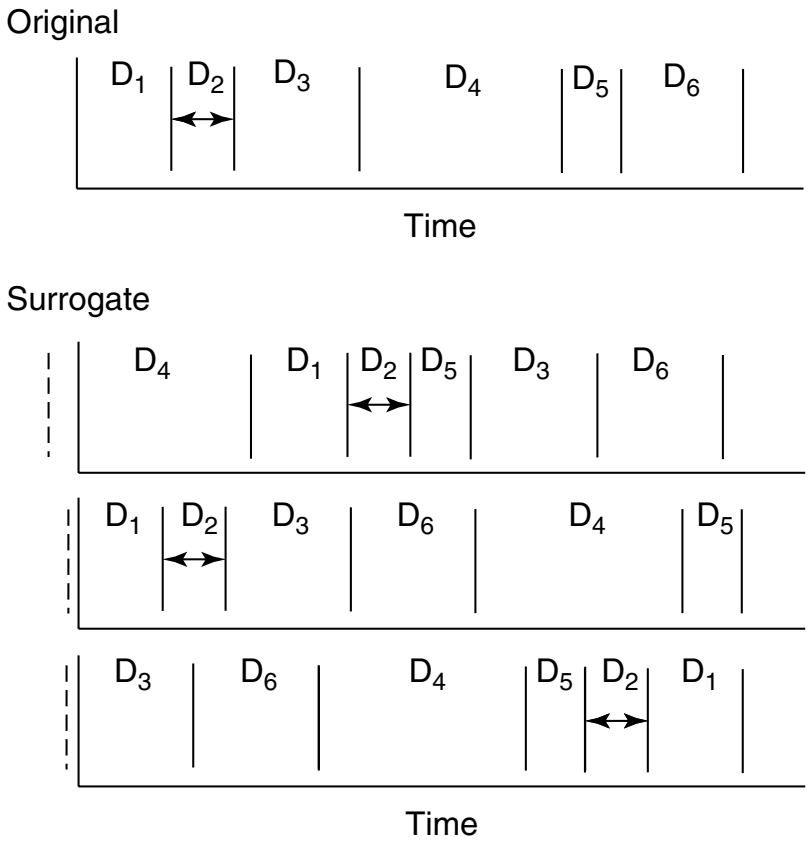


Figure 2.14: Original seizure times and the surrogate times bootstrapped from the inter-seizure intervals. The arbitrary onset times for the surrogates are obtained from a uniform distribution and are indicated by the dashed vertical lines. Source: Schelter et al. (2008) [69].

2.5 Concept Drifts

A vital problem that seizure prediction and forecasting algorithms face is the existence of Concept Drifts (CDs). In real-world cases, the concept of interest may depend on some hidden context that is not given explicitly in the form of predictive features. In other words, some properties of the target variable the algorithm wants to forecast may change unexpectedly over time, consequently influencing the features [18, 75].

The EPILEPSIAE database, like many others, contains EEG data obtained during presurgical monitoring. In this context, due to time constraints, patients undergo sleep deprivation and are subjected to a reduction in AEDs. This results in an uncharacteristically high number of seizures. Under normal circumstances, there is a frequency of around three per month, while in these cases, it can increase to three per day. Additionally, seizures in this high-stress situation may not represent typical events, and reducing the medication may cause measurable changes in the EEG. As most algorithms use these types of databases, the real challenge is ensuring that they still work when using data gathered under realistic conditions [15, 19, 20].

Additionally, advances in diagnostic technology showed daily rhythmic patterns of seizure occurrence and epileptic activity. As defined by Khan et al. (2018) [76] rhythms can be:

- Circadian: "*A biological rhythm is considered to be a circadian rhythm if it meets three criteria: the rhythm should have an endogenous free-running (approximately) 24 h period, should be entrainable (ie, be capable of phase reset by environmental cues and synchronization to the 24 h day), and should exhibit temperature compensation*";
- Diurnal: "*A biological rhythm that is synchronized with the day–night cycle. A diurnal rhythm may or may not be a circadian rhythm*";
- Multidien: "*Refers to rhythms with a time period covering several days*";
- Ultradian: "*Refers to rhythms with periods of less than 24 h; ultradian rhythm cycles can occur with a frequency of more than once per day.*" This includes the non-REM cycle, lasting around 90 minutes.

2.6 Summary

Epilepsy is a disease with substantial clinical heterogeneity. There are several seizure types, depending on the zone of the brain where the seizure onset occurs, and several types of epilepsy, depending on the type of seizures. Furthermore, there are different types of epilepsy syndromes. TLE is the most common, characterized by seizures with temporal lobe focus. The deployment of rescue medication can help control cases of seizure clusters. There exist many other therapeutic and treatment options developed to control seizure activity. The first choice is the administration of AEDs, effective for approximately two-thirds of patients. However, new options must arise for patients suffering from DRE. Resective surgeries are an effective and relatively safe (despite being invasive, the fatality rate is close to zero) alternative for DRE patients. Nevertheless, not all patients are eligible for surgery, so another possibility is neurostimulation.

The brain's electrical activity can be recorded by the EEG, one of the most effective tools for identifying, examining, and interpreting brain disorders. The EEG can capture two different types of phenomena: oscillations (rhythmic fluctuations) and transients (sharp changes). There are two methods for signal acquisition: scalp EEG and iEEG. The iEEG can achieve considerable precision and area specificity, and the ability to reach the deep brain results in a cleaner signal. However, as it is invasive, it brings the risks associated with these sorts of procedures. The EEG can also capture distinctive field potentials produced by the brain with epilepsy called IEDs. This signal can also be segmented into different periods in time: interictal, preictal, ictal, and postictal.

The periodic division of the epileptic EEG is the foundation for seizure prediction, allowing for the development of warning devices. A warning device running on a seizure prediction algorithm should be able to anticipate a seizure by triggering a warning. This warning should be associated with a period containing the onset (SOP) and a time for intervention (SPH). The performance of a seizure prediction algorithm is determined by assessing seizure SS and FPR/h. However, current models do not show promising results. This is where a shift toward seizure forecasting can be advantageous. Shown to be more achievable than seizure prediction, it works by providing constant probabilities of a seizure occurring. Performance evaluations on these models use different metrics: SS, TiW, BS, BSS, and RCs. They also require statistical validation to ensure that they perform above chance. One aspect to be considered when dealing with forecasting algorithms intended for real-world application settings is the existence of CDs. The most common CDs are related to presurgical monitoring and consist of medication tapering, the sleep-wake cycle, and the circadian cycle.

3

State-of-the-Art

This chapter presents an overview of the current state-of-the-art in seizure prediction and forecasting based on the Electroencephalogram (EEG) signal and Machine Learning (ML). Firstly, Section 3.1 will examine the most common framework. Next, Sections 3.2 to 3.5 will describe the commonly used techniques and features. Furthermore, Sections 3.6 and 3.7 will discuss classification and methods for performance evaluation, respectively. Finally, Section 3.8 will summarize the principal concepts and supply final reflections.

3.1 Overview

Most current seizure prediction and forecasting algorithms follow a common framework consisting of signal acquisition, preprocessing, feature extraction, feature selection, classification, regularization, and performance evaluation, as Figure 3.1 shows.

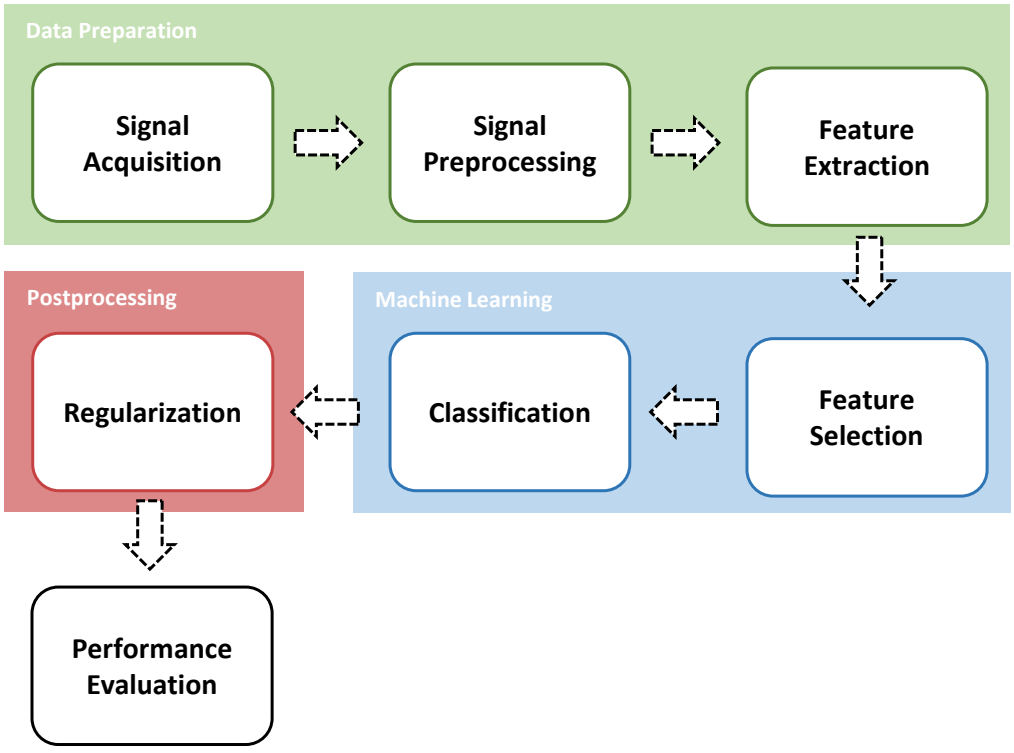


Figure 3.1: Flowchart of the typical pipeline for seizure prediction and forecasting. Adapted: Assi et al. (2017) [12].

Following the collection of EEG recordings, these steps are concisely described as follows:

- Preprocessing: preparing the signal for feature extraction by enhancing the quality through filtering and denoising, followed by segmentation of the data with sliding window analysis;
- Feature extraction: using EEG signal to collect features, individual independent mathematical variables that characterize the signal;
- Feature selection: identifying the most discriminative features to distinguish different epileptic states;
- Classification: training ML models based on the previously selected features;
- Regularization: applying postprocessing methods to smooth the output of the classifier;
- Performance evaluation: testing the performance of the model using appropriate metrics.

However, despite this being the most commonly used pipeline, there has yet to be an agreed-upon standard for seizure prediction and forecasting frameworks. Accordingly, existing approaches encompass a wide variety of possibilities [12].

Differences in Deep Learning approaches

Increasing computational power and quantity of available data during the past years caused the emergence of Deep Learning (DL) models, a more sophisticated version of ML models. The use of DL has advanced to the state-of-the-art of seizure prediction and forecasting.

These models are able to handle raw data with little to no preprocessing and are capable of feature engineering and classification. Therefore, deep learning can move past the need to know the data's structure and detect specific brain signal patterns rather than features. Additionally, due to the size of the data, feature extraction is seen as a form of dimensionality reduction. However, DL's advanced technology allows for bypassing this stage [15, 77–79].

Regardless, there are many challenges with using DL models in a healthcare context. Regarding data volume and quality, DL models prosper with large amounts of data. Nevertheless, comprehensive medical data is often limited and frequently characterized by its heterogeneity, ambiguity, noise, and incompleteness. Moreover, DL models are often treated as black boxes, lacking transparency in their decision-making processes. In healthcare, interpretability is crucial for understanding the reasoning behind the model's predictions and gaining the trust of medical professionals. Providing explanations and insights into the factors influencing predictions is vital for guiding medical actions and decision-making [80].

The easiest option (A in Figure 3.2) is to provide the raw data, allow the models to tend to the preprocessing and feature engineering steps, and supply the classifier's output [81]. Nevertheless, there are other alternatives. For instance, Stirling et al. (2021) [64] (B in Figure 3.2) only applied the DL model after processing the signal. Moreover, it is also possible to use these models only for the feature engineering step, acquiring features and using them on a different classifier (C in Figure 3.2) [82, 83].

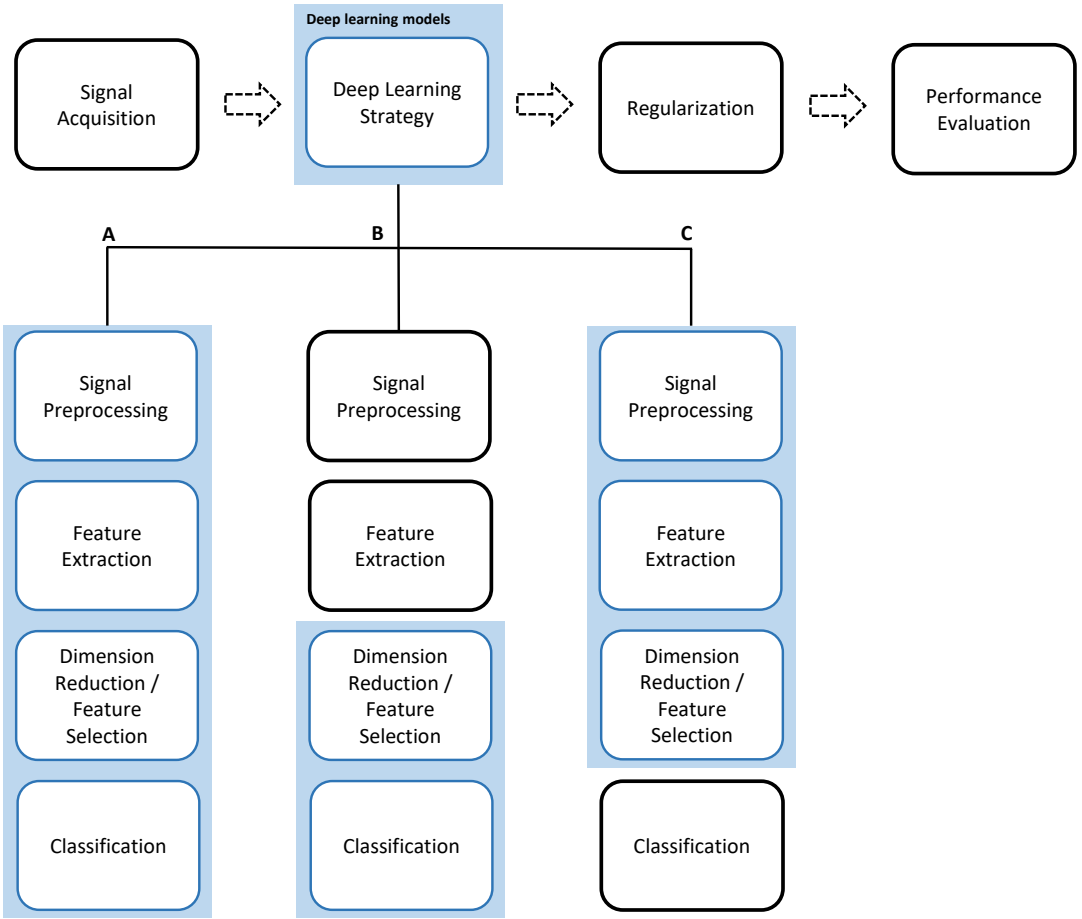


Figure 3.2: Flowchart of the variations on the seizure prediction and forecasting pipeline using Deep Learning models. Adapted: Assi et al. (2017) [12].

3.2 Signal Acquisition

The database chosen and the signal type greatly influence the results of any prediction and forecasting algorithm. Table 3.1 shows an overview of the data used in this field for the past 11 years. It begins with Cook et al. (2013) [60], considered a study of substantial importance as it was the first successful human clinical trial of a seizure prediction algorithm implemented in a wearable warning device.

Table 3.1: Overview of the signal acquisition from seizure prediction and forecasting over the past 11 years.

Study	Database	Patients	No. of Seizures (analyzed time)	Signal	Electrodes
Pinto et al. (2023) [84]	EPILEPSIAE	40	224 (135.6 days)	Scalp EEG	-
Lopes et al. (2023) [85]	EPILEPSIAE	41	227 (233.3 days)	Scalp EEG	-
Xu et al. (2023) [86]	CHB-MIT	4	27 (-)	Scalp EEG	-
Hu et al. (2023) [87]	CHB-MIT	22	198 (26.8 days)	Scalp EEG	-
Li et al. (2023) [88]	Kaggle (AES) CHB-MIT	4 dogs+18	41+90 (-)	iEEG Scalp EEG	-
Viana et al. (2022) [89]	ZUH KCL's clinical trial	6	82 (594 days)	Sub-scalp EEG	-

Table 3.1 – *Continued from previous page*

Study	Database	Patients	No. of Seizures (analyzed time)	Signal	Electrodes
Pal Attia et al. (2022) [90]	ZUH KCL's clinical trial	6	- (409 days)	Sub-scalp EEG	-
Zhang et al. (2022) [91]	SeizeIT1 SeizeIT2	42+39	182+67 (168.1+133.3 days)	Scalp EEG BTE EEG	-
Pinto et al. (2022) [92]	EPILEPSIAE	93	238t (153.6t days)	Scalp EEG	-
Proix et al. (2021) [73]	NeuroPace	18	- (> 6 months p.p.)	iEEG	-
Stirling et al. (2021a) [9]	Personal	1	134 (6 months)	Sub-scalp EEG	-
Usman et al. (2021) [83]	CHB-MIT	22	198 (26.8 days)	Scalp EEG	-
Stirling et al. (2021b) [64]	Personal	11	1493 (13.5 years)	BVP Sleep stages	Smartwatch
Nasseri et al. (2021) [93]	NeuroPace	6	278 (4 years)	ACC BVP, EDA TEMP	Wristband
Vandecasteele et al. (2021) [94]	SeizeIT1 EPILEPSIAE	42+93	221+675 (-)	Scalp EEG BTE EEG ECG	-
Pinto et al. (2021) [95]	EPILEPSIAE	19	49t (29.6t days)	Scalp EEG	-
Xu et al. (2020) [81]	Kaggle (AES) CHB-MIT	5 dogs+22	44+45 (1.85+1.18 days)	iEEG Scalp EEG	-
Meisel et al. (2020) [71]	Personal	69	-	ACC BVP, EDA TEMP	Wristband
Zhang et al. (2019) [96]	CHB-MIT	22	182 (-)	Scalp EEG	-
Truong et al. (2019) [97]	Freiburg CHB-MIT EPILEPSIAE	13+13+30	59+64+261 (12.95+8.7+120 days)	Scalp EEG iEEG	6, 22, 19
Nejedly et al. (2019) [72]	NeuroVista Canines	4 dogs	75 (1608 days)	iEEG	16
Daoud and Bayoumi (2019) [82]	CHB-MIT	8	43 (-)	Scalp EEG	-
Kiral-Kornek et al. (2018) [98]	NeuroVista	15	2817 (16.29 years)	iEEG	16
Tsiouris et al. (2018) [99]	CHB-MIT	12	185 (40 days)	Scalp EEG	-
Truong et al. (2018) [100]	Freiburg CHB-MIT Kaggle (AES)	13+13+5 dogs+2	59+64+48 (13+8.7+26 days)	Scalp EEG iEEG	6, 22, 19
Kuhlmann et al. (2018) [14]	NeuroVista	3	211 (442 days)	iEEG	16
Karoly et al. (2017) [17]	NeuroVista	9	1458 (10.35 years)	iEEG	-
Direito et al. (2017) [101]	EPILEPSIAE	216	1206t (697t days)	Scalp EEG iEEG	F7, FZ, F8, T5, PZ, T6 6 random 6 in focal region
Bandarabadi et al. (2015) [102]	EPILEPSIAE	24	183t (150t days)	Scalp EEG iEEG	3 in focal region and 3 far from local region
Assi et al. (2015) [103]	Kaggle (AES)	5 dogs	44 (-)	iEEG	16
Rasekhi et al. (2015) [104]	EPILEPSIAE	10	86 (58 days)	Scalp EEG iEEG	3 in focal region and 3 far from local region
Teixeira et al. (2014) [105]	EPILEPSIAE	278	2702 (2031 days)	Scalp EEG iEEG	F7, FZ, F8, T5, PZ, T6 6 random 6 in focal region
Alvarado-Rojas et al. (2014) [106]	EPILEPSIAE	53	558 (531 days)	iEEG	-
Moghim and Corne (2014) [47]	Freiburg	21	- (24 days)	iEEG	3 in focal region and 3 far from local region
Rasekhi et al. (2013) [107]	EPILEPSIAE	10	46t (31t days)	Scalp EEG iEEG	3 in focal region and 3 far from local region
Rabbi et al. (2013) [108]	EPILEPSIAE	1	7 (1.5 days)	iEEG	2
Cook et al. (2013) [60]	NeuroVista	15	1392 (≈ 16 years)	iEEG	16

AES stands for American Epilepsy Society. CHB-MIT for the Children's Hospital Boston from the Massachusetts Institute of Technology, ZUH for Zealand University Hospital, and KCL for King's College London. BTE for Behind-The-Ear. In analyzed time and seizures, "t" stands for testing data. BVP, ACC, EDA, and TEMP stand for blood volume pulse, accelerometry, electrodermal activity, and temperature.

3.2.1 Signal Type

The EEG signal is the most widely used for seizure prediction and forecasting studies. Both scalp EEG and Intracranial EEG (iEEG) are used, as there is no conclusion regarding their comparison. The former can provide information regarding a general brain state, not just localized information. However, the latter has greater proximity to deep brain structures, reduced noise, and fewer artifacts, making it more suitable for intervention devices [12, 13].

However, with the rise in interest concerning seizure forecasting algorithms with recordings extending long periods comes the need to find more comfortable alternatives for data acquisition. New studies [64, 71, 93] have used different biosignals, such as Blood Volume Pulse, Accelerometry, Electrodermal Activity, and Temperature acquired from smartwatches or wristbands. The use of Sub-scalp EEG (ssEEG) [9, 89, 90] has also grown.

3.2.2 Databases

Several EEG databases are available. The most commonly used are the ones from the University of Freiburg [47,97,100], the Children's Hospital Boston (CHB-MIT) [81–83,86–88,96, 97,99,100,109], and the European Database on Epilepsy (EPILEPSIAE) [84,85,92,95,101,102, 104–108], all containing data obtained in a presurgical monitoring context. The popularity of the CHB-MIT is partly due to it being open-source. The most extensive database is EPILEPIAE. It contains an average of 165 hours of EEG recordings from 275 patients (217 with scalp EEG and 58 with iEEG) suffering from Drug-Resistant Epilepsy (DRE) [20].

The NeuroVista clinical trial by Cook et al. (2013) [60] originated a database from the 15 patients that underwent the trial, meaning that it contains data for up to two years per patient, depicting real-life conditions. Later studies use this database [14,17,98]. Smartwatches or wristbands recording different biosignals may be a more practical, comfortable, and optimal alternative. Other popular databases include the Kaggle database, comprised of data from five dogs and two patients [81,88,100,103], SeizeIT1 [91,94] and SeizeIT2 [91], the NeuroPace database, with data from the Responsive Neurostimulation (RNS) system [73,93], and the database from the Zealand University Hospital [89,90]. The use of ssEEG allows for the development of new ultra-long-term databases ideal for forecasting algorithms, such as the former. The evident rise in the use of ultra-long-term databases indicates the present shift from prediction to forecasting, as the large quantities of data allow seizure cycles and additional information to be considered.

3.2.3 Electrode Selection

Most scalp EEG databases use the 10-20 convention for signal acquisition [9, 73, 83–92, 94, 95]. However, not all authors use all electrodes to improve comfort and simulate real-life applications. Some studies use only six electrodes. Among these, some studies place the electrodes in the focal region [101, 105], and others place three in the focal region, and three in the afocal [47, 102, 104]. The idea is that the focal electrodes are near the seizure focus, and the remaining three will only be involved subsequently during the seizure spread. Nevertheless, many authors still utilize all available electrodes [83, 87, 88, 96, 99, 109].

3.3 Preprocessing

To ensure practicality in the real world, recommended solutions for developing a tool capable of receiving and analyzing online data in real-time must consider their real-life feasibility. The general pipeline for signal preprocessing often starts with data segmentation using a sliding window, with further optional steps like denoising, filtering, and artifact removal employed to enhance signal quality. Signal decomposition through frequency decomposition into frequency bands of interest or wavelet decomposition is another potential option [110]. The final step involves defining the preictal period, Seizure Occurrence Period (SOP), and Seizure Prediction Horizon (SPH) (see Figure 3.3). Even though these values can be defined later, during the classification stage, presenting them during preprocessing will prevent this choice from influencing the ML model's performance [12].

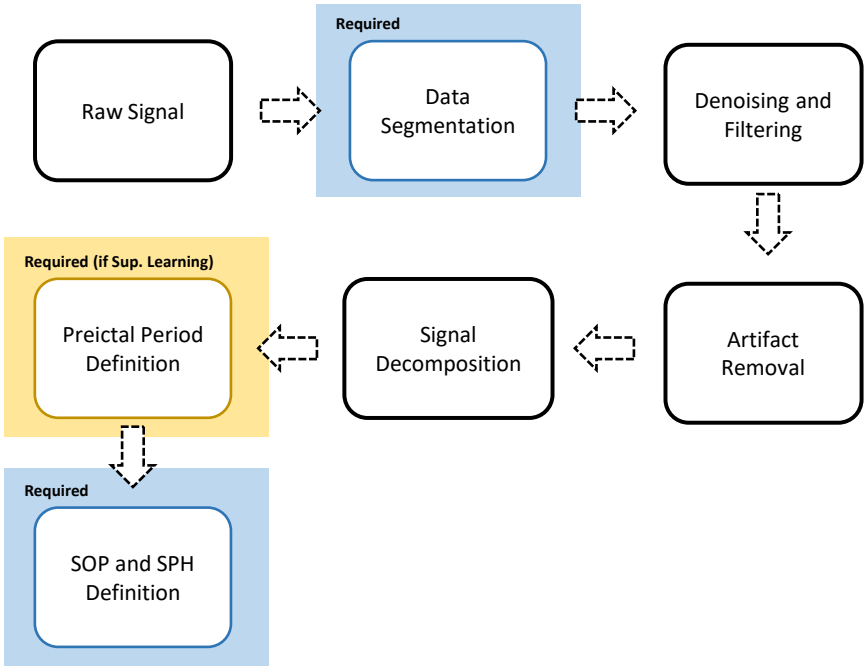


Figure 3.3: Flowchart of the typical seizure prediction and forecasting preprocessing pipeline. The definition of the preictal period is mandatory if a supervised learning approach is used.

Table 3.2 shows an overview of the preprocessing steps used in this field for the past 11 years. Each will be further discussed ahead.

Table 3.2: Overview of the signal preprocessing steps, preictal period, and SPH duration over the last 11 years.

Study	Sliding Window	Filtering	Preictal Period	SPH
Pinto et al. (2023) [84]	5s No overlap	-	30, 40, 50, 60 min	10 min
Lopes et al. (2023) [85]	10s No overlap	0.5–100Hz 4th-order band-pass filter 50Hz 2nd-order notch filter	30 min	10 min
Assali et al. (2023) [109]	2s No overlap	-	30 min and 60 min	-
Xu et al. (2023) [86]	30s 50% overlap	Trap filter 57-63Hz and 117-123Hz	4 hours	5 min

Table 3.2 – *Continued from previous page*

Study	Sliding Window	Filtering	Preictal Period	SPH
Hu et al. (2023) [87]	5s N.A. overlap	5th-order Butterworth band-pass filter 5-50Hz	25 min	5 min
Li et al. (2023) [88]	30s No overlap	-	30 min	1 min (CHB-MIT) 5 min (Kaggle)
Viana et al. (2022) [89]	60s No overlap	0.5-48Hz band-pass and 25Hz low-pass filters 40dB attenuation filter	60 min	5 min
Pal Attia et al. (2022) [90]	60s No overlap	0.5-48Hz band-pass filter 40dB attenuation filter	60 min	5 min
Zhang et al. (2022) [91]	2s 50% overlap	1-25Hz band-pass filter	-	-
Pinto et al. (2022) [92]	5s No overlap	50Hz notch 0.5Hz high-pass	30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90 min	-
Stirling et al. (2021a) [9]	N.A.	0th-order Butterworth band-pass filter Hilbert transform	-	-
Usman et al. (2021) [83]	29s No overlap	Empirical Mode Decomposition	32 min	-
Stirling et al. (2021b) [64]	5s and 60s No overlap	Butterworth band-pass filter Hilbert transform	60 min and 24 hours	-
Nasseri et al. (2021) [93]	1s and 4s N.A. overlap	-	60 min	15 min
Vandecasteele et al. (2021) [94]	2s and 60s 50 and 17% overlap	-	-	-
Pinto et al. (2021) [95]	5s No overlap	50Hz notch 0.5Hz high-pass	30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90 min	-
Xu et al. (2020) [81]	20s No overlap	-	30 min	5 min
Meisel et al. (2020) [71]	30s No overlap	-	60 min	-
Zhang et al. (2019) [96]	5s No overlap	5th-order Butterworth band-pass filter 5-50Hz	30 min	-
Truong et al. (2019) [97]	28s No overlap	Band-pass filters as notch filters 47-53Hz and 97-103Hz	30 min	5 min
Nejedly et al. (2019) [72]	30s 15s overlap	-	60 min	-
Daoud and Bayoumi (2019) [82]	5s No overlap	-	60 min	-
Kiral-Kornek et al. (2018) [98]	5s No overlap	Octave-wide digital and notch filters 8Hz-128Hz	15 min	-
Tsiouris et al. (2018) [99]	5s No overlap	-	15, 30, 60, 120 min	-
Truong et al. (2018) [100]	30s No overlap	Notch-Filters DC removed	30 min	5 min
Kuhlmann et al. (2018) [14]	0s to 600s 0 to 50% overlap	-	55 min	5 min
Karoly et al. (2017) [17]	60s 50% overlap	1-140Hz band-pass filter	30 min	1 min
Direito et al. (2017) [101]	5s No overlap	50Hz notch filter	10:10:40 min	10s
Bandarabadi et al. (2015) [102]	5s No overlap	50Hz notch filter	10:10:40 min	-
Assi et al. (2015) [103]	5s No overlap	50Hz notch 0.5 - 180Hz band-pass	60 min	5s
Rasekhi et al. (2015) [104]	5s No overlap	50Hz notch filter	10:10:40 min	-
Teixeira et al. (2014) [105]	5s No overlap	50Hz notch filter	10:10:40 min	10s
Alvarado-Rojas et al. (2014) [106]	5s No overlap	8th-order Butterworth filter in bands of interest from 0.5Hz to 140Hz Hilbert transform	60 min	1 min
Moghim and Corne (2014) [47]	5s and 9s No overlap	Artifact removal with EEGLAB	5 min	-
Rasekhi et al. (2013) [107]	5s No overlap	50Hz notch filter	10:10:40 min	-
Rabbi et al. (2013) [108]	10 seconds 50% overlap	60Hz notch 0.5 - 100Hz band-pass	15, 30, 45 min	-
Cook et al. (2013) [60]	5s No overlap	Octave-wide digital and notch filters 8Hz-128Hz	minutes to hours	-

3.3.1 Data Segmentation

To extract features from the EEG data, it is common practice to divide the signal into windows (sliding-window analysis), with most selected studies varying from 5 to 60 seconds. The same happens for the overlap, as most studies choose either no overlap or 50%.

As Table 3.2 shows, given the number of electrodes, sampling frequency, and recording duration, several researchers have utilized a 5-second window with no overlap [60, 82, 87, 92, 95, 96, 98, 99, 101, 109]. This window size is believed to strike a balance between capturing specific patterns in the EEG signal and accommodating signal stationarity assumptions. A crucial factor is the trade-off between computational cost and execution speed [12].

3.3.2 Denoising, Filtering and Artifact Removal

This step encompasses the removal of powerline interference (50Hz [92, 95, 101, 107] or 60Hz [108]), band-pass filtering, and artifact removal (abnormal transients).

There has yet to be a consensus on the cut-off frequencies for band-pass filters. The norm is to remove the low-frequency components (below 0.5Hz) associated with breathing artifacts and the high-frequency components (this limit varies between studies) associated with noise.

An option widely employed is temporal filtering with digital filters, namely Infinite Impulse Response (IIR) and Finite Impulse Response (FIR). FIR filters allow for zero-phase distortion and induce a linear phase response, contrasting with IIR filters, which cause barely any ripples in the frequency of interest [12].

Concerning artifacts, there are several ways to deal with the problem. Firstly, it is possible to assume that the artifact's influence on feature engineering will be minimal and ignore it. An alternative is to search for windows or channels containing artifacts and reject them from the analysis. On the other hand, instead of excluding the contaminated windows, it is possible to use methods such as Independent Component Analysis (ICA) and wavelet filtering to separate the artifacts from the signal and lose minimal data. Finally, one last option is training one model to identify and deal with more typical artifacts [34].

3.3.3 Preictal Period Duration, SOP, and SPH

As of yet, there is no standard or optimal preictal duration. Some authors use fixed preictal periods, such as 5 [47], 30 [17, 81, 96, 97], 60 [89, 90, 93], and even 240 minutes [86], and others experiment with different time intervals [64, 92, 95, 99, 101, 108, 109]. Teixeira et al. [105] tested preictal periods of 10, 20, 30, and 40 minutes and found that despite the lack of relevant changes to Sensitivity (SS) values, longer preictal periods significantly lowered the False Positive Rate per Hour (FPR/h).

Values for the SPH also vary between authors, with some being 5 seconds [103] and others 5 minutes [14, 86, 87]. However, many studies completely discard this metric, becoming a severe limiting factor [60, 64, 82, 83, 92, 98, 108]. With no value for the SPH, the real-life applications of these studies are unknown, as there is uncertainty about the ability of the models to predict seizures with sufficient time for an intervention to happen. The same goes for prediction system studies using exceedingly low values. For instance, a scalp EEG prediction model with a SPH of only under one minute leaves very little time for preparation or intervention.

3.4 Feature Extraction

Feature extraction is the most heterogeneous step due to the wide range of proposed methodologies. The extracted features' goal is to capture three behaviors concerning seizure activity: i) an increase of energy caused by the brain's electrical discharges; ii) a shift in spectral power from low to high frequencies; iii) a rise in neuronal synchronization [12, 13].

Features are typically placed in four categories according to the linearity and number of channels. Regarding the former, these can be linear or nonlinear. Concerning the latter, if only one channel is employed, it is univariate; if multiple channels are employed, it is multivariate (emphasis on features that employ two channels, classified as bivariate). No type of feature has been considered optimal. However, in a 2005 study by Mormann et al. [111], a combination of multivariate and univariate features reportedly reached better predictive results. Figure 3.4 lists some features classified into the aforementioned categories. Appendix B presents a more detailed description of these features.

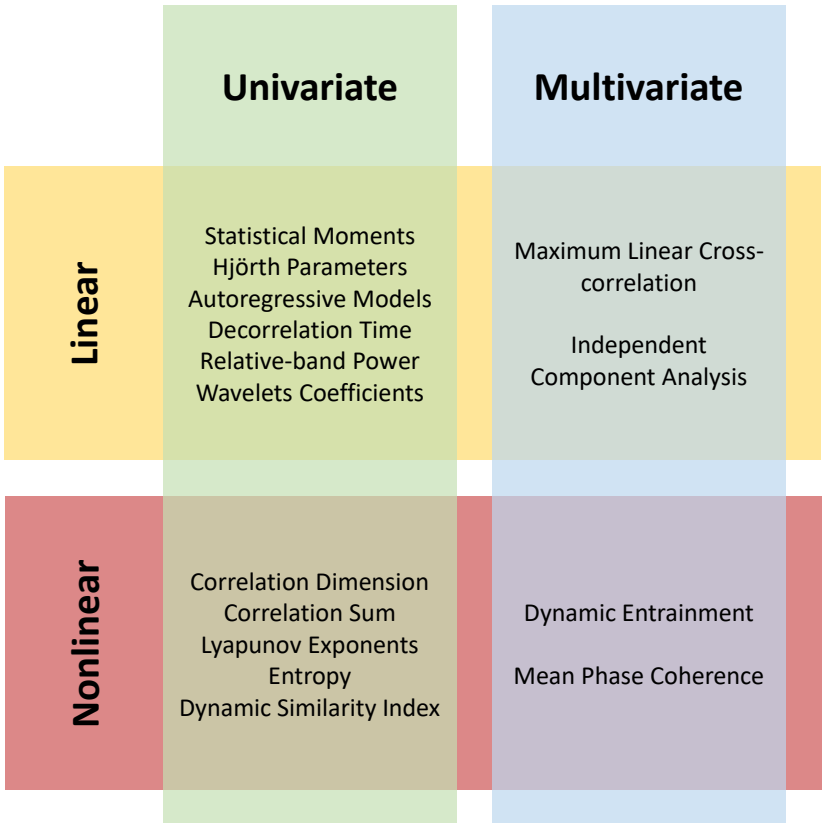


Figure 3.4: Categorization of some typical seizure prediction and forecasting features into univariate, multivariate, linear, and nonlinear.

Table 3.3 shows an overview of the features used in seizure prediction and forecasting studies for the past 11 years.

Table 3.3: Overview of the used features from seizure prediction and forecasting over the past 11 years.

		Linear Univariate Features							Nonlinear Univariate Features						Linear Multivariate Features	Nonlinear Multivariate Features	
Study	**Other**	**Statistical Moments**	**Spectral Band related**	**Wavelets**	**Autoregressive Modelling**	**Energy**	**Hjörth Parameters**	**Decorrelation Time**	**Phase-space and Chaos**	**Lyapunov Exponent**	**Dynamic Similarity Index**	**Line-length**	**Energy**	**Entropy**	**Ratio**	**Mean Phase Coherence**	**Synchrony**
Pinto et al. (2023) [84]	Raw data	X	X	X		X	X	X									
Lopes et al. (2023) [85]	Raw data	X	X	X			X	X									
Assali et al. (2023) [109]	From raw data to STFT Stability Index		X											X			
Xu et al. (2023) [86]	Raw data																
Hu et al. (2023) [87]	Raw data and STFT data		X														
Li et al. (2023) [88]	Raw data																
Viana et al. (2022) [89]	Raw data and FFT data		X														
Pal Attia et al. (2022) [90]	Raw data, FFT data, and TOD		X														
Zhang et al. (2022) [91]	Total Power	X	X											X			
Pinto et al. (2022) [92]		X															
Proix et al. (2021) [73]	Temporal Features																
Stirling et al. (2021a) [9]	Event-based cycles Seizure-based cycles																
Usman et al. (2021) [83]	From raw data to STFT		X														
Stirling et al. (2021b) [64]	HR features Time of the day Sleep features																
Nasseri et al. (2021) [93]	Raw data HR Time of the day																
Vandecasteele et al. (2021) [94]	Time domain HR/HRV		X											X			
Xu et al. (2020) [81]	Raw data																
Meisel et al. (2020) [71]	Raw data HR																
Zhang et al. (2019) [96]	From raw data to CSP																
Truong et al. (2019) [97]	From raw data to STFT		X														
Nejedly et al. (2019) [72]	Raw data and STFT data																
Daoud and Bayoumi (2019) [82]	Raw data																
Kiral-Kornek et al. (2018) [98]	From raw data to Spectograms Time of the day		X														
Tsiouris et al. (2018) [99]	Raw data																
Truong et al. (2018) [100]	From raw data to STFT		X														
Kuhlmann et al. (2018) [14]		X	X	X	X	X	X	X	X				X			X	X
Karoly et al. (2017) [17]						X						X	X				
Direito et al. (2017) [101]		X	X	X	X	X	X	X									
Bandarabadi et al. (2015) [102]			X												X		
Assi et al. (2015) [103]			X				X	X									
Rasekhi et al. (2015) [104]		X	X	X	X	X	X	X							X		
Teixeira et al. (2014) [105]		X	X	X	X	X	X	X									
Alvarado-Rojas et al. (2014) [106]	Phase interaction with HFO																
Moghim and Corne (2014) [47]		X	X	X		X			X	X							
Rasekhi et al. (2013) [107]		X	X	X	X	X	X	X									
Rabbi et al. (2013) [108]											X					X	X
Cook et al. (2013) [60]						X						X	X				

STFT stands for Short-Time Fourier Transform, FFT for Fast-Fourier Transform, TOD for Time Of Day, HR for Heart Rate, CSP for Common Spatial Patterns, and HFO for High-Frequency Oscillations.

There is an evident predominance of univariate and linear features over multivariate and nonlinear ones. The need for greater computational power for multivariate features might explain this trend. Additionally, nonlinear features are less understood than linear ones, less intuitive for clinicians, and are computationally heavier, thus requiring more computation time.

In recent years, there has been a visible rise in DL techniques that use time series data and handle feature engineering automatically [82, 83, 86–88, 99]. However, extracting hand-crafted features may be advantageous instead of automatically extracting ones when the objective is to learn more about the underlying problem.

3.5 Feature Selection

Due to the complexity of brain dynamics, prediction and forecasting algorithms tend to combine numerous features, resulting in high-dimensional feature spaces. In these cases, there might be redundant or confounding features, which can damage the classifier's performance and significantly increase computation time. Therefore, feature selection is a critical step [12].

Generally, feature selection methods look for the most discriminative features, maximizing relevance, and eliminate redundant ones, minimizing similarity. Studies have used several different methods such as ReliefF [47], minimum Redundancy Maximum Relevance (mRMR) [103], maximum Difference Amplitude Distribution histograms (mDAD) [102], minimum normalized difference of percentiles [102], forward selection [112], and Genetic Algorithms (GA) [103]. An alternate strategy to deal with the high dimensional feature spaces is dimensionality reduction using methods such as Principal Component Analysis (PCA). PCA projects the data onto an orthogonal space and selects the projections with higher variance values [113]. DL approaches use convolutional layers [81, 83, 86, 87, 97] or autoencoders [82] to perform feature reduction.

3.6 Classification

A classification model is trained to detect the preictal period and distinguish it from the interictal one based on the selected features [12]. There is a significant heterogeneity regarding the types of models used among authors, as shown in Table 3.4. A shift from Support Vector Machines (SVMs) [47, 91, 94, 101, 102, 107] to Convolutional Neural Networks (CNNs) [72, 81, 83, 98, 109] and Long Short-Term Memory Networks (LSTMs) [71, 89, 93, 99] has been observed. Other classifiers include Deep Residual Shrinkage Network (DRSN) [86] and Transformer [87] based ones, random forests and decision trees [9, 14, 60, 64, 94], k-Nearest Neighbors (kNNs) [60], Adaptive Neuro-Fuzzy Inference System (ANFIS) [103, 108], and Logistic Regressions [9, 17, 64, 92, 95].

A significant problem with training a seizure prediction or forecasting model is that interictal samples are more abundant than preictal ones, causing data imbalance. The way authors deal with this varies. Some address this by undersampling (removing part of the interictal samples) [101, 102, 105, 107]. Others adapted the models to become cost-sensitive classification algorithms [12, 114]. Finally, some authors use Generative Adversarial Networks (GANs) to generate new preictal periods artificially [100].

Table 3.4: Overview of the classification, regularization, performance, and statistical validation over the past 11 years.

Study	Training data (testing data)	Classification (regularization)	Performance	Statistical Validation
Pinto et al. (2023) [84]	First 3 seizures (remaining seizures)	Log Reg, SVMs, CNNs (Firing Power)	SS=0.17 FPR/h=0.87	7 in 40 (0.175) Surrogate Analysis
Lopes et al. (2023) [85]	First 60% of seizures (last 40% of seizures)	CNN-BiLSTM, SNN	SS=0.34 FPR/h=0.90	21 in 41 (0.51) Surrogate Analysis
Assali et al. (2023) [109]	80% samples (20% samples)	CNN	SS=0.93 ACC=0.945	No
Xu et al. (2023) [86]	Leave-One-Out with seizures	DRSN-GRU	SS=0.90 FPR/h=0.025	No
Hu et al. (2023) [87]	80% samples (20% samples)	Transformer	SS=0.92 FPR/h=0.00	No
Li et al. (2023) [88]	Leave-One-Out with seizures	MLPs	SS=0.93 FPR/h=0.11	No
Viana et al. (2022) [89]	Initial 1/3 of data (last 2/3 of data)	LSTM (1h smooth)	SS=0.74 TiW=0.31	5 in 6 (0.83) Surrogate Analysis
Pal Attia et al. (2022) [90]	k-fold cross validation with patients	LSTM (1h smooth)	SS=0.69 TiW=0.37	4 in 6 (0.67) Surrogate Analysis
Zhang et al. (2022) [91]	SeizeIT1 dataset (SeizeIT2 dataset)	SVM	SS=0.88 FPR/h=1.93	4 in 6 (0.67) Surrogate Analysis
Pinto et al. (2022) [92]	First 3 seizures (remaining seizures)	Logistic Regression (Firing Power)	SS=0.16 FPR/h=0.21	30 in 93 (0.32) Surrogate Analysis
Proix et al. (2021) [73]	At least 60% of data	PP-GLMs	AUC=0.74 BSS=0.23	15 in 18 (0.83) Surrogate Analysis
Stirling et al. (2021a) [9]	Retraining and testing chronologically and iteratively	Random Forest+Log Reg	AUC=0.88	No
Usman et al. (2021) [83]	k-fold cross validation with seizures	CNN+LSTM	SS=0.93 SP=0.92	No
Stirling et al. (2021b) [64]	Retraining and testing chronologically and iteratively	LSTM+Random Forest+Log Reg (Kalman Filter)	AUC=0.74	11 in 11 (1.00) Random Forecast
Nasseri et al. (2021) [93]	First 2/3 of data (last 1/3 of data)	LSTM (Kalman Filter)	AUC=0.80	5 in 6 (0.83) Random Predictor
Vandecasteele et al. (2021) [94]	Leave-One-Out with seizures/ with patients	SVM, Random Forest	SS=0.79 (EEG) SS=0.64 (ECG) FPR/h=1.00	No
Pinto et al. (2021) [95]	First 60% of seizures (last 40% of seizures)	Logistic Regression (Firing Power)	SS=0.37 FPR/h=0.79	6 in 19 (0.32) Surrogate Analysis
Xu et al. (2020) [81]	80% samples (20% samples)	CNN	SS=0.96 FPR/h=0.07	No
Meisel et al. (2020) [71]	Leave-One-Out with patients	LSTM	SS=0.51 TiW=0.44	30 in 69 (0.43) Random Predictor
Truong et al. (2019) [97]	Leave-One-Out with seizures	GAN, CNN, NN	AUC=0.81	51 in 56 (0.91) Hanley-McNeil AUC test
Zhang et al. (2019) [96]	Leave-One-Out with seizures	CNN (Kalman Filter)	SS=0.92 FPR/h=0.12	Statistical comparison between methods
Nejedly et al. (2019) [72]	Increases over time (after training epochs)	CNN	SS=0.79 TiW=0.18	4 in 4 (1.00) Random Predictor
Daoud and Bayoumi (2019) [82]	Leave-One-Out with seizures	CNN, Bi-LSTM	SS=0.99 FPR/h=0.004	No
Kiral-Kornek et al. (2018) [98]	First 2 months (remaining duration)	CNN	SS=0.69 FPR/h=0.00	15 in 15 (1.00) Random Predictor
Tsiouris et al. (2018) [99]	K-fold with recordings	LSTM	SS=0.99 FPR/h=0.02	No
Truong et al. (2018) [100]	Leave-One-Out with seizures	CNN (Kalman Filter)	SS=0.79 FPR/h=0.14	28 in 31 (0.90) Random Predictor
Kuhlmann et al. (2018) [14]	Training and testing clips	GLMs, SVM, CNN Ensembles, Boosting, Trees	AUC=0.75 FPR/h=0.58	No
Karoly et al. (2017) [17]	Day 100-200 (Day 200 onwards)	Logistic Regression (Bin width of 1h)	SS=0.55 TiW=0.25 BSS=0.05	9 in 9 (1.00) Time-matched predictor
Direito et al. (2017) [101]	2 - 3 seizures / patient (Remaining seizures)	SVM (Firing Power)	SS=0.38 FPR/h=0.20	24 in 216 (0.11) Random Predictor
Bandarabadi et al. (2015) [102]	First 3 seizures / patient (Remaining seizures)	SVM (Firing Power)	SS=0.76 FPR/h=0.10	23 in 24 (0.96) Random Predictor
Assi et al. (2015) [103]	80% segments (Remaining segments)	SVM, ANFIS	SS=0.85 SP=0.80	No
Rasekhi et al. (2015) [104]	First 3 seizures / patient (Remaining seizures)	SVM (Firing Power)	SS=0.61 FPR/h=0.11	5 in 10 (0.50) Random Predictor
Teixeira et al. (2014) [105]	2 - 3 seizures / patient (Remaining seizures)	SVM, ANN (Firing Power)	SS=0.74 FPR/h=0.28	Statistical comparison between methods
Alvarado-Rojas et al. (2014) [106]	First 4 seizures / patient and at least 10 hours of data (Remaining seizures)	Thresholding (Kalman Filter)	SS=0.68 FPR/h=0.33	7 in 53 (0.13) Random Predictor
Moghim and Corne (2014) [47]	10-fold cross validation with 70%/30% samples	SVM	SS=0.91 SP=1.00	Unspecific predictors
Rasekhi et al. (2013) [107]	First 3 seizures / patient (Remaining seizures)	SVM (Firing Power)	SS=0.74 FPR/h=0.15	No

Table 3.4 – *Continued from previous page*

Study	Training data (testing data)	Classification (regularization)	Performance	Statistical Validation
Rabbi et al. (2013) [108]	1 seizure / patient (5 seizures)	ANFIS	SS=0.80 FPR/h=0.46	No
Cook et al. (2013) [60]	First 4 months (Remaining duration)	kNN+Decision Tree (Smoothing)	SS=0.61 TiW=0.23	9 in 10 (0.90) Time-matched predictor

Log Reg stands for Logistic Regression, LSTM for Long Short-Term Memory, CNN for Convolutional Neural Network, SNN for Shallow Neural Network, DRSN for Deep Residual Shrinkage Network, GRU for Gated Recurrent Unit, MLP for Multi-Layer Perceptron, SVM for Support Vector Machine, ANFIS for Adaptive Neuro-Fuzzy Inference Systems, GLM for Generalized Linear Model, PP-GLM for Point Process Generalized Linear Model, GAN for Generative Adversarial Network, ANN for Artificial Neural Network, kNN for k-Nearest Neighbor, SS for Sensitivity, FPR/h for False Positive Rate per Hour, ACC for Accuracy, AUC for Area Under the Curve, BSS for Brier Skill Score, TiW for Time in Warning.

3.6.1 Data Partitioning

When testing a model, authors should not use the training data, including segments from the same ictal event. Randomly selecting segments from the entire dataset should also be avoided to prevent bias. This ensures that the model's performance is evaluated on unseen data and that the results accurately reflect its ability to generalize to new data. So, several methods for data partitioning have been used.

The different methods take on different assumptions about the seizure-generating process. One assumption is that it is not patient-specific, so they choose a specific number of seizures from the entire set of patients for training and use the remaining for testing [16, 90, 115]. However, most studies take the opposite approach, training and testing the model for each patient [47, 101, 108, 109]. Additionally, recently there has been a rise in the Leave-One-Out strategy, where authors use N-1 seizures (or patients) for training and the remaining one for testing (N is the number of seizures/patients) [71, 82, 86, 88, 94, 96, 97]. Furthermore, for the case of studies such as Cook et al. (2013) [60], Kiral-Kornek et al. (2018) [98], and Nasseri et al. (2021) [93], concept drifts have to be taken into account, so authors retrain the classifiers periodically.

3.6.2 SVMs

SVMs [14, 47, 91, 94, 101–105, 107] are a widely popular technique in supervised ML, as they produce good results in data classification for seizure prediction and forecasting problems. It is known for having a suitable capability of generalization and few parameters that require optimization [12, 101, 104].

One of the options of this algorithm is to employ nonlinear kernel functions (Radial Basis Function (RBF), for instance) to produce nonlinear decision boundaries. The SVM algorithm determines this decision boundary by maximizing the margin of separation between two classes. Therefore, SVMs maximize the shortest distance between the decision function and the closest data points (support vectors) for each class [101, 116].

The ability of SVMs to linearize the feature space and interpret the produced support vectors makes it a desirable choice for interpretable models. However, when dealing with a large number of features, the interpretability of the classifiers, including SVMs, may be lost. Therefore, it is essential to keep in mind the trade-off between the number of features used and the interpretability of the model.

3.6.3 CNNs

CNNs [14, 72, 81–83, 96–98, 100, 109] are DL models capable of learning optimal features from input data. These algorithms are designed to process data containing multiple dimensions or arrays, such as images. By detecting patterns within the input data, neural networks can effectively train themselves to automatically extract features that directly address the temporal properties of the input [77, 117]. For instance, in seizure prediction and forecasting, the time series is transformed into a suitable format that can be fed as input to the neural network. This transformation can be done using raw data or applying techniques such as Fast Fourier Transform (FFT) or wavelet decomposition. Subsequently, the neural networks are trained to detect short-term temporal dependencies in the EEG data, which helps automate feature engineering [97, 118].

Regarding architecture, CNNs usually comprise several convolutional layers that generate feature maps through kernel filtering operations. These layers are then followed by pooling layers that extract features from the maps generated by the previous ones. Classification layers can use these. To prevent overfitting during training, dropout layers are frequently included, randomly setting some units' output to 0 [118].

3.6.4 LSTMs

LSTM networks [64, 71, 83, 89, 90, 93, 99] are prevalent in DL models. These networks, a type of Recurrent Neural Networks (RNNs), include unique units known as gates that regulate, by learning the corresponding weights, which information is stored in memory and which is forgotten by the classifier.

Unlike CNNs, LSTMs are able to learn the temporal characteristics of brain activity during different states while retaining long-term dependencies. This is advantageous in seizure prediction and forecasting. However, like CNNs, LSTMs need significant amounts of data and are prone to overfitting [77, 82].

3.6.5 Postprocessing

A regularization function can be used as the postprocessing step to smooth the output of the classifier [12]. Most selected studies chose not to perform any regularization. Among the ones that did, the most common options were the Firing Power [84, 92, 95, 101, 102, 104, 105, 107] or the Kalman Filter [64, 93, 96, 100, 106].

3.7 Performance Assessment

Finally, the last step is to evaluate the model according to specific metrics. The seizure prediction characteristic, suggested by Winterhalder et al. (2003) [19], recommends using SS and FPR/h, followed by a statistical validation, such as surrogate analysis or an unspecific random predictor. However, many authors stray from this strategy. Some use metrics such as Area Under the Curve (AUC) [9, 14, 64, 73, 93, 97], accuracy [109], sample sensitivity, and sample

specificity [47, 83, 103]. Others shift from seizure prediction to forecasting, so they substitute FPR/h for Time In Warning (TiW) [17,60,71,72,89,90] or add the Brier Skill Score (BSS) [17,73].

The choice of database severely influences the performance of the model. The CHB-MIT database produces the highest performance and most homogeneous results. EPILEPSIAE, on the other hand, perhaps due to the significant number of patients, generates a high heterogeneity in results. Authors use different patient numbers, ranging from only 1 [108] to all 278 patients [105]. This heterogeneity suggests that this database may better demonstrate the predictive power of new methods as it provides data from a wide scope of patients with different characteristics. However, the data is acquired in a presurgical monitoring context, straying from day-to-day patient reality. On the other hand, ultra-long-term databases, such as NeuroVista [14, 17, 60, 72, 98], NeuroPace [73, 93], or others [9, 64, 71], may be the best option for developing seizure warning devices, as they display more realistic performances. However, unlike databases such as EPILEPSIAE and NeuroVista, CHB-MIT is free and open-access.

3.8 Summary

A seizure prediction or forecasting model typically follows the same pipeline consisting of the following steps: signal acquisition, signal preprocessing, feature extraction, feature selection/dimensionality reduction, classification, regularization, and performance assessment. The emergence of DL models has given researchers another option. These can handle raw data and perform feature engineering and automatic classification.

Several authors use databases with presurgical monitoring data, which may not fully represent real-life conditions. Options like ultra-long-term databases and ones with different biosignals acquired are often not public.

Feature engineering and classification can be the most complex steps. However, a predominance of univariate linear features is clear. Feature selection deals with the high-dimensional feature spaces created by the prediction and forecasting algorithms, removing redundant or confounding features and choosing the most discriminative ones. A classification model then uses the selected features to detect preictal segments and distinguish them from interictal ones. Often, authors use a regularization function to smooth the output of the classifier.

Finally, the authors evaluate the performance of the classifier. Most seizure prediction studies choose SS and FPR/h as the performance metrics. However, the recent shift from seizure prediction to seizure forecasting required a more suitable way to evaluate performance. So, new metrics were brought forward, namely TiW and BSS.

4

Methodology

This chapter describes the methodology deployed for the development of an algorithm for epileptic seizure risk assessment. Firstly, Section 4.1 presents an overview of the overall pipeline. Secondly, Section 4.2 describes the developed epileptic seizure forecasting and prediction models.

4.1 Pipeline Overview

The present work aims to develop an Electroencephalogram (EEG)-based algorithm capable of assessing the probability of an epileptic seizure occurring.

Accordingly, a seizure forecasting patient-specific algorithm was developed based on the typical framework presented in Section 3.1 of the state-of-the-art. Three classifier options were explored: a Logistic Regression, an ensemble of 15 Support Vector Machines (SVMs), and an ensemble of 15 Shallow Neural Networks (SNNs). These are among the most widely used, all with different learning mechanisms. Following the evaluation of the model using forecasting metrics, a comparison with a seizure prediction model was made. Figure 4.1 shows an overview of the specific pipeline used for this study.

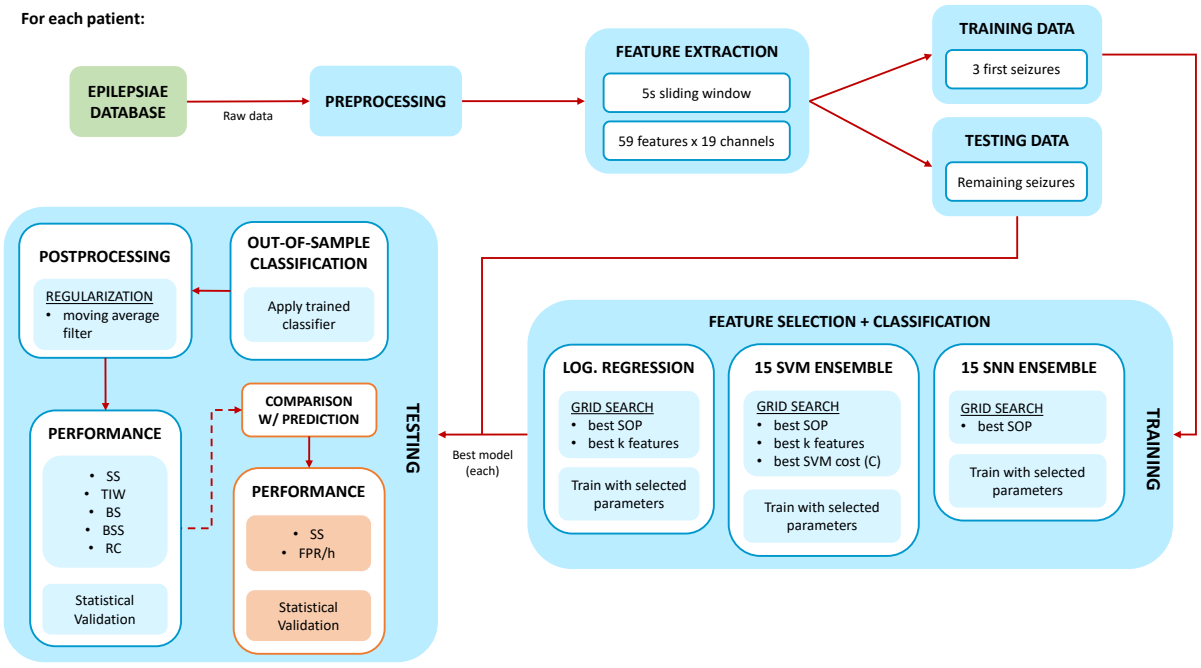


Figure 4.1: General overview of the proposed pipeline.

To summarize the process, all the raw scalp EEG data from the EPILEPSIAE database [20] was preprocessed in order to remove artifacts. It was then divided into 5-second windows to extract relevant features. The data was split into training data, using each patient's first three seizures, and testing data, using the remaining ones. Subsequently, the most discriminative features were selected and used to train the classifier. Finally, the trained classifier was applied, and a moving average filter was applied for a regularization step. The performance was evaluated using specific prediction and forecasting metrics, and the results were compared. Statistical validation was included.

4.2 Seizure Forecasting and Prediction

4.2.1 Data

The current study used a subset of 40 Drug-Resistant Epilepsy (DRE) patients (17 females and 23 males, with a mean age of 41.4 ± 15.9 years) from the European Database on Epilepsy (EPILEPSIAE). The data is collected from patients during presurgical monitoring. The choice of patients rested on multiple criteria. Firstly, the number of independent seizures. Specifically, only patients who had experienced at least four lead seizures, with a minimum separation of 4.5 hours between each seizure, were included [84, 85, 92]. This selection criterion aimed to avoid analyzing clustered seizures. Consequently, 224 seizures met the criteria and were deemed suitable for analysis. Secondly, only patients with Temporal Lobe Epilepsy (TLE) were selected, as it is the most common focal epilepsy syndrome. Finally, the data must be collected using scalp EEG with a sampling rate of 256 Hz. Placement of the 19 EEG electrodes followed the International 10-20 System with the following channels: FP1, FP2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, and Pz. Table 4.1 shows each patient's information.

Table 4.1: Information for the 40 studied patients.

Patient ID	Age	Sex	Number of seizures (train/test)	Seizure classification	Seizure activity pattern	Vigilance at seizure onset	Recording duration (h)
402	55	f	3	FOIA, FBTC, FOIA	t, t, t	A, A, A	103.81
			2	FBTC, FOIA	t, t	A, A	29.66
8902	67	f	3	UC, FOIA, FOIA	a, b, a	A, A, A	133.91
			2	FOIA, FOIA	m, a	A, A	22.5
11002	41	m	3	UC, FOIA, FOIA	?, s, a	A, R, A	97.16
			1	FOIA	t	A	11.7
16202	46	f	3	UC, FBTC, UC	r, ?, r	A, A, A	201.32
			4	FOIA, FOIA, FOIA, FOIA	r, r, ?, r	A, A, A, A	34.45
21902	47	m	3	UC, FOIA, FOIA	t, t, t	A, A, A	67.08
			1	FOIA	b	R	9.76
23902	36	m	3	FOA, FOA, FOA	t, t, t	A, A, A	70.74
			2	FOA, FOA	d, t	A, A	33.95
26102	65	m	3	FOIA, FOIA, FOIA	m, t, t	A, A, A	60.65
			1	FOIA	t	A	22.58
30802	28	m	3	FOA, FOA, FOA	t ,t ,t	R, A, 2	87.57
			5	FOA, FOA, FOA, FOA, FOA	t, t, t, t, t	A, A, R, 2, 2	61.71
32702	62	f	3	FOIA, FOIA, FOIA	t ,t ,t	A, A, A	117.38
			2	FOIA, FOIA	r, a	A, A	20.49
45402	41	f	3	FOIA, FOIA, FOA	t, t, t	A, A, A	71.98
			1	FOIA	t	A	22.31
46702	15	f	3	FOA, FOIA, FOIA	a, a, t	A, 2, A	47.46
			2	FBTC, FOIA	b, t	2, A	12.6
50802	43	m	3	FOIA, UC, UC	t, t, t	A, 2, 2	165.93
			2	FOIA, FBTC	t, t	2, A	35.6
52302	61	f	3	UC, FOA, UC	?, ?, d	A, A, 1	76.45
			1	UC	t	A	6.85

Table 4.1 – *Continued from previous page*

Patient ID	Age	Sex	Number of seizures (train/test)	Seizure classification	Seizure activity pattern	Vigilance at seizure onset	Recording duration (h)
53402	39	m	3	FOA, FOA, FOA	?, ?, ?	A, 2, A	70.31
			1	FOIA	t	A	13.73
55202	17	f	3	FOIA, FOIA, FOA	t, d, t	A, A, A	47.05
			5	UC, UC, FOA, UC, FOIA	t, t, t, r, r	A, A, A, A, A	65.37
56402	47	m	3	UC, UC, UC	t, ?, ?	A, A, A	184.22
			1	FBTC	a	A	20.25
58602	32	m	3	FOIA, FOIA, FOIA	r, t, t	A, R, A	96.94
			3	FOIA, FOIA, FOIA	r, r, t	A, A, 2	23.34
59102	47	m	3	FOA, FOIA, FOIA	?, t, t	A, A, A	65.83
			2	FOIA, FOA	t, t	A, A	82.22
60002	55	m	3	FOIA, FOIA, FOIA	d, c, t	1, A, A	208.11
			3	UC, FOIA, FOIA	t, d, d	R, R, 1	152.4
64702	51	m	3	FOA, FBTC, FBTC	?, m, t	A, A, A	75.91
			2	FBTC, FBTC	t, t	A, 2	31.59
75202	13	m	3	FOA, FOA, UC	t, t, t	2, 2, A	100.94
			4	FOA, FOA, FOA, FOA	t, t, ?, t	A, A, A, A	52.63
80702	22	f	3	FOIA, FOIA, UC	b, b, ?	A, A, A	49.4
			3	FOIA, FBTC, FOIA	c, c, c	A, A, A	29.55
85202	54	f	3	FOIA, FOIA, UC	m, c, m	2, A, A	53.49
			2	UC, UC	m, m	A, A	20.42
93402	67	m	3	FBTC, FOIA, FOIA	t, t, t	2, 2, 2	98.0
			2	UC, UC	t, t	2, 2	54.07
93902	50	m	3	FOA, FOIA, FBTC	t, t, d	A, A, 2	370.83
			3	FOIA, FOIA, UC	d, d, d	A, 2, A	20.29
94402	37	f	3	FOA, UC, FOIA	?, d, b	A, A, A	120.23
			4	UC, FOA, UC, FOA	t, ?, b, ?	2, A ,2, A	30.37
95202	50	f	3	FBTC, FOIA, FOIA	b, b, b	2, 2, 2	57.6
			4	FOIA, UC, FOIA, UC	m, b, b, t	2, 2, 2, 2	89.53
96002	58	m	3	FOIA, FOIA, FOIA	t, t, t	A, A, A	48.4
			4	FOIA, UC ,FOIA, FOIA	d ,a ,t ,a	A, A, A, A	82.2
98102	36	m	3	FOA, UC, UC	?, ?, ?	A, A, A	108.61
			2	UC, FBTC	?, ?	A, A	45.68
98202	39	m	3	FOIA, FOIA, FOIA	t, a, t	A, A, A	111.33
			5	FBTC, FOIA, FOIA, FOIA, UC	t, t, t, t, t	A, A, A, A, A	49.88
101702	52	m	3	FOIA, FOIA, FOIA	t, t, t	A, A, A	28.41
			2	FOIA, FOIA	r, r	2, A	23.83
102202	17	m	3	FOA, UC, FOIA	b, ?, t	2, A, 2	57.45
			4	UC, FOA, FOIA, UC	?, t, t, t	A, A, 2, A	51.41
104602	17	f	3	FOIA, FBTC, FBTC	t, a, t	A, 2, 2	87.87
			2	FBTC, UC	t, d	2, 2	15.25
109502	50	m	3	FOIA, FOIA, UC	t, t, t	A, A, A	76.8
			1	UC	t	A	41.94
110602	56	m	3	FOIA, FOIA, FOIA	t, t, t	A, A, A	89.63
			2	FOIA, FOA	t, t	A, A	25.92
112802	52	m	3	UC, FOIA, UC	t, t, t	A, A, A	71.58
			3	FOIA, FOIA, UC	t, t, t	A, A, A	111.5
113902	29	f	3	UC, FOIA, FOIA	t, d, t	A, A, 2	61.98
			3	FOIA, UC, FOIA	t, t, t	A, 2, A	22.73
114702	22	f	3	FOIA, FOIA, UC	t, t, t	A, A, A	68.39
			5	FOIA, FOIA, FOIA, FOIA, FOIA	t, d, t, d, t	A, A, A, A, A	34.04
114902	16	f	3	FOA, FOIA, FOIA	s, b, s	A, A, A	26.55
			4	FBTC, UC, FOIA, FOIA	t, r, a, t	2, A, A, A	50.66
123902	25	f	3	FBTC, FBTC, FOIA	t, t, t	2, 2, R	152.11
			2	FOIA, FOA	t, t	A, A	30.15

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

4.2.2 Preprocessing

This study's EEG data underwent preprocessing using a Convolutional Neural Network (CNN)-based artifact removal model. Lopes et al. (2021) [119] proposed this model to automatically eliminate artifacts from EEG signals, including eye blinks, eye movements, muscular activity, heart activity, and electrode connection interference, achieving results comparable to those of experts.

The model was trained on raw and manually preprocessed EEG segments to replicate the experts' actions during data preprocessing. The experimental findings demonstrated that the proposed model effectively reduced the impact of EEG signal artifacts without human intervention, making it well-suited for long-term real-time scenarios. It should be noted that the data used in this study consisted of long-term EEG recordings from a subset of the group of epilepsy patients used in this thesis.

4.2.3 Feature Extraction

Following the preprocessing phase, the EEG signals were divided into non-overlapping 5-second windows to extract relevant features. The choice of a 5-second window aligns with current practices in seizure prediction and forecasting [84, 87, 92, 96, 101]. It is appropriate for capturing relevant EEG variations, considering stationarity, temporal, and spectral resolution.

To ensure computational efficiency, univariate linear features were extracted. These features require relatively lower computational power than multivariate features and are more well-understood than nonlinear ones. Additionally, all available electrodes were utilized in feature extraction, recognizing that different brain areas may be involved in seizure generation.

Consequently, a sliding window analysis was used to compute 59 univariate linear features on each of the 19 EEG channels in each window. Several different types of features were extracted. Concerning the frequency domain, it was extracted the relative spectral power of bands delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and four gamma sub-bands: gamma band 1 (30-47 Hz), gamma band 2 (53-75 Hz), gamma band 3 (75-97 Hz), and gamma band 4 (103-128 Hz), the ratio between the bands, spectral edge frequency and power at 50%. Regarding the time domain, the features consisted of the four statistical moments (mean, variance, skewness, kurtosis), Hjörth parameters (activity, mobility, complexity), and decorrelation time. Moreover, the energy of five wavelet detail coefficients (from D1 to D5, using the db4 mother wavelet) was extracted. The choice of features was based on the studies by Pinto et al. (2023) [84] and Lopes et al. (2023) [85], as they used the same patient data from the EPILEPSIAE database. For additional information regarding the extracted features, please refer to Appendix B.

4.2.4 Data Splitting

Each patient's set of features was divided into a group for training and another for testing. The training set, composed of the first three seizures, was used to learn the optimal parameters (the Seizure Occurrence Period (SOP) duration for every classifier, the number of features for the Logistic Regression and the SVM ensemble, and the C parameter for the SVM ensemble) and to train the classifier. The testing set, composed of the remaining seizures per patient, was used to evaluate the classifier's performance. Consequently, the training phase used 120 seizures across all patients, and the testing phase 104.

The presence of time dependence is considered by employing earlier seizures for model training and later seizures for testing. This chronological division acknowledges the dynamic nature of seizure patterns and enables a realistic seizure forecasting and prediction scenario. In this scenario, the model is initially trained using collected seizures, trying to learn specific brain dynamics, and subsequently applying this knowledge to upcoming data in order to predict or forecast future seizures, mimicking the mechanism employed in a possible future warning device.

4.2.5 Training

4.2.5.1 Class Labeling

The samples were divided into two distinct classes: preictal and interictal. The preictal class refers to the moment preceding the seizure onset, which contains the Seizure Prediction Horizon (SPH). During training, the remaining preictal length will correspond to the length of the SOP. In other words, the preictal period is comprised of the total duration of the SPH and SOP combined.

This thesis aims to integrate the developed algorithm into a warning device, allowing users to take rescue medication to eliminate/reduce the effects of a seizure. McIntyre et al. (2005) [120] define the therapeutic success of rescue medication as the "*cessation of seizure within 10 min of drug administration.*" As a result, the SPH will be 10 minutes for this model. The algorithm is supposed to forecast the seizure before the SPH, i.e., during the remaining preictal period, so any samples from this period were discarded from training and testing.

Regarding the SOP, a grid-search was performed, and several values were analyzed: 20, 25, 30, 35, 40, 45, and 50 minutes. All these values were studied and tested for each patient, and the one that optimized the performance was chosen. So, the length of the preictal period varied from 30 to 60 minutes in 5-minute intervals. Higher values would result in a preictal period longer than one hour, which may not be practical for clinical usage.

4.2.5.2 Feature Standardization

Following the class labeling step, the range of independent features extracted from the raw data underwent a standardization process. For the training data, each feature value was normalized using the z-score normalization method, which adjusts the mean of all values to 0 and sets the standard deviation to 1. For the testing data, each feature value was normalized with the training values.

4.2.5.3 Class Balancing

Due to the relative rarity of seizures, a significant imbalance exists between interictal and preictal samples. Class balancing procedures were implemented during the training phase to mitigate bias and specialization of the classifier. This procedure helps address the class imbalance issue and ensures fair interictal and preictal sample representation in the training process. The procedure varied between classifiers.

Class Weights

For the Logistic Regression, a deterministic class balancing process was done through class weights. In the case of this algorithm, as interictal samples are much more frequent than preictal samples, the weight for the interictal class will be smaller. This relationship is inversely proportional. Equation 4.1 shows the formula used to calculate each weight:

$$w_i = \frac{N_{Total\ Samples}}{2 \cdot N_{i\ Samples}}, i = 0, 1. \tag{4.1}$$

Random Undersampling

For the SVM and SNN ensembles, a systematic random undersampling was performed. This procedure results in an equal number of interictal and preictal samples while sampling more or less equally spaced samples from the most represented class, allowing for it to cover, approximately, the full temporality of the signal. It does so by separating the interictal class into n groups, where n is the number of preictal samples, and randomly selecting a sample from each group (see Figure 4.2). Maintaining samples from every interictal interval guarantees higher representativeness.

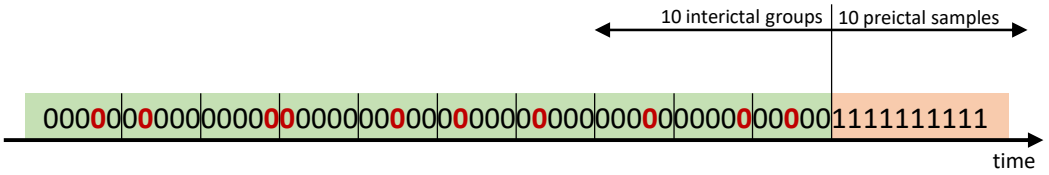


Figure 4.2: Random undersampling of the interictal class on a hypothetical seizure with ten preictal samples. The red-colored samples represent the randomly selected samples from each group.

4.2.5.4 Feature Selection

As with the SOP, a grid-search was performed to find the optimal values for the k most discriminative features per patient. In this model, k varied between 3, 5, 7, 10, 15, 20, and 30 features. Like feature extraction, the feature selection was based on Pinto et al. (2023) [84].

For the Logistic Regression pipeline, the ANOVA (Analysis of Variance) f-test was used. This method assesses the level of linear dependency between each feature and the target. It then ranks the features according to the dependence degree, selecting the k most discriminative features. The SVM ensemble pipeline performs feature selection based on the importance scores obtained from a Random Forest classifier, selecting the top k most discriminative features. The SNN ensemble pipeline did not perform feature selection to allow the SNN to determine the most discriminative features for the problem at hand.

4.2.5.5 Classifier

This work used three classifiers to compare results: a Logistic Regression, a voting system of 15 SVMs, and a voting system of 15 SNNs.

The first approach is the most straightforward one. It uses a Logistic Regression classifier with an L-BFGS (Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm) optimizer, an L2 penalty regularization term, and the class weights considered.

The SVM classifier was used with a linear kernel, which is simple, computationally light, and the performance is comparable to more complex kernels. For better optimization, the SVM 's parameter C (cost) was tuned using a grid-search, along with the SOP duration and k number of features. Considered values were: 2^{-10}, 2^{-8}, 2^{-6}, 2^{-4}, 2^{-2}, 2^{0}, 2^{2}, 2^{4}, 2^{6}, and 2^{8}. This parameter controls the trade-off between the classification margin and non-separable data [101]. A higher value of C generates more intricate decision curves, aiming to closely fit all the training points, possibly leading to overfitting. The SVM also presents an L2 penalty regularization term and a Hinge loss function. An ensemble learning approach was adopted to account for the inherent stochasticity of the random undersampling performed during class balancing. In this approach, 15 SVM classifiers were trained using different data samples.

The final classifier option also took the ensemble approach, but this time using 15 SNNs. The SNN employed was based on Lopes et al. (2023) [85], which used the same patients from this thesis. Its architecture consists of four layers (see Figure 4.3): an Input Layer with an input dimension of 59x19x1 (59 features over 19 channels), a Dropout (Drop) Layer with a 50% rate (to deal with overfitting caused by the limited number of training samples), a Fully Connected (FC) Layer with two neurons, and an Activation Layer with the softmax function. The SNN used an Adam optimizer with a $3 \cdot 10^{-4}$ learning rate and a Binary Cross Entropy loss function.

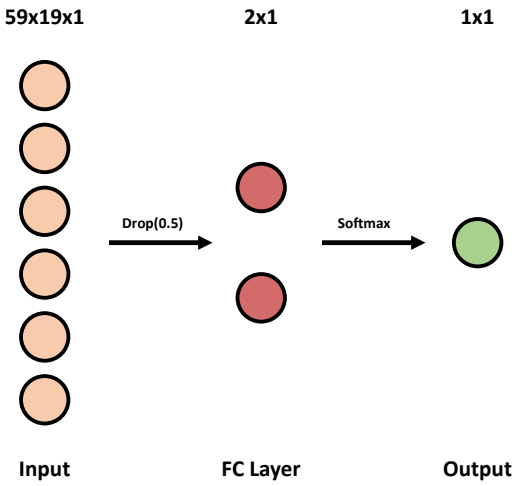


Figure 4.3: Shallow Neural Network architecture based on Lopes et al. (2023) [85]. It is composed of an input layer (Input), a dropout layer with a 50% rate (Drop(0.5)), a fully connected layer (FC layer), and an output layer (Output).

4.2.5.6 Grid-Search

A grid-search strategy was used to determine the optimal parameters for the classifiers. The parameters consist of the optimal SOP (20, 25, 30, 35, 40, 45, 50), k number of features (3, 5, 7, 10, 15, 20, 25), and SVM cost parameter C (2^{-10}, 2^{-8}, 2^{-6}, 2^{-4}, 2^{-2}, 2^{0}, 2^{2}, 2^{4}, 2^{6}, 2^{8}). Different classifiers require different parameters (see Table 4.2).

Table 4.2: Summary of the parameters for the grid-search.

	Parameters		
Classifier	**SOP Duration**	**k-features**	**C-value**
Logistic Regression	•	•	
SVM Ensemble	•	•	•
SNN Ensemble	•		

During the grid-search, a Leave-One-Out Cross-Validation (LOOCV) was implemented to find the optimal parameters. This strategy is leave-one-out seizure-wise, dividing the training set into three seizures, two for training and the remaining for validation. The process has three total iterations, repeating until each set is used for validation. This partitioning strategy guarantees that training and validation sets contain samples from preictal and interictal classes. The ensemble method was applied for the SVM and SNN pipelines. Regarding the former, each iteration of the LOOCV was executed 15 times for the 15 SVMs used in training. Due to the considerable computational power and time required to train 15 SNNs (average of 27.04 minutes), only five SNNs were used to determine the optimal SOP, so each iteration was executed five times. However, the number of SNNs in the ensemble was increased to 15 for the testing phase. This process is demonstrated in Figure 4.4.

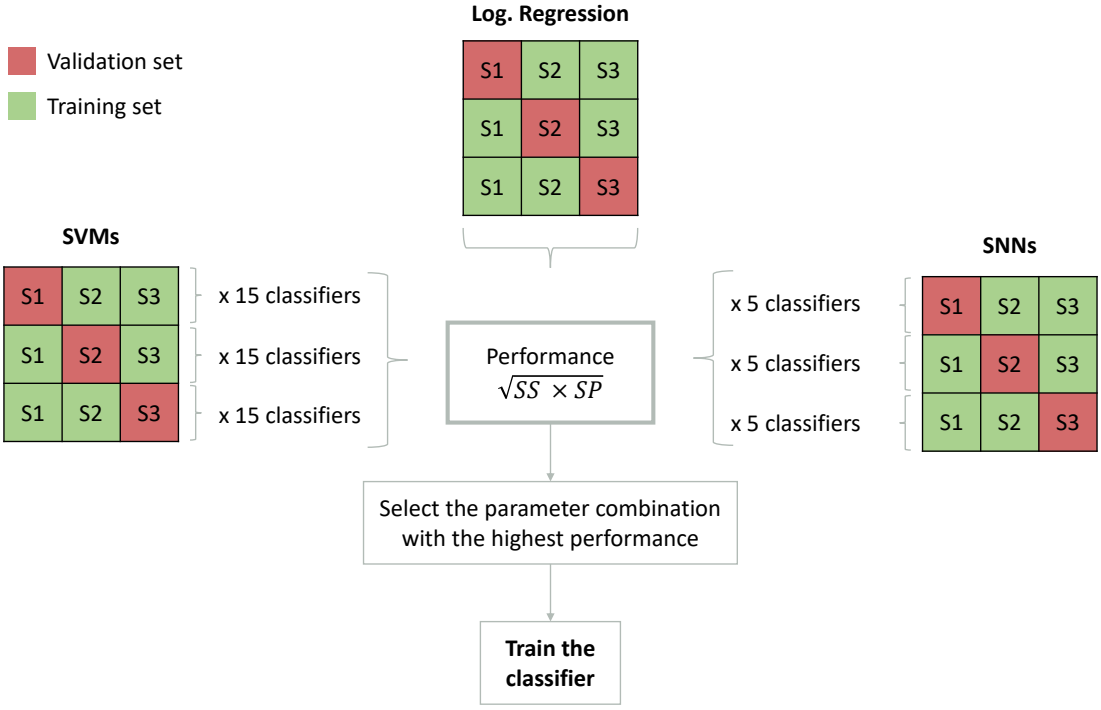


Figure 4.4: Leave-One-Out Cross-Validation procedure implemented to select optimal training parameters.

To evaluate the model, a performance metric capable of revealing the trade-off between the SS_{Sample} (2.3) and the SP_{Sample} (2.4) was required. The chosen metric used to determine which combination of parameters (or the SOP, in the case of the SNN ensemble) is: $\sqrt{SS_{Sample} \cdot SP_{Sample}}$. After finding the optimal parameters for each classifier, the model is trained using each patient's training data.

4.2.6 Testing

Following the training phase, the model was applied to the testing set to make forecasts using an out-of-sample classification approach. The procedure applied to the testing data was similar to that of the training set, except for class balancing (see Figure 4.5). The testing set was standardized using the z-score parameters derived from the training set, ensuring consistency in the feature scaling. The most relevant features identified during the training phase were also the ones selected. Finally, the classifier was employed to generate seizure forecasts.

As for the models using the SVM and SNN classifiers, the procedure was executed for each of the 15 classifiers in the ensemble. This resulted in 15 outputs per sample, so a voting system strategy was used. In this strategy, the most predominant class determined the final output.

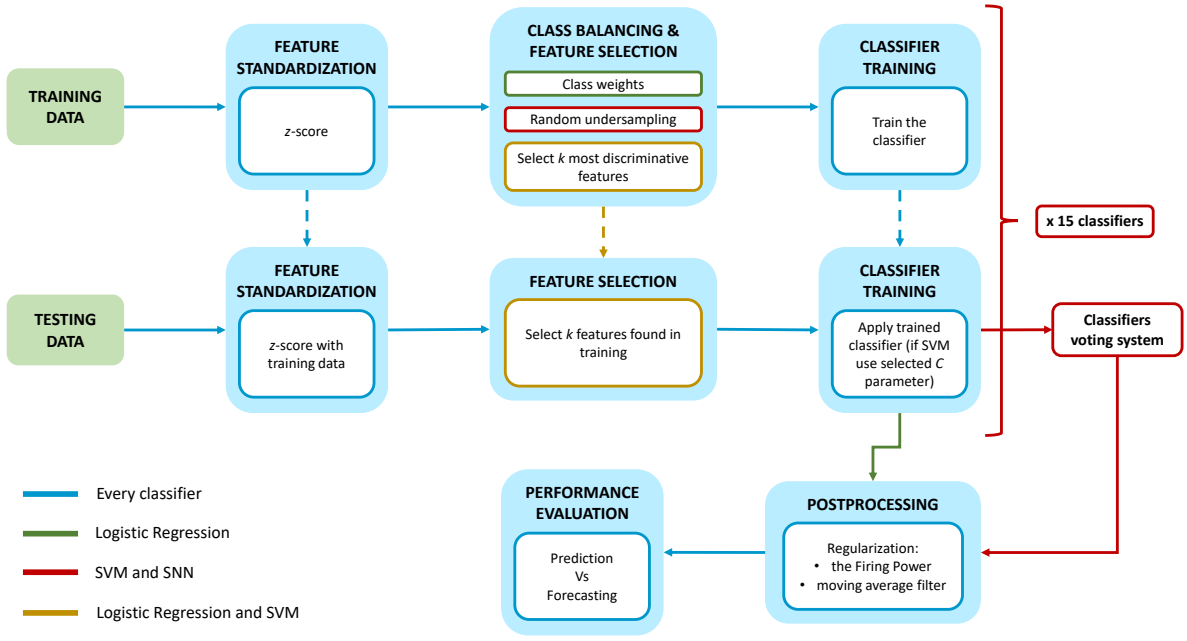


Figure 4.5: Procedure applied to train and test the model for each patient.

4.2.7 Postprocessing

After classification, a regularization step was performed to reduce the noise and smooth the output given, considering the temporal dynamics of the classifier. The chosen method was different for seizure prediction and forecasting.

Regarding seizure prediction, the Firing Power method, described in Section 2.4.6, was used. This method involves applying a moving average low-pass filter to smooth the signal and raising an alarm once the value surpasses a certain threshold. A refractory period (typically SPH+SOP) where no other alarms can be raised is associated with each alarm. The threshold chosen was a conservative value of 0.7, based on several studies [84, 92, 95] using the same database.

For seizure forecasting, only the moving average filter was employed for regularization. For the sake of this model, a patient had a low risk of seizure occurrence if the output was below a reasonable 0.3, a moderate risk between 0.3 and 0.7, and a high risk above 0.7, taking a more conservative approach. Figure 4.6 shows the difference between the methods employed.

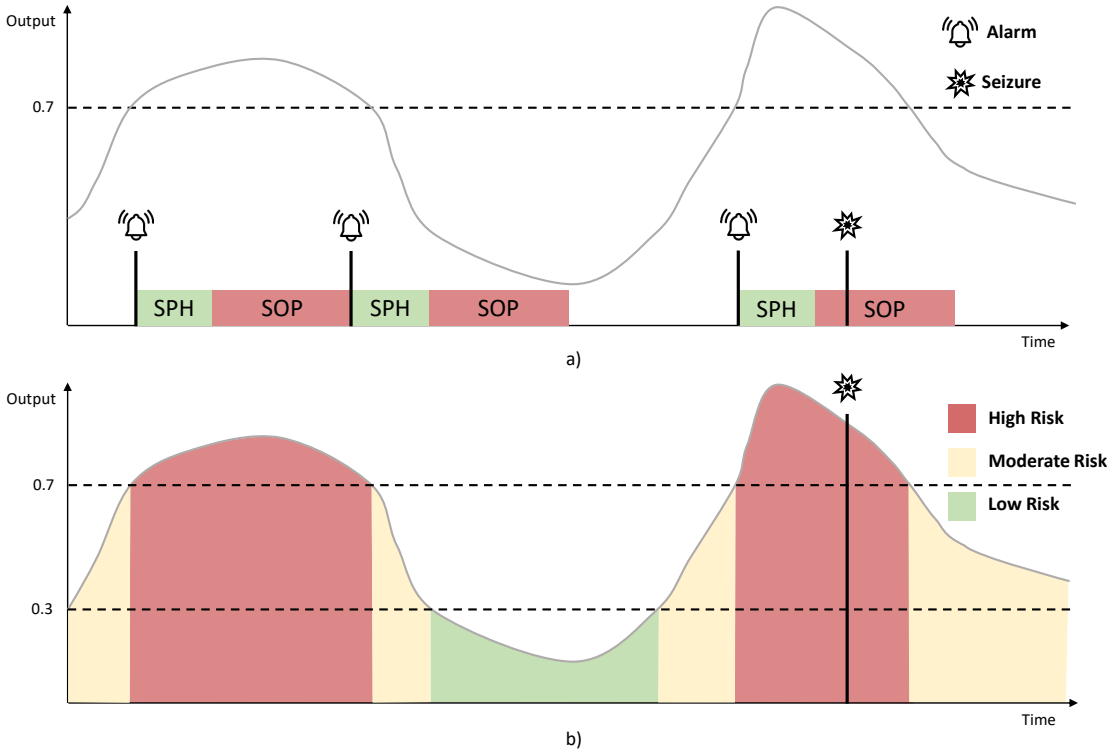


Figure 4.6: Illustration of the postprocessing techniques used. In a) there is the Firing Power with a threshold of 0.7, used for seizure prediction. In b) there is a moving average filter with a determination of the level of risk based on the output (under 0.3 low, between 0.3 and 0.7 moderate, over 0.7 high) used for seizure forecasting.

4.2.8 Performance Evaluation

4.2.8.1 Forecasting

The performance of the seizure forecasting model was evaluated using the metrics described in Section 2.4.7: Sensitivity (SS), Time In Warning (TiW), Brier Score (BS) (Equation 2.5), Brier Skill Score (BSS) (Equation 2.7), and Reliability Curves (RCs). The BS_{ref} used for the BSS was based on Karoly et al. (2017) [17] and was calculated through the mean of 1000 surrogate forecasts.

4.2.8.2 Prediction

The performance of the seizure prediction model, calculated for comparison to the forecasting approach, used the standard metrics for evaluation, also described in Section 2.4.7: SS (Equation 2.1) and False Positive Rate per Hour (FPR/h) (Equation 2.2).

4.2.8.3 Statistical Validation

The surrogate time series analysis (described in Section 2.4.7.3) was employed to assess the statistical validity of the developed algorithm and determine the Improvement over Chance (IoC). This approach aimed to confirm whether the algorithm performed above chance. Here, the original onset time of each seizure was randomly placed within the interictal period. This process was done seizure-by-seizure to ensure that the artificial seizure times respect the seizure distribution over time. The resulting surrogate seizure times were then used to calculate the algorithm's SS. This process was repeated 30 times, and the average SS obtained from the surrogate analysis was compared to the SS achieved by the proposed methodology. The algorithm performed better than chance if its SS was higher and statistically significant. A one-sample t-test with a significance level of 0.05 was conducted to evaluate statistical significance. The null hypothesis stated that "the sensitivity of the proposed methodology is not superior to the sensitivity of the surrogate predictor."

An additional statistical validation technique was made. Another surrogate analysis was made to compare the forecasting models, but this time to check if the BS obtained was better than chance. Here, the data of every seizure was randomly shuffled, and a surrogate forecast drew probabilities in the same distribution as the model's output. This process was repeated 1000 times, and the average BS obtained from this analysis was compared to the proposed model's BS. The same one-sample t-test was performed. The algorithm performed better than chance if the obtained BS is higher than the surrogate and statistically relevant.

5

Results and Discussion

This chapter displays, analyzes, and discusses the results obtained from the proposed methodology. Section 5.1 concerns the results obtained in the training phase, and Section 5.2 the ones from the testing phase. In Section 5.3, there is a comparison of the proposed pipeline's results with similar studies from other authors, both for seizure prediction and forecasting. Finally, Section 5.4 exposes the limitations this present work faced.

5.1 Training Results

During the training phase, three patient-tailored algorithms were developed using different Machine Learning (ML) strategies: a Logistic Regression, a Support Vector Machine (SVM) ensemble, and a Shallow Neural Network (SNN) ensemble. A grid-search was performed to determine different optimal parameters for each model, indicated in Table 4.2 of Section 4.2.5.6. The first three seizures of each patient were used for the classifier's training.

Tables 5.1, 5.2, and 5.3 summarize the training parameters determined during the grid-search method and each patient's respective Sample Sensitivity (SS_{Sample}) and Sample Specificity (SP_{Sample}) values. Additionally, the average SS_{Sample} and SP_{Sample} values were calculated per classifier. The tables concern the Logistic Regression model, 15 SVM ensemble model, and five SNN ensemble model, respectively.

There is no clear tendency for the optimal Seizure Occurrence Period (SOP) and k number of features. These vary significantly from the minimum (20 mins and three features) to the maximum (50 mins and 30 features, respectively). However, despite the C-values ranging from 2^{-10} to 2^{8}, there is a tendency for the smaller ones, mainly 2^{-10}.

The results for the mean SS_{Sample} and SP_{Sample} differ between classifiers: for the Logistic Regression, 0.44±0.19 and 0.69±0.11; for the SVM ensemble, 0.55±0.19 and 0.64±0.10; for the SNN ensemble, 0.73±0.22 and 0.47±0.22, respectively. An increase in the SS_{Sample} accompanies a decrease in the SP_{Sample}. As the SP_{Sample} is higher when using the first two classifiers, it is implied that these are better suited to classify interictal samples than preictal ones. However, as the SS_{Sample} is higher in the SNN ensemble, this classifier is more accurate for preictal samples.

Table 5.1: Training results for the Logistic Regression model. SOP stands for Seizure Occurrence Period, SS_{Sample} for Sample Sensitivity, and SP_{Sample} for Sample Specificity.

Patient	**SOP (mins)**	**k-features**	SS_{Sample}	SP_{Sample}
402	50	20	0.35	0.52
8902	25	7	0.87	0.81
11002	20	7	0.45	0.73
16202	20	30	0.64	0.82
21902	40	10	0.65	0.76
23902	45	15	0.71	0.55
26102	50	30	0.29	0.66
30802	50	5	0.74	0.83
32702	20	7	0.72	0.72
45402	25	20	0.69	0.53
46702	30	20	0.19	0.72
50802	20	10	0.74	0.84
52302	45	30	0.55	0.69
53402	40	20	0.37	0.70
55202	45	3	0.50	0.75
56402	20	10	0.51	0.76
58602	20	3	0.09	0.77
59102	35	30	0.40	0.55
60002	35	7	0.57	0.59
64702	40	3	0.36	0.69
75202	25	30	0.73	0.83
80702	40	30	0.30	0.82
85202	20	20	0.31	0.70
93402	20	3	0.39	0.66
93902	45	5	0.60	0.63
94402	45	30	0.51	0.53
95202	25	15	0.48	0.71
96002	25	3	0.70	0.74
98102	25	20	0.36	0.62
98202	25	30	0.17	0.73
101702	50	30	0.65	0.37
102202	50	3	0.31	0.73
104602	35	30	0.24	0.79
109502	30	20	0.35	0.66
110602	40	10	0.51	0.67
112802	20	3	0.33	0.66
113902	45	30	0.32	0.61
114702	35	30	0.13	0.76
114902	25	10	0.55	0.63
123902	25	3	0.66	0.88
Overall	33.13±10.96	16.05±10.82	0.47±0.19	0.69±0.11

Table 5.2: Training results for the SVM ensemble model. SOP stands for Seizure Occurrence Period, SS_{Sample} for Sample Sensitivity, and SP_{Sample} for Sample Specificity.

Patient	**SOP (mins)**	**k-features**	**C-value**	SS_{Sample}	SP_{Sample}
402	50	30	2^{-2}	0.39	0.58
8902	20	20	2^{-8}	0.90	0.79
11002	20	15	2^{-10}	0.54	0.68
16202	20	15	2^{0}	0.69	0.75
21902	35	7	2^{-2}	0.80	0.61
23902	45	10	2^{-8}	0.74	0.50
26102	50	30	2^{2}	0.37	0.64
30802	50	15	2^{-10}	0.91	0.77
32702	20	30	2^{8}	0.69	0.69
45402	20	15	2^{2}	0.68	0.59
46702	50	30	2^{4}	0.21	0.71
50802	20	30	2^{-8}	0.79	0.83
52302	50	3	2^{6}	0.68	0.71
53402	40	3	2^{-10}	0.53	0.54
55202	20	20	2^{-10}	0.54	0.68
56402	20	15	2^{-10}	0.57	0.66
58602	20	3	2^{-10}	0.36	0.75
59102	50	15	2^{-4}	0.35	0.49
60002	25	10	2^{8}	0.53	0.61
64702	35	3	2^{-10}	0.77	0.43
75202	30	30	2^{-10}	0.77	0.78
80702	45	15	2^{-10}	0.28	0.77
85202	20	20	2^{-10}	0.37	0.54
93402	20	5	2^{-8}	0.53	0.62
93902	35	3	2^{-10}	0.53	0.66
94402	20	20	2^{-4}	0.47	0.67
95202	25	30	2^{-10}	0.58	0.49
96002	45	20	2^{8}	0.83	0.63
98102	25	15	2^{-10}	0.50	0.54
98202	20	10	2^{8}	0.29	0.63
101702	20	15	2^{6}	0.46	0.52
102202	50	5	2^{-6}	0.32	0.62
104602	25	3	2^{-10}	0.41	0.63
109502	50	5	2^{-4}	0.39	0.60
110602	25	30	2^{-10}	0.76	0.66
112802	35	3	2^{4}	0.44	0.52
113902	20	7	2^{-10}	0.36	0.51
114702	35	10	2^{-2}	0.40	0.65
114902	20	3	2^{-10}	0.69	0.51
123902	20	3	2^{-2}	0.74	0.85
Overall	31.13±12.27	14.28±9.78	$2^{-4.30\pm6.57}$	0.55±0.19	0.64±0.10

Table 5.3: Training results for the SNN ensemble model. SOP stands for Seizure Occurrence Period, SS_{Sample} for Sample Sensitivity, and SP_{Sample} for Sample Specificity.

Patient	**SOP (mins)**	SS_{Sample}	SP_{Sample}
402	45	0.56	0.44
8902	35	0.48	0.83
11002	20	0.68	0.36
16202	50	0.37	0.77
21902	20	1.00	0.33
23902	40	1.00	0.34
26102	30	0.79	0.37
30802	40	0.51	0.88
32702	30	0.97	0.37
45402	45	0.98	0.36
46702	25	0.75	0.24
50802	20	0.52	0.74
52302	45	0.27	0.81
53402	35	0.80	0.34
55202	20	0.77	0.39
56402	20	0.88	0.35
58602	30	0.45	0.76
59102	25	0.90	0.43
60002	50	0.93	0.35
64702	50	0.97	0.26
75202	50	0.91	0.29
80702	35	0.82	0.28
85202	50	0.35	0.82
93402	25	0.54	0.44
93902	45	0.94	0.31
94402	50	0.88	0.33
95202	20	0.47	0.78
96002	50	0.87	0.33
98102	45	0.37	0.78
98202	35	0.27	0.81
101702	25	0.90	0.34
102202	50	0.88	0.32
104602	40	0.77	0.22
109502	30	0.76	0.34
110602	50	0.86	0.37
112802	20	0.89	0.41
113902	50	0.78	0.32
114702	20	0.76	0.25
114902	50	0.99	0.38
123902	20	0.66	0.92
Overall	35.88±11.92	0.73±0.22	0.47±0.22

5.2 Testing Results

During the testing phase, each patient's remaining seizures were used to evaluate the performance of the developed patient-tailored algorithms. Different metrics were used to evaluate the seizure prediction and forecasting performances.

Tables 5.4, 5.5, and 5.6 summarize the prediction and forecasting results, as well as the statistical validation results and number of tested seizures, for each patient. The tables concern the Logistic Regression model, 15 SVMs ensemble model, and 15 SNNs ensemble model, respectively.

Table 5.4: Testing results for the Logistic Regression model. SS stands for Sensitivity, FPR/h for False Positive Rate per Hour, TiW for Time in Warning, BS for Brier Score, BSS for Brier Skill Score, and IoC for Improvement over Chance.

		Prediction			**Forecasting**					
Patient	**Tested Seizures**	**SS**	**FPR/h**	**IoC SS**	**SS**	**TiW**	**BS**	**BSS**	**IoC SS**	**IoC BS**
402	2	0.00	0.00		0.00	0.00	0.17	-0.05		
8902	2	1.00	0.10	X	1.00	0.10	0.08	0.21	X	X
11002	1	0.00	0.71		0.00	0.16	0.28	0.00		X
16202	4	0.00	0.03		0.00	0.01	0.07	-0.08		
21902	1	0.00	0.00		0.00	0.00	0.10	-0.15		
23902	2	0.50	1.09		1.00	0.25	0.31	0.13	X	X
26102	1	0.00	0.00		0.00	0.00	0.15	-0.02		
30802	5	0.20	0.37		0.80	0.41	0.30	-0.23	X	
32702	2	0.50	0.05	X	0.50	0.04	0.07	0.22	X	X
45402	1	0.00	0.55		0.00	0.09	0.23	0.01		X
46702	2	0.00	0.00		0.00	0.00	0.20	0.00		
50802	2	0.00	0.26		0.00	0.05	0.07	0.23		X
52302	1	0.00	0.94		0.00	0.43	0.51	-0.06		
53402	1	0.00	0.27		1.00	0.14	0.17	0.16	X	X
55202	5	0.20	0.52		0.80	0.29	0.28	-0.01	X	
56402	1	0.00	0.51		0.00	0.11	0.19	0.00		
58602	3	0.00	0.52		0.00	0.14	0.25	-0.04		
59102	2	0.50	0.99		1.00	0.40	0.37	-0.14	X	
60002	3	0.00	0.05		0.00	0.06	0.18	-0.09		
64702	2	0.00	0.52		0.50	0.15	0.19	0.19	X	X
75202	4	0.00	0.04		0.00	0.01	0.08	-0.07		
80702	3	0.33	0.27	X	0.33	0.09	0.24	0.09	X	X
85202	2	0.00	0.11		0.00	0.03	0.08	-0.03		
93402	2	1.00	0.46	X	1.00	0.33	0.31	-0.60	X	
93902	3	0.00	0.12		0.33	0.04	0.15	0.16	X	X
94402	4	0.00	0.71		0.25	0.20	0.31	0.00		
95202	4	0.00	0.34		0.25	0.08	0.13	0.12	X	X
96002	4	0.25	0.52		0.25	0.18	0.23	0.09		X
98102	2	0.00	0.12		0.50	0.03	0.07	0.16	X	X
98202	5	0.00	0.02		0.00	0.01	0.15	0.02		X

Table 5.4 – *Continued from previous page*

Patient	Tested Seizures	SS	FPR/h	IoC SS	SS	TiW	BS	BSS	IoC SS	IoC BS
101702	2	0.00	0.66		0.50	0.32	0.27	0.02		X
102202	4	0.00	0.04		0.00	0.02	0.09	0.10		X
104602	2	0.00	0.37		0.00	0.10	0.20	0.09		X
109502	1	0.00	1.95		0.00	0.43	0.44	-0.02		
110602	2	0.50	0.31	X	0.50	0.17	0.19	0.10	X	X
112802	3	0.33	0.68		0.67	0.27	0.28	-0.19	X	
113902	3	0.00	0.05		0.00	0.04	0.17	0.01		X
114702	5	0.00	0.00		0.00	0.00	0.13	-0.06		
114902	4	0.00	0.00		0.00	0.00	0.05	0.17		X
123902	2	0.00	0.00		0.00	0.00	0.03	-0.08		

Table 5.5: Testing results for the Support Vector Machine ensemble model. SS stands for Sensitivity, FPR/h for False Positive Rate per Hour, TiW for Time in Warning, BS for Brier Score, BSS for Brier Skill Score, and IoC for Improvement over Chance.

		Prediction			Forecasting					
Patient	Tested Seizures	SS	FPR/h	IoC SS	SS	TiW	BS	BSS	IoC SS	IoC BS
402	2	0.00	0.00		0.00	0.00	0.13	-0.50		
8902	2	0.50	0.20	X	1.00	0.15	0.11	0.00	X	
11002	1	0.00	2.69		1.00	0.46	0.45	-0.01	X	
16202	4	0.00	0.06		0.00	0.02	0.08	-0.01		
21902	1	0.00	0.00		0.00	0.00	0.14	-0.16		
23902	2	0.00	1.38		0.50	0.31	0.37	0.08		X
26102	1	0.00	0.10		0.00	0.04	0.15	0.00		X
30802	5	0.60	0.38	X	0.80	0.44	0.36	-0.25	X	
32702	2	0.00	0.11		0.00	0.01	0.07	0.03		X
45402	1	0.00	0.85		0.00	0.16	0.24	-0.05		
46702	2	0.00	0.00		0.00	0.00	0.19	-0.04		
50802	2	0.00	0.16		0.00	0.04	0.06	0.21		X
52302	1	0.00	1.05		0.00	0.19	0.31	-0.15		
53402	1	0.00	0.27		1.00	0.16	0.23	0.11	X	X
55202	5	0.20	1.06		0.20	0.26	0.28	-0.09		
56402	1	0.00	3.45		0.00	0.53	0.48	-0.02		
58602	3	0.00	0.00		0.00	0.00	0.06	-0.14		
59102	2	0.00	0.59		0.00	0.27	0.31	-0.13		
60002	3	0.33	0.32		0.33	0.28	0.31	-0.04		
64702	2	0.00	1.40		0.50	0.29	0.28	0.23		X
75202	4	0.00	0.04		0.00	0.02	0.09	-0.13		
80702	3	0.67	0.48	X	0.67	0.21	0.27	0.11	X	X
85202	2	0.50	1.21	X	1.00	0.27	0.29	0.05	X	X
93402	2	0.50	1.20		1.00	0.30	0.31	-0.12	X	
93902	3	0.00	0.06		0.00	0.01	0.13	0.04		X
94402	4	0.00	1.54		0.00	0.29	0.33	0.02		X
95202	4	0.25	0.62		0.25	0.19	0.20	0.10		X
96002	4	0.00	0.94		0.25	0.42	0.46	-0.05		

Table 5.5 – *Continued from previous page*

Patient	Tested Seizures	SS	FPR/h	IoC SS	SS	TiW	BS	BSS	IoC SS	IoC BS
98102	2	0.50	0.17	X	0.50	0.05	0.10	0.09	X	X
98202	5	0.00	1.55		0.00	0.23	0.27	0.18		X
101702	2	0.00	0.87		0.50	0.19	0.22	0.04		X
102202	4	0.00	0.36		0.50	0.18	0.28	0.00	X	X
104602	2	0.50	2.67		1.00	0.41	0.38	0.08	X	X
109502	1	0.00	0.29		0.00	0.15	0.23	-0.05		
110602	2	0.00	0.45		0.50	0.16	0.20	0.04	X	X
112802	3	0.00	2.22		0.67	0.65	0.60	-0.12		
113902	3	0.33	0.21	X	0.33	0.05	0.17	0.03	X	X
114702	5	0.00	0.00		0.00	0.00	0.13	-0.02		
114902	4	0.25	0.20	X	0.25	0.06	0.08	0.24	X	X
123902	2	0.00	0.00		0.00	0.00	0.04	-0.09		

Table 5.6: Testing results for the Shallow Neural Network ensemble model. SS stands for Sensitivity, FPR/h for False Positive Rate per Hour, TiW for Time in Warning, BS for Brier Score, BSS for Brier Skill Score, and IoC for Improvement over Chance.

		Prediction			Forecasting					
Patient	Tested Seizures	SS	FPR/h	IoC SS	SS	TiW	BS	BSS	IoC SS	IoC BS
402	2	1.00	0.91	X	1.00	0.36	0.33	0.28	X	X
8902	2	0.00	0.22		0.50	0.33	0.32	-1.12		
11002	1	1.00	3.15	X	1.00	0.60	0.55	0.02	X	X
16202	4	0.25	0.19		0.25	0.12	0.18	-0.02		
21902	1	0.00	1.89		0.00	0.40	0.42	-0.07		
23902	2	0.50	6.93		1.00	0.86	0.72	0.03		X
26102	1	1.00	2.95	X	1.00	0.61	0.54	0.02	X	X
30802	5	0.20	0.16		0.40	0.13	0.16	-0.17	X	
32702	2	1.00	1.19	X	1.00	0.36	0.30	0.02	X	X
45402	1	0.00	1.52		1.00	0.53	0.50	0.04	X	X
46702	2	0.50	11.21		0.50	0.85	0.84	-0.10		
50802	2	0.50	0.60	X	0.50	0.14	0.13	0.31	X	X
52302	1	0.00	0.94		0.00	0.37	0.38	-0.11		
53402	1	0.00	2.64		1.00	0.63	0.53	0.03	X	X
55202	5	0.80	3.87	X	1.00	0.66	0.64	-0.07	X	
56402	1	0.00	8.55		0.00	0.79	0.71	-0.02		
58602	3	0.00	0.10		0.00	0.01	0.11	-0.14		
59102	2	1.00	4.82	X	1.00	0.82	0.72	-0.12	X	
60002	3	0.33	0.32		0.33	0.50	0.45	-0.07		
64702	2	0.00	0.84		0.50	0.27	0.30	0.18		X
75202	4	0.75	0.86	X	0.75	0.46	0.41	-0.05	X	
80702	3	0.33	0.81		0.33	0.45	0.43	-0.22		
85202	2	0.00	0.00		0.00	0.00	0.09	0.06		X
93402	2	0.00	0.45		0.50	0.18	0.17	-0.03		
93902	3	0.33	0.47		0.67	0.23	0.20	0.26	X	X
94402	4	0.50	3.02		0.75	0.61	0.54	0.03		X

Table 5.6 – *Continued from previous page*

Patient	Tested Seizures	SS	FPR/h	IoC SS	SS	TiW	BS	BSS	IoC SS	IoC BS
95202	4	0.00	0.43		0.00	0.11	0.14	0.08		X
96002	4	0.50	2.21		1.00	0.71	0.56	-0.02		
98102	2	0.00	0.05		0.00	0.01	0.05	0.39		X
98202	5	0.00	0.26		0.00	0.20	0.26	-0.47		
101702	2	0.50	1.97		0.50	0.46	0.45	0.02		X
102202	4	0.50	1.36		0.75	0.51	0.51	0.00		
104602	2	0.50	0.90		0.50	0.34	0.36	-0.01		
109502	1	0.00	2.03		1.00	0.52	0.51	0.01	X	X
110602	2	0.00	1.71		1.00	0.56	0.45	0.06	X	X
112802	3	0.33	4.82		0.67	0.76	0.68	-0.09		
113902	3	0.33	11.69		1.00	0.88	0.70	-0.02		
114702	5	0.60	5.10		1.00	0.70	0.66	-0.01	X	
114902	4	0.50	0.81		0.75	0.41	0.37	0.06	X	X
123902	2	0.00	0.00		0.00	0.00	0.02	-0.06		

The total number of tested seizures spanned from one to five. This discrepancy causes difficulties when comparing Sensitivity (SS) values between patients—for instance, in the case of patients 26102 and 30802. The SNN ensemble's forecasting results (Table 5.6) show that patient 26102 displays a SS of 1.00 while patient 30802's SS is 0.40, significantly lower. However, patient 26102 only had one seizure evaluated, while patient 30802 had five. Consequently, the algorithm was able to forecast one of the first patient's seizures and two of the second one. Therefore, even though the algorithm could forecast twice as many seizures from patient 30802, its SS is less than half. For patients like 26102, with only one tested seizure, the SS is either 1.00 if the seizure is correctly predicted/forecast or 0.00 if not.

Table 5.7 displays the average values for all applied metrics. It can be used to compare the classifiers and the approaches.

Table 5.7: Overall results for the three Machine Learning pipelines. SVM stands for Support Vector Machine, SNN for Shallow Neural Network, SS for Sensitivity, FPR/h for False Positive Rate per Hour, TiW for Time in Warning, BS for Brier Score, BSS for Brier Skill Score, and IoC for Improvement over Chance.

	Prediction			Forecasting					
Model	**SS**	**FPR/h**	**IoC SS**	**SS**	**TiW**	**BS**	**BSS**	**IoC SS**	**IoC BS**
Logistic Regression	0.13±0.26	0.36±0.40	12.5% (5 in 40)	0.28±0.37	0.13±0.14	0.19±0.11	0.01±0.15	37.5% (15 in 40)	50% (20 in 40)
15 SVMs Ensemble	0.13±0.21	0.73±0.85	17.5% (7 in 40)	0.32±0.37	0.19±0.17	0.23±0.13	-0.01±0.14	32.5% (13 in 40)	47.5% (19 in 40)
15 SNNs Ensemble	0.34±0.35	2.30±2.89	20% (8 in 40)	0.58±0.39	0.44±0.26	0.41±0.21	-0.03±0.23	40% (16 in 40)	45% (18 in 40)

5.2.1 Seizure Prediction

Starting with the comparison between seizure prediction models. The lowest SS values belong to the Logistic Regression and the SVM ensemble (0.13), and the highest to the SNN ensemble (0.34). As expected, an increase in the False Positive Rate per Hour (FPR/h) accompanied the increase in SS. The Improvement over Chance (IoC) also followed this trend.

This algorithm used a Firing Power threshold of 0.7, taking a more conservative approach, so lower SS values are to be expected. However, the values for the FPR/h raise concerns about the practical applicability of the system in real-life scenarios [19].

An interesting comparison between the Logistic Regression and SVM ensemble models can be made. They both produce the same mean SS, but there is a significant increase in the FPR/h from the Logistic Regression to the SVM ensemble (0.36±0.40 to 0.73±0.85). This disparity, supported by the fact that the Logistic Regression is far computationally lighter than the 15 SVM ensemble (average of 7.34 seconds vs. 196.18 seconds per patient, respectively), suggests that the former may be a better option.

On the other hand, the SNN ensemble produced a considerable increase in the SS, but a comparable increase in the FPR/h also followed. Winterhalder et al. (2003) [19] proposed a FPR/h_{max} of 0.15 in a presurgical monitoring context. This value strays considerably from the 2.30±2.89 obtained using this model, suggesting that it might not be suitable for patient use. Moreover, the conservative approach taken for the model also shows itself in the FPR/h. The choice of a 0.5 threshold, as is the case for many studies, would further increase this value.

Results for the IoC were also poor for every model, with only a three-patient difference from the worst to the best classifier.

The choice of the best proposed methodology is not straightforward. It was already determined that the SVM ensemble model produced the worst results, having the same SS value as the Logistic Regression one while doubling the FPR/h. Deciding between the remaining two methodologies poses a problem. The Logistic Regression methodology has a lower SS than the SNN one. Nevertheless, it also displays a much lower FPR/h. On the one hand, the high 2.30 value for the SNN ensemble's FPR/h is unsuitable for real-life use. On the other hand, neither is the low 0.13 value for the Logistic Regression's SS.

5.2.2 Seizure Forecasting

Regarding the forecasting approach, the SS values were all different between models, ranging from 0.28 for the Logistic Regression to 0.58 for the SNN ensemble. However, similar to prediction, Time In Warning (TiW) accompanies the SS increase, from 0.13 (Logistic Regression) to 0.44 (SNN ensemble). This tendency is exhibited when looking at the Brier Score (BS) and the Brier Skill Score (BSS), as these metrics' results progressively worsen. According to the BSS, only the Logistic Regression shows an improvement relative to the reference, with a score of 0.01±0.15. The closer the score is to 0, the less improvement over the reference it has. So, for this model, the improvement is low. Both the SVM and the SNN ensembles are worse than the uninformed reference, displaying negative scores (-0.01±0.14 and -0.03±0.23, respectively). The BSSs also show that the SVM ensemble is better than the SNN ensemble.

Concerning the statistical validation, the IoCs were relatively close to each other, with only a three-patient difference for the SS IoC and two for the BS. However, better IoC results were obtained when assessing the BS. These scored between 45% and 50% of patients, while the SS IoC were between 32.5% and 40%. This increase can be explained by the fact that the BS is the mean squared error of the forecasts relative to the observations. So, while statistical validation using the SS only assesses the classifier's ability to detect the preictal state correctly, the one using the BS also considers correct identification of the interictal period with low seizure probabilities. This way, it can validate patients based on the model's aptness to correctly identify interictal samples as low-risk states, even if they cannot forecast the seizure. Furthermore, the BS is a more independent metric than the SS, as the latter is entirely dependent on the determination of a threshold, and the former only deals with the probabilistic output.

Reliability Curves (RCs) were plotted for every patient, depicting the forecast likelihood of the preictal period against the observed frequency. An ideal RC is a diagonal line where the forecast probabilities equal their observed frequency.

The obtained BSSs can be confirmed by visualization of the RCs. Figure 5.1 shows the RCs obtained using the data from every testing seizure of every patient. The fact that all these curves more closely resemble a flat line indicates that the forecasts are not significantly better than an uninformed forecast. This conclusion is supported by all the BSSs being close to zero, implying that the models are either slightly better or slightly worse than the reference. The SNN model's higher mean SS can be observed in Figure 5.1, as the samples that have a higher probability of being preictal appear slightly more often than in the other models.

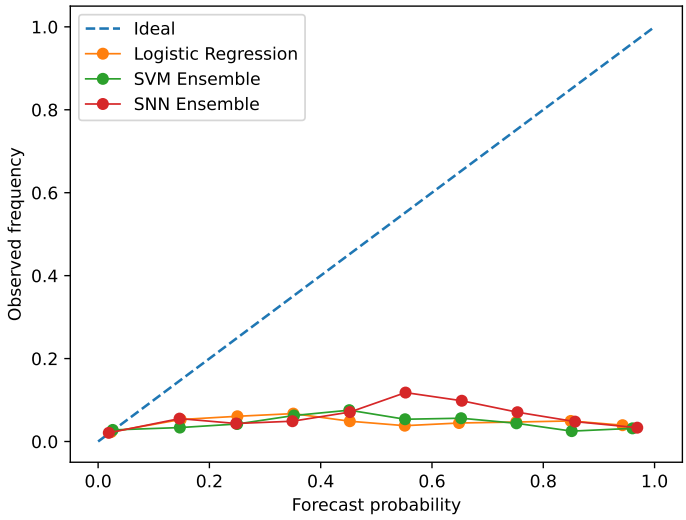


Figure 5.1: Reliability Curves for the three models, using data from every patient.

Taking a closer look at patient 8902. This patient has a SS of 1.00 for the Logistic Regression and 0.50 for the SNN ensemble, respectively. These results show that, for patient 8902, the Logistic Regression classifier is better at classifying the preictal period than the SNN ensemble classifier. This relationship is evident in Figure 5.2. Here, the samples of higher preictal likelihood have a higher observed frequency for the Logistic Regression, meaning these samples are classified as preictal more often than in the model with the lower sensitivity. Furthermore, the

RC can help visualize why, for the SNN ensemble, a higher TiW does not result in a higher SS for this patient. The almost flat line above 0.5 shown in the SNN ensemble's RC demonstrates that despite being longer under warning, these do not match the true preictal samples. On the other hand, the RC also shows that the SNN ensemble is better at classifying interictal samples, as the curve relating to samples with a low probability of being preictal (high probability of being interictal) is closer to the ideal RC than in the other model.

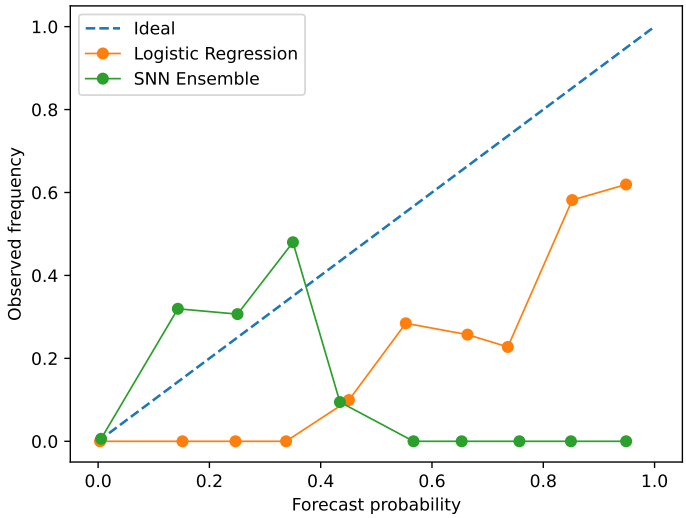


Figure 5.2: Reliability Curves for patient 8902 using the Logistic Regression and the Shallow Neural Network ensemble.

Remarks regarding the quality of seizure prediction results hold up for seizure forecasting. The threshold to distinguish between moderate and high seizure likelihood is 0.7, a conservative value. As a result, the obtained SS values will be lower than if a less conservative approach was selected, for instance, with a 0.5 threshold [102].

Defining one of the three proposed methodologies as the best one is easier in the case of forecasting than it is in prediction. Only one model produced a positive BSS, meaning only one model performed better than an uninformed forecast. Consequently, even though the other two models have higher SSs, they are worse for practical use. So, the best methodology is the Logistic Regression model.

5.2.3 Prediction vs Forecasting

The main goal of this study is to develop an algorithm capable of assessing the likelihood of a seizure and to determine if this probabilistic technique produces better results than a crisp prediction view. The way to address this challenge is to compare the SS values. A significant improvement in the SS can be seen in all three models when moving from prediction to forecasting. There is a 115% increase on the Logistic Regression model (0.13±0.26 to 0.28±0.37), a 146% increase on the SVM ensemble (0.13±0.21 to 0.32±0.37), and a 70% increase on the SNN ensemble (0.34±0.35 to 0.58±0.39). The same can be said about the statistical validation, showing an IoC increase of 300% on the Logistic Regression (12.5% to 50%), 171% on the SVM ensemble (17.5% to 47.5%), and 125% on the SNN ensemble (20% to 45%).

The definition of SS for both approaches can explain this considerable increase. Even though they share the same name, it is not the same metric. In prediction, it is defined by Equation 2.1, which corresponds to the ratio between true alarms and total seizures. However, in forecasting, it corresponds to the ratio of seizures that occur after the forecast of a high-risk state happens during the SOP.

In practical terms, there is a reason why a seizure forecasting algorithm can have better SS results than a prediction one. When an alarm is raised, it is associated with a refractory period, usually the duration of the Seizure Prediction Horizon (SPH) and SOP combined—for instance, in the case of patient 30802's 8th seizure. Figure 5.3 shows the Firing Power output for the seizure prediction model and the moving average filter output for the forecasting model with the Logistic Regression classifier. In this case, an alarm was raised right before the SOP. Even though the probabilistic output was above 0.7, due to the refractory period and the deterministic approach characteristic of seizure prediction, no alarm was raised during the SOP, and the algorithm failed to warn the patient in the designated period. However, seeing as the forecasting model was in a high-risk state during part of the SOP, the algorithm succeeded.

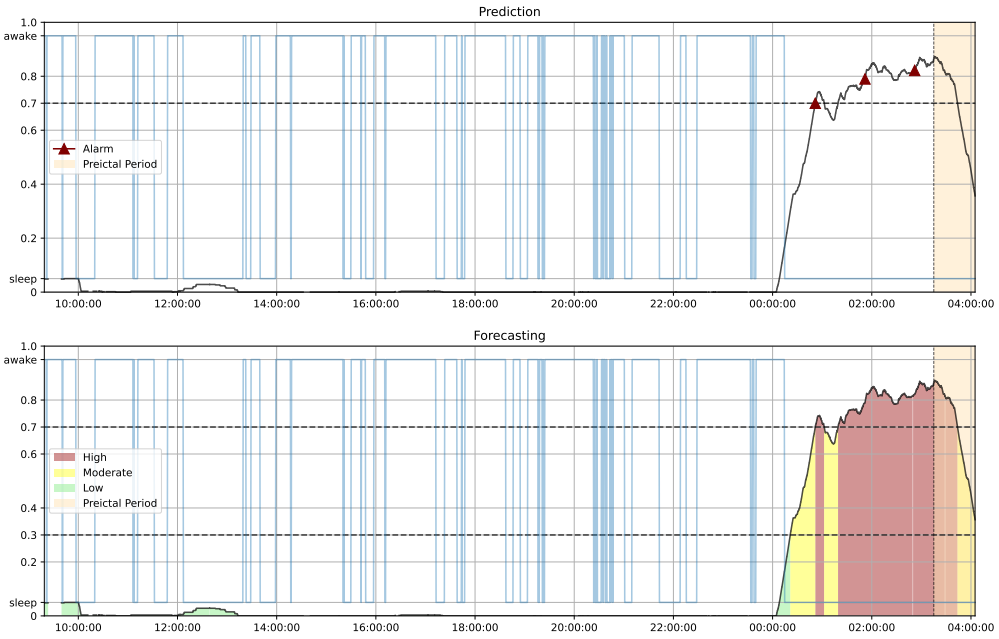


Figure 5.3: Plot of the Firing Power's output with the alarms (seizure prediction) and the moving average filter's output with the risk zones (seizure forecasting) for seizure 8 of patient 30802 using the Logistic Regression model. The light blue curve represents the sleep/awake state.

Seizure prediction algorithms also fail when dealing with potential seizure self-termination brain mechanisms. Figure 5.4 shows the SVM ensemble output of patient 114902's 6th seizure. There is a clear spike slightly after 20:00:00 and two between around 09:00:00 and 11:00:00. These spikes may be caused by a seizure close to occurring that was terminated by specific brain dynamics. Even though using the probabilistic approach of forecasting a patient is rarely sure of a seizure, it can avoid these occasions where the model is wrong.

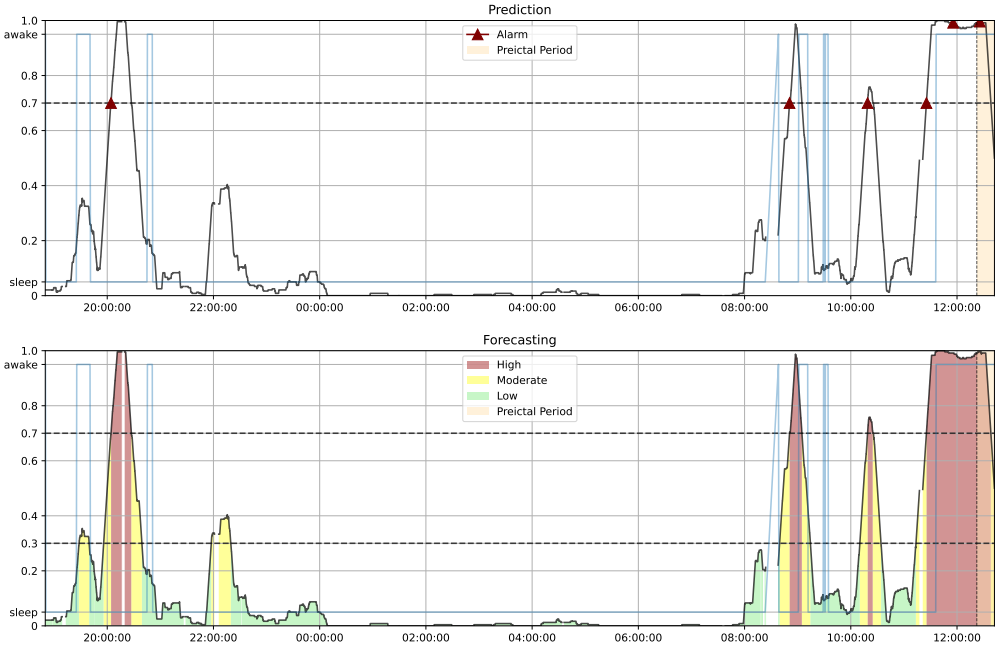


Figure 5.4: Plot of the Firing Power's output with the alarms (seizure prediction) and the moving average filter's output with the risk zones (seizure forecasting) for seizure 6 of patient 114902 using the Support Vector Machine model. The light blue curve represents the sleep/awake state.

A shared difficulty between seizure prediction and forecasting algorithms is the heterogeneity in epilepsies and seizures. Even with patient-tailored algorithms, as is the case with this present work, this heterogeneity of results in different patients is evident. Figure 5.5 a) shows the moving average filter's output from the SNN ensemble model for patient 23902's two testing seizures and Figure 5.5 b) from patient 123902's. There is a clear difference in the models' outputs, as patient 23902's model overly classifies samples as preictal and patient 123902's as interictal.

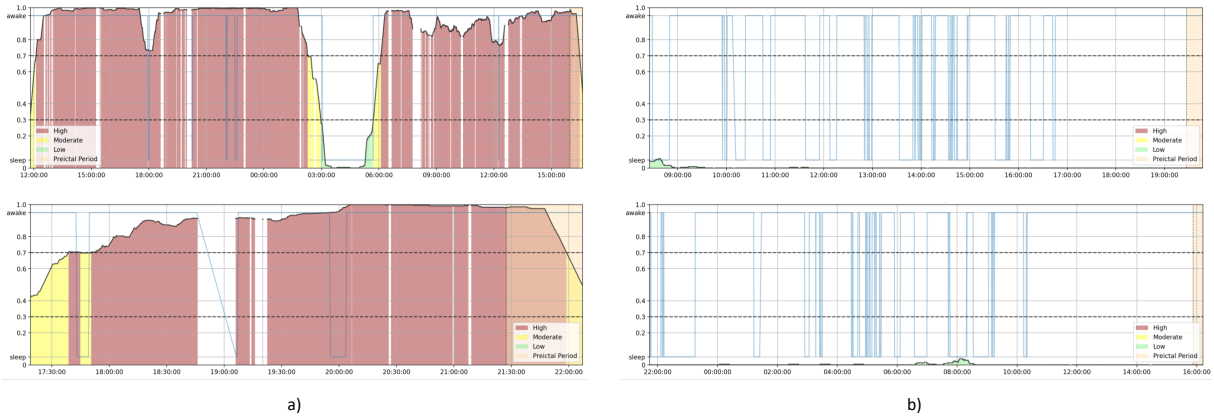


Figure 5.5: Plot of the moving average filter's output with the risk zones for seizures 4 (top) and 5 (bottom) of patient 23902 a) and 123902 b) using the Shallow Neural Network model. The light blue curve represents the sleep/awake state.

There is even a sizeable difference between the same patient's results. Figure 5.6 shows two seizures from patient 98202, and the same problem transpires. On one seizure, the algorithm mainly classifies preictal samples; on the other, primarily interictal ones.

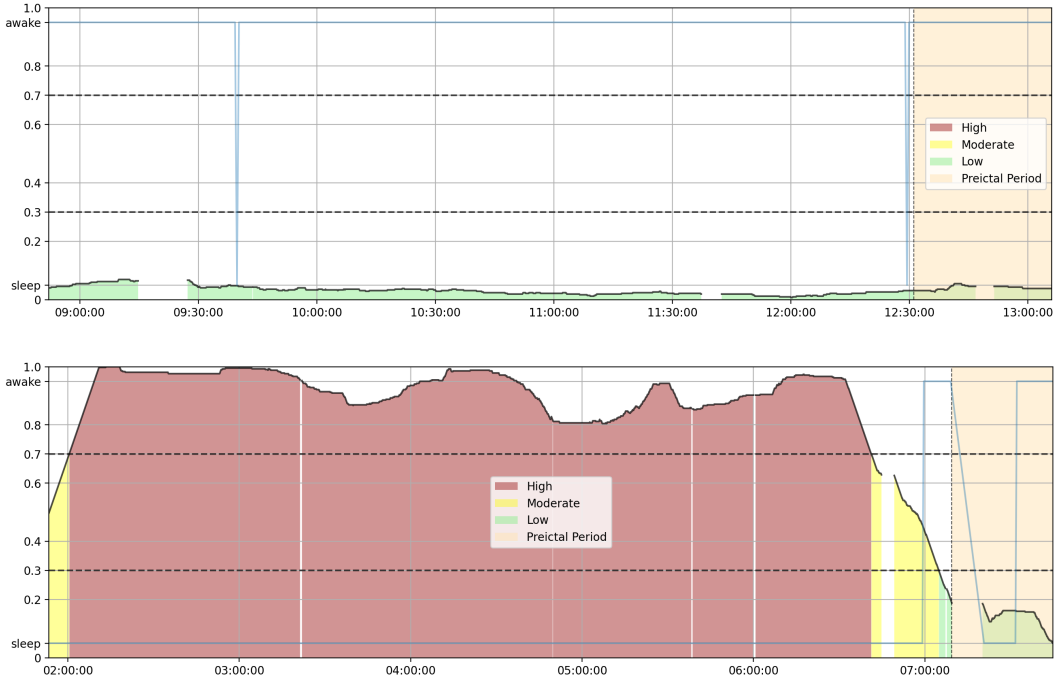


Figure 5.6: Plot of the moving average filter's output with the risk zones for seizures 4 (top) and 7 (bottom) of patient 98202 using the Shallow Neural Network model. The light blue curve represents the sleep/awake state.

An additional testing run was made for every patient with every model and altered thresholds. These were changed to 0.25 and 0.5 for forecasting and 0.5 for prediction to investigate the differences between the algorithm detecting a high-risk and medium-risk state. Tables C.1, C.2, and C.3 in Appendix C display the full results, and Table 5.8 has the average values for all applied metrics.

Table 5.8: Overall results for the three Machine Learning pipelines with a high-risk/alarm threshold of 0.5. SVM stands for Support Vector Machine, SNN for Shallow Neural Network, SS for Sensitivity, FPR/h for False Positive Rate per Hour, TiW for Time in Warning, BS for Brier Score, BSS for Brier Skill Score, and IoC for Improvement over Chance.

	Prediction			**Forecasting**					
Model	**SS**	**FPR/h**	**IoC SS**	**SS**	**TiW**	**BS**	**BSS**	**IoC SS**	**IoC BS**
Logistic Regression	0.24±0.30	1.96±5.62	25% (10 in 40)	0.44±0.37	0.28±0.21	0.19±0.11	0.01±0.15	40% (16 in 40)	50% (20 in 40)
15 SVMs Ensemble	0.32±0.37	1.68±1.95	27.5% (11 in 40)	0.51±0.39	0.34±0.22	0.24±0.13	-0.01±0.14	40% (16 in 40)	45% (18 in 40)
15 SNNs Ensemble	0.50±0.40	3.42±4.95	40% (16 in 40)	0.68±0.36	0.51±0.27	0.41±0.21	-0.03±0.23	50% (20 in 40)	45% (18 in 40)

Starting with prediction, it is clear that the new threshold, while increasing the SS, also increases the FPR/h up to 444%. Therefore, the seizure prediction algorithm using a 0.5 alarm threshold is unsuitable for clinical use. There was also a significant increase in the number of statistically validated patients. These results, that ranged from 12.5% to 20%, now range from 25% to 40%

Moving to forecasting, as expected, lowering the threshold to a medium-risk one increased both the SS and TiW. The SS increased up to 59%, with the SNN ensemble achieving a value of 0.68. However, this model also displayed a TiW of 0.51, meaning that the algorithm is half of the time in, at least, medium risk. As the BS and the BSS are not dependent on any threshold, as they only consider the model's forecasts, these scores showed no changes. Accordingly, the statistical validation using the BS remained unchanged. On the other hand, statistical validation using SS increased from a 32.5% to 40% interval to a 40% to 50% one.

5.3 Comparison with Other Studies

5.3.1 Seizure Prediction

The results of the proposed best methodologies were compared with four studies selected among the ones mentioned in Chapter 3. Table 5.9 exhibits the performance and validation results for the best approaches of the considered studies and all developed methodologies.

Table 5.9: Seizure prediction performance for studies under comparison. The models marked with * indicate the best proposed methodologies. SVM stands for Support Vector Machine, SNN for Shallow Neural Network, SS for Sensitivity, FPR/h for False Positive Rate per Hour, and IoC for Improvement over Chance.

Study	No. of Patients	SS	FPR/h	IoC
Lopes et al. (2023) [85]	41	0.34	0.90	51%
Pinto et al. (2022) [92]	93	0.16	0.21	32%
Pinto et al. (2021) [95]	19	0.37	0.79	32%
Alvarado-Rojas et al. (2014) [106]	53	0.47	0.94	13%
Logistic Regression*	**40**	**0.13**	**0.36**	**12.5%**
15 SVM Ensemble	**40**	**0.13**	**0.73**	**17.5%**
15 SNN Ensemble*	**40**	**0.34**	**2.30**	**20%**

The chosen studies used the EPILEPSIAE database and implemented statistical validation. Except for Lopes et al. (2023) [85], the selected studies took the same approach as the present work regarding the determination of the SOP by testing a range of values.

By observing the performances of the selected studies, it is notable that superior SS values were obtained compared to both approaches. The SNN ensemble approach is the exception, achieving a higher SS (0.34) than Pinto et al. (2022) [92] (0.16). Regarding the FPR/h, the Logistic Regression and SVM ensemble ranked among the best, only outperformed by Pinto et al. (2022) [92]. However, the FPR/h of 2.30 obtained by the proposed SNN ensemble methodology was, by far, the largest in the selection, followed by Lopes et al. (2023) [85] with a value of 0.87.

Concerning statistical validation, two strategies were used. Alvarado-Rojas et al. (2014) [106] used a random predictor, and the remaining three studies used a surrogate time series analysis. Most studies displayed a higher IoC than both best proposed methodologies. However, Alvarado-Rojas et al. (2014) [106] (13%) performed worse than the SNN ensemble model (20%). The highest IoC was 51%, achieved by Lopes et al. (2023) [85].

Particular attention should be given to Lopes et al. (2023) [85], as the data used in this thesis was also used in the referenced study. The Logistic Regression model achieved a FPR/h value of 0.36, much lower than the 0.90 from the study. However, it also came with a significantly lower SS (0.13 compared with 0.34). The SNN ensemble model matched the SS at 0.34 but caused a considerable increase in the FPR/h to 2.30. Furthermore, statistical validation of the present work was also considerably lower than Lopes et al. (2023) [85] for every model (12.5% to 20% compared with 51%).

It is worth noting that comparing studies in seizure prediction poses a significant challenge due to substantial heterogeneity in patient selection and the wide range of parameters and options incorporated throughout the development of a ML methodology.

5.3.2 Seizure Forecasting

The results of the proposed best methodology were compared with six studies selected among the ones mentioned in Chapter 3. Table 5.10 exhibits the performance and validation results for the best approaches of the considered studies and all developed methodologies.

Table 5.10: Seizure forecasting performance for studies under comparison. The models marked with * indicate the best proposed methodologies. SVM stands for Support Vector Machine, SNN for Shallow Neural Network, SS for Sensitivity, TiW for Time in Warning, BSS for Brier Skill Score, and IoC for Improvement over Chance.

Study	Database	No. of Patients	SS	TiW	BSS	IoC
Viana et al. (2022) [89]	ZUH KCL's clinical trial	6	0.74	0.31	-	83%
Pal Attia et al. (2022) [90]	ZUH KCL's clinical trial	6	0.69	0.37	-	67%
Proix et al. (2021) [73]	NeuroPace	18	-	-	0.23	83%
Nejedly et al. (2019) [72]	NeuroVista Canines	4 (dogs)	0.79	0.18	-	100%
Karoly et al. (2017) [17]	NeuroVista	9	0.55	0.25	0.05	100%
Cook et al. (2013) [60]	NeuroVista	10	0.61	0.23	-	90%
Logistic Regression*	**EPILEPSIAE**	**40**	**0.28**	**0.13**	**0.01**	**50%**
15 SVM Ensemble	**EPILEPSIAE**	**40**	**0.32**	**0.19**	**-0.01**	**47.5%**
15 SNN Ensemble	**EPILEPSIAE**	**40**	**0.58**	**0.44**	**-0.03**	**45%**

Unlike the present work, which used the EPILEPSIAE database, all selected studies used ultra-long-term Electroencephalogram (EEG) databases. Whereas the database used for this study comprises presurgical monitoring data, all others contain daily activity recordings. These recordings contain brain patterns better associated with the real-life conditions of patients with

epilepsy and thus are better suited for seizure forecasting studies. The databases include the NeuroPace database, with data from the RNS System, a neurostimulation device [73]; the database from the Zealand University Hospital and the King's College London's clinical trial, with data from the 24/7 EEG SubQ system, recording Sub-scalp EEG (ssEEG) [89, 90]; and the NeuroVista databases, recorded using the NeuroVista Seizure Advisory System, with data both from the patients in the renowned Cook et al. (2013) [60] trial [17, 60], and four canines [72].

Not all studies use well-defined SOPs. Cook et al. (2013) [60] defined the SOP from minutes to hours, and Proix et al. (2021) [73] did not define one at all, simply using both diaries of self-reported seizures and electrographic seizure timestamps with the help of a Board-certified epileptologist analyzing the model's output to assess the performance of the model. The other studies chose 60 min [72, 89, 90] and 30 min [17]. Karoly et al. (2017) [17] used a SPH of only 1 minute, which may not give enough time for patients or users to take preventive action.

The selected studies' performances show that the proposed model achieved the lowest SS value of 0.28, followed by Karoly et al. (2017) [17] (0.55). The performance obtained by the SNN ensemble of 0.58 was higher than that of the previously mentioned study. However, this approach was discarded as the best one due to the negative BSS, meaning it performed worse than an uninformed reference. However, the TiW took the opposite trend. The value obtained by the proposed methodology was the best among all studies at 0.13, followed by Nejedly et al. (2019) [72] at 0.18. This TiW is expected considering the conservative approach taken for the high-risk threshold. Regarding the BSS, only Proix et al. (2021) [73] and Karoly et al. (2017) [17] studied this metric. Among these, the BSS obtained by the best proposed methodology was the worst at 0.01, close to the 0.05 obtained by Karoly et al. (2017) [17]. This close-to-zero score indicates that the model only performed slightly better than the reference.

Statistical validation of the selected studies was made using three strategies. Three used a surrogate analysis, two used a time-matched predictor, and one used a random predictor. Out of the two surrogate analysis techniques used in the seizure forecasting model, the one assessing the BS was chosen as it obtained better results than the one assessing SS. The best proposed methodology's statistical validation results were the lowest among the group at 50%, followed by Pal Attia et al. (2022) [90] at 63%. However, it is noteworthy that this present work's statistical validation of 50% corresponds to a total of 20 validated patients. As a comparison, five of the six selected studies have a sample size of 10 patients or under; the remaining study counts 18 total patients. The two studies with the best statistical validation were Karoly et al. (2017) [17] and Nejedly et al. (2019) [72] at 100%, with data from four dogs and nine patients, respectively.

It is noteworthy that a comparison between the present work and the selected studies is made considerably more difficult due to the chosen database. All selected studies use ultra-long-term databases, with data of up to months of regular patient activity. On the other hand, the EPILESIAE database contains hours to days of data from presurgical monitoring, where patients are placed in conditions designed to trigger seizures, and their brains are constantly in a higher risk state.

5.4 Limitations

The most significant limiting factor for this study is the database used. EPILEPSIAE is a presurgical monitoring database with only a few hours/days of data. The recordings belong to patients suffering from medication withdrawal and sleep deprivation to forcefully trigger and analyze seizures. As the days in these conditions pass and the effects of the medication leave the patients' systems, the brain function will change over time, and the number of seizures will increase. This procedure results in a more active brain state and a higher number of seizures than that characteristic of everyday life. Using ultra-long-term databases with recordings from days to months of everyday life would assess the algorithm's performance more realistically.

Moreover, this present work used fixed thresholds for low and high-risk states in seizure forecasting. A patient-specific threshold definition algorithm based on the number of seizures in each state or percentage of time in a high-risk state could significantly improve the model's performance, as it would adapt and learn from the patient's seizure frequency. However, the EPILEPSIAE database data, with insufficient seizures and interictal time per patient, does not allow this personalized approach to threshold definition.

6

Conclusion

This thesis aimed to develop methodologies capable of forecasting epileptic seizures and explore possible improvements compared to seizure prediction. To this end, three patient-tailored models with different classifiers were developed based on the most common framework in the literature for each approach to the problem in question: forecasting and prediction.

To establish a comparison to seizure forecasting, three of said methodologies concerned a prediction approach. Three Machine Learning (ML) strategies were explored, consisting of a Logistic Regression, a voting system of 15 Support Vector Machines (SVMs), and a voting system of 15 Shallow Neural Networks (SNNs). None of the three methodologies proved to be the overall best. The SNN ensemble model achieved the highest Sensitivity (SS) result at 0.34±0.35 but also had the highest False Positive Rate per Hour (FPR/h) result at 2.30±2.89. Inversely, the Logistic Regression model had the lowest FPR/h result at 0.36±0.40. However, it was accompanied by the lowest SS result at 0.13±0.26. The SNN ensemble model is not suitable for real-life applications due to the high FPR/h performance. Conversely, due to the low SS score, neither is the Logistic Regression model. Statistical validation ranged from 12.5% to 20%.

The remaining three methodologies consisted of seizure forecasting algorithms using the same three classifiers as the seizure prediction pipelines. The overall best methodology used the Logistic Regression as the classifier. This methodology achieved results of 0.28±0.37 for SS, 0.13±0.14 for Time In Warning (TiW), 0.19±0.11 for Brier Score (BS), 0.01±0.15 for Brier Skill Score (BSS), and 50% of the patients were statistically validated. The value of the BSS shows that this was the only methodology that performed better, even if only slightly, than an uninformed forecasting reference. Two surrogate analysis techniques were used for statistical validation. The SS and the BS were tested to determine whether these metrics were statistically significant and higher than chance. The surrogate time series analysis assessing the BS achieved results up to 46% better.

Comparisons between the results generated by the seizure prediction and forecasting methodologies show a solid conclusion. Looking at the problem through the lens of forecasting produces more optimistic results than seizure prediction. The average SS for the methodologies increased up to 146%, and the number of statistically validated patients likewise showed an improvement of up to 300%. The probabilistic approach of seizure forecasting entirely eliminates the presence of alarms that may even be correctly triggered by brain activity but are rendered useless as standard brain mechanisms could have terminated the seizure it detected. Allowing patients, caretakers, and clinicians to assess the probability of a seizure and decide accordingly what action to take may be more straightforward and stress-free than often triggering wrong warnings [10].

It is essential that, even though the conclusions regarding the comparison between the different fields can be extrapolated, the actual performance results should only be viewed as proof of concept, as the study was based on presurgical monitoring data. Furthermore, this work did not consider the existence of other Concept Drifts (CDs), such as circadian rhythms or the sleep-wake cycle.

Compared to the current literature, despite being closer than the prediction results, the forecasting results obtained in this thesis are still lower. Even with the shift to the forecasting approach, results are far from ideal. This discrepancy is caused by the data used, which significantly hinders the development of effective methodologies. Therefore, the developed methodology should be applied to real-life data from an ultra-long-term database, like the one used in the significant first-in-man study published by Cook et al. (2013) [60]. Such data would contribute to advancing the wider clinical acceptability of these algorithms and facilitate the evaluation and improvement of current methodologies. Including naturally occurring seizures in long-term datasets provides a realistic and comprehensive foundation for research, offering valuable insights into the challenges and opportunities in seizure forecasting.

Bibliography

[1] P. Van Mierlo, M. Papadopoulou, E. Carrette, P. Boon, S. Vandenberghe, K. Vonck, and D. Marinazzo, "Functional brain connectivity from eeg in epilepsy: Seizure prediction and epileptogenic focus localization," *Progress in neurobiology*, vol. 121, pp. 19–35, 2014. 1, 5

[2] Y. Wang and Z. Chen, "An update for epilepsy research and antiepileptic drug development: Toward precise circuit therapy," *Pharmacology & therapeutics*, vol. 201, pp. 77–93, 2019. 1, 14, 15

[3] M. Sazgar and B. F. Bourgeois, "Aggravation of epilepsy by antiepileptic drugs," *Pediatric neurology*, vol. 33, no. 4, pp. 227–234, 2005. 1, 8, 15

[4] B. K. Steiger and H. Jokeit, "Why epilepsy challenges social life," *Seizure*, vol. 44, pp. 194–198, 2017. 1, 14

[5] M. R. Sperling, "The consequences of uncontrolled epilepsy," *CNS spectrums*, vol. 9, no. 2, pp. 98–109, 2004. 1, 14, 17

[6] C. Rathore and K. Radhakrishnan, "Concept of epilepsy surgery and presurgical evaluation," *Epileptic disorders*, vol. 17, no. 1, pp. 19–31, 2015. 1, 14, 15

[7] C. Baumgartner, J. P. Koren, and M. Rothmayer, "Automatic computer-based detection of epileptic seizures," *Frontiers in neurology*, vol. 9, p. 639, 2018. 1, 17

[8] F. A. Lado and S. L. Moshé, "How do seizures stop?," *Epilepsia*, vol. 49, no. 10, pp. 1651–1664, 2008. 2, 20

[9] R. E. Stirling, M. J. Cook, D. B. Grayden, and P. J. Karoly, "Seizure forecasting and cyclic control of seizures," *Epilepsia*, vol. 62, pp. S2–S14, 2021. 2, 19, 20, 34, 35, 37, 40, 41, 42, 44, 45

[10] M. O. Baud, T. Proix, N. M. Gregg, B. H. Brinkmann, E. S. Nurse, M. J. Cook, and P. J. Karoly, "Seizure forecasting: bifurcations in the long and winding road," *Epilepsia*, 2022. 2, 18, 24, 25, 77

[11] S. Noachtar and J. Rémi, "The role of eeg in epilepsy: a critical review," *Epilepsy & Behavior*, vol. 15, no. 1, pp. 22–33, 2009. 2, 13

[12] 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. 2, 18, 22, 31, 32, 33, 35, 36, 38, 39, 41, 43, 44, 99, 100

[13] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2007. 2, 18, 21, 25, 35, 39, 97, 100

[14] L. Kuhlmann, P. Karoly, D. R. Freestone, B. H. Brinkmann, A. Temko, A. Barachant, F. Li, G. Titericz Jr, B. W. Lang, D. Lavery, *et al.*, "Epilepsyecosystem. org: crowd-sourcing reproducible seizure prediction with long-term human intracranial eeg," *Brain*, vol. 141, no. 9, pp. 2619–2630, 2018. 2, 34, 35, 37, 38, 40, 41, 42, 43, 44, 45, 97, 98, 100

[15] 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. 2, 28, 32

[16] 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. 2, 13, 43

[17] 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. 2, 24, 25, 34, 35, 37, 38, 40, 41, 42, 45, 56, 74, 75, 98

[18] 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. 2, 28

[19] 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. 2, 20, 21, 22, 23, 26, 28, 44, 67

[20] J. Klatt, H. Feldwisch-Drentrup, M. Ihle, V. Navarro, M. Neufang, C. Teixeira, C. Adam, M. Valderrama, C. Alvarado-Rojas, A. Witon, *et al.*, "The epilepsiae database: An extensive electroencephalography database of epilepsy patients," 2012. 2, 28, 35, 48

[21] J. Engel, "What can we do for people with drug-resistant epilepsy?: the 2016 wartenberg lecture," *Neurology*, vol. 87, no. 23, pp. 2483–2489, 2016. 5, 14, 15

[22] 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. 5

[23] 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. 5, 6, 8

[24] 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. 6, 7, 8

[25] S. T. Sarmast, A. M. Abdullahi, and N. Jahan, "Current classification of seizures and epilepsies: scope, limitations and recommendations for future action," *Cureus*, vol. 12, no. 9, 2020. 6

[26] J. J. Falco-Walter, I. E. Scheffer, and R. S. Fisher, "The new definition and classification of seizures and epilepsy," *Epilepsy research*, vol. 139, pp. 73–79, 2018. 6

[27] 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. 7, 8

[28] E. Foundation, "Temporal Lobe Epilepsy (TLE)." https://www.epilepsy.com/what-is-epilepsy/syndromes/temporal-lobe-epilepsy, 2019. 8

[29] A. L. Hartman, C. D. L. Devore, P. B. Kang, D. Gilbert, A. Gropman, S. Joshi, S. Partap, J. Okamoto, M. Allison, R. Ancona, *et al.*, "Rescue medicine for epilepsy in education settings," *Pediatrics*, vol. 137, no. 1, 2016. 8, 16

[30] M. Gaínza-Lein, R. Benjamin, C. Stredny, M. McGurl, K. Kapur, and T. Loddenkemper, "Rescue medications in epilepsy patients: a family perspective," *Seizure*, vol. 52, pp. 188–194, 2017. 8, 9, 16

[31] S. Jafarpour, L. J. Hirsch, M. Gaínza-Lein, C. Kellinghaus, and K. Detyniecki, "Seizure cluster: definition, prevalence, consequences, and management," *Seizure*, vol. 68, pp. 9–15, 2019. 9, 18

[32] J. Cloyd, S. Haut, E. Carrazana, and A. L. Rabinowicz, "Overcoming the challenges of developing an intranasal diazepam rescue therapy for the treatment of seizure clusters," *Epilepsia*, vol. 62, no. 4, pp. 846–856, 2021. 9, 16

[33] E. Trinka, H. Cock, D. Hesdorffer, A. O. Rossetti, I. E. Scheffer, S. Shinnar, S. Shorvon, and D. H. Lowenstein, "A definition and classification of status epilepticus–report of the ilae task force on classification of status epilepticus," *Epilepsia*, vol. 56, no. 10, pp. 1515–1523, 2015. 9

[34] A. Varsavsky, I. Mareels, and M. Cook, *Epileptic seizures and the EEG: measurement, models, detection and prediction.* Taylor & Francis, 2011. 9, 11, 12, 38

[35] M. X. Cohen, *Analyzing neural time series data: theory and practice.* MIT press, 2014. 9

[36] J. W. C. Medithe and U. R. Nelakuditi, "Study of normal and abnormal eeg," in *2016 3rd International conference on advanced computing and communication systems (ICACCS)*, vol. 1, pp. 1–4, IEEE, 2016. 9

[37] I. Osorio, H. P. Zaveri, M. G. Frei, and S. Arthurs, *Epilepsy: the intersection of neurosciences, biology, mathematics, engineering, and physics*. CRC press, 2016. 9, 10, 12, 13, 18, 23, 25

[38] S. Sanei and J. Chambers, "Introduction to eeg," *EEG signal processing*, pp. 1–34, 2007. 9, 10, 11

[39] W. Tatum, G. Rubboli, P. Kaplan, S. Mirsatari, K. Radhakrishnan, D. Gloss, L. Caboclo, F. Drislane, M. Koutroumanidis, D. Schomer, *et al.*, "Clinical utility of eeg in diagnosing and monitoring epilepsy in adults," *Clinical Neurophysiology*, vol. 129, no. 5, pp. 1056–1082, 2018. 10, 13

[40] D. Wang, D. Ren, K. Li, Y. Feng, D. Ma, X. Yan, and G. Wang, "Epileptic seizure detection in long-term eeg recordings by using wavelet-based directed transfer function," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 11, pp. 2591–2599, 2018. 10

[41] M. Sazgar and M. G. Young, "Overview of eeg, electrode placement, and montages," in *Absolute epilepsy and EEG rotation review*, pp. 117–125, Springer, 2019. 10, 11

[42] J. Duun-Henriksen, M. Baud, M. P. Richardson, M. Cook, G. Kouvas, J. M. Heasman, D. Friedman, J. Peltola, I. C. Zibrandtsen, and T. W. Kjaer, "A new era in electroencephalographic monitoring? subscalp devices for ultra–long-term recordings," *Epilepsia*, vol. 61, no. 9, pp. 1805–1817, 2020. 12

[43] J. Parvizi and S. Kastner, "Promises and limitations of human intracranial electroencephalography," *Nature neuroscience*, vol. 21, no. 4, pp. 474–483, 2018. 12

[44] J. P. Lachaux, D. Rudrauf, and P. Kahane, "Intracranial eeg and human brain mapping," *Journal of Physiology-Paris*, vol. 97, no. 4-6, pp. 613–628, 2003. 12

[45] A. K. Health, "Invasive electroencephalography (EEG) monitoring before epilepsy surgery." https://www.aboutkidshealth.ca/article?contentid=2056&language=english, 2017. 12

[46] Z. Haneef, K. Yang, S. A. Sheth, F. Z. Aloor, B. Aazhang, V. Krishnan, and C. Karakas, "Sub-scalp electroencephalography: A next-generation technique to study human neurophysiology," *Clinical Neurophysiology*, 2022. 12, 20

[47] N. Moghim and D. W. Corne, "Predicting epileptic seizures in advance," *PloS one*, vol. 9, no. 6, p. e99334, 2014. 14, 34, 35, 37, 38, 40, 41, 42, 43, 45, 97, 98, 100

[48] A. Jacoby and J. K. Austin, "Social stigma for adults and children with epilepsy," *Epilepsia*, vol. 48, pp. 6–9, 2007. 14, 17

[49] E. Perucca, "An introduction to antiepileptic drugs," *Epilepsia*, vol. 46, pp. 31–37, 2005. 14

[50] D. Schmidt and S. C. Schachter, "Drug treatment of epilepsy in adults," *Bmj*, vol. 348, 2014. 14

[51] 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," 2010. 15

[52] P. Ryvlin, J. H. Cross, and S. Rheims, "Epilepsy surgery in children and adults," *The Lancet Neurology*, vol. 13, no. 11, pp. 1114–1126, 2014. 15

[53] J. Engel Jr, "The current place of epilepsy surgery," *Current opinion in neurology*, vol. 31, no. 2, p. 192, 2018. 15

[54] G. K. Bergey, "Neurostimulation in the treatment of epilepsy," *Experimental neurology*, vol. 244, pp. 87–95, 2013. 16

[55] Y. Lin and Y. Wang, "Neurostimulation as a promising epilepsy therapy," *Epilepsia open*, vol. 2, no. 4, pp. 371–387, 2017. 16

[56] A. Wallace, E. Wirrell, and E. Payne, "Seizure rescue medication use among us pediatric epilepsy providers: a survey of the pediatric epilepsy research consortium," *The Journal of pediatrics*, vol. 212, pp. 111–116, 2019. 16

[57] S. Sharma and K. Detyniecki, "Rescue therapies in epilepsy," *Current Opinion in Neurology*, vol. 35, no. 2, pp. 155–160, 2022. 16

[58] 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. 17, 18

[59] E. Bruno, S. Simblett, A. Lang, A. Biondi, C. Odoi, A. Schulze-Bonhage, T. Wykes, M. P. Richardson, R.-C. Consortium, *et al.*, "Wearable technology in epilepsy: the views of patients, caregivers, and healthcare professionals," *Epilepsy & Behavior*, vol. 85, pp. 141–149, 2018. 17

[60] 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. 17, 18, 19, 20, 33, 34, 35, 37, 38, 40, 41, 43, 45, 74, 75, 78, 98

[61] L. Drew, "Like taking away a part of myself"—life after a neural implant trial," *Nat Med*, vol. 26, no. 8, pp. 1154–1156, 2020. 17

[62] A. Shoeibi, P. Moridian, M. Khodatars, N. Ghassemi, M. Jafari, R. Alizadehsani, Y. Kong, J. M. Gorriz, J. Ramírez, A. Khosravi, *et al.*, "An overview of deep learning techniques for epileptic seizures detection and prediction based on neuroimaging modalities: Methods, challenges, and future works," *Computers in Biology and Medicine*, p. 106053, 2022. 18

[63] H.-H. Chen and V. Cherkassky, "Performance metrics for online seizure prediction," *Neural Networks*, vol. 128, pp. 22–32, 2020. 18

[64] R. E. Stirling, D. B. Grayden, W. D'Souza, M. J. Cook, E. Nurse, D. R. Freestone, D. E. Payne, B. H. Brinkmann, T. Pal Attia, P. F. Viana, *et al.*, "Forecasting seizure likelihood with wearable technology," *Frontiers in neurology*, p. 1170, 2021. 18, 32, 34, 35, 37, 38, 40, 41, 42, 44, 45

[65] 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. 18

[66] M. Bialer, S. I. Johannessen, R. H. Levy, E. Perucca, T. Tomson, H. S. White, and M. J. Koepp, "Seizure detection and neuromodulation: A summary of data presented at the xiii conference on new antiepileptic drug and devices (eilat xiii)," *Epilepsy research*, vol. 130, pp. 27–36, 2017. 18

[67] S. B. Dumanis, J. A. French, C. Bernard, G. A. Worrell, and B. E. Fureman, "Seizure forecasting from idea to reality. outcomes of the my seizure gauge epilepsy innovation institute workshop," *Eneuro*, vol. 4, no. 6, 2017. 19

[68] P. J. Karoly, M. J. Cook, M. Maturana, E. S. Nurse, D. Payne, B. H. Brinkmann, D. B. Grayden, S. B. Dumanis, M. P. Richardson, G. A. Worrell, *et al.*, "Forecasting cycles of seizure likelihood," *Epilepsia*, vol. 61, no. 4, pp. 776–786, 2020. 20

[69] B. Schelter, R. G. Andrzejak, and F. Mormann, "Can your prediction algorithm beat a random predictor?," *Seizure prediction in epilepsy: from basic mechanisms to clinical applications*, pp. 237–248, 2008. 21, 26, 27

[70] 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*, pp. 6530–6533, IEEE, 2012. 22, 99

[71] C. Meisel, R. El Atrache, M. Jackson, S. Schubach, C. Ufongene, and T. Loddenkemper, "Machine learning from wristband sensor data for wearable, noninvasive seizure forecasting," *Epilepsia*, vol. 61, no. 12, pp. 2653–2666, 2020. 24, 34, 35, 37, 40, 41, 42, 43, 44, 45

[72] P. Nejedly, V. Kremen, V. Sladky, M. Nasseri, H. Guragain, P. Klimes, J. Cimbalnik, Y. Varatharajah, B. H. Brinkmann, and G. A. Worrell, "Deep-learning for seizure forecasting in canines with epilepsy," *Journal of neural engineering*, vol. 16, no. 3, p. 036031, 2019. 24, 34, 37, 40, 41, 42, 44, 45, 74, 75

[73] T. Proix, W. Truccolo, M. G. Leguia, T. K. Tcheng, D. King-Stephens, V. R. Rao, and M. O. Baud, "Forecasting seizure risk in adults with focal epilepsy: a development and validation study," *The Lancet Neurology*, vol. 20, no. 2, pp. 127–135, 2021. 25, 34, 35, 40, 42, 44, 45, 74, 75

[74] M. G. Leguia, V. R. Rao, T. K. Tcheng, J. Duun-Henriksen, T. W. Kjaer, T. Proix, and M. O. Baud, "Learning to generalize seizure forecasts," *Epilepsia*, 2022. 25

[75] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE transactions on knowledge and data engineering*, vol. 31, no. 12, pp. 2346–2363, 2018. 28

[76] S. Khan, L. Nobili, R. Khatami, T. Loddenkemper, C. Cajochen, D.-J. Dijk, and S. H. Eriksson, "Circadian rhythm and epilepsy," *The Lancet Neurology*, vol. 17, no. 12, pp. 1098–1108, 2018. 28

[77] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015. 32, 44

[78] E. Nurse, B. S. Mashford, A. J. Yepes, I. Kiral-Kornek, S. Harrer, and D. R. Freestone, "Decoding eeg and lfp signals using deep learning: heading truenorth," in *Proceedings of the ACM international conference on computing frontiers*, pp. 259–266, 2016. 32

[79] R. Jana and I. Mukherjee, "Deep learning based efficient epileptic seizure prediction with eeg channel optimization," *Biomedical Signal Processing and Control*, vol. 68, p. 102767, 2021. 32

[80] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018. 32

[81] Y. Xu, J. Yang, S. Zhao, H. Wu, and M. Sawan, "An end-to-end deep learning approach for epileptic seizure prediction," in *2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, pp. 266–270, IEEE, 2020. 32, 34, 35, 37, 38, 40, 41, 42, 44

[82] 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. 32, 34, 35, 37, 38, 40, 41, 42, 43, 44

[83] S. M. Usman, S. Khalid, and Z. Bashir, "Epileptic seizure prediction using scalp electroencephalogram signals," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 1, pp. 211–220, 2021. 32, 34, 35, 37, 38, 40, 41, 42, 44, 45, 98

[84] M. F. Pinto, J. Batista, A. Leal, F. Lopes, A. Oliveira, A. Dourado, S. I. Abuhaiba, F. Sales, P. Martins, and C. A. Teixeira, "The goal of explaining black boxes in eeg seizure prediction is not to explain models' decisions," *Epilepsia Open*, 2023. 33, 35, 36, 40, 42, 44, 48, 50, 52, 56

[85] F. Lopes, A. Leal, M. F. Pinto, A. Dourado, A. Schulze-Bonhage, M. Dümpelmann, and C. Teixeira, "Removing artefacts and periodically retraining improve performance of neural network-based seizure prediction models," *Scientific Reports*, vol. 13, no. 1, p. 5918, 2023. 33, 35, 36, 40, 42, 48, 50, 53, 73, 74

[86] X. Xu, Y. Zhang, R. Zhang, and T. Xu, "Patient-specific method for predicting epileptic seizures based on drsn-gru," *Biomedical Signal Processing and Control*, vol. 81, p. 104449, 2023. 33, 35, 36, 38, 40, 41, 42, 43

[87] S. Hu, J. Liu, R. Yang, Y. Wang, A. Wang, K. Li, W. Liu, and C. Yang, "Exploring the applicability of transfer learning and feature engineering in epilepsy prediction using hybrid transformer model," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023. 33, 35, 37, 38, 40, 41, 42, 50

[88] C. Li, C. Shao, R. Song, G. Xu, X. Liu, R. Qian, and X. Chen, "Spatio-temporal mlp network for seizure prediction using eeg signals," *Measurement*, vol. 206, p. 112278, 2023. 33, 35, 37, 40, 41, 42, 43

[89] P. F. Viana, M. Nasseri, J. Duun-Henriksen, A. Biondi, J. Winston, P. Martins, E. Nurse, M. Dümpelmann, G. Worrell, A. Schulze-Bonhage, *et al.*, "Seizure forecasting using minimally invasive, ultra-long-term subcutaneous electroencephalography: Individualized intrapatient models.," *Epilepsia*, 2022. 33, 35, 37, 38, 40, 41, 42, 44, 45, 74, 75, 98

[90] T. Pal Attia, P. F. Viana, M. Nasseri, J. Duun-Henriksen, A. Biondi, J. S. Winston, I. P. Martins, E. S. Nurse, M. Dümpelmann, G. A. Worrell, *et al.*, "Seizure forecasting using minimally invasive, ultra-long-term subcutaneous eeg: Generalizable cross-patient models," *Epilepsia*, 2022. 34, 35, 37, 38, 40, 42, 43, 44, 45, 74, 75, 98

[91] J. Zhang, C. Chatzichristos, K. Vandecasteele, L. Swinnen, V. Broux, E. Cleeren, W. Van Paesschen, and M. De Vos, "Automatic annotation correction for wearable eeg based epileptic seizure detection," *Journal of Neural Engineering*, vol. 19, no. 1, p. 016038, 2022. 34, 35, 37, 40, 41, 42, 43, 97, 98, 99

[92] M. Pinto, T. Coelho, A. Leal, F. Lopes, A. Dourado, P. Martins, and C. Teixeira, "Interpretable eeg seizure prediction using a multiobjective evolutionary algorithm," *Scientific Reports*, vol. 12, no. 1, pp. 1–15, 2022. 34, 35, 37, 38, 40, 41, 42, 44, 48, 50, 56, 73, 97

[93] M. Nasseri, T. Pal Attia, B. Joseph, N. M. Gregg, E. S. Nurse, P. F. Viana, G. Worrell, M. Dümpelmann, M. P. Richardson, D. R. Freestone, *et al.*, "Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning," *Scientific reports*, vol. 11, no. 1, pp. 1–9, 2021. 34, 35, 37, 38, 40, 41, 42, 43, 44, 45

[94] K. Vandecasteele, T. De Cooman, C. Chatzichristos, E. Cleeren, L. Swinnen, J. Macea Ortiz, S. Van Huffel, M. Dümpelmann, A. Schulze-Bonhage, M. De Vos, *et al.*, "The power of ecg in multimodal patient-specific seizure monitoring: added value to an eeg-based detector using limited channels," *Epilepsia*, vol. 62, no. 10, pp. 2333–2343, 2021. 34, 35, 37, 40, 41, 42, 43, 99

[95] M. Pinto, A. Leal, F. Lopes, A. Dourado, P. Martins, C. A. Teixeira, *et al.*, "A personalized and evolutionary algorithm for interpretable eeg epilepsy seizure prediction," *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021. 34, 35, 37, 38, 41, 42, 44, 56, 73

[96] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy seizure prediction on eeg using common spatial pattern and convolutional neural network," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 465–474, 2019. 34, 35, 37, 38, 40, 42, 43, 44, 50

[97] N. D. Truong, L. Kuhlmann, M. R. Bonyadi, D. Querlioz, L. Zhou, and O. Kavehei, "Epileptic seizure forecasting with generative adversarial networks," *IEEE Access*, vol. 7, pp. 143999–144009, 2019. 34, 35, 37, 38, 40, 41, 42, 43, 44, 98

[98] 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. 34, 35, 37, 38, 40, 41, 42, 43, 44, 45, 98

[99] 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. 34, 35, 37, 38, 40, 41, 42, 44

[100] 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. 34, 35, 37, 40, 41, 42, 44, 98

[101] 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. 34, 35, 37, 38, 40, 41, 42, 43, 44, 50, 53, 97, 98

[102] 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. 34, 35, 37, 40, 41, 42, 43, 44, 69, 98

[103] E. B. Assi, M. Sawan, D. K. 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)*, pp. 1–4, IEEE, 2015. 34, 35, 37, 38, 40, 41, 42, 43, 45, 97, 98

[104] J. Rasekhi, M. R. K. Mollaei, M. Bandarabadi, C. A. Teixeira, and A. Dourado, "Epileptic seizure prediction based on ratio and differential linear univariate features," *Journal of medical signals and sensors*, vol. 5, no. 1, p. 1, 2015. 34, 35, 37, 40, 42, 43, 44, 97, 98

[105] 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. 34, 35, 37, 38, 40, 41, 42, 43, 44, 45, 97, 98

[106] 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, no. 1, pp. 1–9, 2014. 34, 35, 37, 40, 42, 44, 73, 74

[107] 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. 34, 35, 37, 38, 40, 41, 42, 43, 44, 97, 98

[108] 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)*, pp. 2100–2103, IEEE, 2013. 34, 35, 37, 38, 40, 41, 43, 45, 99, 100

[109] I. Assali, A. G. Blaiech, A. B. Abdallah, K. B. Khalifa, M. Carrère, and M. H. Bedoui, "Cnn-based classification of epileptic states for seizure prediction using combined temporal and spectral features," *Biomedical Signal Processing and Control*, vol. 82, p. 104519, 2023. 35, 36, 38, 40, 41, 42, 43, 44, 98, 99

[110] 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. 36

[111] 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. 39, 97, 98, 100

[112] J. J. Howbert, E. E. Patterson, S. M. Stead, B. Brinkmann, V. Vasoli, D. Crepeau, C. H. Vite, B. Sturges, V. Ruedebusch, J. Mavoori, *et al.*, "Forecasting seizures in dogs with naturally occurring epilepsy," *PloS one*, vol. 9, no. 1, p. e81920, 2014. 41

[113] U. R. Acharya, S. V. Sree, A. P. C. Alvin, and J. S. Suri, "Use of principal component analysis for automatic classification of epileptic eeg activities in wavelet framework," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9072–9078, 2012. 41

[114] 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. 41

[115] 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. 43, 99

[116] 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. 43

[117] 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. 44

[118] 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. 44

[119] F. Lopes, A. Leal, J. Medeiros, M. F. Pinto, A. Dourado, M. Dümpelmann, and C. Teixeira, "Automatic electroencephalogram artifact removal using deep convolutional neural networks," *IEEE Access*, vol. 9, pp. 149955–149970, 2021. 49

[120] J. McIntyre, S. Robertson, E. Norris, R. Appleton, W. P. Whitehouse, B. Phillips, T. Martland, K. Berry, J. Collier, S. Smith, *et al.*, "Safety and efficacy of buccal midazolam versus rectal diazepam for emergency treatment of seizures in children: a randomised controlled trial," *The Lancet*, vol. 366, no. 9481, pp. 205–210, 2005. 51

[121] 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. 98

[122] K. Gadhoumi, J.-M. Lina, and J. Gotman, "Seizure prediction in patients with mesial temporal lobe epilepsy using eeg measures of state similarity," *Clinical Neurophysiology*, vol. 124, no. 9, pp. 1745–1754, 2013. 98

[123] E. Oja and A. Hyvarinen, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000. 99

[124] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical review letters*, vol. 50, no. 5, p. 346, 1983. 99

[125] L. D. Iasemidis, D.-S. Shiau, J. C. Sackellares, P. M. Pardalos, and A. Prasad, "Dynamical resetting of the human brain at epileptic seizures: application of nonlinear dynamics and global optimization techniques," *IEEE transactions on biomedical engineering*, vol. 51, no. 3, pp. 493–506, 2004. 100

[126] 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. 100

Appendices

A

Conference Paper

Machine Learning Methods in Seizure Prediction and Forecasting: What Is the Best Approach?

Gonçalo S. Costa[1], Mauro F. Pinto[1], and César A. Teixeira[1]

***Abstract*— Traditional treatments do not work on 33% of epileptic patients. Warning devices employing seizure prediction or forecasting algorithms could bring patients a newfound quality of life. These algorithms would attempt to detect the preictal period, a transitional moment between regular brain activity and the seizure and warn the user. Several past methodologies have been developed, triggering an alarm when detecting the preictal period, but few have been clinically applicable. Recent studies have suggested a paradigm change to seizure forecasting that takes a probabilistic approach instead of the crisp one of seizure prediction. The alarm is substituted by a constant risk assessment analysis. To the best of our knowledge, no direct comparison between prediction and forecasting using the same database has been made. This paper explores methodologies capable of seizure forecasting and compares them with seizure prediction ones. Using data from the EPILEPSIAE database, we developed several patient-specific prediction and forecasting algorithms with different classifiers (a Logistic Regression, a 15 Support Vector Machines ensemble, and a 15 Shallow Neural Networks ensemble). Results show an increase of the sensitivity in forecasting relative to prediction of up to 146% and in the number of patients displaying an improvement over chance of up to 200%. These results suggest that seizure forecasting may be more suitable for seizure warning devices than seizure prediction.**

I. INTRODUCTION

Epilepsy, characterized by recurring seizures and affecting approximately 1% of the global population, is one of the most common neurological diseases. A significant challenge faced by epilepsy patients is the apparent unpredictability of seizures [1]. About one-third of epilepsy patients suffer from Drug-Resistant Epilepsy, requiring alternatives beyond Anti-Epileptic Drugs for seizure management [2]. In these cases, the objective becomes informing the patients when a seizure will occur or estimating the seizure likelihood within a specific period before the seizure through warning devices.

One approach is seizure prediction, which entails algorithms identifying pre-seizure moments to alert patients of an imminent seizure. Though conceptually optimal, current performance results are low for real-life applications, as seizure generation is very complex [3]. So, as of today, seizure forecasting offers a more viable option [4]. Here, the algorithm displays a continuous seizure likelihood without guaranteeing its occurrence. By taking a probabilistic approach, forecasting avoids the binary view of prediction, which is incompatible with any variation. In prediction, missing the SOP by a few minutes, or even seconds, results in a false alarm.

The Electroencephalogram (EEG) emerges as the prime biosignal for an epilepsy diagnosis. Segmentation into preictal (preceding a seizure), ictal (during a seizure), postictal (succeeding a seizure), and interictal (between the postictal and preictal periods of consecutive seizures) periods aids in identifying distinctive brain patterns, specifically those in the preictal phase [5]. Recently, the focus has shifted towards forecasting algorithms, as evidenced by a growing body of research in this direction. However, an essential step involves directly comparing prediction and forecasting performance on the same dataset, which is yet to be explored [3], [4].

This study bridges this gap by assessing the transition from seizure prediction to forecasting using patient-specific Machine Learning algorithms with multiple classifiers.

II. METHODS

We developed three patient-specific prediction methods and three for forecasting, using a Logistic Regression, a 15 Support Vector Machines (SVMs) ensemble, and a 15 Shallow Neural Networks (SNNs) ensemble as classifiers.

The EEG data from a subset of 40 patients from the EPILEPSIAE database [6] underwent preprocessing suing Convolutional Neural Network that mimics experts' behavior in identifying and removing artifacts [7]. We extracted 59 univariate linear features from time and frequency domains using a 5-second sliding window [8]. The first three chronological seizures per patient served as training data. We performed class balancing, standardization, and feature selection and trained the model to differentiate preictal and interictal samples. A grid search determined optimal training parameters: preictal length for every classifier, number of selected features for the Logistic Regression and SVM ensemble, and SVM cost value for the SVM ensemble.

We used the remaining seizures per patient for testing (out-of-sample testing). In the forecasting pipeline, we applied a moving average filter to smooth the classifier's output and set a high-risk threshold of 0.7 and a moderate-risk threshold of 0.3. We evaluated the performance by assessing the Sensitivity (SS), Time In Warning (TiW), Brier Score (BS), and Brier Skill Score (BSS) [4]. In the prediction pipeline, we employed the Firing Power [9] with an alarm threshold of 0.7 [10] and evaluated the performance by assessing the SS and the False Positive Rate per Hour (FPR/h) [11].

To gauge model performance against chance and quantify the percentage of patients that show Improvement over

*This research is funded by Foundation for Science and Technology, I.P./MCTES through project RECoD - PTDC/EEI-EEE/5788/2020 financed with national funds (PIDDAC) via the Portuguese State Budget

[1]University of Coimbra, Center for Informatics and Systems of the University of Coimbra, Department of Informatics Engineering, 3030-290, Coimbra, Portugal (glpscosta00@gmail.com, {mauropinto,cteixei}@dei.uc.pt)

Chance (IoC), we adopted a surrogate time series analysis. The model is statistically validated if its performance surpasses the surrogate one with statistical significance. This assessment is conducted under the null hypothesis that "the performance of each model is not higher than the chance level ($\alpha = 0.05$)."

III. RESULTS AND DISCUSSION

After finding the optimal parameters for every patient in each methodology, we used the testing seizures to evaluate the performance of the algorithms. Table I displays the average values for all metrics and the statistical validation results for the prediction and forecasting algorithms.

In this work, we evaluate the shift from a prediction paradigm to a forecasting one by comparing the models' SS values and IoC percentage. We can observe an increase of the SS in forecasting relative to prediction of 115% using the Logistic Regression, 146% using the SVM ensemble, and 70% using the SNN ensemble. Likewise, the IoC shows improvement of 200%, 86%, and 100%, respectively. These results suggest that forecasting algorithms can produce better results than prediction algorithms. Not only do they forecast a higher percentage of seizures, they also produce more statistically significant results.

The definition of SS for both approaches can give us a reason for this considerable increase. In prediction, it is the ratio between true alarms and total seizures. A true alarm is one triggered during a predefined period in the preictal stage, the Seizure Occurrence Period (SOP) [11]. However, in forecasting, SS is the ratio of seizures in which the model outputs a high-risk state during the SOP, which could last only a few seconds, for example. So, in cases where brain dynamics start changing before the SOP, a prediction algorithm will trigger the alarm too early, which is not considered a true alarm. A forecasting algorithm would detect the high-risk state before the SOP. However, the continuous analysis makes it so that the algorithm would continue to output this information during the SOP. This distinction is essential, as researchers believe the preictal period duration could vary between seizures in the same patient [12].

IV. CONCLUSIONS

By comparing the prediction and forecasting results, we can conclude that the latter shows significant improvements in the SS of the model and the degree of statistical validation. Allowing patients, caretakers, and clinicians to assess the probability of a seizure and decide accordingly what action to take may be more straightforward and stress-free than often triggering wrong warnings. However, we used presurgical monitoring data, where patients were subject to medication withdrawal, sleep deprivation, and high stress to increase the seizure rates. So, the values for each performance metric should only be taken as a proof-of-concept. We must apply the developed methodologies to real-life data from ultra-long-term databases to achieve more realistic results [13].

TABLE I
OVERALL RESULTS FOR THE THREE MACHINE LEARNING PIPELINES

	Prediction			**Forecasting**				
Model	**SS**	**FPR/h**	**IoC**	**SS**	**TiW**	**BS**	**BSS**	**IoC**
Logistic Regression	0.13±0.26	0.36±0.40	12.5% (5 in 40)	0.28±0.37	0.13±0.14	0.19±0.11	0.01±0.15	37.5% (15 in 40)
15 SVMs Ensemble	0.13±0.21	0.73±0.85	17.5% (7 in 40)	0.32±0.37	0.19±0.17	0.23±0.13	-0.01±0.14	32.5% (13 in 40)
15 SNNs Ensemble	0.34±0.35	2.30±2.89	20% (8 in 40)	0.58±0.39	0.44±0.26	0.41±0.21	-0.03±0.23	40% (16 in 40)

REFERENCES

[1] Van Mierlo, Pieter, et al. "Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization." *Progress in neurobiology* 121 (2014): 19-35.

[2] Sazgar, Mona, and Blaise FD Bourgeois. "Aggravation of epilepsy by antiepileptic drugs." *Pediatric neurology* 33.4 (2005): 227-234.

[3] Mormann, Florian, et al. "Seizure prediction: the long and winding road." *Brain* 130.2 (2007): 314-333.

[4] Baud, Maxime O., et al. "Seizure forecasting: bifurcations in the long and winding road." *Epilepsia* (2022).

[5] Moghim, Negin, and David W. Corne. "Predicting epileptic seizures in advance." *PloS one* 9.6 (2014): e99334.

[6] Klatt, Juliane, et al. "The EPILEPSIAE database: An extensive electroencephalography database of epilepsy patients." (2012): 1669-1676.

[7] Lopes, Fábio, et al. "Automatic electroencephalogram artifact removal using deep convolutional neural networks." *IEEE Access* 9 (2021): 149955-149970.

[8] Pinto, Mauro F., et al. "The goal of explaining black boxes in EEG seizure prediction is not to explain models' decisions." *Epilepsia Open* (2023).

[9] Teixeira, C. A., et al. "EPILAB: A software package for studies on the prediction of epileptic seizures." *Journal of Neuroscience Methods* 200.2 (2011): 257-271.

[10] Pinto, Mauro F., et al. "A personalized and evolutionary algorithm for interpretable EEG epilepsy seizure prediction." *Scientific reports* 11.1 (2021): 3415.

[11] Winterhalder, M., et al. "The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods." *Epilepsy & Behavior* 4.3 (2003): 318-325.

[12] Acharya, U. Rajendra, et al. "Automated EEG analysis of epilepsy: a review." *Knowledge-Based Systems* 45 (2013): 147-165.

[13] Cook, Mark J., 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* 12.6 (2013): 563-571.

B

Features Description

This chapter provides a detailed description of the most extracted features from the state-of-the-art and the most used features in this thesis.

Univariate Linear Features

Linear features are mathematical measures that use the amplitude and phase/frequency information to extract linear dynamics from signals. These features assume that the Electroencephalogram (EEG) signal is quasi-stationary within each time window.

Statistical Moments

Statistical moments [14,47,91,92,101,104,105,107] can determine the amplitude distribution of the EEG time series. The first four statistical moments are:

- Mean (First): the central tendency of the samples' amplitude;
- Variance (Second): the dispersion of the samples' amplitude around its mean;
- Skewness (Third): the degree of asymmetries of the amplitude distribution;
- Kurtosis (Forth): the relative flatness or peakedness of the amplitude distribution [13].

Comparing the preictal period with the interictal one shows significant changes to these statistical measures [107, 111]. Specifically, the preictal stage is connected to an increase in kurtosis and a decrease in variance [105].

Hjörth Parameters

Hjörth parameters [14,101,103–105,107] detect an increase of energy in the brain caused by an intensification of brain activity. These concern activity, mobility, and complexity, which are measures of mean power, root-mean-squared frequency, and root-mean-square frequency spread, respectively. Studies [107, 111] show that during the preictal period, there is a substantial rise in the mobility and complexity of the EEG.

Autoregressive Models

Autoregressive models [14, 101, 104, 105, 107] are used to inspect neural synchronization by modeling the EEG. The way these models are applied differs between authors. Some use the modeling coefficient values as features. Others use the modeling error as a result of a seizure-generation process [107, 121].

Decorrelation Time

Decorrelation time [14, 101, 103–105, 107] represents the first zero-crossing of the autocorrelation function. It is a measure that detects repeating patterns or identifies the fundamental frequency underlying its harmonic frequencies. Additionally, decorrelation time provides information about the typical time scale of data variability and can be used to measure signal stochasticity. A value of 0 indicates that a signal is entirely stochastic or white noise. Decreases in decorrelation time have been observed before seizures, indicating its potential for detecting preictal periods [107, 111].

Frequency Band related Features

Features in the frequency domain [14, 47, 83, 89–91, 97, 98, 100–103, 105, 107, 109] can be used to capture shifts from low to high frequencies. The EEG signal is decomposed into frequency bands (delta, theta, alpha, beta, and gamma), and their spectral power is computed. These are the most widely used features and can be determined using the Power Spectral Density (PSD) of the time series within a time window. One of the ways to calculate PSD is to perform the Fast Fourier Transform (FFT) on the time series and average the squared coefficients of the frequency range of interest. This calculation assumes the signal in each window is long enough to capture the low-frequency activity of the brain but short enough to be considered quasi-stationary. Several studies [102, 105, 107, 111] suggest a transfer of PSD from lower to higher frequencies before seizure onset.

Wavelet Transform

Wavelet transform [14, 47, 101, 104, 105, 107] is a time-frequency domain transform that decomposes the signal in different resolution levels according to different frequency ranges. It can be an alternative to the FFT. Wavelet analysis decomposes a signal into time-variant frequency components. It provides a high-frequency resolution for lower frequencies and a high time resolution for higher ones, capturing sudden changes and minor details in the signal. By computing the energy of the signal components generated by the decomposition, it is possible to measure the energy in different frequency ranges. This approach can provide insight into the dominant frequency ranges of the signal and their relative strengths. The wavelet coefficients acquired from the decomposed signal can be used to compute other features. One of these is the signal energy [14, 17, 47, 60, 101, 104, 105, 107] in different frequency ranges [122].

Multivariate Linear Features

Multivariate linear features are useful in characterizing the interactions between different electrodes recording different brain regions. Given the complexity of the preictal stage and the fact that seizures are caused by brain synchronization, such features can effectively capture and quantify this state [12, 108].

Maximum Linear Cross-correlation

Maximum linear cross-correlation measures the similarity between two time series. It does so by analyzing the extent to which two identical signals are shifted by a temporal lag. A value close to 1 indicates a similar profile between the channels, while a value close to 0 indicates asynchronous signals [12, 115].

Independent Component Analysis

Independent Component Analysis (ICA) assumes that each measured signal is a linear combination of independent signals. It decomposes multidimensional data into statistically independent components that can be used to extract features [123].

Univariate Nonlinear Features

To account for the noisy and non-stationary nature of EEG signals, measures based on chaos theory can be used to describe brain dynamics. These measures are particularly useful in detecting upcoming seizures, as a decrease in chaos may indicate increased predictability before a seizure. However, nonlinear features may be too computationally expensive to be used in an online system. The following measures aim to capture the increased brain synchronization that occurs before seizures [12].

Correlation Sum and Dimension

The correlation dimension is a method for assessing the fractal dimension of a signal, which measures the space dimension occupied by signal samples. It estimates the complexity of attractors mathematically and provides valuable information about the signal. The correlation sum quantifies the probability of two state space trajectory vectors lying within a given distance of each other [124].

Entropy

Entropy [91, 94, 109] measures can be used to quantify the degree of predictability and regularity in EEG data. It is useful in detecting changes from the interictal to the preictal state, as seizures are characterized by synchronous brain activity [70].

Lyapunov Exponents

The Lyapunov exponent [47], which measures the exponential divergence of two state-space trajectories that start close to each other, is an indicator of a system's chaotic behavior. Chaos theory posits that the predictability of a system is highly dependent on its initial conditions, and the exponential divergence or convergence of nearby trajectories in the state space reflects the inherent chaos of a system [47, 125]. Iasimidis et al. (2003) [125] showed a decrease in the largest Lyapunov exponent before the seizure onset, but Mormann et al. (2005) [111] suggested an increase 30 minutes before the start of the seizure.

Dynamic Similarity Index

The dynamic similarity index [108] is a measure that calculates the degree of similarity between two segments of an EEG signal. One segment is a reference segment, usually chosen from an interictal period, while the other is a segment from a moving test window. By comparing the dynamics of these two windows, the measure can identify when the preictal period begins if the difference in dynamics exceeds a certain threshold [12, 13, 126].

Multivariate Nonlinear Features

Multivariate nonlinear features examine input from several electrodes simultaneously, intending to capture changes in synchrony using similarity and mutual information measures.

Dynamic Entrainment

Dynamical entrainment, proposed by Iasemidis et al. (2004) [125], aims to gauge the nonlinear interaction between two time series. It is rooted in chaos theory and is a multivariate adaptation of the Lyapunov exponent.

Mean Phase Coherence

Mean Phase Coherence (MPC) [14, 108] is a method for evaluating the level of phase synchronization between two time series, outputting values within the range of 0 to 1. Higher values indicate a greater degree of synchronization [13, 108]. Rabbi et al. (2013) [108] documented a decrease in MPC values immediately before seizure onset.

C

Additional Testing Results

This chapter provides tables containing the testing results with a high-risk/alarm threshold of 0.5 for the three models, using a Logistic Regression, a 15 Support Vector Machine (SVM) ensemble, and a 15 Shallow Neural Network (SNN) ensemble, respectively.

Table C.1: Testing results for the Logistic Regression model with a high-risk/alarm threshold of 0.5. SS stands for Sensitivity, FPR/h for False Positive Rate per Hour, TiW for Time in Warning, BS for Brier Score, BSS for Brier Skill Score, and IoC for Improvement over Chance.

		Prediction			**Forecasting**					
Patient	**Tested Seizures**	**SS**	**FPR/h**	**IoC SS**	**SS**	**TiW**	**BS**	**BSS**	**IoC SS**	**IoC BS**
402	2	0.00	0.22		0.00	0.10	0.17	-0.05		
8902	2	0.50	0.20	X	1.00	0.19	0.08	0.22	X	X
11002	1	0.00	3.17		1.00	0.44	0.28	0.00	X	X
16202	4	0.00	0.13		0.00	0.04	0.07	-0.08		
21902	1	0.00	0.12		0.00	0.00	0.10	-0.15		
23902	2	1.00	3.31	X	1.00	0.59	0.31	0.13	X	X
26102	1	0.00	0.30		1.00	0.13	0.15	-0.02	X	
30802	5	0.80	0.41	X	0.80	0.46	0.30	-0.23	X	
32702	2	0.50	0.23	X	0.50	0.11	0.07	0.22	X	X
45402	1	0.00	1.90		1.00	0.50	0.23	0.01	X	X
46702	2	0.50	0.83	X	0.50	0.20	0.20	0.00		
50802	2	0.00	0.38		0.00	0.08	0.07	0.23		X
52302	1	0.00	35.25		0.00	0.70	0.51	-0.06		
53402	1	0.00	0.53		1.00	0.22	0.17	0.16	X	X
55202	5	0.60	0.85		0.80	0.46	0.28	-0.01	X	
56402	1	0.00	1.36		0.00	0.28	0.19	0.00		
58602	3	0.00	2.87		0.33	0.48	0.25	-0.04		
59102	2	1.00	2.31	X	1.00	0.67	0.37	-0.14	X	
60002	3	0.00	0.25		0.33	0.25	0.18	-0.09		
64702	2	0.00	0.93		0.50	0.29	0.19	0.19	X	X
75202	4	0.00	0.18		0.00	0.04	0.08	-0.07		
80702	3	0.33	1.25		0.67	0.34	0.24	0.09		X
85202	2	0.00	0.23		0.00	0.08	0.08	-0.03		
93402	2	0.50	1.24		1.00	0.49	0.31	-0.60	X	
93902	3	0.33	0.47		0.67	0.20	0.15	0.16	X	X
94402	4	0.50	2.80		0.50	0.59	0.31	0.00		

Table C.1 – *Continued from previous page*

Patient	Tested Seizures	SS	FPR/h	IoC SS	SS	TiW	BS	BSS	IoC SS	IoC BS
95202	4	0.25	0.62		0.25	0.19	0.13	0.12		X
96002	4	0.25	1.15		0.25	0.31	0.23	0.09		X
98102	2	0.50	0.25	X	0.50	0.11	0.07	0.16	X	X
98202	5	0.20	0.87		0.40	0.19	0.15	0.02		X
101702	2	0.50	0.99		0.50	0.47	0.27	0.02		X
102202	4	0.25	0.12	X	0.25	0.06	0.09	0.10	X	X
104602	2	0.00	0.95		0.50	0.31	0.20	0.09		X
109502	1	0.00	8.92		0.00	0.75	0.44	-0.02		
110602	2	0.50	0.73	X	0.50	0.26	0.19	0.10		X
112802	3	0.67	1.40	X	0.67	0.47	0.28	-0.19		
113902	3	0.00	0.52		0.00	0.14	0.17	0.01		X
114702	5	0.00	0.15		0.00	0.03	0.13	-0.06		
114902	4	0.00	0.11		0.25	0.03	0.05	0.17	X	X
123902	2	0.00	0.00		0.00	0.00	0.03	-0.08		

Table C.2: Testing results for the Support Vector Machine ensemble model with a high-risk/alarm threshold of 0.5. SS stands for Sensitivity, FPR/h for False Positive Rate per Hour, TiW for Time in Warning, BS for Brier Score, BSS for Brier Skill Score, and IoC for Improvement over Chance.

		Prediction			Forecasting					
Patient	Tested Seizures	SS	FPR/h	IoC SS	SS	TiW	BS	BSS	IoC SS	IoC BS
402	2	0.00	0.08		0.00	0.04	0.14	-0.53		
8902	2	0.50	0.25	X	1.00	0.20	0.10	0.02	X	X
11002	1	0.00	6.89		1.00	0.65	0.36	0.00	X	
16202	4	0.00	0.20		0.25	0.05	0.08	-0.01	X	
21902	1	0.00	0.67		0.00	0.28	0.21	-0.15		
23902	2	1.00	3.25	X	1.00	0.71	0.33	0.09		X
26102	1	0.00	0.46		1.00	0.19	0.15	-0.01	X	
30802	5	0.80	0.42	X	0.80	0.48	0.36	-0.25	X	
32702	2	0.00	0.11		0.00	0.03	0.07	0.03		X
45402	1	0.00	1.69		0.00	0.35	0.24	-0.05		
46702	2	0.00	0.55		0.50	0.20	0.20	-0.07		
50802	2	0.00	0.30		0.00	0.06	0.06	0.21		X
52302	1	0.00	2.17		0.00	0.50	0.31	-0.15		
53402	1	1.00	1.20	X	1.00	0.34	0.22	0.12	X	X
55202	5	0.40	1.38		0.60	0.42	0.28	-0.10		
56402	1	1.00	7.30	X	1.00	0.71	0.53	-0.02	X	
58602	3	0.33	0.05	X	0.33	0.00	0.06	-0.14	X	
59102	2	0.50	1.19		1.00	0.50	0.32	-0.14	X	
60002	3	0.00	0.54		0.33	0.51	0.29	-0.04		
64702	2	1.00	1.93	X	1.00	0.40	0.29	0.23	X	X
75202	4	0.00	0.13		0.00	0.06	0.09	-0.13		
80702	3	0.00	1.32		0.67	0.34	0.25	0.11		X

Table C.2 – *Continued from previous page*

Patient	Tested Seizures	SS	FPR/h	IoC SS	SS	TiW	BS	BSS	IoC SS	IoC BS
85202	2	1.00	4.35	X	1.00	0.56	0.34	0.02	X	X
93402	2	1.00	1.89	X	1.00	0.46	0.30	-0.15	X	
93902	3	0.00	0.42		0.33	0.14	0.13	0.02		X
94402	4	0.25	3.70		0.50	0.53	0.34	0.02		X
95202	4	0.25	0.80		0.50	0.26	0.18	0.10	X	X
96002	4	0.25	1.51		0.25	0.58	0.45	-0.04		
98102	2	0.50	0.38	X	0.50	0.13	0.10	0.09	X	X
98202	5	0.40	6.44		0.80	0.61	0.38	0.17		X
101702	2	0.50	1.82		0.50	0.40	0.23	0.03		X
102202	4	0.50	1.62		0.75	0.45	0.28	0.00		
104602	2	0.50	5.35		1.00	0.59	0.36	0.11	X	X
109502	1	0.00	0.80		0.00	0.34	0.25	-0.05		
110602	2	0.50	1.04	X	0.50	0.28	0.20	0.04	X	X
112802	3	0.67	3.25		0.67	0.75	0.62	-0.11		
113902	3	0.00	1.16		0.33	0.25	0.17	0.01		X
114702	5	0.00	0.15		0.00	0.02	0.12	-0.02		
114902	4	0.00	0.33		0.25	0.08	0.08	0.24		X
123902	2	0.00	0.00		0.00	0.00	0.04	-0.08		

Table C.3: Testing results for the SNN ensemble model with a high-risk/alarm threshold of 0.5. SS stands for Sensitivity, FPR/h for False Positive Rate per Hour, TiW for Time in Warning, BS for Brier Score, BSS for Brier Skill Score, and IoC for Improvement over Chance.

		Prediction			Forecasting					
Patient	Tested Seizures	SS	FPR/h	IoC SS	SS	TiW	BS	BSS	IoC SS	IoC BS
402	2	1.00	1.01	X	1.00	0.44	0.33	0.28	X	X
8902	2	0.00	0.34		0.50	0.43	0.32	-1.12		
11002	1	0.00	4.41		1.00	0.64	0.55	0.02	X	X
16202	4	0.25	0.20		0.25	0.15	0.18	-0.02		
21902	1	0.00	1.89		0.00	0.42	0.42	-0.07		
23902	2	0.50	11.18		1.00	0.93	0.72	0.03		X
26102	1	1.00	3.77	X	1.00	0.67	0.54	0.02	X	X
30802	5	0.20	0.35		0.80	0.26	0.16	-0.17	X	
32702	2	1.00	1.60	X	1.00	0.52	0.30	0.02	X	X
45402	1	1.00	2.79	X	1.00	0.68	0.50	0.04	X	X
46702	2	0.00	12.57		0.50	0.89	0.84	-0.10		
50802	2	0.00	0.80		0.50	0.19	0.13	0.31	X	X
52302	1	0.00	0.94		0.00	0.42	0.38	-0.11		
53402	1	1.00	3.73	X	1.00	0.69	0.53	0.03	X	X
55202	5	0.60	4.89		1.00	0.73	0.64	-0.07	X	
56402	1	1.00	12.39	X	1.00	0.84	0.71	-0.02	X	
58602	3	0.00	0.21		0.00	0.06	0.11	-0.14		
59102	2	0.50	5.27		1.00	0.85	0.72	-0.12		
60002	3	0.33	0.38		0.33	0.59	0.45	-0.07		
64702	2	0.50	0.85	X	0.50	0.32	0.30	0.18	X	X

Table C.3 – *Continued from previous page*

Patient	Tested Seizures	SS	FPR/h	IoC SS	SS	TiW	BS	BSS	IoC SS	IoC BS
75202	4	0.75	1.01	X	0.75	0.52	0.41	-0.05	X	
80702	3	0.67	0.92	X	0.67	0.49	0.43	-0.22	X	
85202	2	0.00	0.06		0.00	0.02	0.09	0.06		X
93402	2	0.00	0.51		0.50	0.23	0.17	-0.03	X	
93902	3	0.67	0.48	X	0.67	0.31	0.20	0.26	X	X
94402	4	0.50	4.70		0.75	0.70	0.54	0.03		X
95202	4	0.00	0.61		0.25	0.17	0.14	0.08		X
96002	4	1.00	3.10	X	1.00	0.85	0.56	-0.02		
98102	2	1.00	0.13	X	1.00	0.09	0.05	0.39	X	X
98202	5	0.00	0.39		0.20	0.22	0.26	-0.47		
101702	2	0.50	2.75		0.50	0.55	0.45	0.02		X
102202	4	0.75	2.51		0.75	0.64	0.51	0.00		
104602	2	0.50	1.52		1.00	0.45	0.36	-0.01	X	
109502	1	1.00	2.67	X	1.00	0.60	0.51	0.01	X	X
110602	2	1.00	2.08	X	1.00	0.68	0.45	0.06	X	X
112802	3	0.67	5.95	X	0.67	0.81	0.68	-0.09		
113902	3	1.00	25.29	X	1.00	0.96	0.70	-0.02		
114702	5	0.80	11.54		1.00	0.83	0.66	-0.01		
114902	4	0.50	1.06		1.00	0.47	0.37	0.06	X	X
123902	2	0.00	0.00		0.00	0.00	0.02	-0.06		