UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CÂMPUS CORNÉLIO PROCÓPIO
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA
MESTRADO EM ENGENHARIA ELÉTRICA

CAROLLINE VICTOR GOMES DA SILVA

UMA NOVA ABORDAGEM PARA A PREDIÇÃO DE CRISES EPILÉPTICAS BASEADA NAS TÉCNICAS DE PADRÕES ESPACIAIS COMUNS E APRENDIZAGEM DE MÁQUINA

DISSERTAÇÃO

CORNÉLIO PROCÓPIO
2022

CAROLLINE VICTOR GOMES DA SILVA

UMA NOVA ABORDAGEM PARA A PREDIÇÃO DE CRISES EPILÉPTICAS BASEADA NAS TÉCNICAS DE PADRÕES ESPACIAIS COMUNS E APRENDIZAGEM DE MÁQUINA

A NOVEL APPROACH FOR EPILEPTIC SEIZURES PREDICTION BASED ON COMMON SPATIAL PATTERNS AND MACHINE LEARNING TECHNIQUES

Dissertação apresentada como requisito para obtenção do título de Mestra em Engenharia Elétrica da Universidade Tecnológica Federal do Paraná (UTFPR).
Área de Concentração: Controle e Automação de Sistemas.

Orientador: Prof. Dr. Cristiano Marcos Agulhari
Coorientador: Prof. Dr. Paulo Rogério Scalassara

CORNÉLIO PROCÓPIO

2022

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Cornélio Procópio

CAROLLINE VICTOR GOMES DA SILVA

UMA NOVA ABORDAGEM PARA A PREDIÇÃO DE CRISES EPILÉPTICAS BASEADA NAS TÉCNICAS DE PADRÕES ESPACIAIS COMUNS E APRENDIZAGEM DE MÁQUINA

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestra Em Engenharia Elétrica da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Sistemas Eletrônicos Industriais.

Data de aprovação: 22 de Fevereiro de 2022

Prof Cristiano Marcos Agulhari, Doutorado - Universidade Tecnológica Federal do Paraná

Prof Andre Kazuo Takahata, Doutorado - Fundação Universidade Federal do Abc (Ufabc)

Prof.a Elisangela Aparecida Da Silva Lizzi, Doutorado - Universidade Tecnológica Federal do Paraná

Prof Jefferson Tales Oliva, - Universidade Tecnológica Federal do Paraná

Prof Paulo Rogerio Scalassara, Doutorado - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 22/02/2022.

AGRADECIMENTOS

Considerando a concepção dessa dissertação em meio a pandemia da COVID-19, essa seção de agradecimentos se torna ainda mais especial.

Agradeço aos meus pais, Silmara e Elcio, minha irmã Maria Gabriella e avó Neide por todo o suporte incondicional e compreensão.

À todos os meus amigos pela torcida. Em especial à minha amiga-irmã Samira, por compreender meus anseios e me fortalecer a cada momento.

Agradeço ao meu orientador Cristiano Marcos Agulhari pelos quase quatro anos de parceria – foi responsável por materializar minhas ideias e embarcar comigo nesse universo (até então inexplorado) dos sinais de eletroencefalograma. Obrigada, Cristiano, por ser sempre solícito e me dar segurança nos momentos de incertezas, mas principalmente por ser esse ser-humano e profissional tão inspirador!

Ao professor Paulo Rogério Scalassara por acompanhar de perto meus passos iniciais na pesquisa, ao longo do PIF, sempre contribuindo para o enriquecimento dos meus trabalhos. Agora como coorientador, agradeço pelas contribuições e entusiasmo a cada resultado apresentado durante nossas reuniões remotas, e por sempre ter me encorajado a dar passos maiores.

Aos amigos do LACOS e LPSA pelas conversas, onde discutíamos nossos dilemas de pesquisas e compartilhávamos ideias, que me trouxeram conforto nesse período tão difícil.

A todos os professores e demais funcionários da UTFPR-CP com os quais tive a oportunidade de conviver, aprender e pelos quais tenho profunda admiração.

Aos membros da banca, obrigada pelas contribuições desde a qualificação e pelos caminhos indicados. Em especial gostaria de agradecer à professora Elisângela Ap. da Silva Lizzi por todo o suporte desde a graduação. Nossa convivência me fez adimirá-la ainda mais e reconhecer sua tamanha importância para a representatividade feminina nas ciências exatas.

Ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Tecnológica Federal do Paraná – Campus Cornélio Procópio, pela oportunidade.

E por fim, à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) pelo apoio – Código de Financiamento 001.

Every great dream begins with a dreamer. Always remember, you have within you the strength, the patience and the passion to reach for the stars to change the world.
Harriet Tubman

RESUMO

SILVA, Carolline V. G.. **Uma nova abordagem para a predição de crises epilépticas baseada nas técnicas de Padrões Espaciais Comuns e Aprendizagem de Máquina** . 2022. 79 f. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Tecnológica Federal do Paraná, Cornélio Procópio, 2022.

A epilepsia é uma das doenças neurológicas mais comuns, caracterizada por convulsões recorrentes causadas por breves distúrbios nas funções elétricas do cérebro. Em 30% dos casos, esta condição não pode ser tratada com sucesso por medicação ou ressecção, impactando diretamente na qualidade de vida do paciente. Assim, há um interesse significativo em desenvolver ferramentas confiáveis para prever convulsões, permitindo a tomada de decisões, ou pelo menos alertar os pacientes para estarem preparados quando uma convulsão se aproxima. O método proposto para previsão de convulsões é baseado na análise tempo-frequência do eletroencefalograma de escalpo (EEG) e no uso de técnicas de filtragem espacial para extrair características capazes de discriminar as atividades interictal e preictal. Os coeficientes dos ritmos teta, alfa e beta do EEG, obtidos pela decomposição da Transformada Discreta *Wavelet*, são submetidos à técnica de filtragem dos Padrões Espaciais Comuns. Atributos estatísticos e relacionados à entropia são extraídos e, em seguida, as características são selecionadas e aplicadas no classificador SVM com *kernel* Gaussiano, a fim de discriminar o estado cerebral como preictal ou não-preictal. O algoritmo proposto é avaliado em registros de superfície multicanal de 17 sujeitos com epilepsia refratária do banco de dados do *Children's Hospital Boston and Massachusetts Institute of Technology* (CHB-MIT). Duas técnicas, Filtro de Kalman e o Filtro de Mediana, são também comparadas em uma etapa de pós-processamento para suavizar os resultados do classificador. Uma decisão final de cada época do EEG foi tomada após um processo de nivelamento. Os melhores resultados mostraram uma precisão média de 68,8% para a classificação da amostra. O gerador de alarme reportou uma taxa de falso-positivo de 0,334 por hora.

Palavras-chave: Eletroencefalograma. Wavelet. Máquina de Vetor de Suporte. Filtro de Kalman. Filtro de Mediana.

ABSTRACT

SILVA, Carolline V. G.. **A novel approach for epileptic seizures prediction based on common spatial patterns and machine learning techniques** . 2022. 79 p. Dissertation (Mestrado em Engenharia Elétrica) — Universidade Tecnológica Federal do Paraná, Cornélio Procópio, 2022.

Epilepsy is one of the most common neurological diseases characterized by recurrent seizures caused by brief disturbances in the brain's electrical functions. In 30% of the cases, this condition cannot be successfully treated by medication or resection, directly affecting the quality of life of these individuals. Thus, there is a significant interest in developing reliable tools for predicting seizures, enabling decision making, or alerting patients to be prepared when a seizure is approaching. The proposed method for seizure prediction is based on time-frequency analysis of the scalp electroencephalogram (EEG) and spatial filtering techniques to extract features capable of discriminating the interictal and preictal activities. The coefficients of the theta, alpha, and beta EEG rhythms, obtained by the decomposition of the Wavelet Discrete Transform, are subjected to the Common Spatial Patterns filtering technique. Statistical and entropy-related attributes are extracted, and then features are selected and applied to the SVM classifier with Gaussian kernel to discriminate cerebral state as preictal or non-preictal. The proposed algorithm is evaluated on multichannel surface recordings of 17 subjects with refractory epilepsy from the Children's Hospital Boston and Massachusetts Institute of Technology (CHB-MIT) database. Two techniques, namely Kalman Filter and Median Filter, are used to smooth the classifier's outputs. A final decision of each EEG epoch is yielded after a thresholding process. The best results have shown an average precision of 68.8% for sample classification. The alarm generator reported a false-positive rate of 0.334 per hour.

Keywords: Electroencephalogram. Wavelet. Support Vector Machine. Kalman Filter. Median Filter.

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS, INITIALS, AND ACRONYMS

INITIALS

ADF	Augmented Dickey-Fuller
APT	Average prediction time
BCI	Brain-Computer Interface
CHB-MIT	Children's Hospital Boston and Massachusetts Institute of Technology
CNN	Convolutional Neural Network
CSP	Common Spatial Pattern
CWT	Continuous Wavelet Transform
DF	Dickey-Fuller
DL	Deep Learning
DWT	Discrete Wavelet Transform
ECoG	Electrocorticography
EEG	Electroencephalogram
EMG	Electromyogram
FNR	False-negative rate
FPR	False-positive rate
FT	Fourier Transform
HPF	High-pass filter
iEEG	Intercranial EEG
ILAE	International League Against Epilepsy
KF	Kalman Filter
KNN	K-nearest neighbors
LDA	Linear Discriminant Analysis
LPF	Low-pass filter
MF	Median Filter
ML	Machine Learning
MLP	Multilayer Perceptron
PCA	Principal Component Analysis
PSD	Power Spectrum Density
RBF	Radial Basis Function
SNR	Signal-to-noise ratio
SPH	Seizure prediction horizon
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TBI	Traumatic brain injury
UTFPR	Universidade Tecnológica Federal do Paraná
WPD	Wavelet Packet Decomposition
WT	Wavelet Transform

SUMMARY

1 INTRODUCTION

Epilepsy is a neurological disorder characterized by a transient alteration of neuronal activities, causing a lasting predisposition to epileptic seizures (THURMAN et al., 2011). The most common epilepsy symptom is the seizure, an event defined by signals and/or symptoms stemming from abnormally excessive or synchronous neuronal activity, causing loss of awareness or consciousness and disturbances of movement, sensation, or other cognitive functions (FISHER et al., 2005).

This condition affects about 50 million people around the world. The incidence rate of epilepsy is expected to increase further due to rising life expectancy, and an increasing proportion of people surviving events that often lead to epilepsy, such as birth trauma, traumatic brain injuries (TBI), infections of the brain and stroke (WHO, 2019).

Epilepsy has a higher prevalence in children and in the elderly. Considering socio-economic aspects, three-quarters of people with epilepsy live in low-income countries and do not get the treatment they need – due to sanitary hygiene conditions, inadequate health care delivery services, and even higher risk of infections and brain infestations (DUNCAN et al., 2006). However, up to 70% of people with epilepsy could become seizure-free with the appropriate use of cost-effective antiseizure medicines (WHO, 2019).

For the other 30% of cases, a condition called refractory epilepsy, treatment possibilities include brain surgery and electrical brain stimulation. Surgical resection of epileptic parts of the brain may be proposed if the seizure focus can be located and safely removed. However, not all patients are suitable surgical candidates, and surgery does not always provide a complete absence of seizures (TÉLLEZ-ZENTENO; DHAR; WIEBE, 2005). One of the most debilitating aspects of epilepsy is the unpredictability of seizure onset. Despite medical costs linked with the treatment of epilepsy, the injuries resulting from uncontrolled seizures represent an even higher cost to the society (STRZELCZYK et al., 2013).

Electroencephalogram (EEG) recordings by the electroencephalography method are a helpful tool for diagnosing and managing epileptic syndromes. The measurement of electrical potentials can be done either on the scalp or directly on the cortex, in intracranial areas of the brain with invasive EEG. The scalp EEG is an inexpensive and non-invasive method in which recordings are made through electrodes arranged in well-

defined positions, the most commonly used of which is the 10-20 system (LIBENSON, 2010).

Spikes and sharp waves distinguish epilepsy in the EEG, usually on almost all channels. From the appearance of these morphologies, the epileptic seizure events can present the following states: interictal (characterized by the absence of seizures), preictal (starting before the seizure, ending the interictal state), ictal (through the seizure), postictal (immediately after the seizure, occurring or reestablishment of normal interictal conditions) (FISHER et al., 2005). Also, similar patterns can be found in normal and interictal EEG segments, which makes it difficult to distinguish between the two in the time domain (OLIVA; ROSA, 2017).

The primary goal of seizure prediction is to distinguish the preictal period from the rest of these states. In contrast, seizure detection comprises the proper identification of the ictal state among these four states.

1.1 PREDICTABILITY OF SEIZURES

Many of the early studies focused on seizure detection rather than prediction - while seizure detection aims to identify seizures shortly before or after onset, seizure prediction seeks to recognize seizures minutes in advance. According to Acharya, Sree, et al. (2013), better processing for seizure detection can be handled by single-channel or multi-channel processing techniques.

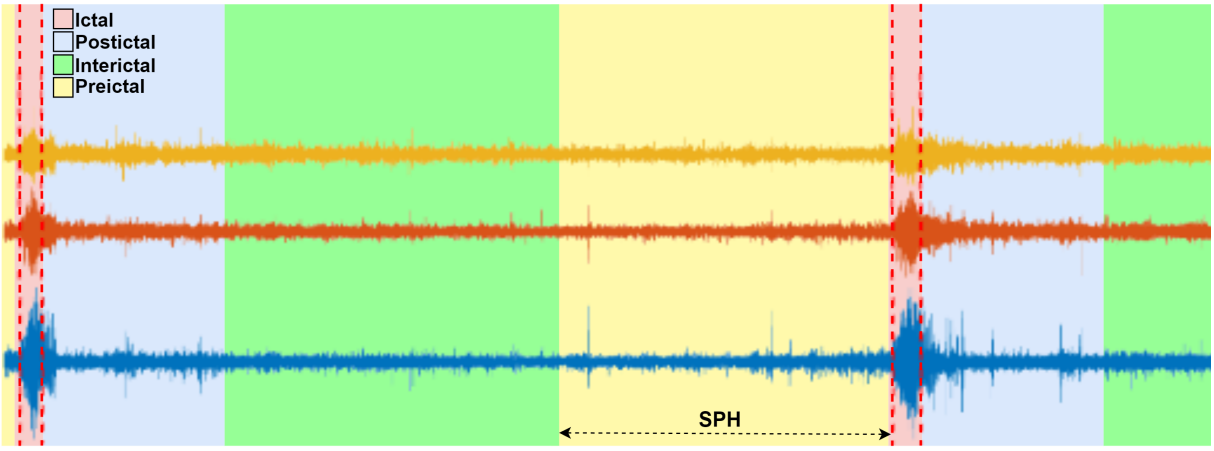
At the beginning of the present century, after the work of Adeli, Z. Zhou, and Dadmehr (2003), meaningful advances have been seen in the automated diagnosis and detection of epilepsy. Several studies have been conducted on EEG seizure detection (SIDDIQUI et al., 2020). Their performance is considerably higher since detecting an impending seizure is easier than predicting it before onset - making seizure prediction a more challenging problem.

In the biomedical engineering field, the development of techniques for the problem of predicting epileptic seizures onset has the goal of significantly contributing to improving diagnostic and therapeutic methods, ensuring the quality of life for many patients (ACHARYA; HAGIWARA; ADELI, 2018). On the other hand, researchers are still far from a complete and reliable approach that can practically be used in actual medical applications (FREESTONE; KAROLY, P. J.; COOK, 2017).

The seizure prediction system is often treated as a classification problem whose objective is to distinguish preictal and interictal brain activities. By neuroscientific knowledge, the length of the preictal period is unclear, ranging from several minutes to hours, so it has often been left as a design choice (BANDARABADI; RASEKHI, et al., 2015; MORMANN; ANDRZEJAK, et al., 2007).

The prediction system is expected to trigger an alarm when a threshold is crossed, indicating the possible approach to a seizure event. If the alarm is within the time interval of the seizure prediction horizon (SPH), it is classified as a true alarm (true positive). Otherwise, if triggered within the interictal period, it is considered a false alarm (false-positive) (RASHEED et al., 2021). Figure 1 illustrates a typical SPH from a 3-channel EEG.

Figure 1 – Concept of seizure (preictal) prediction horizon (SPH) in a 3-channel EEG



Source: Authors (2022).

Patients with refractory epilepsy would approve of a prediction algorithm if it could satisfactorily provide a better quality of life. A reliable alarm system would allow the management of preventive interventions, possibly dealing with the seizure before it happens (WINTERHALDER et al., 2003).

1.2 PAST EFFORTS IN SEIZURE PREDICTION

The interest in detection of the preictal/ictal stages started during the 1970s with Viglione and Walsh (1975), to identify epileptic seizure precursors with a linear approach. Later, using nonlinear models, Rogowski, Gath, and Bental (1981) and Salant, Gath, and Henriksen (1998) could better represent the nonlinear characteristic of EEG recordings.

Over the past decade, a wide variety of methods have been proposed to perform the task of detecting the preictal stage from an assortment of EEG datasets, but achieving high sensitivity and a low false-positive rate remains a significant challenge (USMAN, S. M.; KHALID; AKHTAR, et al., 2019; ACHARYA; HAGIWARA; ADELI, 2018; ALOTAIBY; ALSHEBEILI; ALSHAWI, et al., 2014). Notable among them are traditional Machine Learning (ML) and Deep Learning (DL) models that include the steps of EEG signal preprocessing, feature extraction, classification between preictal and interictal seizure states, and postprocessing for decreasing false positives.

The preprocessing step by attribute extraction have greatly influence the maximization of prediction time and true positive rate. Many approaches enable signal decomposition on their temporal and spectral components. Several researchers have used the Wavelet Transform for the EEG signals preprocessing, where mainly Discrete Wavelet Transform (DWT) features are used to extract information from signals in different frequency sub-bands (ADELI; ZHOU, Z.; DADMEHR, 2003). This technique is very efficient when it is desired to detect seizure onset (OCAK, 2009; KHAN et al., 2017).

Gadhoumi, Lina, and Gotman (2013), through signal decomposition into wavelet coefficients and entropy measurements, obtained significant results in 7 out of 17 patients for a 20-minute preictal period. The classifier training method selected the best channels, which were then submitted to cross-validation. The algorithm achieved sensitivities of 85% and a false positive rate of 0.1/h.

Instead of tasking the classifier to select the best channels, it is possible to handle multichannel problems using other techniques. One such method, which can result in a better signal-to-noise ratio (SNR), removing internal noise from the EEG signal, and decreasing computational cost by reducing the number of channels is Common Spatial Patterns (CSP) filtering. CSP is a feature extraction algorithm used in different applications, such as EEG signal analysis for motor imagery purposes (WANG, Y.; GAO, S.; GAO, X., 2006), seizure detection (FU et al., 2020) and prediction (ALOTAIBY; ALSHEBEILI; ALOTAIBI, et al., 2017; ZHANG, Y. et al., 2019). The main idea is to use a linear transformation to project multichannel EEG data into low-dimensional subspaces to generate spatial pattern vectors. This transformation can maximize the variance of one class and minimize the variance of another, being a useful tool for discriminating between different EEG activities. In contrast, the Principal Component Analysis (PCA) maximizes the signal variance (WANG, Y.; GAO, S.; GAO, X., 2006).

A challenging issue is using spatial filters to extract robust and representative features for seizure discrimination. One example of CSP application for seizure prediction is from Alotaiby, Alshebeili, Alotaibi, et al. (2017), who extracted features from the CSP projection matrix for training a linear discriminant classifier, being able to achieve an average accuracy of 89%.

Syed Muhammad Usman, Muhammad Usman, and Fong (2017) have applied CSP in a pre-processing step Children's Hospital Boston and Massachusetts Institute of Technology (CHB-MIT) database (GOLDBERGER et al., 2000). Then, a surrogate channel was obtained to increase SNR and high variance between preictal and interictal states. Empirical Mode Decomposition has applied to the surrogate channel, and spectral and statistical features were extracted and fed into three classifiers: Naive Bayes, k-nearest neighbors (KNN) and Support Vector Machine (SVM) (the latter presenting the best performance).

The most common classifier used to distinguish between preictal and interictal stages has been the Support Vector Machine (USMAN, S. M.; KHALID; AKHTAR, et al., 2019; RASHEED et al., 2021). Yun Park, Luo, et al. (2011) used cost-sensitive SVM on measures derived from frequency analysis, namely spectral density at various frequencies, in 18 patients from the Freiburg database. The assumed preictal period was 30 minutes, and sensitivity values of 98.3% are reported in this study.

Teixeira et al. (2014) developed a study on seizure prediction in continuous and long-term intracranial EEG recordings of 280 patients. The method comprised the analysis of 22 univariate features per channel, being 6 channels per subject. In this approach, the seizure occurrence periods were optimized for each patient. The method was validated on three machine learning methods: SVM, Multilayer Perceptron and Radial Basis Function (RBF) architectures. In this study, the temporal dynamics of the samples were emphasized through the development of the "Firing Power method", which smooths the output of the classifiers, reporting 100% sensitivity for 16% of the individuals.

Deep Learning methods result from advancements in Machine Learning research. DL methods comprise multiple layers of computational (non-linear) modules that work mutually to process large amounts of data and extract features for classification tasks. Current studies have also successfully used Convolutional Neural Networks (CNN) for classification. Yuan Zhang et al. (2019) proposed a CNN for patient-specific seizure

prediction, using also CSP and Wavelet Packet Decomposition (WPD) to reduce the input dimension of the network. The resulting classifier achieved 92.2% accuracy with a 30 minutes early prediction window for 23 subjects from the CHB-MIT EEG dataset.

Convolutional neural networks are proving to be a suitable feature extraction method. Syed Muhammad Usman, Khalid, and Aslam (2020) used CNN-based features as input for three types of classifiers: Naïve Bayes, KNN, and SVM. The SVM topology gave the highest sensitivity compared to others, predicting seizures 23.6 min before onset with 92.2% sensitivity and 90.8% specificity.

Deploying neural networks in a portable system is still a challenge. General-purpose Central Process Units (CPUs), even with their architectural transformation over the years, cannot fulfill the high computational demand of deep learning models (AJANI; IMOIZE; ATAYERO, 2021).

Therefore, implementing these classifiers may require adaptations that considerably affect system performance, especially within resource-limited embedded and mobile environment applications. Classification models ratify these requirements since they are simple models, such as logistic regression and decision trees. Some models have already been successfully employed in embedded applications, such as Multilayer Perceptron (MLP) and SVMs (HUSSAIN; PARK, S. J., 2020; CONDORI; URQUIZO; DIAZ, 2016; RÚA et al., 2012).

A post-processing step is usually needed to decide, given a temporal sequence of classifications, if an alarm is triggered. Different post-processing researches have applied Kalman filtering (CHISCI et al., 2010; PARK, Y.; LUO, et al., 2011), moving mean/median filters (ALOTAIBY; ALSHEBEILI; ALSHAWI, et al., 2014), and statistical validation methods (KIRAL-KORNEK et al., 2018). Alotaiby, Alshebeili, Alshawi, et al. (2014) used a seventh-order median filter with a patient-dependent threshold approach to smooth Linear Discriminant Analysis (LDA) classifier's results, reporting a 89% average sensitivity and 0.39/h average false prediction rate.

The present work proposes to bring potential improvements to current methods by applying two processing techniques, DWT and CSP, along with an SVM model to detect seizures before onset with sufficient time. Post-processing techniques must smooth the temporal dynamics of the classifier outputs, minimizing the false alarm rate. The proposed procedure may enable the future development of low-cost devices that can improve the quality of life of refractory epilepsy patients living in economically

disadvantaged areas.

1.3 HYPOTESIS AND OBJECTIVES

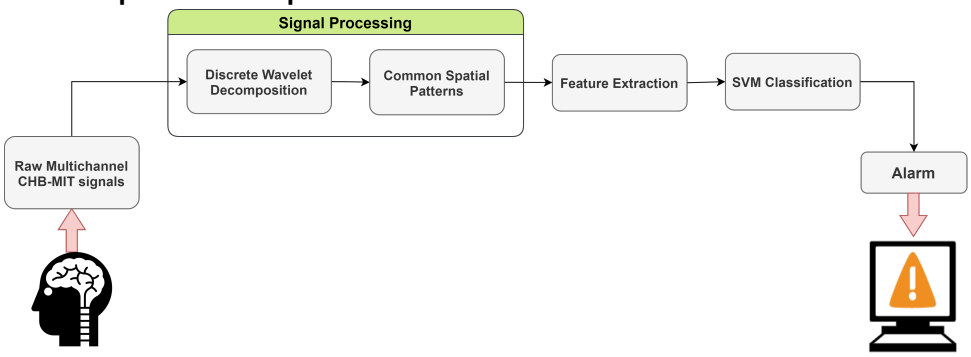
This work hypothesizes that integrating the Wavelet Decomposition and Common Spatial Patterns filtering techniques for processing EEG signals, focusing on the SVM classifier outputs regularization, can enable the development of an efficient seizure prediction system, both in terms of implementation cost and precision. Therefore, this study aims to contribute to the development of methodologies with appropriate performance to be integrated into a transportable device to alarm the impending seizures or just occurring seizures in refractory epileptic patients in real-time.

From the hypothesis, the following specific objectives were defined:

- Perform the time-frequency analysis of the EEG signals in preictal and interictal states, the Discrete Wavelet Transform is used to extract the features of CHB-MIT database signals in the frequency sub-bands of interest (theta, alpha, and beta);
- Extract statistical attributes from Wavelet Decomposition coefficients' resulting CSP filtering process to train a patient-independent SVM structure to discriminate between interictal and preictal periods.
- Develop an alarm protocol by applying techniques for post-processing the model outputs and threshold evaluation.

The general workflow of a epileptic seizure prediction model is presented in Figure 2.

Figure 2 – Seizure predictor simplified workflow



Source: Authors (2022).

All of the algorithms of this thesis were simulated using the MATLAB software

package from MathWorks Inc. The summary of our contribution to these epileptic seizure problems is detailed as follows.

1.4 LIMITATIONS

There are some limitations linked to databases or the behavior of epilepsy. Analyzing the definition of the preictal state, as previously described, the prediction of seizures implies the detection of the preictal period, which is not well defined or characterized in the literature, being therefore considered an ambiguous state. This results in one of the critical challenges to seizure prediction since it implies identifying something that is conceptually difficult to define because it varies substantially among patients. The choice of preictal length is an essential issue with seizure prediction algorithms and can differ from patient to patient, and even between a patient's seizures (BANDARABADI; RASEKHI, et al., 2015).

The significance of publicly available datasets is that they provide a reference to analyze and compare results with others. The availability of continuous, annotated, long-term datasets allows researchers to generate hypotheses and statistically validate prospective algorithms (KUHLMANN et al., 2018). Many research groups have used non-public data so that other groups' results are not reproducible. There is no widely available open-source multichannel EEG database covering numerous epilepsy syndromes and different seizure types, and providing clues about the development of epilepsy with various therapies. As a consequence, comparisons are still scarce.

1.5 TEXT OUTLINE

This dissertation presents the results from the master's research and is organized into five chapters.

This first chapter contextualizes and characterizes the problem of seizure prediction. With this, the motivations for implementing a method for prediction using multichannel EEG recordings are highlighted.

In Chapter 2, the theoretical and technical aspects of seizure prediction are presented. It starts with an introduction on EEG recordings: origin, brain rhythms, technical aspects necessary for signal acquisition, brain rhythms, abnormalities, among

others. The key concepts of epilepsy are then briefly explored. Subsequently, the main characteristics of EEG signals are explained. Next, a brief description of the primary techniques used in the study is presented: Discrete Wavelet Transform, Common Spatial Patterns Filtering, Support Vector Machine classifier, and post-processing techniques for alarm generation.

The main contributions to seizure prediction are demonstrated in Chapter 3. All the steps of the adopted methodology are covered, and the performance metrics in seizure prediction are addressed.

In Chapter 4, quantitative and qualitative analyses of the results of the application of the proposed methodology are presented, as well as a comparison with some studies in the literature.

Finally, Chapter 5 comprise the last considerations about the study. Suggestions for future work and ways to continue this research are also presented.

2 BACKGROUND

This chapter comprises a survey of the relevant concepts that guide the development of this work. The literature review covers the main biological and mathematical properties of the electroencephalogram signals, the Discrete Wavelet Transform and Common Spatial Patterns processing techniques for feature extraction, the Support Vector Machine classifier, and the post-processing approaches for the prediction protocol.

2.1 OVERVIEW ON ELECTROENCEPHALOGRAM

Electroencephalography is a method that has enabled the development of a medical technique for measuring the brain's electrical activities through the scalp, which results in the reading is called electroencephalogram. The EEG was developed to be a tool for clinical neurological and neurophysiological research because of its ability to detect brain activity (FREEMAN; QUIROGA, 2012).

Electroencephalogram signals reflect the electrical activity of large populations of brain neurons. The synapse's electric current is generated in the dendrite when these neurons are activated. This current generates a magnetic field measurable by electromyogram (EMG) machines and a secondary electrical field over the scalp, measurable by EEG systems (SANEI; CHAMBERS, 2013).

Many electrodes are placed on the scalp surface during the EEG acquisition procedure. Each electrode detects tiny electrical charges resulting from the brain cells' activity. The electric differentials between the electrodes are amplified, converted into wavy lines on a computer screen, and then recorded (SIULY; LI; ZHANG, Y., 2016).

Even with patients under similar circumstances, the behavior of the captured EEG signal is different. For this and other reasons, unlike other bioelectrical signals such as the electrocardiogram, the encephalogram recording is irregularly, thus not presenting an exact pattern.

2.1.1 Brain Rhythms

Spectral analysis of the EEG shows specific peaks, and studies about the EEG suggest that characteristics of the waveforms can also be associated with specific mental

states (SANEI; CHAMBERS, 2013).

The frequency components of different waves seen in a routine EEG recording are usually employed for analysis: delta, theta, alpha, beta, and gamma. Waveform activities differ according to the brain function related to the specific mental and physical tasks. For instance, low-frequency waves (delta and theta) predominate during sleep. In contrast, an EEG signal acquired during awake periods includes a higher percentage of high-frequency waves (alpha and beta) (HENRY, 2006).

Frame 1 presents the behavior of the main brain rhythms, organized from the lowest to the highest frequency, as well as their amplitude levels.

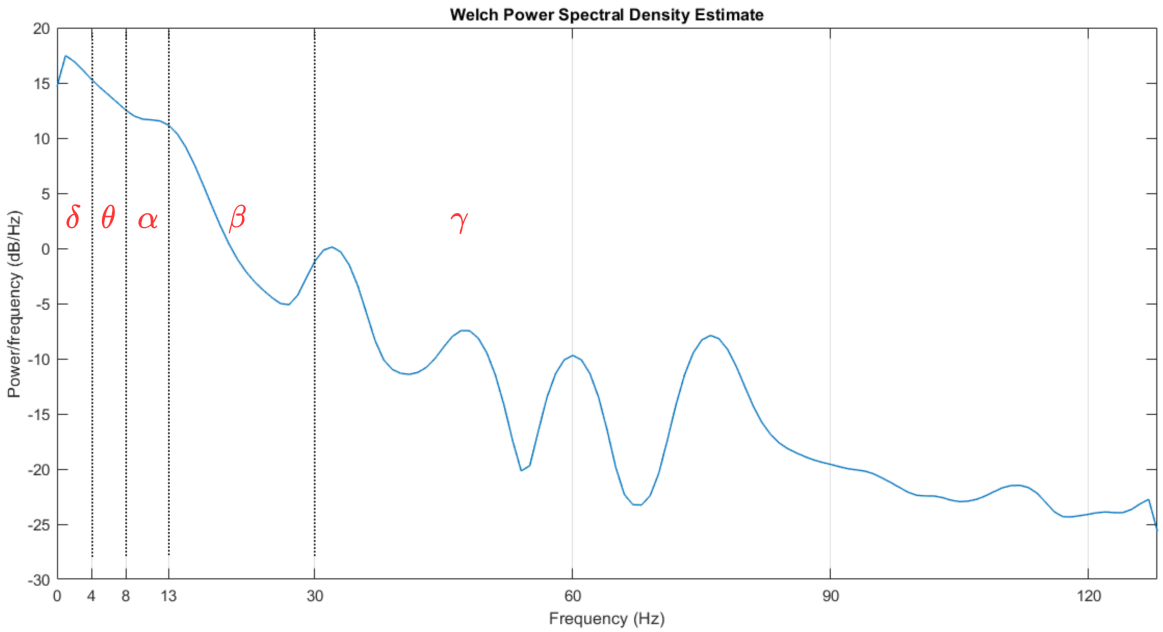
Frame 1 – Main EEG rhythms

Wave	Frequency (Hz)	Predominancy	Abnormalities
Delta (δ)	1 – 4	Deep sleep stages.	White Matter Subcortical Lesions Diffuse Lesions Encephalopathy
Theta (θ)	4 – 8	The transition from the conscious state to the drowsy state. Play a significant role in infancy and childhood. In awake adult, high theta activity is abnormal and related to brain disorders such as epilepsy.	Encephalopathy Lesions
Alpha (α)	8 – 13	Appear spontaneously in normal adults during wakefulness, under relaxation, and mental inactivity conditions.	Coma Ictal rhythm with seizures
Beta (β)	13 – 30	Attention states and active thoughts.	Drug overdose Seizures
Gamma (γ)	>30	Intense mental activity, high concentration moments.	Seizures

Source: Adapted from Sanei and Chambers (2013) and Tatum (2014).

Figure 3 shows the Power Spectrum Density (PSD) of an EEG using a 20-second Welch window, which transforms a signal from the time domain to the frequency domain (WELCH, 1967). The y-axis represents the normalized amplitude of the power spectral densities on a decibel scale; the vertical dashed lines delimit the standard EEG frequency bands. It is observed, that there is a predominance of frequencies up to around 30Hz.

Figure 3 – Frequency sub-bands on a single-channel EEG Power Spectrum Density



Source: Authors (2022).

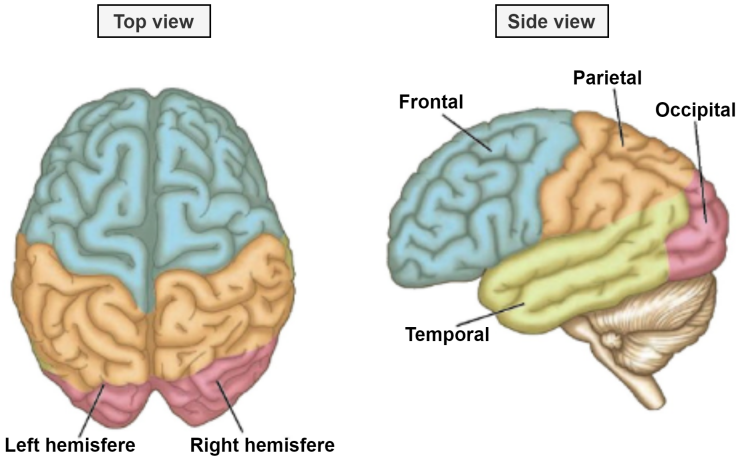
2.1.2 EEG Recording and Capture

Recording electrical brain signals can be done both in an invasive and a non-invasive way. Electrocorticography (ECoG) is an example of invasive recording also known as intercranial EEG (iEEG) in which electrodes are inserted into the brain and remain in direct contact with the brain tissue. Scalp EEG recording is noninvasive, where electrodes are placed over the scalp (SANEI; CHAMBERS, 2013). Knowing where to place the electrodes is necessary because different cortex lobes process different activities (SIULY; LI; ZHANG, Y., 2016).

The cortex is the outermost layer of the brain and is subdivided symmetrically into two hemispheres: the right and the left. Each hemisphere is divided into four lobes, represented in Figure 4.

These lobes contain regions where the main brain activities, vital and cognitive, are processed. The frontal lobe is involved with personality, emotions, problem-solving, motor development, reasoning, planning, speech, and movement. The parietal lobe handles sensation, sensory comprehension, recognition, perception of stimuli, orientation and movement. The occipital lobe is responsible for visual processing. The temporal lobe is involved in recognizing auditory stimuli, speech, perception, and memory (SIULY; LI; ZHANG, Y., 2016).

Figure 4 – Anatomy of the cerebral cortex presenting the four lobes



Source: Adapted from Lim et al. (2018).

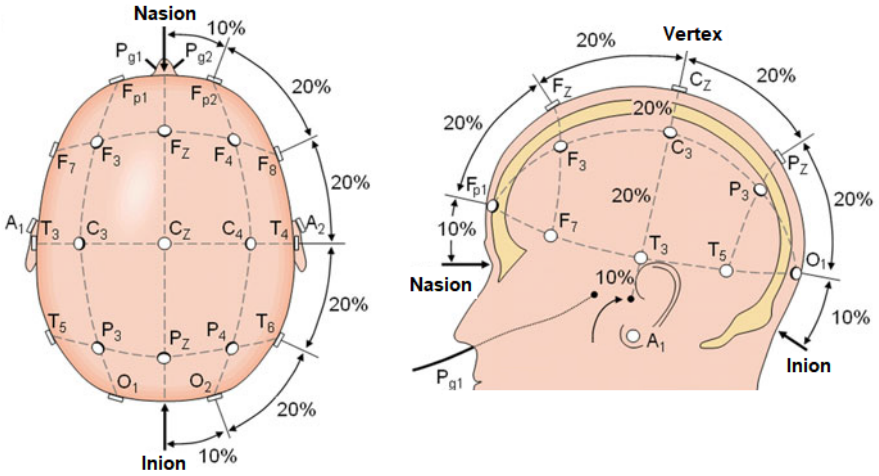
The location of the electrodes is pre-established by the 10-20 international system. It ensures standardization, reproducibility, and comparison of studies and scientific papers in the literature. The positioning uses 21 points, as illustrated in Figure 5, which are determined by dividing the scalp in proportions of 10% or 20% of the distance between the reference points: nasion and inion — front and back parts of the head, respectively (MALMIVUO; PLONSEY, 1995).

The electrodes nomenclature is also defined by the 10-20 system from letters and numbers. The letters correspond to the scalp section where the electrode is positioned: C for central, F for frontal, Fp for frontal-polar, P for parietal, T for temporal, and O for occipital. The sub-index "z" refers to an electrode placed on the middle of the scalp; even numbers refer to electrode positions on the right hemisphere, whereas odd numbers refer to those on the left hemisphere (MALMIVUO; PLONSEY, 1995).

Recording the EEG from distinct regions of the skull makes it feasible to identify a functional disorder in the brain and judge its severity, location, and expansion. In this context, it is significant to note that the same type of damage can affect the EEG in different ways due to individual diversities (FREEMAN; QUIROGA, 2012).

Since typical scalp EEG recording has a peak-to-peak amplitude of less than 200μV, it becomes necessary to use amplifiers and filters for artifacts reduction. Usually, artifacts have an extra-brain origin, either technical or physiological activities, which can hinder or even prevent the correct interpretation of EEG signals (*e.g.*, electrode movement, muscle contraction, eye blinking, and power lines noise at 50 or 60 Hz) (FREEMAN; QUIROGA, 2012). All of those can have a significant impact on the EEG

Figure 5 – International 10-20 system of electrode placement



Source: Adapted from Malmivuo and Plonsey (1995) .

analysis. Thus several types of research aim at minimizing such effects (GOTMAN; IVES; GLOOR, 1981; VERGULT et al., 2007). Therefore, the seizure prediction algorithm has to handle various types of noise and ensure that even high amplitude noise does not interfere with the detection process.

2.1.3 EEG and Epilepsy

The EEG acquisition system is one of the oldest electrical devices used in medical technology, with the first human EEG recorded by Hans Berger in 1929 (SANEI; CHAMBERS, 2013). A decade later, it was already shown that, for patients having epilepsy, certain patterns called epileptiform pattern activities were present. Epileptiform appears during the onset (the transition from the preictal to the ictal state) of epilepsy and refers mostly to spike waves and sharp waves (SIULY; LI; ZHANG, Y., 2016). Some examples of such shapes are pictured in Figure 6.

Figure 6 – Examples of (a) epileptic spikes, (b) sharp waves and (c) spike and wave complex in the EEG

(a) **(b)** **(c)**

Source: Adapted from Malaver (2017).

Both waveforms are generally described as having a sudden initial upstroke,

being their duration the only difference. A sharp wave duration is between 70–200 ms, whereas a spike is no longer than 70 ms. Specific artifacts such as cardiac activity can affect a normal EEG, so an ECG artifact can sometimes be incorrectly classified as these types of waves (SÖRNMO; LAGUNA, 2005). These patterns in EEG recordings have been used to diagnose and classify epilepsy since their discovery.

The flow of a naturally occurring seizure comprises an interictal (baseline activity), preictal, ictal, and postictal section. The interictal stage also contains epileptic patterns that change according to the patient's consciousness and brain activity. Then, before the seizure, the brain transitions to the preictal stage. Unlike the preictal state, the ictal period is well distinguished because a seizure causes a noticeable change in the EEG (LODDENKEMPER; LÜDERS, 2008).

The most widely accepted classification of epileptic seizures is defined by the Commission on Classification and Terminology of International League Against Epilepsy (ILAE) (FISHER et al., 2005; LODDENKEMPER; LÜDERS, 2008).

- Clinical seizures: epileptic seizures that show clinical manifestations as behavioral events characterized by involuntary movements like the flexing of lower and upper limbs, eyes rolling towards the back of the head, facial twitches or shaking.
- Sub-clinical seizures: seizures with no clinical manifestations, but with recorded abnormalities in the EEG. These electrographic events are usually of shorter duration and remain more localized in the brain when compared to clinical seizures.

The second type of classification is based on the epileptic focus and has two major categories:

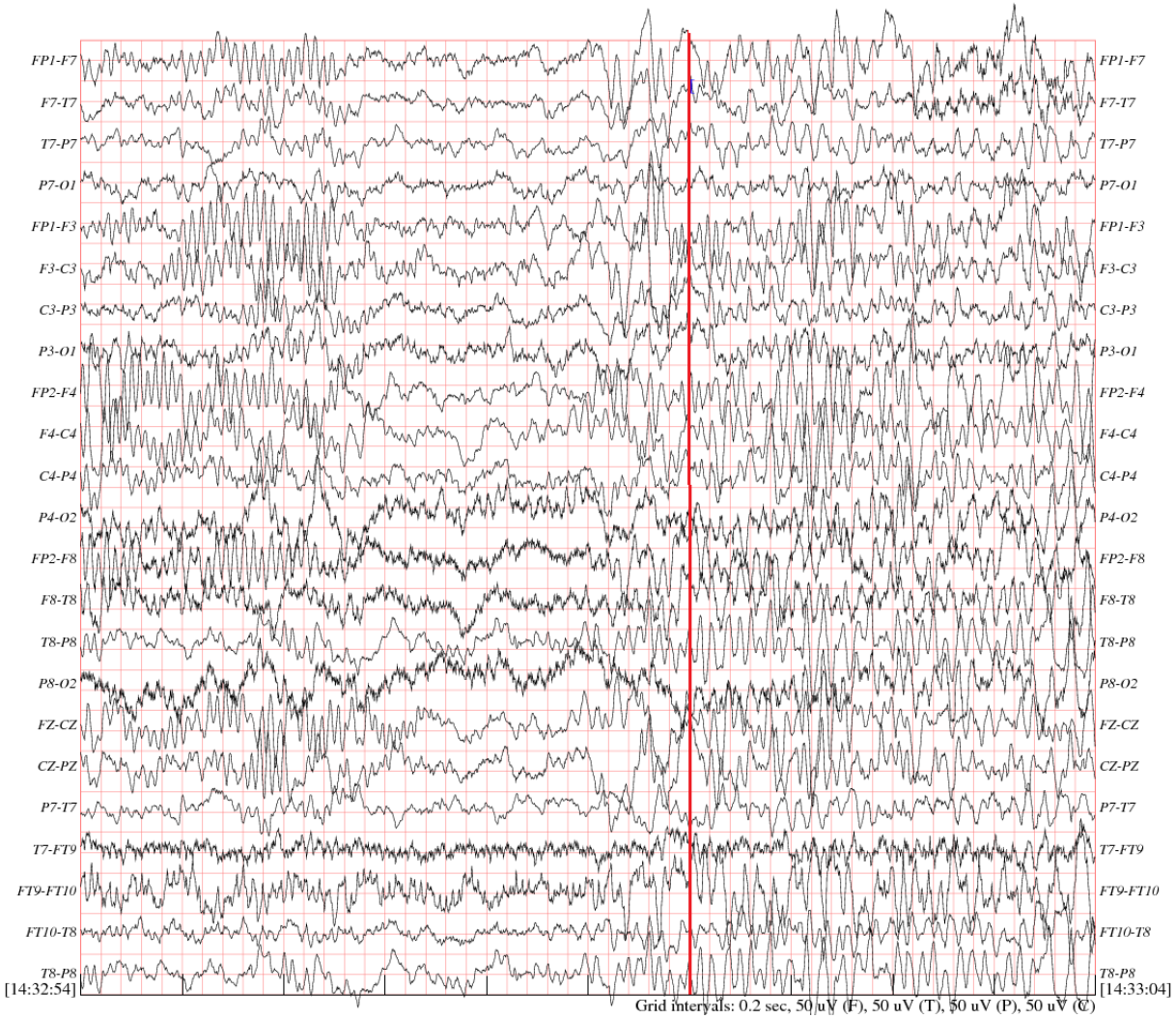
- Generalized seizures: typically affect both cerebral hemispheres simultaneously. Such seizures do not have a recognizable focus at onset and usually cause loss of consciousness.
- Partial seizures: the most common type of seizure in children, manifesting in only one hemisphere of the brain. They are further classified as simple partial seizures (if there is no impairment of the patient's consciousness and if they are limited to a small region), or as complex partial seizures if patients lose consciousness.

Partial seizures are segmented into simple partial seizures, complex partial

seizures (with impairment of consciousness at onset or followed by loss of consciousness), and partial seizures resulting in generalized tonic-clonic convulsions (preceded either by a simple or complex partial seizure). Generalized seizures are further classified into absence seizures, myoclonic seizures, clonic seizures, tonic seizures, tonic-clonic seizures, and atonic seizures (LODDENKEMPER; LÜDERS, 2008).

Figure 7 illustrates one segment of multi-channel EEG seizure onset. An epileptic seizure began at the sixth second (highlighted by the red bar), followed by a dramatic oscillation in the EEG signal.

Figure 7 – A 10-second sample of multichannel EEG record from the CHB-MIT dataset. The red bar marks the beginning of a seizure.



Source: Adapted from Goldberger et al. (2000) and Shoeb (2009).

2.2 EEG SIGNAL PROCESSING

Biomedical signals are a specific class of signals that contain information about biological systems. Thus, biomedical signals processing presents typical issues, derived from the complexity of the human body, and the need to make non-invasive indirect measurements in most applications.

The nature of the EEG signal and its importance in research on brain functioning, as well as its clinical applications, has made the introduction of signal analysis methods indispensable. In this chapter, we will explain some mathematical characteristics of EEG signals, as well as some possibilities of digital processing.

2.2.1 Signal Characterization and Preprocessing

Signals can be generically divided into continuous and discrete. Bioelectric signals, as well as electroencephalogram recordings, are essentially continuous. However, during their capture, they are sampled and become discrete. To provide a consistent notation throughout this study, we define continuous signals as $x(t)$ and its discrete time indexed by n, as $x[n]$, given by

$$x[n] = x(n\,T_s) \tag{1}$$

where T_s is the sampling period, and its reciprocal is the sampling frequency (OPPENHEIM, A., 1997).

Another approach to classifying bioelectric signals comprises two groups: deterministic and stochastic. Deterministic signals are those that can be represented by mathematical or graphical means, *i.e.* their current and past values in time can determine all their future values precisely (VAN DRONGELEN, 2018).

By the same definition, a stochastic signal comes from a stochastic process, which cannot be expressed analitically, being described by its statistical properties. From this simple definition one can conclude that all observed signals, whether man-made or otherwise, should be classified as random.

In the analysis of random sequence, it is frequently convenient to represent finite-length sequence as vectors. Thus, a discrete signal $x[n]$ defined on the interval

$[0,\ N-1]$ can be represented as a vector x with components $x[n]$ as shown in Equation (2) (THERRIEN, 1992).

$$\mathbf{x} = \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} \tag{2}$$

As shown in the previous sections, the EEG signals do not represent only brain activities. This signal is corrupted by random noise and artifacts introduced by distinct sources — it is not possible to predict the amplitude, duration, and morphology of this type of signal, which makes it reasonable to consider its stochasticity. Several researchers have proven that the EEG is not necessarily random in nature, but has such high complexity that analysis by statistical tools is convenient (VAN DRONGELEN, 2018).

The underlying model representing the random sequence is known as a random process or a stochastic process. Consider a sequence $x[n]$ such that its value for any choice of the parameter n is a random variable, hence a collection of random variable values determines the formation of a random process. Mathematically, the set of all outcomes of an experiment can be thought of as comprising the sample space S. For the event ζ defined on S their respective probabilities, if for any $t \in I$ and $\zeta \in S$ one sets a random variable $X(t,\zeta)$ to the set $\{X(t,\zeta) : t \in I\}$, then one has a random process (LEON-GARCIA, 2008). If the real variable t is a time variable, the random process is called a stochastic process.

In recent decades, many systems based on continuous-time analog circuits have been implemented using discrete-time digital systems. This phenomenon is mainly due to the more accessible access to digital signal processors (DSP) and general-purpose microcomputers (LATHI, 1998).

Using a digital filter in DSP has numerous advantages over the analog filter. Among them is that digital filters are programmable without hardware modifications. Analog filters, usually composed of active components, are subject to variations due to temperature changes or other components. Variations that digital filters are not subject to, which makes them much more stable (OPPENHEIM, A. V.; SCHAFER, 2009).

The discrete-time approach often represent naturally occurring events at discrete points in time, *i.e.*, they result from sampling continuous random signals. This concept is of paramount importance in signal processing and other engineering applications.

Stochastic processes consider the uncertainty and randomness often inherent in the phenomenon itself. Biomedical applications have a complex characteristic; most of the results that come from interactions of many variables that are very difficult - or even impossible in principle - to isolate and control in most procedures (KUTZ, 2009).

2.2.2 Discrete Wavelet Transform

Due to their particular characteristics (low amplitudes, non-stationarity for long recordings, existence of artifacts, etc.), the electroencephalogram signals present processing problems. The Fourier Transform (FT), widely used for processing stationary signals, does not provide enough information when applied to non-stationary signals.

The FT determines the frequency components of the signal, but not how they are distributed in the time domain (WEEKS, 2010), which is a disadvantage of the method since EEG recordings are susceptible to abrupt changes caused by various events.

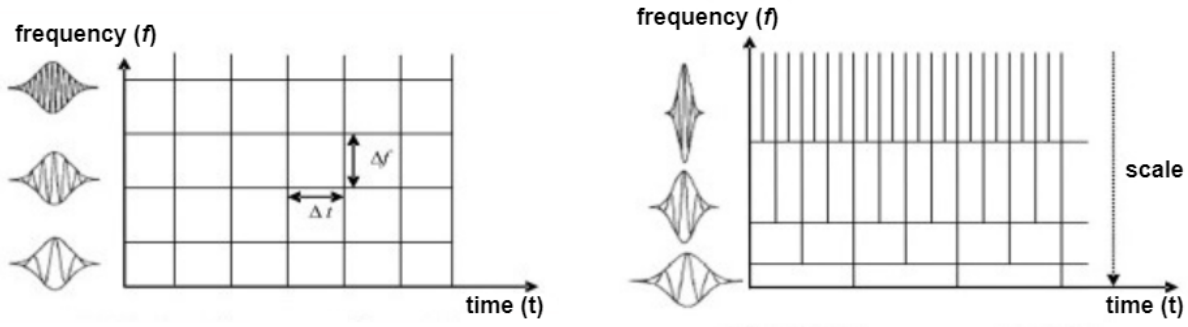
A solution to this problem would be to partition this complete time interval into smaller subintervals. The FT would be calculated for each time window - a procedure referred to as the Short-Time Fourier Transform (STFT). However, this method has problems regarding the size of the time window, since higher frequency waves have short periods, while low-frequency signals have longer periods. Thus, choosing the window size in time would impair its resolution in frequency and vice versa (MISITI et al., 2008; WEEKS, 2010).

For these reasons, a type of mathematical analysis was developed and formalized in the mid-1980s that was called Wavelet Transform (WT) (MISITI et al., 2008). WT provides a more flexible way of representing the time-frequency domains of a signal, as illustrated in Figure 8. This is achieved by using a variable windowing technique, where the time windows can vary their width depending on the frequency band being analyzed.

When low frequencies are observed, a long observation in time is required. In contrast, when high frequencies are observed, only a quick observation in time is needed. According to Daubechies (1992), WT studies each different frequency component with a resolution appropriate to its scale. Thus, one has short basis functions for high frequency

and long ones for low frequencies.

Figure 8 – Time-frequency windows of STFT (a) and WT (b)



Source: Adapted from Poisson, Rioual, and Meunier (1999).

From Figure 8(a), it can be seen that the STFT has a fixed time-frequency window (Δt and Δf), which means a lack of flexibility. However, the WT, Figure 8(b), can provide high-frequency components with sharper time resolution than the low-frequency components. Such a feature enables users to choose a proper window to examine signals at different resolutions, especially in the analysis of fast transient waveforms, such as spikes present in EEG signals.

The Continuous Wavelet Transform (CWT) decomposes a signal in terms of functions called wavelets resultant from scaling and shifting an original function called mother wavelet $\Psi(t)$ — an arbitrary small wave of limited duration and concentrated energy, designed to afford the analysis of nonstationary signals and transients. Unlike STFT, the CWT of a function of one variable is represented by a function of two continuous variables, scale (a) and translation (b). The wavelet function, depending on the scale and translation factors, can be defined as:

$$\Psi_{a,b} = \frac{1}{\sqrt{|a|}} \Psi^* \left(\frac{t-b}{a} \right) \tag{3}$$

where a and b are a set of real numbers and * is a complex conjugate.

If $x(t)$ is a continuous function in time t, then the CWT coefficients are defined as:

$$W_{a,b} = \int_{-\infty}^{+\infty} x(t) \Psi_{a,b} \, dt \tag{4}$$

The wavelet function becomes narrower with the increase of a and is displaced

in time with varying values of b. Therefore, wavelets can monitor different frequency ranges by varying the width of the window. The translation parameter relates to the location of the window. When this parameter is varied, the wavelets are shifted in time, allowing them to cover the whole signal in the time domain. The signal is analyzed piecewise, essentially processing the content that is within the window (DAUBECHIES, 1992).

CWT is not very versatile because the calculation of the analysis scales requires much time and computational resources and generates redundant information (MALLAT, 1989). Therefore, the WT admits another main approach for signal inspection: the Discrete Wavelet Transform.

In the DWT procedure, the wavelets are translated and scaled in discrete intervals, $a = a_0^m$ and $b = nb_0a_0^m$, reducing the computational effort. Initially, this process is done by fixing two positive constants a_0 and b_0 and defining:

$$\Psi_{m,n} = a_0^{-m/2}\,\Psi\left(a_0^{-m}t - nb_0\right) \tag{5}$$

where m, $n \in \mathbb{Z}$.

DWT represents a function through a measurable set of wavelet coefficients, which corresponds to points on a two-dimensional grid of discrete points in the time-scale domain, indexed by m and n.

A computationally efficient approach involves setting $a_0 = 2$ and $b_0 = 1$. We can observe that for these values, for any increment of m, the scaling value doubles, which implies doubling the width in the time domain and halving the width in the frequency domain (DAUBECHIES, 1992). Thus, for every increment of m, a enhanced frequency resolution is obtained, and for every decrement, a enhanced time resolution is yielded. For different values of m and n, a multi-resolution analysis is possible (MISITI et al., 2008).

From the DWT multi-resolution analysis, the signal is decomposed into approximations and details and may be reconstructed with no loss. The approximation refers to the high scales (low-frequency components), while the detail corresponds to the low scales, high-frequency components of the signal (MALLAT, 1989).

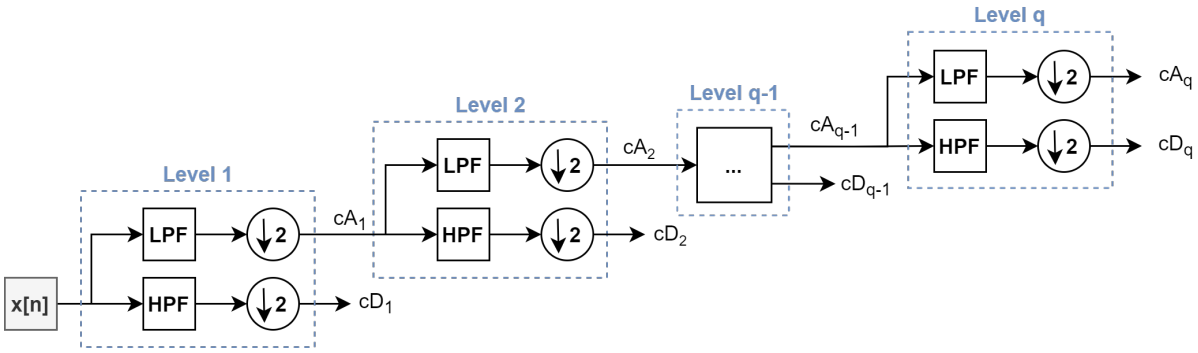
According to Mallat (2009), the calculation of DWT, through the use of filters, is the most efficient way of applying WT. The wavelet decomposition can be done through

banks of high-pass and low-pass filters, providing representations of the original signal relative to different scale/frequency levels.

Figure 9 shows the q-level wavelet decomposition scheme of a signal $x[n]$ sampled at a frequency f_s. As noted, the signal $x[n]$ is applied to a pair of low-pass and high-pass filters (LPF and HPF respectively) and then sub-sampled, producing the approximations (cA) and details (cD) with n_q samples - half the amount of samples from the previous stage $q-1$ (MALLAT, 1989).

The coefficients of the filters depend on the mother wavelet function used. The low-pass channel wrappers are expansions/compressions of a single scaling function $\phi(t)$ and the high-pass channel wrappers are also expansions/compressions of a single wavelet function $\Psi(t)$.

Figure 9 – DWT q-level filter bank structure



Source: Authors (2022).

The approximations are again subject to the filters at the next level, producing another set of approximations and details. The coefficients of approximations (cA_q) and details (cD_q), of level q, are obtained by convolution of the approximation coefficients (cA_{q-1}), with the LPH and HPF filters, respectively.

This process is repeated until the desired level of decomposition is completed, resulting in a binary tree – which can also be seen as a filter bank structure. Features are extracted from each node in this tree (MISITI et al., 2008). Each node represents a frequency interval half as wide as the level above and twice as wide as the level below. The time scale at each level is twice the level below and half above (MALLAT, 1989).

The choice of wavelet basis function and the number of decompositions imply the success of DWT and are determined by the characteristics of the signal or image and the nature of the application (RAFIEE et al., 2011).

In practice, a satisfying number of levels is based on the frequency ranges that need to be analyzed and the signal nature. Here, the coefficients must discriminate the spike-like events of the signal and be useful for further analysis: spatial filtering, feature extraction, and classification steps.

2.3 COMMON SPATIAL PATTERNS

Many real-world data, such as noise signals, multichannel signals, images, or corpus data, exhibit the characteristic of high dimensionality. Keeping in mind the goal of dealing with such information successfully, it is desirable to reduce the dimensions of the data in many situations (VAN DER MAATEN; POSTMA; VAN DEN HERIK, et al., 2009).

One of the major tools used for dimensionality reduction is the method called Principal Components Analysis (PCA). PCA is a multivariate statistical technique that comprises exploring the variance and covariance structures of a random vector, composed of random variables, employing principal components. The principal components are linear combinations of all the original variables, are decorrelated from each other and are estimated to keep, in order of estimation, the maximum information, in terms of the total variation in the data (JOLLIFFE, 1972).

The Common Spatial Pattern (CSP) technique is another feature extraction method that can self-design spatial filters to maximize the separation between two classes. This strategy was first introduced by Koles, Lazar, and Steven Z. Zhou (1990) to extract discriminative EEG features from two human populations.

CSP is one of the most popular and efficient Brain-Computer Interface (BCI) systems methods commonly applied in BCI competitions (WANG, Y.; GAO, S.; GAO, X., 2006; WANG, L. et al., 2013). However, several other applications extract signal features that best represent the underlying physiologic activity for a specific task, such as hand/foot movements (FENG et al., 2019; PARK, Y.; CHUNG, 2019) and seizures detection/prediction EEG signals (ALOTAIBY; ALSHEBEILI; ALOTAIBI, et al., 2017; USMAN, S. M.; USMAN, M.; FONG, 2017; ZHANG, Y. et al., 2019).

The aim of using CSP in this study is to distinguish between interictal and preictal states. As an extension of PCA, CSP can find a projection matrix $\mathbf{W}$ that minimizes the variance for preictal activity and maximizes it for the other class. The mathematical formulation of the algorithm is described as follows (WANG, Y.; GAO, S.; GAO, X., 2006).

2.3.1 CSP Algorithm

CSP algorithm performs spatial filtering of the EEG signals and provides a metric for choosing the subspace projection that contains more class-related information. The transformation is made so that, for each CSP filter, the variance of the resulting signal is maximized if the trial is from one class and minimized if it is from the other class (KOLES; LAZAR; ZHOU, S. Z., 1990).

For this work, the variables $\mathbf{X}_I{}^{(k)}$ and $\mathbf{X}_P{}^{(k)}$ denote each trial of wavelet decomposition matrices under two EEG states (interictal and preictal) with dimensions $N \times n_q$, representing a set of q-th level wavelet detail coefficients of the k-th 5-second window, with N channels and n_q samples. For CSP projection, a normalized covariance matrix of each class is required for each segment given in Equation (6), where $\{.\}^T$ is the matrix transpose operation and the $Trace\{\mathbf{X}\}$ operation computes the sum of the diagonal elements of $\mathbf{X}$:

$$\mathbf{R}_I{}^{(k)} = \frac{\mathbf{X}_I{}^{(k)}\mathbf{X}_I{}^{(k)T}}{Trace\left\{\mathbf{X}_I{}^{(k)}\mathbf{X}_I{}^{(k)T}\right\}} \qquad \mathbf{R}_P{}^{(k)} = \frac{\mathbf{X}_P{}^{(k)}\mathbf{X}_P{}^{(k)T}}{Trace\left\{\mathbf{X}_P{}^{(k)}\mathbf{X}_P{}^{(k)T}\right\}} \tag{6}$$

The $\mathbf{R}$ matrices are calculated for each multichannel samples. The algorithm also computes the average of these covariance matrices for each class $\overline{\mathbf{R}_I}$ and $\overline{\mathbf{R}_P}$

$$\overline{\mathbf{R}_I} = \sum_{k=1}^{m_I} \frac{1}{m_I} \mathbf{R}_I{}^{(k)} \qquad \overline{\mathbf{R}_P} = \sum_{k=1}^{m_P} \frac{1}{m_P} \mathbf{R}_P{}^{(k)} \tag{7}$$

where m_I and m_P are the total five-second windows corresponding to the interictal and preictal periods, respectively.

The following step consists in computing the global covariance of both classes as $\mathbf{R} = \overline{\mathbf{R}_I} + \overline{\mathbf{R}_P}$. Covariance matrices are symmetric and positive semi-definite by construction (FUKUNAGA, 1990). Thus the spectral decomposition of $\mathbf{R}$ is given by

$$\mathbf{R} = \mathbf{U}_0 \, \Sigma \, \mathbf{U}_0^T \tag{8}$$

where $\mathbf{U}_0$ is the eigenvectors matrix and Σ is the diagonal matrix of eigenvalues of the covariance matrix $\mathbf{R}$. Using the previous factorization, the next step consists of defining the whitening matrix $\mathbf{P} = \Sigma^{-1/2} \, \mathbf{U}_0^T$, which will transform the class related covariance matrices as:

$$\mathbf{S}_I = \mathbf{P}\,\overline{\mathbf{R}_I}\,\mathbf{P}^T \quad \mathbf{S}_P = \mathbf{P}\,\overline{\mathbf{R}_P}\,\mathbf{P}^T \tag{9}$$

These matrices share the same eigenvectors whose concatenation will form the projection matrix $\mathbf{W}$. Nevertheless, they are sorted in a descending order since their eigenvalues are complimentary in the sense that $\Sigma_I + \Sigma_P = \mathbf{I}$ (WANG, Y.; GAO, S.; GAO, X., 2006). A high eigenvalue means that the filter output based their respective filter vector in $\mathbf{W}$ yields a high variance for input signals in one class and a low variance for signals from the other class (and vice versa. Thus, this can be expressed mathematically as:

$$\mathbf{S}_I = \mathbf{U}\,\Sigma_I\,\mathbf{U}^T \quad \mathbf{S}_P = \mathbf{U}\,\Sigma_P\,\mathbf{U}^T \tag{10}$$

Thus, the CSP projection matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$ is formulated as

$$\mathbf{W} = \mathbf{U}^{-1}\mathbf{P} \tag{11}$$

whose filtering process results in a matrix $\mathbf{Z} \in \mathbb{R}^{N\times n_q}$ represented by Equation (12).

$$\mathbf{Z} = \mathbf{W}\,\mathbf{X} \tag{12}$$

Matrix $\mathbf{Z}$ can be seen as source components including common and specific components of different EEG periods, $\mathbf{X}_I$ and $\mathbf{X}_P$, for interictal and preictal states, respectively. Then each trial of wavelet coefficients data $\mathbf{X}$ can be reconstructed by

$$\mathbf{X} = \mathbf{W}^{-1}\mathbf{Z} \tag{13}$$

The rows of $\mathbf{W}$ represent spatial filters of a selected type, while the columns of $\mathbf{W}^{-1}$ represent the common spatial patterns. The first and last columns of the projection matrix are the most significant spatial patterns that explain one task's largest variance and the smallest variance of the other (KOLES; LAZAR; ZHOU, S. Z., 1990).

Assuming that the vectors $\mathbf{w}_i$ are the optimal filters in $\mathbf{W} = [\mathbf{w}_1,...,\mathbf{w}_N]^T$, CSP can provide N spatial filters. A common practice is to reduce the dimensionality of the output by selecting the P filters that best discriminate between classes. If P is too small, discrimination between classes may not be effective; on the other hand, high values of P can lead to an overfitting condition (BLANKERTZ et al., 2008).

The chosen P filters are the eigenvectors associated with the most and least significant eigenvalues related to the preictal class. Assuming P to be an even number smaller than N, the set $\mathbf{W}_R \in \mathbb{R}^{P \times n_q}$ is composed of the first $P/2$ and last $P/2$ optimal filters:

$$\mathbf{W}_R = [\mathbf{w}_1,...,\mathbf{w}_{\frac{P}{2}},\mathbf{w}_{N-\frac{P}{2}+1},...,\mathbf{w}_N]^T \tag{14}$$

Therefore, the filtering process also reduces the dimension from N EEG channels to P spatially filtered data and gathering the relevant information that is spread over different channels.

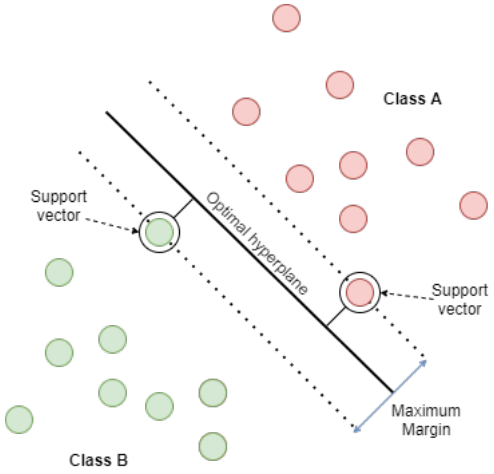
2.4 SUPPORT VECTOR MACHINE

Support Vector Machine is a supervised learning model for pattern analysis and recognition. They are used for binary classification and regression of data and aim to solve the problem through the concept of margin, which comprises the smallest distance between the decision surface and any of the samples (BISHOP, 2006).

Support vector machines, developed by Vapnik (2000), can solve classification and regression problems, acquiring through a learning process, in the training stage, the ability to generalize. Considering a binary problem, the goal of the SVM is to separate the instances of the two classes by hyperplanes that will be obtained from the known instances in the training phase. The goal is to produce a classifier that works adequately with unfamiliar examples, *i.e.*, examples that were not applied during training, thus acquiring the ability to predict the outputs of future new inputs (GUNN et al., 1998).

Training an linear SVM is equivalent to solving a quadratic programming problem. The goal is to find an optimal separating hyperplane for a feature set. In this approach, the decision surface (or hyperplane) is chosen to be the one with the largest margin, as illustrated in Figure 10. The optimal hyperplane is determined from the reference of two other hyperplanes, called auxiliary hyperplanes. These connect a set of points (support vectors) (BISHOP, 2006).

Figure 10 – Optimal hyperplane of the SVM algorithm



Source: Adapted from Duda, Hart, and Stork (2001).

The maximum margins are calculated by minimizing $\|w\|$, the norm of w, where w is the normal vector of the separation hyperplane. Thus, we have the primal optimization problem in variables w and the bias parameter b, for a training set $X = (x_i, y_i)$, where x_i is the feature vector and y_i the labels:

$$\begin{aligned} &\text{Min } \frac{1}{2}\|w\|^2 \\ &\text{subject to} \\ &y_i(wx_i + b) \geq 0, \quad i = 1,2,...,N \end{aligned} \tag{15}$$

where N is the cardinality of the training set.

The classification is then done by assigning the data point to the class that is associated with the side of the hyperplane where the point lies. An essential property of this classifier is that the determination of the parameters corresponds to a convex optimization problem. Therefore, any global solution is also an optimal global value.

2.4.1 Soft Margin SVM

The problem illustrated in Figure 10 stands for linearly separable patterns. In situations where the data is non-linearly separable but can be divided into two classes by a hyperplane, a classifier that accepts such errors is adopted. The optimization problem shown in Equation (15) is updated, by adding the slack variable ξ_i, making the model tolerable to classification errors, as well as to samples between the edges, thus enabling SVMs to classify non-linearly separable data. These errors are controlled by a parameter

called the regularization parameter C, and these classifiers are known as smooth-margin SVMs (BISHOP, 2006).

The primal optimization problem for smooth-margin SVMs is given by Equation (16).

$$\begin{aligned} &\text{Min } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \\ &\text{subject to} \\ &y_i(wx_i + b) + \xi_i - 1 \geq 0 \\ &\xi_i \geq 0 \end{aligned} \tag{16}$$

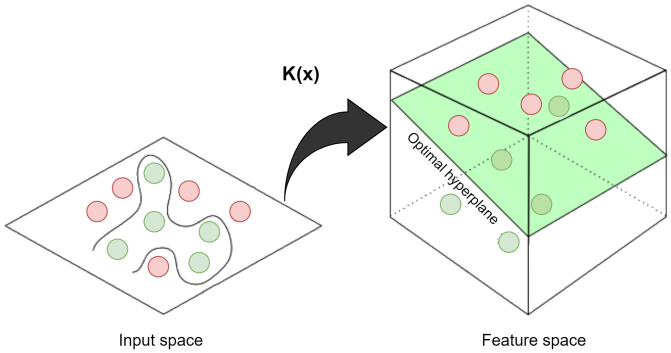
Equation (16) gives the Generalized Optimal Separating Hyperplane – a Quadratic Programming problem that can be solved here using the method of Lagrange multipliers (BURGES, 1998).

2.4.2 Kernel Functions

In genuine problems, as with EEG patterns, the features domain is not linearly separable. So even soft-margin SVMs do not produce satisfactory results. Thus, nonlinear classifiers are implemented for this study, which performs a dimensionality change through kernel functions, allowing linear classification techniques in higher dimensions (BISHOP, 2006).

Figure 11 illustrates the transformation of a non-linearly separable domain into a linearly separable problem by increasing the dimensionality, where mapping is made by a kernel function $K(x)$.

Figure 11 – Kernel transformation: non-linearly separable problem into a linearly separable problem



Source: Adapted from Tinghua Wang, Lin Zhang, and Hu (2021).

However, the kernel function $K(x)$ to be used needs to be defined in implementing an SVM (BURGES, 1998). Different kernels can assign various types of decision boundaries, where the most common include a polynomial kernel or Gaussian radial basis function (RBF) kernel, being the latter chosen for this work. The RBF kernel on two samples $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$, represented as feature vectors in some input space, is defined as

$$K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = e^{-\gamma \left\| \mathbf{x_i} - \mathbf{x_j} \right\|^2} \tag{17}$$

where γ is the scale parameter that controls the width of the Gaussian surface of the RBF kernel; it replaces each point in the feature space by the Gaussian of its squared Euclidean distance from support vectors.

With the Gaussian kernel, there are two parameters to control the classification performance: the scale parameter γ and the soft margin parameter of C. Most Machine Learning algorithms make use of heuristics and meta-heuristics to select the best parameters, because exhaustive search is very time consuming, especially when the set of examples has numerous instances and/or features. The gamma value was automatically specified by the software, according to nearest neighbor heuristics (MATLAB, 2018). This parameter was not changed as that did not seem to lead to an overall improvement in results in this preliminary analysis.

Within the MATLAB Statistics and Machine Learning Toolbox, the default error cost hyper-parameter is set to $C = 10^2$ after some preliminary experiments. The C parameter adds a penalty for each misclassified data: if a small value of C is chosen, the penalty for misclassified samples is low so a decision boundary with a large margin is chosen. On the other hand, a large C value will lead SVM to minimize the number of misclassified examples, which results in a decision boundary with a smaller margin, thus a minor expected generalization capability (QUITADAMO et al., 2017).

2.5 CLASSIFIER POSTPROCESSING TECHINIQUES

Classifiers are essentially defined to generate independent outputs for the input test samples, and not all samples are usually correctly classified. This issue becomes serious with EEG prediction, where the classification is performed between few seconds intervals. As a result, classification schemes may eventually lead to a significant number of false alarms.

Practical systems are subject to noise and random disturbances that can hinder the analysis and identification procedures. To avoid these problems, filters have been used to capture only the signals with the dynamics of interest present in the system (HAYKIN, S., 2004). The regularization methods are therefore designed to reduce the number of false alarms by taking the most recent classifications into account by making some corrective manipulations on the classifier outputs.

The following are the filters that will be considered in the prediction design from the SVM classifiers' output smoothing. These are the Moving Average Filter, the Median Filter, and the Kalman Filter.

2.5.1 Moving Average Filter

Suppose that we want to remove the noise from the data, to get a closer representation of the expected underlying smooth curve. Consider a time series y, $t = 1,...,N$. A symmetric (centered) moving average filter of window length $m = 2k + 1$ is given by:

$$MA = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j} \tag{18}$$

Intuitively, this could be accomplished by averaging values of the time series within k periods of t. Observations that are nearby in time are also likely to be close in value. Therefore, the average eliminates some high-frequency randomnesses in the data, leaving a smooth trend-cycle component (SMITH, 2003).

The centralized moving average smooths the series since it keeps the lower frequencies and attenuates the higher frequencies (LYNCH, 2015).

This class of filters change the spectral characteristics of a time series predictably, requiring that the series exhibit stationarity, has non-time-dependent mean and

autocovariance (LÜTKEPOHL, 2004).

2.5.2 Median Filter

The Median Filter (MF) is the filtering technique used for noise removal from images and signals. The median filters are considered good alternatives because they have some very interesting properties (PITAS, 1990; NODES; GALLAGHER, 1982):

1. They can smooth the transient changes in signal intensity;
2. The edge information is preserved in the filtered signal;
3. They can be implemented by very simple digital nonlinear operations.

Due to these properties, they are frequently used in various signal and image processing applications, such as seismic signal processing, speech processing, medical imaging, robotic vision, and pattern recognition (NODES; GALLAGHER, 1982; ALI, 2018).

The median of a group, containing an odd number of elements, is defined as the middle element when the elements of the group are sorted into either ascending or descending order. If the set comprises an even number, the median value is the arithmetic mean of all the values in the range $(k - r)$ to $(k + r)$.

A median filter finds the median of several elements at its input. For a set of N samples of a time series y, the median of k-th element is calculated as:

$$m = \begin{cases} y_k; \ k = \frac{N+1}{2} & \text{(N odd)} \\ \dfrac{y_k + y_l}{2}; \ k = \dfrac{N}{2}; \ l = \dfrac{N}{2} + 1 & \text{(N even)} \end{cases} \tag{19}$$

In the standard median filtering applications, a moving "window" of rank r and width $(2r + 1)$ is moved along the sampled values of the signal or the image. For each position of the window, the median of the elements within the window is computed and then written as the output value at the same position as the central element of the window.

2.5.3 Kalman Filter

The Kalman Filter (KF) is a recursive solution for linear filtering of discrete data, using mathematical equations to estimate the state of a process to minimize the mean square error. The technique was developed in the decade of 1960 by Hungarian engineer Rudolf Kalman within the field of electrical engineering, applying to stationary and non-stationary environments (KALMAN, 1960).

To understand how a Kalman filter works, it is important to first understand its equation, which considers observations got through noisy measurements. This type of filter is widely used in applications such as tracking the position, velocity, and acceleration of airplanes, drones and even atomic vehicles (LEFFERTS; MARKLEY; SHUSTER, 1982).

First, let's introduce a generic system described as state variables. Let $\mathbf{x_k} = [\mathbf{dk}\ \mathbf{d_k}]'$ denote the state vector ($\dot{d}_k$ represents the rate of change of $\mathbf{d_k}$), $\mathbf{y_k}$ is the measured variable (*i.e*, SVM output). Then, $\mathbf{y_k}$ is represented by the following state-space model:

$$\begin{cases} \mathbf{x_{k+1}} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \mathbf{x_k} + \mathbf{w_k} \\ \mathbf{y_k} = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x_k} + \mathbf{v_k} \end{cases} \tag{20}$$

where T is the time interval of an EEG epoch (a 5-second window is considered in this work); $\mathbf{w_k}$ and $\mathbf{v_k}$ are zero-mean white noise vectors with covariance

$$\mathbf{Q} = \begin{bmatrix} \sigma_w^2 \frac{T^3}{3} & \sigma_w^2 \frac{T^2}{2} \\ \sigma_w^2 \frac{T^2}{2} & \sigma_w^2 T \end{bmatrix} \tag{21}$$

where σ_w the assumed standard deviation of the random fluctuations of $\dot{\mathbf{d}}_\mathbf{k}$.

The KF underlying idea is to obtain a filtered version of $\mathbf{d_k}$ ($\hat{\mathbf{d}}_\mathbf{k}$) (CHISCI et al., 2010). The smoother variable $\hat{\mathbf{d}}_\mathbf{k}$ will then be used as the output in place of $\mathbf{y_k}$. The ratio $\sigma_{KF} = \sigma_w/\sigma_v$ can be adjusted to tune the filter performance.

The complete algorithm is described in Simon Haykin (2004). Shortly, the Kalman estimator presents two steps for estimation: the prediction and the correction. In the prediction step, one must predict the projection of the state in the forward step (a priori),

as well as the uncertainty in the estimation. The correction step, necessary to reach the estimated state, is formulated through an equation that associates the estimated state in the measure's prediction with the difference between the real measured value and the predicted value (a posteriori).

2.5.4 Stationarity Test

Stationarity is a significant factor for the analysis of time series, as it is an indicator for choosing the best statistical method, in this study, to smooth the behavior of the SVM classifiers' outputs. From all the techniques previously considered, the moving average filter is the only one that has a restriction on the stationarity of the time series. In order to verify if the dynamics of the outputs of the classifiers come from stationary processes, a stationarity test will be performed.

Several types of tests are found in the literature. In this study, the Augmented Dickey-Fuller (ADF) test was chosen por ser um famoso método formal para a verificação de estacionariedade (DICKEY; FULLER, 1979, 1981).

The development of the test was motivated by the need to check whether a series needs to be differentiated to become stationary. A time series is stationary if the random data oscillate around a constant value. This is found when the probability distribution parameters mean, variance, and covariance are fixed and constant over time (WOOLDRIDGE, 2015; MORETTIN, 2017).

The first thing to note about the ADF test is that it seeks to check whether or not a series is stationary by testing whether it has a unit root. The initial approach is to assume the following model.

$$y_t = \phi y_{t-1} + e_t \tag{22}$$

where e_t is the random error term known as white noise, which has zero mean, constant variance σ^2 and is non-self-correlated. The term ϕ stands for the adjustment speed coefficient.

When $\phi = 1$, Equation (22) becomes a random walk model (no displacement) and then the process has a unit root and the series is said to be nonstationary. If $|\phi| \leq 1$, the time series y_t is said to be stationary and consequently has no unit root. Therefore,

called the unit root test, ADF uses hypotheses, which are:

$$\begin{cases} H_0 : \ \phi = 1, \quad y_t \text{ is non-stationary (has unit root)} \\ H_1 : \ |\phi| \leq 1, \quad y_t \text{ is stationary} \end{cases} \tag{23}$$

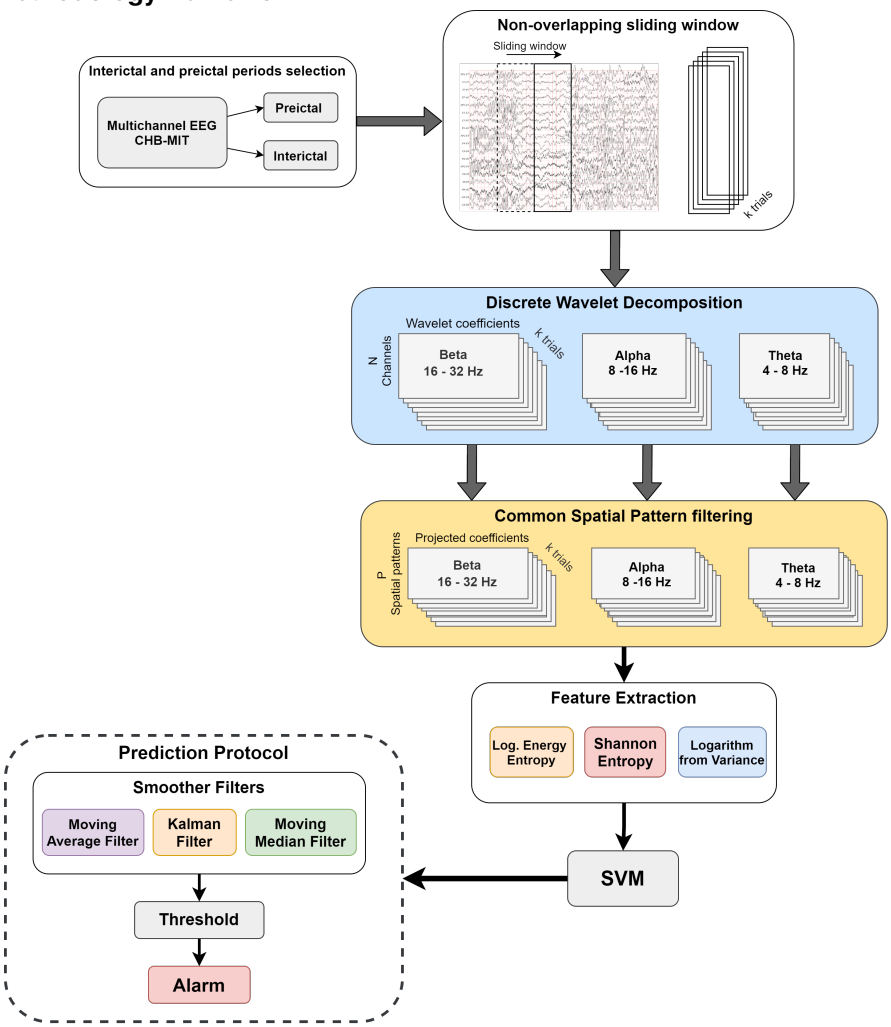
Dickey and Fuller (1979) registered critical values through Monte Carlo simulations and developed a τ statistic to formally test the unit root problem. If the absolute value of the calculated τ statistic is greater than the absolute value tabulated by DF, there is evidence to accept the null hypothesis, considering the series as non-stationary.

3 MATERIAL AND METHODS

This chapter discusses the materials and methods employed to develop the research. In the first section, there is a description of the CHB-MIT EEG data set, commenting on the acquisition procedure. The criteria for the selection of patients are also presented, as well as the criteria for the determination of the interictal and preictal periods. In Sections 3.2 and 3.3, the DWT EEG processing technique and the spatial filtering procedure via CSP are discussed, respectively. Sections 3.4 and 3.5 explains the feature extraction step, also covering the classification performance metrics used in the study. Finally, Section 3.6 discusses the procedures adopted for the design and validation of the prediction protocol.

Figure 12 shows an overview of the proposed process for the seizure prediction method.

Figure 12 – Methodology framework



Source: Authors (2022).

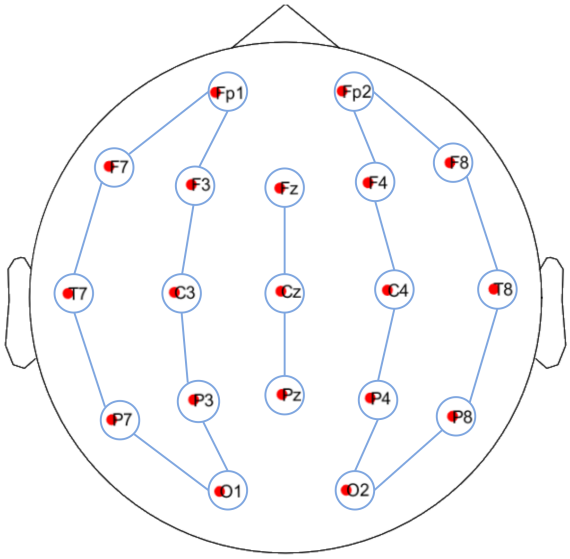
All of the computational codes are written in MATLAB. The MATLAB environment is a well-known suite for digital signal processing and advanced mathematical applications. The tool has also a large set of specialized support programs, called toolboxes, which make it suitable for solving particular classes of problems. All algorithms were developed using the Wavelet and Statistics, and Machine Learning Network Toolboxes.

3.1 DATABASE DESCRIPTION

The scalp EEG signals used in this study are available in the PhysioNet open-source CHB-MIT database created by a group of researchers from Boston Children's Hospital and Massachusetts Institute of Technology (MIT). The recordings are from 23 patients, whose ages range from 1.5 years to 22 years. Furthermore, the gender of the patients in each case is specified, 5 males and 17 females. All 23 individuals have intractable seizures and underwent continuous monitoring for several days following withdrawal of anti-seizure medication to characterize their seizures and assess their candidacy for surgical intervention. An annotation document clearly states the time of seizure start and end for each subject (GOLDBERGER et al., 2000; SHOEB, 2009).

The EEG signals were sampled at 256 Hz with 16-bit resolution. The electrodes were placed on the scalp according to the 10-20 international system with bipolar longitudinal setup (Figure 13) – where each channel measures the difference between two adjacent electrodes (HENRY, 2006; FISCH; SPEHLMANN, 1999).

Figure 13 – CHB-MIT Database electrode's setup



Source: Adapted from Libenson (2010).

As a patient-independent study, and since seizures often occur in different regions across patients (as mentioned in Section 2.1), is relevant to consider the records from electrodes placed all over the scalp. Hence, a subject's 18 common channels were selected (FP1-F7, F7-T7, T7- P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ and CZ-PZ), covering large area records of the scalp, which can provide a more consistent analysis in both time-frequency and spatial domains.

According to the number of crises available and their respective distributions over the recorded period, we have considered 17 out of 23 patients. The subjects' information as their seizure condition is presented in Table 1.

Table 1 – Detailed description of CHB-MIT EEG Database

Subject (ID)	Gender[1]	Age	Seizure Type[2]	Brain Location	No. of seizures[3]
01	F	11	SP, CP	Frontal	4
02	M	11	SP, CP, GTC	Temporal	1
03	F	14	SP, CP	Frontal	1
04	M	22	SP, CP, GTC	Temporal	2
05	F	7	CP, GTC	Frontal	4
06	F	1.5	CP, GTC	Temporal/Occipal	6
07	F	14.5	SP, CP, GTC	Temporal	2
08	M	3.5	SP, CP, GTC	Frontal	2
09	F	10	CP, GTC	Temporal/Occipal	3
10	M	3	SP, CP, GTC	Temporal	2
13	F	3	SP, CP, GTC	Temporal/Occipal	1
14	F	9	CP, GTC	Frontal/Temporal	2
15	M	16	SP, CP, GTC	Temporal	5
16	F	7	SP, CP, GTC	Temporal	2
18	F	18	SP, CP	Frontal	2
21	F	13	SP, CP	Temporal/Parietal	1
22	F	9	-	Temporal	2
Total					42

[1] Female (F) and Male (M)
[2] Simple partial seizure (SP), Complex partial seizure (CP) and Generalized tonic-clonic seizure (GTC)
[3] The number of seizures considered for the study of the total available, which met the criteria explained in subsection 3.1.1

Source: Yuan Zhang et al. (2019), Goldberger et al. (2000) and Shoeb (2009).

3.1.1 Interictal and Preictal Periods Selection

The choice of preictal length is a major issue with seizure prediction algorithms, and may differ from patient to patient, and even among the seizures of a patient. Based on the literature, it has been reported that there are electrophysiological changes that

might develop minutes to hours before the actual seizure onset (MORMANN; KREUZ, et al., 2005; DIREITO et al., 2017). As there is no prior marking by the experts, the definition of preictal and interictal periods comprises the initial stage of this work.

The preictal period should be selected carefully, so to cover the information and patterns best reflecting the preictal activities. To tackle this issue, the length of the preictal period in this study, also called seizure prediction horizon (SPH), is defined by AES (2014). Daoud and Bayoumi (2019) also believe that 60 minutes is an appropriate range that can effectively provide significant time for effective patient care.

The period of 60 minutes was also chosen for the interictal period. This strategy was assumed because of the limited number of seizures available as well as to avoid the unbalanced data problem.

Thus, the continuous recordings were truncated and the interictal and preictal periods were extracted, taking as reference the seizure onset and termination marks. Hence, each seizure in the database could only be considered for the study if it had at least 120 minutes of prior seizure-free recordings, justifying the subjects and the number of seizures presented in Table 1.

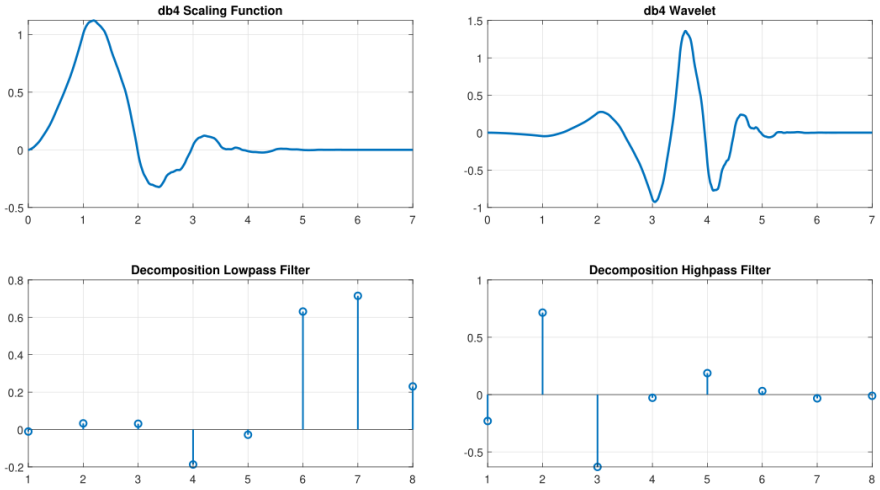
3.2 DISCRETE WAVELET DECOMPOSITION

The preprocessing stage generally consists of filtering the EEG signals to limit them within frequency ranges of interest through DWT high and low-pass filters. In the specific case of application in EEG recordings, the choice of wavelet function assumes that it should resemble, as much as possible, the epileptogenic event that one wishes to highlight.

Given the respective approaches described in Adeli, Z. Zhou, and Dadmehr (2003) and Gandhi, Panigrahi, and Anand (2011), the Daubechies 4 orthonormal wavelet function (db4), which is constructed from an eighth-order filter, is used on the proposed system. Both Halford (2009) and Indiradevi et al. (2008) argued the db4 wavelet yields the highest correlation coefficients with the epileptic behavior among the wavelet bases available in the MATLAB Toolbox.

Figure 14 presents the scale function, Daubechies 4 wavelet function, and the coefficients of the low-pass and high-pass filters, respectively.

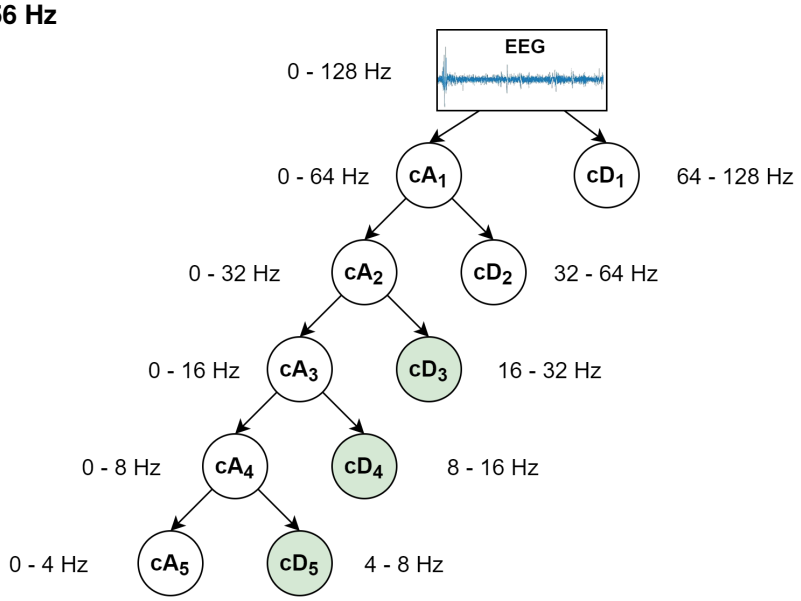
Figure 14 – Scaling and Wavelet functions for Daubechies 4 and its low-pass and high-pass filters



For the EEG segmentation as trials, a non-overlapping sliding windowing technique with a fixed width of 5 seconds (1280 samples) was used to minimize the effect of non-stationarity (BANDARABADI; TEIXEIRA, et al., 2015; RUKHSAR et al., 2019). According to the Nyquist-Shannon sampling theorem (SHANNON, 1949), one can only represent frequencies up to half of the sampling frequency, thus the frequencies of the signals range between 0 and 128 Hz.

Knowing that the seizure in recorded EEGs usually occurs between 3-29 Hz (GOTMAN, 1982), five-level Wavelet decomposition was set to match well with the aforementioned frequency range of the EEGs sub-bands. Figure 15 shows the corresponding wavelet decomposition tree.

Figure 15 – Five-level decomposition for Daubechies order 4 wavelet with a sampling frequency of 256 Hz



Source: Authors (2022).

Frame 2 presents the bandwidth and the five sub-bands corresponding to the respective levels of decomposition.

Frame 2 – Frequencies corresponding to the 5-level wavelet decomposition

Level	Coef.	Frenquency range	EEG sub-band
1	cD1	64 - 128 Hz	-
2	cD2	32 - 64 Hz	gamma
3	cD3	16 - 32 Hz	beta
4	cD4	8 - 16 Hz	alpha
5	cD5	4 - 8 Hz	theta
5	cA5	0 - 4 Hz	delta

The detail coefficients, highlighted in green, correspond to the sub-bands beta, alpha and theta, and will be considered for the spatial filtering step because they comprise the range of interest for pathological studies in electroencephalography. Figure 16 illustrates the single-channel decomposition process of a subject 1 5-second epoch.

The multiresolution decomposition allows highlighting significant variations in the behavior of the signals in the considered frequency range. To further highlight these variations within the range of 4 and 32 Hz and, mainly, to significantly reduce the dimensional problem, the spatial filtering technique via CSP is applied.

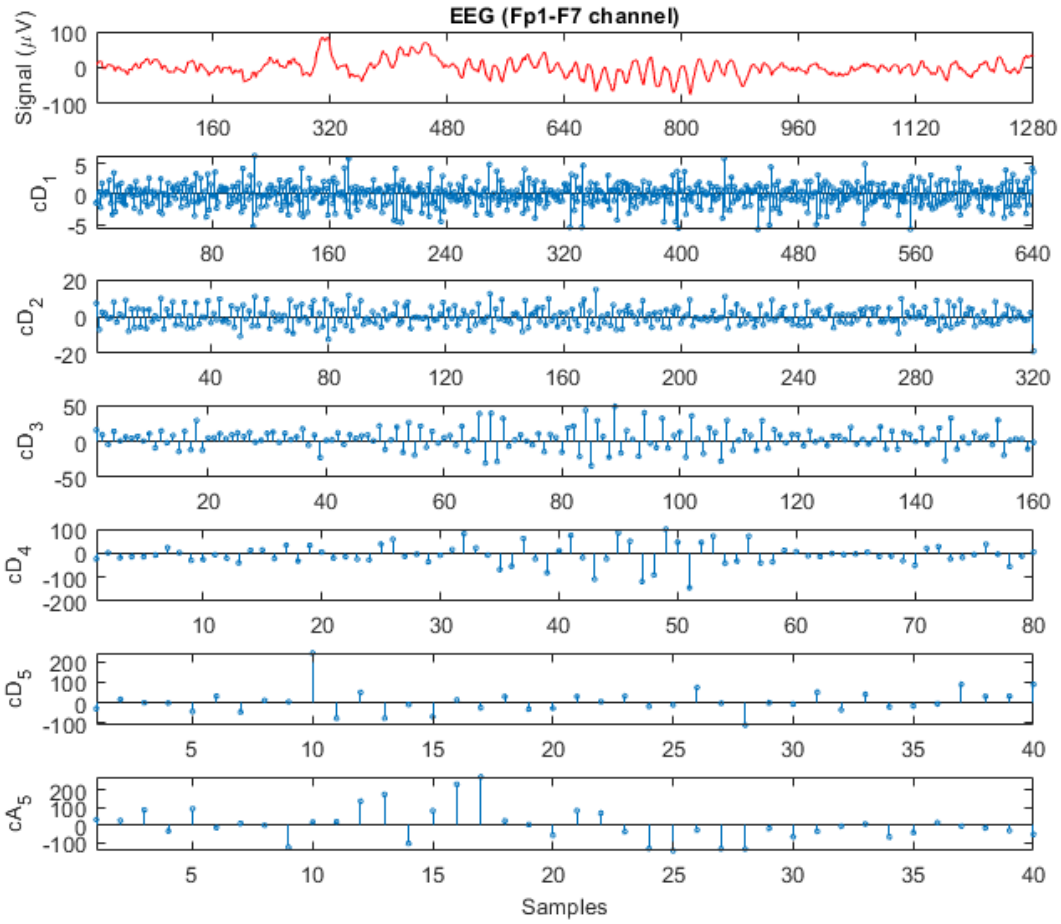
3.3 COMMON SPATIAL PATTERNS FILTERING

In the present CSP filtering process, three spatial filters are obtained, one for each frequency sub-band (beta, alpha and theta). The parameter was set experimentally to $P = 4$ spatial filters ($\mathbf{W}_R = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_{17}, \mathbf{w}_{18}]^T$). In other words, as explained in the CSP method, it is assumed that the first two and last two spatial patterns are related to specific sources of the two EEG periods: preictal and interictal, respectively. Analogously to the Equation (12), the reduced filtered components of each trial $\mathbf{X}$ of wavelet coefficients of a given frequency sub-band, are given by

$$\mathbf{Z}_R = \mathbf{W}_R \mathbf{X} \tag{24}$$

By reducing the data dimension from 18 channels to 4 components, the features extraction can be more convenient, as well as the processing time can be decreased.

Figure 16 – Wavelet decomposition of a EEG sample up to the fifth level using DB4 wavelet. The signals displayed in the following order from top to bottom: single-channel EEG, cD1, cD2, cD3, cD4, cD5 and cA5 coefficients



Source: Authors (2022).

3.4 FEATURE EXTRACTION

The feature extraction of EEG signals for classification aims to represent the signals compactly and with properties that highlight differences between different classes of interest (DUDA; HART; STORK, 2001).

In similar employs, DWT metrics related to entropy and statistical moments are considered, while when there is spatial transformation via CSP, they usually extract attributes from the W projection matrix (ALOTAIBY; ALSHEBEILI; ALOTAIBI, et al., 2017). In the present study, feature extraction will be from the $P = 4$ vectors that compose the compact projection of the coefficient matrices resulting from DWT. Thus, the reduced matrix $\mathbf{Z}_R$ (Equation(25)) will be used for feature extraction.

$$\mathbf{Z}_R = [\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4]^T \tag{25}$$

Three different attributes have been selected, which act as inputs to the classifiers: Shannon Entropy, Logarithm Energy Entropy and Variance logarithm. Entropy has been used as a measure of the complexity or uncertainty of a signal, which allows efficient characterization of the chaotic behavior reflected in the EEG signal (GREENE et al., 2008; SANEI; CHAMBERS, 2013). The three attributes are detailed as follows.

- Shannon Entropy for wavelet decomposition has been presented by Coifman and Wickerhauser (1992). As $\mathbf{z_j}$ is the j-th filtered vector of n_q coefficients resulting from wavelet decomposition on an orthonormal basis, the (unnormalized) entropy F1 of a given EEG sub-band is denoted by

$$\mathrm{F1}(\mathbf{z}_j) = -\sum_{k=1}^{n_q} {\mathbf{z}_{jk}}^2 \log({\mathbf{z}_{jk}}^2) \tag{26}$$

- Logarithm Energy Entropy: is a parameter that has been used in recent studies and has proved to be an effective tool for extracting information about EEG signals (AYDIN; SARAOĞLU; KARA, 2009; GÖKSU, 2018). As well as in Equation (26), the Logarithm Energy Entropy F2 is given by

$$\mathrm{F2}(\mathbf{z}_j) = \sum_{k=1}^{n_q} log({\mathbf{z}_{jk}}^2) \tag{27}$$

- Logarithm of Variance: variance is a statistical parameter that captures the deviation from the mean value of random data. Extracting the logarithm of this measure, Equation (28), is a widely used approach in data from CSP filtering (BLANKERTZ et al., 2008).

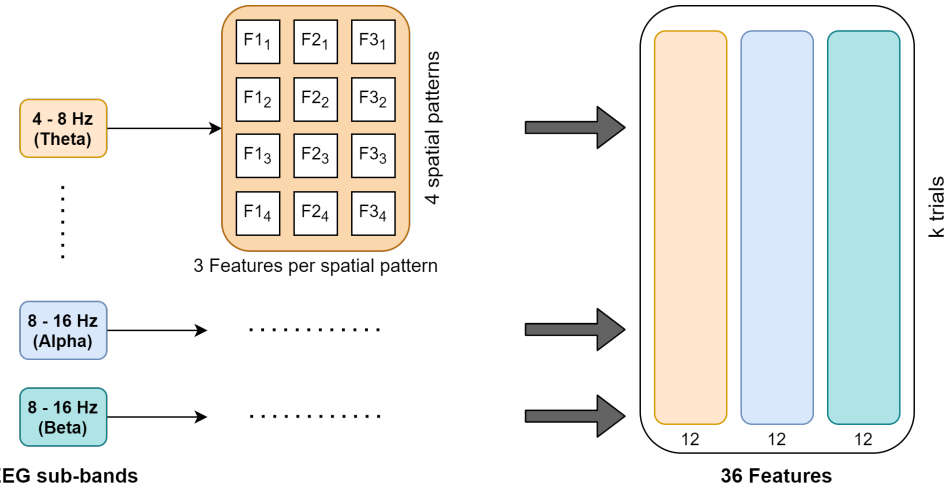
$$F3(\mathbf{z}_j) = log(E[\mathbf{z}_j^2] - \mu^2) \tag{28}$$

where μ is the mean of $\mathbf{z}_j$ and the $E[.]$ operation corresponds to the expected value.

Taking the set of 4 components in $\mathbf{Z}_R$, we have 12 attributes for each frequency sub-band of interest (beta, alpha and theta), totaling 36 attributes. These are extracted from k trials and grouped into a single matrix. Figure 17 represents the formation of a feature matrix for each EEG state, where the blocks represent attribute vectors of the k

trials.

Figure 17 – Feature matrix process for each window of EEG activity



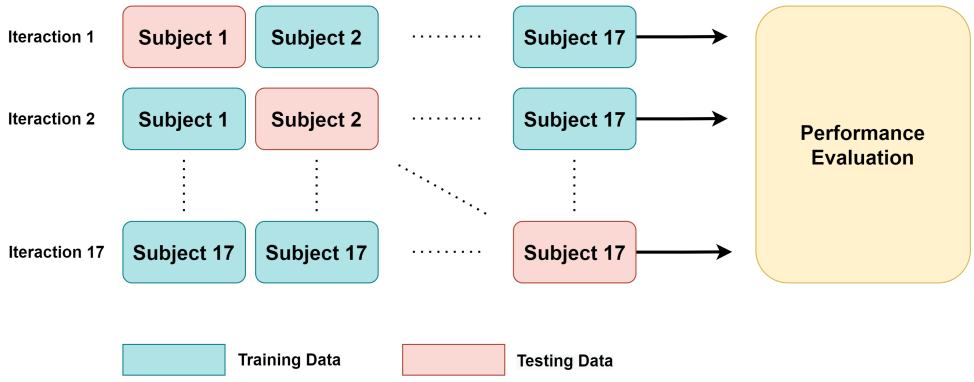
Source: Authors (2022).

3.5 SVM CLASSIFIER

A universal SVM-RBF model is trained to distinguish the state of a five seconds EEG trial, from preictal to interictal, across 17 subjects. Each trial is evaluated regardless of its temporal organization.

The non-specific model approach is carried out using the leave-one-subject-out (LOSO) cross-validation method (Figure 18), which is performed by leaving a different subject out for testing in each iteration, to check whether the model performance is stable for the entire dataset (PEREIRA et al., 2018), and is intended to derive a generalized prediction.

Figure 18 – Protocol of the leave-one-subject-out cross-validation



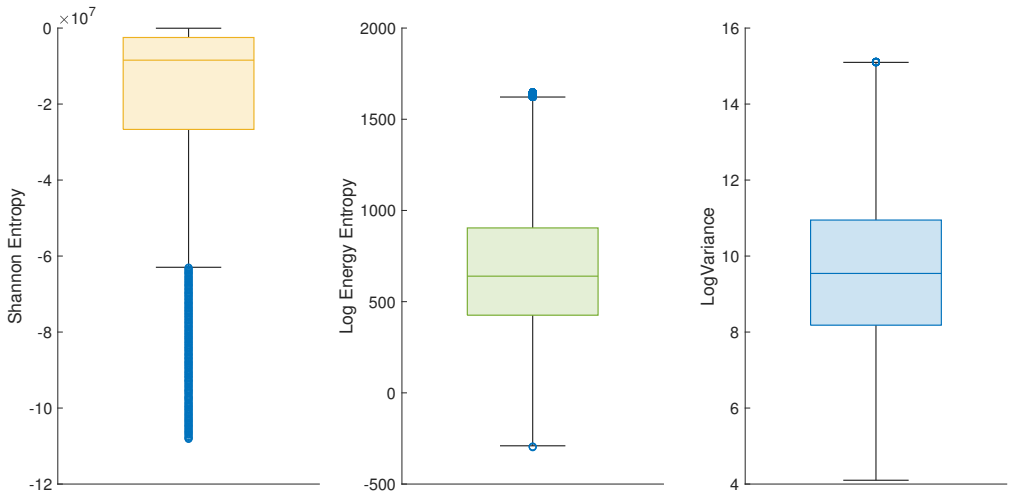
Source: Adapted from Ma, Liu, and Cai (2020).

3.5.1 Features' Standardization

One of the most common steps when designing pattern recognition systems is feature scaling. Ideally, all attributes should vary within the same boundaries, thus avoiding introducing bias in the classifier. Such a strategy can help reduce the convergence time of the (HAYKIN, S. S., 2009) algorithm.

The boxplot is a very useful tool for visualizing and comparing distributions between features, highlighting useful information such as outliers (KING, 2019). In addition, each quartile, median, maximum and minimum of the three different feature types can be observed, as shown in Figure 19.

Figure 19 – Boxplot for the three chosen features



Source: Authors (2022).

The behavior of the EEG signals implies that entropy features have higher magnitudes compared to the rest, whereas logarithm of the variance may be small.

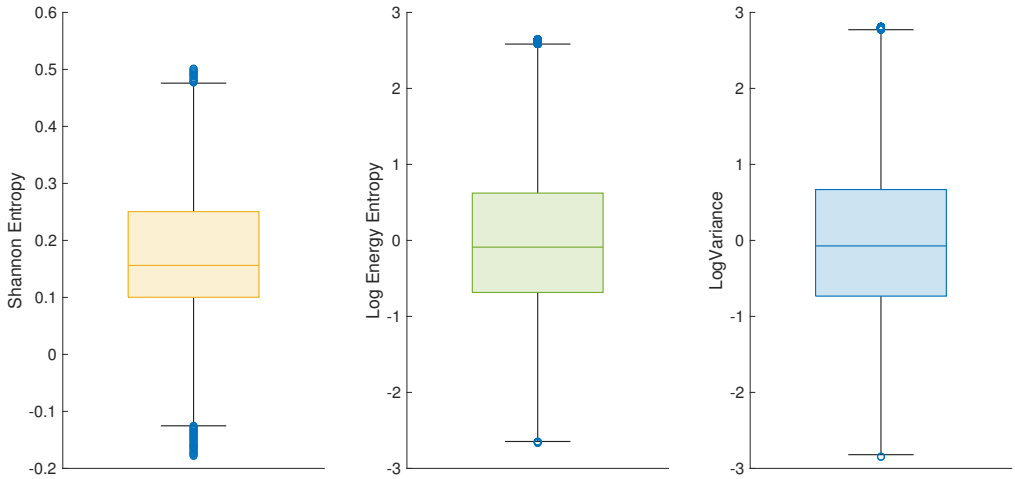
In Machine Learning, standardization is the best practice to oversee the dataset to transform the features to compatible scales. The proposed method uses the z-score – a very useful statistical tool because it allows us to compare two different values from different normal distributions (MOLUGARAM; RAO, 2017).

Z-score is a function provided by MATLAB. For a data with mean μ and standard deviation σ, the z-score of a data point x is computed as:

$$z = \frac{x - \mu}{\sigma} \tag{29}$$

In summary, it can be said that standardization gives the features a comparable scale, but without highlighting outliers. Figure 20 presents a distribution of the feature values after the standardization procedure.

Figure 20 – Features' bloxplot after z-score



Source: Authors (2022).

The standardization process has enabled features to be worked out on very close scales, as seen in Figure 20.

3.5.2 Performance Evaluation

The trained classifiers should return an output 1 for the preictal state prediction (positive label) and 0 for the interictal period (negative label).

Performance evaluation is an essential step and must be executed before any algorithm can be implemented as a medical diagnostic aid. In this study, performance will be quantified by the classifier's ability to discriminate between the two grasping states of interest.

A meaningful evaluation for a classifier response can also be performed according to some metrics, which are given by the information in the confusion matrix (POWERS, 2011), where TP = true-positive, TN = true-negative, FP = false-positive and FN = false-negative.

- Precision: the fraction of trials correctly classified from all those predicted as positive. Precision gives an indication of how certain the model is when

predicting a positive class.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{30}$$

- Recall: the correctly classified fraction of all trials that actually have a positive label. This measurement gives an indication to what rate of a class labels that the model was able to predict.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{31}$$

- F1-score: precision and recall are often combined into a single measure using their harmonic mean.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{32}$$

The range for the F1-score is $[0, 1]$, where 0 is rated the worst and 1 the best. It states the performance of a classifier and also its robustness, meaning that it should not fail with a significant number of samples. A high precision but lower recall would give high accuracy, but fail many samples. The greater F1, the better the performance of a model is (MURPHY, 2013).

3.6 PREDICTION PROTOCOL

The classifier used in this study does not consider the temporal dynamics of the samples. The samples are considered as independent from each other, without the classifier knowing that they are part of a time series.

It is unclear how accurate prediction of brief windows is effective in the clinical goal of prediction when there is no protocol for the regularization of the SVM model output. Hence, once smoothed by a post-processing filter, a specific threshold could determine the patient's state: when the filter output is above the threshold, an alarm will be raised to acknowledge seizure, enabling decision making.

3.6.1 Stationarity Test

The Augmented Dickey-Fuller tests were run at a significance level of 10% and it was verified that not all series (SVM outputs) followed a stationary trend, most times, there was not enough evidence to reject the null hypothesis. This caused the Moving Average method to be discarded for this work. Therefore, only Kalman Filter and Median Filter are used to smooth the outputs of the classifiers.

3.6.2 Filters' parameters settings

The standard deviations are the only design parameter of the Kalman Filter. The value of $\sigma_{KF} = \sigma_w/\sigma_v$ (Equation (21)) was set according to Chisci et al. (2010), being $\sigma_{KF} = 5 \times 10^{-5}$.

For the Moving Median filter parameter r was set to $r = 30$ after several tests. This gives a window length of $m = 2 \times 30 + 1 = 61$, the calculation includes the element in the current position, 30 elements before the current position (corresponding to a period of $30 \times 5 = 150$ seconds), and 30 elements after the current position.

3.6.3 Performance Evaluation

To reinforce correct decision-making, besides the post-processing result, threshold evaluation law was created. The predictors' alarms are generated using the post-processed outputs y_{pp} values according to:

$$a[k] = \begin{cases} 1, & \text{if } y_{pp}[i] \geq L \\ 0, & \text{if } y_{pp}[i] < L \end{cases} \quad \forall i \in [k-5,\ k] \tag{33}$$

where L is an arbitrary threshold, which in this work the values $\{0.7,\ 0.9\}$ are analyzed. Both L and values were set experimentally and consisted of a bold approach to smooth out false alarms.

An alarm can only be raised if, at the current time and in the 5 previous times, $a[k]$ crosses the threshold, therefore strongly suggesting a preictal state. This means that if at least 6 epochs (30 seconds) led to a positive prediction, then the alarm is set.

To evaluate the performance of the proposed prediction protocol, another sta-

tistical measurement of false positive/negative rate and anticipation time to assess the performance of our approach.

- The false-positive rate (FPR) represents the ratio of interictal trials which are currently misclassified as preictal trials to all the interictal horizon. It is an important metric for a seizure predictor, especially when applied to patients with refractory epilepsy. For those patients who suffer constantly from seizures, the wrong prediction can increase the patient's anxiety.
- The false-negative rate (FNR) represents the ratio of preictal trials which are currently misclassified as interictal to all the preictal trials. Both FNR, FPR demonstrate how many false alarms the proposed algorithm would generate.
- The average prediction time (APT) is the ratio of difference between the algorithm-driven alarms within the preictal period and their corresponding actual seizure onsets (in minutes).

4 RESULTS AND DISCUSSION

In this chapter, the final results of the epileptic seizure prediction system are presented as a progression of the research and experiments described in Chapter 3.

4.1 SVM CLASSIFIER

Table 2 presents the results obtained from the proposed method for the SVM classification step. An average precision of 68.8% is achieved.

Table 2 – SVM classifier performance on 17 subjects

Subject (ID)	Precision	Recall	F1-score
1	0.621	0.549	0.583
2	0.728	0.942	0.821
3	0.505	0.771	0.610
4	0.520	0.817	0.636
5	0.611	0.921	0.735
6	0.701	0.880	0.780
7	0.680	0.791	0.731
8	0.672	0.894	0.767
9	0.642	0.914	0.754
10	0.870	0.834	0.852
13	0.646	0.576	0.609
14	0.672	0.741	0.705
15	0.478	0.812	0.602
16	0.654	0.833	0.733
18	0.873	0.853	0.863
21	0.965	0.912	0.938
22	0.859	0.946	0.900
Average	0.688	0.823	0.742
Standard deviation	0.136	0.115	0.109

A heterogeneity in performance was found. The large plurality can justify this regarding the affected regions and the generation mechanisms of these seizures across the subjects. The accuracy rate is lower than the other metrics, showing that the algorithm found it more challenging to detect the preictal activity in most patients.

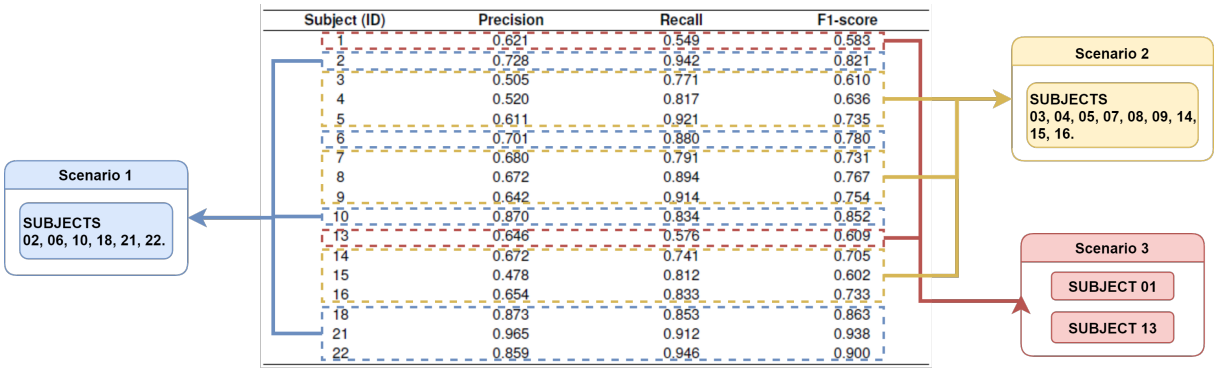
Looking strictly at the results, we can group the subjects into three distinct scenarios:

1. Scenario 1: best performances, precision and recall metrics with high values, resulting in remarkable F1-scores;
2. Scenario 2: low precision, discrepant of the higher remaining metrics;

3. Scenario 3: worst performances, with all metrics having low values.

Figure 21 presents a schematic of the patients studied in each of the respective scenarios presented above.

Figure 21 – Representation of the three distinct behaviors by the studied patients



Subject (ID)	Precision	Recall	F1-score
1	0.621	0.549	0.583
2	0.728	0.942	0.821
3	0.505	0.771	0.610
4	0.520	0.817	0.636
5	0.611	0.921	0.735
6	0.701	0.880	0.780
7	0.680	0.791	0.731
8	0.672	0.894	0.767
9	0.642	0.914	0.754
10	0.870	0.834	0.852
13	0.646	0.576	0.609
14	0.672	0.741	0.705
15	0.478	0.812	0.602
16	0.654	0.833	0.733
18	0.873	0.853	0.863
21	0.965	0.912	0.938
22	0.859	0.946	0.900

Source: Authors (2022).

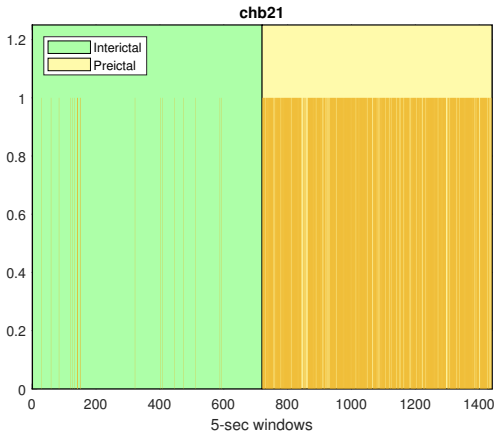
From the scenario segmentation in Figure 21, it can be observed that in only six subjects, the classifier presents a remarkable performance. More than half of the patients fall into scenario 2, where high recall rates are found, evidencing higher difficulty in detecting the interictal period in the samples of these individuals.

Organizing the test samples according to their respective window in time gives continuous 2 hours test segments (both interictal and preictal) per seizure. Examples of the behavior of the classified samples from four subjects' outcomes are presented in Figure 22. Both interictal and preictal periods are highlighted for reference, outlined by the colors green and yellow, respectively, showing that alarms outside this window will be considered as misclassification.

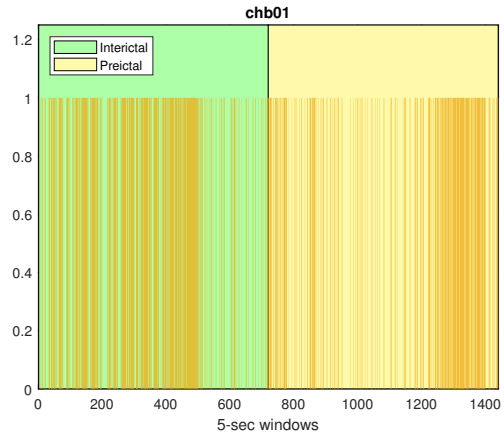
For subject 21 (Figure 22(a)), whose behavior also represents those patients who provided the highest F1-score, it can be observed that positive detection occurs throughout the SPH. The low false positive rate is also observed, reflecting the high precision rates.

For subject 1 (Figure 22(b)), a lot of false-positive labels frequently appeared within the interictal horizon. At the preictal period, the density of true positives increases slightly as the seizure approaches but loses density in the final windows of the prediction horizon. A prediction system capable of generating alarms even a few minutes before a seizure will be entirely satisfactory for keeping patients safe from some dangerous situations.

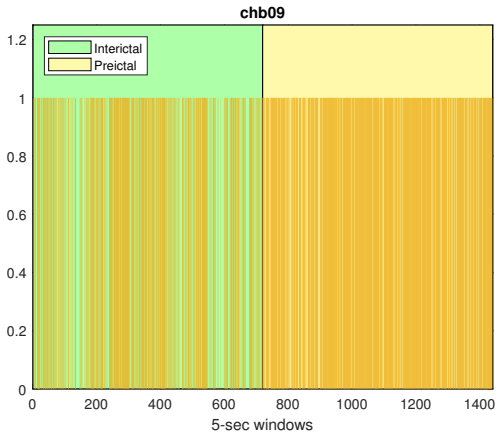
Figure 22 – Temporal dynamics of SVM classifier outputs



(a) Subject 21, Scenario 1



(b) Subject 1, Scenario 3



(c) Subject 9, Scenario 2

Source: Authors (2022).

However, it was expected that recall could vary for different subjects due to the varying quality of their preictal activity behavior. One of the most significant divergences between precision and recall of scenario 2 is found in samples from subject 9 (Figure 22(c)), where the high density of false positives is visible. The prediction horizon has a low false-negative density, reflecting the high recall rate. The temporal dynamics analysis does not show any correlation between the behavior of the classifiers and the evolution between the interictal and preictal periods, which is detrimental to decision making.

For patients with refractory epilepsy, for such excessive false prediction rates, patients may either not consider generated alarms seriously, even those of accurate predictions, or go through intensive psychological stresses.

To eliminate some false positives, and properly arrive at a classification decision to detect seizure events, a post-processing method, called SVM output regularization, needs to be incorporated into the system – we post-process the SVM classification

output using the smoothing techniques Median and Kalman Filters.

4.2 POSTPROCESSING AND ALARM GENERATION

Summary results for all the patients after the two preprocessing methods and alarm-rule (Equation (33)) for both threshold values set are shown in Tables 3 and 4.

Table 3 – Predictor Performance - Threshold at 0.7

Subject	Kalman Filter			Median Filter		
(ID)	FPR/h	FNR/h	APT(min)	FPR/h	FNR/h	APT(min)
1	0.049	0.836	59.321	0.151	0.559	59.021
2	0.024	0.276	59.333	0.189	0.000	59.917
3	0.381	0.553	59.917	0.893	0.179	59.917
4	0.535	0.433	58.542	0.767	0.099	59.417
5	0.220	0.320	59.604	0.544	0.004	59.667
6	0.030	0.440	59.347	0.211	0.001	59.861
7	0.043	0.575	57.583	0.347	0.046	58.833
8	0.088	0.250	59.333	0.360	0.089	59.833
9	0.105	0.355	58.944	0.581	0.002	59.806
10	0.003	0.526	59.042	0.004	0.007	59.500
13	0.014	0.921	57.000	0.082	0.232	57.667
14	0.022	0.697	59.458	0.258	0.110	59.917
15	0.617	0.562	59.517	0.979	0.033	59.917
16	0.034	0.557	57.958	0.312	0.000	59.917
18	0.000	0.422	59.250	0.000	0.078	58.500
21	0.000	0.357	59.333	0.000	0.007	59.500
22	0.000	0.674	59.208	0.002	0.005	59.625
Average	0.127	0.515	58.982	0.334	0.085	59.460
Standard deviation	0.195	0.189	0.776	0.317	0.140	0.626

We can observe that the adoption of a default threshold at L=0.7 could mitigate the false positive problems by the Kalman Filter approach, which were evidenced by the precision metric in Table 2. On the other hand, the false-negative rate was noticeably impaired, given the high recall values, which evidenced the low false-negative rate in the classification step.

With the average false positive rate at 0.334, it can be seen that the Median Filter was able to reduce the false positive rate to the point that it did not hurt the low false-negative rate.

A more striking threshold L=0.9 did not provide a significant improvement in alarm performance. For the Kalman Filter, FNR was reduced by an average of 0.085, but the FNR rate increased slightly to 0.609. The performance for the Median Filter remained intact, evidencing its better suitability for the proposed method.

Table 4 – Predictor Performance - Threshold at 0.9

Subject	Kalman Filter			Median Filter		
(ID)	FPR/h	FNR/h	APT(min)	FPR/h	FNR/h	APT(min)
1	0.023	0.932	47.011	0.151	0.559	59.021
2	0.000	0.378	59.250	0.189	0.000	59.917
3	0.225	0.667	59.917	0.893	0.172	59.917
4	0.503	0.499	56.208	0.767	0.099	59.417
5	0.146	0.440	58.312	0.544	0.004	59.667
6	0.006	0.597	57.861	0.211	0.001	59.861
7	0.013	0.719	56.750	0.347	0.046	58.833
8	0.033	0.278	58.917	0.360	0.089	59.833
9	0.034	0.517	56.028	0.581	0.002	59.806
10	0.000	0.665	58.042	0.004	0.007	59.500
13	0.001	0.994	28.500	0.082	0.233	57.667
14	0.001	0.875	59.292	0.258	0.110	59.917
15	0.481	0.701	58.317	0.979	0.033	59.917
16	0.002	0.724	56.417	0.312	0.000	59.917
18	0.000	0.532	59.000	0.000	0.078	58.500
21	0.000	0.500	59.250	0.000	0.007	59.500
22	0.000	0.328	58.625	0.002	0.005	59.625
Average	0.086	0.609	55.747	0.334	0.085	59.460
Standard deviation	0.164	0.205	7.611	0.317	0.140	0.626

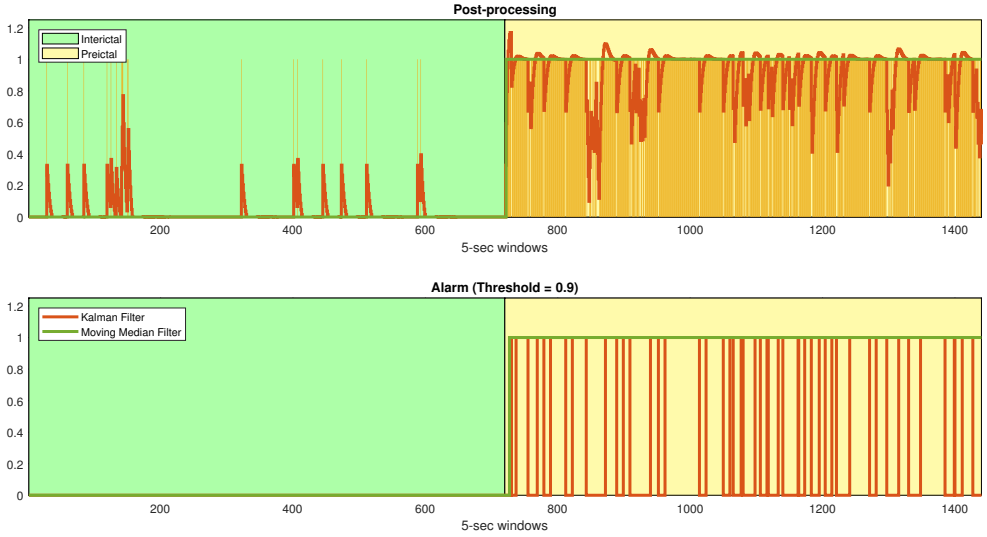
It is essential to note that the high values of the standard deviation of the FNR metric show a high variation of the score across seizures from the patients or seizures from a particular location. For the patients with much smaller FNR, no specific characteristics were found that distinguished them from the others.

Unfortunately, both FNR and FPR parameters cannot be optimally set simultaneously, as improving one means worsening the other. Therefore, a trade-off has to be made in their selection. For example, during a preictal period where the amplitude of the signals increases significantly, one can make predictions by choosing a threshold level. All seizures will be correctly predicted for a relatively low threshold level while FPR will increase. In contrast, higher thresholds will decrease FPR while decreasing sensitivity.

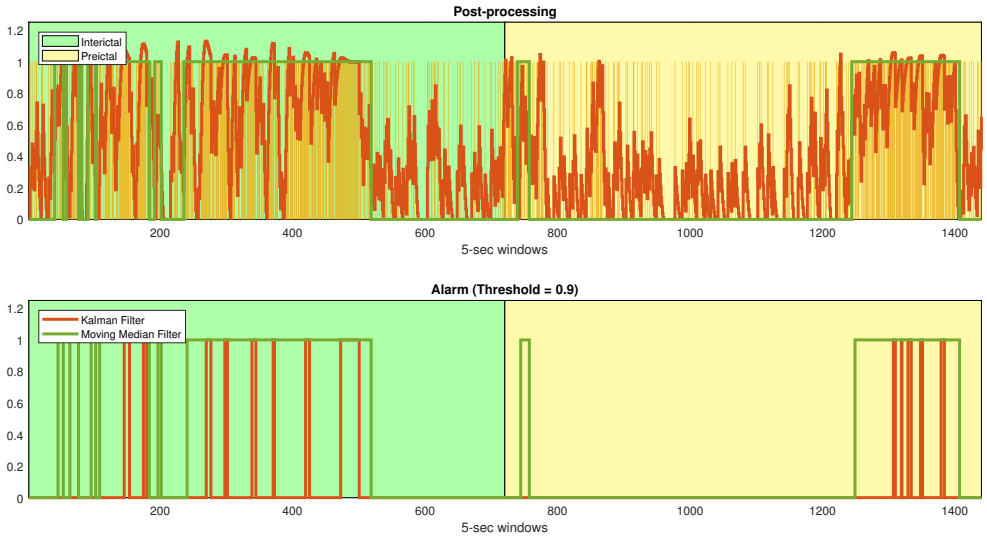
The final stage of alarms generation post-processing methods is illustrated in Figure 23. As shown in this figure, upon exceeding the threshold value 0.9, while fulfilling the conditions presented in Equation (33), an alarm is triggered. At the top panels, the decision values from the SVM classification (in mustard bars) and their Median and Kalman-smoothed outputs (olive and orange colors, respectively) are shown. Meanwhile, the bottom plot stands for final outputs after the alarm rule is evaluated. Once post-processed and classified as positive, the positive output must continue in the next 25-sec prediction horizon, alarming evidence of an epileptic seizure; otherwise, it is classified

into a negative (interictal) group.

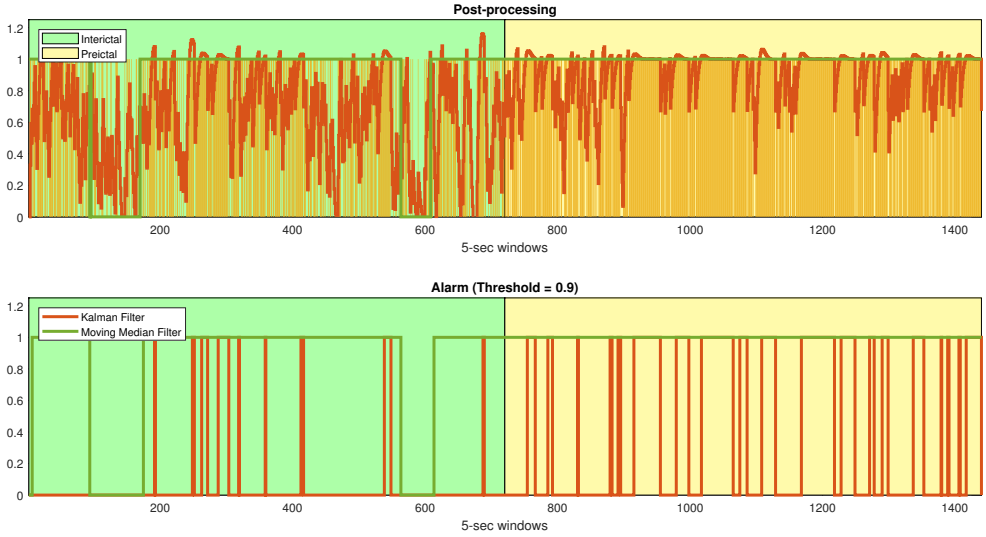
Figure 23 – Examples of postprocessing by the Kalman and Moving Median filters and posterior threshold evaluation in subjects 21 (a), 01 (b) and 09 (c)



(a) Subject 21



(b) Subject 1



(c) Subject 9

Source: Authors (2022).

As shown in Figure 23(a), after the smoothing process, the chattering behavior imposed on the SVM decision variable has been mitigated. For the other cases (Figures 23(b) and 23(c)), where the false detection density is very high, the presence of false alarms is inevitable.

In gereral, the Median filter is less volatile, while the Kalman filter is sensitive to the slightest change in the density of positive outputs of the classifier. However, the number of false alarms raised by KF does not hinder its applicability for some patients.

It can be concluded that the MF approach is more conservative concerning the raising of alarms because it considers a longer memory, and because of its particular constraints (rules) on the times where alarms are possible to be raised. While MF considers a range from 180 to 360 seconds, the KF approach is based on just the past output sample, and its derivative (rate of change), *i.e.*, a much shorter memory is considered.

Another important aspect is how patient-specific characteristics, such as seizure types, age, and gender, can directly affect seizure prediction performance. Correlations on the seizure type or focal aware seizures were not found to account for high or low performance in certain subjects. This observation could be helpful for clinical trial consideration; *e.g.*, focus on patients with aware focal seizures first.

Frame 3 summarizes some state-of-art methods using the same CHB-MIT

database in chronological order. N/M represents "not mentioned" entries.

Frame 3 – Comparison of summary approaches for epileptic seizure predictors

Method	Features's nature	Classifier	Precision	FPR/h	Postprocessing	Statistical Validation
Khan et al. (2017)	DWT	CNN	0.878	0.14	N/M	Random Predictor
Alotaiby, Alshebeili, Alotaibi, et al. (2017)	CSP	LDA	0.81	0.47	Median Filter	N/M
Syed Muhammad Usman, Muhammad Usman, and Fong (2017)	Hjorth parameters Approximate entropy Statistical moments Spectral moments	SVM	0.922	N/M	N/M	N/M
Syed Muhammad Usman and Hassan (2018)	Hjorth parameters Statistical moments	KNN Naive Bayes SVM	0.974 0.907 0.971	N/M	N/M	t-test
Yuan Zhang et al. (2019)	CSP WPD	CNN	0.922	0.12	Kalman Filter	t-test
Syed Muhammad Usman, Khalid, and Aslam (2020)	CNN	SVM	0.927		N/M	N/M
Proposed method	DWT CSP Entropy features Statistical moment	SVM	0.688	0.086 0.334	Median Filter Kalman Filter	No

This study produced a generalized seizure prediction with low average precision relative to related work. On the other hand, the false positive rate can be mitigated in the post-processing stage, which compares favorably with the results from the patient-specific studies listed, including those using Deep Learning techniques. It is worth repeating that generalized prediction algorithms are at a performance disadvantage because of the heterogeneity of seizures between patients.

It can be observed that the vast majority of notable works with the CHB-MIT dataset do not use post-processing techniques and focus only on the classifiers' performance. As seen in the methodology of this paper, such techniques are of paramount importance for understanding and smoothing the temporal dynamics of a predictor.

Because this is a primary study of a seizure predictor, an important mechanism proposed to evaluate the performance of predictors concerning their evaluation for

clinical application was out of the project scope: the comparison of predictors with a random predictor. It is a recent trend and should also be implemented to understand if the proposed methods perform above a reference probability (MORMANN; ANDRZEJAK, et al., 2007).

5 CONCLUSIONS

This work proposes a novel approach for non-patient-specific epileptic seizure prediction by scalp EEG multichannel records.

The primary goal of this paper is to explore the roles of four preprocessing techniques to determine their classification performance. Preprocessing EEG signals is performed by applying the DWT for extracting features in a range of frequency sub-bands of interest. CSP-based features extraction made it possible to combine the information from the 18 available channels to maximize the variance between the interictal and preictal periods, reducing the problem dimensionality to 4 vectors. This approach contributes to developing robust classification models suitable for practical seizure prediction systems, where information about the entire scalp is combined and, even with reduced dimensions, the attributes essential for class discrimination remain.

The number of different EEG features used in the literature is considerable. Considering the primary goal of this research, this paper does not determine the optimal type or combination of features. However, it focuses on constructing robust models based on two types of entropy features and a statistical one that is well known and relatively simple to extract.

With an average precision of 68.8%, the SVM-RBF classifier demonstrates that new procedures should be addressed to enable the non-patient-specific approach. On the other hand, the soft post-processing techniques demonstrated upward directions for seizure prediction, managing to achieve an false-positive rate of 0.334 for the Median Filter.

Our work is not yet comparable with the-state-of-art methods using the same CHB-MIT dataset. Nevertheless, this method opens multiple avenues for further studies to investigate more suitable techniques for patient-independent seizure alarm systems.

Complementary approaches are therefore needed to acquire additional insights into the preictal period. We have assumed a uniform preictal period (60 minutes) for all patients in this study, but this might not be true as the preictal period may vary from patient to patient. A better understanding of the dynamic transitions into and out of seizures will contribute to the prediction and comprehension of what composes a seizure. An exploratory study regarding the quality of the features and a further validation via another machine learning technique is also a valid alternative.

A common issue with initial studies concerned the statistical validity of the algorithms' performance (MORMANN; ANDRZEJAK, et al., 2007). According to Andrzejak et al. (2009), many early seizure prediction models often claimed excellent results but were later proven to be irreproducible or were unsubstantiated. In future investigations, it is suggested to refine post-processing parameters and apply another statistical validation technique to verify the superiority of the proposed method for seizure prediction, comparing our results with those achieved from the analytical random predictor introduced by Schelter et al. (2006).

Another procedure often carried out as part of the performance measurement of seizure prediction systems is the statistical validation of the applicability of the systems. This statistical test aims to verify whether the developed algorithm has sensitivity higher than the acute sensitivity, *i.e.* sensitivity under the same conditions as the random predictor. Thus, a specific predictor, such as those developed in this study, should have sensitivity superior to that of non-specific predictors, such as a random predictor (WINTERHALDER et al., 2003).

REFERENCES

ACHARYA, U. R.; HAGIWARA, Y.; ADELI, H. Automated seizure prediction. **Epilepsy & Behavior**, Elsevier, v. 88, p. 251–261, 2018.

ACHARYA, U. R.; SREE, S. V., et al. Automated EEG analysis of epilepsy: a review. **Knowledge-Based Systems**, Elsevier, v. 45, p. 147–165, 2013.

ADELI, H.; ZHOU, Z.; DADMEHR, N. Analysis of EEG records in an epileptic patient using wavelet transform. **Journal of Neuroscience Methods**, v. 123, n. 1, p. 69–87, 2003.

AES. **American Epilepsy Society (AES) Seizure Prediction Challenge**. Chicago, IL: Kaggle, Aug. 2014. Accessed: 14-10-2020. Available from: https://www.kaggle.com/c/seizure-prediction.

AJANI, T. S.; IMOIZE, A. L.; ATAYERO, A. A. An overview of machine learning within embedded and mobile devices–optimizations and applications. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 21, n. 13, p. 4412, 2021.

ALI, H. M. MRI medical image denoising by fundamental filters. **High-Resolution Neuroimaging-Basic Physical Principles and Clinical Applications**, Intechopen, v. 14, p. 111–124, 2018.

ALOTAIBY, T. N.; ALSHEBEILI, S. A.; ALSHAWI, T., et al. EEG seizure detection and prediction algorithms: a survey. **EURASIP Journal on Advances in Signal Processing**, Springer Science and Business Media LLC, v. 2014, n. 1, p. 1–21, 2014.

ALOTAIBY, T. N.; ALSHEBEILI, S. A.; ALOTAIBI, F. M., et al. Epileptic Seizure Prediction Using CSP and LDA for Scalp EEG Signals. **Computational Intelligence and Neuroscience**, Hindawi Limited, v. 2017, p. 1–11, 2017.

ANDRZEJAK, R. G. et al. Seizure prediction: Any better than chance? **Clinical Neurophysiology**, Elsevier, v. 120, n. 8, p. 1465–1478, 2009.

AYDIN, S.; SARAOĞLU, H. M.; KARA, S. Log Energy Entropy-Based EEG Classification with Multilayer Neural Networks in Seizure. **Annals of Biomedical Engineering**, Springer Science and Business Media LLC, v. 37, n. 12, p. 2626–2630, 2009.

BANDARABADI, M.; RASEKHI, J., et al. On the proper selection of preictal period for seizure prediction. **Epilepsy & Behavior**, Elsevier, v. 46, p. 158–166, 2015.

BANDARABADI, M.; TEIXEIRA, C. A., et al. Epileptic seizure prediction using relative spectral power features. **Clinical Neurophysiology**, Elsevier, v. 126, n. 2, p. 237–248, 2015.

BISHOP, C. M. **Pattern Recognition and Machine Learning (Information Science and Statistics)**. Berlin, Heidelberg: Springer, 2006.

BLANKERTZ, B. et al. Optimizing Spatial filters for Robust EEG Single-Trial Analysis. **IEEE Signal Processing Magazine**, IEEE, v. 25, n. 1, p. 41–56, 2008.

BURGES, C. J. C. A Tutorial on Support Vector Machines for Pattern Recognition. **Data Mining and Knowledge Discovery**, Springer Science and Business Media LLC, v. 2, n. 2, p. 121–167, 1998.

CHISCI, L. et al. Real-time epileptic seizure prediction using AR models and support vector machines. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 57, n. 5, p. 1124–1132, 2010.

COIFMAN, R. R.; WICKERHAUSER, M. V. Entropy-based algorithms for best basis selection. **IEEE Transactions on Information Theory**, IEEE, v. 38, n. 2, p. 713–718, 1992.

CONDORI, K. A.; URQUIZO, E. C.; DIAZ, D. A. Embedded Brain Machine Interface based on motor imagery paradigm to control prosthetic hand. In: 2016 IEEE ANDESCON. Arequipa, Peru: IEEE, 2016. P. 1–4.

DAOUD, H.; BAYOUMI, M. A. Efficient Epileptic Seizure Prediction Based on Deep Learning. **IEEE Transactions on Biomedical Circuits and Systems**, IEEE, v. 13, n. 5, p. 804–813, 2019.

DAUBECHIES, I. **Ten Lectures on Wavelets**. USA: Society for Industrial and Applied Mathematics, 1992.

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American statistical association**, Taylor & Francis, v. 74, 366a, p. 427–431, 1979.

DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. **Econometrica: journal of the Econometric Society**, JSTOR, p. 1057–1072, 1981.

DIREITO, B. et al. A Realistic Seizure Prediction Study Based on Multiclass SVM. **International Journal of Neural Systems**, World Scientific Pub Co Pte Lt, v. 27, n. 03, p. 1750006, 2017.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. 2. ed. New York: Wiley, 2001.

DUNCAN, J. S. et al. Adult epilepsy. **The Lancet**, Elsevier, v. 367, n. 9516, p. 1087–1100, 2006.

FENG, J. K. et al. An optimized channel selection method based on multifrequency CSP-rank for motor imagery-based BCI system. **Computational intelligence and neuroscience**, Hindawi, v. 2019, 2019.

FISCH, B. J.; SPEHLMANN, R. EEG primer: basic principles of digital and analog EEG. **Elsevier Health Sciences**, p. 23–27, 1999.

FISHER, R. S. et al. Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). **Epilepsia**, v. 46, n. 4, p. 470–472, 2005.

FREEMAN, W.; QUIROGA, R. Q. **Imaging brain function with EEG: advanced temporal and spatial analysis of electroencephalographic signals**. New York, NY: Springer Science & Business Media, 2012.

FREESTONE, D. R.; KAROLY, P. J.; COOK, M. J. A forward-looking review of seizure prediction. **Current opinion in neurology**, LWW, v. 30, n. 2, p. 167–173, 2017.

FU, R. et al. Automatic Detection of Epileptic Seizures in EEG Using Sparse CSP and Fisher Linear Discrimination Analysis Algorithm. **Journal of Medical Systems**, Springer, v. 44, n. 2, p. 1–13, 2020.

FUKUNAGA, K. **Introduction to statistical pattern recognition**. San Diego, CA, USA: Elsevier, 1990.

GADHOUMI, K.; LINA, J.-M.; GOTMAN, J. Seizure prediction in patients with mesial temporal lobe epilepsy using EEG measures of state similarity. **Clinical Neurophysiology**, Elsevier, v. 124, n. 9, p. 1745–1754, 2013.

GANDHI, T.; PANIGRAHI, B. K.; ANAND, S. A comparative study of wavelet families for EEG signal classification. **Neurocomputing**, Elsevier, v. 74, n. 17, p. 3051–3057, 2011.

GÖKSU, H. EEG based epileptiform pattern recognition inside and outside the seizure states. **Biomedical Signal Processing and Control**, Elsevier, v. 43, p. 204–215, 2018.

GOLDBERGER, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet. **Circulation**, Ovid Technologies (Wolters Kluwer Health), v. 101, n. 23, 2000.

GOTMAN, J.; IVES, J.; GLOOR, P. **Electroencephalography and Clinical Neurophysiology**, Elsevier, v. 52, n. 6, p. 626–639, 1981.

GOTMAN, J. Automatic recognition of epileptic seizures in the EEG. **Electroencephalography and Clinical Neurophysiology**, Elsevier, v. 54, n. 5, p. 530–540, 1982.

GREENE, B. et al. A comparison of quantitative EEG features for neonatal seizure detection. **Clinical Neurophysiology**, Elsevier BV, v. 119, n. 6, p. 1248–1261, June 2008.

GUNN, S. R. et al. Support vector machines for classification and regression. **ISIS technical report**, v. 14, n. 1, p. 5–16, 1998.

HALFORD, J. J. Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ECG interpretation. **Clinical Neurophysiology**, Elsevier, v. 120, n. 11, p. 1909–1915, Nov. 2009.

HAYKIN, S. **Kalman filtering and neural networks**. USA: John Wiley & Sons, 2004. v. 47.

HAYKIN, S. S. **Neural networks and learning machines**. Third. Upper Saddle River, NJ: Pearson Education, 2009.

HENRY, J. C. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, Fifth Edition. **Neurology**, Ovid Technologies (Wolters Kluwer Health), v. 67, n. 11, p. 2092–2092, 2006.

HUSSAIN, I.; PARK, S. J. HealthSOS: Real-Time Health Monitoring System for Stroke Prognostics. **IEEE Access**, IEEE, v. 8, p. 213574–213586, 2020.

INDIRADEVI, K. et al. A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram. **Computers in Biology and Medicine**, Elsevier BV, v. 38, n. 7, p. 805–816, July 2008.

JOLLIFFE, I. T. Discarding variables in a principal component analysis. I: Artificial data. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, Wiley Online Library, v. 21, n. 2, p. 160–173, 1972.

KALMAN, R. E. A new approach to linear filtering and prediction problems. **Journal of Basic Engineering**, ASME International, v. 82, n. 1, p. 35–45, 1960.

KHAN, H. et al. Focal onset seizure prediction using convolutional networks. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 65, n. 9, p. 2109–2118, 2017.

KING, A. **Statistics for Biomedical Engineers and Scientists**. London: Academic Press, 2019. ISBN 9780081029398.

KIRAL-KORNEK, I. et al. Epileptic seizure prediction using big data and deep learning: toward a mobile system. **EBioMedicine**, Elsevier, v. 27, p. 103–111, 2018.

KOLES, Z. J.; LAZAR, M. S.; ZHOU, S. Z. Spatial patterns underlying population differences in the background EEG. **Brain Topography**, Springer Science and Business Media LLC, v. 2, n. 4, p. 275–284, 1990.

KUHLMANN, L. et al. Seizure prediction—ready for a new era. **Nature Reviews Neurology**, Nature Publishing Group, v. 14, n. 10, p. 618–630, 2018.

KUTZ, M. **Biomedical Engineering and Design Handbook, Volume 1**. New York, USA: McGraw-Hill Education, 2009.

LATHI, B. P. **Modern Digital and Analog Communication Systems 3e Osece**. 3rd. USA: Oxford University Press, Inc., 1998.

LEFFERTS, E. J.; MARKLEY, F. L.; SHUSTER, M. D. Kalman filtering for spacecraft attitude estimation. **Journal of Guidance, Control, and Dynamics**, v. 5, n. 5, p. 417–429, 1982.

LEON-GARCIA, A. **Probability, Statistics, and Random Processes for Electrical Engineering**. Third. Upper Saddle River, NJ: Pearson/Prentice Hall, 2008.

LIBENSON, M. **Practical approach to electroencephalography**. Philadelphia, PA: Elsevier/Saunders, 2010.

LIM, S. H. et al. A novel method for tracking and analysis of EEG activation across brain lobes. **Biomedical Signal Processing and Control**, Elsevier, v. 40, p. 488–504, 2018.

LODDENKEMPER, T.; LÜDERS, H. History of epilepsy and seizure classification. In: LUDERS, H. O. (Ed.). **Textbook of Epilepsy Surgery**. 1. ed. London, UK: CRC Press, 2008. P. 160–173.

LÜTKEPOHL, H. **Applied time series econometrics**. Cambridge, UK New York: Cambridge University Press, 2004. ISBN 052183919X.

LYNCH, K. **Embedded computing and mechatronics with the PIC32 microcontroller**. Waltham, MA: Newnes is an imprint of Elsevier, 2015. ISBN 978-0-12-420165-1.

MA, Y.; LIU, D.; CAI, L. Deep Learning-Based Upper Limb Functional Assessment Using a Single Kinect v2 Sensor. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 20, n. 7, p. 1903, 2020.

MALAVER, W. J. L. Wavelet multiresolution analysis and dyadic scalogram for detection of epileptiform paroxysms in electroencephalographic signals. **Research on Biomedical Engineering**, SciELO Brasil, v. 33, n. 3, p. 195–201, 2017.

MALLAT, S. G. **A wavelet tour of signal processing : the sparse way**. Amsterdam Boston: Elsevier/Academic Press, 2009.

MALLAT, S. G. A theory for multiresolution signal decomposition: the wavelet representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE, v. 11, n. 7, p. 674–693, 1989.

MALMIVUO, J.; PLONSEY, R. **Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields**. UK: Oxford University Press, 1995.

MATLAB. **Statistics and Machine Learning Toolbox User's Guide**. Natick, MA: Mathworks Inc., 2018.

MISITI, M. et al. **Wavelet Toolbox MATLAB users guide Version 4**. Natick, MA: Mathworks Inc., 2008.

MOLUGARAM, K.; RAO, G. S. Random Variables. In: STATISTICAL Techniques for Transportation Engineering. Oxford, UK: Elsevier, 2017. P. 113–279.

MORETTIN, P. A. **Econometria financeira: um curso em séries temporais financeiras**. São Paulo: Editora Blucher, 2017.

MORMANN, F.; ANDRZEJAK, R. G., et al. Seizure prediction: the long and winding road. **Brain**, Oxford University Press, v. 130, n. 2, p. 314–333, 2007.

MORMANN, F.; KREUZ, T., et al. On the predictability of epileptic seizures. **Clinical neurophysiology**, Elsevier, v. 116, n. 3, p. 569–587, 2005.

MURPHY, K. P. **Machine learning : a probabilistic perspective**. Cambridge, MA: MIT Press, 2013.

NODES, T.; GALLAGHER, N. Median filters: Some modifications and their properties. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, IEEE, v. 30, n. 5, p. 739–746, 1982.

OCAK, H. Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. **Expert Systems with Applications**, Elsevier, v. 36, n. 2, p. 2027–2036, 2009.

OLIVA, J. T.; ROSA, J. L. G. The use of one-class classifiers for differentiating healthy from epileptic EEG segments. In: IEEE. 2017 International Joint Conference on Neural Networks (IJCNN). [S.l.: s.n.], 2017. P. 2956–2963.

OPPENHEIM, A. **Signals & systems**. Upper Saddle River, N.J: Prentice Hall, 1997.

OPPENHEIM, A. V.; SCHAFER, R. W. **Discrete-Time Signal Processing**. 3rd. USA: Prentice Hall Press, 2009.

PARK, Y.; CHUNG, W. Frequency-optimized local region common spatial pattern approach for motor imagery classification. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, IEEE, v. 27, n. 7, p. 1378–1388, 2019.

PARK, Y.; LUO, L., et al. Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. **Epilepsia**, Wiley Online Library, v. 52, n. 10, p. 1761–1770, 2011.

PEREIRA, E. T. et al. Empirical evidence relating EEG signal duration to emotion classification performance. **IEEE Transactions on Affective Computing**, IEEE, v. 12, n. 1, p. 154–164, 2018.

PITAS, I. **Nonlinear Digital Filters : Principles and Applications**. Boston, MA: Springer US, 1990. ISBN 978-1-4757-6017-0.

POISSON, O.; RIOUAL, P.; MEUNIER, M. New signal processing tools applied to power quality analysis. **IEEE Transactions on Power Delivery**, IEEE, v. 14, n. 2, p. 561–566, 1999.

POWERS, D. M. W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. **Journal of Machine Learning Technologies**, Bioinfo Publications, v. 2, p. 37–63, 2011.

QUITADAMO, L. et al. Support vector machines to detect physiological patterns for EEG and EMG-based human–computer interaction: a review. **Journal of Neural Engineering**, IOP Publishing, v. 14, n. 1, p. 011001, 2017.

RAFIEE, J. et al. Wavelet basis functions in biomedical signal processing. **Expert systems with Applications**, Elsevier, v. 38, n. 5, p. 6190–6201, 2011.

RASHEED, K. et al. Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review. **IEEE Reviews in Biomedical Engineering**, IEEE, v. 14, 2021.

ROGOWSKI, Z.; GATH, I.; BENTAL, E. On the prediction of epileptic seizures. **Biological cybernetics**, Springer, v. 42, n. 1, p. 9–15, 1981.

RÚA, S. et al. Machine learning algorithms for real time arrhythmias detection in portable cardiac devices: microcontroller implementation and comparative analysis. In: 2012 XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA). Medellin, Colombia: IEEE, 2012. P. 50–55.

RUKHSAR, S. et al. Patient-Specific Epileptic Seizure Prediction in Long-Term Scalp EEG Signal Using Multivariate Statistical Process Control. **IRBM**, Elsevier, v. 40, n. 6, p. 320–331, 2019.

SALANT, Y.; GATH, I.; HENRIKSEN, O. Prediction of epileptic seizures from two-channel EEG. **Medical and Biological Engineering and Computing**, Springer, v. 36, n. 5, p. 549–556, 1998.

SANEI, S.; CHAMBERS, J. A. **EEG Signal Processing**. Hoboken, USA: Wiley-Blackwell, 2013.

SCHELTER, B. et al. Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, American Institute of Physics, v. 16, n. 1, p. 013108, 2006.

SHANNON, C. E. Communication in the presence of noise. **Proceedings of the IRE**, IEEE, v. 37, n. 1, p. 10–21, 1949.

SHOEB, A. H. **Application of machine learning to epileptic seizure onset detection and treatment**. 2009. PhD thesis – Massachusetts Institute of Technology.

SIDDIQUI, M. K. et al. A review of epileptic seizure detection using machine learning classifiers. **Brain informatics**, Springer, v. 7, p. 1–18, 2020.

SIULY, S.; LI, Y.; ZHANG, Y. Electroencephalogram (EEG) and Its Background. In: EEG Signal Analysis and Classification. Cham, Switzerland: Springer, 2016. P. 3–21.

SMITH, S. **Digital signal processing : a practical guide for engineers and scientists**. Amsterdam Boston: Newnes, 2003. ISBN 978-0-7506-7444-7.

SÖRNMO, L.; LAGUNA, P. **Bioelectrical signal processing in cardiac and neurological applications**. San Diego, CA: Academic Press, 2005. v. 8.

STRZELCZYK, A. et al. Costs of epilepsy and their predictors: Cross-sectional study in Germany and review of literature. **Epileptology**, v. 1, n. 1, p. 55–60, 2013. Pharmacological Treatment Strategies. ISSN 2212-8220.

TATUM, W. O. Ellen r. grass lecture: Extraordinary eeg. **The Neurodiagnostic Journal**, Taylor & Francis, v. 54, n. 1, p. 3–21, 2014.

TEIXEIRA, C. A. et al. Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients. **Computer methods and programs in biomedicine**, Elsevier, v. 114, n. 3, p. 324–336, 2014.

TÉLLEZ-ZENTENO, J. F.; DHAR, R.; WIEBE, S. Long-term seizure outcomes following epilepsy surgery: a systematic review and meta-analysis. **Brain**, Oxford University Press, v. 128, n. 5, p. 1188–1198, 2005.

THERRIEN, C. W. **Discrete random signals and statistical signal processing**. Englewood Cliffs, NJ: Prentice Hall, 1992.

THURMAN, D. J. et al. Standards for epidemiologic studies and surveillance of epilepsy. **Epilepsia**, v. 52, p. 2–26, 2011.

USMAN, S. M.; HASSAN, A. Efficient Prediction and Classification of Epileptic Seizures Using EEG Data Based on Univariate Linear Features. **J. Comput.**, v. 13, n. 6, p. 616–621, 2018.

USMAN, S. M.; KHALID, S.; AKHTAR, R., et al. Using scalp EEG and intracranial EEG signals for predicting epileptic seizures: Review of available methodologies. **Seizure**, Elsevier, v. 71, p. 258–269, 2019.

USMAN, S. M.; KHALID, S.; ASLAM, M. H. Epileptic Seizures Prediction Using Deep Learning Techniques. **IEEE Access**, IEEE, v. 8, p. 39998–40007, 2020.

USMAN, S. M.; USMAN, M.; FONG, S. Epileptic Seizures Prediction Using Machine Learning Methods. **Computational and Mathematical Methods in Medicine**, Hindawi, v. 2017, p. 1–10, 2017.

VAN DER MAATEN, L.; POSTMA, E.; VAN DEN HERIK, J., et al. Dimensionality reduction: a comparative. **J Mach Learn Res**, v. 10, n. 66-71, p. 13, 2009.

VAN DRONGELEN, W. **Signal processing for neuroscientists**. London: Academic Press, 2018.

VAPNIK, V. N. **The nature of statistical learning theory**. New York: Springer, 2000.

VERGULT, A. et al. Improving the Interpretation of Ictal Scalp EEG: BSS-CCA Algorithm for Muscle Artifact Removal. **Epilepsia**, Wiley, v. 48, n. 5, p. 950–958, 2007.

VIGLIONE, S.; WALSH, G. Proceedings: Epileptic seizure prediction. **Electroencephalography and clinical neurophysiology**, v. 39, n. 4, p. 435–436, 1975.

WANG, L. et al. Analysis and classification of speech imagery EEG for BCI. **Biomedical signal processing and control**, Elsevier, v. 8, n. 6, p. 901–908, 2013.

WANG, T.; ZHANG, L.; HU, W. Bridging deep and multiple kernel learning: A review. **Information Fusion**, Elsevier, v. 67, p. 3–13, 2021.

WANG, Y.; GAO, S.; GAO, X. Common Spatial Pattern Method for Channel Selelction in Motor Imagery Based Brain-computer Interface. In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. Shanghai, China: IEEE, 2006. P. 5392–5395.

WEEKS, M. **Digital signal processing using MATLAB & Wavelets**. Burlington, MA: Jones & Bartlett Publishers, 2010.

WELCH, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. **IEEE Transactions on Audio and Electroacoustics**, IEEE, v. 15, n. 2, p. 70–73, 1967.

WHO. **Epilepsy: a public health imperative**. Geneva: World Health Organization, 2019. 146 p.

WINTERHALDER, M. et al. The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. **Epilepsy & Behavior**, Elsevier, v. 4, n. 3, p. 318–325, 2003.

WOOLDRIDGE, J. M. **Introductory econometrics: A modern approach**. Boston, MA: Cengage Learning, 2015.

ZHANG, Y. et al. Epilepsy Seizure Prediction on EEG Using Common Spatial Pattern and Convolutional Neural Network. **IEEE Journal of Biomedical and Health Informatics**, IEEE, v. 24, n. 2, p. 465–474, 2019.