UNIVERSIDADE D
COIMBRA

Fábio André da Costa Lopes

Towards Adaptive Seizure Prediction using Automatic Electroencephalogram Denoising and Deep Learning Approaches

PhD Thesis in Informatics Engineering, Intelligent Systems, supervised by Professor César Alexandre Domingues Teixeira and Doctor Engineer Matthias Dümpelmann, and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra

August 2023

Towards Adaptive Seizure Prediction using Automatic Electroencephalogram Denoising and Deep Learning Approaches

Fábio André da Costa Lopes

August, 2023

The studies presented in this thesis were carried out at the Centre for Informatics and Systems of the University of Coimbra (CISUC) at the Informatics Department, Faculty of Sciences and Technology, University of Coimbra, Portugal and at the Epilepsy Center at the Department of Neurosurgery, University Medical Center Freiburg, Germany. The research was performed with financial support of the following institutions/programs:

- Grant from European Union project WELMO (H2020-825572).
- Ph.D. grant 2020.04537.BD from the FCT – Foundation for Science and Technology, I.P..
- Project CISUC (UID/CEC/00326/2020) and project RECoD (PTDC/EEI-EEE/5788/2020), both financed by national funds through the FCT – Foundation for Science and Technology, I.P..

Faculty of Sciences and Technology

University of Coimbra

Towards Adaptive Seizure Prediction using Automatic Electroencephalogram Denoising and Deep Learning Approaches

Fábio André da Costa Lopes

Supervisor:
César Teixeira

Co-supervisor:
Matthias Dümpelmann

PhD Thesis presented to obtain a Ph.D. degree in Informatics Engineering, Intelligent Systems, at the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra

Tese de Doutoramento apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra, para prestação de provas de Doutoramento em Engenharia Informática, Sistemas Inteligentes

August, 2023

Abstract

Seizure prediction aims to anticipate rare events known as epileptic seizures. State of the art methods are based on electroencephalograms (EEGs) and present several challenges. Firstly, long-term acquisition leads to several noisy artefacts, which must be reduced. Due to the large amount of data, this reduction must be automatically performed. Secondly, seizure risk varies over time. Seizure prediction models should be able to adjust themselves to adapt to new data distributions. Lastly, according to the literature, models must be patient-specific, limiting their optimisation, particularly for deep neural networks (DNNs) that require many samples to achieve peak performance. Transfer learning emerges as a solution, enabling the exploration of diverse seizures from multiple patients while maintaining the patient-specific optimisation requirement.

The thesis aims to advance adaptive seizure prediction models using DNNs. Research focused on denoising EEGs automatically, adapting prediction models over time, and using transfer learning to optimise DNN-based models. This thesis comprises three main contributions.

The first contribution is a dataset with denoised EEGs. To create this dataset, an artefact removal algorithm that incorporates automatic and manual parts was built. The automatic part removes experimental errors, while the manual part consists in decomposing the EEGs using independent component analysis (ICA), visually inspecting and removing the non-brain related sources. Data were obtained from patients with epilepsy available in the EPILEPSIAE database. The preprocessed dataset contains 612.88 hours of data with 77,426 independent components (ICs) and is publicly available.

The second contribution involves the development of deep learning-based EEG artefact removal models, comprising an EEG reconstruction model and an IC classifier. The EEG reconstruction model, based on a deep convolutional neural net-

work (DCNN), effectively removed noise from nineteen EEG channels, achieving a root mean squared error (RMSE) of 4.83 μV, a relative root mean squared error (RRMSE) of 0.52, a Pearson correlation coefficient (PCC) of 0.86, and a signal-to-noise ratio (SNR) difference of 8.81. Its low computational time (0.29 seconds to reconstruct 10 minutes of data) makes it possible to use it in quasi-real-time scenarios, e.g., in seizure prediction warning systems. The IC classifier incorporates spectral, spatial, and temporal information of ICs. The dataset prepared in the first contribution and an external dataset, containing ICs obtained from controlled acquisition environments, were used. Applying both datasets, it was found that temporal information is crucial. Exclusively using time series yields comparable results to using both spectral and spatial information. A transfer learning approach further enhanced performance, increasing the geometric mean between sensitivity and specificity from 92.98% to 94.18%. Despite the obtained results, real-time deployment remains challenging due to the computational time required by the ICA.

The third contribution involves tuning seizure prediction models. The impact of denoising EEGs and adapting models over time to possible concept drifts were studied. Models using artificial neural networks (ANNs) with handcrafted features and DNNs using EEG time series as input were developed. Denoising data and retraining models after each new seizure improved performance, though without statistical significance. Comparing model complexity, DNNs predicted half the seizures predicted by the feature-based models while presenting an average false prediction rate per hour (FPR/h) about three times lower. A transfer learning approach was then developed, using a deep convolutional autoencoder trained on EEG data from 41 patients from the EPILEPSIAE database. Its weights were transferred to a DNN used as the basis of the seizure prediction model, and a bidirectional long short-term memory layer and a classifier layer were added. It was personalised for 24 patients from the Epilepsy Center of the Universitätsklinikum Freiburg. Results showed that pre-training the convolutional layers significantly reduced false alarms while maintaining the seizure prediction ability.

The exploration of automatic denoising, temporal retraining, model complexity and transfer learning using patients from another database evidenced that several details still need to be improved before obtaining robust seizure prediction models. Researchers must continue studying to obtain better approaches because the

breakthrough of seizure prediction models can potentially transform the perception of epilepsy within society.

Keywords: Epilepsy, Electroencephalogram, Artefact Removal, Seizure Prediction, Deep Learning

Resumo

A previsão de crises tem como objetivo antecipar eventos raros conhecidos como crises epilépticas. O estado da arte é baseado em eletroencefalogramas (EEGs) e apresenta vários desafios. Primeiramente, a aquisição de longo termo leva a vários artefactos ruidosos, que devem ser reduzidos. Devido ao vasto conjunto de dados, esta redução deve ser automática. Em segundo lugar, o risco de crise varia ao longo do tempo. Os modelos de previsão de crises devem ser capazes de se ajustar a novas distribuições de dados. Por fim, segundo a literatura, os modelos devem ser específicos a cada doente, limitando a otimização dos mesmos, particularmente no caso das redes neuronais de aprendizagem profunda (DNNs), que requerem muitas amostras. Transfer learning surge como uma solução permitindo a exploração de diversas crises de múltiplos doentes mantendo o requisito do treino personalizado.

O objetivo da tese é contribuir para o avanço de modelos adaptativos de previsão de crises usando DNNs. A investigação concentrou-se na redução automática do ruído em EEGs, na adaptação dos modelos de previsão ao longo do tempo e no uso de transfer learning para otimizar os modelos baseados em DNNs. Esta tese compreende três contribuições.

A primeira contribuição é um conjunto de dados com EEGs com ruído reduzido. Para criar estes dados, foi desenvolvido um algoritmo de remoção de artefactos com uma parte automática e uma manual. A parte automática remove erros experimentais, enquanto a parte manual consiste em decompor os EEGs usando a análise independente de componentes (ICA), inspecionar visualmente e remover fontes não relacionadas com o cérebro. Os dados foram obtidos de doentes com epilepsia da base de dados EPILEPSIAE. O conjunto de dados pré-processado contém 612,88 horas com 77.426 componentes independentes (ICs) e está disponível publicamente.

A segunda contribuição envolve o desenvolvimento de modelos de remoção de

artefactos de EEGs baseados em DNNs, compreendendo um modelo de reconstrução de EEG e um classificador de ICs. O modelo de reconstrução de EEG, baseado em redes neuronais convolucionais de aprendizagem profunda, remove efetivamente o ruído de dezanove canais de EEG, alcançando uma raíz do erro médio quadrático (RMSE) de 4,83 μV, uma RMSE relativa de 0,52, um coeficiente de correlação de Pearson de 0,86 e uma diferença da relação sinal-ruído de 8,81. O baixo tempo computacional (0,29 segundos para reconstruir 10 minutos de dados) possibilita o seu uso em cenários quase em tempo real, como em sistemas de alerta de crises. O classificador de ICs incorpora informação espetral, espacial e temporal das ICs. Foram usados os dados preparados na primeira contribuição e um conjunto de dados externo, contendo ICs adquiridos em ambientes controlados. Ao aplicar ambos os conjuntos, observou-se que a informação temporal é decisiva. O uso exclusivo das séries temporais produz resultados comparáveis ao uso de informação espetral e espacial. Uma abordagem de transfer learning melhorou o desempenho, aumentando a média geométrica entre sensibilidade e especificidade de 92,98% para 94,18%. Apesar dos resultados obtidos, a implementação em tempo real continua desafiadora devido à complexidade do ICA.

A terceira contribuição envolve o ajuste de modelos de previsão de crises. Foi estudado o impacto da redução de ruído de EEGs e da adaptação dos modelos ao longo do tempo para possíveis mudanças de conceito. Foram desenvolvidos modelos usando redes neuronais artificiais (ANNs) com features extraídas manualmente e DNNs usando janelas de EEG como input. A redução de ruído e o ajuste dos modelos após cada nova crise melhoraram o desempenho destes, embora sem significância estatística. Ao comparar a complexidade do modelo, as DNNs previram metade das crises previstas pelas ANNs, apresentando um rácio de falsos positivos por hora (FPR/h) médio cerca de três vezes inferior. Em seguida, foi desenvolvida uma abordagem de transfer learning, usando um autoencoder convolucional treinado com dados de EEG de 41 doentes da base de dados EPILEPSIAE. Os pesos aprendidos foram transferidos para uma DNN utilizada como base do modelo de previsão. A esta DNN foram adicionadas uma camada long short-term memory bidirecional e um classificador. Esta DNN foi personalizada para 24 doentes do Centro de Epilepsia de Universitätsklinikum Freiburg. Os resultados mostraram que o pré-treino das camadas convolucionais reduziu significativamente a FPR/h mantendo a capacidade

da previsão de crises.

A exploração da redução automática de ruído, reaprendizagem temporal, complexidade do modelo e transfer learning usando doentes de outra base de dados evidenciou que vários detalhes ainda precisam ser aprimorados antes de obter modelos de previsão de crises robustos. Os investigadores devem continuar a pesquisar para obter melhores abordagens, dado que a inovação dos modelos de previsão pode potencialmente transformar a perceção da epilepsia.

Palavras-chave: Epilepsia, Eletroencefalograma, Remoção de Artefactos, Previsão de Crises Epilépticas, Aprendizagem Profunda

Acknowledgements

I would like to express my gratitude to my supervisors, Professor César Teixeira and Doctor Engineer Matthias Dümpelmann, for their guidance throughout the last four years. I also extend my thanks to Professors António Dourado and Andreas Schulze-Bonhage for their thoughtful revisions and meetings.

I am grateful to the Centre of Informatics and Systems from the University of Coimbra (CISUC) and the Epilepsy Center at the Universitätsklinikum Freiburg for hosting my research over the past four years.

I acknowledge Fundação para a Ciência e Tecnologia (FCT), the Human Capital Operational Program (POCH), and the European Union (EU) for supporting this research work under Ph.D. grant 2020.04537.BD.

Special thanks to Adriana Leal and Mauro Pinto for their friendship, team work, brainstorms, and support during my PhD. I also thank Júlio Medeiros, Diogo Pessoa, Francisco Valente, and Nelson Monteiro for the wonderful moments in the laboratory during my first two years in Coimbra. Likewise, I express my gratitude to Nicolas Zabler, Sebastian Böttcher, and Farrokh Manzouri for the great moments and assistance during the last two years in Freiburg, Germany. I am also grateful to all my other friends, especially Ricardo Pereira, Gonçalo Santos, and the ones from the best group of friends that Coimbra has ever known, for all the amazing moments.

I sincerely thank Daniela for supporting me throughout the last four years. Thank you for understanding the decisions made during this journey, especially during the last two years away from Portugal. Your value to me is beyond words.

Last but not least, I express my heartfelt thanks to my parents, my sister, and her husband for their constant support, especially during the difficult moments. I firmly believe that without your encouragement, it would not have been possible to complete this thesis.

Contents

List of Acronyms

AC alternate current 50

Adam adaptive moment estimation 100, 124, 140, 155

AE autoencoder 58, 98

AED anti-epileptic drug 13, 19, 20, 24

AES American Epilepsy Society 60, 61

AMICA adaptive mixture independent component analysis 53, 54

ANN artificial neural network vi, xxxviii, 4, 5, 37–40, 42, 135, 138, 139, 141–146, 149, 161, 222

ANT-DBS deep brain stimulation of the anterior nucleus of the thalamus 22, 23

AUC area under the receiver operating characteristic curve 30, 84, 147

BiLSTM bidirectional long short-term memory vii, 120, 122, 135, 138, 154

BSS blind source separation 52–56, 58, 85, 117

CV cross-validation 83, 158

DC direct current 2, 18, 50, 89, 134

DCAE deep convolutional autoencoder 58, 98, 151

DCNN deep convolutional neural network vi, 4, 38, 40, 47, 55, 58, 83, 97–100, 102–105, 110–115, 118, 120, 129, 134, 135, 138, 145, 153, 161, 228

DNN deep neural network v–vii, 3–5, 9, 37–40, 44, 55, 58, 70, 81, 82, 84, 115, 117–120, 122–131, 135, 138, 142–146, 149–153, 155, 159, 161, 162, 222, 236

DRE drug-resistant epilepsy 1, 13, 19, 20, 22, 25–28, 46, 59, 88

DRNN deep recurrent neural network 38, 42, 43

DWT discrete wavelet transform 51, 52

ECG electrocardiography 18, 58, 60, 88

ECoG electrocorticogram 14, 26

EEG electroencephalogram v, vi, 1–5, 7, 9, 11–18, 21, 22, 26–29, 32–34, 36, 37, 43, 46, 47, 49, 50, 52, 54, 55, 57, 58, 60–63, 69, 70, 81, 82, 84, 85, 87–89, 92–105, 110, 112–115, 117–119, 130–136, 138, 140–143, 145–149, 152, 154, 158, 159, 161–163, 201, 204, 213, 224, 228, 229

EMD empirical mode decomposition 51, 52, 55, 69

EMG electromyogram 37, 58, 88

EOG electrooculogram 58, 88

ESI electric source imaging 21

EU european union 25

FDA food and drug administration 24, 25

FFT fast Fourier transform 69

FIR finite impulse response 49, 50

fMRI functional magnetic resonance imaging 21

FN false negative 30

FOA focal onset aware 11, 12

FOIA focal onset impaired awareness 11, 12, 28

FP false positive 30

FPR/h false prediction rate per hour vi, 29, 31–33, 47, 84, 85, 141–144, 146, 147, 155–158, 224, 237

GA genetic algorithm 82

GAN generative adversarial network 55, 118, 147

GRU gated recurrent unit 43

IC independent component v, vi, 4, 55, 56, 87–89, 93–96, 98–100, 102, 104, 105, 113, 117–120, 122–125, 127–131, 161, 162

ICA independent component analysis v, 4, 53–55, 69, 89, 91, 93–95, 98, 99, 102, 105, 110, 112–114, 117, 119, 130, 131, 161

iEEG intracranial electroencephalography 14–16, 60, 61

IIR infinite impulse response 49, 50

ILAE International League Against Epilepsy 9–13, 19, 213, 229

IMF intrinsic mode function 52

KCL King's College London 60, 61, 85

LMS least mean squares 50

LR logistic regression 83

LSTM long short-term memory 42, 43, 47, 83, 122, 129, 138

MEG magnetoencephalogram 55, 118

MRI magnetic resonance imaging 21

mRMR minimum redudancy maximum relevance 82

MSI magnetic source imaging 21

MTLE mesial temporal lobe epilepsy 13, 20

NREM non-rapid eye movement 37, 213, 229

OUH Odense University Hospital 60, 61, 85

PCA principal component analysis 54

PCC Pearson correlation coefficient vi, 82, 98, 101, 102, 105, 110–112, 161, 202

PDF probability density function 54

PET positron emission tomography 21

PSD power spectrum density 4, 51, 55, 88, 94–96, 104, 117, 118, 120, 123, 126–129, 131, 162

PSO particle swarm optimisation 82

ReLU rectified linear unit 39, 99, 100, 120, 122

REM rapid eye movement 36, 37, 213, 229

REST reference electrode standardisation technique 14

RMS root mean square 101, 110

RMSE root mean squared error vi, 98, 100–102, 105, 110–112, 161, 202

RNS responsive neurostimulation system 22, 23

RRMSE relative root mean squared error vi, 98, 100–102, 105, 110–112, 161, 202

SNR signal-to-noise ratio vi, 16, 69, 98, 101, 102, 105, 110–112, 161, 202

SOBI second-order blind identification 53, 54

SOP seizure occurrence period 29–32, 35, 46, 62, 69, 70, 137, 140, 141, 147–149, 153, 159

SPECT single-photon emission computed tomography 21

SPH seizure prediction horizon 29–31, 46, 62, 69, 70, 137, 141, 147, 149, 153

SS seizure sensitivity 29, 31, 32, 35, 47, 84, 141–144, 146–148, 155–158, 162, 224, 237

ssEEG subscalp electroencephalogram 14, 16, 26, 61

STFT short-term Fourier transform 228

SUDEP Sudden Unexpected Death in Epilepsy 11

SVMs support vector machines 83, 146

TiW time in warning 85

TLE temporal lobe epilepsy 13, 22, 46, 61, 135, 152

TN true negative 30

TP true positive 30

VNS vagus nerve stimulation 22, 23

ZUH Zealand University Hospital 60, 61, 85

List of Figures

List of Tables

Chapter 1

Introduction

Seizure prediction is a research area related to epilepsy disease, which emerged about 50 years ago. Since then, several approaches, mostly based on the electroencephalogram (EEG), have been proposed by different research groups. However, even after this time, there is still no effective solution to predict any type of seizure. Different factors affect their performance, such as the amount of data available, the quality of the EEG or the ability of the algorithms to be sufficiently dynamic to adapt to different data distributions over time. This chapter introduces the main motivations of seizure prediction and the limitations of developing approaches using EEGs. In addition, the main goals and contributions are also presented.

1.1 Motivation

Epilepsy is one of the most common neurological disorders, affecting people of different ages. Approximately 1% of the population in the world have epilepsy which represents about 70 million people. Furthermore, nearly 80% of this population live in low or middle-income countries, meaning they do not have easy access to adequate treatment [3, 4].

Although epilepsy can be controlled with antiepileptic drugs or other treatments, only about two-thirds of the patients respond satisfactorily to this procedure. The other one-third are considered drug-resistant epilepsy (DRE) patients [5–7]. DRE patients live a seriously limited life because they are not allowed to do daily activities such as driving a car and often have restrictions in their professional life. Moreover, beyond the physical injuries (brain and body lesions) experienced by these patients and the increased risk of sudden unexpected death, they may suffer from discrimination and tend to develop educational problems [8, 9]. These consequences could lead to psychological diseases as depression or even suicide [10].

The unpredictability of the seizure onset is seen by the patients as the main disease burden [11]. Seizure prediction approaches emerge as a solution to mitigate it. These could be achieved by real-time EEG analysis and triggering an alarm if

an upcoming seizure is predicted. One of the outcomes is that these models could be integrated into closed-loop systems that would automatically be able to perform neuromodulation to suppress seizures or used by the patients to decide when to take rescue medication to cancel seizures [12–15]. Therefore, a fully automated seizure prediction model able to acquire, process, and predict upcoming seizures without human intervention would allow them to live with fewer restrictions.

1.1.1 Gaps/Needs

Several different seizure prediction approaches based on EEG data have been published. However, some gaps still require further effort, such as accurately handling the noisy artefacts present in EEGs and the limitations on developing seizure predictors.

1.1.1.1 Noisy artefacts present in long-term non-invasive EEG signals

Researchers are moving towards minimally invasive ultra-long-term EEG recordings acquired daily to develop seizure prediction approaches [16]. As epileptic seizures are rare, it is necessary to acquire long-term EEG to capture epileptic brain signal [17,18]. Therefore, it is impossible to make a controlled acquisition as the patients must maintain their daily activities, such as eating, talking, moving, and sleeping, while being monitored. These activities usually produce several noisy artefacts. EEG artefacts could be responsible for increasing false alarm rate and, therefore, should be removed before creating the seizure prediction models [19–21]. Generally, researchers remove EEG data considerably contaminated with noisy artefacts [22]. However, discarding data may lead to a high loss of information and, therefore, should be avoided whenever possible. Simple digital filtering is often used to denoise EEGs, e.g., high-pass filters to remove the direct current (DC) component, notch filters to remove powerline interference, and low-pass filters to diminish the influence of high-frequency noise. Although being a simple and fast approach to handling noise, simple digital filtering suppresses whole frequency bands, not being capable of removing physiological artefacts such as eye and muscle artefacts, which usually present spectral bands overlapping important brain information [23, 24]. Consequently, other techniques, such as decomposition algorithms [25–29], have been used to reduce the influence of noisy artefacts on seizure prediction. Nevertheless, these require visual inspection or are too slow and may not be used in real-time scenarios. Therefore, developing automatic and fast EEG artefact removal methods may bring enormous benefits to the seizure prediction field.

1.1.1.2 Limitations on seizure prediction approaches

Patients with epilepsy usually present a great heterogeneity of epilepsy and seizure types. This heterogeneity makes it almost impossible to create a general solution to

predict any upcoming seizure. Therefore, seizure prediction models are commonly developed following a patient-specific approach [20]. Seizures are events that usually do not happen very often. Therefore, the acquisition of long-term EEGs is necessary not only to capture seizures but also to study what happens before the onsets [14, 30]. However, even with the acquisition of long-term EEG signals, the number of seizures per patient is often quite limited, and, therefore, the need to develop patient-dependent models may reduce the prediction efficacy [18].

Seizure prediction models based on shallow classifiers are built using handcrafted features extracted using signal processing methods. This feature engineering could lead to a loss of information if the proper features are not used. Deep learning architectures have increasingly been used because they can extract meaningful features directly from the data [31, 32]. These architectures can automatically extract features and classify samples without human intervention. Therefore, these are of great value because they may be easily adapted to handle possible concept drifts that may appear over extensive data periods [33, 34]. Also, deep neural networks (DNNs) can easily perform transfer learning. It means deep learning architectures may be trained with data obtained from several patients and perform fine-tuning whenever one wants to develop a seizure prediction model for a new patient. Therefore, it could solve the problem of the limited number of existing seizures per patient to train effective patient-specific models [35].

1.2 Goals and contributions

In this thesis, it was hypothesised that, by using deep learning-based techniques, it would be possible to develop adaptive seizure prediction solutions capable of automatically predicting upcoming seizures. To this end, investigations were made on important steps of the seizure prediction pipeline, such as removing noisy EEG artefacts and training classifiers for seizure prediction. The main dataset used in the development of this thesis was obtained from the European Epilepsy Database (EPILEPSIAE) [13]. In addition, long-term EEGs acquired from patients with epilepsy at the Epilepsy Center at the Department of Neurosurgery, University Medical Center Freiburg, Germany, were also used. It is worth noting that both datasets were obtained under pre-surgical conditions. Additionally, short-term EEGs acquired from software engineers during code comprehension tasks on behalf of the Biofeedback Augmented Software Engineering (BASE) project [36] were used for one of the EEG preprocessing tasks. The investigation carried out in this thesis was divided into three main contributions described in the following sections.

1.2.1 A curated EEG dataset for artefact handling

The first part of this thesis describes the whole process of EEGs curation. In this study, the removal of artefacts from the EEGs of patients with epilepsy was performed using several techniques such as frequency filtering, removal of experimental errors and removal of physiological artefacts through visual inspection of the various independent sources that originate the multi-channel EEGs. It contributed to the availability of a dataset containing EEGs before and after removing artefacts, thus allowing the development of automatic EEG preprocessing techniques. The study is openly available in [37].

1.2.2 EEG artefact removal using deep learning approaches

The second contribution of this document concerns the development of EEG artefact removal using deep learning approaches. The EEGs used in the various studies presented in this thesis were acquired over several days in a pre-surgical situation, where the patients were lying in bed but performing daily routine activities such as eating, reading, or chatting with neighbours or visitors. Thus, the visual inspection of these data would be very time-consuming. Furthermore, suppose the artefact removal is not automatic. In that case, it can not be used in a real-time scenario where the patient would use a seizure prediction device based on EEGs without human intervention. Therefore, using the dataset built in the first contribution, two approaches were developed to automatically simulate the decisions made by the experts when visually inspecting the EEGs. One approach is based on deep convolutional neural networks (DCNNs) that take as input the noisy EEGs. The other approach consists of an ensemble of three DNNs, each handling different aspects of the independent component (IC) data, such as time series, power spectrum density (PSD), and topographic map. The first approach can remove artefacts from several minutes of EEG in less than a second. Thus, it allows quick data preparation for seizure prediction tasks or may be integrated into a real-time seizure prediction device. The second approach lasts several minutes because of the independent component analysis (ICA) complexity. Nevertheless, it is more interpretable than the EEG reconstruction approach and may assist researchers while classifying the ICs. The studies are openly available in [38] and [39], respectively.

1.2.3 Tuning seizure prediction models based on deep neural networks

The third contribution of this thesis comprises the tuning of seizure prediction models based on DNNs. Two different studies were performed. The first one assessed the impact of removing EEG artefacts and adapting seizure prediction models over time to possible concept drifts. Two types of neural networks were used, one based on artificial neural networks (ANNs) with handcrafted features and one based on DNNs

with EEG segments as input. The first comparison evaluated the difference between using EEGs, denoised applying the model developed in the second study, with EEGs that were not denoised. Additionally, the performance of the models trained only once and tested in the following seizures was compared with models which were retrained every new seizure. Finally, a comparison was also made between the two types of used ANNs to verify whether there were differences between models developed with knowledge-based features and models capable of extracting their features automatically. The second study concerns developing a deep convolutional autoencoder (DCAE), which may be used to improve the performance of patient-specific approaches. The model was trained using EEGs acquired from forty-one patients. Afterwards, the obtained weights were used to produce patient-specific seizure prediction models by applying a transfer learning approach. Transfer learning makes it possible to use information from other patients and reduce training computation time. Finally, the model was shared with the scientific community to allow other researchers to use it on novel seizure prediction methods. The first study is openly available in [40].

1.3 Scientific outcomes

During this thesis, several contributions to the seizure prediction field were made. These include publications as main author or co-author in international peer-reviewed journals, presentations at national and international conferences, and co-supervision of master's degrees thesis. The ones containing **Lopes, F.** as first author resulted from this document. The contributions are listed in the next sections.

1.3.1 Peer-reviewed journal articles

J1 **Lopes, F.**, Leal, A., Medeiros, J., Pinto, M. F., Dourado, A., Dümplemann, M., and Teixeira, C. A. Automatic Electroencephalogram Artifact Removal using Deep Convolutional Neural Networks. *IEEE Access*, 2021, 9, 149955-149970. **(Impact factor = 3.476; Scimago Quartile = Q1 in "Engineering (miscellaneous)")**

J2 **Lopes, F.**, Leal, A., Pinto, M. F., Dourado, A., Dümplemann, M., and Teixeira, C. A. Ensemble Deep Neural Network for Automatic Classification of EEG Independent Components. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022, 30, 559-568. **(Impact factor = 4.900; Scimago Quartile = Q1 in "Biomedical Engineering")**

J3 **Lopes, F.**, Leal, A., Medeiros, J., Pinto, M. F., Dourado, A., Dümplemann, M., and Teixeira, C. A. EPIC: Annotated epileptic EEG independent components for artifact reduction. *Scientific Data*, 2022, 9, 512. **(Impact factor = 8.501; Scimago Quartile = Q1 in "Computer Science Applications")**

J4 **Lopes, F.**, Leal, A., Pinto, M. F., Dourado, A., Dümplemann, M., Schulze-Bonhage, A., and Teixeira, C. A. Removing artefacts and periodically retraining improve performance of neural network-based seizure prediction models. *Scientific Reports*, 2023, 13, 5918. **(Impact factor = 4.996; Scimago Quartile = Q1 in "Multidisciplinary')**

J5 Leal, A., Pinto, M. F.,**Lopes, F.**, Bianchi, A. M., Henriques, J., Ruano, M. G., Carvalho, P., Dourado, A., and Teixeira, C. A. Heart rate variability analysis for the identification of the preictal interval in patients with drug-resistant epilepsy. *Scientific Reports*, 2021, 11, 5987. **(Impact factor = 4.996; Scimago Quartile = Q1 in "Multidisciplinary')**

J6 Pinto, M. F., Leal, A., **Lopes, F.**, Dourado, A., Martins, P., and Teixeira, C. A. A personalized and evolutionary algorithm for interpretable EEG epilepsy seizure prediction. *Scientific Reports*, 2021, 11, 3415. **(Impact factor = 4.996; Scimago Quartile = Q1 in "Multidisciplinary')**

J7 Pinto, M. F., Leal, A., **Lopes, F.**, Pais, J., Dourado, A., Sales, F., Martins, P., and Teixeira, C. A. Interpretable EEG seizure prediction using a multiobjective evolutionary algorithm. *Scientific Reports*, 2022, 12, 4420. **(Impact factor = 4.996; Scimago Quartile = Q1 in "Multidisciplinary')**

J8 Pinto, M. F., Leal, A., **Lopes, F.**, Pais, J., Dourado, A., Sales, F., Martins, P., and Teixeira, C. A. On the clinical acceptance of black-box systems for EEG seizure prediction. *Epilepsia Open*, 2022, 00, 1-13. **(Impact factor = 4.026; Scimago Quartile = Q2 in "Neurology')**

J9 Pinto, M. F., Batista, J., Leal, A., **Lopes, F.**, Oliveira, A., Dourado, A., Sales, F., Martins, P., and Teixeira, C. A. The goal of explaining black boxes in EEG seizure prediction is not to explain models' decisions. *Epilepsia Open*, 2023, 8, 285-297. **(Impact factor = 4.026; Scimago Quartile = Q2 in "Neurology')**

J10 Leal, A., Curty, J., **Lopes, F.**, Pinto, M. F., Oliveira, A., Sales, F., Bianchi, A. M., Ruano, M. G., Dourado, A., Henriques, J., and Teixeira, C. A. Unsupervised EEG preictal interval identification in patients with drug-resistant epilepsy. *Scientific Reports*, 2023, 13, 784. **(Impact factor = 4.996; Scimago Quartile = Q1 in "Multidisciplinary')**

1.3.2 Other scientific publications

O1 *Presentation in national conference:* **Lopes, F.**, Leal, A., Pinto, M. F., Dourado, A., Dümpelmann, M., and Teixeira, C. A. "Long-term EEG Artifact Removal in Real-time", *33º Encontro Nacional de Epileptologia (ENE)*

– *CONGRESSO VIRTUAL DA Liga Portuguesa Contra a Epilepsia (LPCE)*, 2021

O2 *Poster presentation in international conference:* **Lopes, F.**, Leal, A., Pinto, M. F., Dourado, A., Schulze-Bonhage, A., Dümpelmann, M., and Teixeira, C. A. "Evaluating the Influence of Retraining Seizure Prediction Models over Time", *International Conference for Technology and Analysis of Seizures (ICTALS2022)*, 2022

O3 *Poster in international conference:* Leal, A., Pinto, M. F., **Lopes, F.**, Curty, J., Oliveira, A., Sales, F., Ruano, M. G., Dourado, A., Bianchi, A. M., Henriques, J., and Teixeira, C. A. "Can unsupervised preictal labelling improve seizure prediction?", *International Conference for Technology and Analysis of Seizures (ICTALS2022)*, 2022

O4 *Poster in international conference:* Pinto, M. F., Leal, A., **Lopes, F.**, Dourado, A., Martins, P., and Teixeira, C. A. "Can we explain how Machine Learning Models predict seizures? Towards an appropriate explainability of EEG seizure prediction models", *International Conference for Technology and Analysis of Seizures (ICTALS2022)*, 2022

O5 *Poster in international conference:* **Lopes, F.**, Leal, A., Pinto, M. F., Dourado, A., Schulze-Bonhage, A., Dümpelmann, M., and Teixeira, C. A. "Transfer learning on seizure prediction: Does information from several patients improve patient-specific approaches?", *4th International Congress on Mobile Health and Digital Technology in Epilepsy (MHDTE2023)*, 2023

O6 *Poster in international conference:* Zabler, N., Böttcher S., Manzouri, F., **Lopes, F.**, Epitashvili, N., Van Paesschen, W., Schulze-Bonhage, A., and Dümpelmann, M. "ECG artefact detection in low channel EEG systems for robust multimodal seizure detection", *4th International Congress on Mobile Health and Digital Technology in Epilepsy (MHDTE2023)*, 2023

1.3.3 Master's degree co-supervision

M1 Paiva, V. "EEG Pre-processing using Deep Learning Networks: An Unichannel Approach", *Master thesis dissertation, Faculty of Science and Technology of the University of Coimbra* (2022).

M2 Curty, J. "Unsupervised Preictal Estimation Based on Univariate Features and Dimensionality Reduction", *Master thesis dissertation, Faculty of Science and Technology of the University of Coimbra* (2023).

1.4 Thesis outline/structure

The remainder of this thesis is structured as follows.

Chapter 2 provides background information related to epilepsy, EEG and artefacts, seizure prediction, DNNs, and possible transfer learning approaches.

Chapter 3 presents a literature overview on EEG artefact removal and EEG seizure prediction.

Chapter 4 refers to the manual artefact removal performed to prepare the dataset to develop automatic artefact removal approaches.

Chapter 5 presents the development of an automatic EEG artefact reduction method.

Chapter 6 includes the process of producing an EEG independent component classifier.

Chapter 7 analyses the impact of EEG artefact reduction and periodic retraining on seizure prediction approaches.

Chapter 8 presents the effect of using a transfer learning approach based on the parameters of a DCAE trained using data from a database on improving the performance of new patient-specific seizure prediction models.

Chapter 9 concludes this thesis by summarising the main findings highlighting their added value. Additionally, it provides directions for future work.

Chapter 2

Background concepts

This chapter introduces the main concepts required to understand this thesis. Section 2.1 includes a brief description of the epilepsy disease and the classification system used for categorising the different types of epilepsy. Then, electroencephalograms (EEGs) and its artefacts are explained in Section 2.2. Afterwards, a short explanation about the available epilepsy treatments is introduced in Section 2.3. After that, the concepts needed to develop seizure prediction models are covered in Section 2.4. Concept drifts are introduced in Section 2.5. Section 2.6 explains deep neural networks (DNNs) that are going to be used in this thesis. Afterwards, transfer learning is covered in Section 2.7. Finally, a summary with the key concepts of this chapter is presented in Section 2.8.

2.1 Epilepsy

Epilepsy is a chronic neurological disease characterised by brief and recurrent episodes known as seizures [41, 42], which affects 1% of the world population. Epileptic seizures may be characterised, for example, by involving involuntary movements of the human body, losing consciousness or loss of awareness. These episodes could also include loss of bowel and bladder functions. These events result from abnormal synchronisation of neuronal activity of the brain [43], which leads to an intense hypersynchronous state with increased neuron excitability, i.e., the neurons become excited at the same time in different areas of the brain leading to a high release of energy [44, 45].

2.1.1 Definition of epilepsy and seizures

In 2005, International League Against Epilepsy (ILAE) described epilepsy disease and epileptic seizures [43]. Definitions are presented in Boxes 1 and 2, respectively.

Later, in 2014, ILAE proposed an operational clinical definition of epilepsy disease to be applied in clinical diagnosis [42], presented in Box 3. It was suggested to

Box 1 - Conceptual definition of epilepsy (as defined by the Task Force of the ILAE in 2005 [43])

"Epilepsy is a disorder of the brain characterised by an enduring predisposition to generate epileptic seizures, and by the neurobiologic, cognitive, psychological, and social consequences of this condition. The definition of epilepsy requires the occurrence of at least one epileptic seizure."

Box 2 - Conceptual definition of seizure (as defined by the Task Force of the ILAE in 2005 [43])

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

Box 3 - Diagnostic criteria for epilepsy (as defined by the Task Force of the ILAE in 2014 [42])

"Epilepsy is a disease of the brain defined by any of the following conditions:

1. *At least two unprovoked (or reflex) seizures occurring >24 h apart.*
2. *One unprovoked (or reflex) seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after two unprovoked seizures, occurring over the next 10 years.*
3. *Diagnosis of an epilepsy syndrome."*

emphasise the risk of experiencing new seizures after an unprovoked seizure. It may lead to an earlier diagnosis and, therefore, to an earlier treatment.

The term "provoked seizure" refers to seizures caused by some provocative factor. An example of a provoked seizure is one that occurs after a concussion. Around 10% of the global population suffer at least one seizure during their lifetime without an epilepsy diagnosis [3]. It is worth noting that epilepsy may not last for the whole patient's lifetime as it is considered to be resolved for the ones who either had an age-dependent epilepsy syndrome but are not that age anymore or by the ones who did not have any seizures during the last 10 years and did not take any antiseizure medication over the last 5 years. Although the patient does not experience a seizure or has a high risk of having a seizure over several years without taking any medicine, no one can ensure that the patient will not evidence a future seizure event [6].

2.1.2 Classification of seizures and epilepsies

In 2017, ILAE updated the classification of epilepsy types [46] and also the operational classification of seizure types [47]. According to the new classification, there are three levels for classifying epilepsy disease: seizure type, epilepsy type and epilepsy syndrome. The classification may also include the aetiology of each level, such as structural, genetic, infectious, metabolic, and immune; and comorbidities.

Figure 2.1 summarises this new classification.

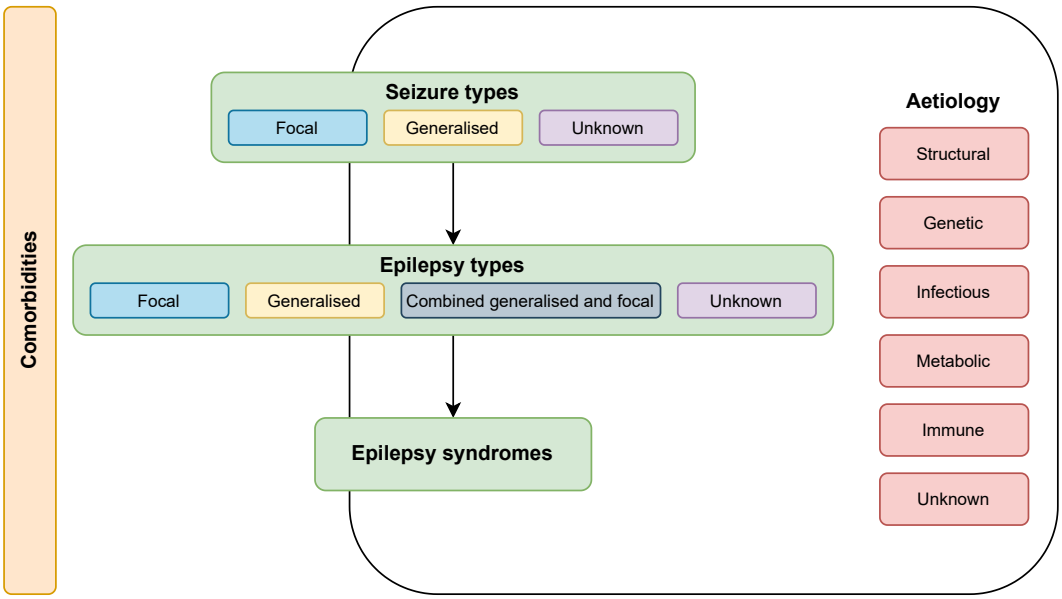


Figure 2.1: Summary of epilepsy categorisation following ILAE classification. Adapted from [46].

2.1.2.1 Seizure type

Seizure type is the first level for classifying epilepsy. After identifying the seizure as epileptic, the clinician has to define its type. A seizure may be focal, generalised or unknown, depending on which part of the brain it is generated. To perform this classification, the clinician must analyse seizures using video-EEG recordings [6, 46]. Figure 2.2 summarises seizure type classification.

Focal seizures can be classified according to the awareness during the event, i.e., whether the patient is aware of self and the environment during the seizure. In the case of awareness, the seizure is classified as focal onset aware (FOA), otherwise focal onset impaired awareness (FOIA). Seizure onset may also be defined as motor or nonmotor onset, depending on the symptoms experienced by the patients. Moreover, there is a special type of focal seizure named "focal to bilateral tonic-clonic". This classification occurs when the seizure onset is limited to one hemisphere but quickly propagates to the other. Tonic (tense body) and clonic (jerking movements) are symptoms usually evidenced during these seizures [6, 47].

Generalised seizures are generated within the neural connections of both hemispheres. The most feared type of generalised seizure is the tonic-clonic because it involves the entire body; generally, patients lose their conscience. Furthermore, Sudden Unexpected Death in Epilepsy (SUDEP) is also frequent in patients with this type of seizure. Generalised tonic-clonic seizures are also known as *grand mal* seizures [48].

The seizure is classified as unknown whenever the clinicians can classify it as

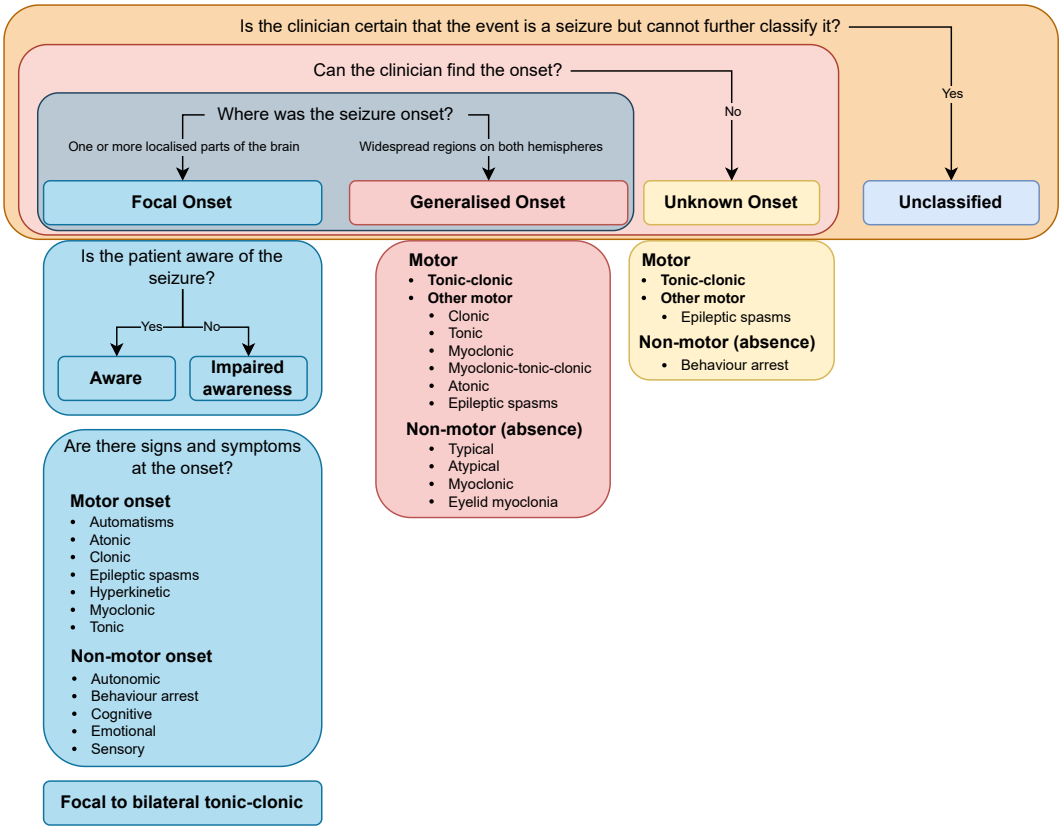


Figure 2.2: Summary of seizure categorisation following expanded version of ILAE seizure type classification. Adapted from [47].

motor or non-motor but are not certain of the onset. Unclassified seizures are those that can not be classified due to a lack of information or not being correctly placed in one of the other cases [47].

This level may be the final classification of epilepsy when the clinician has no access to video-EEG and neuroimaging data [46].

2.1.2.2 Epilepsy type

Epilepsy type is the second level of the definition of epilepsy. This level is only considered when the patient is already diagnosed with epilepsy based on the 2014 ILAE definition [42]. This level comprises three classes: focal epilepsy, generalised epilepsy and combined generalised and focal epilepsy. Whenever the clinician has insufficient information for classifying the epilepsy type, it is labelled as unknown. Also, when the seizure types are unknown, epilepsy is usually classified as unknown.

Focal epilepsy includes both unifocal and multifocal disorders, as well as focal seizures. Patients with focal epilepsy usually experience FOA and FOIA seizures, focal motor seizures, focal non-motor seizures and focal to bilateral tonic-clonic seizures. Additionally, the interictal EEG usually exhibits focal epileptiform discharges.

Patients with generalised epilepsy evidence spike-wave activity on EEG and the most common seizure types are absence, myoclonic, tonic, atonic and tonic-clonic

seizures.

Combined generalised and focal epilepsy is a new type of epilepsy introduced in 2014 [42]. Patients with this type of epilepsy experience both focal and generalised epilepsy. The interictal EEG can show generalised spike-wave activity and/or focal epileptiform discharges.

Considering focal epilepsy types, temporal lobe epilepsy (TLE) is the most common epilepsy type diagnosed in adults. TLE covers about 60% of all epileptic patients with focal epilepsy. It is usually diagnosed in childhood and teenage years. TLE can be divided into mesial, neocortical, or lateral. Mesial temporal lobe epilepsy (MTLE) is characterised by the involvement of the medial or internal parts of the temporal lobe. About 80% of all temporal lobe seizures are MTLE. Patients with MTLE usually do not become seizure-free just by using anti-epileptic drugs (AEDs), i.e., seizure medicines just lower the number of events. Therefore, these patients are frequently submitted to resective surgery [49].

2.1.2.3 Epilepsy syndromes

The final level of epilepsy classification is the diagnosis of the epilepsy syndrome. Epilepsy could be defined with characteristics that usually occur together, e.g., seizure types, EEG and imaging findings, seizure triggers and vigilance states. It is worth noting that although some syndromes are known, ILAE does not have a formal classification for epilepsy syndromes [46].

People with epilepsy syndromes usually do not become seizure-free. However, in the case of early detection and control using antiseizure medication or if the person has had surgery to remove the seizure focus, it is possible to decrease or even stop the number of seizures events. Therefore, early identification of epilepsy syndrome is crucial for its treatment [50].

2.1.3 Seizure clusters

Although there is no clear definition of ILAE, seizure clusters are phenomena that derive from the occurrence of several seizures over a certain period which can be minutes or even hours. One of the most adopted definition for seizure cluster is when a patient with drug-resistant epilepsy (DRE) suffers at least 3 seizures within a 24-hour interval. Seizure clusters should be treated with seizure-suppression medication in order to reduce the threat of prolonged seizures and to prevent them from evolving into status epilepticus [51–53].

2.2 EEG

EEG measures the electrical activity of the brain, i.e., it records the summed energy activity resulting from several synchronised excitatory and inhibitory synapses

within brain cells which have similar spatial orientation [54,55]. It is a nonlinear and nonstationary signal, making it truly complex to analyse [17,24]. It is used for presurgical evaluation, epilepsy diagnosis, and long-term continuous monitoring [56, 57]. EEG can be acquired using non-invasive (scalp EEG) or invasive acquisition electrodes (intracranial electroencephalography (iEEG) or electrocorticogram (ECoG) and subscalp electroencephalogram (ssEEG)).

2.2.1 Scalp EEG

Scalp EEG or simply EEG is the most used method for acquiring neural activity because it is not invasive and can be more easily accommodated by patients, given the current technological advances, to monitor their neurological condition even during their day-to-day routines [58]. Furthermore, it captures information from the entire scalp being a useful signal for studying neural mechanisms that propagate in different brain areas.

Scalp EEG comprises mainly three types of spatial setups: the International 10-20 System with at most 21 electrodes; the International 10-10 System, which usually contains a number of electrodes between 64 and 85 and the high-density International 10-5 System that is able to use up to over 300 electrodes [59]. The numbers associated with the acquisition system represent how the electrodes are placed on the skull, e.g., the electrodes of the International 10-20 System are placed at 10%, 20%, 20%, 20%, 20% and 10% from inion (Iz) to nasion (Nz) over its length. Figure 2.3 shows the location of the electrodes used by each international montage.

For measuring EEGs it is necessary to have a reference location. It is used to obtain the EEG potentials at each electrode location. Thus, the optimal reference would be a location with no potential, i.e., a zero reference, to not produce any distortion effect in the desired measures. Considering when the reference is applied, there are two types of references: the online references also known as physical reference electrodes such as the left earlobe or any central electrode between Fz and Oz, and the offline references such as linked mastoids, average reference and reference electrode standardisation technique (REST) [61–63]. The latter ones are usually used for re-referencing the EEGs, a reconstruction procedure, performed after EEG acquisition, used to convert a non-zero reference to an approximate zero reference. Although average reference and REST present the smallest reconstruction error, they require a large number of recording electrodes to have a better scalp coverage. Therefore, they are more suitable for high-density EEG architectures.

2.2.2 Invasive EEG

Brain electrical activity may be also recorded invasively using (i) intracranial electrodes that acquire data directly from the brain or (ii) subscalp electrodes which are placed between the scalp and the bone.

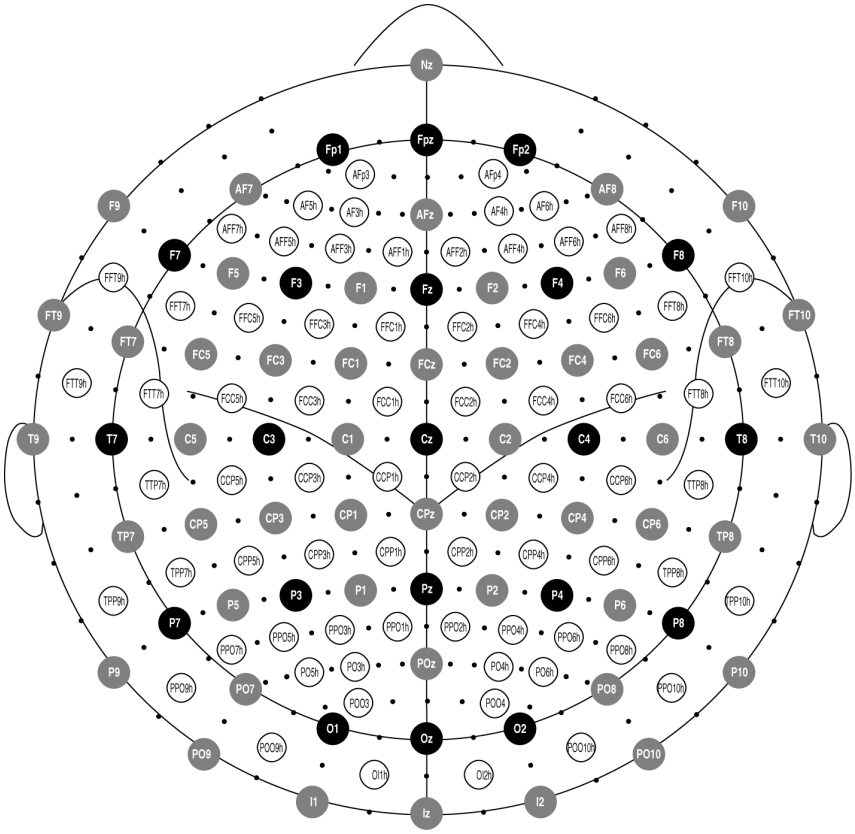


Figure 2.3: Representation of the three scalp EEG architectures. The International 10-5 System is represented by all the electrodes. The International 10-10 System is constituted by the grey and black locations. Finally, the International 10-20 System uses the electrodes represented by the black dots. Extracted from [60].

In iEEG or ECoG, the electrodes are placed directly on the brain surface or in deep brain structures accessed during surgery, to measure the activity from the brain cortex [64]. It is usually used as part of the presurgical evaluation in patients, when non-invasive studies remained non-concordant or inconclusive regarding the irritative area, the seizure onset zone or the eloquent cortex [65]. Compared to scalp EEG, this method does not follow an international standardised spatial setup. However, as presented in Figure 2.4, there are two different methods for acquiring these signals: subdural or depth electrodes. Subdural electrodes are placed on the surface of the cortex. Depending on the area the clinicians would like to evaluate, they consist of a series of electrodes configured as linear strips or square and rectangular grids. Depth electrodes are intended to capture brain signal from deep brain areas. They are implanted inside those areas such as amygdala, hippocampus, orbitofrontal, and medial occipital regions. These electrodes are thin and flexible, embedded in silastic material, with contact points located on their surface. Depth electrodes are usually placed through a stereotactic surgery, which uses a three-dimensional coordinate system to identify a small target inside the skull. Therefore, this aquisition method

is usually named stereoelectroencephalography [66].

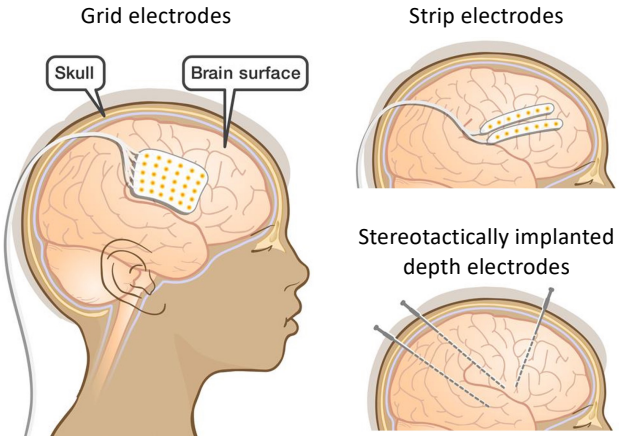


Figure 2.4: Different architectures for acquiring iEEG. Adapted from [67]

The iEEG provides better signal-to-noise ratio (SNR), and the presence of artefacts is practically null because the number of layers between the signal source and the sensors is much lower than in scalp EEG. These layers also diminish the magnitude of the brain waves [68]. Therefore, iEEG may be better than scalp EEG for analysing neural biomarkers. Furthermore, it can acquire signal directly from the seizure-source region. However, due to its invasive nature, acquiring iEEG has risks such as intracranial hematomas, haemorrhage, infection or even death [69].

In recent years, researchers have presented EEG acquisition systems based on electrodes placed under the scalp but outside the skull, i.e., ssEEG [70, 71]. Their implantation is minimally invasive and carries less risk for the patient than iEEG. ssEEG presents spectral characteristics similar to scalp EEG. However, as subscalp electrodes are placed under the scalp, they avoid skin abrasion, reduce electrode contact problems, and require less electrode maintenance. Therefore, they can be used to acquire EEG over several months, also known as ultra-long-term EEG [57, 72, 73]. These acquisition devices may revolutionise the diagnosis and treatment of patients with epilepsy because they allow scientists to investigate the occurrence of seizures in day-to-day scenarios. Moreover, capturing a higher number of seizures would increase the chances of developing improved seizure prediction approaches. However, the number of electrodes used in these is quite limited, leading to a low spatial resolution. Figure 2.5 shows different devices to acquire ssEEG.

2.2.3 Neural activity and artefacts

Neural signals are mainly composed of two types of potentials: the rhythmic ones known as oscillations and the transients. The excitability of brain cells generates rhythmic fluctuations. They are usually divided in frequency bands such as delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz) and gamma (above 30 Hz) [68, 74, 75]. Transient potentials can be divided in normal or abnormal type.

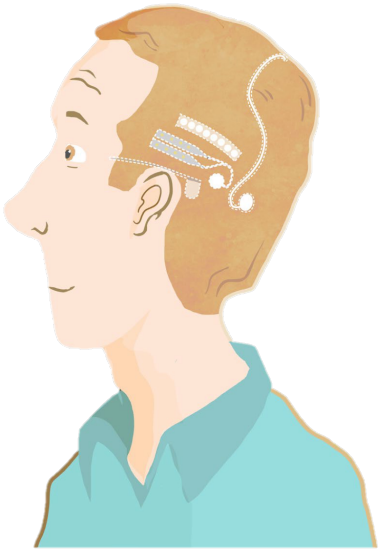


Figure 2.5: Electrodes to acquire ssEEG.

The first ones are related to sleep potentials while the abnormal ones may represent seizure or interictal activity in epileptic patients [24, 75].

As neural signals propagate throughout the scalp, they require several electrodes to be acquired with high spatial resolution [59]. However, the large number of electrodes leads to a high probability of measuring variations that do not come from the neural connections.

EEG artefacts can be reduced or even avoided when signal acquisition is performed under controlled procedures. However, in epilepsy, seizures are usually very unpredictable [76, 77] taking some minutes, hours, days, or even months to occur. Therefore long-term recordings are required. These recordings are acquired throughout several continuous days and, consequently, it is impossible to not measure noisy data such as the environmental artefacts, experimental errors and physiological artefacts [23].

Environmental artefacts comprehend noisy data related to the acquisition system. They could be generated by interferences such as electrical, magnetic, sound or even the electromagnetic waves such as the ones generated by the communication devices [24]. Furthermore, these artefacts include the instrumentation disturbances such as pink noise (small variations of the condensed-matter materials) or poor electrical grounding. These type of artefacts generally does not cause any trouble during the EEG analysis because they usually present fixed frequencies, e.g., the powerline noise, generated by the alternated current power, has a fixed frequency of 50 Hz or 60 Hz depending on the country, or they have a spectrum that do not overlap the desired signal frequencies [23].

Experimental errors are usually related to poor electrode adhesion, incorrect scalp cleansing and subject motion related to daily life situations. These factors can change the position of those electrodes and produce high electrical impedance

causing signal distortion. This kind of artefacts are usually hard to remove even with artefact removal algorithms because they typically do not follow any pattern, overlaps the frequencies of interest and occasionally distorts the signal eliminating any neural information in the signal. Therefore, they definitely should be avoided by correctly preparing the recording system scalp before placing the electrodes [23, 78].

Physiological artefacts are alterations on the desired signals generated from physiological processes such as eye blinks, eye saccades, muscle activity, e.g. chewing and swallowing, and cardiac activity also known as pulse artefacts. Furthermore, there are other less common artefacts such as breathing and perspiration.

Eye movements generate electrical activity which are strong enough to be measured by the EEG electrodes. Although these alterations cause a larger variation in the frontal electrodes, they propagate throughout the scalp with a magnitude higher than the EEGs amplitude. Although they cause a larger variation in the frontal electrodes these type of activity propagate throughout the scalp. Furthermore, these artefacts generally present a spectrum that overlaps brain activity frequencies useful for the study [79, 80]. Thus, they can not be removed using simple digital filtering without losing useful neural information.

Muscle activity signals can be generated in any part of the human body including the head. Thus, they can be produced very closely to any electrode. This activity is mostly a consequence of chewing, swallowing, talking or even scalp contraction, i.e., actions related to each subject daily life. These signals are generally difficult to remove because of their transient behaviour and wide spectral distribution overlapping all the frequencies from 0 to >200 Hz, specially the beta waves [24, 78, 81].

Among the three major types of physiological artefacts, cardiac activity is the most regular, having a well-known characteristic and repetitive pattern. Normal pulse artefacts are nonstationary. Therefore, although they generally have a frequency around 1.2 Hz [23, 82], this frequency it not fixed, changing according to the subject activities making it difficult to remove from the EEGs. However, its amplitude depends on the location of the electrodes, i.e., if they are placed over a scalp artery its amplitude would be high, otherwise it will be practically unnoticeable [83]. Thus, this type of activity can be reduced by a proper sensor positioning. Also, as EEG is usually recorded at the same time as electrocardiography (ECG), researchers could use this latter for detecting the pulse waves [78].

Breathing and perspiration artefacts are also related with physiological activity. They have a frequency spectrum near 0 Hz and are removed simultaneously with direct current (DC) component, i.e., using a simple high-pass filter.

Figure 2.6 summarises the activity in EEGs.

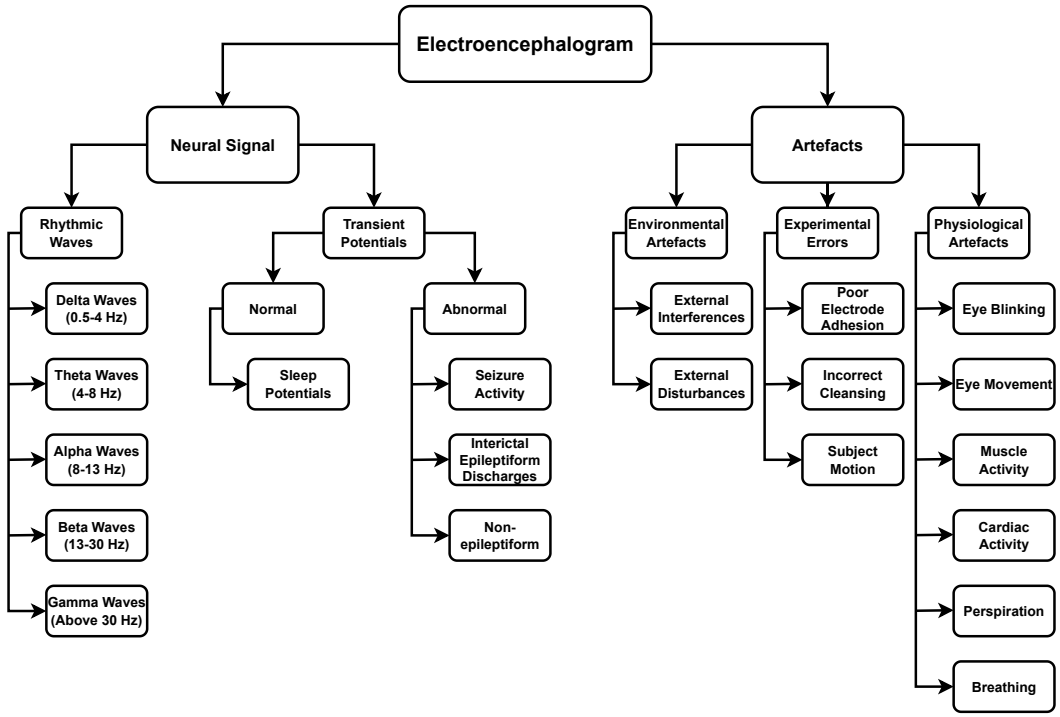


Figure 2.6: Detailed description of the neural signal and artefacts present in EEG signals.

2.3 Epilepsy treatment and therapeutics

The primary form of treatment for patients with epilepsy is the administration of AEDs. However, this type of treatment is not always efficient. Even if the patients take AEDs, they do not achieve a seizure-free condition. This state usually worries them, as they can not perform normal daily tasks. Additionally, treatment with AEDs often comes along with side effects affecting their behaviour. Therefore, resective surgery, neurostimulation, dietary therapies, and warning devices emerge as possible ways to improve patients' quality of life [6].

2.3.1 Antiepileptic drugs and drug-resistant epilepsy

Epilepsy is defined by hyperexcitatory or hypersynchronous neuronal activity resulting from perturbations on the normal balance between excitation and inhibition [84]. The development of AEDs aims to create drugs to control these disturbances. Currently there are around 30 AEDs. These work by enhancing inhibitory mechanisms or blocking excitatory mechanisms [85, 86].

Currently, about one-third of patients with epilepsy are diagnosed with DRE [5–7]. According to ILAE, DRE is identified whenever the treatment using two AEDs do not lead to a seizure freedom condition (see Box 4). Seizure freedom occurs when the patient does not suffer any seizure over one year or three times the longest interval between seizures in the previous year before starting the treatment [87].

Box 4 - Diagnostic criteria for drug-resistant epilepsy (as defined by the Task Force of the ILAE in 2010 [87]).

"Drug-resistant epilepsy may be defined as failure of adequate trials of two tolerated and appropriately chosen and used antiepileptic drugs schedules (whether as monotherapies or in combination) to achieve sustained seizure freedom."

This condition is a serious limitation because patients are not allowed to perform normal daily tasks such as driving and usually have restrictions in professional life. Additionally, as these patients continue to have seizures, there is a high risk of getting brain injuries, fractures and, even, sudden unexpected deaths (SUDEP) [5]. These patients experience a mortality rate 5-10 times higher than that of the general population [88]. Among all types of epilepsies, patients with focal seizures are more susceptible of being diagnosed with DRE [89]. Specifically, patients with MTLE are more prone to drug resistance.

Even after several years of AEDs research, the ratio of patients with DRE remained practically unchanged. However, it should be pointed that the side effects have decreased, which means an improvement for the condition of patients who need this type of drugs [7, 84].

Diagnosis and treatment of DRE at an earlier stage leads to a greater chance of seizure control. However, most patients with DRE are not followed up by an expert epilepsy team. In the United States of America, it is estimated that less than 1% of patients with DRE are evaluated by an epilepsy center. However, on average, they begin to be monitored, about 20 years after the first seizure onset, making seizure control much more difficult [90].

2.3.2 Surgery

One of the solutions for patients with DRE is brain surgery. This procedure consists of removing the part of the brain causing the seizures also known as the epileptogenic zone. However, not all patients with DRE can undergo this procedure. A compreensive assessment is required to identify the epileptogenic zone. This evaluation must be performed within an epilepsy center while presurgical monitoring [91, 92].

Figure 2.7 summarises the evaluation procedure performed to decide whether the patient is able to be undergoing resective surgery. Firstly, clinicians verify whether the patients are suitable to go through presurgical monitoring. During presurgical monitoring, patients are subjected to reduced dosages of AEDs and sleep deprivation to increase the rate of seizures in a short interval of time [6, 93]. Clinicians expect to observe seizures with the same onset characteristics. However, these procedures may have serious consequences. Medication reduction and sleep deprivation may trigger tonic-clonic seizures in patients who have never experienced them. In addition, the patient may experience a seizure cluster. Both procedures should therefore be

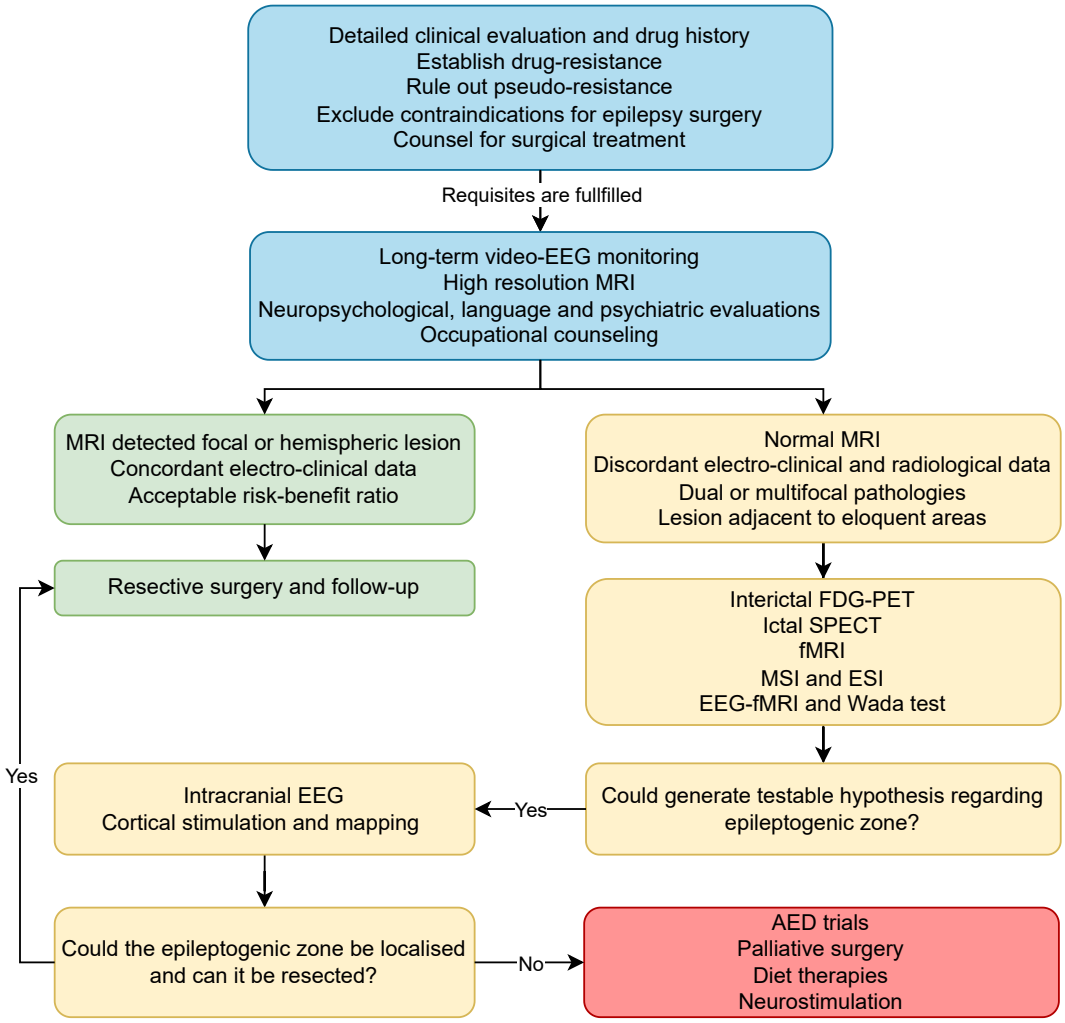


Figure 2.7: Flowchart followed to decide whether the patient with DRE. Blue boxes represent evaluation processes followed for all patients. Green boxes describe the easiest scenarios in which the patients do not have to perform many exams before undergoing resective surgery. Yellow boxes consist of the scenarios in which the clinicians were not certain about the epileptogenic zone at the beginning. Red box details the therapies followed by patients unable to undergo resective surgery. Adapted from [92].

carried out carefully and always under medical supervision [93].

The patient is evaluated in three ways: long-term video-EEG, high resolution magnetic resonance imaging (MRI) and neurophysiological evaluation. If the epileptogenic zone is detected, the clinical information obtained is consistent, and the risk-benefit to the patient is acceptable, then the patient will undergo surgery. Otherwise, further evaluations must be carried out before making a final decision. The new assessments are based on further exams such as positron emission tomography (PET), single-photon emission computed tomography (SPECT), functional magnetic resonance imaging (fMRI), electric source imaging (ESI), and magnetic source imaging (MSI). A further assessment may also be performed based on the acquisition of EEG-fMRI and the Wada test. It aims at minimising possible problems caused by the surgery. After that, the clinicians evaluate the possibility of generating a

testable hypothesis about the epileptogenic zone. If so, the patient is submitted to intracranial EEG, cortical stimulation, and mapping. If the epileptogenic zone is identified and can be removed, the patient undergoes surgery. Otherwise, the patient receives other types of treatment [6, 92].

A high percentage of patients undergoing resective surgery manage to become seizure-free. For patients with intractable TLE, the percentage of individuals, who achieved the seizure freedom condition, varies between studies from two thirds [94] to about 80% [95]. In any case, it is worth noting that the chance of getting rid of seizures after surgery is about ten times greater than if they did not [96]. However, it should be noted that surgeries do not always go well. There are cases where patients experience a kind of "double hit", which means that in addition to not getting rid of seizures, they start to suffer memory decline [97]. This effect should be considered during decision making and when providing information to the patients.

2.3.3 Neurostimulation

Neurostimulation emerges as an alternative therapeutics for patients with DRE who cannot undergo resective surgery. However, contrary to the results obtained after the resective surgery, only a few patients achieve seizure freedom for more than one year after starting a neurostimulation device.

There are two types of neurostimulation devices: invasive and noninvasive. It depends on the type of clinical procedure needed. Noninvasive devices are not clinically validated [98]. The invasive devices comprise implanting a device that transmits electrical pulses to peripheral nerves or specific parts of the brain to impede potential seizures [4, 98, 99]. Furthermore, devices can be classified as open or closed-loop devices. Open-loop devices are based on planned stimulations whereas closed-loop ones automatically analyse biosignals and trigger an electrical impulse whenever they detect seizure activity. The most well-known neurostimulation devices include vagus nerve stimulation (VNS), responsive neurostimulation system (RNS), and deep brain stimulation of the anterior nucleus of the thalamus (ANT-DBS) [4, 98, 99]. Figure 2.8 and Table 2.1 summarise the three devices.

VNS was firstly created as an open-loop system that stimulated the left vagus nerve for 30 seconds every 5 minutes [100]. However, it was later converted into a closed-loop system. For that, the device analyses the heart rate and, if it detects any variation associated to a seizure, it starts the stimulation of the vagus nerve [101]. Even though it has been converted into a closed-loop system, it is possible to trigger a stimulation by placing a specific magnet in front of the device [102]. Of the three devices presented here, the VNS is the only one implanted completely outside the brain, thus having a lower risk.

RNS is a closed-loop system comprising intracranial electrodes placed directly on at most two seizure foci [103]. Therefore, it continuously analyses the brain signals

and starts the stimulation as soon as it detects seizure activity. The objective of the electrical stimulation is to block an upcoming seizure or simply avoid seizure related activity [98,104]. RNS is based on a threshold-based model which uses as features the amplitude, frequency, and rhythmicity of the brain signals. Physicians can control the parameters used by the system to decide when to trigger the stimulation [105].

ANT-DBS consists in an open-loop system that uses intracranial electrodes to deliver scheduled stimulus to the anterior nucleus of the thalamus [106, 107]. This technology is based on studies reporting that the intervention in this part of the brain may avoid the spreading of a seizure over the neural networks [108].

Table 2.1 compares the accepted neurostimulation devices. It should be pointed that there are no studies yet comparing the performance of the three already accepted neurostimulation devices in a randomised form. Literature evidences that both RNS and ANT-DBS present a percentage of patients achieving $\geq$50% reduction in seizure frequency after 1 year (50% responder rate at 1 year) higher than the one obtained by VNS. Furthermore, some patients achieved a seizure freedom condition after 2 years of follow-up using RNS and ANT-DBS. However, it is worth noting that VNS requires no intracranial surgery and it is typically low-cost making it a possible solution for patients from lower income countries [98, 99]. There are concerns regarding the software implemented in the RNS and the closed-loop VNS devices. These technologies report very high false detection rates, leading to a high amount of stimulations compared to seizure frequency. Their results may not be due to the prediction capacity of the models but rather to the long-term neuromodulatory effect provoked by long periods of stimulation [109]. In addition, frequent stimulation leads to a greater battery drain. Thus, the patient will need to go to hospital more frequently to replace it.

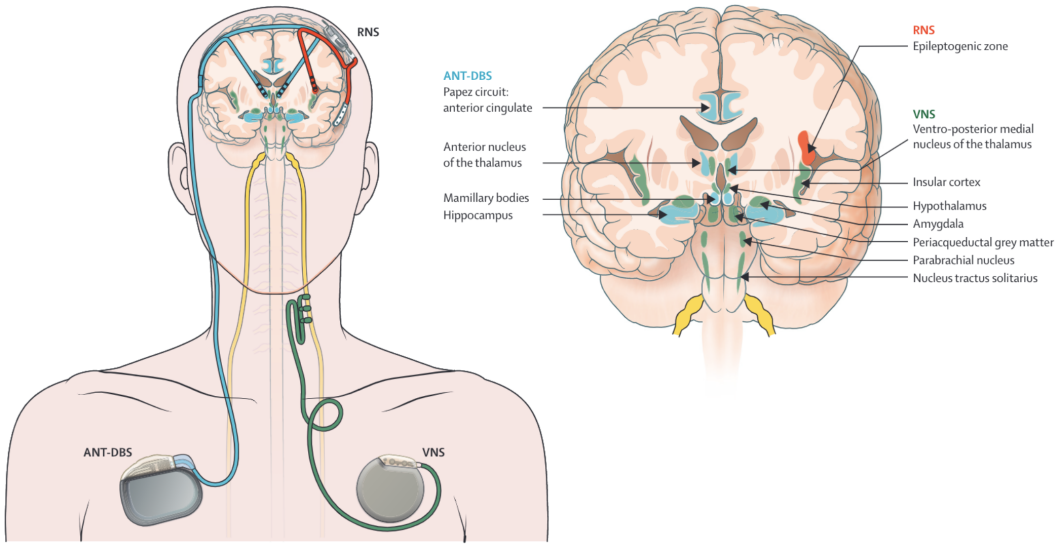


Figure 2.8: Clinically approved neurostimulation devices. Extracted from [98].

Table 2.1: Neurostimulation therapies already approved for patients with DRE. Based on Ryvlin *et al.* [98].

	VNS	**RNS**	**ANT-DBS**
Stimulation target	Left vagus nerve	Ictal onset yone (cortex)	Anterior nucleus of the thalamus
Stimulation system	Open-loop; Closed-loop based on heart rate; On demand	Closed-loop based on features extracted from ECoG	Open-loop
Age	Children $\geq$ 4 years; adults	Adults	Adults
Seizure types	Focal and generalised	Focal	Focal
Epileptogenic focus	Non-localisable; Multifocal; Not resectable	Bitemporal; Eloquent focus	Multifocal; Non-localisable
Level of accessibility	Moderate	Low	Low
Risk of therapy-induced brain lesion	None	Low	Low
Short-term infection rate	1% at 3 months	3% at 3 months	Not available
Long-term infection rate	Not available	12% at 9 years	13% at 10 years
Material dysfunction	1%	5% lead damage or revision	8% lead replacement
50% responder rate at 1 year	31%	42%	44%
Seizure free at 2 years of follow-up	0%	9%	13%

2.3.4 Rescue medication

Rescue medication may help controlling seizures. These could provide three main advantages: (i) patients may achieve a seizure freedom state when combined with other AEDs; (ii) reduce the continuous intake of AEDs diminishing their long-term side effects; and (iii) avoid seizure clusters and prolonged seizures [110].

Benzodiazepines are the most used rescue medication due to their rapid effect (at most 10 minutes). However, these should not be taken continuously due to their secondary adverse effects [110]. Diazepam was the first approved acute medication. Rectal Diazepam was authorised in 1997 by the food and drug administration (FDA) for patients older than 2 years old for seizure cluster treatment outside the clinic [111]. As a consequence of its difficult route of administration, it was not well

accepted by the adult patients and, thus, it was mostly used in children [112]. Diazepam nasal spray was approved 2020 by the FDA to be used in patients older than 5 years [113]. Midazolam is a rescue medication for seizure clusters and prolonged seizures. It can be either taken by via oral or nasal. Oral Midazolam was approved in 2011 by the european union (EU) for prolonged seizures in people until 18 years old and Midazolam nasal spray was authorised in 2019 by the FDA for seizure clusters in patients with 12 years old or more [114–116]. Currently, other rescue therapies are under investigation such as sublingual Diazepam, oral Diazepam, or sublingual Lorazepam. Table 2.2 summarises the already approved drugs.

Table 2.2: Rescue medication therapies already approved for patients with DRE. Based on Wolf *et al.* [110] and Cloyd *et al.* [115].

Drug	Route	Waiting time	Peak level	Adversal effects	First approval
Diazepam	Rectal	5-10 min	10-45 min	Somnolence; diarrhea; headache	FDA in 1997 for seizure clusters for patients older than 2 years
Midazolam	Oral	$<$ 5 min	20-30 min	Sedation; somnolence; depressed levels of consciousness; respiratory depression; nausea; vomiting	EU in 2011 for prolonged seizures in patients younger than 18 years
Midazolam	Intranasal	$<$ 10 min	15-120 min	Somnolence; headache; nasal discomfort; throat irritation; rhinorrhea	FDA in 2019 for seizure clusters for patients older than 12 years
Diazepam	Intranasal	$<$ 5 min	$>$ 60 min	Somnolence; headache; nasal discomfort	FDA in 2020 for seizure clusters for patients older than 6 years

It should be noted that these drugs may be a life saviour when linked to prediction devices as patients could take them before an upcoming seizure as long as they are warned with at least 10 minutes anticipation.

2.3.5 Dietary therapies

Ketogenic diet, modified Atkins diet and the low-glycaemic diet emerge as possible alternative solutions for patients with DRE [117]. Ketogenic diet is the most used therapy for epilepsy caused by metabolic disorders such as GLUT1 deficiency syndrome. More than half of the patients following a ketogenic diet have a reduction

of more than half of the seizures. Actually, one-third of them have a reduction of more than 90% and some achieve seizure freedom condition [6]. However, ketogenic diet causes long-term complications such as lack of nutrients, growth disorders in children, abnormal lipid profile, bone disorders, and kidney stones [6].

2.3.6 Warning devices

Warning devices emerge as a further solution to patients with epilepsy. There are two types of warning devices: the detection devices aiming at detecting and annotating every seizure and the prediction devices, which are designed to notify the patient that a seizure will happen [118].

Detection devices would be useful to complete patients' seizure diaries leading to a better follow-up of the disease [119]. Although the great potential of EEGs to detect seizures [120], currently only non-EEG devices such as acelerometry [121], electromyogram [122], photoplethysmography [123], or electrocardiography [124], were considered for phase III validation [118].

Prediction devices would notify the patient some minutes before the seizure onset. Thus, patients would be able to take precaution to minimise consequences. Patients may also be able to take a rescue medication and block an upcoming seizure. NeuroVista Seizure Advisory System [125] was one of the most relevant prediction devices. In a phase I clinical trial, it was validated on 15 patients with DRE over two years. It comprised ECoG electrodes on the epileptogenic zone connected to a telemetry unit implanted on the chest which transmited wireless data to a personal advisory device. The personal advisory device notified the patient about their seizure susceptibility. This device was a breakthrough to the seizure prediction research area because it showed that it is possible to correctly predict seizures for some patients and improve their lives. Seer health developed a prediction software based on seizure diaries. The app uses information from previous seizures to notify patients about the seizure probability [126].

Although no EEG-based warning device have been studied beyond phase I clinical trial, this may change in the future. In recent years, several ssEEG acquisition systems have appeared such as the UNEEG SubQ [72], the EpiMinder Subscalp [127] the Neuroview [128], the Soenia Ultimate EEG [70], the Epios [129], and the EASEE [130]. These ssEEG acquisition systems have minimal intrusion compared to ECoG and overcome some of the problems of scalp EEG, thus making it possible to collect brain signals over several months. It could enable developing warning devices based on brain signals and improve their performance.

2.4 Seizure prediction

Seizure prediction is the main topic of this thesis. It is an active research theme studied since 1970 [131]. Its main goal is developing a system able to anticipate upcoming seizures.

As shown in Figure 2.9, epileptic EEG may be partitioned into four stages: the interictal, the preictal, the ictal, and the postictal periods. Despite this division, the intervals' location are not well-defined. The ictal stage is the period when the seizure occurs. It is the only epileptic stage identified by clinicians. It is usually found by inspecting video-EEG [132]. The interictal corresponds to the normal stage, varying from hours to days depending on the frequency of seizures. The postictal matches the period after the seizure until both EEG and behaviour return to the normal stage. Finally, preictal stage is the most important period in this research area. It is not clearly defined as it varies from seconds to minutes depending on the patient and the seizure [18].

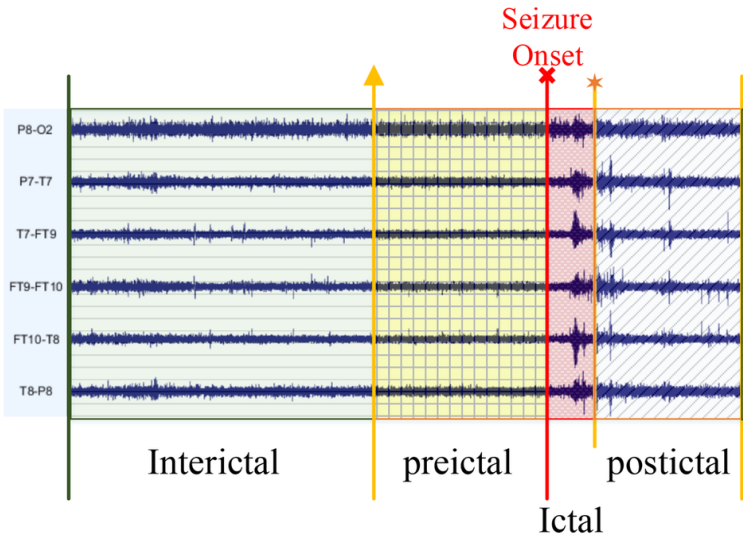


Figure 2.9: Representation of the four seizure stages in the epileptic EEG signal. The interictal matches the interval between the postictal and preictal stages of consecutive seizures. The preictal corresponds to the transition period between the normal and seizure stages. The ictal stage is when the seizure occurs. Finally, the postictal period matches the interval after the seizure. Extracted from [133].

Seizure prediction models are developed to find the preictal period, i.e., identify the transition from interictal to ictal period. It is not a straightforward task because of the rarity of seizure events. In fact, Cook *et al.* [125] reported that on average patients with DRE only experience seizures during less than 0.05% of total acquisition time. Therefore, a high class imbalance between interictal and preictal periods makes it harder to develop intelligent models because they use far more interictal samples than preictal samples to tune the decision boundary. Furthermore, preictal period is uncertain. Thus, researchers have to estimate it considering the ictal onset. An ideal seizure prediction model would give the patient the exact time when the

upcoming seizure happens. However, since the preictal period is not clearly defined, it is nearly impossible to develop such a method.

2.4.1 Seizure onset

Seizure onset times are crucial to develop seizure prediction solutions. In presurgical conditions, these are usually obtained through the inspection of video-EEG. There are two types of seizure onset: (i) the EEG onset, which corresponds to the moment in which the first changes relative to the ictal period start in the EEG; and (ii) the clinical onset, which corresponds to the moment when the patients manifest symptoms derived from the seizure they are suffering [134, 135]. Not all seizures present both onsets. In the case of the subclinical seizures, there are only changes in the EEG and therefore only EEG onset can be found. Furthermore, neurophysiologists sometimes struggle at finding the clinical onset for seizures classified as FOIA type [136]. Seizure prediction models generally use the EEG onset as the point of transition from the preictal to the ictal period [137].

2.4.2 Lead seizure

Seizure prediction models require data from several seizures to be correctly developed. However, the considered seizures must be independent events with no relation [138]. Thus, in the case of seizure clusters, only the first seizure, also called lead seizure, should be considered for developing the seizure prediction approaches since all the others that occur right after might be triggered by the first one [136]. Therefore, if one can predict and avoid the lead seizure, it is probable that none of the following would occur. Furthermore, it would not make sense to use the other seizures as it is difficult to distinguish the four epileptic stages.

There is no consensus on how far apart the seizures should be in time to be considered lead seizures. One of the most used definition for seizure cluster considers that a patient with DRE suffers a seizure cluster whenever at least three seizures occur within a 24-hour interval [51]. However, this definition may lead to a great loss of data since, on average, a patient in presurgical conditions experience 3.6 seizures per day [139]. Therefore, researchers have considered lead seizure as consecutive seizures separated by: (i) 1 hour [140], 1.5 hours [141], 2 hours [32], 4 hours [142], 4.5 hours [143], 5 hours [76], and 8 hours [125].

2.4.3 Seizure prediction vs detection

Seizure detection frameworks aim at detecting the onset of the ictal period. Suppose seizure detection is possible at an early stage. The detector model may be integrated into a closed-loop neurostimulation device to cancel seizures before they spread and begin clinical manifestations. Seizure detection is also very useful to complete patients' seizure diaries or to help clinicians to find the seizure onsets, making their

work simpler because they would no longer need to analyse the whole long EEG, but only the segments where a seizure was detected [144]. Seizure prediction is based on a similar concept to seizure detection. However, instead of finding the onset of the ictal period, it aims to find the preictal period. Therefore, its goal is to inform the patients that a seizure is about to happen in the next minutes/hours, giving them more time to take preventive measures in order to minimise the consequences of possible accidents or to take rescue medication to try to avoid the seizure [137, 145].

2.4.4 Seizure prediction vs forecasting

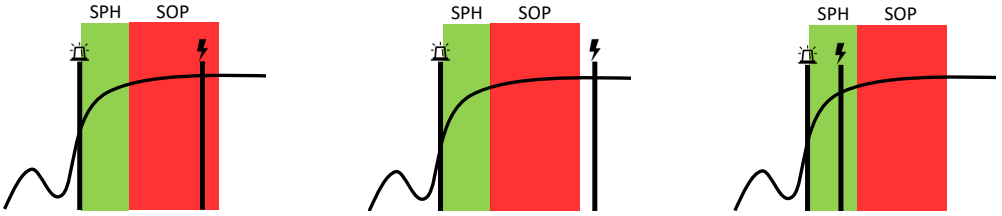
Seizure forecasting has recently appeared as a new concept for seizure prediction. It aims to model the probability of the patient having a seizure [127]. Thus, there is no longer the need to find out the transition instant between the interictal and preictal periods and the objective is to obtain the seizure risk over time [18]. On days with high seizure risk (proictal state), patients may prepare themselves to reduce the consequences of seizures. However, unlike traditional seizure prediction, even if the patient is warned of a high-risk period, it does not mean a seizure will happen. The high-risk period only means a high probability of suffering a seizure. As a consequence of the seizure risk modelling, the researchers no longer look at the signal on a minute scale rather on an hour/day scale since the goal is to find out periods where there is a high risk and not to find that a seizure is going to happen in the following minutes. Based on longer periods, seizure forecasting allows clinicians to understand the relationship between seizures and biological cycles such as the circadian and multidien cycles [16].

2.4.5 Seizure prediction characteristic

By the early 2000s several seizure prediction approaches had already been presented. However, how these models were evaluated led to doubts about whether these approaches could be used in clinical applications. In 2003, Winterhalder *et al.* [139] proposed the seizure prediction characteristic. This framework is comparable to an alarm system which triggers an alarm whenever the system predicts a seizure. In this way, the authors proposed evaluating the models using seizure sensitivity (SS) and false prediction rate per hour (FPR/h). In addition, this alarm system needs two intervals: the seizure prediction horizon (SPH) and the seizure occurrence period (SOP).

SPH or intervention time is the period between the alarm and the starting of the SOP. This period consists of an interval where researchers consider that there is not any risk of seizure occurrence. Therefore, this time is used to diminish seizure's consequences or cancel it using for example seizure-suppression drugs. The length of this period depends on the type of system which researchers are developing. In case of an implanted device, it could consist of just a few seconds because it will cancel

the seizure automatically. However, in the case of a warning system, this period has to be as long as some minutes to give the patient time to prevent dangerous scenarios, such as driving a car. It should be noted that when authors identify this period as intervention time, the SOP is named SPH [146]. SOP is when a seizure is expected to occur. Consequently, high values of SOP could put the patients under psychological stress and anxiety. Furthermore, systems developed under the condition of a long SOP could automatically fail the prediction of some seizures, e.g., a prediction model designed considering a SOP that lasts 6 hours will not be able to predict seizures spaced 4 hours apart. Therefore, this period should be minimised. Only alarms where seizures happen inside SOP are considered true (see Figure 2.10).



(a) An alarm is raised and the seizure happens after a portion of time equal to SPH and before the ending of SOP, thus it is a true alarm.

(b) An alarm is raised and the seizure happens after the ending of SOP, thus it is a false alarm.

(c) The alarm is raised inside the SPH, thus it is a false alarm.

Figure 2.10: Requirements needed to be considered a true alarm.

SOP and SPH are fundamental for the development and performance assessment of seizure prediction models. As presented in Figure 2.11, the SPH allows the model to give the patient a certain period to prepare for a seizure, whereas SOP is the period when the seizure happens. When training the models, the preictal samples correspond to an interval with the same duration as SOP [137, 146]. The samples following the training preictal samples and ending at the seizure onset correspond to the SPH and are discarded from the analysis. In this way, in case of a true alarm, the patient will have an interval equal to the SPH to prepare for the upcoming seizure and the SOP when the seizure is expected to occur [58].

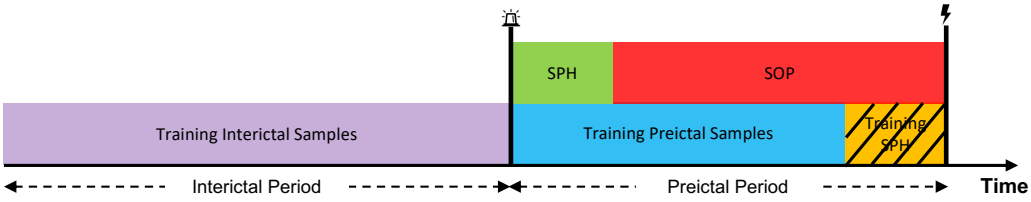


Figure 2.11: Representation of how to train a seizure prediction approach. The training preictal period has the same duration as the SOP. The training SPH corresponds to the interval before the seizure. Thus, if an alarm is triggered at the beginning of the preictal interval, the model will wait a period equal to the training SPH, and the seizure will occur within the SOP, in this case, at the end.

2.4.5.1 Performance assessment

Standard classification approaches are evaluated using the confusion matrix presented in Table 2.3 considering true positive (TP) as a well predicted preictal sample, true negative (TN) as a well predicted interictal sample, false positive (FP) as a interictal sample predicted as preictal and false negative (FN) as a preictal sample predicted as interictal. From confusion matrix, researchers may compute sensitivity (see Equation 2.1) and specificity (see Equation 2.2). Some authors used those metrics to compute area under the receiver operating characteristic curve (AUC). Although these metrics can be used to evaluate classifiers, they are not able to inform whether the system is predicting well upcoming seizures. Therefore, they had to be adapted to correctly evaluate the seizure prediction approaches.

Table 2.3: Confusion matrix for evaluating sample predictions given by seizure prediction models.

		Predicted Labels	
		Interictal	**Preictal**
True Labels	**Interictal**	True Negative	False Positive
	Preictal	False Negative	True Positive

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{2.1}$$

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives} \tag{2.2}$$

Sensitivity was replaced by the SS in order to verify how many seizures were well predicted and specificity was replaced by FPR/h because there was not any advantage of measuring the number of true negatives, i.e., the number of well predicted interictal samples. Equations 2.3 and 2.4 present how to calculate the SS and FPR/h of a seizure prediction model [139]. SS is the ratio between the number of true alarms ($\#\,True\ Alarms$) and the number of seizures ($\#\,Number\ of\ Seizures$), in other words, alarm's accuracy, whereas the FPR/h is the ratio between the number of false alarms ($\#\,False\ Alarms$) and the time length when false alarms could be triggered. Whenever an alarm is triggered, there is usually a period with a time length equals to the duration of SPH plus SOP, also known as refractory period, when the predictor stops analysing. Therefore, this period should be subtracted from the FPR/h denominator since it is not possible to fire false alarms during those periods ($Interictal_{Duration} - \#\,False\ Alarms \times Refractory_{Duration}$).

$$Seizure\ Sensitivity = \frac{\#\,True\ Alarms}{\#\,Number\ of\ Seizures} \tag{2.3}$$

$$FPR/h = \frac{\#\,False\ Alarms}{Interictal_{Duration} - \#\,False\ Alarms \times Refractory_{Duration}} \tag{2.4}$$

Both metrics must always be used together because they complement each other. A prediction system could achieve a high SS and yet not performing well, e.g., considering a model with 100% of SS and a FPR/h of 3. Although it predicts all upcoming seizures, it fires excessive false alarms comparing to true alarms. Therefore, it could not be implemented in a real-world scenario.

The optimal values for SS and FPR/h would be 100% and 0, respectively. However, achieving those values is a hard task. Therefore, more tolerant standard results should be specified. Regarding SS, Schulze-Bonhage *et al.* [147] presented that the majority of epileptic patients would only use a seizure prediction device if it has a SS higher than 90%. Considering FPR/h, Winterhalder *et al.* [139] suggested a FPR/h of 0.15 for patients under presurgical conditions and 0.0042 for patients under normal conditions. These values are based on the number of seizures experienced by these patients in both conditions, i.e., patients under presurgical circumstances who experience a reduction of antiepileptic medication suffer an average of 0.15 seizures per hour comparing to 0.0042 under normal condition, i.e. when they take medicine regularly and are not under presurgical stress.

2.4.5.2 Statistical validation

Beyond the performance metrics, researchers also use statistical tests to validate their models. These tests are often used to prove that the models are reliable because non-zero SS and low FPR/h values can be achieved without having a trustworthy model [148]. The most used validation tests are the comparison with an unspecific random predictor and the seizure time surrogate analysis [137].

Unspecific random predictor Unspecific random predictor was firstly proposed by Winterhalder *et al.* [139]. It consists in two equations used to calculate the probability of triggering an alarm without using any information from the EEG. Equation 2.5 describes the probability p of firing an alarm during a small interictal interval I. Equation 2.6 calculates the probability P of raising at least an alarm in a longer interval W.

$$p = FPR/h \times I \tag{2.5}$$

$$P = 1 - (1 - FPR/h \times I)^{W/I} \approx 1 - e^{-FPR/h \times W} \; for \; I \ll W \tag{2.6}$$

When considering $W = SOP$, P represents the probability of triggering an alarm inside the SOP, i.e., the SS of a random prediction method. Therefore, not only the FPR/h but also the SOP increases the performance of a random predictor, i.e., the chance of a random method to have an equal or better performance than a seizure prediction model.

The problem of the abovementioned equations is that those only consider one seizure event. Consequently, Schelter *et al.* [2] introduced a method that not only

considers several seizures but also several seizure prediction models. It is based on a homogeneous Poisson process for false predictions. Accordingly, the probability of firing an alarm is given by Equation 2.7.

$$P_{Poisson} = \frac{\#\ False\ Alarms}{\#\ Samples} \tag{2.7}$$

Considering $W = SOP$ and $FPR/h \times SOP \ll 1$, the probability P from Equation 2.6 may be approximated to Equation 2.8.

$$P \approx FPR/h \times SOP \approx FPR/h \times SOP \tag{2.8}$$

The probability P is the basis for the Equation 2.9. It is based on the cumulative binomial distribution giving the probability of predicting at least k upcoming seizures using at least one of the d seizure prediction models. Therefore, increasing the number of models also increase the probability of predicting seizures by chance. However, in a real-time scenario, warning devices only use one seizure prediction model ($d = 1$).

$$P_{binom,d}\{k; K; P\} = 1 - (\sum_{j<k} \binom{K}{j} P^j (1-P)^{K-j})^d \tag{2.9}$$

The critical value to test statistical significance is given by Equation 2.10. Basically, it is the ratio between the maximum k to which the probability given by Equation 2.9 is higher than a previously chosen significance level (α) and the total number of seizure events.

$$\sigma = \frac{argmax_k\{P_{binom}\{k; K; P\} > \alpha\}}{K} \times 100\% \tag{2.10}$$

Unspecific random predictor does not consider the EEG patterns. Furthermore, this method may not be appropriate whenever a low number of testing seizures is available, which is often a case, because seizures are rare events and methodologies tend to be patient-specific (see Figure 2.12).

Surrogate analysis The surrogate analysis [148] uses Monte Carlo simulation to randomly shuffle the labels while maintaining the input's order. Then, the new seizure prediction model is tested in this new time series. If the model tested on the false preictal times gets equal or better than the one tested on the true data it is considered that the model is not following any pattern (see Figure 2.13). Unlike the unspecific random predictor, this validation analysis considers the EEG for calculating the sensitivity and FPR/h. Furthermore, it may be used either with a low or a high number of testing seizures. Therefore, it is more useful than the unspecific random predictor when analysing whether the seizure prediction model follows any pattern.

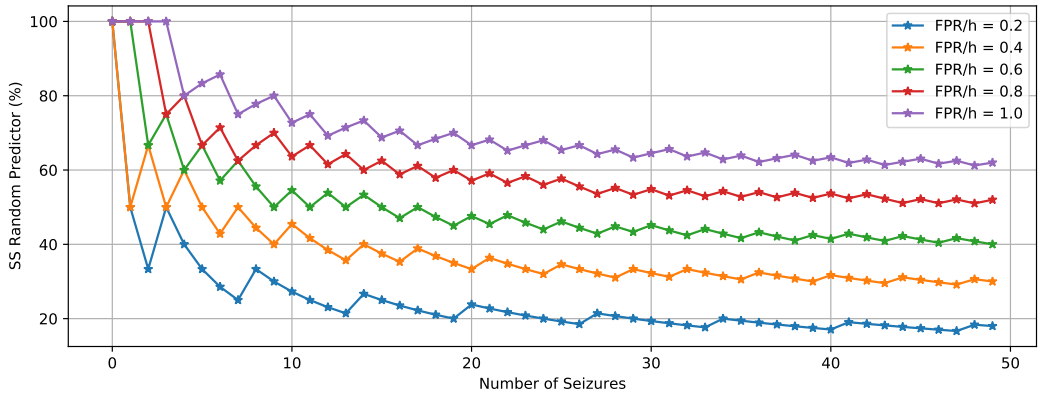


Figure 2.12: SS of the unspecific random predictor for different number of seizures and FPR/h considering a SOP of 30 minutes. Adapted from Schlter *et al.* [1].

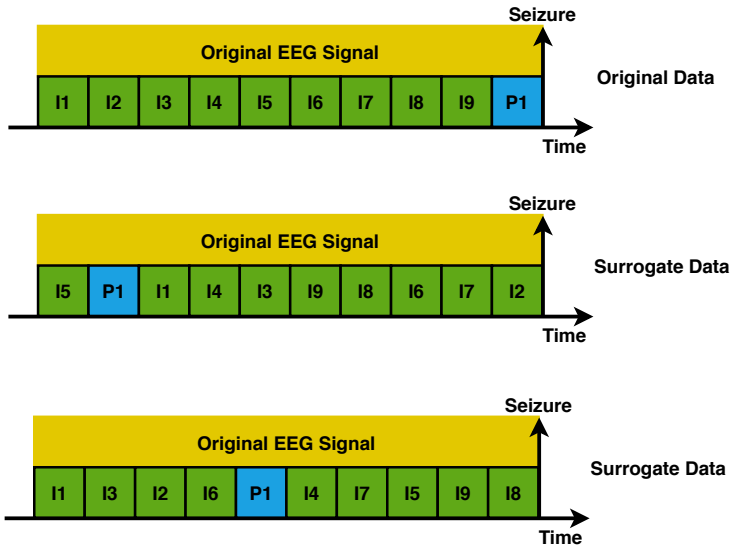


Figure 2.13: Seizure time surrogate analysis examples. The first schema presents the original data composed by the original EEG signals and the original labels. The other two schemas present two different surrogate time series, i.e., the original EEG signal with two different false labelling vectors. Legend: I: Interictal Window; P: Preictal Window. Adapted from Schlter *et al.* [2].

2.4.6 Postprocessing

Seizure prediction models based on machine learning algorithm usually classify EEG segments independently without considering their temporal relation. This characteristic makes them highly susceptible to false predictions, e.g., if noise is acquired over the long-term recording [137]. Therefore, postprocessing methods, such as Kalman filter [149] and the firing power regularitation [150], should be used to smooth the classifier's output.

2.4.6.1 Kalman filter

Kalman filter consists of the computation of the states s_k of a linear dynamic system, at the instant k, where y_k is the output variable, w_k and v_k are the zero mean white noise vectors, and T_p is the prediction interval. It is computed following the system of equations presented in Equation 2.11. This filter was first applied as a postprocessing method for seizure prediction in Chisci *et al.* [149]. An alarm was raised every time the output variable becomes positive [150].

$$\begin{cases} s_{k+1} = \begin{bmatrix} 1 & T_p \\ 0 & 1 \end{bmatrix} s_k + w_k \\ y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} s_k + v_k \end{cases} \tag{2.11}$$

2.4.6.2 Firing power filter

Firing power regularisation [150] consists of applying a moving window to the seizure prediction classifier's output, equal to SOP. Therefore, it accumulates the predicted output of several samples. Equation 2.12 gives the mathematical formulation of this postprocessing method where $fp[n]$ is the firing power regularisation output, τ is the number of samples of the moving window, and $o[k]$ is the seizure prediction model output at time k.

$$fp[n] = \frac{\sum_{k=n-\tau}^{n} o[k]}{\tau} \tag{2.12}$$

An alarm is triggered as soon as a ratio of instants inside the moving window, classified as preictal exceeds a defined threshold. This threshold should be defined before using it. A high threshold means that an alarm is only triggered when a high concentration of preictal samples leads to a lower number of false alarms. However, SS may also be low because many consequent preictal samples are required to trigger an alarm. Compared with the Kalman filter, the Firing Power produces fewer false alarms [150].

2.5 Concept drifts

Concept drift occurs when there is a change in the relation between the input variables and the target, in other words, it occurs when data distribution fluctuates over time [151]. Common machine learning algorithms assume that data characteristics keep stationary, i.e., that the learned rules during training can be used to make correct future predictions. However, as seen in Figure 2.14 there are different types of concept drift such as sudden, incremental, gradual drifts and reoccurring concepts. These changes are usually responsible for decreasing the model performance since it is not prepared to handle them. Thus, intelligent models must be able to adapt to these fluctuations in the data.

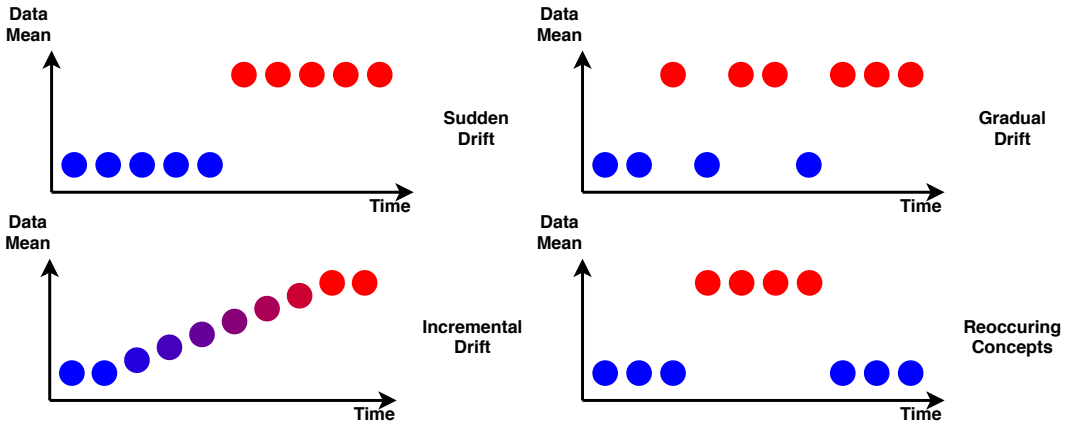


Figure 2.14: Different types of concept drift. Adapted from [151].

Possible solutions for adapting the models to these changes are: periodically refitting the model from the scratch; periodically update model weights; weight the training data according to its importance; create new models able to correct the old model; training several models for different concepts and detect and choose the model that handles better the future data concepts.

EEG acquired from patients with epilepsy in presurgical conditions may contain several distribution drifts. These drifts result from alterations on the biological rhythms and the sleep-wake cycle provoked by modifications in the drug dosage, sleep deprivation, and stress. Baud *et al.* [77] reported that the interictal epileptiform activity, an EEG biomarker observed between seizures changes over time. Furthermore, the number of seizures experienced by a patient diagnosed with epilepsy could vary from 3.6 per day to 3 per month depending on the dose of AED taken [139]. Therefore, researchers must collect EEG for several weeks or even months in order to be able to study all kinds of concept drifts existing in the problem under study and, consequently develop adaptive seizure prediction models able to handle all these changes while predicting upcoming seizures.

2.5.1 Rhythms in epilepsy

Circadian, multidien, or ultradian rhythms are biological rhythms that may alter the seizure susceptability [77, 152] (see Box 5). Circadian rhythms comprehend 24-hour physiological cycles such as heart rate, blood pressure, hormonal production, body temperature and sleep-wake cycle. Ultradien rhythms include the non-rapid eye movement (NREM) - rapid eye movement (REM) cycle which, in humans, lasts approximately 90 minutes. Multidien rhythms comprise cycles that repeat after several days.

2.5.2 Sleep-wake cycle

Sleep is a primary brain function important in regeneration and learning. Unhealthy sleep can lead to disorders such as sleepness, depression, and even death [153]. Sleep is controlled by the sleep-wake cycle which is directly connected to the human circadian cycle [154]. Sleep-wake cycle comprehends awake and sleep stages. Awake stage lasts on average two-third of the cycle and is easily distinguished on the EEG due to the low-amplitude high-frequency waves and the high number of artefacts, such as electromyogram (EMG) contaminations, which are almost absent during sleep. Sleep stage comprises NREM and REM stages. During sleep, adults experience four to six cycles. Each cycle lasts about 90 minutes [155]. According to the American Academy of Sleep Medicine [153], a health sleep starts with the first substage of the NREM (N1) that is mainly composed by theta waves. Afterwards, it moves to substage N2 that manifests K-complexes and sleep splindles. Then, NREM stage ends with the N3 substage, also called deep sleep, which is easily distinguished by slow-wave activity and delta rhythms. Finally, the cycle finishes with the REM stage which comprises theta and beta waves making it similar to the EEG acquired during the awake stage [156].

Sleep is highly related to seizures. In fact, NREM is a stage with high cerebral hypersynchrony facilitating seizure onset. Seizures are caused by the transition between the substages being the N2 the one with higher seizure susceptibility. On the other hand, REM is a stage of seizure suppression linked to very few seizures [157]. The lack of sleeping is also related to triggering seizures. Sleep deprivation is commonly used in presurgical conditions to provoke more seizures in a short period. It may be explained by the fact that sleep deprivation induces NREM stage which facilitates seizure onset and the propagation of seizures [158].

Box 5 - Definition of circadian, multidien, and ultradian rhythms (proposed by Khan *et al.* [152] in 2018).

Circadian rhythm: *"A biological rhythm is considered to be a circadian rhythm if it meets three criteria: the rhythm should have an endogenous free-running (approximately) 24h period, should be entrainable (i.e., be capable of phase reset by environmental cues and synchronisation to the 24h day) and should exhibit temperature compensation."*

Multidien rhythm: *"Refers to rhythms with a time period covering several days."*

Ultradien rhythm: *"Refers to rhythms with periods of less than 24h; ultradian rhythm cycles can occur with a frequency of more than once per day."*

2.6 Deep neural networks

DNNs are advanced machine learning models based on artificial neural networks (ANNs) inspired on the biological processes of the human brain [159, 160]. Figure

2.15 puts these models into a hierarchy including machine learning and artificial intelligence.

DNNs have recently become popular as a result of the increase of the amount of available data, the development of new weight optimisation approaches and also because of the upgrade of the technology in terms of hardware and software [161]. The great advantage of these architectures is that unlike the classical machine learning algorithms, they can automatically learn how to classify data without human expert knowledge.

There are several different DNNs such as deep convolutional neural network (DCNN) and deep recurrent neural network (DRNN). They have been used in different domains such as bioinformatics [162, 163], speech recognition [164, 165], image classification [166, 167], among others. It should be noted that all DNNs belong to ANN class. However, to simplify the explanation, in the following sections, ANN term will not be used as a class but as a feed-forward neural network, which may also be considered a DNN if it contains several layers.

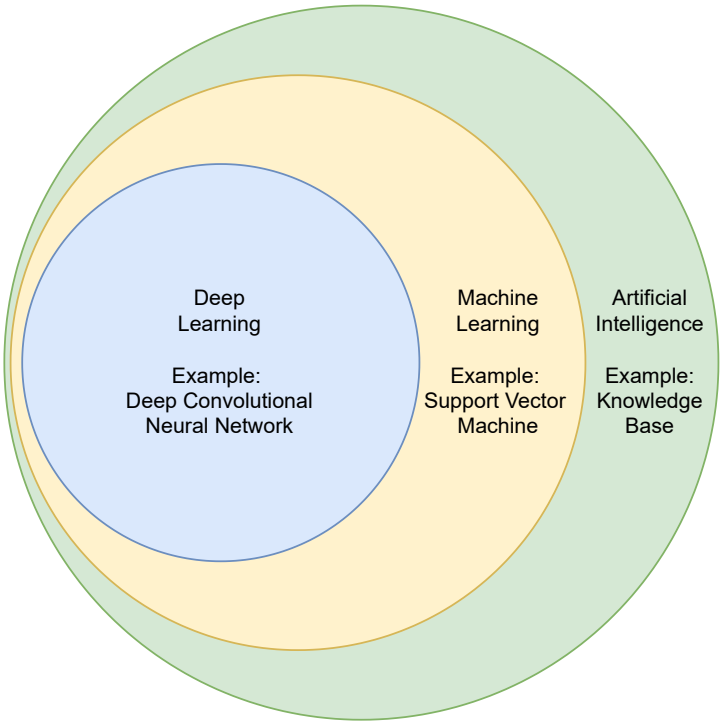


Figure 2.15: Venn diagram representing the hierarchy of artificial intelligence, machine learning and deep learning. Adapted from Figure 1.4 of [161].

2.6.1 Artificial neural networks

ANN is the most simple neural network architecture [161]. It is made of several artificial neurons connected with each other. As presented in Figure 2.16 an artificial neuron is inspired in biological neurons.

The different inputs simulate the various signals collected in dendrites. Each

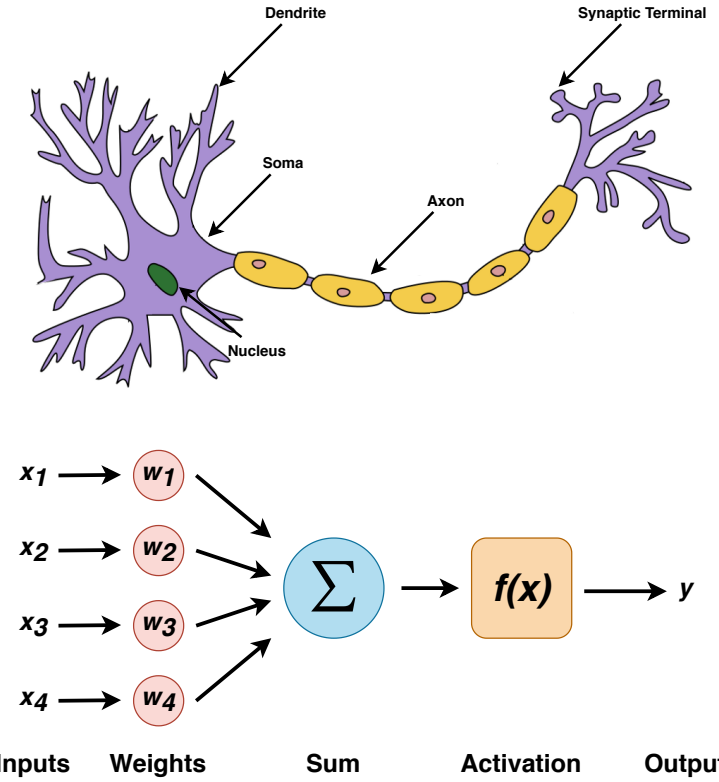


Figure 2.16: Comparison between biological and artificial neurons. Adapted from [168].

weight represents a different synaptic coupling' strength received by the neuron. The sum function symbolises the integration of all received synapses performed at soma. This sum goes through an activation function, similar to the action potential of a neuron, and its transmitted to other neurons (synapse). There are two types of activation functions: the linear function and the nonlinear ones. The nonlinear functions are usually preferred to develop DNNs because they enable the models to learn more complex patterns (nonlinear behaviour). Examples of such functions are logistic function, hyperbolic tangent, softmax, rectified linear unit (ReLU), leaky ReLU, and swish [169].

A biological neural network is composed by several layers of neurons. To simulate the brain's behaviour, an ANN should be composed by an input layer, an output layer and several hidden layers, as seen in Figure 2.17. The input layer takes and transmits the input data to the hidden layers. Consequently, these hidden layers explore the data performing different transformations to find patterns. It should be pointed out that all the neurons of a certain layer transmit responses to all the neurons of the following layer. However, the following layer weights modulate the strength of each response, i.e., there could be some neurons that practically do not influence the output. Finally, the output layer merges all the responses transmitted by the last hidden layer to accomplish a certain task. It is worth noting that an ANN with multiple layers is able to generate abstract features from the input data. Therefore, it is no longer considered a simple neural network but a DNN [170].

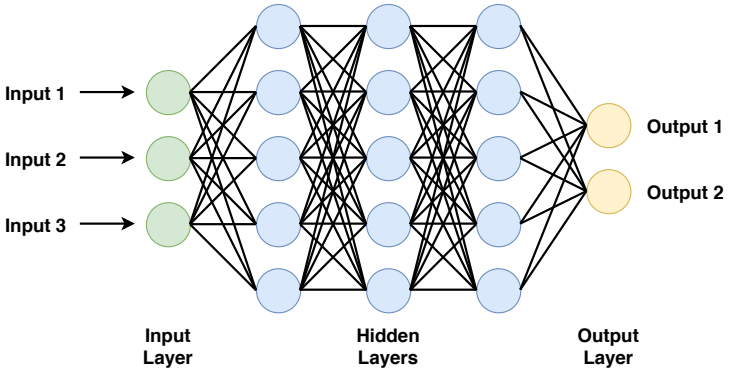


Figure 2.17: Example of an ANN. The circles represent the artificial neurons and the connections correspond to the different weights attributed to previous layer's neurons response.

2.6.2 Deep convolutional neural networks

DCNN is a special type of DNN able to automatically extract features from complex data such as biosignals, audio, images or even three-dimensional (3D) data. It simulates the behaviour of the visual cortex, i.e., first it divides the data in small regions, then it analyses this subregions individually and finally it merges these subareas maintaining the spatial characteristics (see Figure 2.18).

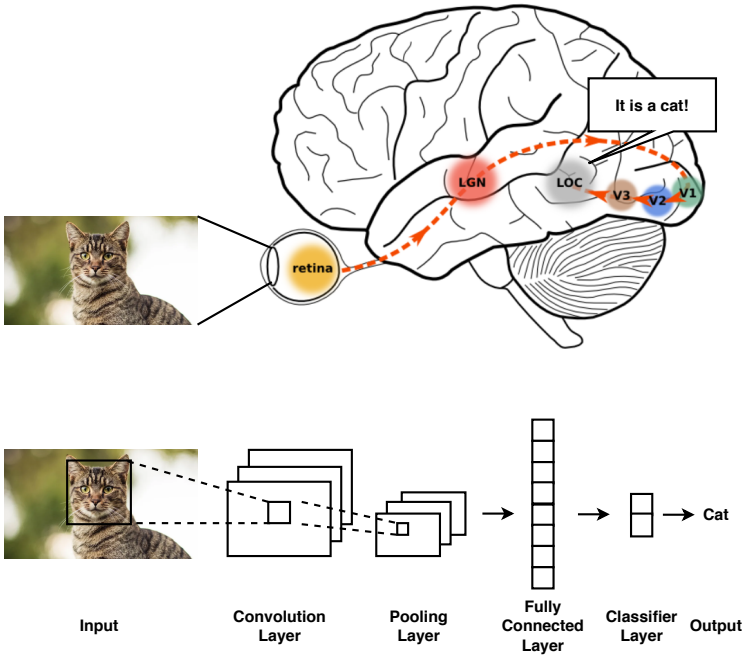


Figure 2.18: Comparison between visual cortex and DCNN. Adapted from [171].

Contrary to the ANN architecture, DCNN does not connect all the neurons, which reduces its complexity. Furthermore, this architecture is translation invariant, enabling it to operate effectively even when the sample is changed from its initial position (e.g., reflection, rotation, or translation) [172]. As presented in Figure 2.18, DCNN simulates the processes performed by visual cortex to recognise an image. For that, it uses convolutional layers, pooling layers and dense layers.

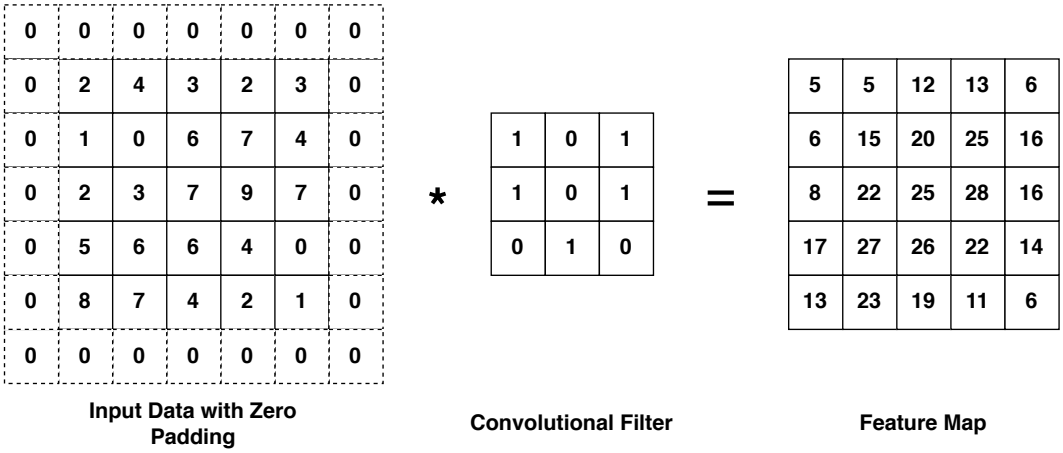


Figure 2.19: Example of convolutional layer with a 2D input. The filter is convolved with the image from the upper left to the lower left moving one pixel at a time. Example: Upper Left Convolution: $0 \times 1 + 0 \times 0 + 0 \times 1 + 0 \times 1 + 2 \times 0 + 4 \times 1 + 0 \times 0 + 1 \times 1 + 0 \times 0 = 5$. It should be noted that zero padding is using for maintaining the spatial information.

The Convolutional layer is composed of several filters, used for extracting features from the input data, optimised during the learning process. As observed in Figure 2.19 each filter performs a moving convolution covering the entire input data step by step to create a feature map. The number of pixels the filter moves at a time is named stride. This method is repeated as often as the number of filters used. Then, every feature map is passed through an activation function and converted to another set of values. Furthermore, authors usually use batch normalisation layers to normalise every convolutional layer output [173].

Pooling layer is a downsampling layer with no weights. There are different types of pooling layers being maximum and average pooling frequently used. Figure 2.20 presents an example for both types of pooling layer using a stride equals to 1.

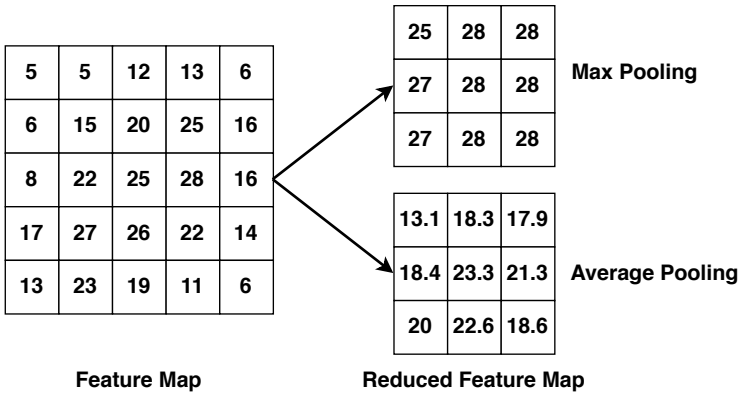


Figure 2.20: Example of maximum and average 3x3 pooling layers with a 2D input. Maximum pooling layer selects the maximum value using a 3x3 window whereas the average pooling layer performs the average of all 9 values.

Although this layer is specified for downsampling, some researchers prefer to use convolutional layers with stride higher than one and without padding to reduce the number of values in the input data. Consequently, the model can automatically

learn the optimal weights for downsampling instead of simply removing information.

Finally, the dense layers are usually used at the network's end. Contrary to the preceding layers, these are only able to analyse 1D data. These layers usually have three main goals: merging all the information available in the feature maps, reducing the number of used features to lower the probability of model overfitting, and classifying the samples. Contrary to convolutional and pooling layers, every neuron of one layer is connected to every neuron of the preceding and following ones. Therefore, they are also named fully connected layers. It should be pointed out that although these layers are the most commonly used to transform the data from feature maps to a feature vector, they can be replaced with global pooling layers (layers which output one value per feature map) drastically reducing the number of parameters while increasing the performance in some scenarios [174].

2.6.3 Deep recurrent neural networks

DRNN is a special type of neural network which has internal memory. As seen in Figure 2.21, it can keep an internal memory that is maintained over timesteps making it able to create sequential rules, i.e., exploring current data using information from the past [175], e.g., text translation and video classification.

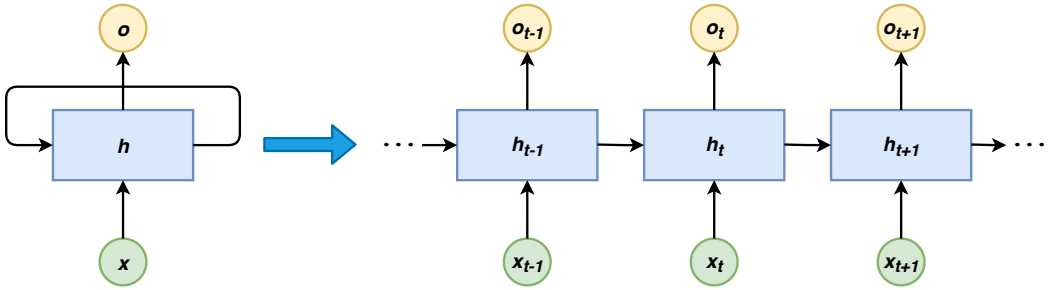


Figure 2.21: Example of simple DRNN architecture. Adapted from [170].

However, simple DRNN architecture suffers from vanishing gradient during its training with gradient-based learning methods. Besides, it has a long-term dependency problem since the information learned from past sequence samples does not disappear [176]. In order to overcome such disadvantages, Hochreiter and Schmidhuber presented long short-term memory (LSTM) [177], a type of DRNN that is able to control the flow of the information over the timesteps, e.g. over the images of a video.

As presented in Figure 2.22, this architecture is more sophisticated than the simple DRNN. The cell state, C, works as a memory, in other words, it keeps the old information. However, this memory is controlled by the forget gate, f. If the forget gate outputs a vector of zeros the multiplication with the old cell state will be zero and consequently the memory will be erased. On the other hand, if the output of this gate is a vector of ones, all the old information flows through the cell. Also, the input gate, i, controls how much information goes to the cell state in each

timestep. Supposing this gate's result is a vector of numbers near zero, almost no information is passed to the new cell state. The output gate, o, controls how much information from the hidden, h, and input states, x, is used to compute the new hidden state. All the processes being performed inside the LSTM are described by Equations 2.13, 2.14, 2.15, 2.16, and 2.17. During training, the network optimises the weight matrixes W. There are two types of matrixes W, the ones that multiply with the input vectors, W_x, and the ones that multiply with the hidden state vectors W_h. The model also learns the bias vector, b.

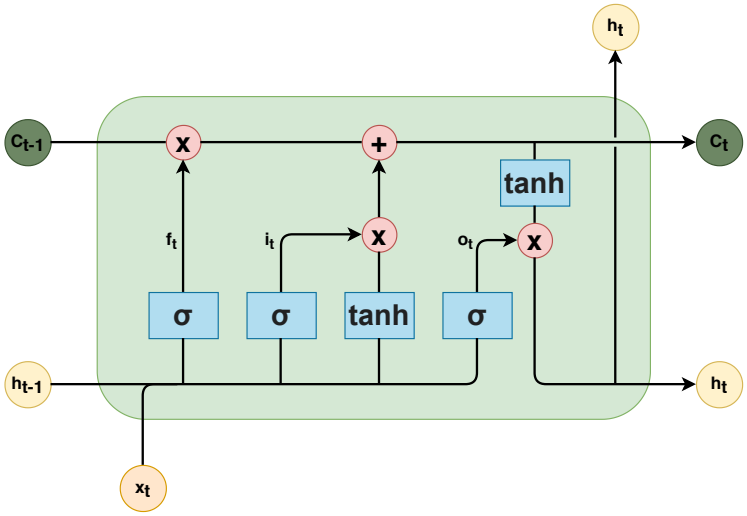


Figure 2.22: LSTM unit. The forget gate, input gate, output gate and cell states are represented by f_t, i_t, o_t, C_{t-1}, C_t, respectively. Adapted from [178].

$$f^{(t)} = \sigma(W_{fx} \cdot x^{(t)} + W_{fh} \cdot h^{(t-1)} + b_f) \tag{2.13}$$

$$i^{(t)} = \sigma(W_{ix} \cdot x^{(t)} + W_{ih} \cdot h^{(t-1)} + b_i) \tag{2.14}$$

$$o^{(t)} = \sigma(W_{ox} \cdot x^{(t)} + W_{oh} \cdot h^{(t-1)} + b_o) \tag{2.15}$$

$$C^{(t)} = f^{(t)} \odot C^{(t-1)} + i^{(t)} \odot tanh(W_{Cx} \cdot x^{(t)} + W_{Ch} \cdot h^{(t-1)} + b_c) \tag{2.16}$$

$$h^{(t)} = tanh(C^{(t)}) \odot o^{(t)} \tag{2.17}$$

In summary, the high complexity of the LSTM layer turns the DRNNs able to learn more complex rules. However, it requires high computational time due to the many trainable parameters.

Gated recurrent unit (GRU) architecture [179] is a lighter version of LSTM (see Figure 2.23). Contrary to the LSTM, this architecture does not have cell state and only uses two gates to control the information flow, the reset gate (r_t) and the update gate (z_t). Reset gate controls the past information flow whereas the update gate decides what new information to add. Consequently, it has a lower number of parameters than a LSTM layer making its optimisation faster. All the processes

being performed inside the GRU are described by Equations 2.18, 2.19, and 2.20. As in the LSTM, during training, the network optimises the weight matrixes W and the bias vector b.

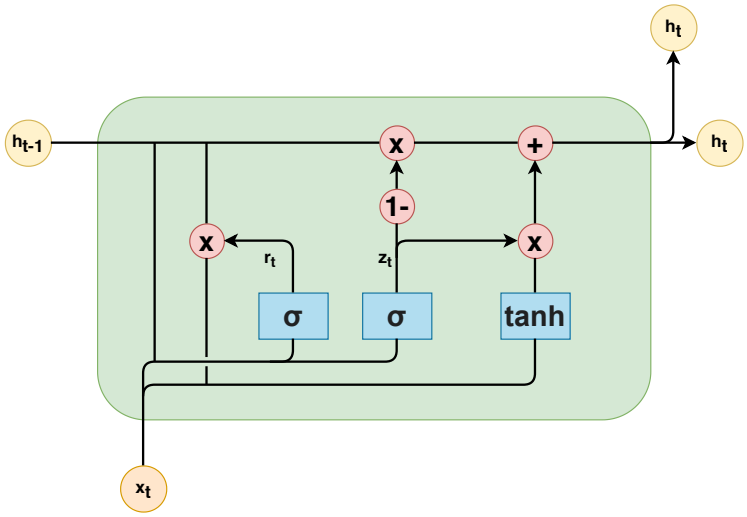


Figure 2.23: GRU. The update gate and reset gate are represented by z_t and r_t, respectively. Adapted from [178].

$$z^{(t)} = \sigma(W_{zx} \cdot x^{(t)} + W_{zh} \cdot h^{(t-1)} + b_z) \tag{2.18}$$

$$r^{(t)} = \sigma(W_{rx} \cdot x^{(t)} + W_{rh} \cdot h^{(t-1)} + b_r) \tag{2.19}$$

$$h^{(t)} = (1 - z^{(t)}) \odot h^{(t-1)} + z^{(t)} \odot tanh(W_{hx} \cdot x^{(t)} + W_{hh} \cdot (r^{(t)} \odot h^{(t-1)})) \tag{2.20}$$

Beyond the possibility of using past information, all DRNN architectures can also use information from future timesteps, e.g., analysing an EEG window from the left to the right and from the right to the left. For that, bidirectional layers must be used [175].

2.7 Transfer learning

Transfer learning is a method that allows using knowledge acquired from a large dataset (source domain) to develop a new model in other domain (target domain) [180]. Therefore, it is often used when the training dataset is not large enough to create a model from scratch, as in the case of seizure prediction where the models must be trained in a patient-specific way and the number of seizures per patient is quite limited [181].

Over the last years, with the rapid growth of deep learning methods, researchers have focused on using transfer learning to develop new DNNs, also known as deep transfer learning. It can be divided into four categories: instance-based, mapping-based, network-based, and adversarial-based [182, 183].

Instance-based approach uses some samples from the source domain to train the model on the target domain (see Figure 2.24). For that, it selects samples from the source domain similar to the target domain and weights them accordingly.

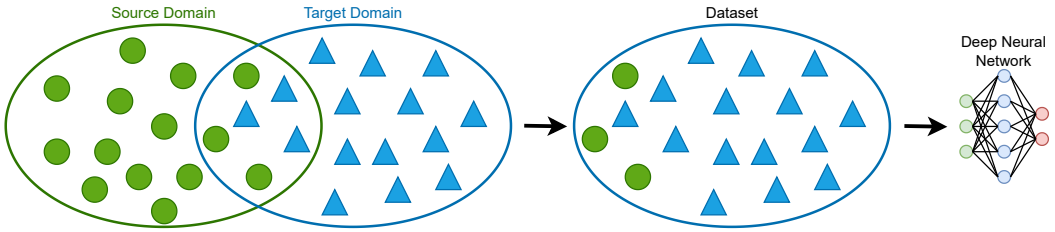


Figure 2.24: Instance-based transfer learning. Adapted from [183].

Mapping-based approach converts samples from both source and target domains into a new data space where both have similar data distributions. Afterwards, they use this new dataset to train the model (see Figure 2.25).

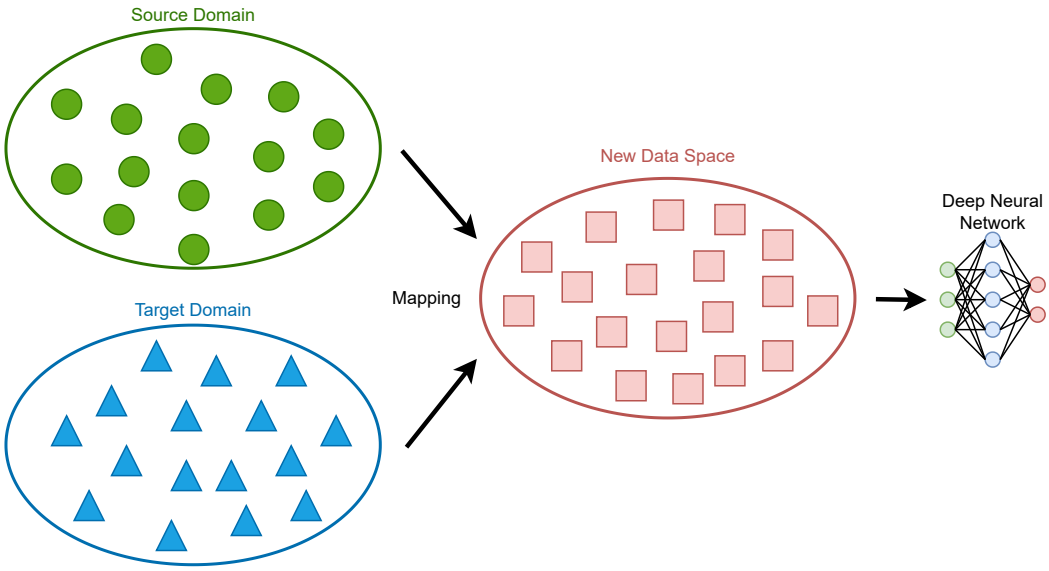


Figure 2.25: Mapping-based transfer learning. Adapted from [183].

Network-based consists of reusing a part of the deep learning architecture previously trained in the source domain and optimising the new layers on the target domain (see Figure 2.26). DNNs developed using ImageNet dataset [184–187] are often used in this type of approach [188, 189]. This approach is also going to be used in Chapter 8.

Adversarial-based approach compreehends training a model that can extract features that follow similar distributions for both source and target domains so that there are no significant differences between the features extracted in both domains. Simultaneously, researchers must be able to use those features to perform accurate classifications on the main task (see Figure 2.27).

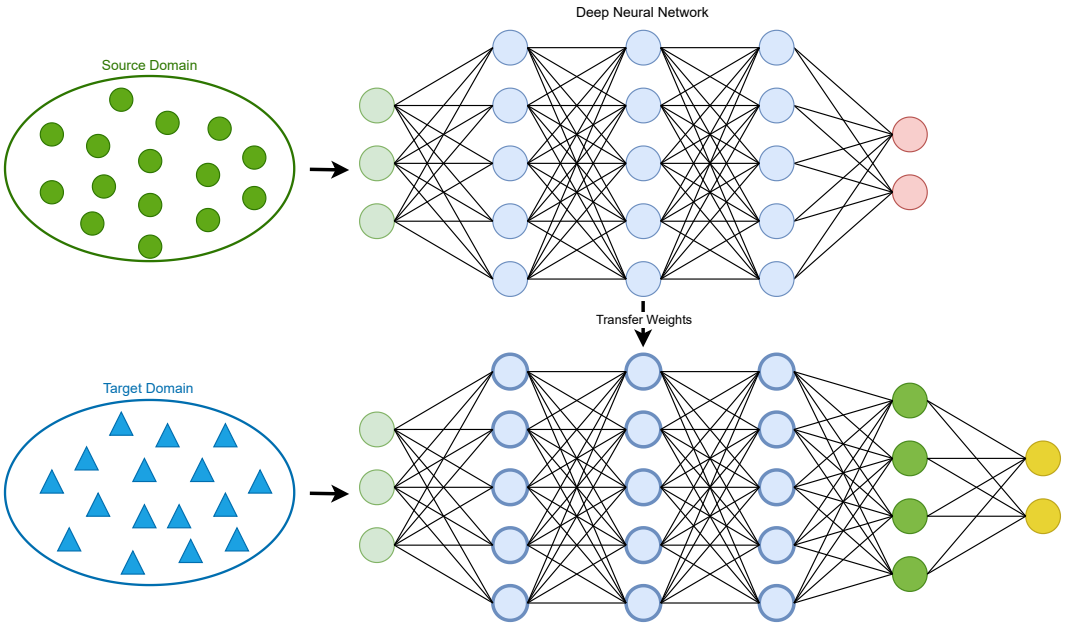


Figure 2.26: Network-based transfer learning. Adapted from [183].

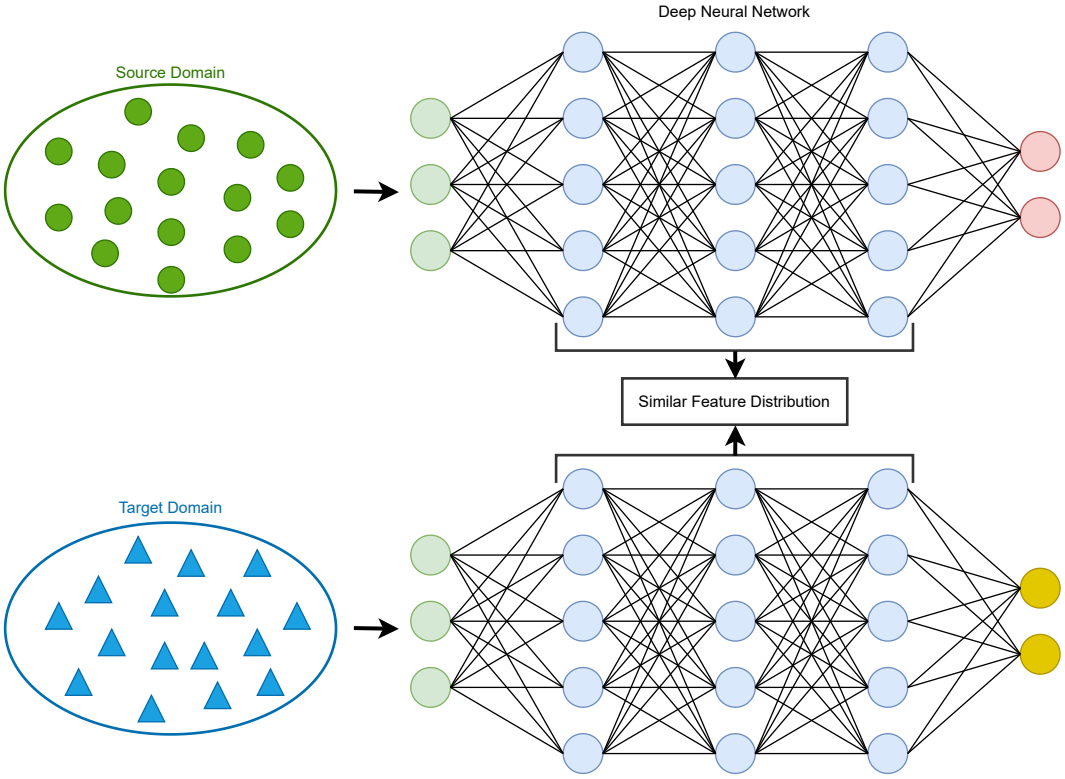


Figure 2.27: Adversarial-based transfer learning. Adapted from [183].

2.8 Summary

Epilepsy is a chronic neurological disease characterised by brief and recurrent seizures. It comprises three classification levels: seizure type, epilepsy type, and epilepsy syndrome. However, filling in all these three levels is not always possi-

ble due to a lack of information leaving the classification incomplete for a subset of patients. Although there are drugs to control seizures, about one-third of the patients can not get the seizure freedom condition, also known as patients with DRE. It is a very severe condition as the patients may randomly suffer a seizure.

Patients with focal seizures are generally prone to have DRE. Specifically, patients with TLE are more susceptible to drug resistance. There are some treatments to take care of these patients. For that, patients have to go to an epilepsy center to be assessed. During the diagnostic, experts decide whether the patient is recommended for resective surgery or referred to palliative treatment. The main source of acquired data is the EEG. This signal acquires the brain's electrical activity and is widely used to study seizures. EEG represents two types of potentials: rhythmic potentials related to brain waves of various frequencies and transients related to the activity that appears spontaneously and is easily distinguishable from background activity. EEG also captures a lot of artefacts from different sources, such as experimental errors, acquisition environment, and physiological errors, which may provoke false results.

EEG is the most used signal to develop seizure prediction models. Seizure prediction aims to find the period before the seizure (preictal interval) and trigger an alarm as soon as it happens. Each alarm is associated with an SOP and an SPH. The SOP is when the seizure is expected to occur, whereas the SPH is the interval between the alarm and the SOP. The SPH should be long enough to give the patient time to prepare for the seizure, whereas in the case of the SOP, it should be short enough to avoid anxiety. Regarding the evaluation, seizure prediction models should be assessed using SS and FPR/h. Furthermore, statistical tests should be performed to verify whether the models perform better than a random approach.

Seizure characteristics change a lot from patient to patient, and therefore, models should be developed following a patient-specific approach. Moreover, being a task based on nonstationary time series, several concept drifts occur over time, leading to the need for approaches that adapt to the several shifts.

Researchers have been developing seizure prediction models based on deep learning architectures. The most used architectures are DCNN and LSTM since they allow exploring local and global details within each EEG window. The great advantage of these architectures is their ability to learn directly from the data. In addition, it is relatively simple to perform transfer learning using these. Thus, researchers may, for example, train a general model using data from several patients and then fine-tune the model for each patient, turning it into a patient-specific approach.

Chapter 3

State of the Art

This chapter reviews the previous and current state of the art related to electroencephalogram (EEG) artefact removal methods and epileptic seizure prediction. It is divided into three sections. Section 3.1 presents approaches to remove noise from EEGs preparing them for further analysis. Then, Section 3.2 describes databases used to study epilepsy and the methods utilised to prepare the data and develop seizure prediction models. Finally, Section 3.3 presents a summary of the literature review.

3.1 EEG artefact removal

Seizure prediction studies require long-term EEGs. Therefore, patients must keep their daily basic tasks such as eating, sleeping, talking and walking while being monitored. As expected, these tasks negatively affect the acquisition because they frequently produce patterns unrelated to neural information (artefacts) such as electrode movement or activation of the head and neck muscles. Furthermore, EEG electrodes usually capture other physiological signals such as eye movements and cardiac pulse [23, 78]. A possible solution would be to detect and remove those segments. However, this removal would result in a high loss of information and, therefore should only be carried out in cases where nothing can be extracted from the signal (electric flatlines, saturated segments and excessive interference). Thus, researchers have developed several techniques for removing noisy data from the EEGs while trying to maintain neural information [190–197].

3.1.1 Simple digital filtering

Simple digital filtering is a highly used technique for removing undesired frequency bands from the EEGs. This technique comprises simple finite impulse response (FIR) and infinite impulse response (IIR) filters which could be divided in low-pass, band-pass, high-pass, band-stop and notch filters. Regarding studies using scalp EEG, the data are usually filtered using low-pass, high-pass, and notch filters [23, 78].

Low-pass filters reduce the effect of noisy high-frequencies whereas high-pass filters remove low-frequency components such as the direct current (DC) component and breathing artefacts. There is no standard cutoff frequency for both types of filters, depending only on the problem and what the authors choose. The notch filter removes the powerline component introduced by the alternate current (AC) power supply. Filtering is usually performed using IIR filters [74, 137]. Compared to FIR filters, IIR are faster and usually require less coefficients to obtain a similar performance [198]. Although simple digital filters are highly used to prepare EEG data, these operate over complete frequency bands and, therefore, can not separate signals with overlapped frequency spectra. Thus, they may not be used to separate EEG from other physiological signals [23, 24].

3.1.2 Linear regression

Linear regression was the most used method for artefact removal until the 1990s as a result of its lower computational complexity [78]. It was highly used for removing ocular activity from the EEGs [190,199,200]. Linear regression computes the filtered signal $\hat{s}$ by subtracting noise n multiplied by a previously optimised coefficient β from the measured signal x (see Equation 3.1).

$$\hat{s_k} = x_k - \beta \times n_k \tag{3.1}$$

Despite its simple formula, it presents two major drawbacks that led to its replacement as the state of the art of artefact removal algorithms. First, the linear behaviour is not able to adapt perfectly to the nonlinearity of physiological processes, and the requirement of a reference noise signal to train the models makes them unsuitable when template signals are not available [24, 78].

3.1.3 Advanced filtering

Advanced filters adjust their weights to reduce the difference between the estimated and desired clean signals. They tend to generate better results than the linear regression [24, 83]. The three main types of advanced filtering methods usually used to remove artefacts from EEGs are adaptive filtering, Wiener filtering and Bayesian filtering [23, 78].

Adaptive filters assume that both signal and artefacts are not correlated. They require an artefact reference for generating an approximation signal, very similar to that template, to subtract it from the input [191, 201–204]. Therefore, they require extra electrodes to be able to operate [23, 78]. Equation 3.2 presents the adaptive filter formula using least mean squares (LMS) optimisation algorithm, where the $\hat{s}$ is the filtered signal, the x is the measured signal, the w are the filter weights, the n is the noise signal, and the μ is the adaptation step size.

$$\begin{aligned} \hat{s_k} &= x_k - w_k^T n_k \\ w_{k+1} &= w_k + 2\mu \hat{s_k} n_k \end{aligned} \tag{3.2}$$

Contrary to adaptive filtering, Wiener filtering is based on the estimation of the artefacts [192]. The filter uses a statistical procedure to find an approximation to the artefacts and generate a linear time-invariant filter that produces an output similar to the pure signal. The power spectrum density (PSD) is an example of statistical information that may be used to estimate the artefacts [205]. Though it does not require a reference, overcoming the drawback of the adaptive filters, it always needs to be calibrated before its usage which may not be possible in real-time scenarios [23, 78]. Equation 3.3 presents the Wiener filter formula using the PSDs as statistical information about the signals, where the $H(f)$ is the filter in the frequency domain, the $|X(f)|^2$ is the PSD of the measured signal, the $|N(f)|^2$ is the PSD of the noise signal, the $\hat{s}$ is the filtered signal, the x is the measured signal, and the h is the filter in the discrete time domain.

$$\begin{aligned} H(f) &= \frac{|X(f)|^2 - |N(f)|^2}{|X(f)|^2} \\ \hat{s}_k &= x_k * h_k \end{aligned} \tag{3.3}$$

Finally, Bayesian filtering is a method based on Markov dynamics able to work in real-time without any reference signal overcoming the aforementioned advanced filtering techniques. However, due to its complexity, it needs to be approximated using Kalman filters [193, 206] or particle filters [207]. As in the case of the Wiener filters, Bayes filters require calibration before its usage which may lead to some problems in non-controlled environments [23, 78]. Equation 3.4 presents the Kalman filter formula for noise removal, where the $\hat{s}$ is the filtered signal, the P is the state variance variable, the Q is the process variance constant, the K is the Kalman gain, the x_k is the measured signal, and the R is the measurement constant.

$$\begin{aligned} \hat{s}_{k|k-1} &= \hat{s}_{k-1|k-1} \\ P_{k-1} &= P_{k-1|k-1} + Q_k \\ \hat{s}_{k|k} &= \hat{s}_{k|k-1} + K_k(x_k - \hat{s}_{k|k-1}) \\ K_k &= P_{k|k-1} + (P_{k|k-1} + R)^{-1} \\ P_{k|k} &= (1 - K_k) P_{k|k-1} \end{aligned} \tag{3.4}$$

3.1.4 Source decomposition

Source decomposition methods, such as wavelets and empirical mode decomposition (EMD) approaches, aim at separating the neural information from the artefacts by decomposing each signal channel into different waveforms.

Wavelet decomposition consists of dividing the signal into several components [208]. The most used wavelet technique is the discrete wavelet transform (DWT). It is de-

fined by Equation 3.5.

$$\begin{aligned} x_{a,L} &= \sum_{k=1}^{N} g[k] * x_{a-1,L}[2n-k] \\ x_{a,H} &= \sum_{k=1}^{N} h[k] * x_{a-1,L}[2n-k] \end{aligned} \tag{3.5}$$

It splits the signal into different levels using scaling functions $g[n]$ (low-pass filters) and wavelet functions $h[n]$ (high-pass filters). The components $x_{a,H}$ generated after applying the wavelet functions are called detail coefficients whereas the components $x_{a,L}$ generated by applying the scaling functions are called approximation coefficients. The approximation coefficients $x_{a,L}$ are downsampled by two and used to compute another level of coefficients. The process is performed until the required decomposition level A is reached. To perform the artefact removal, the detail coefficients $x_{a,H}$ below a defined threshold are removed, and then the inverse DWT is applied to obtain the denoised signal $\hat{s}[n]$ [194, 209–212] (see Equation 3.6).

$$\hat{s}[n] = \sum_{a=1}^{A} \sum_{k=1}^{N} g[k] * x_{a,L}[\frac{n}{2} - k] + \sum_{a=1}^{A} \sum_{k=1}^{N} h[k] * x_{a,H}[\frac{n}{2} - k] \tag{3.6}$$

Similarly to simple digital filtering, wavelets cannot remove artefacts that overlap frequencies of interest without removing important data [78, 82].

EMD is a heuristic algorithm that aims to decompose the signal into different time series named intrinsic mode functions (IMFs) [213]. It is able to adapt itself to nonlinear and nonstationary signals, such as EEG. These components are obtained through an iterative pipeline. This pipeline first identifies the local maxima and minima and connects them using interpolation, generating two envelopes. Afterwards, the average of both envelopes is computed and subtracted from the original signal obtaining a residual signal called IMF. This process is performed until no more extremes are found on the residual signal. After that, the IMFs containing artefacts are removed whereas the others are summed to form the reconstructed signal $\hat{s}[n]$ [82, 195, 214–216] (see Equation 3.7).

$$\hat{s}[n] = \sum_{m=1}^{M} IMF_m[n] + Residue_M[n] \tag{3.7}$$

However, as in the case of the wavelets, the selection of the noisy IMFs is not automatic. Therefore, before the implementation, it is necessary to define an automatic evaluation to classify the IMFs. Furthermore, EMD is computationally complex and thus difficult to be used in real-time.

3.1.5 Blind source separation

Blind source separation (BSS) methods aim at the separation of the signals X in their independent sources S assuming that the measured signals X result from the

multiplication of a mixing matrix A by the independent sources S plus a noise component N, as described by equation 3.8.

$$X = AS + N \tag{3.8}$$

To compute the unmixing matrix W, BSS methods suppose that the noise and error values N are contained in the independent sources S and that the combination of the sources S is linear. Afterwards, the unmixing matrix $W = A^{-1}$ is used to obtain an approximation of the independent sources $\hat{S}$. Equation 3.9 presents the main formula for performing a BSS.

$$\hat{S} = WX \tag{3.9}$$

It is worth noting that these methods do not find more independent sources than the number of measured signal channels. Thus, they tend to not perform satisfactorily when few channels are available [24]. Figure 3.1 presents an example of how BSS works.

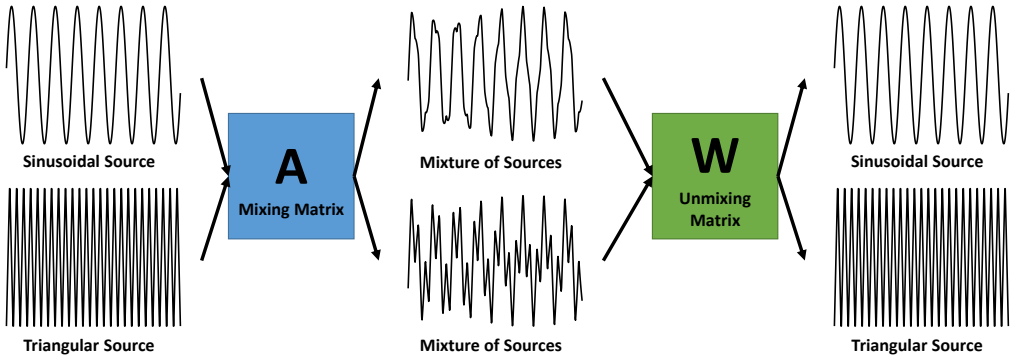


Figure 3.1: BSS example.

The unmixing matrix W is calculated using different statistics depending on the algorithm. There are four classes of BSS algorithms: high-order statistic methods, second-order statistic methods, second-order statistic methods with nonstationary properties and methods that use temporal and spectral information to separate the sources [220]. The most used ones are described below, namely the restricted and extended Infomax independent component analysis (ICA), adaptive mixture independent component analysis (AMICA), FastICA and second-order blind identification (SOBI) [23, 78, 221, 222].

All the aforementioned BSS algorithms assume temporal and spatial stationarity throughout the entire recording and that the sources are statistically independent. Therefore, for using these algorithms the data should be divided into quasi-stationary segments, and the acquisition setup configuration must be maintained throughout the recording [78, 223]. For making a more accurate unmixing matrix calculation, these methods require a preprocessing step where the data is whitened, usually using principal component analysis (PCA), in order to cancel the correlation between

the signal's channels. Then, as stated previously, the unmixing matrix is calculated depending on the type of algorithm [23, 220]. Infomax ICA is a high-order statistic method. It searches for the best unmixing matrix using maximum likelihood estimation. There are two different types of Infomax ICA algorithm: the restricted version which can separate leptokurtic sources, e.g., physiological sources, and the extended version which is able to distinguish both leptokurtic and platykurtic sources, e.g., the electrical artefacts. Therefore, both methods are able to separate brain signals from the physiological artefacts [224]. AMICA is an improved version of the Infomax ICA. It calculates both probability density function (PDF) and spatial projection for every component while the Infomax ICA assume that all the components follow one or two template PDFs. Furthermore, this method is able to combine more than one ICA model to achieve a better source separation. Therefore, this model is able to give a better approximation of each source [217]. FastICA is also a high-order statistics method based on maximising the negentropy of the source distributions. However, unlike the Infomax and AMICA algorithms, this method always converges to the same point. It is faster than the previous ones when the components are well-defined non-Gaussian sources, i.e., when there are no near-Gaussian components [225]. Finally, SOBI algorithm is a second-order statistics algorithm. It calculates the optimal unmixing matrix by reducing the correlation between time-lagged versions of the source activities. Compared to the high-order statistics algorithms, this is not only faster but also has a satisfactory performance turning it a potential algorithm for real-time situations in which computational time should be minimised [78, 217, 226].

Although there are different approaches for getting the independent sources of a signal, all of them share the same purpose: obtaining artefact-free signals. Delorme *et al.* [217] presents a large comparison of different BSS algorithms where they report that AMICA shows the best performance for source separation. However, it presents results similar to the Infomax-based methods which require less computational power and present fewer hyperparameters to tune.

BSS methods are considerably used for EEG artefact removal [24, 78, 217] even in epilepsy studies [25, 218, 219]. Generally, EEGs are considered to be generated by independent dipolar sources that mix together following a linear model. Thus, BSS algorithms tend to perform well for separating brain signals from artefacts [217, 220]. Furthermore, these methods do not require any external information about the type of artefacts before their usage, making them an important solution when artefact references are unavailable.

3.1.6 Hybrid blind source separation

Hybrid BSS methods combine the advantages of nonlinear and linear algorithms to better separate the independent sources. Moreover, these can also remove artefacts

from unichannel signals, making them valuable whenever a few channels are available, e.g., ultra-long-term EEG acquisition using UNEEG [57]. However, merging two complex algorithms increase the time required to process the data. Extensive computational time may be a problem in real-time operation [23]. Wavelet-BSS [227–230] and EMD-BSS [231–234] are examples of hybrid BSS pipelines.

3.1.7 Automatic blind source separation

Despite the satisfactory performance of BSS methods on separating the neural and artefact components, they are not automatic, i.e., the selection of the neural components has to be performed manually. Automatic BSS approaches classify each independent component (IC) based on features containing temporal, spectral, or spatial information. These methods receive the ICs, classify them and perform the signal reconstruction by discarding the artefact components.

Automatic BSS approaches have been created using temporal, spectral and spatial handcrafted features extracted from the time series, PSD and topographic map of each IC, respectively [218,235–242]. Recently, new approaches based on deep neural networks (DNNs) were formulated [196, 243, 244]. Croce *et al.* [243] developed deep convolutional neural networks (DCNNs) using the PSDs and the topographic maps. To develop their models, they used 1,067 ICs from EEGs and 4,749 ICs from magnetoencephalograms (MEGs), labelled by trained experts. They reported that their approach outperformed classifiers based on manually extracted features. Lee *et al.* [244] introduced a Bayesian deep learning model based on convolutional layers and an attention module using only topographic maps. They used a dataset comprising 12,020 ICs classified by trained experts. Their results indicate that the attention module improved the classifier performance. Pion-Tonachini *et al.* [196] presented the current state of the art model for automatically preprocessing EEG data using ICA. They compared three different deep learning architectures to label ICs: (i) a DCNN; (ii) a DCNN trained with a cross-entropy function that double weighted brain IC classification errors (weighted DCNN); and (iii) a generative adversarial network (GAN). They trained these architectures using PSDs, topographic maps, and autocorrelation functions. The authors used EEG recordings from the Swartz Center for Computational Neuroscience[1] database. These recordings were converted in ICs using ICA. Training and test sets were prepared differently. The training set comprised 5,937 ICs labelled using a crowd labelling approach whereas the testing set contained 130 ICs labelled by six trained experts. They reported that the weighted DCNN outperformed the other two neural network models. Additionally, they developed a faster version of the weighted DCNN which does not use the autocorrelation function. They named the weighted DCNN as ICLabel and the faster weighted DCNN model as ICLabel$_{Lite}$. Finally, they showed that these

[1] https://sccn.ucsd.edu

two models outperformed previously developed handcrafted feature methods. Curiously, they reported that the ICLabel$_{Lite}$ performed better than the default ICLabel for the binary classification of ICs. ICLabel and ICLabel$_{Lite}$ models are currently implemented in ICLabel toolbox available in EEGLAB [245].

Automatic BSS approaches should be considered whenever a BSS model is used because they allow its usage in real-time ambulatory scenarios. Table 3.1 summarises the automatic BSS methodologies proposed in the referenced scientific articles.

Table 3.1: Summary of the EEG preprocessing models based on automatic BSS.

Authors	Method	Database	Dataset Description	Artefacts
Shoker *et al.* [246]	SOBI-SVM	London King's College Hospital EEG Database	200 EEG segments that last 10 seconds each. Each EEG segment contains 16 channels.	Eye Blinking
LeVan *et al.* [218]	FastICA-Bayesian Network	Personal Database collected at the Montreal Neurological Institute	EEG data from 46 patients. Number of channels varied from 21 to 39 electrodes. The signals were converted in 30-second segments. Each segment was divided in 2-second epochs comprising 15 epochs for each IC.	Every type of artefacts
Nolan *et al.* [235]	ICA-Threshold-based Model	Personal Simulated and Real Databases	The first dataset contains 200 simulated epochs with artefacts in randomly selected channels. The second dataset contains 47 healthy patients. All signals from this second dataset contain 134 channels. Each EEG epoch lasts 2 seconds.	Eye Blinking, Eye Movement, Muscle, Discontinuities, Linear Trends and White Noise
Mognon *et al.* [237]	Infomax ICA-Threshold-based Model	Two Personal Databases	The first dataset contains EEG signals from 16 subjects. Each EEG segment contains 128 channels which were reduced to 63 for performing ICA. Regarding the second dataset, the EEG Segments contain 64 channels. On both datasets, the signals were segmented in 2.5-second epochs.	Eye Blinking, Eye Movement and Generic Descontinuities
Winkler *et al.* [236]	TDSEP-LDA	Personal Database	23 EEG recordings from 12 subjects with 121 channels reduced to 30 using PCA. Each recording lasts 10 minutes.	Every type of artefacts

SOBI: Second-order Blind Identification; SVM: Support Vector Machines; EEG: Electroencephalogram; MEG: Magnetoencephalogram; IC: Independent Component; TDSEP: Temporal Decorrelation source Separation; LDA: Linear Discriminant Analysis; MNR: Multinomial Regression; PCA: Principal Component Analysis; ANN: Artificial Neural Network; DCNN: Deep Convolutional Neural Network wDCNN: Weighted Deep Convolutional Neural Network

Continues on next page

Authors	Method	Database	Dataset Description	Artefacts
Frølich *et al.* [240]	Infomax ICA-MNR	Personal Databases	The first database contains 250-channel EEG signals from 34 subjects lasting between 8 and 88 minutes. The second database contains 64-channel EEG signals from 12 subjects lasting between 56 and 66 minutes. In total, the dataset comprises 8,023 ICs.	Eye Blinking, Cardiac Pulse, Eye Movement, and Muscle
Chaumon *et al.* [238]	Extended Infomax ICA-Threshold-based Model	Personal Databases	Training dataset with 8 EEG sets containing between 64 and 128 channels each and test dataset with 13 EEG sets containing 64 channels each. The EEG recordings last from about 2 minutes to about 7 minutes.	Eye Blinking, Eye Movement, Muscle, Bad Channel, Rare Event and Non-Dipolar Component
Radüntz *et al.* [239]	ICA-LDA	Personal Database	The dataset includes 57 subjects. Each EEG signal contains 25 channels and lasts between 1.5 and 20 minutes.	Eye Blinking, Eye Movement, Cardiac Pulse, Muscle and Impedance
Radüntz *et al.* [241]	ICA-ANN	Personal Database	The dataset includes 57 subjects. Each EEG signal contains 25 channels and lasts between 1.5 and 20 minutes.	Eye Blinking, Eye Movement, Cardiac Pulse, Muscle and Impedance
Tamburro *et al.* [242]	ICA-Nonlinear SVM	Personal Database	The dataset includes 62 subjects. The dataset contains both dry and wet EEG signals. The first ones comprehend 97 channels whereas the second ones contain 128 channels. Each EEG recording lasts from 42 seconds to 250 seconds.	Eye Blinking, Eye Movement, Cardiac Pulse and Muscle
Croce *et al.* [243]	FastICA-DCNN	Personal Databases	The dataset includes EEG and MEG signals. EEG signals contain 128 channels, last between 5 and 22 minutes, and comprise 1,067 ICs. MEG signals contain 153 magnetometers, last between 5 and 15 minutes, and comprise 4,749 ICs.	Eye Movement, Cardiac Pulse and Sensor Malfunction
Pion-Tonachini *et al.* [196]	ICA-wDCNN	Personal Database	The dataset includes 6352 EEG recordings with the number of channels ranging from 32 to 256. Each recording lasts 5 seconds. From these 6352 EEGs, there are only 5937 labelled ICs.	Every type of artefacts

SOBI: Second-order Blind Identification; SVM: Support Vector Machines; EEG: Electroencephalogram; MEG: Magnetoencephalogram; IC: Independent Component; TDSEP: Temporal Decorrelation source Separation; LDA: Linear Discriminant Analysis; MNR: Multinomial Regression; PCA: Principal Component Analysis; ANN: Artificial Neural Network; DCNN: Deep Convolutional Neural Network wDCNN: Weighted Deep Convolutional Neural Network

Continues on next page

Authors	Method	Database	Dataset Description	Artefacts
Lee *et al.* [244]	Infomax ICA-Bayesian Deep Learning	Personal Database	The dataset includes 19-channel EEG data. Data was converted in 12,020 ICs.	Eye Movement and Muscle

SOBI: Second-order Blind Identification; SVM: Support Vector Machines; EEG: Electroencephalogram; MEG: Magnetoencephalogram; IC: Independent Component; TDSEP: Temporal Decorrelation source Separation; LDA: Linear Discriminant Analysis; MNR: Multinomial Regression; PCA: Principal Component Analysis; ANN: Artificial Neural Network; DCNN: Deep Convolutional Neural Network wDCNN: Weighted Deep Convolutional Neural Network

3.1.8 Neural network models

Recently, authors have reported new EEG artefact removal methods, based on deep learning architectures, that aim at solving the drawbacks of the aforementioned methods [197, 247–250]. Ghosh *et al.* [248] and Yang *et al.* [247] developed autoencoders (AEs), based on fully connected layers, to automatically remove ocular artefacts from EEGs. Later on, Leite *et al.* [249], Zhang *et al.* [250] and Sun *et al.* [197] proposed models, based on DCNNs, which are able to extract spatiotemporal features, and, therefore, are more robust than the traditional fully connected neural networks. Leite *et al.* developed a deep convolutional autoencoder (DCAE) for removing eye blink and jaw clenching artefacts previously added to clean EEG signals. Zhang *et al.* [250] proposed a DCNN that gradually increases its width to remove muscle artefacts from EEG signals. They reported that this architecture prevents the occurrence of overfitting. Their model was trained using a publicly available benchmark dataset [251]. Sun *et al.* [197] presented a DCNN, based on residual connections, for removing ocular, muscle and cardiac artefacts from noisy EEGs. These signals were generated from summing clean epileptic EEG segments from the CHB-MIT Scalp EEG Database, with electromyogram (EMG), electrocardiography (ECG) and electrooculogram (EOG) signals from Physionet. To develop these approaches, authors require large datasets and considerable computational time. However, compared to the current state of the art BSS methods, these approaches present some main advantages: minor loss of relevant information; faster signal reconstruction; no need for several channels to get cleaner signals; and fully automatic output.

Table 3.2 summarises the EEG artefact removal methods based on deep learning architectures. In conclusion, researchers are currently exploring the potential of DNNs to eliminate artefacts from EEGs. They report that these approaches can learn the complex patterns and the high-dimensional characteristics of the EEGs, being able to separate these from noisy disturbances. However, despite the high performances obtained using deep learning methods, studies were evaluated using either simulated data or data acquired under controlled environments. Therefore, these studies do not completely simulate artefact removal from realistic long-term EEG acquisitions.

Table 3.2: Summary of the EEG preprocessing models based on neural networks.

Authors	Method	Database	Dataset Description	Artefacts
Yang *et al.* [247]	SAE	"Datasets 1" for BCI Competition IV and Personal Database	7 subjects from the first database and 3 subjects from the second, each one with 200 trails lasting more than 6 seconds. The EEG from the first database contains 59 channels whereas the EEG from the second only contains 32.	Ocular Artefacts
Leite *et al.* [249]	DCAE	Personal Database (Noisy Signals) and DEAP EEG Dataset for Emotion Analysis (Noiseless Signals)	The authors do not report how many subjects are present in the first database. The second database contains 32 subjects. Only 19 channels from the 10-20 International System are used. Each EEG segment lasts 1 minute.	Eye Blinking and Jaw Clenching
Gandhi *et al.* [252]	GAN	Personal Database (Noisy Signals) and US Army Research Laboratory EEG database	The dataset contains 4 patients with noiseless data and 10 patients with noisy signals. All signals contain 64 channels. Each input EEG segment has 1000 samples (about 2 seconds).	Jaw Clenching and Moving, Ocular Artefacts, raising and lowering eyebrows and rotating the head for both sides
Sun *et al.* [197]	1D Residual DCNN	CHB-MIT Database and Physionet Noise Dataset (EMG, ECG and EOG)	The first dataset contains 20 patients. All signals contain 23 channels. Each input EEG segment has 6 minutes and 40 seconds. The dataset was divided into 80% for training and 20% for testing.	Muscle, Cardiac and Ocular Artefacts
Zhang *et al.* [250]	DCNN	EEGDenoiseNet Dataset	5,598 epochs obtained from merging 4,514 EEG epochs with 5,598 EMG epochs	Muscle

SAE: Stacked Autoencoder; BCI: Brain-Computer Interface; EEG: Electroencephalogram; DCNN: Deep Convolutional Neural Network DCAE: Deep Convolutional Autoencoder; GAN: Generative Adversarial Networks; EMG: Electromyogram; ECG: Electrocardiogram; EOG: Electrooculogram

3.2 Seizure prediction

Seizure prediction models are a possible solution to improve drug-resistant epilepsy (DRE) patient's quality of life. Their main goal is to find the transitional brain state between the normal and the seizure state, i.e., to find the preictal period. These models may be integrated into warning devices to send alarms before seizures. With this information, patients can leave an ongoing activity and prevent, for example, accidents. Seizure prediction models might also allow for the suppression of the seizure discharge by attempting brain electrical stimulation or administering timely targeted seizure-suppressive medication.

Figure 3.2 presents the common seizure prediction framework. The seizure pre-

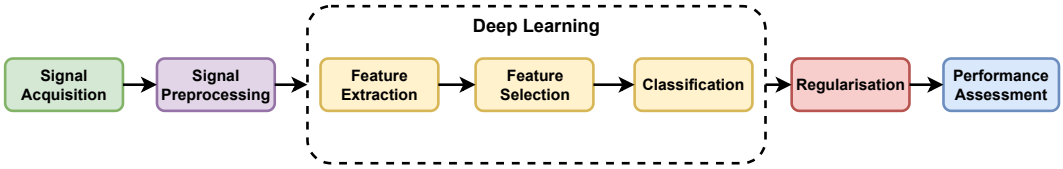


Figure 3.2: Seizure prediction framework.

diction framework comprises signal acquisition, signal preprocessing, feature extraction, feature selection, classification, postprocessing, and performance assessment. The signal acquisition includes retrieving data from databases. Signal preprocessing aims to denoise, segment the data in smaller windows, and define the preictal period for training the predictors. Feature extraction transforms the data into sets of biomarkers that can be used to build predictions. Feature selection picks the biomarkers with high discriminative power and removes the redundant ones. The classification consists in developing predictive models using supervised learning algorithms. It should be noted that when using deep learning architectures, the feature extraction, feature selection, and classification steps are often merged into one. After the classification of the various windows, postprocessing, also known as regularisation, smooths the classifier output and creates a temporal dependency between the various windows. Finally, the prediction is evaluated using different performance metrics.

3.2.1 Epilepsy databases

Data is essential to develop seizure prediction models [253]. Over the years, several databases have been used such as Children's Hospital Boston-Massachusetts Institute of Technology (CHB-MIT) database [254], Universität Bonn database [255], Universität Freiburg database [139], EPILEPSIAE database [13], NeuroVista database [33,125], American Epilepsy Society (AES) database [256], Epilepsy Ecosystem database [257], a database from the Zealand University Hospital (ZUH), the Odense University Hospital (OUH), and the King's College London (KCL) (ZUH-OUH-KCL database) [72], SeizeIT2 database [258], and a database from the St.Vincent's Hospital Melbourne [127]. It is worth noting that the Universität Freiburg database is discontinued as the EPILEPSIAE database superseded it. Table 3.3 summarises the aforementioned databases in seizure prediction containing EEG data.

As presented in Table 3.3, the Universität Bonn's and the Universität Freiburg's databases were two of the first databases containing epileptic EEG. These databases contained intracranial electroencephalography (iEEG) from a small number of patients in presurgical monitoring. Additionaly, the Universität Bonn's database also contained EEGs collected using an external acquisition system (scalp EEG). Afterwards, CHB-MIT was created containing epileptic scalp EEG from 23 pediatric patients under presurgical conditions. CHB-MIT is currently one of the most used databases because it can be freely accessed on the Physionet website.

Later on, EPILEPSIAE was set up. This repository is significantly larger than the remaining ones. It comprises data collected from 275 patients in presurgical monitoring, using different physiological signals including scalp EEG, iEEG, ECG, among others. Data was acquired at Universitätsklinikum Freiburg, Centro Hospitalar e Universitário de Coimbra and Hôpital de la Pitié-Salpêtrière in Paris. Moreover, it is a standard database as it contains long-term annotated recordings (each patient's repository contains on average 165 hours of data making a total of 40,000 hours of signal). 68% of all patients in this database evidence temporal lobe epilepsy (TLE) being in line with its prevalence in the population with focal epilepsy. Each patient contains at least three leading seizures [259], i.e., three seizures separated by 4 hours of interictal data. The patient's metadata includes vigilance stages at seizure onset, seizure type, seizure localization, and drug dosage.

In 2013, Cook *et al.* presented the NeuroVista database. Contrary to the previous databases, this repository includes iEEG from 10 patients collected over a maximum of 2 years. This database covers a large period, providing sufficient information for developing a concept drift study. However, the study would only be a proof of concept because of the small number of subjects.

Afterwards, two more databases appeared: AES database and Epilepsy Ecosystem repository. AES database included iEEG data from five dogs and two humans whereas the Epilepsy Ecosystem repository included iEEG data from three patients from the NeuroVista database. Both databases were used in public seizure prediction competitions proposed by Kaggle. Therefore, these are also known as Kaggle databases.

Recently, three more databases were created: a database from ZUH, OUH, and KCL (ZUH-OUH-KCL database), the SeizeIT2 database, and a database from the St.Vincent's Hospital Melbourne. SeizeIT2 database contains scalp EEG and EEG captured with behind-the-ear electrodes acquired in presurgical conditions. The other two databases contain a new type of EEG. They contain subscalp electroencephalogram (ssEEG) collected over several months inside and outside the hospital. Therefore, they comprehend a new idea of analysing epileptic EEG data which will certainly provide better insights about the seizures. Data from the ZUH-OUH-KCL database were acquired using UNEEG SubQ device [72] whereas data from the St.Vincent's Hospital Melbourne were acquired using EpiMinder Subscalp device [127]. Currently, more data are being collected by KCL [260] in order to increase the ZUH-OUH-KCL database.

Beyond the aforementioned databases, there are other repositories containing other types of biosignals such as NeuroPace database [261], a database from the Children's Hospital Boston [262], and another database from the St.Vincent's Hospital Melbourne [140]. However, as this thesis focuses on developing seizure prediction models using only EEG data, they are not presented in the Table 3.3.

Table 3.3: Summary of epileptic EEG databases found in the literature.

Database (Year)	Epileptic Subjects	EEG Type	Number of Electrodes	Sampling Rate (Hz)	Hours of Signal	Number of Seizures
Universität Bonn [255] (2001)	5	iEEG	128	173.61	35	5
Universität Freiburg [139] (2003)	21	iEEG	128	256	582	88
CHB-MIT [254] (2009)	23	Scalp EEG	18	256	844	163
EPILEPSIAE [13] (2012)	275	Scalp EEG and iEEG	Scalp EEG: At least 19; iEEG: Up to 125	250-2500	More than 40,000	2,662
NeuroVista [125] (2013)	10	iEEG	16	400	About 140,000	2,817
AES [256] (2014)	5 Dogs and 2 Humans	iEEG	16	Dogs: 400; Humans: 5000	NA	NA
Epilepsy Ecossystem [257] (2018)	3	iEEG	16	400	About 30,000	633
ZUH-OUH-KCL [72] (2019)	9	ssEEG	2	207	11,774	338
SeizeIT2 [258] (2020)	54	Scalp EEG and Behind-the-ear EEG	Scalp EEG: 19; Behind-the-ear EEG: 4	250	5,284	182
St.Vincent's Hospital Melbourne [127] (2021)	5	Scalp EEG and ssEEG	Scalp EEG: 23; ssEEG 2	250	Scalp EEG: About 336 per patient; ssEEG: At least 5,760 per patient	55 (2 weeks where both scalp EEG and ssEEG were acquired)

EEG: Electroencephalogram; iEEG: Intracranial Electroencephalogram; ssEEG: Subscalp Electroencephalogram; NA: Not Available

3.2.2 Signal preprocessing

The signal preprocessing aims to prepare EEG data for information extraction, analysis, and development of seizure predictors. It comprehends filtering and artefact removal and data segmentation [137]. Both tasks must be quick and autonomous because the main objective of seizure prediction models is to continuously acquire and process data in real time. Furthermore, researchers also specify the three periods needed to develop seizure predictors: preictal period, seizure occurrence period (SOP), and seizure prediction horizon (SPH). These periods may be also determined while developing the classifiers. However, the definition of the periods must be com-

pletely independent from the out-of-sample results obtained by the developed models. Table 3.4 summarises the signal preprocessing settings used to develop seizure prediction models based on EEG.

Table 3.4: Summary of signal preprocessing procedures found in the literature.

Study (Year)	Database (# Patients, # Seizures, Duration, Signal)	Artefact Removal Methods	Window Length (Seconds)	SOP/SPH (Minutes)
Mormann *et al.* [263] (2005)	Personal (5, 46, 331 hours, iEEG)	Removal of corrupted data	17, 20.5	5, 30, 120, 240 / 0
Le Van Quyen *et al.* [264] (2005)	Personal (5, 52, 305 hours, iEEG)	NA	5	Variable / 0
Mirowski *et al.* [265] (2009)	Freiburg (21, NA, NA, iEEG)	Band-pass filter (0.5-49 Hz and 51-120 Hz)	5	50-120 / 0
Park *et al.* [266] (2011)	Freiburg (18, 80, NA, iEEG)	Removal of corrupted data and band-stop filter (47-53 Hz, 97-103 Hz)	20 with 50% overlap	30 / 0
Cook *et al.* [125] (2013)	NeuroVista (10, 477, 8.4 years, iEEG)	Octave-wave digital filters (8-128 Hz) and notch filters	5	Minutes to hours / 0
Rabbi *et al.* [267] (2013)	EPILEPSIAE (1, 7, 35, iEEG)	Band-pass filter (0.5-100 Hz) and notch filter (60 Hz)	10 with 50% overlap	15, 30, 45 / 0
Rasekhi *et al.* [268] (2013)	EPILEPSIAE (10, 76, 1,388 hours, iEEG+scalp EEG	Notch filter (50 Hz)	5	10, 20, 30, 40 / 0
Alvarado-Rojas *et al.* [141] (2014)	EPILEPSIAE (53, 558, 531 days, iEEG)	Band-pass filter (0.5-8 Hz, 40-140 Hz)	60	10, 30, 60 / 1
Teixeira *et al.* [269] (2014)	EPILEPSIAE (278, 2,702, 2,031 days, iEEG+scalp EEG)	Notch filter (50 Hz)	5	10, 20, 30, 40 / 0.167
Moghim and Corne [270] (2014)	Freiburg (21, NA, 24 days, iEEG)	Visual inspection	5	5 / 1
Parvez *et al.* [271] (2015)	EPILEPSIAE (21, 87, 509 hours, iEEG)	Differential window and phase correlation	10	5 / 0
Rasekhi *et al.* [272] (2015)	EPILEPSIAE (10, 76, 1,388 hours, iEEG+scalp EEG)	Notch filter (50 Hz)	5	10, 20, 30, 40 / 0
Bandarabadi *et al.* [74] (2015)	EPILEPSIAE (24, 159, 3,565 hours, iEEG+scalp EEG)	Notch filter (50 Hz)	5	10, 20, 30, 40 / 0
Assi *et al.* [273] (2015)	AES (5 dogs, 44, NA, iEEG)	Band-pass filter (0.5-180 Hz) and notch filter (50 Hz)	5	60 / 1/12
Myers *et al.* [22] (2016)	CHB-MIT (10, 30, 61 hours, scalp EEG)	Removal of corrupted data, band-pass filters		60 / 2-62

EEG: Electroencephalogram; iEEG: Intracranial Electroencephalogram; ssEEG: Subscalp Electroencephalogram; NA: Not Available; ICA: Independent Component Analysis; EMD: Empirical Mode Decomposition; FFT: Fast Fourier Transform

Continues on next page

Study (Year)	Database (# Patients, # Seizures, Duration, Signal)	Artefact Removal Methods	Window Length (Seconds)	SOP/SPH (Minutes)
Liang *et al.* [274] (2016)	AES (5 dogs+2, 50, 512.58 hours, iEEG)	Band-stop filter (47-53 Hz, 97-103 Hz), low-pass filter (200 Hz), high-pass filter (0.1 Hz)	5	60 / 0
Chu *et al.* [275] (2017)	CHB-MIT (14, 79, 434.3 hours, scalp EEG) Personal (3, 17, 148.4 hours, scalp EEG)	Band-stop filter (57-63 Hz, 117-123 Hz)	20 with 50% overlap	85 / 1/2
Direito *et al.* [276] (2017)	EPILEPSIAE (216, 1800, NA, iEEG+scalp EEG)	Notch filter (50 Hz)	5	10, 20, 30, 40 / 1/6
Khan *et al.* [277] (2017)	Personal (28, 61, NA, scalp EEG) CHB-MIT (22, 68, NA, scalp EEG)	Low-pass filter (128 Hz)	1	10 / 0
Karoly *et al.* [76] (2017)	NeuroVista (9, 1,383, 10.35 years, iEEG)	Band-pass filter (1-140 Hz)	60 with 50% overlap	30 / 1
Kuhlmann *et al.* [257] (2018)	NeuroVista (3, 633, 3.63 years, iEEG)	A different algorithm per team	15-600	60 / 5
Shahbazi *et al.* [278] (2018)	CHB-MIT (14, 49, NA, scalp EEG)	Band-stop filter (57-63 Hz, 117-123 Hz) and notch filter (0 Hz)	10 with 50% overlap (interictal) and 75% (preictal)	30 / 0
Eberlein *et al.* [279] (2018)	AES (4 dogs, NA, NA, iEEG) NeuroVista (3, NA, NA, iEEG)	NA	15	60 / 5
Truong *et al.* [280] (2018)	Freiburg (13, 59, 311.4, iEEG) CHB-MIT (13, 64, 209 hours, scalp EEG) AES (5 dogs+2, 48, 627.7 hours, iEEG)	Band-stop filter (47-53 Hz, 97-103 Hz; 57-63 Hz, 117-123 Hz) and notch filter (0 Hz)	30 with custom overlap (preictal)	30 / 5
Kiral-Kornek *et al.* [33] (2018)	NeuroVista (10, 2,817, 16.29 years, iEEG)	Removal of corrupted data, band-pass filtering (1-140 Hz), notch filter (50 Hz)	30	15 / 1
Sun *et al.* [281] (2018)	AES (5 dogs+2, NA, 1,412 hours, iEEG)	Band-pass filter (0.1 Hz-180 Hz)	30	60 / 5
Abdelhameed *et al.* [282] (2018)	CHB-MIT (12, 56, NA, scalp EEG)	NA	4	60 / 0
Tsiouris *et al.* [283] (2018)	CHB-MIT (22, 185, 980 hours, scalp EEG)	NA	5	15, 30, 60, 120 / 0

EEG: Electroencephalogram; iEEG: Intracranial Electroencephalogram; ssEEG: Subscalp Electroencephalogram; NA: Not Available; ICA: Independent Component Analysis; EMD: Empirical Mode Decomposition; FFT: Fast Fourier Transform

Continues on next page

Study (Year)	Database (# Patients, # Seizures, Duration, Signal)	Artefact Removal Methods	Window Length (Seconds)	SOP/SPH (Minutes)
Truong *et al.* [284] (2019)	Freiburg (13, 59, 311.4 hours, iEEG) CHB-MIT (13, 64, 209 hours, scalp EEG) EPILEPSIAE (30, 261, 2,881.4 hours, scalp EEG)	Band-stop filter (47-53 Hz, 97-103 Hz; 57-63 Hz, 117-123 Hz) and notch filter (0 Hz)	28	30 / 5
Nejedly *et al.* [285] (2019)	AES (4 dogs, 75, 1,021 hours, iEEG)	NA	30 with 50% overlap	55 / 5
Dauod and Bayoumi [286] (2019)	CHB-MIT (8, 43, NA, scalp EEG)	NA	5	60 / 0
Wei *et al.* [219] (2019)	Personal (15, 164, 540 hours, scalp EEG)	Removal of corrupted data, ICA, band-pass filter (1-35 Hz)	10	30 / 0
Zhang *et al.* [287] (2019)	CHB-MIT (23, 182, NA, scalp EEG)	Band-pass filter (5-50 Hz)	5	30 / 0
Xu *et al.* [288] (2020)	CHB-MIT (23, 182, NA, scalp EEG) AES (5 dogs, 44, NA, iEEG)	NA	20 with 25% overlap (preictal)	30 / 5
Islam *et al.* [25] (2020)	Freiburg (5, 68, NA, iEEG)	ICA, band-pass filter (0.5-64 Hz) and notch filter (50 Hz)	5	NA / NA
Das *et al.* [26] (2020)	Personal (103, NA, 154.5 hours, scalp EEG)	Wavelet decomposition, band-pass filter (0.1-70 Hz)	10	NA / NA
Borhade *et al.* [289] (2020)	CHB-MIT (23, NA, NA, scalp EEG)	Preprocessing module capable of separating artefacts from the neural activity	NA	NA / NA
Usman *et al.* [290] (2021)	CHB-MIT (22, 198, 644 hours, scalp EEG)	EMD	29	32 / 0
Chen *et al.* [34] (2021)	AES (4 dogs, 43, 4.42 years, iEEG)	Band-pass filter (0.1-180 Hz)	20	240 / 30
Prathaban *et al.* [28] (2021)	CHB-MIT (22, NA, NA, scalp EEG) Personal (20, NA, NA, scalp EEG) Personal (8, NA, NA, scalp EEG)	Sparsity-based algorithm	1	60-66 / 0
Chen *et al.* [291] (2021)	CHB-MIT (22, NA, NA, scalp EEG)	NA	4 with 50% (preictal)	55 / 5
Cheng *et al.* [292] (2021)	CHB-MIT (4, 13, NA, scalp EEG)	Band-pass filter (0.5-64 Hz)	4 with 50% overlap	30 / 0

EEG: Electroencephalogram; iEEG: Intracranial Electroencephalogram; ssEEG: Subscalp Electroencephalogram; NA: Not Available; ICA: Independent Component Analysis; EMD: Empirical Mode Decomposition; FFT: Fast Fourier Transform

Continues on next page

Study (Year)	Database (# Patients, # Seizures, Duration, Signal)	Artefact Removal Methods	Window Length (Seconds)	SOP/SPH (Minutes)
Dissanayake *et al.* [181] (2021)	CHB-MIT (23, NA, NA, scalp EEG)	NA	10 with 20% overlap (preictal) for dataset 1 and 2 with 40% overlap (preictal) for dataset 2	60 / 0
Dissanayake *et al.* [293] (2021)	CHB-MIT (23, NA, NA, scalp EEG) Personal (15, NA, NA, scalp EEG)	NA	10 with 50% overlap (preictal)	60 / 0
Hussein *et al.* [294] (2021)	CHB-MIT (22, 198, NA, scalp EEG) AES (5 dogs+2, NA, 47.13 days, iEEG) NeuroVista (3, 633, 1,326 days, iEEG)	Removal of corrupted data	30	60 / 0
Wang *et al.* [295] (2021)	CHB-MIT (7, 27, NA, scalp EEG)	NA	5	30 / 5
Zhao *et al.* [296] (2021)	CHB-MIT (10, NA, NA, scalp EEG) AES (5 dogs+2, NA, NA, iEEG) NeuroVista (3, NA, NA, iEEG)	NA	Variable	30 and 60 / 5
Usman *et al.* [27] (2021)	CHB-MIT (22, 198, 644 hours, scalp EEG) AES (5 dogs+2, 198, 479.3 hours, iEEG)	EMD and band-stop filter (50-60 Hz)	30 with 50% overlap	33 (average) / 0
Pal Attia *et al.* [260] (2022)	ZUH-OUH-KCL (6, NA, 409 days, ssEEG)	Band-pass filter (0.5-48 Hz)	60	60 / 5
Viana *et al.* [297] (2022)	ZUH-OUH-KCL (6, NA, 594 days, ssEEG)G	Band-pass filter (0.5-48 Hz), low-pass filter (25 Hz)	60	60 / 5
Müller *et al.* [298] (2022)	Personal (5, 53, 991 hours, iEEG) AES (4 dogs, NA, 1,143 hours, iEEG) NeuroVista (3, NA, 1,159.2 hours, iEEG)	Removal of corrupted data	15	60 / 5
Peng *et al.* [299] (2022)	Freiburg (20, 82, NA, iEEG) CHB-MIT (16, 74, NA, scalp EEG)	Band-stop filter (47-53 Hz, 97-103 Hz; 57-63 Hz, 117-123 Hz)	5	30 / 0

EEG: Electroencephalogram; iEEG: Intracranial Electroencephalogram; ssEEG: Subscalp Electroencephalogram; NA: Not Available; ICA: Independent Component Analysis; EMD: Empirical Mode Decomposition; FFT: Fast Fourier Transform

Continues on next page

Study (Year)	Database (# Patients, # Seizures, Duration, Signal)	Artefact Removal Methods	Window Length (Seconds)	SOP/SPH (Minutes)
Affes *et al.* [300] (2022)	CHB-MIT (23, 198, 1,081 hours, scalp EEG)	Band-stop filter (57-63 Hz, 117-123 Hz), notch filter (0 Hz)	30	30 / 5
Liang *et al.* [301] (2022)	CHB-MIT (13, 64, 209 hours, scalp EEG) AES (5 dogs+2, NA, 1,333.7 hours, iEEG)	Band-stop (57-63 Hz, 117-123 Hz), notch filter (0 Hz)	30 with overlap window selected per patient (preictal)	30 / 0
Uvaydov *et al.* [302] (2022)	EPILEPSIAE (30, NA, NA, iEEG)	NA	4	60 / 0
Jemal *et al.* [303] (2022)	CHB-MIT (23, 163, 940 hours, scalp EEG)	Band-pass filter (0.5-70 Hz) and notch filter (50 Hz)	5	30 / 0
Assali *et al.* [304] (2023)	CHB-MIT (17, 61, NA, scalp EEG)	NA	2	30 and 60 / 0
Li *et al.* [305] (2023)	CHB-MIT (17, 90, 384 hours, scalp EEG) AES (4 dogs, 41, 569.1 hours, iEEG)	Learnable filter based on FFTs	30 with T overlap chosen by patient (preictal)	30 and 34 / 5 and 1
Xu *et al.* [306] (2023)	CHB-MIT (4, 27, NA, scalp EEG)	Band-stop filter (57-63 Hz, 117-123 Hz)	30 with 50% overlap	30 / 0

EEG: Electroencephalogram; iEEG: Intracranial Electroencephalogram; ssEEG: Subscalp Electroencephalogram; NA: Not Available; ICA: Independent Component Analysis; EMD: Empirical Mode Decomposition; FFT: Fast Fourier Transform

3.2.2.1 Filtering and artefact removal

Currently, seizure prediction models are mainly developed using non-invasive EEG data [16]. Although they are easier to capture and generally more accepted by the patients, they present low signal-to-noise ratio (SNR) compared to invasive setups. EEG artefacts may be responsible for the increase of false alarms and, therefore, should be reduced before developing seizure prediction models [19–21].

According to Table 3.4, some researchers remove all the data containing artefacts [22, 33, 219, 263, 266, 294, 298, 307]. However, it may cause loss of information and, therefore, should not be the first considered option. Regarding artefact removal approaches, researchers mainly use digital filters. They focus on using low-pass filters to remove high-frequency noise and high-pass filters to remove low-frequency artefacts. Additionally, notch filters are used to remove the powerline interference. Although there is no consent regarding the cutoff frequencies, the frequencies of interest are generally between 0.5 and around 100 Hz. There are other artefact removal algorithms used to remove artefacts such as ICA [25, 219], EMD [27, 290], wavelet decomposition [26], learnable filter based on fast Fourier transforms (FFTs) [305], differential window and phase correlation [271], preprocessing module capable of separating artefacts from the neural activity [289], and sparsity-based algorithm [28].

3.2.2.2 Data segmentation

The EEG is usually a multi-channel time series lasting several minutes, hours, or even days. Therefore, it must be segmented into smaller windows to be analysed. According to Table 3.4, the window size used in seizure prediction studies usually varies between 1 and 60 seconds. There are also approaches using windows longer than 1 minute [257]. However, large windows are not common and are usually not recommended due to the high nonstationary behaviour of the EEGs [137]. The data segmentation may also be performed using overlapping. Some authors use it to perform data augmentation of preictal samples increasing the number of samples and thus obtaining a dataset with a lower interictal-preictal ratio [181, 280, 288, 291, 293, 301, 305].

3.2.2.3 Preictal period duration, seizure occurrence period, and seizure prediction horizon

The preictal period is a transition stage between the interictal and ictal stages. Although several authors found some evidence of the existence of this stage, there is no clear clinical definition about how and when it begins, how long it lasts and whether it is continuous in time [14]. For designing seizure prediction models, SOP and SPH must be defined. The duration of both periods together is usually equal to the preictal period. According to Table 3.4, different approaches have been followed for performing predictions over the years. Some authors considered a fixed SOP

while others performed a specific search for each patient in order to find the most optimal SOP duration. Recently, researchers mostly considered a fixed SOP lasting either 30 or 60 minutes. Regarding the SPH, it was almost always not considered or lasted at most one minute on seizure prediction studies before 2018. However, that is impractical because the patients would not have time to prepare for an upcoming seizure. Therefore, recently several authors start considering it lasting about 5 minutes.

3.2.3 Computational models

Computational models are typically based on EEG screening as it can record the elctrical brain activity [17, 64]. Over the years several different methodologies have been proposed. These range from simple methods such as threshold-based classifiers to complex methods such as DNNs. Four steps have been common to most of them: feature extraction, feature selection, classification, and the definition of training strategies. Table 3.5 summarises the seizure prediction approaches proposed by several different authors.

Table 3.5: Summary of seizure prediction approaches found in the literature.

Study	Training Strategy	Input Data	Classifier (Postprocessing)	Approach	Performance	Statistical Validation
Mormann *et al.* [263] (2005)	No data partition	Univariate linear features, univariate nonlinear features, bivariate linear features, bivariate nonlinear features	Threshold-based	Prediction	AUC=0.63-0.90	Yes (Surrogate analysis)
Le Van Quyen *et al.* [264] (2005)	No data partition	Phase-locking values from 15 frequency bands	Threshold-based using pre-determined interictal clusters	Prediction	SS=0.70	No
Mirowski *et al.* [265] (2009)	For each patient, the earlier ones were used to train and the last 1-2 were used to test	Linear bivariate features and nonlinear bivariate features (L1-norm)	LR, DCNN and SVM	Prediction	SS=52%, 71% and 38% FPR/h=0	Yes (Surrogate analysis)
Park *et al.* [266] (2011)	Double CV using N-1 folds to train and one fold to test. Training set was further divided using 80/20 ratio to validate the approach before testing	Relative spectral band powers and total power	SVM (Kalman filter)	Prediction	SS=97.5% FPR/h=0.27	Yes (Surrogate analysis)
Cook *et al.* [125] (2013)	First 4 months for training and remaining duration for testing	Average energy, Teager-Kaiser energy and line-length (Backward elimination based on Hilbert-Schmidt Independence Criterion)	kNN - DT (Smoothing)	Forecasting	SS=0.61 TiW=0.23	Yes (Time-matched predictor)

EMD: Empirical mode decomposition; ICA: Independent component analysis; EEG: Electroencephalogram; AUC: Area Under ROC Curve; SS: Sensitivity; SP: Specificity; LR: Logistic Regression; DCNN: Deep Convolutional Neural Network; SVM: Support Vector Machine; CV: Cross-validation; FPR/h: False Positive Rate per hour; kNN: k-Nearest Neighbors; DT: Decition Tree; TiW: Time in Warning; ANFIS: Adaptive Neuro Fuzzy Inference System; ANN: Artificial Neural Network; mRMR: minimum Redundancy Maximum Relevance; ADH: Amplitude Distribution Histogram; GA: Genetic Algorithm; PLV: Phase Locking Value; ALV: Amplitude Locking Value; FFT: Fast Fourier Transform; GLM: Generalised Linear Model; XGBoost: Extreme Gradient Boosting; LSTM: Long Short-term Memory; LDA: Linear Discriminant Analysis; DRNN: Deep Recurrent Neural Network; BiLSTM: Bidirectional Long Short-term Memory; RF: Random Forest; GAN: Generative Adversarial Network; DCAE: Deep Convolutional Autoencoder; SASO: Search Optimisation; GRU: Gated Recurrent Unit; CS: Channel Selection; GNN: Graph Neural Network; GSN: Graph Synthesizing Network; ICA: Independent Component Analysis; SDCN: Semi-dilated Convolutional Network; PCA: Principal Component Analysis; MLP: Multilayer Perceptron; DA: Domain Adaptation

Continues on next page

Study	Training Approach	Input Data (Feature Selection)	Classifier (Postprocessing)	Approach	Performance	Statistical Validation
Rabbi *et al.* [267] (2013)	One seizure for training, one for validation and five for testing	Dynamical similarity index, nonlinear interdependence, and mean phase coherence	ANFIS	Prediction	SS=0.80 FPR/h=0.46	No
Rasekhi *et al.* [268] (2013)	First three seizures used for training and the others for testing for each patient.	Univariate linear features	SVM (Firing power)	Prediction	SS=0.74 FPR/h=0.15	No
Alvarado-Rojas *et al.* [141] (2014)	First 2-4 seizures for training and the others for testing for each patient.	Mean coupling phases between different frequency bands	Threshold-based (Kalman filter)	Prediction	SS=0.68 FPR/h=0.33	Yes (Random predictor)
Teixeira *et al.* [269] (2014)	First 2-3 seizures per patient for training and the others for testing.	Univariate linear features	SVM, ANN (Firing power)	Prediction	SS=0.74 FPR/h=0.28	Kruskal-Wallis between methods
Moghim and Corne [270] (2014)	10 times K-Fold CV for searching hyperparameters and 70/30 holdout validation for evaluation	Univariate linear features and univariate nonlinear features (ReliefF)	SVM	Prediction	S1-score=0.95	Unspecific predictors
Parvez *et al.* [271] (2015)	10-fold CV	Features based on phase correlation	Least squares SVM (k-of-n filter)	Prediction	Accuracy=0.92 False alarms=2.14	No
Rasekhi *et al.* [272] (2015)	First three seizures used for training and the others for testing for each patient	Univariate linear features (mRMR)	SVM (Firing power)	Prediction	SS=0.61 FPR/h=0.11	Random predictor
Bandarabadi *et al.* [74] (2015)	First three seizures used for training and the others for testing for each patient	Univariate linear features (Based on ADH)	SVM (Firing power)	Prediction	SS=0.76 FPR/h=0.10	Random predictor

EMD: Empirical mode decomposition; ICA: Independent component analysis; EEG: Electroencephalogram; AUC: Area Under ROC Curve; SS: Sensitivity; SP: Specificity; LR: Logistic Regression; DCNN: Deep Convolutional Neural Network; SVM: Support Vector Machine; CV: Cross-validation; FPR/h: False Positive Rate per hour; kNN: k-Nearest Neighbors; DT: Decition Tree; TiW: Time in Warning; ANFIS: Adaptive Neuro Fuzzy Inference System; ANN: Artificial Neural Network; mRMR: minimum Redundancy Maximum Relevance; ADH: Amplitude Distribution Histogram; GA: Genetic Algorithm; PLV: Phase Locking Value; ALV: Amplitude Locking Value; FFT: Fast Fourier Transform; GLM: Generalised Linear Model; XGBoost: Extreme Gradient Boosting; LSTM: Long Short-term Memory; LDA: Linear Discriminant Analysis; DRNN: Deep Recurrent Neural Network; BiLSTM: Bidirectional Long Short-term Memory; RF: Random Forest; GAN: Generative Adversarial Network; DCAE: Deep Convolutional Autoencoder; SASO: Search Optimisation; GRU: Gated Recurrent Unit; CS: Channel Selection; GNN: Graph Neural Network; GSN: Graph Synthesizing Network; ICA: Independent Component Analysis; SDCN: Semi-dilated Convolutional Network; PCA: Principal Component Analysis; MLP: Multilayer Perceptron; DA: Domain Adaptation

Continues on next page

Study	Training Approach	Input Data (Feature Selection)	Classifier (Postprocessing)	Approach	Performance	Statistical Validation
Assi *et al.* [273] (2015)	80% segments for training and 20% for testing	Univariate linear features (GA and mRMR)	SVM, ANFIS	Prediction	SS=0.90 SP=0.89	No
Myers *et al.* [22] (2016)	No data partition	Phase locking value (PLV) and amplitude lock value (ALV)	Threshold-based	Prediction	SS=0.67 FPR/h=0.167	Random predictor
Liang *et al.* [274] (2016)	4-fold CV for dogs and 3-fold CV for patients	FFT segments	DCNN-SVM with pretrained weights	Prediction	AUC=0.724	No
Chu *et al.* [275] (2017)	For each patient, 3 seizures were used for training and the remaining ones for testing	Spectral bands ratios	Threshold-based	Prediction	SS=0.87 FPR/h=0.367	Random predictor
Direito *et al.* [276] (2017)	2-3 seizures for training (3-Fold CV) and remaining for testing for each patient	Univariate linear features	SVM (Firing power)	Prediction	SS=0.38 FPR/h=0.20	Random predictor
Khan *et al.* [277] (2017)	Holdout validation and 10-fold CV	Wavelet tensors	DCNN	Prediction	SS=0.88 FPR/h=0.142	Random predictor
Karoly *et al.* [76] (2017)	For each patient, 100 days for training (average 38 seizures) and 0.5-1 year for testing (average 116 seizures). Training was performed using 10-fold CV. Model was updated over the time using circadian profile to deal with concept drifts.	Line length and frequency bands energy (Kullback-Leibler distance)	LR (Bin width of 1h)	Forecasting	SS=0.61 TiW=0.25	Time-matched predictor

EMD: Empirical mode decomposition; ICA: Independent component analysis; EEG: Electroencephalogram; AUC: Area Under ROC Curve; SS: Sensitivity; SP: Specificity; LR: Logistic Regression; DCNN: Deep Convolutional Neural Network; SVM: Support Vector Machine; CV: Cross-validation; FPR/h: False Positive Rate per hour; kNN: k-Nearest Neighbors; DT: Decition Tree; TiW: Time in Warning; ANFIS: Adaptive Neuro Fuzzy Inference System; ANN: Artificial Neural Network; mRMR: minimum Redundancy Maximum Relevance; ADH: Amplitude Distribution Histogram; GA: Genetic Algorithm; PLV: Phase Locking Value; ALV: Amplitude Locking Value; FFT: Fast Fourier Transform; GLM: Generalised Linear Model; XGBoost: Extreme Gradient Boosting; LSTM: Long Short-term Memory; LDA: Linear Discriminant Analysis; DRNN: Deep Recurrent Neural Network; BiLSTM: Bidirectional Long Short-term Memory; RF: Random Forest; GAN: Generative Adversarial Network; DCAE: Deep Convolutional Autoencoder; SASO: Search Optimisation; GRU: Gated Recurrent Unit; CS: Channel Selection; GNN: Graph Neural Network; GSN: Graph Synthesizing Network; ICA: Independent Component Analysis; SDCN: Semi-dilated Convolutional Network; PCA: Principal Component Analysis; MLP: Multilayer Perceptron; DA: Domain Adaptation

Continues on next page

Study	Training Approach	Input Data (Feature Selection)	Classifier (Postprocessing)	Approach	Performance	Statistical Validation
Kuhlmann *et al.* [257] (2018)	Holdout validation	Univariate linear features, univariate nonlinear features, multivariate features	Ensemble algorithm (XGBoost, kNN, GLM, SVM)	Prediction	AUC=0.75	No
Shahbazi *et al.* [278] (2018)	70% of interictal files used for training and 30% for testing. Preictal samples were selected using leave-one-seizure-out CV for each patient.	Spectrograms	DCNN-LSTM (k-of-n filter)	Prediction	SS=0.98 FPR/h=0.13	No
Eberlein *et al.* [279] (2018)	Holdout validation	Time series	DCNN	Prediction	AUC=0.74	No
Truong *et al.* [280] (2018)	Leave-one-seizure-out CV for each patient	Spectrograms	DCNN (k-of-n filter)	Prediction	SS=0.79 FPR/h=0.14	Random predictor
Kiral-Kornek *et al.* [33] (2018)	First 2 months containing at least 1 seizure for training and remaining duration for testing for each patient. Retrained the model after a certain interval to deal with concept drifts.	Spectrograms and time of the day	DCNN (Postprocessing layer included in model)	Prediction	SS=0.69 TiW=0.27	Random predictor

EMD: Empirical mode decomposition; ICA: Independent component analysis; EEG: Electroencephalogram; AUC: Area Under ROC Curve; SS: Sensitivity; SP: Specificity; LR: Logistic Regression; DCNN: Deep Convolutional Neural Network; SVM: Support Vector Machine; CV: Cross-validation; FPR/h: False Positive Rate per hour; kNN: k-Nearest Neighbors; DT: Decition Tree; TiW: Time in Warning; ANFIS: Adaptive Neuro Fuzzy Inference System; ANN: Artificial Neural Network; mRMR: minimum Redundancy Maximum Relevance; ADH: Amplitude Distribution Histogram; GA: Genetic Algorithm; PLV: Phase Locking Value; ALV: Amplitude Locking Value; FFT: Fast Fourier Transform; GLM: Generalised Linear Model; XGBoost: Extreme Gradient Boosting; LSTM: Long Short-term Memory; LDA: Linear Discriminant Analysis; DRNN: Deep Recurrent Neural Network; BiLSTM: Bidirectional Long Short-term Memory; RF: Random Forest; GAN: Generative Adversarial Network; DCAE: Deep Convolutional Autoencoder; SASO: Search Optimisation; GRU: Gated Recurrent Unit; CS: Channel Selection; GNN: Graph Neural Network; GSN: Graph Synthesizing Network; ICA: Independent Component Analysis; SDCN: Semi-dilated Convolutional Network; PCA: Principal Component Analysis; MLP: Multilayer Perceptron; DA: Domain Adaptation

Continues on next page

Study	Training Approach	Input Data (Feature Selection)	Classifier (Postprocessing)	Approach	Performance	Statistical Validation
Sun *et al.* [281] (2018)	Holdout validation	Spectral band features	LDA, LR, DCNN, RNN	Prediction	AUC (LDA)= 0.802 AUC(LR)= 0.787 AUC(DCNN)=0.797 AUC(DRNN)=0.557	No
Abdelhameed *et al.* [282] (2018)	To train the DCAE, the dataset of two patients is split in training (80%) and test (20%). To evaluate the model, they used k-fold CV	Time series	pretrained DCNN-BiLSTM	Prediction	SS=0.946 FPR/h=0.04	No
Tsiouris *et al.* [283] (2018)	For each patient, 10-fold CV	Univariate and multivariate features	LSTM	Prediction	SS=0.99 FPR/h=0.02	No
Truong *et al.* [284] (2019)	Leave-one-seizure-out CV for each patient	Spectrograms	GAN-NN, DCNN (k-of-n filter)	Prediction	AUC (DCNN)=0.81	Hanley-McNeil AUC test
Nejedly *et al.* [285] (2019)	First 2 months used for training, third month for validation and the remaining for testing. They retrained every month to handle concept drifts	Time series	DCNN (Median filter)	Prediction	SS=0.79 TiW=0.18	Monte-Carlo simulations
Dauod and Bayoumi [286] (2019)	Leave-one-seizure-out CV for each patient	Time series	DCAE-BiLSTM-CS	Prediction	SS=0.997 FPR/h=0.004	No

EMD: Empirical mode decomposition; ICA: Independent component analysis; EEG: Electroencephalogram; AUC: Area Under ROC Curve; SS: Sensitivity; SP: Specificity; LR: Logistic Regression; DCNN: Deep Convolutional Neural Network; SVM: Support Vector Machine; CV: Cross-validation; FPR/h: False Positive Rate per hour; kNN: k-Nearest Neighbors; DT: Decition Tree; TiW: Time in Warning; ANFIS: Adaptive Neuro Fuzzy Inference System; ANN: Artificial Neural Network; mRMR: minimum Redundancy Maximum Relevance; ADH: Amplitude Distribution Histogram; GA: Genetic Algorithm; PLV: Phase Locking Value; ALV: Amplitude Locking Value; FFT: Fast Fourier Transform; GLM: Generalised Linear Model; XGBoost: Extreme Gradient Boosting; LSTM: Long Short-term Memory; LDA: Linear Discriminant Analysis; DRNN: Deep Recurrent Neural Network; BiLSTM: Bidirectional Long Short-term Memory; RF: Random Forest; GAN: Generative Adversarial Network; DCAE: Deep Convolutional Autoencoder; SASO: Search Optimisation; GRU: Gated Recurrent Unit; CS: Channel Selection; GNN: Graph Neural Network; GSN: Graph Synthesizing Network; ICA: Independent Component Analysis; SDCN: Semi-dilated Convolutional Network; PCA: Principal Component Analysis; MLP: Multilayer Perceptron; DA: Domain Adaptation

Continues on next page

Study	Training Approach	Input Data (Feature Selection)	Classifier (Postprocessing)	Approach	Performance	Statistical Validation
Wei *et al.* [219] (2019)	10-fold CV for each patient	Time series images	DCNN-LSTM (k-of-n filter)	Prediction	SS=1.00 FPR/h=0.04	No
Zhang *et al.* [287] (2019)	Leave-one-seizure-out for each patient	Common spatial pattern from different spectral bands	DCNN (Kalman filter)	Prediction	SS=0.92 FPR/h=0.14	No
Xu *et al.* [288] (2020)	80% samples for training and 20% samples for testing for each patient	Time series	DCNN	Prediction	SS=0.96 FPR/h=0.07	No
Islam *et al.* [25] (2020)	80% of the segments used for training and 20% for testing	Anomalies ratio	RF	Prediction	SS=0.903 FPR=0.330	No
Das *et al.* [26] (2020)	Holdout validation	Current max value, lower threshold, homogeneity, neighbour connectivity, and power and energy	Sequential algorithm that uses the features step by step	Prediction	F1 Score=0.949	No
Borhade *et al.* [289] (2020)	Holdout and K-fold CV.	Univariate linear and nonlinear features (fuzzy information gain)	Search optimisation-based (SASO-based) DRNN	Prediction	SS=0.965 SP=0.975 K-Fold SS=0.965 SP=0.937	No
Usman *et al.* [290] (2021)	K-fold CV	STFTs	GAN-DCNN-LSTM	Prediction	SS=0.93 SP=0.92	No

EMD: Empirical mode decomposition; ICA: Independent component analysis; EEG: Electroencephalogram; AUC: Area Under ROC Curve; SS: Sensitivity; SP: Specificity; LR: Logistic Regression; DCNN: Deep Convolutional Neural Network; SVM: Support Vector Machine; CV: Cross-validation; FPR/h: False Positive Rate per hour; kNN: k-Nearest Neighbors; DT: Decition Tree; TiW: Time in Warning; ANFIS: Adaptive Neuro Fuzzy Inference System; ANN: Artificial Neural Network; mRMR: minimum Redundancy Maximum Relevance; ADH: Amplitude Distribution Histogram; GA: Genetic Algorithm; PLV: Phase Locking Value; ALV: Amplitude Locking Value; FFT: Fast Fourier Transform; GLM: Generalised Linear Model; XGBoost: Extreme Gradient Boosting; LSTM: Long Short-term Memory; LDA: Linear Discriminant Analysis; DRNN: Deep Recurrent Neural Network; BiLSTM: Bidirectional Long Short-term Memory; RF: Random Forest; GAN: Generative Adversarial Network; DCAE: Deep Convolutional Autoencoder; SASO: Search Optimisation; GRU: Gated Recurrent Unit; CS: Channel Selection; GNN: Graph Neural Network; GSN: Graph Synthesizing Network; ICA: Independent Component Analysis; SDCN: Semi-dilated Convolutional Network; PCA: Principal Component Analysis; MLP: Multilayer Perceptron; DA: Domain Adaptation

Continues on next page

Study	Training Approach	Input Data (Feature Selection)	Classifier (Postprocessing)	Approach	Performance	Statistical Validation
Chen *et al.* [34] (2021)	Initial training performed using 16 interictal segments and 2 preictal segments. Testing using the following 7 days. Retrained every 7 days to deal with concept drifts.	Spectral band features	SVM (SVM outputs histograms)	Prediction	SS=0.84 TiW=0.27	No
Prathaban *et al.* [28] (2021)	Holdout validation and 5-fold CV is used in the training dataset.	2D EEG images	DCNN optimised using Fletcher-Reeves	Prediction	SS=0.993 FPR/h=0.033	Random predictor
Chen *et al.* [291] (2021)	NA	Wavelet decomposition features	Graph-DCNN-GRU	Prediction	SS=0.986 FPR/h=0.011	No
Cheng *et al.* [292] (2021)	Leave-one-seizure-out CV	Wavelet decomposition features	Bi-LSTM (Moving average filter)	Prediction	SS=0.993 FPR/h=0.00	No
Dissanayake *et al.* [181] (2021)	Leave-one-patient-out (transfer learning) and 10-fold CV (patient-specific model).	Mel frequency cepstral coefficients	Multitask DCNN and Siamese DCNN	Prediction	AUC=0.9273 AUC=0.969	No
Dissanayake *et al.* [293] (2021)	Leave-one-patient-out (transfer learning) and 10-fold CV (patient-specific model).	Mel frequency cepstral coefficients	Graph neural network (GNN)-Graph Synthesizing Network (GSN)	Prediction	AUC=0.988 AUC=0.992	No
Hussein *et al.* [294] (2021)	Leave-one-patient-out CV for first dataset and holdout validation for the others	Scalogram	Prediction	Semi-dilated convolution network (SDCN)	AUC=0.97 AUC=0.93 AUC=0.88	No

EMD: Empirical mode decomposition; ICA: Independent component analysis; EEG: Electroencephalogram; AUC: Area Under ROC Curve; SS: Sensitivity; SP: Specificity; LR: Logistic Regression; DCNN: Deep Convolutional Neural Network; SVM: Support Vector Machine; CV: Cross-validation; FPR/h: False Positive Rate per hour; kNN: k-Nearest Neighbors; DT: Decition Tree; TiW: Time in Warning; ANFIS: Adaptive Neuro Fuzzy Inference System; ANN: Artificial Neural Network; mRMR: minimum Redundancy Maximum Relevance; ADH: Amplitude Distribution Histogram; GA: Genetic Algorithm; PLV: Phase Locking Value; ALV: Amplitude Locking Value; FFT: Fast Fourier Transform; GLM: Generalised Linear Model; XGBoost: Extreme Gradient Boosting; LSTM: Long Short-term Memory; LDA: Linear Discriminant Analysis; DRNN: Deep Recurrent Neural Network; BiLSTM: Bidirectional Long Short-term Memory; RF: Random Forest; GAN: Generative Adversarial Network; DCAE: Deep Convolutional Autoencoder; SASO: Search Optimisation; GRU: Gated Recurrent Unit; CS: Channel Selection; GNN: Graph Neural Network; GSN: Graph Synthesizing Network; ICA: Independent Component Analysis; SDCN: Semi-dilated Convolutional Network; PCA: Principal Component Analysis; MLP: Multilayer Perceptron; DA: Domain Adaptation

Continues on next page

Study	Training Approach	Input Data (Feature Selection)	Classifier (Postprocessing)	Approach	Performance	Statistical Validation
Wang *et al.* [295] (2021)	Leave-one-seizure-out CV	Single channel time series	Ensemble DCNN-BiLSTM (merging output from all channels)	Prediction	SS=0.827 SP=0.724	No
Zhao *et al.* [296] (2021)	Holdout validation	Time series	DCNN	Prediction	SS=0.998 FPR/h=0.005 SS=0.935 FPR/h=0.063 SS=0.852 FPR/h=0.116	No
Usman *et al.* [27] (2021)	Holdout validation	Univariate linear features and features obtained from STFTs using DCNNs (Pearson correlation coefficient and Particle swarm optimisation)	Model agnostic meta learning that receives input from SVM, DCNN and LSTM	Prediction	SS=0.963 SP=0.957 SS=0.942 SP=0.958	No
Pal Attia *et al.* [260] (2022)	Leave-one-patient-out CV	Time series, FFT and time of day	BiLSTM (1h smooth)	Forecasting	SS=0.55 TiW=0.34	Surrogate analysis
Viana *et al.* [297] (2022)	Initial 1/3 data for training and remaining of testing for each patient.	Time series, FFT and time of day	BiLSTM (1h smooth)	Forecasting	SS=0.73 TiW=0.34	Time-matched predictor
Müller *et al.* [298] (2022)	Holdout validation.	Univariate and bivariate features. Time series	MLP. DCNN	Prediction	AUC=0.793 AUC=0.68	Hanley-McNeil method

EMD: Empirical mode decomposition; ICA: Independent component analysis; EEG: Electroencephalogram; AUC: Area Under ROC Curve; SS: Sensitivity; SP: Specificity; LR: Logistic Regression; DCNN: Deep Convolutional Neural Network; SVM: Support Vector Machine; CV: Cross-validation; FPR/h: False Positive Rate per hour; kNN: k-Nearest Neighbors; DT: Decition Tree; TiW: Time in Warning; ANFIS: Adaptive Neuro Fuzzy Inference System; ANN: Artificial Neural Network; mRMR: minimum Redundancy Maximum Relevance; ADH: Amplitude Distribution Histogram; GA: Genetic Algorithm; PLV: Phase Locking Value; ALV: Amplitude Locking Value; FFT: Fast Fourier Transform; GLM: Generalised Linear Model; XGBoost: Extreme Gradient Boosting; LSTM: Long Short-term Memory; LDA: Linear Discriminant Analysis; DRNN: Deep Recurrent Neural Network; BiLSTM: Bidirectional Long Short-term Memory; RF: Random Forest; GAN: Generative Adversarial Network; DCAE: Deep Convolutional Autoencoder; SASO: Search Optimisation; GRU: Gated Recurrent Unit; CS: Channel Selection; GNN: Graph Neural Network; GSN: Graph Synthesizing Network; ICA: Independent Component Analysis; SDCN: Semi-dilated Convolutional Network; PCA: Principal Component Analysis; MLP: Multilayer Perceptron; DA: Domain Adaptation

Continues on next page

Study	Training Approach	Input Data (Feature Selection)	Classifier (Postprocessing)	Approach	Performance	Statistical Validation
Peng *et al.* [299] (2022)	Leave-one-patient-out and 1 seizure from the testing patient. The remaining seizures of the patients are used for testing.	STFT	DCNN-DA	Prediction	SS=0.76 FPR/h=0.19 SS=0.73 FPR/h=0.24	No
Affes *et al.* [300] (2022)	For each patient, dataset was divided using 75/25 ratio.	Spectrograms	DCNN-GRU (Random Forest Ranking to select channels)	Prediction	F1-score=0.768	No
Liang *et al.* [301] (2022)	Leave-one-seizure-out for each patient for first dataset and holdout validation.	STFT. PCA and FFT	DCNN (k-of-n filter). Temporal and spectral DCNN	Prediction	SS=0.883 FPR/h=0.04 AUC=0.86	No
Uvaydov *et al.* [302] (2022)	80/10/10 ratio	Single-channel time series	DCNN (Majority voting multiple single-channel classifiers)	Prediction	AUC=0.82	No
Jemal *et al.* [303] (2022)	Holdout validation and then 5-fold CV (patient-specific model). Holdout validation (patient-independent model)	Filter bank common spatial pattern	DCNN	Prediction	SS=0.909 FPR/h=0.041 SS=0.672 FPR/h=0.6	No
Assali *et al.* [304] (2023)	80/20 ratio for each patient	Relative band powers, sample entropy, stability index	DCNN	Prediction	F1-Score=0.922	No

EMD: Empirical mode decomposition; ICA: Independent component analysis; EEG: Electroencephalogram; AUC: Area Under ROC Curve; SS: Sensitivity; SP: Specificity; LR: Logistic Regression; DCNN: Deep Convolutional Neural Network; SVM: Support Vector Machine; CV: Cross-validation; FPR/h: False Positive Rate per hour; kNN: k-Nearest Neighbors; DT: Decition Tree; TiW: Time in Warning; ANFIS: Adaptive Neuro Fuzzy Inference System; ANN: Artificial Neural Network; mRMR: minimum Redundancy Maximum Relevance; ADH: Amplitude Distribution Histogram; GA: Genetic Algorithm; PLV: Phase Locking Value; ALV: Amplitude Locking Value; FFT: Fast Fourier Transform; GLM: Generalised Linear Model; XGBoost: Extreme Gradient Boosting; LSTM: Long Short-term Memory; LDA: Linear Discriminant Analysis; DRNN: Deep Recurrent Neural Network; BiLSTM: Bidirectional Long Short-term Memory; RF: Random Forest; GAN: Generative Adversarial Network; DCAE: Deep Convolutional Autoencoder; SASO: Search Optimisation; GRU: Gated Recurrent Unit; CS: Channel Selection; GNN: Graph Neural Network; GSN: Graph Synthesizing Network; ICA: Independent Component Analysis; SDCN: Semi-dilated Convolutional Network; PCA: Principal Component Analysis; MLP: Multilayer Perceptron; DA: Domain Adaptation

Continues on next page

Study	Training Approach	Input Data (Feature Selection)	Classifier (Postprocessing)	Approach	Performance	Statistical Validation
Li *et al.* [305] (2023)	Leave-one-seizure-out CV	Time series	MLP (Moving average)	Prediction	SS=0.966 FPR/h=0.060 SS=0.929 FPR/h=0.025	Random predictor
Xu *et al.* [306] (2023)	Leave-one-seizure-out CV for each patient	Time series	deep residual shrinkage network - GRU (DRSN-GRU) (k-of-n filter)	Prediction	SS=0.905 FPR/h=0.11	No

EMD: Empirical mode decomposition; ICA: Independent component analysis; EEG: Electroencephalogram; AUC: Area Under ROC Curve; SS: Sensitivity; SP: Specificity; LR: Logistic Regression; DCNN: Deep Convolutional Neural Network; SVM: Support Vector Machine; CV: Cross-validation; FPR/h: False Positive Rate per hour; kNN: k-Nearest Neighbors; DT: Decition Tree; TiW: Time in Warning; ANFIS: Adaptive Neuro Fuzzy Inference System; ANN: Artificial Neural Network; mRMR: minimum Redundancy Maximum Relevance; ADH: Amplitude Distribution Histogram; GA: Genetic Algorithm; PLV: Phase Locking Value; ALV: Amplitude Locking Value; FFT: Fast Fourier Transform; GLM: Generalised Linear Model; XGBoost: Extreme Gradient Boosting; LSTM: Long Short-term Memory; LDA: Linear Discriminant Analysis; DRNN: Deep Recurrent Neural Network; BiLSTM: Bidirectional Long Short-term Memory; RF: Random Forest; GAN: Generative Adversarial Network; DCAE: Deep Convolutional Autoencoder; SASO: Search Optimisation; GRU: Gated Recurrent Unit; CS: Channel Selection; GNN: Graph Neural Network; GSN: Graph Synthesizing Network; ICA: Independent Component Analysis; SDCN: Semi-dilated Convolutional Network; PCA: Principal Component Analysis; MLP: Multilayer Perceptron; DA: Domain Adaptation; PSO: Particle Swarm Optimisation; PCC: Pearson Correlation Coefficient

3.2.3.1 Feature extraction

Feature extraction converts the EEG time series into tabular values aiming to capture valuable information. This step is one of the most important ones for model development. In the case of extracting features with great separability between different classes, the classification step would be very simple without requiring any complex approach. Specifically, in seizure prediction, the main goal is to obtain information to find the pre-seizure state such as a change of the relative spectral power bands [74].

Features could be extracted from a single channel (univariate) or between two or more channels (multivariate). Beyond that, those could be categorised as linear or nonlinear according to the formula used to compute them. Figure 3.3 presents some examples of features according to each category. Although EEG exhibits a non-

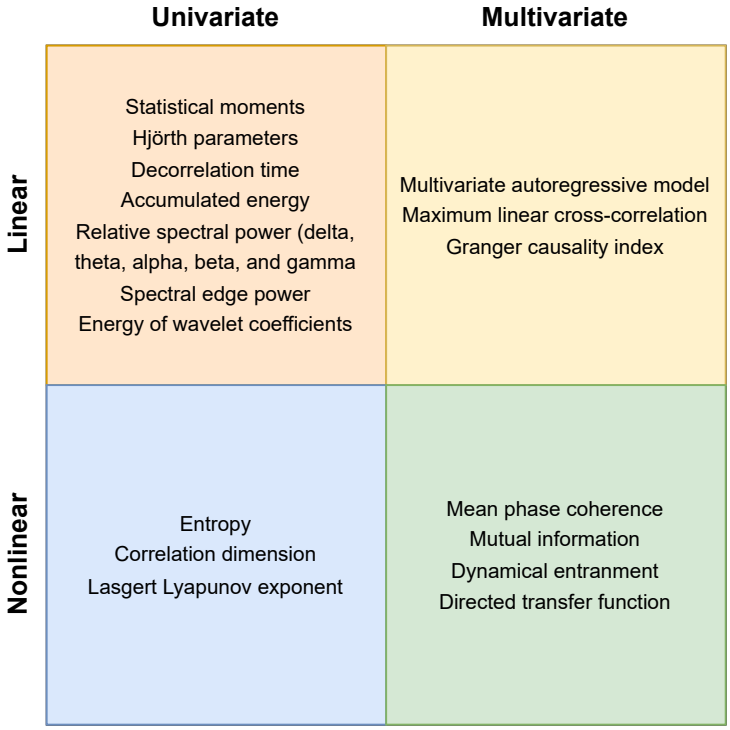


Figure 3.3: Common feature categorisation in seizure prediction studies with some examples for each group.

linear behaviour, researchers mostly use univariate linear features as a consequence of their quick computation making them possible to be used in real-time scenarios. Furthermore, in the comparison studies that were carried out, it was not possible to conclude that the more complex features lead to better results [137, 263, 276]

In deep learning approaches, feature extraction is usually inside the model architecture. Therefore, input data are EEG time series [279, 282] or other type of multidimensional data obtained from the EEGs, e.g., spectrograms [278, 280, 284]. During training, the DNNs optimise the feature extraction layers according to the outputs to extract the most optimal features for the task being performed. Although

this data-driven approach allows extracting the most appropriate features for each dataset, it has the major disadvantage of not giving any information about what each of the obtained features means, which may be a problem for further explanation. It is worth noting that some authors maintain the feature extraction separated from the deep learning architecture, only using the DNNs to classify the data [283].

3.2.3.2 Feature selection

During feature extraction, hundreds or even thousands of features may be extracted from each EEG window. Some features may contain redundant information or confound the classification [137]. Furthermore, the high number of values per sample could lead to overfitting since there will always exist a hyperspace where it is possible to separate training samples from both classes also known as the curse of dimensionality [308]. Therefore, researchers must use feature selection methods to select only the ones that improve the validation performance reducing the multidimensional space.

Feature selection may be categorised in filters, wrappers, or embedded methods. Filters select the most optimal features based on their intrinsic properties. ReliefF [270], minimum redudancy maximum relevance (mRMR) [272,273], and Pearson correlation coefficient (PCC) [27] are examples of filters. Wrappers consist in using the classifier performance to select the best feature set. Compared to filters, wrappers comprise the advantage of selecting features based not only on the relations between them but also with the labels. Additionally, they are based on the performance of a classifier that may consider several features simultaneously as input. However, they are slower since they require more computational power. Backward elimination [125], genetic algorithm (GA) [273], and particle swarm optimisation (PSO) [27] are examples of wrappers. Embedded methods are inside the optimisation function of the classifier and, therefore, assign different weights to the features according to their generalisation capacity. Compared to the previous methods, these comprehend the advantage of assigning different importance and possibly maintaining all of them but with higher weights to the most important ones. Furthermore, the weights are optimised simultaneously as the decision function, obtaining the best weights for the considered labels. L1-normalisation or LASSO [265] is an example of embedded methods.

In the case of deep learning models, although generally no specific method is used to select the most optimal features, one may consider that the architecture can extract and select those that best fit the task being performed. Therefore, it may also be considered an embedded method.

3.2.3.3 Classification

The main objetive of classifiers in seizure prediction is to distinguish between interictal and preictal samples. Over the years several algorithms have been proposed. The simplest ones are threshold-based classifiers. These are linear classifiers that label samples using a threshold on a single biomarker (feature) [141, 263, 264]. Machine learning algorithms such as logistic regression (LR) [76, 281], support vector machines (SVMs) [74, 266, 268], and ensemble methods [25, 257] label the samples based on multidimensional spaces. That is an advantage compared to threshold-based ones since it allows the creation of linear and nonlinear relations between the different features to perform the classification task. Recently, researchers have been using deep learning architectures such as DCNN [33, 280] and long short-term memory (LSTM) [282, 286]. As stated before, these methods are fully data-driven. Therefore, they can optimise the entire pipeline from feature extraction to classification, allowing the creation of a model fully specific to each dataset under analysis.

3.2.3.4 Postprocessing

Postprocessing, also known as output regularisation, is an essential step for classifying samples coming from time series. The main goal is to create a temporal relationship between the several predictions and attenuate the number of false positives [137]. The most used postprocessing methods are the Kalman and the moving average filters. Although the Kalman filter was one of the first postprocessing methods to be used [141, 266], it requires a more complex calibration and does not have a clear advantage over moving average methods [150]. Therefore, the authors have preferred to follow simple approaches by using techniques as smoothing [125], firing power [74, 269, 276], or k-of-n filter [271, 278].

3.2.3.5 Training approach

The development of seizure prediction models generally requires the creation of three sets of data: the training set, the validation set and the test set. These sets may be created using holdout validation [27, 296], cross-validation (CV) [181, 219] or a mixture of both [303]. In this way, it is possible to train and validate the models without compromising their evaluation. Holdout validation splits the dataset into a training and a test set using a ratio defined in advance. The model is developed using the training set, and the performance is obtained using the test set. In CV, the dataset is divided into several sets and each set is used to evaluate the performance at least once. For example, in the case of a k-fold CV, the dataset is divided into k equal groups of which k-1 are used for training and one for testing and is conducted until all groups have been tested at least once. It should be pointed out that if k equals the number of samples in the dataset it is called leave-one-out CV. Although both data partition techniques are widely used in traditional machine

learning, they should not be straightly used in seizure prediction as they may give a wrong notion of the model's performance. The partition should be by seizure and not by sample [280, 295]. The separation may occur randomly, not taking into account the chronology of the seizures or chronologically, similar to what would happen in a real scenario. That strategy allows the researchers to assess the model's ability to predict future events [276].

One of the major problems in the development of seizure prediction models is class imbalance. The number of interictal samples is usually several times larger than the number of preictal samples. Therefore, it is necessary to use methods to attenuate the ratio between interictal and preictal samples such as balancing the data by removing interictal samples [269, 276], increasing the number of preictal samples using sample generation techniques [284, 287], or giving weights to the different classes according to *a priori* probability [266].

Another problem in seizure prediction is the data distribution's high nonstationarity, a phenomenon known as concept drift. Concept drifts usually appear due to changes in medication, vigilance states, sleep deprivation, new seizures that alter brain dynamics, or stress [57, 77, 142, 143, 342–344]. Concept drifts require a different approach for training computational models. Several authors proposed solutions based on simply periodically refitting the models [33, 34, 142, 143, 285]. Kiral-Kornek *et al.* [33], Nejedly *et al.* [285], and Chen *et al.* [34] used EEG data collected over several months. They retrained their models every month and eliminated past data after a certain amount of months, enabling the models to re-adapt over time. Pinto *et al.* [142, 143] used EEG collected under pre-surgical conditions. Therefore, they did not use data from several months, but only from a few days. They retrained their models after testing on a new seizure. Although those studies tried to deal with concept drifts, only Nejedly *et al.* [285] verified whether there was an improvement in the prediction performance.

A great advantage of deep learning architectures over classical machine learning algorithms is the possibility of transferring the weights from one model to another and re-adapting them to a new dataset known as transfer learning. The researchers may train models without starting from scratch, e.g., developing patient-specific models using layers from a DNN already trained with EEGs from other patients. Thus, the training becomes easier because the model's weights do not start at a random point and the model is able to use not only information about the training data but also from other datasets which may lead to a better performance [274, 282, 284, 286].

3.2.3.6 Performance assessment

Seizure sensitivity (SS), false prediction rate per hour (FPR/h) and statistical validation are the basis for evaluating the performance of seizure prediction mod-

els [22, 74, 139, 141, 265, 266, 269, 272, 275–277, 280, 305]. However, not all researchers evaluate their approaches using these metrics. Some prefer to use other metrics such as area under the receiver operating characteristic curve (AUC), sample sensitivity, or sample specificity [26, 257, 271, 289]. Although these assess the classifier performance under analysis, they may not be accurately used to evaluate the approach as a whole where what matters is to understand whether the seizure has been successfully predicted in a certain period. In recent years, a new approach appeared in the seizure prediction field, denominated seizure forecasting. To evaluate seizure forecasting models the researchers replaced the FPR/h by the time in warning (TiW) which consists in evaluating how long the seizure is considered imminent [76, 125, 260, 297].

Several seizure prediction approaches have been proposed. However, it is difficult to compare all of them because not only authors use different evaluation metrics but also different databases. CHB-MIT appears to be the one with which the authors obtained the highest results [278, 283]. On the other side, EPILEPSIAE is a database that shows quite different results from study to study [74, 141, 269, 276]. It may be due to the large number of patients in this database; therefore, using certain patients may lead to better results. However, the aforementioned databases result from the acquisition of signals under pre-surgical conditions, and therefore, the models developed using them may not be evaluated similarly to a real scenario. Therefore, more databases containing ultra-long-term data, such as NeuroVista [76, 125] and ZUH-OUH-KCL [260, 297], are necessary to properly prepare methodologies for a future prediction device.

3.3 Summary

EEG artefact removal has been performed using different approaches such as simple digital filtering, linear regression, adavanced filtering, source decomposition algorithms, BSS algorithms and neural networks. Although BSS algorithms and neural networks are the current state of the art on EEG artefact removal, these algorithms are usually evaluated using data acquired in controlled environments. Therefore, researchers should evaluate the performance of these algorithms on long-term data without any control of the actions performed by the subjects. In addition, researchers should continue prioritising the development of fast and automatic models as only these can be used in real-time scenarios.

The majority of epilepsy databases contain data collected under presurgical conditions. It is a serious limitation for developing seizure prediction models which could be used in real-time conditions because the data does not represent all variations which happen over the months. Furthermore, the patients under presurgical conditions have their medication reduced, being more likely to evidence seizures. Therefore, research developed using presurgical databases should be used to identify

neural mechanisms that lead to the beginning of seizures and to develop proof of concept seizure prediction models. Afterwards, to put acquired knowledge into devices designed for daily life operation, more databases such as NeuroVista repository and ZUH-OUH-KCL database should be created. In this way, the models could be tested in data acquired over several months to evaluate whether they can handle different concept drifts present in the data.

Several studies have been performed in the seizure prediction research field. First, researchers developed threshold-based models that fired alarms whenever a certain feature crossed a predefined threshold. Afterwards, more robust models were created. These were based on machine learning algorithms that analyse data using linear and nonlinear rules developed considering multiple features. Recently, seizure prediction methods based on deep learning approaches have been developed. Contrary to traditional machine learning algorithms, these methods may extract features directly from the data without relying on signal processing techniques to create them. Also, the automatic data processing makes the models easier to be retrained without requiring an expert. This advantage may be used for handling concept drifts over time. Additionally, these methods can easily use the weights from other already trained architectures. It is a great advantage compared to traditional machine learning algorithms because researchers do not always have to train their models from the scratch and may use knowledge from other databases without being concerned with ethical procedures regarding sharing data.

The development of seizure prediction models using different databases and the lack of gold standard metrics make comparing different approaches more difficult. Therefore, researchers must evaluate their approaches with the most used databases and use gold standard metrics for an easier comparison among approaches. Recently researchers have shifted to a new way of performing seizure prediction called seizure forecasting. Unlike traditional seizure prediction, which considers that the brain goes into a point-of-no-return when it enters the preictal period, seizure forecasting considers that the brain of a person with epilepsy comprehends different seizure susceptibility over time. However, since the output is probabilistic, it may not be clear for the patient when an upcoming seizure happens, i.e., the system informs the patient that there is a certain probability of having a seizure in the future but does not provide the exact time. Furthermore, using different metrics makes comparing the studies with traditional seizure prediction approaches difficult.

Chapter 4

EPIC: Annotated epileptic EEG independent components for artefact reduction

This chapter presents the methods performed to build the dataset used for developing the electroencephalogram (EEG) artefact removal models. The content of this chapter is based on the journal article published in *Scientific Data* [37]. Section 4.1 presents a brief context of this study. Section 4.2 describes the materials and methodology followed to obtain the dataset. Section 4.3 explains how the data records were stored and presents the link where the code used to develop the study is available. Section 4.4 provides some final reflections about the study.

4.1 Study context

Developing classifiers to automatically label independent components (ICs) requires annotated datasets. After a thorough search of the relevant literature it was found that only Winkler *et al.* [236] and Pion-Tonachini *et al.* [196] made their datasets publicly available. Winkler *et al.* provided a training set with 690 ICs collected from 23 recordings with 10 minutes of EEG and a test set with 1080 ICs collected from 36 EEG recordings. Data were collected from 12 subjects who had to perform provided tasks, for approximately 5 hours, avoiding producing artefacts. Experts labelled both sets using brain and noise classes (binary classification). However, they only released sets with the best six features instead of the detailed information about the ICs. Pion-Tonachini *et al.* provided a dataset containing ICs collected from 5-second EEG recordings acquired from several studies performed in controlled environments over the past 15 years. The dataset is divided into training and test sets. The training set comprises 5,937 ICs classified by several collaborators through a crown labelling task. The test set includes 130 ICs labelled by seven experts. ICs were

annotated with seven classes: brain and six types of different artefacts. Each sample comprises a scalp topographic map, power spectrum density (PSD), autocorrelation function, equivalent current dipole fits, and hand-crafted features. Although this dataset provided scalp topographic maps and PSDs, it fails at providing IC time series. Therefore, new researchers that would like to use the data are restricted to temporal features contained in the dataset which might limit their approaches.

As a consequence of the aforementioned data restrictions, in this chapter, a dataset with ICs from 19-channel EEG data collected from patients with epilepsy (EPIC dataset) available in the EPILEPSIAE database, was created. Despite only containing data from patients with epilepsy, the EPILEPSIAE database comprises data collected over several days from patients under pre-surgical monitoring. Therefore, data contain several artefacts captured while doing day-to-day activities such as conversation, eating, and sleeping. Every sample in the EPIC dataset includes the time series, PSD, and topographic map of each IC. Furthermore, these samples were classified as brain or noise by two experts.

4.2 Materials and methods

Data from the European Epilepsy Database, also known as the EPILEPSIAE database (www.epilepsy-database.eu) and developed by the FP7 EPILEPSIAE project (www.epilepsiae.eu), were used. EPILEPSIAE database comprises long-term continuous EEG and electrocardiography (ECG) signals simultaneously acquired from 275 patients with drug-resistant epilepsy (DRE) under pre-surgical monitoring over several days [13]. It also contains electromyogram (EMG) and electrooculogram (EOG) signals. However, these are not always available. Data were acquired at Universitätsklinikum Freiburg (Germany), Centro Hospitalar e Universitário de Coimbra (Portugal), and Hôpital de la Pitié-Salpêtrière, Paris (France). Regarding metadata, the database contains the seizure characteristics such as clinical and EEG onsets and offsets, classification, pattern, and vigilance state. It also contains the disease etiology and medication. It is worth noting that over the following chapters, the seizure onset refers to the EEG onset since the models are based on the EEGs. The use of these data for research purposes has been authorised by the Ethical Committee of the three hospitals involved in the EPILEPSIAE database development (Ethik-Kommission der Albert-Ludwigs-Universität, Freiburg; Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé, Pitié- Salpêtrière University Hospital; and Comité de Ética do Centro Hospitalar e Universitário de Coimbra). All studies were performed following the relevant guidelines and regulations. Informed written consent was obtained from the patients and the parents or legal guardians of patients under 18.

The analysed dataset consists of 25 patients (13 males, aged 40±17 years) from the EPILEPSIAE database. Data were curated in the context of epileptic seizure

prediction. To develop the seizure prediction models, it was considered data ranging from 4.5 hours before the beginning of the leading seizure [259] until its onset. This selection was performed considering that EEG data within this interval contain both normal and pre-seizure brain states [142, 269]. Data were collected over several days (accounting for 684 hours of EEG). Typical activities were captured in the signals such as conversation, eating, drinking, washing, and sleeping. Therefore, it may contain several experimental errors such as poor electrode connection and adhesion issues, which must be minimised before performing independent component analysis (ICA). The next section presents the algorithm used to remove experimental errors. Subsequently, in Section 4.2.2, it is described how the manual labelling of the ICs was performed.

4.2.1 Removal of experimental errors

An algorithm to identify and remove data corrupted by experimental errors was developed. Figure 4.1 provides a diagram explaining it. The methods used in the algorithm are ordered from the simplest to the most complex. Therefore, it consists of frequency filtering, identification and removal of flatlines and constant saturated portions in all channels, identification and removal of abnormal peaks, EEG segmentation, removal of noisy EEG segments, removal of electrode pops, interpolation of noisy EEG channels, preparation for ICA, and ICA processing.

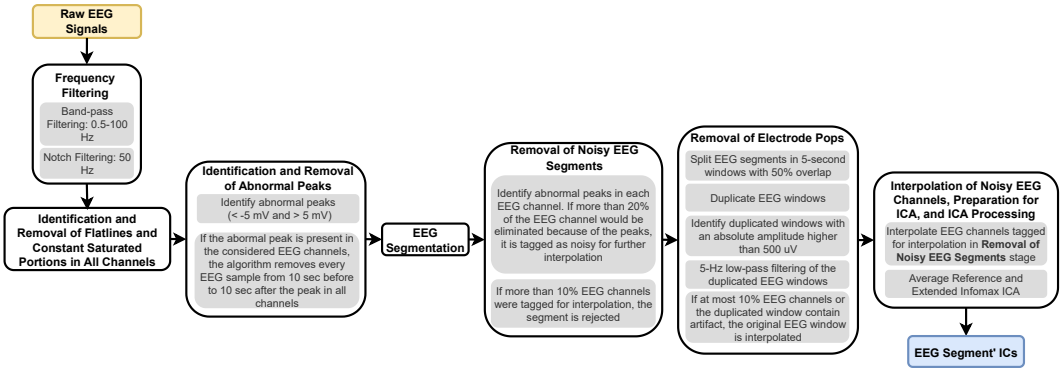


Figure 4.1: Framework followed to obtain the independent components for each EEG segment. It covers the following steps: extraction of signals from EPILEPSIAE database; identification and removal of flatlines, constant saturated portions, and abnormal peaks; EEG segmentation; removal of noisy EEG segments; removal of electrode pops; and interpolation of noisy EEG channels and preparation for ICA.

4.2.1.1 Frequency filtering

The algorithm filtered the data using a 0.5-100 Hz bandpass 4th-order Butterworth filter and a 50 Hz 2nd-order notch filter with the purpose of removing the direct current (DC) component, high-frequency noise, and the powerline interference.

4.2.1.2 Identification and removal of flatlines and constant saturated portions in all channels

Since the data were collected over several days including day-to-day activities, these contained several experimental errors. The algorithm identified and removed every portion of the signal which contained isoelectric flatlines (see Figure 4.2) or constant saturated segments (see Figure 4.3) as well as the 10 seconds of data before and after these errors. These errors were removed for all channels simultaneously.

4.2.1.3 Identification and removal of abnormal peaks

After removing flat and constant segments, the algorithm identified portions of the signal below -5 mV and above 5 mV, named abnormal peaks for easier comprehension (see Figure 4.4). Then, the algorithm verified whether the peaks happened simultaneously in the channels Fp1, Fp2, O1, O2, T5, T6, and Cz. These electrodes were selected according to their geometrical positions, which means that if an abrupt movement affected the system, all of them should capture it. If abnormal peaks appeared in all the aforementioned electrodes, every sample, from 10 seconds before the beginning of the peak until 10 seconds after the peak, was removed. It is worth noting that, to keep the signals' temporal coherence, the data were not concatenated after removing artefacts (see Figure 4.5).

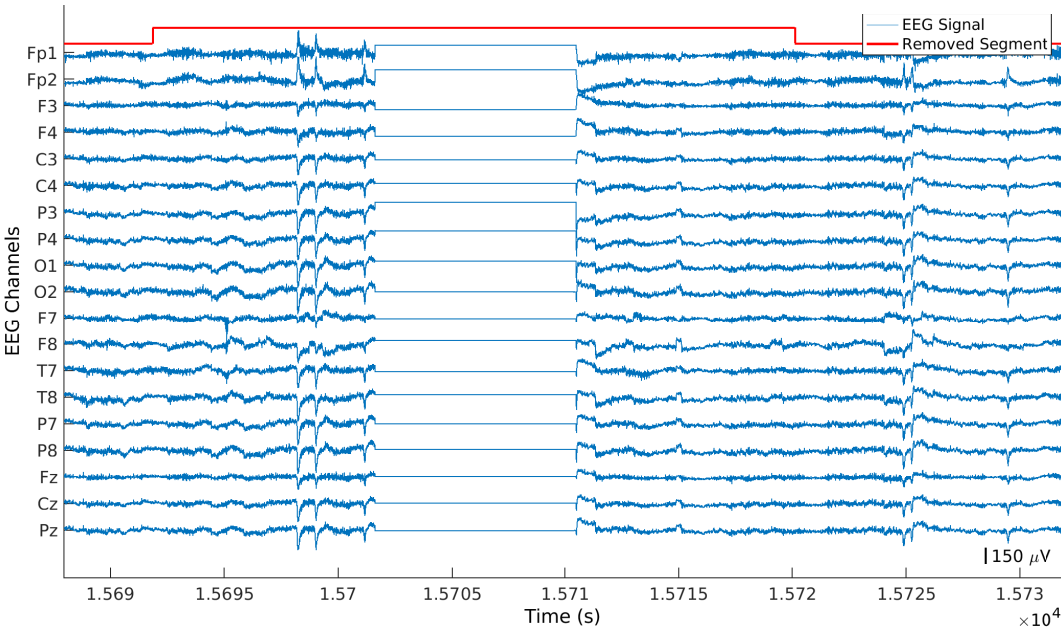


Figure 4.2: Example of an EEG signal with a flat segment. The selected portion was removed over all channels.

4.2.1.4 EEG segmentation

The algorithm divided the remaining data into 10-minute segments (see Figure 4.6). The segmentation in 10-minute portions was performed to prepare the data for the

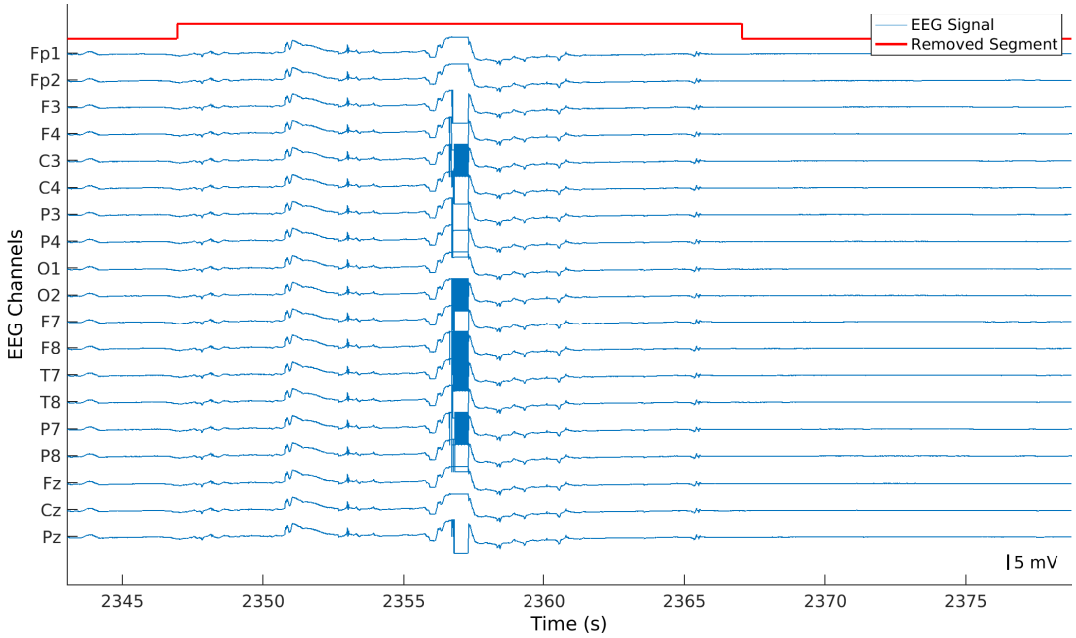


Figure 4.3: Example of an EEG signal containing a saturated segment. The selected portion was removed over all channels.

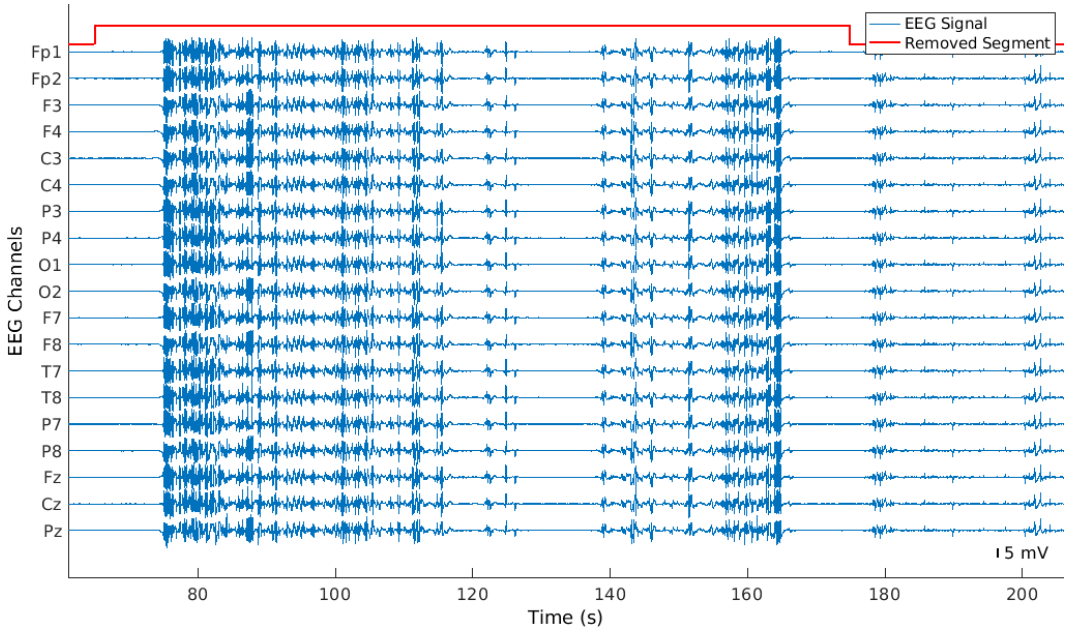


Figure 4.4: Example of an EEG signal with several abnormal peaks. As these artefacts appeared in Fp1, Fp2, T7, T8, O1, O2, and Cz, the selected EEG portion was removed across all channels.

ICA [241, 243]. As the signals were not concatenated after removing errors, there might be segments lasting less than 10 minutes. Despite lasting less than 10 minutes, these segments were kept to obtain the largest possible dataset. The algorithm removed every segment lasting less than 10 seconds as these did not comprise enough data to be properly processed by ICA.

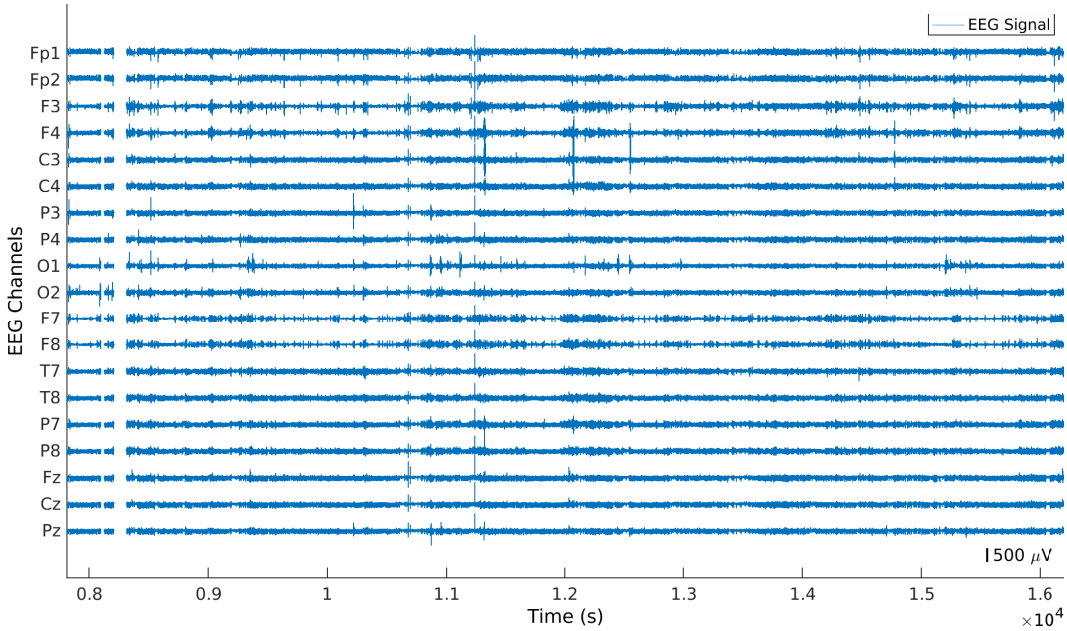


Figure 4.5: Example of a preprocessed EEG signal. The EEG segments were not concatenated after removing noisy data.

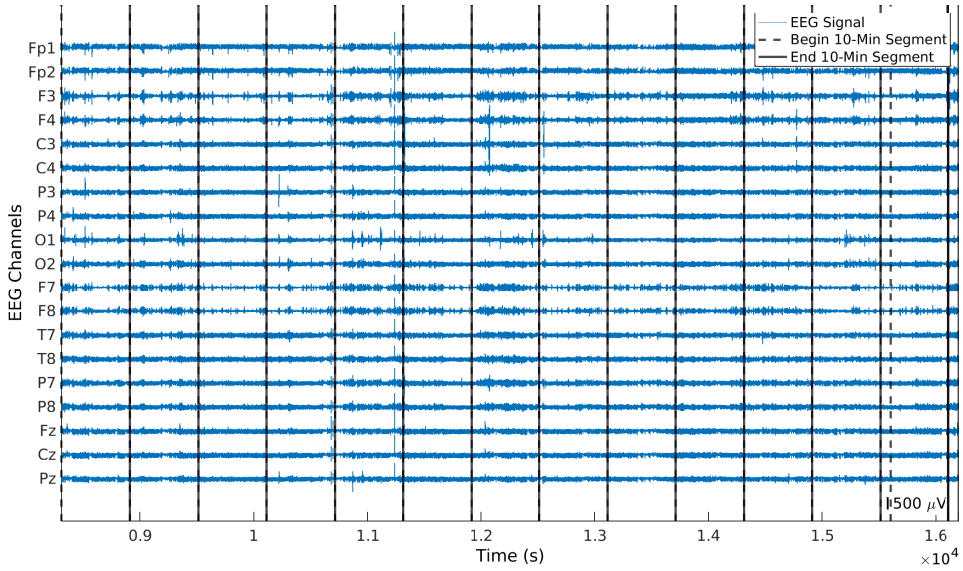


Figure 4.6: Example of a preprocessed EEG signal divided into 10-minute segments. Dashed and continuous lines determine the beginning and the end of the 10-minute segments, respectively. The last segment contains samples from the previous one because the duration of the subsignal is not divisible by 10.

4.2.1.5 Removal of noisy EEG segments

After the EEG segmentation, the algorithm could still identify abnormal peaks in the EEG channels. If more than 20% of the EEG channel would be removed due to abnormal peaks, then the channel was marked to be interpolated. The segment is rejected if more than two EEG channels were marked for interpolation (>10% of all EEG channels).

4.2.1.6 Removal of electrode pops

At this point, the algorithm divided the segments into 5-second windows with a 50% overlap. After that, the algorithm filtered the windows using a 5 Hz low-pass filter. It is worth noting that to not alter the values of the windows when filtering, the original values were saved before performing the analysis. Finally, the algorithm identified filtered windows with an amplitude higher than 0.5 mV to find electrode pops that were not previously removed (see Figure 4.7). The original window was interpolated in the case that at most 10% of channels (two channels) contained electrode pops. Otherwise, it was maintained. It is worth noting that this method was not considered at the beginning of the preprocessing methodology. However, after analysing some ICs, it was noticed that some segments contained one IC only for the electrode pops. Therefore, this step was added to reduce those outputs.

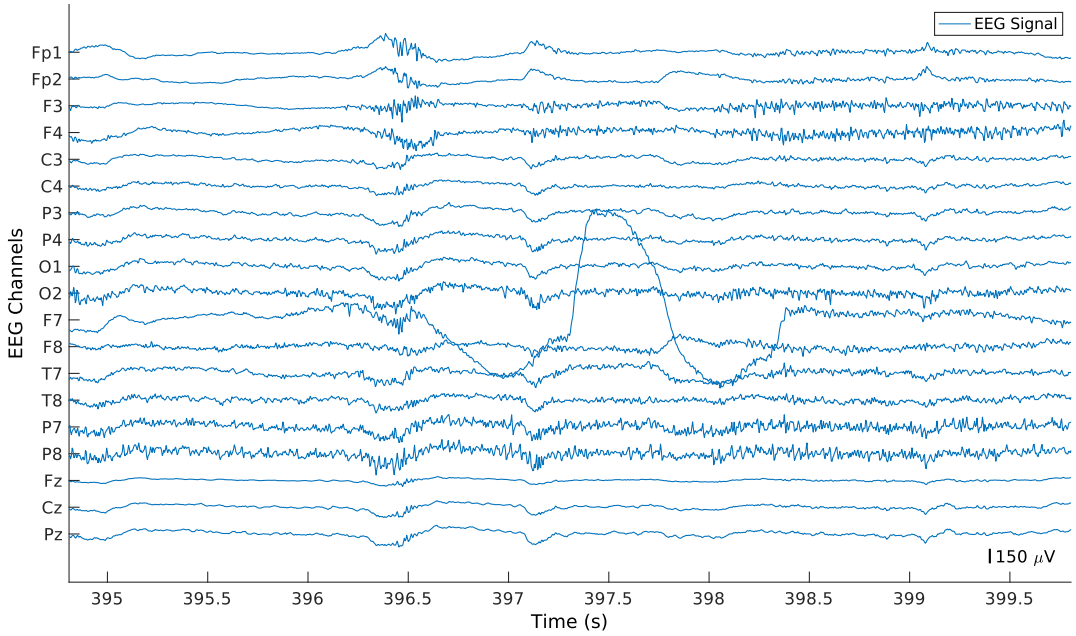


Figure 4.7: Example of EEG segment with an electrode pop present in channel F7.

4.2.1.7 Interpolation of noisy EEG channels, preparation for ICA, and ICA processing

Finally, the channels selected to be interpolated in the phase *Removal of Noisy EEG Segments* were interpolated using the spherical interpolation method [309], available in EEGLAB toolbox [245]. Finally, all segments were re-referenced to the average reference and decomposed by extended infomax ICA algorithm [222].

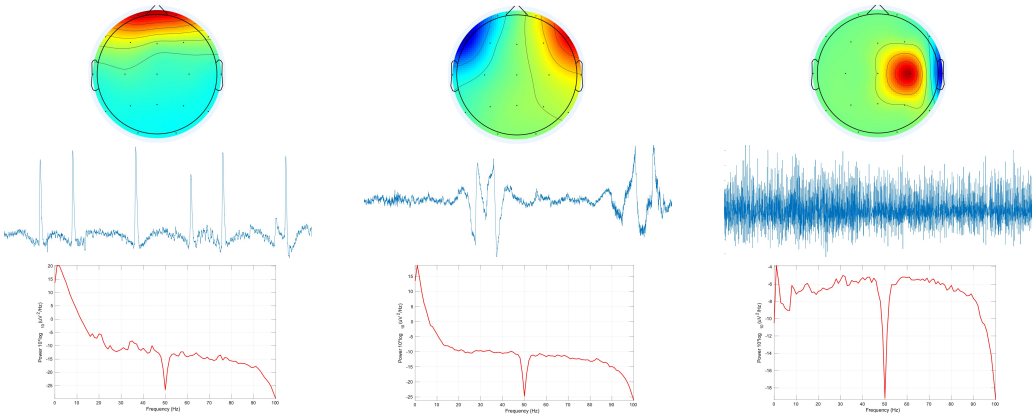
From the original 648 hours of signal, 35.32 hours (5.45%) were removed by the algorithm due to experimental errors. Regarding interpolation steps, 18.25 hours of the preprocessed EEG data (2.98%) contain, at least, one interpolated sample.

4.2.2 Independent components classification

The 612.68 hours of EEG data comprise 77,426 ICs. These data were randomly split into training and test sets. The training set contains 61,092 (78.86%) ICs, from 20 patients, whereas the test set contains 16,334 (21.14%) ICs, from the remaining 5 patients. Two experts visually inspected these segments. This analysis was performed following a semi-supervised approach using the ICLabel toolbox [196] available in the EEGLAB [245]. The ICs were first automatically classified by the ICLabel classifier, as brain component or artefact, and then corrected by the experts, when needed. To make the corrections, they verified the IC time series, PSD, and topographic map of each IC. If they did not agree with the ICLabel classification, they would change it according to their analysis. Figure 4.8 presents some examples of analysed ICs. ICs presented in Figures 4.8a, 4.8b, and 4.8c predominantly present artefacts. Figure 4.8d shows an IC with brain information manifesting (i) on the dipole showing in the topoplot and (ii) on the alpha-band peak showing in the PSD spectrum. Figures 4.8e and 4.8f present ICs containing both brain and noisy data. Despite the existence of noise, experts classified both components as brain in order to maintain neural information that could still be useful in further analysis.

Training and test sets were analysed differently. Each IC of the training set was only inspected by one expert, i.e. if one expert already analysed one IC it was not examined by the other expert. This analysis method was performed to have a dataset classified following different opinions, especially on the doubtful ICs. Training set contains 43,038 (70.44%) brain ICs and 18,054 (29.56%) artefact ICs. The test set was reviewed by both experts independently and finally, the ICs with different classifications were discussed by them to assign a final classification. This approach was made to have a test set validated by both experts with the minimum possible subjectivity. The test set contains 11,437 (70.02%) brain ICs and 4,897 (29.98%) artefact ICs.

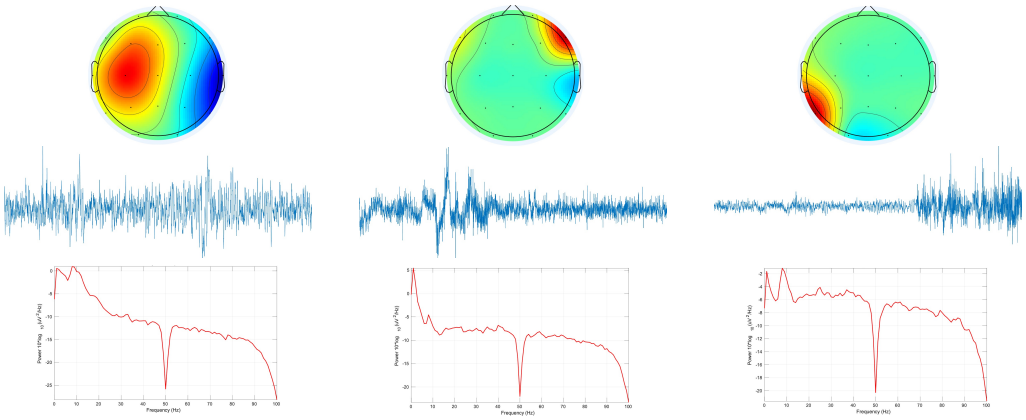
After all the processing methods, the time series, PSDs and topographic maps of all ICs were extracted. IC time series were obtained from the multiplication of the ICA weights with the EEG data. Depending on their length, these may contain 2,560 to 153,600 samples. The IC PSDs were obtained using the *spectopo* function available in the EEGLAB toolbox. These were restricted to frequencies between 1 and 90 Hz to reject the effects of the 0.5-100 Hz band-pass filter. The IC topographic maps were obtained using the *topoplot* function available in the EEGLAB toolbox. These comprise 67x67 pixels. Both PSD and topographic maps were normalised using the maximum and minimum values. The IC time series were not normalised because experts reported that the amplitude of the data was used to decide if the IC should be removed in doubtful cases.



(a) IC classified as noise, containing eye blinks.

(b) IC classified as noise, containing eye saccades.

(c) IC classified as noise, containing muscle activity.



(d) IC classified as brain. It contains mostly brain information.

(e) IC classified as brain. It contains some muscle noise but also contains brain information (peak on the alpha band).

(f) IC classified as brain. It contains some muscle noise but also contains brain information (peak on the alpha band).

Figure 4.8: Time series, topographic maps, and power spectrum densities of example independent components. The presented time series comprises only 5 seconds of the entire IC time series.

4.3 Data records

The dataset comprises two main directories containing the training and test sets. Inside each directory, three subdirectories contain artefact samples, not-artefact samples, and ICA weights. Data is stored in text files. Artefact and not-artefact subdirectories include the IC time series, the PSD, and the topographic map. The files comprising the IC time series and the PSD contain one-dimensional arrays, whereas those comprising topographic maps contain two-dimensional arrays. ICA weights subdirectory contains the weights used to convert the EEG segments into ICs. These comprise two-dimensional matrices with the number of rows equal to the number of ICs and the number of columns equal to the number of EEG channels. Figure 4.9 presents how the data is stored in the files. The dataset and the code used to create

it are available at https://doi.org/10.5281/zenodo.6620655 [310]. It is worth noting that the data were verified with the purpose of finding whether there were corrupted samples. The data is used in Chapters 5 and 6.

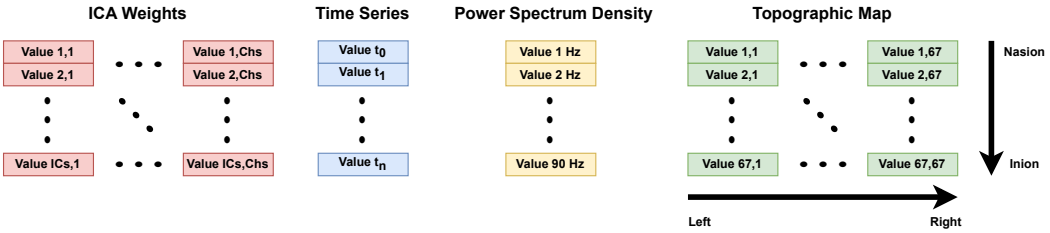


Figure 4.9: Example of text files containing ICA weights, IC time series, PSD, and topographic map. The values of the ICA weights are stored using the number of ICs as rows and the number of EEG channels as columns. The values of the time series are stored chronologically. The values of the PSD begin at 1 Hz and finish at 90 Hz. The values of the topographic map are stored as a square being the upper-left corner the left-nasion, the upper-right corner the right-nasion, the lower-left corner the left-inion and the lower-right corner the right-inion.

4.4 Final reflections

This chapter provides a dataset with ICs from long-term 19-channel EEG collected from patients with epilepsy under pre-surgical monitoring. Every sample in the dataset includes the time series, PSD, and topographic map of each IC. They were also classified as brain or artefact by two experts. The dataset was made publicly available because it is believed that it could be of a great value for the scientific community working on this field. Future work should focus on using the data to develop new IC classifiers or to benchmark the ones that already exist.

Chapter 5

Automatic electroencephalogram artefact removal using deep convolutional neural networks

This chapter concerns the development of the automatic electroencephalogram (EEG) reconstruction method using deep convolutional neural networks (DCNNs). The content of this chapter is based on the journal article published in *IEEE Access* [38]. Section 5.1 presents the study context. Section 5.2 details the methods used to develop the artefact removal model. It also contains the considered evaluation metrics. Section 5.3 describes the results and subsequent analysis using some examples. Section 5.4 discusses the obtained results, presents the advantages and the limitations of the developed approach, and provides some final reflections.

5.1 Study context

Brain potentials propagate over the entire scalp. Therefore, several electrodes are required to capture them with high spatial resolution [59]. Beyond brain information, these electrodes often capture noise, such as environment interference, experimental errors, and physiological artefacts [23].

In general, EEG artefacts can be reduced or avoided when the signal is acquired under controlled conditions. However, in tasks such as epileptic seizure prediction, EEGs must be continuously acquired over several days [76, 77] being practically impossible to avoid artefacts. Although a possible solution would be to detect and remove noisy segments, this removal would result in a high loss of information. Thus, researchers have developed artefact removal techniques to eliminate, or at least

attenuate, noisy data from the EEGs while preserving neural information [23, 78].

Recently, authors have reported new EEG artefact removal methods, based on deep learning architectures such as autoencoders (AEs), deep convolutional autoencoders (DCAEs), and DCNNs [197, 247–250]. These methods are presented to solve the drawbacks of previously developed methods. They perform quick signal estimation; are able to learn nonlinear relations leading to less loss of relevant information; are fully automatic; do not depend on the number of channels; and do not require any calibration before application. However, despite the advantages obtained using deep learning methods, studies using those methods evaluated them using either simulated data or data acquired under controlled environments. Therefore, these analyses do not completely simulate artefact removal from realistic long-term EEG acquisitions.

This chapter presents an EEG reconstruction model, based on DCNNs, developed using long-term data acquired from epileptic patients, in pre-surgical monitoring. It was evaluated using root mean squared error (RMSE), relative root mean squared error (RRMSE), Pearson correlation coefficient (PCC), and signal-to-noise ratio (SNR) difference. Finally, it was compared with the 1D-ResCNN model from [197] and with an automatic independent component analysis (ICA) model based on extended Infomax ICA and MARA classifier [236]. In summary, the main goal is to develop a model that automatically and quickly removes artefacts from long-term EEGs without human intervention, making it suitable to be applied in real-time long-term scenarios such as epileptic seizure prediction.

5.2 Materials and methods

This section presents the methods considered to prepare the dataset used in this study as well as the procedures followed to develop and evaluate the approach.

5.2.1 Data preparation

The dataset presented in Chapter 4 was used to develop the EEG artefact removal algorithm presented in this chapter. The dataset was divided in training and test sets. The first one contains 3,399 segments (486.03 hours), from 20 patients, whereas the second one includes 910 segments (126.65 hours), from the remaining 5 patients. As stated in Chapter 4, two experts visually inspected the independent components (ICs) of the EEG segments of both training and test sets with the purpose of eliminating noisy ICs. However, two different procedures were performed for both sets. EEG segments that were already analysed by one expert, were not analysed by the other, i.e. each expert analysed different segments from the training set. Test set was, firstly, analysed by both experts, independently. Then, discordant samples were inspected by the two experts together with the purpose of producing a set, validated

by both, to evaluate the approach. After the visual inspection, the segments from training and test sets were reconstructed using the non-noisy ICs. Finally, there were a training set and a test set with two different versions for the same EEG segment: the segment before visual inspection of the ICs (noisy segment), and the segment after the visual inspection of the ICs (target segment).

5.2.2 EEG artefact removal deep convolutional neural network

The proposed EEG artefact removal method, based on DCNNs, was designed to automatically remove noise from EEG segments. Although the ICA reconstruction is linear, the decisions performed by the experts to classify the ICs are nonlinear. Therefore, a nonlinear model is required to automatically remove noisy artefacts from the EEG segments.

DCNNs contain convolutional layers and layers with several possible activation functions. Convolutional layers [170] include several filters, used for extracting features from the input data, optimised during learning process. Layers with activation functions are used for controlling the information which is transferred to the following layer. rectified linear unit (ReLU) function is commonly used given its nonlinear behaviour and fast computation [170]. However, this nonlinear function can produce *dead neurons*, which means that some neurons of the network will output a zero value for different inputs. Leaky ReLU function was introduced in order to overcome this disadvantage [311]. It solves the problem by outputting a smaller portion of the negative inputs instead of nullifying them.

As seen in Figure 5.1, the developed architecture is based on three convolutional blocks, i.e., three sets of three convolutional layers followed by leaky ReLU activation function. The convolutional layers, used in each block, become wider as DCNN depth increases. ICA may be viewed as a single convolutional layer with a linear filter that covers all channels at a time. Therefore, more than one nonlinear convolutional layer is required to allow the model to better learn such task.

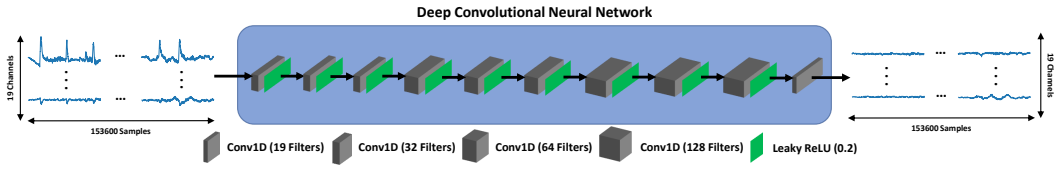


Figure 5.1: DCNN proposed to automatically remove artefacts from EEG segments. Input and output data contain 153600x19 samples. This size corresponds to the number of samples that a 10-minute segment with 19 channels, acquired using a sampling rate of 256 Hz, contains. Convolutional layers are presented as grey parallelipipeds. The larger the number of filters in the layer, the larger the width of the parallelipiped. All convolutional filters were of size 3. Leaky ReLU activation layers are presented green rectangles. All activation layers uses a α value of 0.2.

Since the various scalp EEG channels are not independent from each other, and as ICA processing covers all channels at the same time, the model removes artefacts

from all the channels, simultaneously.

Researchers report that deep learning models improve with the increasing of depth and width [312,313]. Thus, an architecture that combine both factors taking into account the available computational resources was developed (4 GPU NVIDIA Quadro P5000 with 16 GB GDDR5 RAM). The number of filters per layer starts at 32 and doubles from one block to the next. The last convolutional layer is used for converting the data back to the initial dimensions. Small filters are useful for exploring fine details of the data and have less computational cost than large filters [314]. Filters with size 1 were not considered because these ones are not able to analyse the values around the unit under analysis. Filters with an even size were also not used because these ones cannot maintain the symmetry around the unit under analysis resulting in data distortions across the layers. Finally, grid-search experiments were performed using filters with size 3 and filters with size 5. It was verified that the grid search results were similar for both filters' sizes. Therefore, all convolutional layers comprise filters with size 3 making the training of the model faster and less prone to overfit. A stride of 1 was used for every convolutional layer in order to not reduce the sample size across the layers.

All activation layers use leaky ReLU function. All the used leaky ReLU functions consider an α of 0.2 as suggested by Xu *et al.* [315].

5.2.3 Training and validation

The training set was further filtered by the number of eliminated ICs. Therefore, the EEG segments, with more than half of their ICs classified as noise, were discarded. This step was performed in order to remove segments with few brain independent sources, which would not provide enough information for reliable EEG segment reconstruction. After this filtering step, 2,900 EEG segments remained. It is worth noting that this filtering step was not performed in the test set.

The training set was further split in training and validation subsets by performing a random 70/30 holdout partition. Validation aims to prevent overfitting in training. Therefore, training subset contains 2,030 samples whereas the validation subset contains 870 samples. Each sample consists of one noisy segment and one target segment. After that, the samples lasting less than 10 minutes were zero padded. Thereafter, both subsets were standardised using the average and standard deviation calculated using all noisy segments belonging to the training subset.

For training the DCNN, adaptive moment estimation (Adam) optimisation function [316], with an initial learning rate of 3e-4, was used. Regarding the loss function, the usually used RMSE gives more significance to larger reconstruction errors, thus leading the algorithm to focus in artefacts with larger amplitude, independently from the range of values of the target signal (see Equation 5.1). For reducing this bias, the RMSE was replaced by the RRMSE [78], which normalises the RMSE by dividing

it by the root mean square (RMS) of the target EEG segment (see Equations 5.2 and 5.3).

$$RMSE(x,y) = \sqrt{\sum_{i=1}^{N} \frac{(x_i - y_i)^2}{N}} \tag{5.1}$$

$$RMS(x) = \sqrt{\sum_{i=1}^{N} \frac{x_i^2}{N}} \tag{5.2}$$

$$RRMSE(x,y) = \frac{RMSE(x,y)}{RMS(y)} \tag{5.3}$$

where:

x_i = i-th Input Value
y_i = i-th Target Value
N = Number of Samples

The model was trained for 500 epochs. The model was evaluated, every new epoch, using the validation subset, with the purpose of saving the one that obtained the lowest validation loss. The aforementioned procedures were performed ten times. This was intended to decrease the randomness of the training process. At the end of each run, the best model was saved with the intention of being tested with the completely independent test set. Table 5.1 summarises the hyperparameters used for training the models.

Table 5.1: Hyperparameters used to train the DCNN models.

Hyperparameter	**Value**
Dataset Partition	Holdout Validation 70/30
Optimisation Function	Adam
Learning Rate	3.0e-4
Loss Function	RRMSE
Epochs	500
Number of Runs	10

5.2.4 Evaluation metrics

The model was evaluated using standard statistical metrics. As standard statistical metrics, RMSE was used for measuring reconstruction error (see Equation 5.1), RRMSE was used for measuring normalised reconstruction error (see Equation 5.3), PCC was used for measuring the linear correlation between the denoised and the target segments (see Equation 5.4), and SNR difference was used for measuring the noise attenuation [239, 317, 318].

RMSE, RRMSE, and PCC were computed for both noisy and denoised segments. In other words, the noisy segments and the denoised segments were compared with the target segments. In this way, it was possible to inspect whether the DCNN model approximates the noisy data to the target data.

SNR difference is the difference between input and output SNRs (see Equations 5.5 and 5.6). Input SNR was computed under the assumption that noise is equal to the difference between the noisy and target segments. Output SNR was performed under the assumption that noise is equal to the difference between the denoised and target segments.

$$PCC(x,y) = \frac{Covariance(x,y)}{\sigma_x \sigma_y} \tag{5.4}$$

$$SNR(x,y) = 10 \times log_{10} \frac{\sum_{i=1}^{N} y_i^2}{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{5.5}$$

$$SNR_{diff} = SNR(\hat{y},y) - SNR(x,y) \tag{5.6}$$

where:

x = Noisy segment

y = Target segment

$\hat{y}$ = Denoised segment

N = Number of samples

RMSE, RRMSE, PCC, and SNR difference were computed for each EEG channel, independently. In this way, it could be possible to analyse the alterations that the model performed in each channel.

It is worth noting that the SNR difference cannot be performed when there is no difference between noisy and target segments. Thus, for implementing this evaluation metric, test segments containing only brain ICs were removed.

5.2.5 Comparison with different artefact removal models

The DCNN model was compared with 1D-ResCNN model from [197] and with an automatic ICA model based on extended Infomax ICA and MARA classifier [236]. As the 1D-ResCNN is not publicly available, it was reimplemented following the procedures presented by the authors. The MARA model is publicly available in EEGLAB toolbox. These models were chosen because they are also able to automatically remove several different artefacts from the EEGs.

All models were tested in a computer with an AMD Ryzen 5 2600 CPU 3.4 GHz, 64 GB of RAM, NVIDIA RTX 2060 Super, and Linux Ubuntu 20.04 LTS. The extended Infomax ICA-MARA was tested in Matlab 2019b whereas the DCNN and 1D-ResCNN models were tested using Tensorflow 2.0 and Keras 2.3 from Python

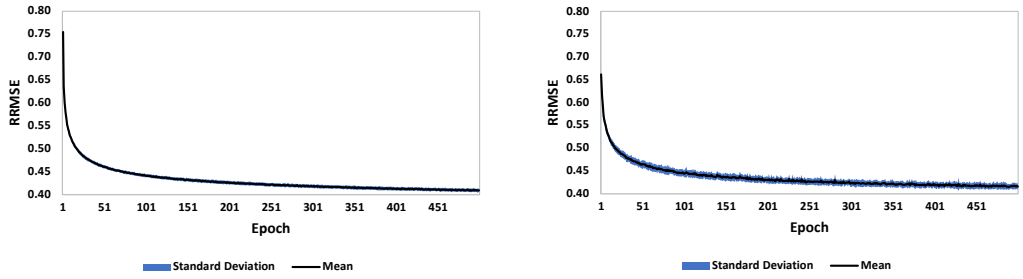
3.8. The inference phase of the DCNN models was performed using CPU rather than GPU with the purpose of comparing it with the extended Infomax-MARA model, which has to be performed in CPU. Additionally, testing the models on the CPU allows to approximate the simulation to a real environment where GPUs are usually less available.

5.3 Results

This section describes the results obtained for the developed approach.

5.3.1 Training and validation learning curves

Figure 5.2 shows the mean and standard deviation of the training and validation learning curves for all the developed models. Figures 5.2a and 5.2b show that the validation learning curve followed the training learning curve. This suggests that the developed models did not overfit the training data. Furthermore, it is seen that the models started to stabilise around the 300th epoch which means that the number of epochs was not a limiting factor to the learning procedure. Moreover, the low standard deviation indicates that all the ten models perform similarly. Therefore, one of them was randomly select for further analysis.



(a) Training learning curve. **(b)** Validation learning curve.

Figure 5.2: Mean and standard deviation of training and validation learning curves obtained from averaging all ten developed models.

5.3.2 Examples of EEG segments reconstructed by the developed approach

In order to demonstrate how the approach performed for the various types of artefacts found in segments, some examples of the noisy segments along with the target and obtained denoised segments are presented. One example for each type of EEG artefact present in the data is shown. More examples are available in Section A.4 in the Appendix A for further exploration.

Figure 5.3a shows three types of artefacts: eye blinks, eye movements, and muscle activity. Figures 5.3b, 5.3c, and 5.3d show these artefacts in detail whereas

Figures 5.3e, 5.3f, and 5.3g present the power spectrum density (PSD) of the time series in the selected time windows.

As can be observed in Figures 5.3b and 5.3c, the model was able to remove the ocular artefacts and returned a denoised segment similar to the target segment. However, Figure 5.3e evidences a loss of information in high frequencies. Eye blinks are typically artefacts with large amplitude and low frequency. Therefore, as the used loss function most strongly penalises the larger differences between the denoised and the target segments, the training of the model tries to find out how to reduce these artefacts before learning how to correct the small details of the data. As the EEG amplitude is, in most cases, inversely proportional to its frequency, in the case of an incomplete training there may be a loss of high frequency information. Figure 5.3d shows that the model attenuated the presence of the muscle activity, but did not remove it completely. This behaviour is confirmed by Figure 5.3g, i.e., there was only an attenuation of the PSD of the noisy channel. This may happen as a result of the difficulty of eliminating this artefact even by visual inspection of the ICs.

Figure 5.4 shows cardiac peaks in channel O1, which were not removed by the DCNN model. As these artefacts appeared rarely in the training set, the model may have had difficulty in considering them as noise.

Figure 5.5 shows pulse artefacts in channel C4. These artefacts resulted from having the EEG electrode on a pulsating vessel on the scalp. It can be observed that the model was not able to remove this interference from the noisy segment. These artefacts also did not occur frequently in the training set. Therefore, similarly to cardiac artefacts, the model may not had learned to consider them as noise.

Figure 5.6 shows electrode movement in all channels. These artefacts usually appear when there is a disturbance in the electrodes which leads to a change of impedance. In this case, the model was able to remove this interference from the noisy segment.

Figure 5.7 evidences that the model was not able to extract brain information from time intervals when there were electrode connection errors. It was expected that this type of artefact would be removed in the first stage of the EEG preprocessing algorithm. The algorithm was designed to remove portions, with an amplitude greater than 5 mV or lower than -5 mV, when the connection error occurred on several channels, simultaneously. Therefore, it is possible that some portions, with connection problems, still remained after the initial EEG preprocessing.

Figure 5.8 shows that the model has learned not to make considerable transformations when noise is not present in the EEG segments. However, it is seen that there was an attenuation of the high frequency waves in EEG channels, especially, where high amplitude artefacts usually appear such as Fp1, Fp2, F7 and F8. This means that the model focused excessively on removing artefacts on these channels containing low frequency artefacts, and thus, failed to learn high frequency details.

Figure 5.9 presents a very important behaviour of the developed model. This figure shows a portion of a segment, with a connection error, that resulted in removing more than half of the ICs by visual inspection. In this case, the model learned to analyse small windows of the entire noisy segment and to keep the data that did not have any noisy artefact. Before developing the models, EEG segment containing more than 50% noisy ICs were removed. Therefore, the models may not have learned to remove excessive information in situations as this one. This is an important advantage comparing to ICA approaches, because it preserved brain information while attenuating the influence of artefacts.

In summary, one may conclude that the model could attenuate artefacts such as eye blinks, eye saccades, muscle activity, and channel movements present in Figures 5.3b, 5.3c, 5.3d, and 5.6, respectively. Furthermore, it could perform reasonable reconstructions when no artefacts were present on the EEG data (see Figures 5.8 and 5.9). However, the model could not handle rare EEG artefacts such as cardiac artefacts, pulse artefacts, and saturated segments present in Figures 5.4, 5.5, and 5.7, respectively.

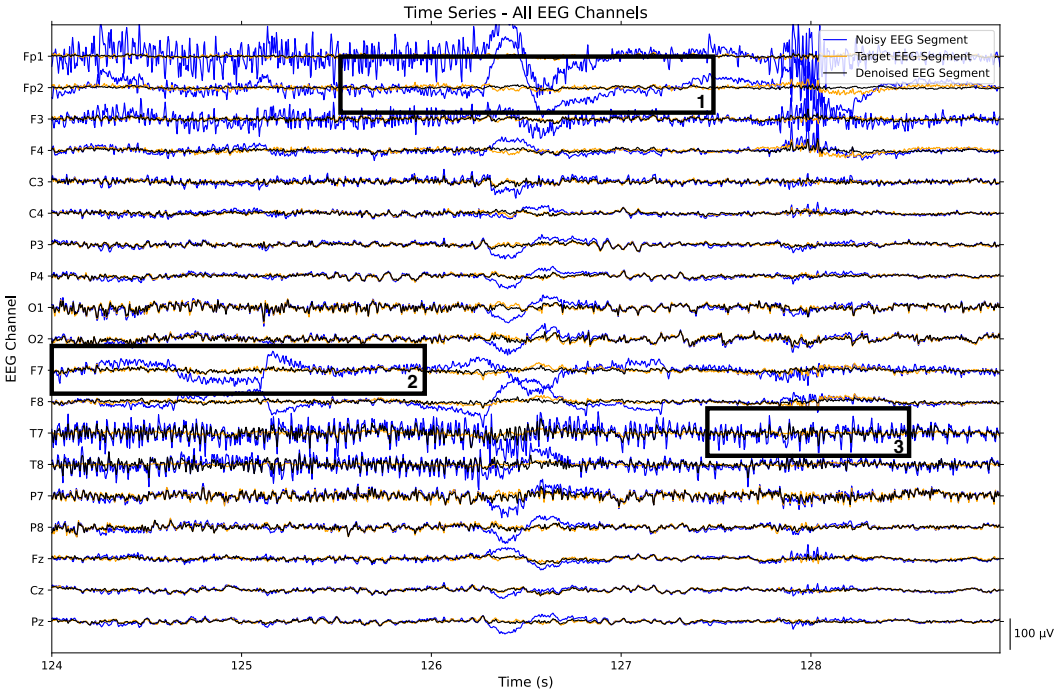
5.3.3 Evaluation metrics

The developed EEG artefact removal model was assessed using the evaluation metrics presented in Section 5.2.4. Thus, the metrics for all independent test samples were computed. The metrics were calculated for each EEG channel, independently. Therefore, for each EEG channel, 910 values for RMSE, RRMSE, and PCC for noisy and denoised segments and 875 values for SNR differences were obtained.

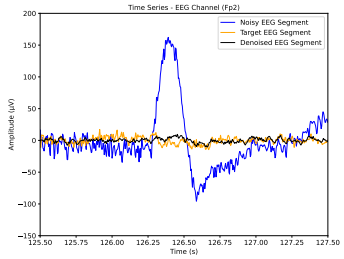
When inspecting the results for RMSE, RRMSE, PCC, and SNR difference, skewed distributions explained by the existence of some outliers in these metrics were observed. These outliers result from some experimental errors that were not removed in the initial preprocessing pipeline. Therefore, instead of using the common central tendency statistics, mean and standard deviation, the median and interquartile range were utilised (see Figure 5.10). Mean and standard deviation are available in Section A.3.

Figures 5.10a and 5.10b present the median and interquartile range values of the RMSE and RRMSE, respectively. As stated in Section 5.2.4, RMSE evaluates the reconstruction error whereas RRMSE measures the normalised reconstruction error. The lower these metrics are, the closer the obtained denoised data are to the target one. In general, these values decreased when using the DCNN model, which suggests that the model learned to approximate the noisy segments to the target ones.

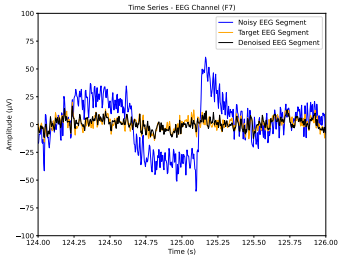
According to Figures 5.10a and 5.10b, Fp1, Fp2, F7, and F8 were the EEG channels associated with a larger decrease in RMSE and RRMSE. This occurred as a result of the removal of ocular artefacts, which typically have an amplitude higher



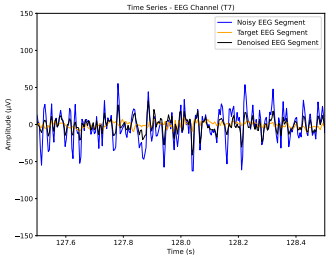
(a) Five seconds of all channels of an example EEG segment containing eye blinks (1), eye movements (2) and muscle activity (3).



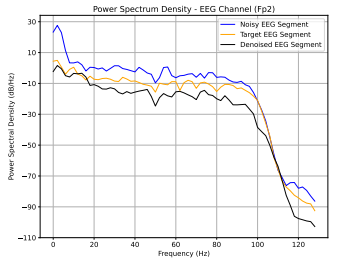
(b) EEG time series of the selected portion of the Fp2 channel of Figure 5.3a.



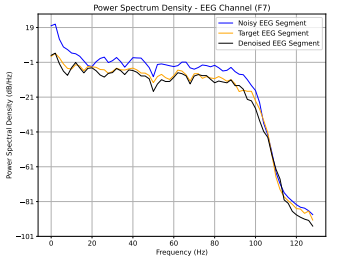
(c) EEG time series of the selected portion of the F7 channel of Figure 5.3a.



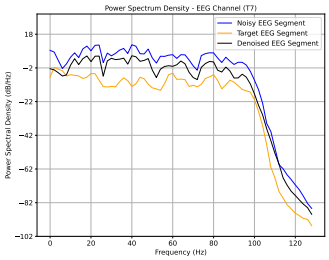
(d) EEG time series of the selected portion of the T7 channel of Figure 5.3a.



(e) PSD of the selected portion of the Fp2 channel of Figure 5.3a.



(f) PSD of the selected portion of the F7 channel of Figure 5.3a.



(g) PSD of the selected portion of the T7 channel of Figure 5.3a.

Figure 5.3: Example EEG segment from the test set. The noisy segment, target segment and denoised segment are presented in blue, orange and black, respectively. The selected portions of Figure 5.3a provide the exact moment of artefact occurrence: in Fp2 channel there is an eye blink, in F7 channel there is an eye saccade and in T7 channel there is muscle activity. These portions are zoomed in in Figures 5.3b, 5.3c and 5.3d. The PSD of the EEG time series, in the selected portions, are provided in Figures 5.3e, 5.3f and 5.3g.

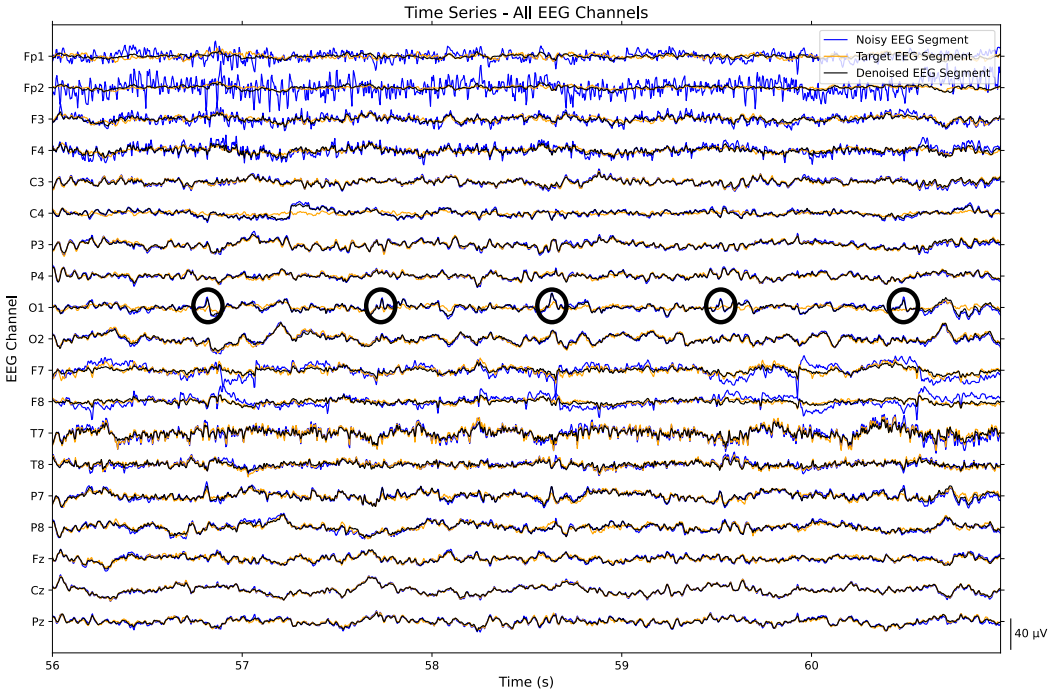


Figure 5.4: Five seconds of all channels of an example EEG segment. The noisy segment, target segment and denoised segment are presented in blue, orange and black, respectively. The selected portions provide the exact moments when the cardiac artefacts occur.

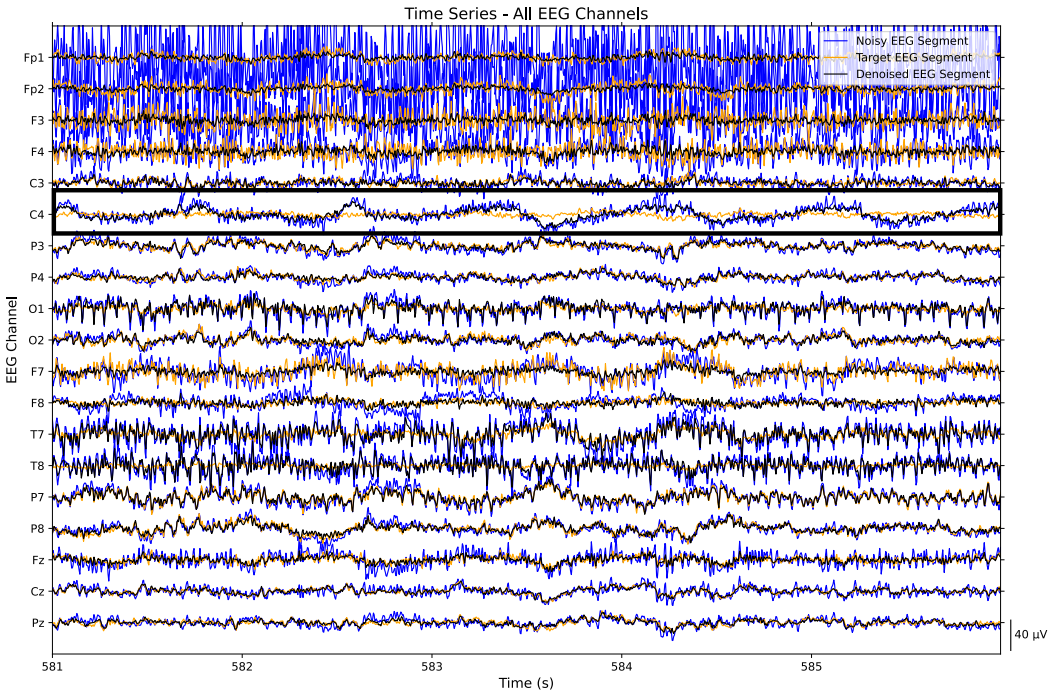


Figure 5.5: Five seconds of all channels of an example EEG segment. The noisy segment, target segment and denoised segment are presented in blue, orange and black, respectively. The selected portion evidences the EEG channel where the pulse artefacts occur.

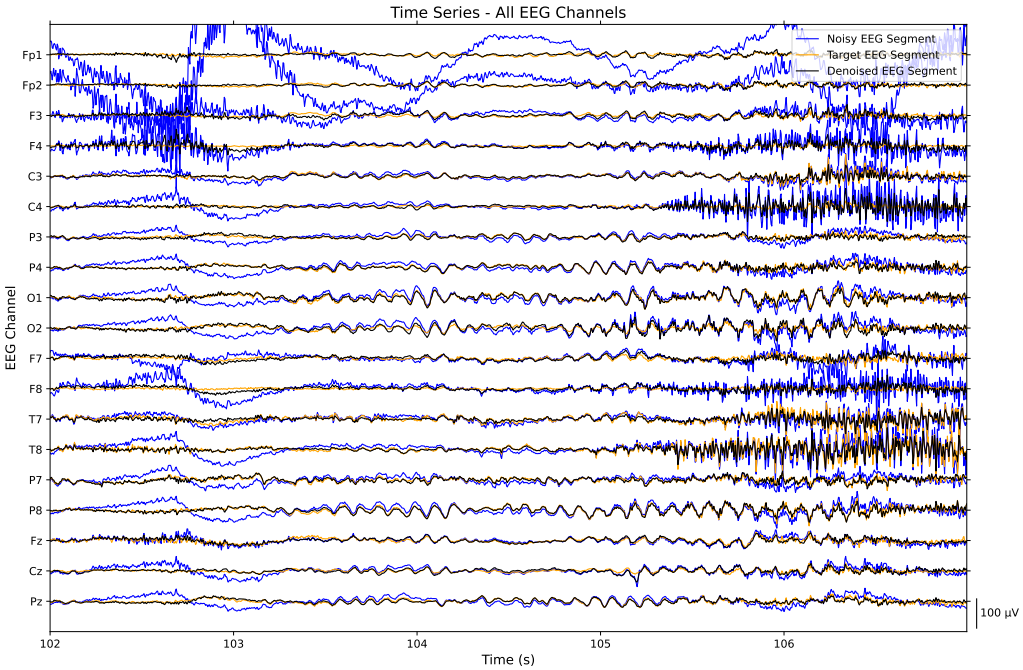


Figure 5.6: Five seconds of all channels of an example EEG segment contaning electrode movements. The noisy segment, target segment and denoised segment are presented in blue, orange and black, respectively.

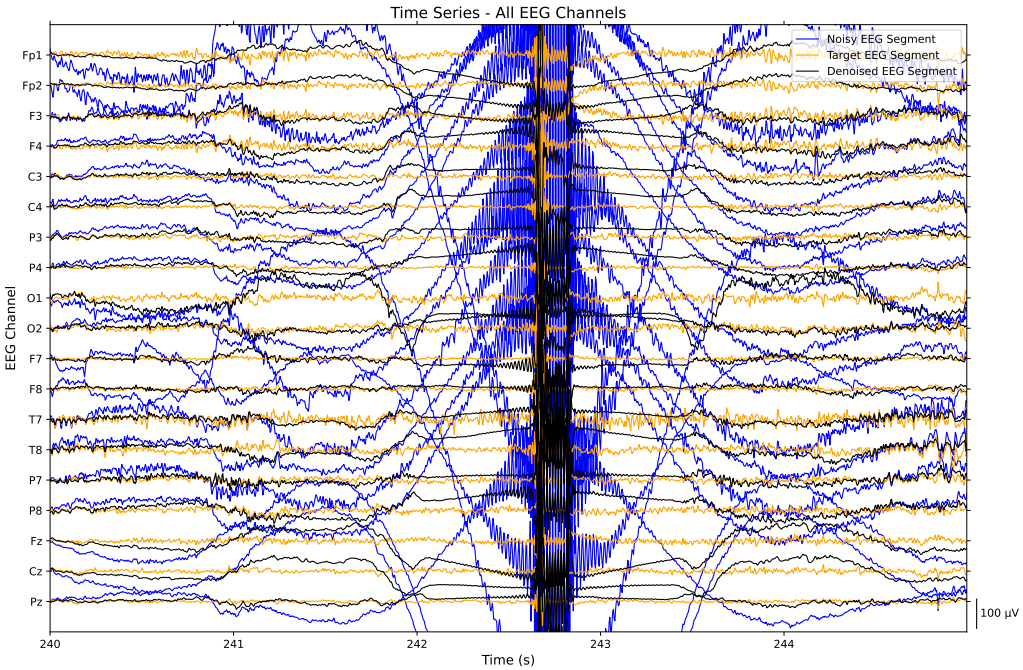


Figure 5.7: Five seconds of all channels of an example EEG segment containing experimental errors which were not removed in the first step of the EEG preprocessing algorithm. The noisy segment, target segment and denoised segment are presented in blue, orange and black, respectively.

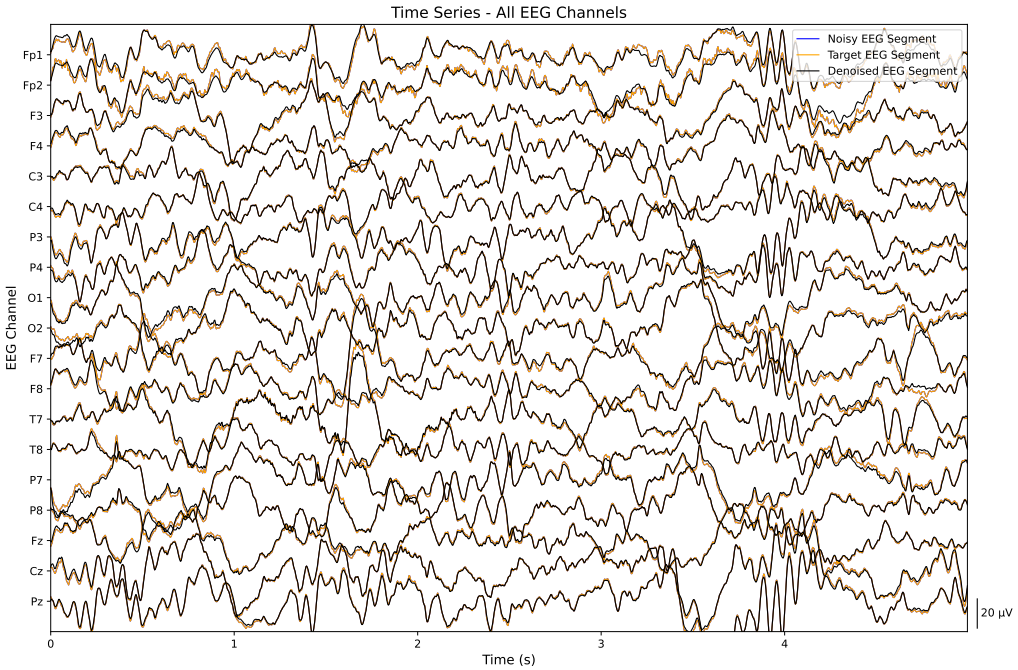


Figure 5.8: Five seconds of all channels of an example EEG segment that does not contain any noisy artefact. The noisy and target segments are equal as there are no artefacts in the EEG data. These segments are presented in orange whereas the denoised segment is presented in black.

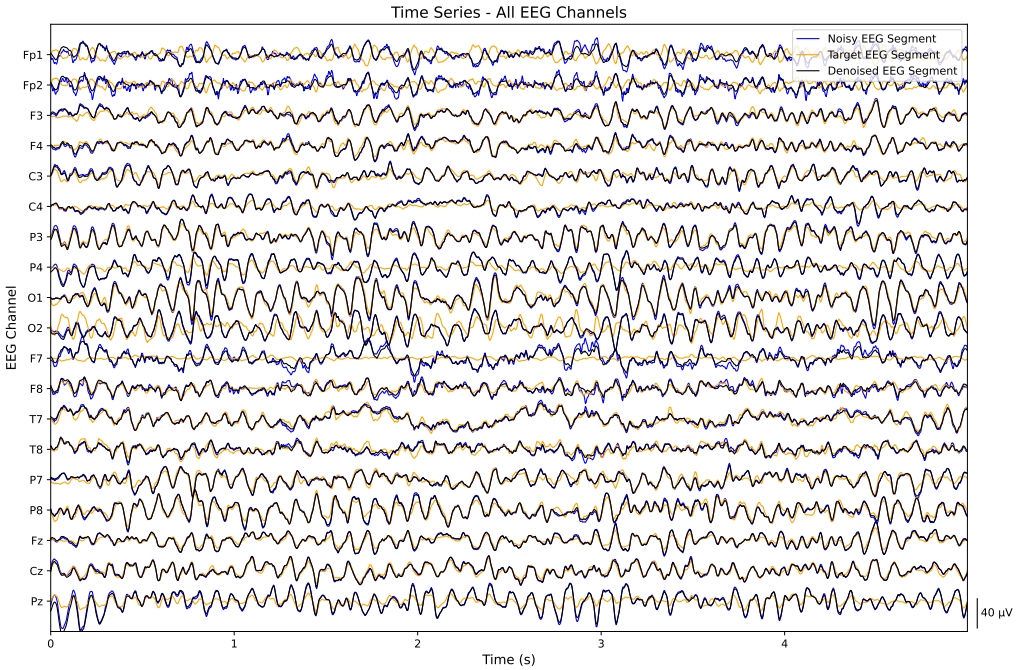


Figure 5.9: Five seconds of all channels of an example EEG segment which had some brain information removed by visual inspection that was not removed by the EEG artefact removal model. The noisy segment, target segment and denoised segment are presented in blue, orange and black, respectively.

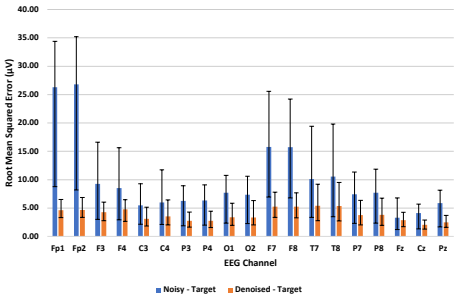
than the brain data. Channels F3, F4, T7, and T8 also evidenced a large reduction of these metrics. Although all EEG channels contained some muscle activity at a certain period of time, F3, F4, T7, and T8 were usually contaminated with this artefact over several segments. Therefore, results suggest that the developed model was able to reduce the presence of these artefacts, even in highly-corrupted channels. Comparing RMSE and RRMSE for the denoised segments, it is seen that although the channels O1, O2, P7, and P8 evidence RMSE values similar to the those obtained for channels C3, C4, P3, and P4, these channels present RRMSE values among the lowest of all channels. This means that the expected RMS of channels O1, O2, P7, and P8 were greater than the expected RMS of the channels C3, C4, P3, and P4. Therefore, it was concluded that the same error value has a lower relevance for those channels.

Figure 5.10c shows the median and interquartile range values of the PCC values. As mentioned in Section 5.2.4, PCC measures the linear correlation between two time series. Therefore, the higher this metric is, the greater is the linear correlation between the obtained denoised segment and the target one. In general, PCC increased after using the DCNN model, which suggests that the noisy segments became more linearly correlated with the target ones after using it. As already verified in Figures 5.10a and 5.10b, the larger PCC increase can be seen for the EEG channels containing ocular artefacts (Fp1, Fp2, F7, and F8).

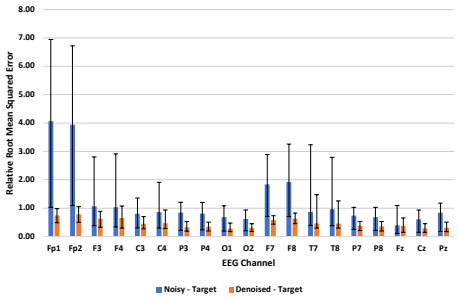
Figure 5.10d shows the median and interquartile range values of the SNR difference values. As mentioned in Section 5.2.4, SNR difference measures the improvement of the SNR after using the DCNN model. Positive values suggest a SNR increase, after using the model, whereas negative values suggest a SNR decrease and, therefore, a lower success in denoising the EEG. Although the overall results evidence the improvement of the SNR for all channels, the Fz channel presents an interquartile range that contains the zero value. As seen in Figures 5.10b and 5.10c, the interquartile range of the results for this channel, before using the DCNNs, contains almost optimal values. This indicates that for the test dataset, Fz channel was less corrupted by artefacts. Therefore, it was practically unchanged by the model.

5.3.4 Comparison with different artefact removal models

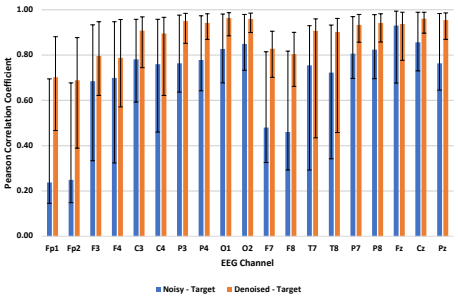
For each artefact removal model, the RMSE, RRMSE, PCC, and SNR difference were computed for all independent test samples. Furthermore, the computation times were also calculated. Contrary to the previous section, to simplify the comparison, these metrics were obtained using all channels. Table 5.2 presents the RMSE, the RRMSE and the PCC for the original data and denoised data reconstructed by the presented approach, 1D-ResCNN, and extended Infomax ICA-MARA. Furthermore, it contains the SNR difference and computation times for each artefact removal model.



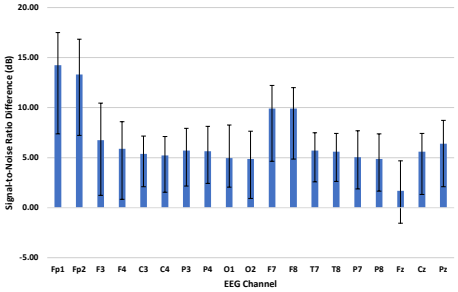
(a) RMSE of the EEG segments, for each EEG channel, before and after applying DCNN model.



(b) RRMSE of the EEG segments, for each EEG channel, before and after applying DCNN model.



(c) PCC of the EEG segments, for each EEG channel, before and after applying DCNN model.



(d) SNR difference of the EEG segments, for each EEG channel.

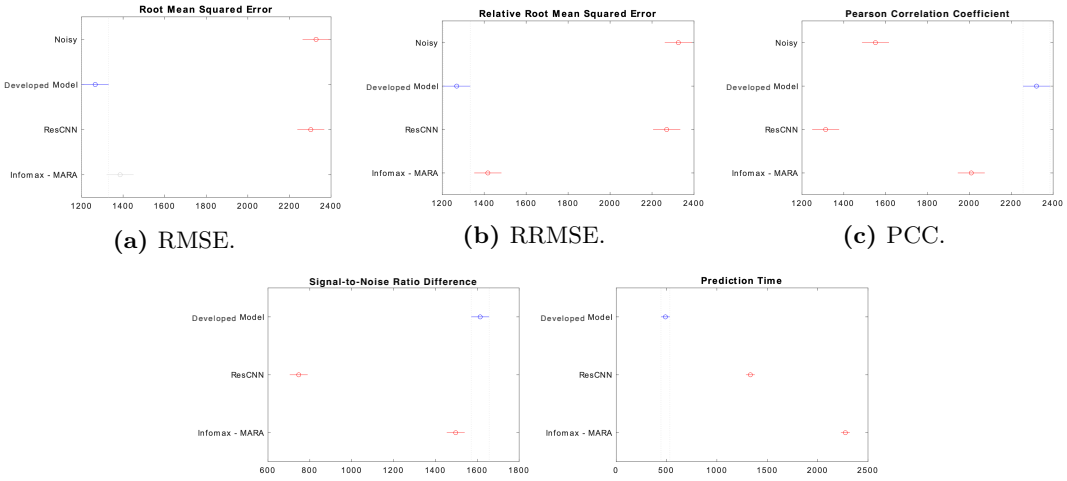
Figure 5.10: Performance results using the standard evaluation metrics reported in EEG-based studies. Each bar corresponds to the median value and the interquartil range.

Table 5.2: Statistical metrics used to compare the EEG artefact removal models. The values are presented in the format median (first quartile - third quartile).

	Noisy	**Denoised (Developed Model)**	**Denoised (1D-ResCNN)**	**Denoised (Extended Infomax ICA - MARA)**
RMSE (Microvolts)	13.88 (4.53 - 19.04)	4.83 (2.92 - 6.90)	9.83 (6.93 - 12.70)	5.28 (3.32 - 7.62)
RRMSE	1.57 (0.50 - 2.29)	0.52 (0.32 - 0.77)	1.03 (0.61 - 1.51)	0.63 (0.38 - 0.81)
PCC	0.54 (0.40 - 0.90)	0.86 (0.70 - 0.95)	0.53 (0.37 - 0.83)	0.78 (0.60 - 0.93)
SNR Difference	-	8.81 (4.83 - 10.05)	3.12 (0.11 - 3.78)	7.58 (3.08 - 10.37)
Computation Time (Seconds)	-	0.29 (0.29 - 0.30)	3.37 (3.35 - 3.38)	384.52 (375.43 - 391.49)

Pairwise comparisons, using non-parametric tests (Kruskal-Wallis [319] and Dunn-Šidák [320]), between all approaches, using all statistical metrics, were also performed to study whether there are statistical differences between them. To compare the performances, a significance level of 0.05 was considered. Figure 5.11 presents those pairwise comparisons. For RMSE and RRMSE, lower values are related with lower noise levels whereas for PCC and SNR difference, higher values are preferred.

Results presented in Table 5.2 evidence that the model obtained considerably lower reconstruction errors and higher PCCs and SNR differences, compared to the 1D-ResCNN. Furthermore, the DCNN model is faster than 1D-ResCNN. Figure 5.11 shows that the differences between both methods were statistical significant (p-value (RMSE) < 0.001; p-value (RRMSE) < 0.001; p-value (PCC) < 0.001; p-value (SNR



(a) RMSE. **(b)** RRMSE. **(c)** PCC.

(d) SNR difference. **(e)** Prediction times.

Figure 5.11: Pairwise comparison plots comparing the statistical metrics of the noisy data and the denoised data reconstructed by the developed model, by 1D-ResCNN and by extended Infomax ICA-MARA. The presented values represent the mean ranks and the standard errors obtained from the statistical test. Overlapping confidence intervals mean that there are no significant statistical differences between the results under analysis. These figures were obtained using the *multcompare* function from MATLAB.

Diff) < 0.001; p-value (Prediction Time) < 0.001).

Comparing the proposed approach with the extended Infomax ICA-MARA, results provided in Table 5.2 evidence that the DCNN model obtained lower median reconstruction error and higher median PCCs and SNR differences. Figure 5.11 also evidences that the DCNN model obtained lower RRMSE, higher PCC and SNR difference with significant statistical differences (p-value (RRMSE) $= 0.015$; p-value (PCC) < 0.001; p-value (SNR Diff) $= 0.003$). However, it shows that both models did not obtain significant statistical differences for RMSE (p-value $= 0.087$). Additionally, results demonstrated that the developed approach is considerably faster than the extended Infomax ICA-MARA which lasted around 6 minutes on average compared to less than a second (p-value < 0.001) in the developed method.

5.4 Discussion

The automatic EEG artefact removal approach, presented in this chapter, is based on DCNNs. It was designed to automatically remove several artefacts, commonly observed in long-term EEGs, such as ocular artefacts, muscle activity, cardiac activity, pulse artefacts, and electrode connection issues, in a similar way to that performed by experts. Studies cited in the literature review, which developed deep learning models that whether only remove one type of artefact [247, 248] or were trained with artificially generated noisy EEG data [197, 249], do not fully simulate real application scenarios such as clinical EEG evaluation. The proposed approach is a step forward because it was able to remove several artefacts present in long-term

signals collected from epileptic patients, in pre-surgical monitoring.

The DCNN model was developed using EEG data previously processed using ICA. ICA is a linear decomposition method. Therefore, the reconstruction of EEG segments, after removing noisy ICs, is performed using linear equations. However, the EEG artefact removal model must not only automate the linear reconstruction of the EEG segments without artefacts but also the nonlinear decisions performed by experts when classifying the ICs.

The developed approach was able to attenuate the influence of artefacts while preserving brain information. Additionally, it was able to recognise artefacts within a segment and minimise information related to these ones keeping the remainder data. Thus, it could preserve more information than ICA [197, 247]. These results could be related to the removal of training data with more than half of their ICs classified as noise, i.e., the model did not learn to excessively remove data. Finally, the model removed artefacts from signals which were not used in training, which means that it did not overfit to signals from patients used in training, and, therefore, may be used in EEGs from new subjects.

It was found out that the model had difficulty in preserving small details of the EEGs in channels where high amplitude artefacts were common, e.g., Fp1 and Fp2 channels. This behaviour was also noticed by Yang *et al.* [247]. They reported that gamma bands (more than 30 Hz) were not perfectly reconstructed when ocular artefacts were present, which means their model also lost high frequency details when signals were corrupted with this type of artefacts. Note that loss functions, which aim at reducing the reconstruction error, firstly find out how to decrease the larger errors, and secondly learn the smaller details. Therefore, as these channels presented high amplitude artefacts, the model learned to remove their influence before learning how to reconstruct low amplitude data. In the EEGs, the frequency is, in most cases, inversely proportional to the amplitude. Thus, although the DCNN model could preserve low frequency EEG data, it may require a different training setup, e.g., increasing the training set, searching for the optimal deep learning architecture, using longer training times or replacing the utilised loss function by another one, in order to improve its high frequency detail reconstruction.

Results evidenced that the proposed approach obtained the greatest performance among the tested artefact removal models. Compared to the 1D-ResCNN, this could be explained by the fact that the latter was developed using simulated noisy EEG data which could not precisely mimic real noisy EEG segments. Considering the extended Infomax ICA-MARA, the DCNN model presented a minor loss of information because instead of removing the entire source related with the noise, it focuses on processing just the time interval when the artefact occurs. Furthermore, the proposed approach is faster than the other evaluated approaches. Therefore, it may be used to remove artefacts from signals in real-time scenarios. It could be, for example, deployed into the IBM's TrueNorth Neurosynaptic System [33, 321],

which is a power-efficient neuromorphic chip that can be adapted to implement DCNNs [322], to remove artefacts from EEGs before epileptic seizure prediction.

The proposed approach was developed using EEG data from patients with epilepsy under pre-surgical monitoring conditions. These data were acquired without conditioning patients' activities. It means that the data contain several artifacts, which are likely to be present in EEGs acquired in uncontrolled environments for other research purposes. Therefore, although it was developed using epileptic EEG data, it may be used for denoising other types of EEGs.

5.4.1 Study limitations

As the proposed approach was developed with EEG segments acquired using a sampling rate of 256 Hz, it is restricted to acquisition systems using the same number of samples per second. However, this may not be seen as a strong limitation because studies often consider scalp EEG that was either obtained using this sampling rate [275, 283, 323–326] or using higher sampling rates which where subsequently downsampled for further analysis [327–330]. Moreover, as it was trained using multichannel EEG segments, it is also restricted to the same channel placement over scalp. Furthermore, it is limited to segments lasting up to 10 minutes, i.e., signals with longer duration must be segmented before being processed.

The model was developed with 19-channel EEG segments previously processed using ICA. ICA can only find a number of independent sources at most equal to the number of used channels. Although some authors state that 19 may be considered as a high number of EEG channels, ICA usually performs more accurately with EEG data with at least 64 channels [331, 332]. Therefore, the reconstruction capability of the DCNN model may be limited by the performance of the ICA decomposition. However, as the main goal of this study was to develop a model that would be able to work in real long-term scenarios, the pipeline was restricted to low-density EEG data that were available in the database.

5.4.2 Final reflections

This work demonstrates the potential of deep learning architectures in developing models that can automatically remove artefacts from EEGs in less than a second.

Removing artefacts present in real long-term EEGs, by visually inspecting the independent sources of the signals, is a time consuming task since it requires examining several hours of data. Therefore, deep learning approach which may automatically and quickly remove artefacts, such as eye blinks, eye movements, muscle activity, cardiac activity, and electrode connection interferences was developed. In this way, it could be used later to automatically eliminate noise from EEGs from other patients, available in the EPILEPSIAE database, or for removing noise in real-time scenarios.

Experimental results suggested that the developed model could attenuate the influence of the artefacts in the EEGs signals. Furthermore, compared to other approaches, the DCNN model could combine a minor reconstruction error with a fast computation, making it suitable to preprocess real-time long-term EEGs signals. It demonstrates that EEG artefact removal models, based on deep neural networks (DNNs), developed using real EEGs, should be taken into consideration when noisy artefacts are present in the EEG data.

Future work should tackle the development of DCNNs models using each EEG channel individually and comparing them with the multi-channel model. In this way, if the new approaches achieve similar or better performance, they could be used to remove artefacts from noisy segments acquired with any type of acquisition system, as long as one provides the same sampling rate.

Chapter 6

Ensemble deep neural network for automatic classification of EEG independent components

This chapter presents the development of a novel ensemble deep neural network (DNN) that combines time series, power spectrum densitys (PSDs), and topoplots to classify independent components (ICs) and the ability of using it in transfer learning approaches. The content of this chapter is based on the journal article published in *IEEE Transactions on Neural Systems and Rehabilitation Engineering* [39]. Section 6.1 presents the study context. Section 6.2 describes the datasets and the methodology followed to develop the study. Section 6.3 describes the results and subsequent analysis. Section 6.4 discusses the obtained results, presents the advantages and the limitations of the developed approach, and provides some final reflections.

6.1 Study context

Independent component analysis (ICA) is the blind source separation (BSS) algorithm most widely used to decompose multi-channel electroencephalograms (EEGs) into their independent sources [78, 217]. After this decomposition, an expert has to review the ICs and to classify them as brain signals or artefacts. This visual inspection is performed by examining the time series, the PSD, and the topographic map of each IC [238, 243]. This procedure is time-consuming and strongly relies on the expert's knowledge. Thus, to automate this task, researchers have developed several models to classify the ICs.

IC classifiers have been created using temporal, spectral, and spatial handcrafted features extracted from the time series, PSD, and topographic map of each IC, respectively [218, 235–242]. Recently, new approaches based on DNNs were formulated [196, 243, 244]. Deep learning approaches outperformed the automated models

developed using manually extracted features. The success of these methods comes from the ability of the DNNs to automatically extract complex nonlinear features directly from data [170]. The presented deep learning approaches achieved satisfactory accuracy performances (around 90%). However, contrary to experts who analyse the ICs, these models do not consider the temporal information of the ICs time series in the classification process. Therefore, their performance may be limited since they cannot fully simulate the visual inspection performed by trained experts.

In this chapter, a novel DNN that fully simulates the labelling of ICs performed by experts is proposed. It combines the three sources of information obtained from an IC: the raw time series, the PSD, and the topographic map. In this way, this study aims at analysing the importance of the temporal information for classifying the ICs. After that, the proposed approach is evaluated in another dataset, different from the one used to develop it, and compared the performance with publicly available approaches. After a thorough search of the relevant literature it was found that only ADJUST [237], MARA [236], ICLabel, and ICLabel$_{Lite}$ [196] can be downloaded and used to classify ICs. Therefore, these were used to perform the comparisons. Furthermore, since obtaining large labelled datasets to train DNNs is difficult, the capability of using the proposed model for transfer learning approaches was evaluated. For that, the performance of retraining a previously trained model was compared with that of a model trained from scratch.

6.2 Material and methods

This section presents the steps followed to develop and evaluate the proposed approach. Firstly, the datasets used to prepare the approach are introduced (see Section 6.2.1). Additionally, the preprocessing steps performed to prepare the data are described. Then, the model's architecture is detailed (see Section 6.2.2). Finally, the procedures performed to develop it; to verify whether the time series improve the performance of the IC classifier; to compare the performance of the presented architecture with publicly available IC classifiers; and to evaluate the capability of using the developed classifier in transfer learning approaches are presented (see Section 6.2.3).

6.2.1 Datasets

The approach was developed and evaluated using two datasets: the EPILEPSIAE dataset and the BASE dataset..

6.2.1.1 EPILEPSIAE dataset

EPILEPSIAE dataset was used to train and validate the IC classifier architecture. It corresponds to the EPIC dataset presented in Chapter 4.

6.2.1.2 BASE dataset

BASE dataset was used to evaluate the obtained architecture and the transfer learning step. It is composed by 2,340 ICs obtained from EEGs recorded during code comprehension in a cognitive load assessment study from BASE project [333, 334]. This dataset contains short-term EEGs, acquired from 30 subjects (24 males and 6 females, with ages ranging from 19 to 42 years), with an average duration of 9.45±3.92 minutes. The EEG data were acquired using a sampling rate of 1000 Hz and 60 electrodes organised according to the 10-10 international system. Contrary to the EEGs available in EPILEPSIAE database, EEG data from the BASE project were acquired in a controlled environment, during a few minutes. Therefore, these data contain less experimental errors. The dataset already contained the ICs and their labels. Therefore, it was not needed to perform ICA on the EEGs. From the 2,340 ICs, 1,296 ICs were classified as brain (55.4%) and 1,044 (44.6%) were classified as artefact. The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Faculty of Medicine of the University of Coimbra (protocol code CE Proc.CE-138/2018 and approved on 19 December 2018).

The procedures used to obtain the time series, PSDs, and topographic maps of each IC from the EPILEPSIAE dataset were also followed to obtain the data of each IC from the BASE dataset. However, since the EPILEPSIAE dataset contains ICs with time series with a sampling rate of 256 Hz, it was necessary to downsample the IC time series from the BASE dataset to 256 Hz.

The proposed approach is able to classify ICs with different duration. Therefore, it is not mandatory to restrict the duration of these time series to at most 10-minute length (maximum duration of IC time series present in EPILEPSIAE dataset).

6.2.2 Ensemble deep neural network architecture

As presented in Figure 6.1, the proposed architecture is an ensemble of DNNs (ensemble DNN). It contains three parallel DNNs: (i) a deep convolutional neural network (DCNN) combined with bidirectional long short-term memorys (BiLSTMs) layers (DCNN-BiLSTMs), to extract temporal features from the IC time series; and (ii) two DCNNs, to extract spectral and spatial features from the PSDs and topographic maps, respectively. For simplicity, the DCNN-BiLSTMs was named as temporal-features DNN (temporal-features DNN), the DCNN that extracts spectral features as spectral-features DNN (spectral-features DNN) and the DCNN that extracts spatial features as spatial-features DNN (spatial-features DNN).

This architecture was partially inspired in classifiers presented in Croce *et al.* [243] and in ICLabel$_{Lite}$ [196] which labelled ICs based on the PSDs and topographic maps. It is worth noting that the architecture of the ICLabel$_{Lite}$ was considered over the architecture of the default ICLabel because higher performance was re-

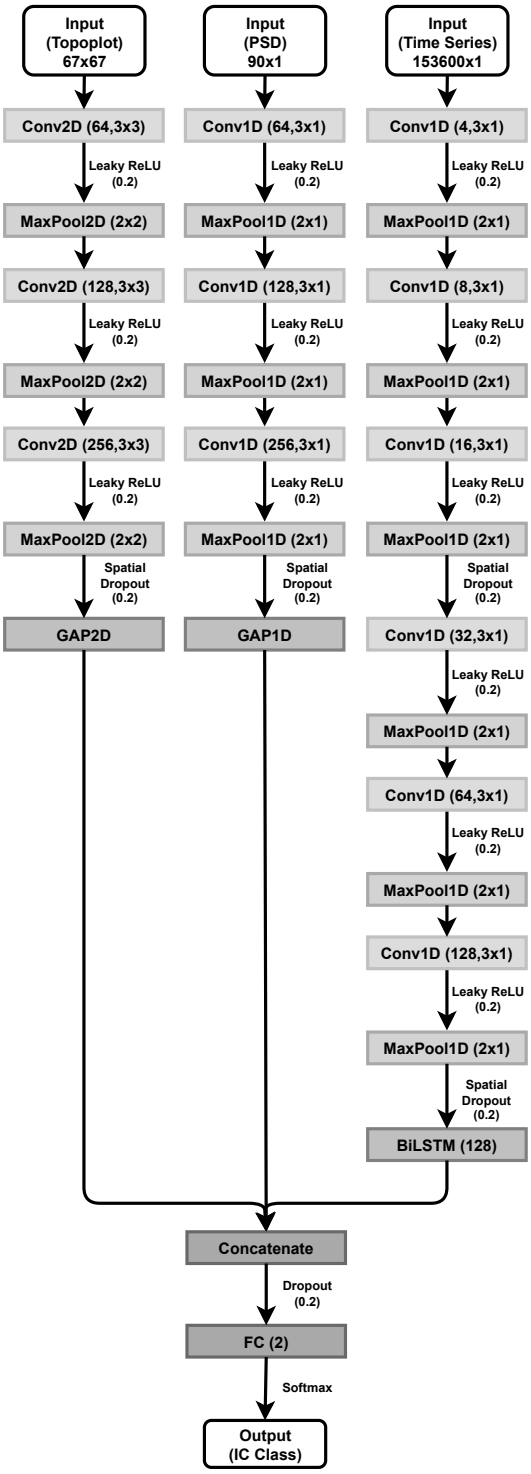


Figure 6.1: Ensemble DNN. It contains three different networks: one DCNN-BiLSTM to extract temporal features from time series and two DCNNs to extract spectral and spatial features from PSDs and topoplots, respectively. Convolutional layers contain the number of filters and the sizes, max pooling layers contain the size of the pooling windows, the leaky ReLU functions contain the used α value, spatial dropout and dropout layers contain the dropout percentage, the BiLSTM layer contains the number of LSTM units per LSTM layer, and the fully connected layer contains the number of output neurons. All convolutional layers and max pooling layers comprise a stride of 1.

ported for the former in the binary classification of the ICs [196]. Additionally, since experts use temporal information to classify the ICs, it was decided to add the time series to the architecture in order to fully simulate the expert's analysis. Therefore, to classify an IC, the model uses information from the corresponding time series, PSD, and topographic map.

The temporal-features DNN contains convolutional layers, activation functions, pooling layers, spatial dropout layers, and a BiLSTM layer. Convolutional layers [172] include several filters used to extract features from the data. Temporal-features DNN contains six convolutional layers. The number of convolutional layers were empirically chosen. The number of convolutional filters begins in 4 and goes up to 128, in powers of two. All filters were of size 3x1 with a stride of 1. The temporal-features DNN utilises the leaky rectified linear unit (ReLU) activation function after every convolutional layer. This function is an improved version of the ReLU function and that can handle the *dead neurons* problem [311] by taking into consideration a fraction of the negative values of the features maps. It is described by Equation 6.1 where x is the input vector and α is a small constant.

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{if } x < 0 \end{cases} \tag{6.1}$$

All the used leaky ReLU functions consider an α of 0.2 as suggested by Xu *et al.* [315]. There are six max pooling layers in the temporal-features DNN. All pooling windows were of size 2x1 with a stride of 1. Dropout layers are regularisation layers that randomly disconnect neurons during the training process. Spatial dropout layers are a particular case of dropout layers that instead of deactivating neurons, deactivate feature maps. Temporal-features DNN contains two spatial dropout layers: one after the third pooling layer and another after the sixth pooling layer. The percentage of spatial dropout during training process is 20%. The temporal-features DNN contains a BiLSTM layer which merges two long short-term memory (LSTM) layers comprising 128 LSTM units, respectively. The two LSTM layers used in the BiLSTM layer process the data in opposing directions. In this way, the BiLSTM layer is able to extract temporal features from both directions. The number of LSTM units was chosen in order to obtain an output with the same size as the other two DNNs.

The spectral-features DNN and the spatial-features DNN contain convolutional layers, activation functions, pooling layers, spatial dropout layers, and a global pooling layer. There are three convolutional layers in both networks. The number of layers was chosen considering the number of layers used by Croce *et al.* [243] and Pion-Tonachini *et al.* [196]. The number of convolutional filters are 64, 128, and 256, respectively. The filters of the spectral-features DNN are of size 3x1 whereas the filters of the spatial-features DNN are of size 3x3. All convolutional filters slide over the data with a stride of 1. Both DNNs utilise leaky ReLU functions with an α of

0.2. Additionally, there are three max pooling layers in both networks. The pooling windows of the spectral-features DNN are of size 2x1 whereas the windows of the spatial-features DNN are of size 2x2. All pooling windows slide over the data with a stride of 1. There is one spatial dropout layer, with a percentage of dropout of 20%, after the third pooling layer of each DNN. Both networks use global pooling layers to produce the features utilised in the classification. Comparing to flatten layers, global pooling layers drastically decrease the number of training parameters, thus reducing the computational time and also the probability of overfitting the training data [174, 335].

The features extracted in the three networks are concatenated and used to classify each IC as brain or artefact. For that, a fully connected layer is applied. Fully connected layers connect all inputs from the previous layer to compute the output values. Finally, the fully connected layer output was converted in probabilities using an activation layer with the softmax function [336].

It is worth noting that although the architecture was developed using a fixed length for the time series, it can classify ICs with different lengths as result of the non-existence of fully connected layers in the feature extraction networks.

6.2.3 Model development and evaluation

After the definition of the general architecture of the proposed approach, the importance of the time series to classify the ICs was evaluated. Afterwards, the proposed architecture was compared with other publicly available IC classifiers, using the BASE dataset. Finally, the capability of using the developed model in transfer learning procedures was evaluated.

6.2.3.1 Importance of the time series

In order to verify whether the temporal information present in the IC time series improves the classification, the ensemble DNN was compared with DNNs that do not use the time series, PSDs and topoplots at the same time. These DNNs comprised (i) only time series and PSDs (temporal-spectral-features DNN); (ii) only time series and topographic maps (temporal-spatial-features DNN); (iii) only PSDs and topographic maps (spectral-spatial-features DNN); (iv) only time series (temporal-features DNN); (v) only PSDs (spectral-features DNN); and (vi) only topographic maps (spatial-features DNN).

To train the aforementioned DNNs the training set of the EPILEPSIAE dataset was divided in a training subset and a validation subset using random 70/30 stratified holdout method. Therefore, 42,764 out of 61,092 ICs were used to train the models, while the other 18,328 ICs were used to constantly evaluate the performance of the models during the training phase. The time series of the ICs that did not contain 10 minutes of duration were zero padded so that it was possible to train the

models using batches with more than one element. Data from both subsets were standardised using the average and standard deviation computed from the training subset.

Training subset includes 30,126 brain ICs and 12,638 artefact ICs. Consequently, batch balancing was performed in order to have an equal number of ICs from both classes in each training batch. First, the ICs present in the training subset were separated according to the labels. A batch size of 128 was considered, made of 64 sequentially chosen ICs classified as brain and 64 ICs classified as artefact. As the number of ICs labelled as artefact was smaller, after using all these ICs once, the production of batches continued using already selected artefact ICs. In this way, it was possible to use all the ICs available in the training subset to develop the approach. The validation subset was not balanced.

Adaptive moment estimation (Adam) optimisation function with an initial learning rate of 3e-4 was used to train the models [316]. This function is able to adapt its learning rate considering the network parameters. Binary cross-entropy function was the loss function which should be optimised during training. The models were trained for 200 epochs using model checkpoint, i.e., they were evaluated, every new epoch, using the validation subset, with the purpose of selecting the ones which did not overfit on training subset.

Table 6.1 summarises the hyperparameters used to develop the IC classifiers.

Table 6.1: Hyperparameters used to train the IC classifiers.

Hyperparameter	**Value**
Dataset Partition	Holdout Validation 70/30
Optimisation Function	Adam
Learning Rate	3e-4
Loss Function	Binary Cross Entropy
Epochs	200

Testing set from the EPILEPSIAE dataset was used to evaluate the performance of the developed models. Contrary to the training and validation subsets, the time series of the testing set were not zero padded. Afterwards, the testing ICs were standardised using the average and standard deviation from the training subset. The models were evaluated using the sensitivity, specificity and geometric mean of sensitivity, and specificity (G-Mean). The G-Mean was used instead of the accuracy because of the unbalanced ratio of testing ICs. With the purpose of reducing the randomness of the model development pipeline, the training and evaluation of the models were repeated 30 times.

6.2.3.2 Comparison to other independent component classifiers

After assessing the importance of the time series to classify the ICs, the performance of the proposed approach was compared with other publicly available models, namely ADJUST [237], MARA [236], ICLabel, and ICLabel$_{Lite}$ [196]. BASE dataset was used to perform the comparisons with other state of the art models because, it was not labelled with the assistance of any IC classifiers. In this way, it was possible to compare with other models without having a bias. The BASE dataset was randomly divided in training subset, validation subset, and testing subset using a stratified 60/20/20 split. Therefore, training subset contains 1,404 ICs whereas the validation and testing subsets contain 468 ICs each. All the three subsets maintain the original dataset classes ratio (55.4% of brain ICs and 44.6% of artefact ICs). Then, the subsets were scaled using the normalisation values of the EPILEPSIAE training subset. Afterwards, the ensemble DNN was trained using the training and validation subsets and evaluated using the testing subset. For simplicity, it was named as BASE model. The optimisation function, learning rate, loss function, and number of epochs used to train the BASE model were equal to the ones presented in Table 5.1. Again, with the aim of reducing the stochasticity of the model development pipeline, these procedures were repeated 30 times.

6.2.3.3 Transfer learning approach

Finally, the capability of using the developed classifier in transfer learning approaches was evaluated. This evaluation was performed in three different ways: (i) the ensemble DNN already trained using the EPILEPSIAE dataset was tested on the BASE testing subset (EPILEPSIAE model); (ii) the already trained BASE model was tested on the BASE testing subset; and (iii) the weights of the already trained EPILEPSIAE model were transferred to a new model, the classifier layer was retrained using the BASE training and validation subsets, and was tested on the BASE testing subset (transfer learning model). In this way, one can study the possibility of using the proposed approach to develop future IC classifiers. Again, to reduce the randomness of the model development pipeline, these procedures were repeated 30 times.

6.3 Results

This section presents the results of the evaluations presented in Section 6.2.3. Firstly, the impact of using the time series for the classification of ICs is presented. Then, the results of the comparison between the performances of the proposed approach and four state of the art models: ADJUST, MARA, ICLabel, and ICLabel$_{Lite}$, using BASE dataset are shown. Finally, the improvement of the performance of the proposed approach when using transfer learning is presented.

6.3.1 Importance of the time series

Different versions of the approach were developed to evaluate the importance of the time series in the classification of the ICs. As stated in Section 6.2.3, these versions differ on the inputs used to classify the ICs. Table 6.2 presents the average results of the sensitivities, specificities, and geometric means of sensitivities and specificities (G-Means) computed for each approach for 30 runs.

Table 6.2: Performance of the IC classifiers with different inputs, using testing set of the EPILEPSIAE dataset. Values are presented in the format median (interquartile range).

Architecture	Sensitivity (%)	Specificity (%)	G-Mean (%)
Ensemble DNN	92.84 (92.25-93.38)	93.82 (93.03-94.18)	93.22 (93.05-93.43)
Temporal-spatial-features DNN	92.79 (91.99-93.65)	93.31 (92.63-93.72)	93.03 (92.78-93.09)
Spectral-spatial-features DNN	86.58 (85.88-87.05)	93.24 (92.82-93.44)	89.76 (89.57-89.92)
Temporal-spectral-features DNN	86.90 (86.39-87.56)	92.51 (91.74-92.94)	89.65 (89.52-89.78)
Temporal-features DNN	85.67 (85.03-86.40)	93.21 (92.76-93.84)	89.31 (89.10-89.44)
Spatial-features DNN	80.83 (78.99-81.52)	91.98 (91.67-92.52)	86.15 (85.66-86.40)
Spectral-features DNN	77.83 (77.38-78.48)	91.33 (90.86-91.74)	84.34 (84.15-84.50)

The median and interquartil ranges were presented instead of the mean and standard deviation because the results are not normally distributed. The pairwise comparisons were performed using Kruskal-Wallis [319] and Dunn-Šidák [320] tests. They were conducted between all architectures, using the G-Mean, in order to study whether there are statistical differences between the performance of different architectures. A significance level of 0.05 was considered. Figure 6.2 presents those pairwise comparisons.

Comparing the single source models, one verifies that the temporal-features DNN obtained higher performance than the others. However, the difference was only statistical significant when compared to the spectral-features DNN (p-value (temporal-features DNN - spatial-features DNN) = 0.264; p-value (temporal-features DNN - spectral-features DNN) < 0.001). The combination of PSD and topoplot improves the results of both single models whereas the combination of the time series and PSD obtained a similar performance to only using the time series. This may indicate that temporal-features DNN were able to extract spectral features from the time series. This evidence was also verified when comparing the ensemble DNN with the temporal-spatial-features DNN. The ensemble DNN and the temporal-spatial-features DNN obtained the highest performances. Although, the ensemble DNN presents a performance slightly higher than the temporal-spatial-features DNN, this difference is not statistical significant (p-value = 0.999). Probably this finding is related to the fact that DNN was able to extract spectral features from the time series, and, therefore, PSD may be discarded.

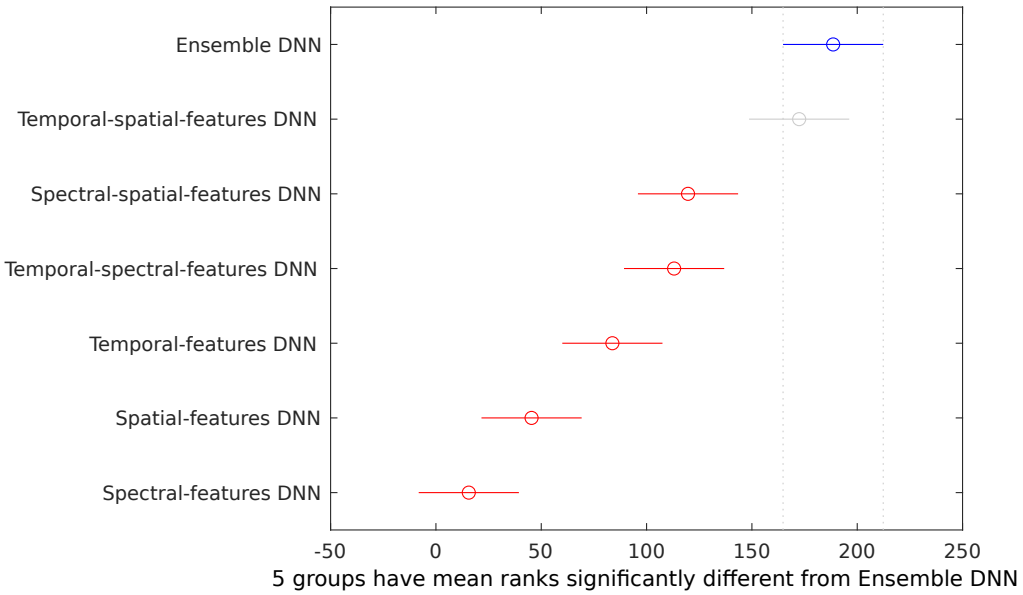


Figure 6.2: Pairwise comparison plot comparing the performances of the different IC classifier architectures, using the testing set of the EPILEPSIAE dataset. The presented values represent the mean ranks and the standard errors obtained from the statistical test. The overlap of the confidence intervals means that the models did not obtain statistical significant differences. This figure was obtained using the *multcompare* function from MATLAB.

6.3.2 Comparison with other independent component classifiers

Table 6.3 presents the average results of the sensitivities, specificities, and G-Means of the BASE model, ADJUST, MARA, ICLabel, and ICLabel$_{Lite}$ computed for 30 runs using BASE dataset.

Table 6.3: Performance of the different IC classifiers, using testing subsets of the BASE dataset. Values are presented in the format average mean ± standard deviation.

IC Classifier	**Sensitivity (%)**	**Specificity (%)**	**G-Mean (%)**
BASE model	92.39 ± 2.43	93.62 ± 2.42	92.98 ± 1.13
ICLabel	97.65 ± 0.99	85.09 ± 2.72	91.14 ± 1.56
ICLabel$_{Lite}$	98.47 ± 0.86	74.14 ± 3.39	85.42 ± 2.05
MARA	85.66 ± 2.53	91.51 ± 1.66	88.52 ± 1.56
ADJUST	80.66 ± 2.10	82.67 ± 2.04	81.65 ± 1.53

As the results follow a normal distribution, they are presented using mean and standard deviation. The pairwise comparisons were performed, using parametric tests (one-way ANOVA and Tukey-HSD tests), between all approaches, using the G-Mean, to study whether there are statistical differences between them. A significance level of 0.05 was used. Figure 6.3 presents those pairwise comparisons.

Default ICLabel performed better than handcrafted-features models MARA and ADJUST (p-value (ICLabel - MARA) < 0.001; p-value (ICLabel - ADJUST) < 0.001). These results are in accordance to the results reported by Pion-Tonachini *et al.* [196]. However, contrary to the results reported by them, default ICLabel

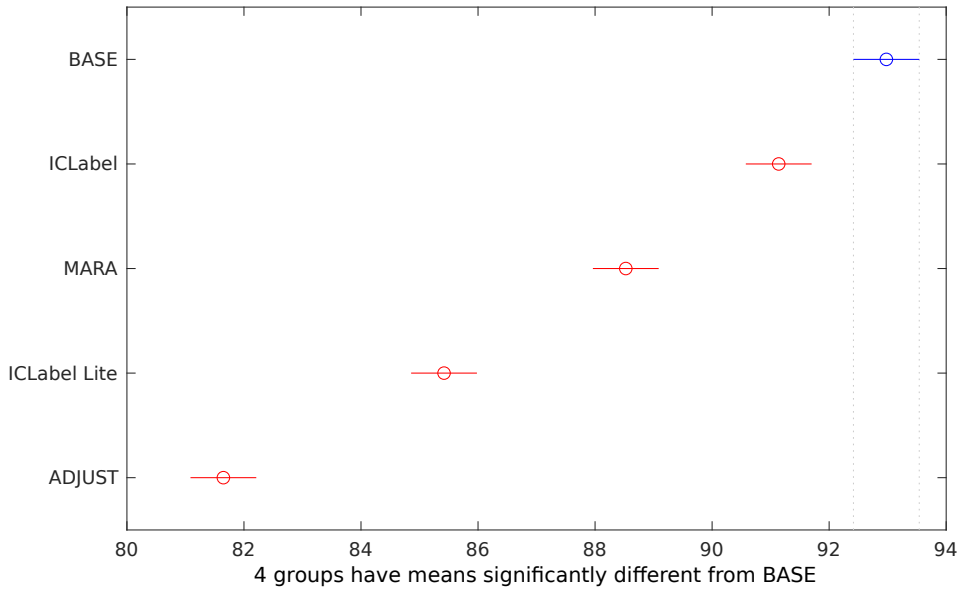


Figure 6.3: Pairwise comparison plot comparing the performances of the BASE model, ADJUST, MARA, ICLabel, and ICLabel$_{Lite}$, using the testing subsets of the BASE dataset. The presented values represent the G-Means and the standard errors obtained, for each model, from the statistical test.

performed better than the ICLabel$_{Lite}$ in binary classification (p-value < 0.001). BASE model obtained the highest G-Mean among all the IC classifiers (p-value (BASE - ICLabel) < 0.001; p-value (BASE - MARA) < 0.001; p-value (BASE - ICLabel$_{Lite}$) < 0.001; p-value (BASE - ADJUST) < 0.001). These results indicate that the time series of the ICs may be important to classify them.

6.3.3 Transfer learning approach

Table 6.4 presents the average results of the sensitivities, specificities, and G-Means of the EPILEPSIAE model, BASE model, and transfer learning model computed for 30 runs, using BASE dataset.

Table 6.4: Performance of the different IC classifiers, using testing subsets of the BASE dataset. Values are presented in the format average mean ± standard deviation.

IC Classifier	Sensitivity (%)	Specificity (%)	G-Mean (%)
Transfer learning model	93.48 ± 1.77	94.91 ± 1.51	94.18 ± 0.99
BASE model	92.39 ± 2.43	93.62 ± 2.42	92.98 ± 1.13
EPILEPSIAE model	74.07 ± 3.40	99.45 ± 0.51	85.80 ± 1.92

The performances follow a normal distribution. Therefore, they are presented using mean and standard deviation. The pairwise comparisons were performed, using parametric tests, between all approaches using the G-Mean, to study whether there are statistical differences between them. Once again, a significance level of 0.05 was used. Figure 6.4 presents those pairwise comparisons.

It was verified that the transfer learning model obtained the highest performance

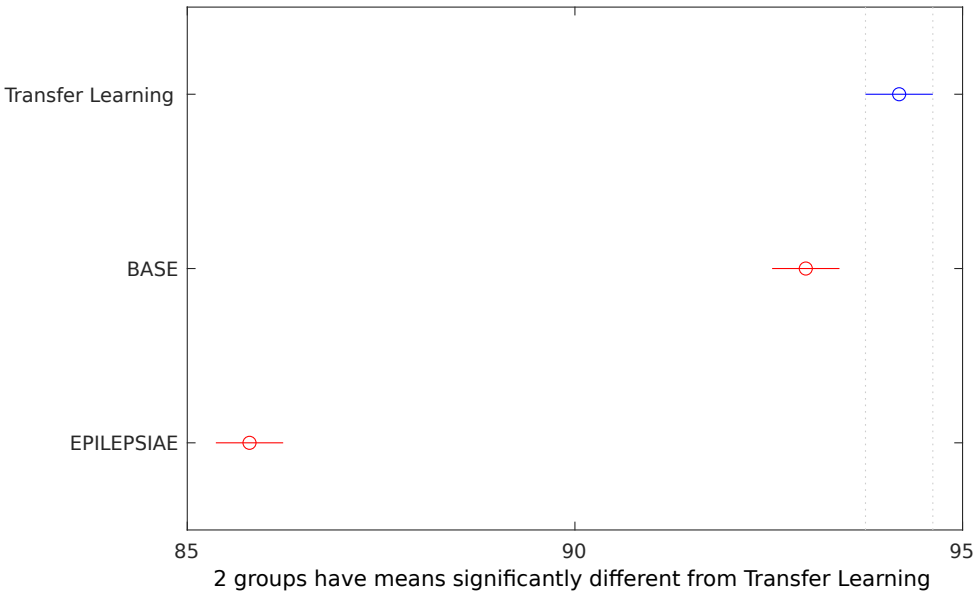


Figure 6.4: Pairwise comparison plot comparing the performances of the EPILEPSIAE model, BASE model, and transfer learning model, using the testing subsets of the BASE dataset. The plot presents the G-Means and the standard errors obtained, for each model, from the statistical test.

(p-value (Transfer Learning - BASE) = 0.004; p-value (Transfer Learning - EPILEPSIAE) < 0.001). This demonstrates that the IC classifier benefited from pretraining with a large dataset before being adjusted for a specific dataset.

6.4 Discussion

Different IC classification approaches have been presented in the previous years. Approaches based on DNNs obtained the highest performances [196, 243]. However, these approaches only used the PSDs and topoplots to classify the ICs. The decision of discarding the IC time series could be performed in order to obtain models prepared for any type of sampling rate. Considering that experts classify the ICs according to the information that is present not only in the PSDs and topoplots but also in the time series, the impact of using the time series in the performance of automatic IC classifiers was assessed. Therefore, an ensemble DNN that attempts to simulate the visual inspection performed by experts was developed. The deep learning architecture presented in this chapter was inspired in the ICLabel$_{Lite}$ architecture [196]. ICLabel$_{Lite}$ was considered over the original ICLabel because higher performance was reported for the former in the binary classification of the ICs. ICLabel$_{Lite}$ used features extracted from PSDs and topoplots to classify the ICs. The new contribution of the proposed approach consists in using a DCNN-LSTM to extract features from the time series. Afterwards, these temporal features are concatenated with the spectral and spatial features extracted from the PSD and

topoplot, respectively, and used to classify the IC under analysis. In this way, the impact of using the temporal information provided by the time series in improving the performance of the ICs classifiers was assessed.

To evaluate the importance of the time series, three types of models were developed: models with only one input source, models with two input sources and the model with the three input sources. Results indicate that using the time series improves the classification of ICs. Additionally, results also suggested that the model may have been able to extract spectral information from the time series. This was already expected since the PSDs are directly computed from the time series.

After assessing the relevance of the time series, the proposed approach was compared with other publicly available IC classifiers: ADJUST [237], MARA [236], ICLabel, and ICLabel$_{Lite}$ [196]. Results evidence that the ensemble DNN architecture obtained the highest performance. Once again, this may indicate that the usage of the time series improves the classification of the ICs. Furthermore, results also show that classifiers based on DNNs obtained better performance than handcrafted-features models. This may be explained by the fact that DNNs are able to better explore every detail present in the data visualised by the experts. Correctly classifying the ICs is mandatory for proper further EEG studies because if the classifiers label important brain data as noise, significant information could be lost and, therefore, not analysed. Also, the wrong classification of noisy ICs could lead to false results [337]. Therefore, it is crucial to continuously enhance the exactness of these automatic IC classifiers [25, 338].

After comparing the ensemble DNN with other publicly available models, one verified whether the transfer learning method improves its performance. For that, using the testing subsets of the BASE dataset, three different models were compared: a model trained with training set of the EPILEPSIAE dataset (EPILEPSIAE model); a model trained with the training and validation subsets of the BASE dataset (BASE model); and a model with all the weights transferred from the EPILEPSIAE model, but with the classification layer retrained using training and validation subsets of the BASE dataset (transfer learning model).

EPILEPSIAE model obtained an almost perfect specificity. However, its sensitivity was the lowest among the three evaluated approaches. EPILEPSIAE model was developed using ICs obtained from long-term 19-channel EEG data acquired during presurgical monitoring. Long-term data are susceptible to several artefacts. To remove these artefacts, ICA requires a large number of channels. Since the number of used EEG channels is not high, there could be some ICs with both brain and artefact information. With the purpose of not having a high loss of valuable information, the experts had to preserve these ICs despite containing some artefacts. BASE dataset contains short-term 60-channel EEG data acquired in a controlled environment. The high number of channels allows a better separation from brain ICs and artefact ICs, i.e., generally, the ICs do not contain both brain and artefact

data. Therefore, the experts that labelled this dataset could be less conservative. This means that some ICs that would not be removed in EPILEPSIAE dataset, were removed in BASE dataset, which could explain the lower sensitivity and higher specificity.

BASE model obtained higher sensitivity than the EPILEPSIAE model, evidencing higher ability to identify artefact ICs from data acquired in controlled environment. This was already expected because BASE model was trained using ICs obtained from high-density short-term EEG with few artefacts.

The classification of the BASE dataset clearly improved after performing transfer learning. Transfer learning model was created using the pretrained EPILEPSIAE model and the training and validation subsets of the BASE dataset. Essentially, the concept of the classifier layer of the EPILEPSIAE model was changed by optimising it with ICs from the BASE dataset. This indicates that the classifier layer of the EPILEPSIAE model may have been biased to the type of ICs available in the EPILEPSIAE dataset and, therefore, not prepared for the ICs provided in the BASE dataset. However, after the retraining, the transfer learning model was even better than the model trained from the scratch using exclusively ICs from the BASE dataset. Consequently, EPILEPSIAE model may be used in transfer learning approaches. This would make the train of new classifiers severely faster, since it would not be necessary to develop these ones from the beginning; would make the model more robust, given that it had already learned several IC details from a large dataset; and would give the opportunity to researchers with small datasets, to develop their own IC classifiers with less probability of overfitting [183].

6.4.1 Study limitations

Beyond the already discussed details, some other limitations must be taken into consideration. The proposed approach does not contain any fully connected layer in the feature extraction layers. Thus, the model is able to classify ICs with different duration than the ones used to train it. However, the time series must have the same sampling rate in order to maintain the temporal order learned from the train dataset. For that, the time series of the ICs have to be resampled before computing the PSDs, the topoplots and the IC classification.

BASE dataset contains few samples, which could lead to biased results. Furthermore, the results highly depend on the labels and, therefore, if errors have been made in the classification of the ICs, these will be reflected in the obtained performance. Moreover, it only comprises EEGs from a certain task. Given that, more datasets should be used to make more robust comparisons, e.g., EEGs collected in controlled and uncontrolled environments to verify if the obtained conclusions are still applicable.

6.4.2 Final reflections

This chapter presents a novel ensemble DNN to classify ICs extracted from EEGs using ICA. The novelty of this approach lies in the use of the temporal information of the time series to improve the classification of the ICs. Results demonstrated that the use of the time series improved the classification of the ICs and, therefore, enhanced the automatic removal of artefacts from EEGs. Additionally, it was verified that the ensemble approach trained using EPILEPSIAE dataset may be used in transfer learning approaches with the purpose of improving the classification of ICs from other datasets collected in completely different acquisition conditions. In this way, the model was made publicly available on GitHub (https://github.com/fabioacl/DeepICClassifier) so that other researchers can use it to improve their approaches using the transfer learning method. Future work should focus on evaluating the proposed approach in larger datasets containing ICs from EEGs from several different tasks.

Chapter 7

Removing artefacts and periodically retraining improve performance of neural network-based seizure prediction models

This chapter presents the development of seizure prediction models. The content of this chapter is based on the journal article published in *Scientific Reports* [40]. Section 7.1 presents the study context. Section 7.2 describes the datasets and the methodology followed to develop the study. Section 7.3 describes the results and subsequent analysis. Section 7.4 discusses the obtained results, compares them with those presented in the state of the art, presents the advantages and the limitations of the developed approaches, and provides some final reflections.

7.1 Study context

Seizure prediction has been an active research theme since 1970 [131]. Several studies have been published in this research area, typically based on electroencephalogram (EEG) [17, 64]. Initially, seizure prediction models were threshold-based [263, 264]. However, these models were linear and based on a single feature, which might not be sufficient to perceive the complexity of the pre-seizure activity [18, 283]. Later, shallow machine learning algorithms were employed with acceptable results for some patients [64, 137, 146]. In recent years, deep learning architectures have been increasingly used in multiple research areas [170]. These architectures are not dependent on the computation of handcrafted features be-

fore classification, as they can extract information directly from the data. Moreover, as these models are able to automatically extract and select the optimal features, less feature engineering and domain knowledge are needed to develop intelligent systems. Although these characteristics turn machine learning models into black-boxes, they could be advantageous when no solid physiological grounding exists. Therefore, authors have started to use them to develop seizure prediction approaches [21, 28, 29, 181, 219, 260, 265, 269, 274, 277–287, 289, 293, 295, 297, 339, 340].

Even though the increasing complexity of seizure prediction algorithms is a significant topic, others are equally important. An example is the EEG preprocessing [137, 341]. Researchers are moving towards non-invasive EEG to develop seizure prediction approaches [16]. However, these signals usually present artefacts. EEG artefacts may be responsible for the increase of false alarms and should be removed before creating the seizure prediction models [19–21]. Although different methods handling artefacts have been published, to the best of found knowledge no one compared the prediction performance of models developed using noisy data with models developed using denoised data [22, 25–28, 74, 271, 289].

Typically, researchers train seizure prediction models using the first chronological seizures and evaluate them on the following seizures without considering concept drifts that occur over time. These changes in data distribution may occur as a result of the seizure events, an alteration of antiepileptic drug type and/or dosage, and biological cycles (e.g., circadian rhythms), which might alter the dynamics of the brain. Dealing with concept drifts requires a different approach for training computational models. The most used technique to deal with this problem is periodically refitting the models [33, 34, 285]. The researchers retrain the models with new data adapting the weights of the approach to new data distributions. In the case of deep learning approaches, this methodology can be automatically performed because they are able to automatically extract and select the most important features directly from the data without the supervision of an expert.

The present chapter addresses some important aspects faced when developing seizure prediction models. The effect of using a deep convolutional neural network (DCNN)-based EEG artefact removal model, able to mimic manual preprocessing made by experts, on the prediction performance was explored. Furthermore, the influence of retraining the models over time to handle possible concept drifts was evaluated. Both comparisons were performed using a DCNN connected to a bidirectional long short-term memory (BiLSTM) layer (DCNN-BiLSTM) using EEG time series as input and a shallow artificial neural network (ANN) trained using established handcrafted features. In summary, this chapter comprehensively assesses the impact of denoising and dealing with the presence of concept drifts in deep learning models fed with EEG time series and in handcrafted feature-based ones.

7.2 Materials and methods

This section starts by presenting the dataset used to develop and evaluate the approaches (see Section 7.2.1). Then, the seizure prediction pipelines for each considered approach are detailed 7.2.2.

7.2.1 Dataset

A group of 41 patients with temporal lobe epilepsy (TLE) (24 males; age range: 13-67 years; mean age: 41±16 years) was selected from the EPILEPSIAE database (see 4.2). TLE is the most frequent type of focal epilepsy in adults and thus the most frequent in the database [269]. All selected data were acquired at the Epilepsy Center of the Universitätsklinikum Freiburg. Only scalp EEG was considered. It comprises 19 EEG electrodes (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, Pz), placed according to the 10-20 international system, and, using a sampling rate of 256 Hz. All patients had at least three leading seizures [259] separated by no less than 4 hours and 30 minutes. The dataset comprises approximately 5,600 hours of recording time containing 227 leading seizures from a total of 338 seizures. More details about the dataset can be found in Tables B.1 and B.2.

7.2.2 Seizure prediction pipeline

The seizure prediction pipeline begins by preprocessing the EEGs using digital frequency filters and removing experimental errors. Afterwards, the pipeline presents two branches: one in which the physiological artefacts are removed (**denoised EEG time series**), and another where they are not removed (**noisy EEG time series**). Then, features were extracted from the resulting datasets (**denoised EEG features** and **noisy EEG features**). EEG time series are used on models based on deep neural networks (DNNs), whereas EEG features are used to develop shallow ANNs. Next, each dataset is similarly divided into training and test sets. The datasets are then used to develop seizure prediction models following two different approaches: the **standard approach**, which consists of training only once and testing on the remaining seizures, and the **chronological approach**, which involves retraining after every new test seizure. Subsequently, the test set predictions were postprocessed. Finally, the performance of the approaches were computed and compared. Figure 7.1 illustrates the pipeline followed in this study. It is worth noting that this pipeline is performed individually for each patient as every model is patient-specific.

7.2.2.1 Preprocessing

Signal preprocessing was performed using the algorithm presented in Chapter 5, which mimics the manual preprocessing made by experts. Subsequently, the first

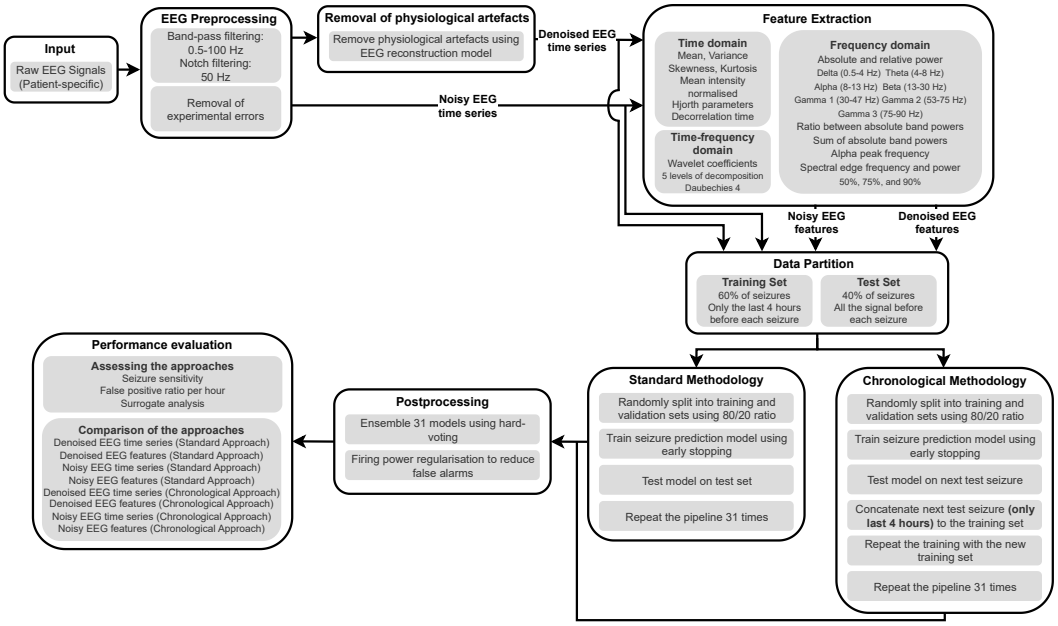


Figure 7.1: Seizure prediction pipeline comprising EEG preprocessing, feature extraction, data partition, training approaches, postprocessing, and evaluation procedure. All models were trained following a patient-specific approach. Therefore, this pipeline was repeated for each patient, individually.

30 minutes of the signal after each seizure onset were discarded to eliminate the influence of a possible postictal state [345]. Finally, the EEG segments were divided into 10-second windows. After the preprocessing methods, the data comprise approximately 4,650 hours of recording time.

7.2.2.2 Feature extraction

After preparing EEG data, established EEG features [137] were extracted using signal processing methods. Time-domain linear univariate features (mean, variance, skewness, kurtosis, mean intensity normalised, Hjorth parameters, and decorrelation time), frequency-domain linear univariate features (absolute and relative band powers of the following bands: 0.5-4 Hz (delta), 4-8 Hz (theta), 8-13 Hz (alpha), 13-30 Hz (beta), 30-47 Hz (gamma 1), 53-75 Hz (gamma 2), and 75-90 Hz (gamma 3); the ratio between every spectral band powers, the sum of all absolute band powers, the alpha peak frequency, and the spectral edge frequency and spectral edge power for 50%, 75%, and 90%), and time-frequency domain linear univariate features (wavelet coefficients computed using mother-wavelet Daubechies 4 with five levels of decomposition) were used. These features were computed for every 10-second window and every channel. Only univariate linear features were considered due to their fast computation time.

7.2.2.3 Seizure occurrence period and seizure prediction horizon

Seizure occurrence period (SOP) and seizure prediction horizon (SPH) are fundamental for developing and assessing seizure prediction models. As presented in Figure 2.11, the SPH allows the model to provide the patient with a period of time to take countermeasures before a seizure, whereas the SOP is the period when the seizure occurs. During training, preictal samples, which are samples taken before the seizure, correspond to an interval with the same duration as the SOP. The samples following the training preictal samples and ending at the seizure's onset correspond to the SPH and are not included in the analysis. This ensures that in the case of a true alarm, the patient will have an interval equal to the SPH to take countermeasures before the upcoming seizure, which is expected to occur within a period of time equal to the SOP.

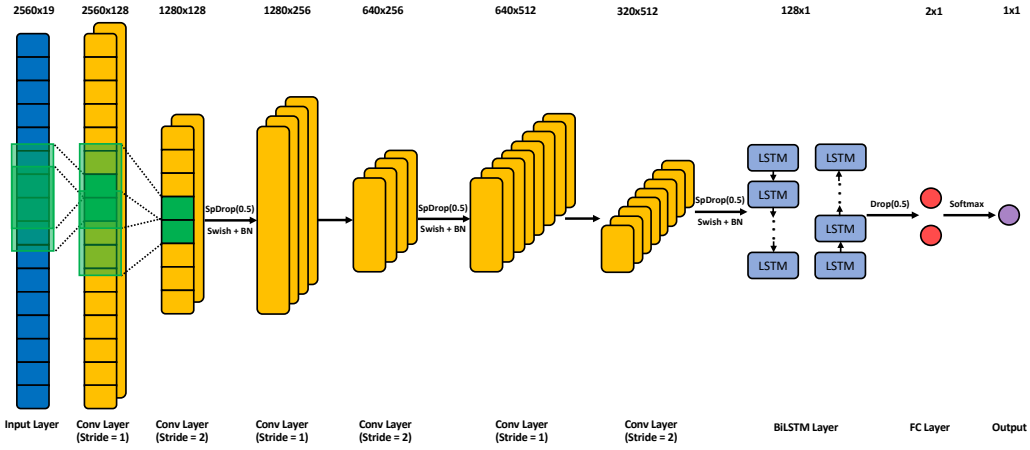
Over the years, there has been no consensus on the most optimal SOP. Extensive research has been conducted to find it, using grid search [74, 142, 143, 269] or unsupervised procedures [307, 346, 347]. According to the aforementioned papers, the optimal SOP is typically between 30 and 60 minutes. Recently, researchers have been using an SOP of 30 minutes not only because it falls within the optimal range of SOPs observed in previous findings, but also because it is short enough to avoid causing anxiety in patients [219,278,284,287,340]. As a result, an SOP of 30 minutes was used in the study. The SPH was set to 10 minutes, allowing patients to take a seizure-suppression drug to prevent the seizure [348]. Accordingly, all samples located up to 40 minutes before the seizure onset were labelled as interictal (class 0). The samples corresponding to the training SOP were considered preictal (class 1).

7.2.2.4 Training and test sets

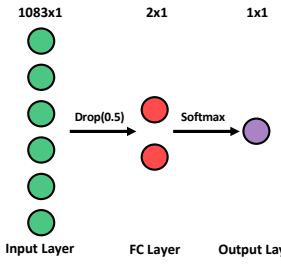
The training set was composed of 60% of the available data, and the remaining 40% was allocated to the test set. This division was performed chronologically using the first 60% of the seizures for training. Data preprocessing, as explained above, involved the removal of some data that could not be used. Therefore, there was insufficient preictal data for some seizures to be correctly predicted during testing. As a result, one test seizure from patient 52302 had to be removed, and both sets from that patient were updated to maintain the 60/40 ratio. Finally, to reduce the training computation time, only the four hours before each seizure's onset were used during the training phase. In the case of the test set, for each seizure, it included all the data from 30 minutes after the previous seizure onset until the onset of the seizure under analysis. Ultimately, the training set contained 540 hours of EEG data and 135 seizures, whereas the test set comprised approximately 1,577 hours of EEG data and 91 seizures. Details of the training and test sets can be found in Table B.2.

7.2.2.5 Artificial neural network architectures

As presented in Figure 7.2, two distinct neural networks were used: a DNN, based on the DCNN-BiLSTM architecture, which is capable of automatically processing EEG time series (**deep classifier**), and a shallow ANN based on fully connected layers with handcrafted features as input (**shallow classifier**).



(a) DNN architecture. It includes an input layer (Input Layer), six convolutional layers (Conv Layer) with a specified stride, a bidirectional LSTM layer (BiLSTM Layer), a fully connected layer (FC Layer), and an output layer (Output Layer). After each Conv Layer with a stride equal to 2, there is also a spatial dropout layer (SpDrop), an activation layer with swish function (Swish), and a batch normalisation layer (BN). After the BiLSTM Layer, there is also a dropout layer (Drop). To convert the output of the FC Layer into classification probabilities, the architecture uses an activation layer with the softmax function (Softmax). The output is the class with the highest probability. The SpDrop and Drop layers contain the dropout ratio. The dimensions of each layer's output are indicated at the top of the diagram.



(b) Shallow ANN architecture. It contains an input layer (Input Layer), a fully connected layer (FC Layer), and an output layer (Output Layer). After the Input Layer, there is a dropout layer (Drop). The architecture uses an activation layer with softmax function (Softmax) to convert output of the FC layer into classification probabilities. The output is the class with the higher probability. The dimensions of each layer's output are indicated at the top.

Figure 7.2: Neuronal network architectures used to develop seizure prediction models. **(a)** DNN, which takes 10-second EEG time series as input. **(b)** Shallow ANN, which is based on EEG features.

The architecture presented in Figure 7.2a is a DCNN-BiLSTM model. It consists of three blocks of convolutional layers and a bidirectional long short-term memory (LSTM) layer. Each block contains two convolutional layers, one of which has a stride of 2 and is used as a learnable pooling layer. Additionally, each block contains

a spatial dropout layer with a 50% rate, an activation layer with the swish function, and a batch normalisation layer. The swish function is described by Equation 7.1 where x is the input vector.

$$f(x) = x \times sigmoid(x) \tag{7.1}$$

The number of filters starts at 128 and doubles with every new block. The filter size for every layer is 3. After the convolutional blocks, the samples are processed by the bidirectional LSTM layers, each containing 64 units. A dropout layer with a 50% rate is then applied. Finally, the samples are classified using an FC layer with two neurons and an activation layer with the softmax function, which is described by Equation 7.2, where x is the input vector.

$$f(x) = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{7.2}$$

The input dimension is 2560x19, which means that each sample consists of 10 seconds of EEG acquired at 256 Hz and 19 channels. All hyperparameter values were obtained from a grid search process. Specifically, the grid search for the deep classifier was performed to find the best hyperparameters out of the following ones:

- Number of filters of the first block: [32, 64, 128];
- Filter size: [3, 5, 7];
- LSTM units: [32, 64, 128].

The architecture presented in Figure 7.2b consists of four layers: an input layer, a dropout layer with a 50% rate, a fully connected layer with two neurons, and an activation layer with the softmax function. The input dimension is 1083x1, which means that it comprises 55 features computed over 19 channels. The grid search for the shallow classifier was conducted to identify the optimal hyperparameters among the following:

- Number of neurons in the FC layer: [None, 8, 16, 32, 64, 128, 256].

No feature selection was performed before classification to enable the shallow ANN to determine which features could contribute more to the prediction performance. Both neural networks comprise dropout layers with a 50% rate, which was selected to address overfitting caused by the limited number of training samples. Both grid searches were conducted using the training set of ten randomly selected patients. To evaluate the hyperparameters, the geometric mean of sensitivity and specificity was computed using the data of the last seizure from the training set. The grid search was repeated three times for each combination and each patient training set, and the results were averaged to compare the performance and select the best hyperparameters. The selected hyperparameters were used in all patient-specific models. Detailed results can be found in Tables B.3 and B.4.

7.2.2.6 Training methodologies

Two methodologies were used to train the patient-specific models: standard and chronological. In the standard training, seizure prediction models were trained using a static training set and tested on subsequent seizures. The chronological training involved training seizure prediction models using the first set of seizures, testing on the following seizure, concatenating it (EEG and labels) to the previous training set, and repeating the process until all seizures were tested. Data partitioning and standardisation were performed each time the training set was updated. Both methodologies were repeated 31 times, resulting in 31 models that were used to perform a majority voting ensemble, whereby an odd number of models avoids ties. Furthermore, in a real-life scenario 31 different performances per patient are unfeasible. A majority voting ensemble helped to mitigate the variability of each trained seizure prediction model and produced a single model for each approach instead of 31 different models.

The neural networks were trained using batches of 64 samples, with each batch containing 32 interictal samples and 32 preictal samples. To address the imbalance between the classes, the minority class was oversampled by replicating preictal samples. The number of training epochs were 500. Adaptive moment estimation (Adam) [316] was defined as the optimisation algorithm, with an initial learning rate of 3e-4. The loss function was binary cross-entropy. Early stopping regularisation with a patience of 50 epochs was used to prevent overfitting. It requires a validation set to constantly verify whether the model is overfitting. Therefore, the training set was randomly divided into a new training set and a validation set according to an 80/20 ratio. In contrast to the data partition step, which was performed on a seizure level, the 80/20 division was performed on the samples. Training, validation, and test sets were normalised using z-score calculated based on the training samples. Table 7.1 provides a summary of the training settings.

Table 7.1: Hyperparameters used to train the neural networks.

Hyperparameter	Value
Dataset Partition	Holdout Validation 80/20
Optimisation Function	Adam
Learning Rate	3e-4
Loss Function	Binary Cross-Entropy
Epochs	500
Patience Epochs (Early Stopping)	50
Runs	31

7.2.2.7 Postprocessing

Firing power regularisation [150] was used to reduce the number of possible false alarms. As presented in 2.4.6.2, the method consists of applying a moving window with a size equal to SOP, which accumulates the predicted output of several samples. An alarm is triggered as soon as the ratio of preictal instants in the moving window exceeds a threshold of 0.5 [74, 269]. After each alarm, a refractory period [142, 143] of 40 minutes, which corresponds to the concatenation of the SPH and SOP intervals, was applied. During this period, the models could not raise any alarm. The refractory periods were implemented to prevent the patient from being overwhelmed with successive alarms in a short period of time. It is worth noting that the firing power implemented in this study is an adaptation of the method proposed by Teixeira *et al.* [150]. The method was adapted to handle temporal gaps resulting from unconnected windows after preprocessing. Thus, when there is a gap, the firing power considers it as several windows with a null value, decreasing until reaching zero when the gap is too long.

7.2.2.8 Performance evaluation

To evaluate the performance of the seizure prediction models, three metrics were used: seizure sensitivity (SS), false prediction rate per hour (FPR/h), and the number of patients with performance above chance level through surrogate analysis, considering a significance level of 0.05.

Pairwise hypothesis testing (with a significance level of 0.05) was performed to compare the different approaches that were developed. These comparisons include:

- Understanding the effect of removing physiological artefacts on the seizure prediction models;
- Comparing standard training with retraining the models over time;
- Comparing deep neural networks using EEG time series with shallow ANNs using handcrafted EEG features.

7.2.2.9 System specifications

All calculations were performed on a computer with two Intel Xeon Silver 4214 12-core 2.2 GHz, ten graphics (five NVIDIA Quadro RTX 5000 and five NVIDIA Quadro P5000), 768 GB of RAM, and Linux Ubuntu 16.04 LTS operating system. The models were developed using the Tensorflow 2.4.1 and Keras 2.4.3 libraries for Python 3.8.

7.3 Results

This section begins by analysing the results for each patient for every approach. Afterwards, the overall results are analysed and the approaches are compared. To facilitate readability, the approaches are presented in the following format:

- **Denoised EEG**$_{Standard}$: DNN, with denoised EEG time series as input, trained using the standard procedure;
- **Denoised EEG**$_{Chronological}$: DNN, with denoised EEG time series as input, trained using the chronological procedure;
- **Denoised Features**$_{Standard}$: Shallow ANN, with handcrafted EEG features, computed from denoised EEG time series, as input, trained using the standard procedure;
- **Denoised Features**$_{Chronological}$: Shallow ANN, with handcrafted EEG features, computed from denoised EEG time series, as input, trained using the chronological procedure;
- **Noisy EEG**$_{Standard}$: DNN, with noisy EEG time series as input, trained using the standard procedure;
- **Noisy EEG**$_{Chronological}$: DNN, with noisy EEG time series as input, trained using the chronological procedure;
- **Noisy Features**$_{Standard}$: Shallow ANN, with handcrafted EEG features, computed from noisy EEG time series, as input, trained using the standard procedure;
- **Noisy Features**$_{Chronological}$: Shallow ANN, with handcrafted EEG features, computed from noisy EEG time series, as input, trained using the chronological procedure.

7.3.1 Individual performance of seizure prediction models

Figure 7.3 shows the SSs and the FPR/h values of the patient-specific models evaluated on the test seizures of each patient. This figure also shows which models performed above chance level. Inspection of the results obtained for each patient leads to the following conclusions:

- For several patient-specific models, all approaches performed similarly. For example, performance above chance level was obtained for all approaches for four patients (9.8%), whereas for twelve (29.3%), no approach performed above chance level.
- For three patients (7.3%), only models trained with denoised data performed above chance level.
- The transition from standard to chronological training decreased the number of false alarms for six patients (14.6%).

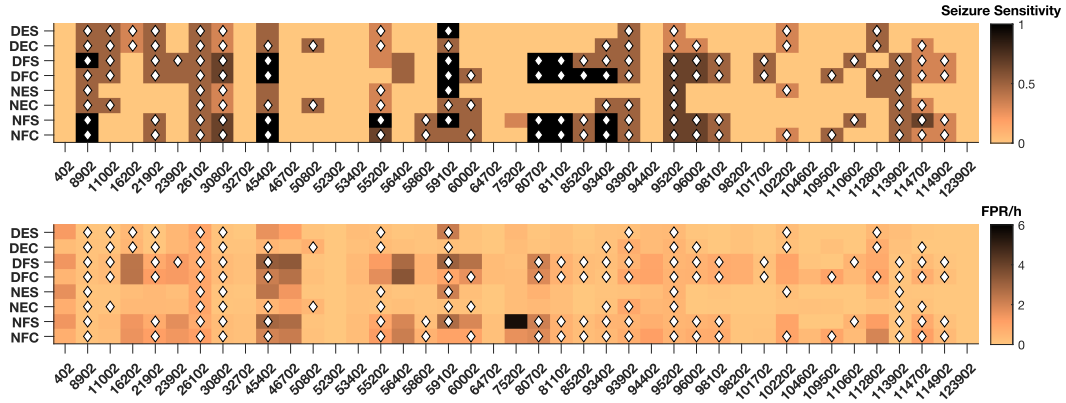


Figure 7.3: Results for each patient for DES (Denoised $\text{EEG}_{Standard}$), DEC (Denoised $\text{EEG}_{Chronological}$), DFS (Denoised $\text{Features}_{Standard}$), DFC (Denoised $\text{Features}_{Chronological}$), NES (Noisy $\text{EEG}_{Standard}$), NEC (Noisy $\text{EEG}_{Chronological}$), NFS (Noisy $\text{Features}_{Standard}$), and NFC (Noisy $\text{Features}_{Chronological}$) approaches. The top subfigure presents the SS obtained for each patient-specific model, while the bottom figure shows the FPR/h. The diamond symbol indicates that the model performed above chance level. The scales of the colours are on the right side of the subfigures.

- For some patients, only one type of neural network obtained performance above chance level: Ten patients (24.4%) in the case of shallow ANNs using features and two patients (4.9%) in the case of DNNs.
- The Noisy $\text{Features}_{Standard}$ approach obtained a very high FPR/h for one patient (2.4%).

Detailed results are presented in Tables B.5 and B.6.

7.3.2 Overall performance of seizure prediction models

Table 7.2 summarises the overall results of all developed approaches. Boxplots with the overall SSs and FPR/h values for all approaches, as well as the distributions of the results, are displayed in Figure 7.4. Additionally, it contains a bar chart with the number of patients with a performance above chance level for each approach. Table 7.3 presents the p-values obtained for pairwise statistical comparisons. Comparisons were made for SS and FPR/h metrics using one-tail Wilcoxon signed test [349] considering an α value of 0.05.

Some approaches based on EEG time series result in SSs with null medians (Figure 7.4), as those approaches scored null SSs for more than half of the patients.

With the exception of the Denoised $\text{Features}_{Standard}$ approach, all models developed with denoised data obtained higher average SSs compared to those using noisy data. Furthermore, the average FPR/h values were mostly lower in models based on denoised data. However, these comparisons did not show statistically significant differences.

Except for the Noisy $\text{Features}_{Chronological}$ approach, all models developed following a chronological methodology performed higher average SSs compared to the ones

Table 7.2: Average results of the seizure prediction models for all approaches, for all 41 patients.

Approach	Seizure Sensitivity	FPR/h	Above Chance Level (%)
Denoised EEG$_{Standard}$	0.15±0.24	0.31±0.48	12 (0.29)
Denoised EEG$_{Chronological}$	0.18±0.22	0.24±0.23	17 (0.42)
Denoised Features$_{Standard}$	0.34±0.35	0.90±0.96	21 (0.51)
Denoised Features$_{Chronological}$	0.37±0.36	0.86±0.77	22 (0.54)
Noisy EEG$_{Standard}$	0.13±0.24	0.35±0.58	8 (0.20)
Noisy EEG$_{Chronological}$	0.16±0.23	0.25±0.26	14 (0.34)
Noisy Features$_{Standard}$	0.36±0.38	0.93±1.09	20 (0.49)
Noisy Features$_{Chronological}$	0.33±0.36	0.83±0.65	21 (0.51)

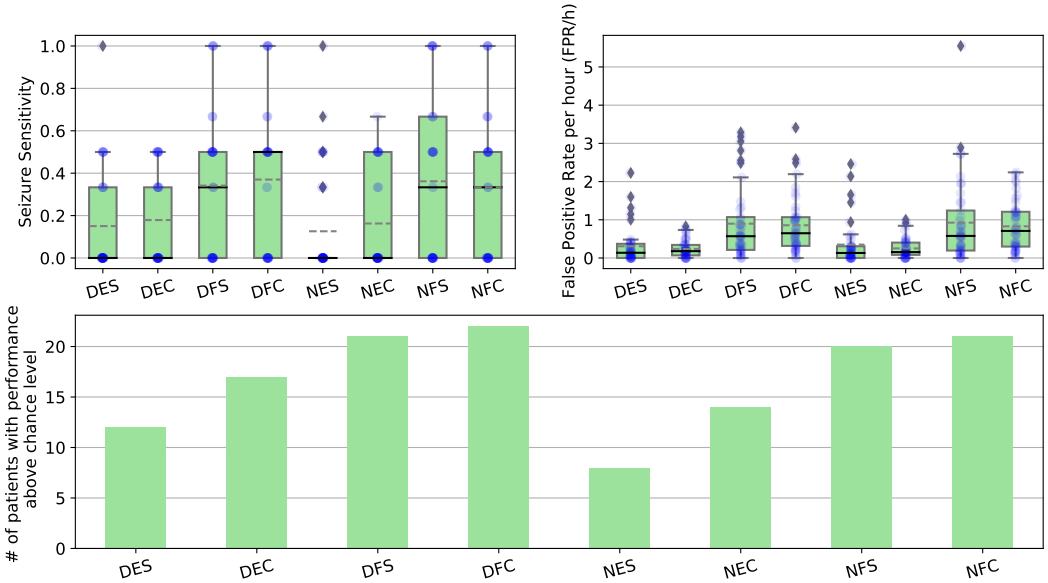


Figure 7.4: Boxplots with the overall seizure sensitivity and FPR/h for the DES (Denoised EEG$_{Standard}$), DEC (Denoised EEG$_{Chronological}$), DFS (Denoised Features$_{Standard}$), DFC (Denoised Features$_{Chronological}$), NES (Noisy EEG$_{Standard}$), NEC (Noisy EEG$_{Chronological}$), NFS (Noisy Features$_{Standard}$), and NFC (Noisy Features$_{Chronological}$) approaches. Continuous black lines represent medians, dashed grey lines correspond to the averages, diamonds symbolise outliers, and the distributions of the results for each patient are presented as blue circles. Bar charts show the number of patients' models with performance over chance using surrogate analysis.

following the standard procedure. The average FPR/h values obtained for models based on the chronological methodology were lower than those for the standard training, whereas the medians showed the opposite behaviour. This was due to the high number of outliers occurring on the approaches based on standard training. Nevertheless, all comparisons did not yield statistically significant differences.

Models based on DNNs returned lower average SSs than those obtained for the models based on shallow ANNs. However, these lower SSs were combined with low FPR/h values, meaning that, on average, deep learning models were more conservative on triggering an alarm. All the comparisons yielded statistically significant differences for both metrics.

The number of patients with performance above chance level increased when

Table 7.3: P-values obtained for the statistical comparisons performed between all developed approaches using SS and FPR/h values. The comparisons were performed using one-tail Wilcoxon signed test, considering an *alpha* of 0.05. For SS, the p-values correspond to the probability of the distribution of group B being greater than the distribution of group A. For FPR/h, the p-values correspond to the probability of the distribution of group B being lower than the distribution of group A, except for the comparison between DNNs (EEG time series) and shallow ANNs (EEG features). In this particular case, the FPR/h values obtained for group B were higher, so the p-values correspond to the probability of group B being greater than group A. Bold values correspond to the statistically significant comparisons.

			P-values	
Approaches	**Group A**	**Group B**	**SS**	**FPR/h**
Noisy EEG / Denoised EEG	**Noisy EEG**$_{Standard}$	**Denoised EEG**$_{Standard}$	0.165	0.314
	Noisy EEG$_{Chronological}$	**Denoised EEG**$_{Chronological}$	0.315	0.433
	Noisy Features$_{Standard}$	**Denoised Features**$_{Standard}$	0.645	0.983
	Noisy Features$_{Chronological}$	**Denoised Features**$_{Chronological}$	0.197	0.781
Standard Training / Chronological Training	**Denoised EEG**$_{Standard}$	**Denoised EEG**$_{Chronological}$	0.167	0.684
	Noisy EEG$_{Standard}$	**Noisy EEG**$_{Chronological}$	0.144	0.464
	Denoised Features$_{Standard}$	**Denoised Features**$_{Chronological}$	0.146	0.802
	Noisy Features$_{Standard}$	**Noisy Features**$_{Chronological}$	0.804	0.698
DNNs (EEG Time Series) / Shallow ANNs (EEG Features)	**Denoised EEG**$_{Standard}$	**Denoised Features**$_{Standard}$	**< 0.001**	**< 0.001**
	Denoised EEG$_{Chronological}$	**Denoised Features**$_{Chronological}$	**< 0.001**	**< 0.001**
	Noisy EEG$_{Standard}$	**Noisy Features**$_{Standard}$	**< 0.001**	**< 0.001**
	Noisy EEG$_{Chronological}$	**Noisy Features**$_{Chronological}$	**0.002**	**< 0.001**

shifting from noisy to denoised data and from the standard to chronological training. The increase was greater in DNNs using EEG time series compared to the shallow ANNs using handcrafted features. However, the shallow ANNs obtained a higher number of patients performing above chance level.

In addition to the machine learning architectures used in this study, the effectiveness of denoising data and chronological training using algorithms presented by other researchers, including a DCNN using spectrograms proposed by Truong *et al.* [280] and a logistic regression using handcrafted features proposed by Karoly *et al.* [76], were also evaluated. These architectures were selected to verify whether the obtained findings were observed in other types of classifiers. The results obtained using the model proposed by Truong *et al.* [280] were similar to those obtained in the performed study. In the case of the model proposed by Karoly *et al.* [76], it was found that using denoised data improved the performance. However, transitioning from standard to chronological training did not result in any improvement. The implementation details and results are described in Section B.3.5.

7.4 Discussion

The impact of two essential aspects for developing patient-specific seizure prediction models were analysed: denoising EEGs and retraining the models over time. The prediction models were developed using DNNs with EEG time series as input and shallow ANNs using widely used EEG features.

The EEG artefact removal model was already proposed and evaluated in Chapter 5 regarding its capacity to reconstruct EEGs. As a next step, it was wanted to evaluate how far the artefact removal method can improve the results for seizure

prediction. For most cases, using the artefact removal model to denoise EEGs before developing the seizure prediction models resulted in an improvement in SSs, FPR/h values and the number of patients with performance above chance level. In the case of DNNs, removing artefacts using the automatic denoising model led to a small reduction in the number of outliers regarding FPR/h. This was expected since artefacts can change signal characteristics and mask some patterns that could be important to assess seizure susceptibility [19–21].

Concerning retraining over time, different behaviours were observed for different metrics. SS did not always increase from standard to chronological training. However, FPR/h and the number of patients with a performance above chance level improved after retraining. Thus, it was concluded that the models benefited from chronological training, either by having more training data available or by adapting to possible concept drifts that occur over time (see Figure B.1). This resulted in a reduction in the number of false alarms and an increase of the number of patients with performance above chance level [33, 34, 285].

Comparing both model types, although deep learning models returned lower SSs, they also yielded lower FPR/h values by being more conservative in firing alarms. The number of patients performing above chance level was also lower for the deep learning models, which is mainly attributed to their lower SS. Models that did not predict any test seizure could not be validated using surrogate analysis, thus leading to a lower number of patients with performance above chance level.

It is worth noting that although both neural network types improved after denoising and chronological training, the improvement was more evident for deep learning models than for shallow ANNs. DNNs are data-driven architectures [170]. Consequently, features are automatically extracted from the EEG time series. On the other hand, feature-based models are trained using values obtained from established equations based on research knowledge acquired over the years. For this reason, each retraining only adapts the model weights used for the classification. Therefore, deep learning architectures, adapting to the input training data distribution, may be more affected by the quantity and quality of input data and the temporal proximity to the next seizure.

7.4.1 Comparison with state of the art studies

After analysing the results, they were compared with other studies that used scalp EEG data from the EPILEPSIAE database to develop seizure prediction models [142, 143, 276, 284]. This paragraph is focused on the ones that do not contain patient identification [276, 284]. Direito *et al.* [276] applied a simple preprocessing method using digital filtering and developed a seizure prediction model based on multiclass support vector machines (SVMs) using handcrafted features. They used firing power regularisation with a threshold of 0.5 to smooth the output of

the classifiers. They considered a set of SOPs between 10 and 40 minutes and an SPH of only 10 seconds. They reported an SS of 39% and an average FPR/h of 0.21. The percentage of patients performing above chance level was approximately 10%. Nevertheless, it is worth noting that they validated their models using the analytical random predictor [1] instead of the surrogate analysis. Although their SS and FPR/h are better than the ones obtained in this study, it should be noted that they considered an SPH of just 10 seconds, which may not be sufficient time for a patient to take countermeasures before an upcoming seizure. Truong *et al.* [284] also performed a simple preprocessing method using digital filtering. They developed a generative adversarial network (GAN) using EEG time series as input and considered an SOP of 30 minutes and an SPH of 5 minutes. They reported an average area under the receiver operating characteristic curve (AUC) of 0.65 for 30 patients and performance above chance level for 23 patients using the Hanley-McNeil AUC Test to compare AUC scores with an AUC of 0.5. Since the authors did not make these metrics available, it is not possible to compare their study with this one regarding SS and FPR/h. In terms of the number of patients with performance above chance level, they obtained a better result than the one obtained in this study. However, it should be pointed out that they did not use the same statistical evaluation method, which precludes fair comparisons.

Concerning the studies with detailed patient information, Pinto *et al.* [142, 143] published two papers with seizure prediction models based on evolutionary algorithms, with a different number of patients analysed in each one. In both studies, the algorithms were trained following a chronological approach. In this chapter, data from some patients who were also included in both Pinto *et al.* studies were used. In both of their experiments, the preprocessing was simple, using only digital filters. In the first study [142], they used data from 19 patients and obtained an average SS of 0.38±0.19 and an average FPR/h of 1.03±0.84, using an SOP of 30 minutes and an SPH of 10 minutes. Performance above chance level was obtained for 32% of the patients. In the second study [143], they used data from 93 patients and obtained an average SS of 0.16±0.11 and an average FPR/h of 0.21±0.08 using a set of SOPs between 20 and 50 minutes and an SPH of 10 minutes. 32% of the patients obtained performance above chance level. Both studies used firing power regularisation with a threshold of 0.7. When analysing individual patients, it was found that twelve patients (29.3%) performed above chance in both Pinto *et al.* studies and in some of the presented approaches. These patients should be selected to further explore preictal changes, as different methods performed similarly. The approaches presented in this chapter and Pinto *et al.* models did not perform above chance level for six patients (14.6%). Common failed predictions could mean no preictal changes at least 10 minutes before the onset of any tested seizures. Pinto *et al.* obtained performance above chance for four patients (9.8%) in at least one of their studies, whereas none of the developed approaches performed above chance level for them.

Also, there were thirteen patients (31.7%) for whom it was obtained models performing above chance level, whereas none of Pinto *et al.* studies obtained significant results. These differences could be related to data preparation details or the type of algorithm used. In any case, the results of the seizure prediction studies seem to be coherent regarding the obtained statistics, even if the approaches were different, which supports the statements reported by Müller *et al.* [298] about the high number of false positives generated by different types of seizure prediction models. Detailed results obtained by Pinto *et al.* [142, 143] are presented in Table B.7.

7.4.2 Study limitations

The study has some limitations that should be highlighted. The first one concerns the use of EEGs acquired in pre-surgical conditions. Pre-surgical conditions involve drug type and/or dosage alteration and possible sleep deprivation as part of the clinical evaluation process, which may cause more concept drifts than expected. Additionally, if the signals are acquired outside the hospital for months, they may contain even more artefacts than the ones seen in the analysed data because there are no technicians to constantly check the equipment. Therefore, pre-surgical data do not fully simulate the daily behaviour of the patients [18], and care should be taken before generalising these results to real-life situations.

Another limitation is the number of seizures per patient. The average number of total seizures per patient was 5.51, whereas the average number of tested seizures was 2.21. The low number of seizures limited the evaluation of the approaches since some patients had only one testing seizure restricting the obtained SS to 0.00 or 1.00. This large difference in possible SSs produced large standard deviations which may have restricted the results of the statistical comparisons. Furthermore, as the amount of data was low, the improvement obtained by training periodically may have been due to the increase in available information rather than the change in concept. Thus, a higher number of seizures would allow for a better evaluation of the effect of retraining over time or even to test other different approaches to handle concept drifts [350].

The seizures used in this study were manually annotated by experts. In real-time scenarios, manual annotation of seizures can be challenging since it is difficult for physicians to review all the acquired EEGs. The solution to this restriction could be the one presented by Pal Attia *et al.* [351], where data would be sent to a cloud along with the outputs of a seizure prediction model, a seizure detection model, and annotations sent by the patient. In this way, the technicians could quickly review only the events that were noted and eliminate all those that were not seizures. Then, the model could be retrained using data from the new seizure events.

The fixed duration of the SOP for all patients was also a constraint. In this case, a fixed SOP of 30 minutes was used for all patients, which is in line with

the SOP duration used in other studies [74, 219, 269, 278, 284, 287, 340]. SOP and SPH determine the considered preictal samples. Therefore, the considered preictal interval was also limited. The preictal interval of seizures could last from few minutes to several hours. Thus, EEG characteristics should be carefully inspected in order to not consider preictal patterns as interictal data [346, 347]. However, the inspection of the optimal SOP was not in the focus of this study.

The fifth limitation was the number of hours used to train the prediction models. In this case, only four hours per seizure were used to train the models. This time was selected to overcome the high imbalance between interictal and preictal periods as well as to reduce computation time. However, it could limit the training of the models because they might not be able to learn critical long-term patterns such as the circadian cycle [77, 342]. Lastly, the models were trained to trigger alarms once they detect any preictal changes in the data. Therefore, they can not measure seizure susceptibility over time and decide if one alarm is more important than others. Additionally, the brain has its own regulatory system. There may be some scenarios in which the model predicts a seizure due to the high seizure susceptibility state, but the brain triggers seizure control mechanisms before reaching the "point of no return" [352,353]. This new perspective has prompted the change from developing seizure prediction models to designing seizure forecasting frameworks. The latter allows researchers to better understand what is happening in the brain because forecasting approaches output seizure risk instead of just alarms. Furthermore, forecasting models are generally less penalised during evaluation since their response is continuous and not based on "all-or-nothing". For example, in the presented approaches, if an alarm were raised 41 minutes before the seizure onset, it would be evaluated as a false alarm even if it was essentially a correct prediction [16].

7.4.3 Final reflections

This chapter explores two essential aspects that should be taken into consideration before developing seizure prediction models: the impact of performing a robust preprocessing to remove noisy artefacts, such as ocular artefacts, from EEGs; and the importance of periodically retraining the seizure prediction models to address the possible presence of concept drifts. The importance of these two variables was investigated using two models: one based on DNNs with EEG time series as input and another based on shallow ANNs using handcrafted EEG features computed using signal processing techniques. The results evidenced that the performance of deep learning approaches improved after denoising the data and periodically retraining the models. On shallow ANNs with handcrafted features as input, the effect of denoising and/or retraining was barely noticeable, which may indicate that handcrafted features were more robust to changes in the data. The results also showed that shallow ANNs using handcrafted features were able to predict twice as many

seizures as deep learning models. However, the number of false alarms was generally approximately four times higher as compared to DNNs. Therefore, when comparing both types of models, it was not possible to conclude which performs better. As future work, these approaches should be tested with more patients and prospective data. Furthermore, it would be beneficial to test these approaches with more test seizures and with data collected over longer periods than just a few days, such as signals obtained by ultra long-term acquisition systems [57].

Chapter 8

Transfer learning on seizure prediction: Does information from several patients improve patient-specific approaches?

This chapter presents the development of seizure prediction models using a transfer learning approach. Section 8.1 presents the study context. Section 8.2 describes the datasets and the methods used in the study. Section 8.3 presents the results and subsequent analysis. Section 8.4 discusses the obtained results, presents the advantages and the limitations of the developed approaches, and provides some final reflections.

8.1 Study context

Deep learning architectures are purely data-driven approaches typically with a large number of parameters to tune. Therefore, those models require large datasets to find patterns in the data and, consequently, accurate generalisations. It is undoubtedly a problem in seizure prediction since models should be trained in a patient-specific manner. Transfer learning has been used as a solution to address this challenge. In the case of deep neural networks (DNNs), transfer learning is mainly used to initialise the weights using large datasets. While these large datasets may not necessarily originate from the same research field as the target dataset, it is expected that both datasets involve using the same type of extracted features. Otherwise, fine-tuning using the target dataset may be difficult or even ineffective [354]. In seizure prediction, researchers have presented promising results using different forms of transfer learning, for example, training a deep convolutional autoencoder (DCAE) with data from different patients to initialise the weights of patient-specific mod-

els or using DNN trained using ImageNet dataset to improve seizure prediction approaches [35, 181, 189, 286, 355].

The present study evaluates the possibility of using electroencephalograms (EEGs) from patients with epilepsy from a large database to improve patient-specific seizure prediction models. Specifically, the performance of patient-dependent approaches developed from scratch are compared with approaches in which weights are already initialised using a DCAE trained with data from other patients (see netword-based transfer learning approach explained in Section 2.7). Finally, the patient-independent model is openly provided to be used by other researchers.

8.2 Materials and methods

This section presents the steps followed to develop and evaluate the proposed methodology. The datasets used to develop and evaluate the approach are presented in Section 8.2.1. Then, the model development pipeline is detailed (see Section 8.2.2).

8.2.1 Datasets

In this study, two datasets were used: data from patients present in the EPILEPSIAE database described in Section 7.2.1 and data from new patients available in the Epilepsy Center of the Universitätsklinikum Freiburg. To make the explanation simpler, the first dataset is named EPILEPSIAE dataset, while the second is named Personal dataset. The Personal dataset contains scalp EEG from 24 patients with temporal lobe epilepsy (TLE) (11 males; age range: 15-67 years; mean age: 38±14 years). The signals comprise the same 19 electrodes used in the EPILEPSIAE dataset, placed according to the 10-20 international system. Data were collected using a sampling rate of 250 Hz. However, data were converted to 256 Hz before proceeding to be coherent with the EPILEPSIAE dataset. All patients had at least three leading seizures separated by no less than 4.5 hours. The dataset comprises approximately 4418 hours of recording time, containing 152 leading seizures from a total of 273 seizures. More details about the dataset can be found in Table C.1. It is worth noting that informed written consent was obtained from the patients and the parents or legal guardians of patients under 18.

8.2.2 Methodology

The pipeline is similar to the one presented in Section 7.2.2. It begins by preprocessing the EEGs from both datasets using digital frequency filters, removing experimental errors and physiological artefacts. Then, the EEGs from the EPILEPSIAE dataset were used to optimise a DCAE. Afterwards, the Personal dataset was used to train two approaches: the **standard approach**, which consists of training the DNN from scratch, and the **transfer learning approach**, which uses the weights from

the DCAE to initialise the convolutional layers. Subsequently, the test set predictions were postprocessed. Finally, both approaches were evaluated and compared. Figure 8.1 illustrates the pipeline followed in this study. It should be pointed out that the seizure prediction models were trained following a patient-specific approach.

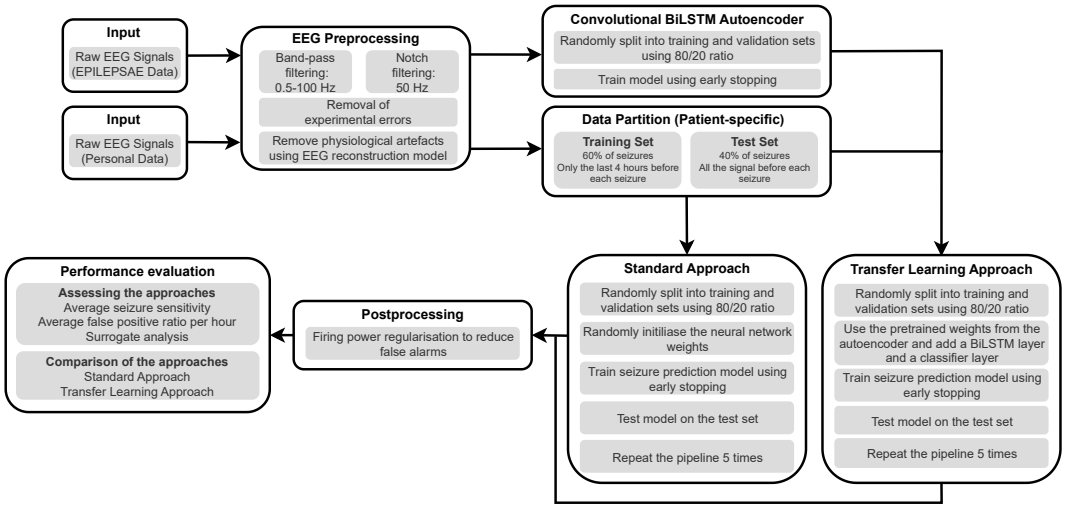


Figure 8.1: Seizure prediction pipeline comprising the EEG preprocessing, the training of the DCAE using the EPILEPSIAE dataset, the split of the Personal dataset, the standard and transfer learning approaches, the postprocessing, and the evaluation procedure. The models were developed following a patient-specific approach.

8.2.2.1 Signal preprocessing

The signal preprocessing was the same presented in Section 7.2.2.1. After the preprocessing methods, the Personal dataset comprises approximately 3,290 hours of recording time, whereas the EPILEPSIAE dataset comprises about 4,744 hours.

8.2.2.2 Seizure occurence period and seizure prediction horizon

The main goal of this study was to compare the performance of seizure prediction models developed following a transfer learning approach with models developed without using pretrained weights similar to the standard methodology presented in Section 7.2.2. The duration of the seizure occurrence period (SOP) and the seizure prediction horizon (SPH), used before, were maintained. The SOP lasts 30 minutes, and the SPH 10 minutes.

8.2.2.3 Transfer learning methodology

The DCAE is an encoder-decoder architecture. The encoder is equal to the deep convolutional neural network (DCNN) part of the DNN presented in Section 7.2.2.5. The decoder contains layers responsible for converting the data to the original size. After optimising the DCAE using the data from all the patients available in the EPILEPSIAE dataset, the decoder was removed, and the weights of the encoder

were used to initialise the weights of the convolutional layers of the seizure prediction models. A bidirectional long short-term memory (BiLSTM) and a classifier layer were added on top of the encoder to include the ability to explore temporal information and to classify samples either as interictal or preictal. The weights of the convolutional layers were frozen, meaning that during the training of the patient-specific approaches, only the weights of the BiLSTM and the classifier layer were optimised. Both architectures are presented in Figure 8.2. More details are described in Section C.2.

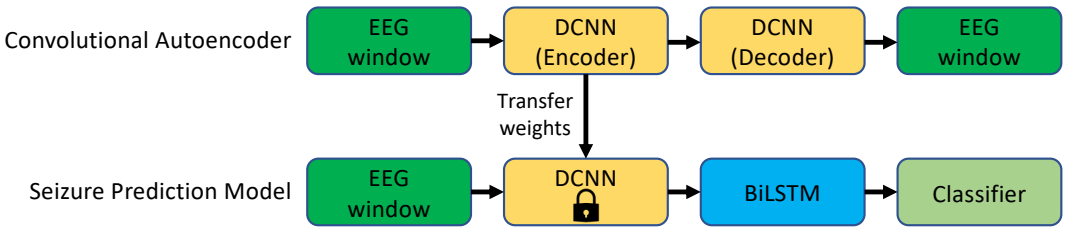


Figure 8.2: Transfer learning approach. On top, there is the DCAE. It contains an encoder with a DCNN that extracts patterns from the data and a decoder with a DCNN that converts the data to the original size. The output is expected to be equal to the input. The encoder weights are transferred to the seizure prediction model, and a BiLSTM and a classifier layer (fully connected layer with a softmax function) are added.

8.2.2.4 Training and test sets

For training the DCAE, the training set contained 80% of the EPILEPSIAE dataset, and the remaining 20% was allocated to the validation set to control the overfitting while training. The split was randomly performed. The training set contained about 3,795 hours of EEG data, whereas the validation set comprised approximately 949 hours of EEG data.

For training the seizure prediction models, the training set comprised 60% of the data for each patient of the Personal dataset, whereas the test set contained 40% of the signals. The split was chronologically performed, considering the seizures and not the samples, meaning that the first 60% of the seizures were used for training. As presented in Section 7.2.2.4, only the four hours before each seizure's onset were used during the training phase, whereas the test set, for each seizure, included all the data from 30 minutes after the previous seizure to the onset of the seizure under analysis. The training set contained 356 hours of EEG data and 89 seizures. The test set included approximately 1,123 hours of EEG data and 62 seizures. Details about the training and test sets can be found in Table C.2.

8.2.2.5 Training methodologies

All sets were standardised using the average and standard deviation from the training set of the EPILEPSIAE dataset. The models were trained using a batch size of 2,048. The number of epochs was 2,000. Early stopping regularisation with a patience of

200 epochs was used to prevent overfitting. Adaptive moment estimation (Adam) function with a learning rate of 3e-4 was used to optimise the parameters. The loss function was the mean squared error function.

The patient-specific seizure prediction models were developed following two different approaches: the standard approach and the transfer learning approach. In the standard approach, the DNNs were trained from scratch, meaning that all the weights were randomly initialised before the optimisation. Additionally, the standardisation of the sets was performed using the average and standard deviation from the training set. In the transfer learning approach, the weights obtained from the optimisation of the DCAE were used to initialise the parameters of the seizure prediction models. In this case, the standardisation was performed using the statistics from the training set used to optimise the DCAE. Both methodologies were repeated five times. It was impossible to perform 31 runs as in Chapter 7 due to computational power restrictions. The hyperparameters used to train both approaches were the same that were used in Chapter 7. Holdout validation 80/20 was used to divide the training set into training and validation subsets. The models were trained using balanced batches of 64 samples. The number of epochs was 500. Early stopping regularisation with a patience of 50 epochs was used to prevent overfitting. Adam was used to optimise the DNNs with a learning rate of 3e-4. The loss function was the binary cross-entropy function.

8.2.2.6 Postprocessing

The postprocessing was the same presented in Section 7.2.2.7. Thus, it consisted of the firing power regularisation [356] with a threshold of 0.5 adapted to consider time instead of the number of samples on the denominator.

8.2.2.7 Performance evaluation

The seizure sensitivity (SS), the false prediction rate per hour (FPR/h), and the number of patients with performance above chance level obtained through surrogate analysis, considering a significance level of 0.05, were used to evaluate the seizure prediction models. It should be noted that the evaluation metrics consist of the average over the five runs. The majority voting ensemble was not used because as the number of runs was low, the results could depend on the result of only one run.

Pairwise hypothesis testing (with a significance level of 0.05) was performed to compare the standard approach with the transfer learning approach.

8.2.2.8 System specifications

The DCAE was developed in a computer with two Intel Xeon Silver 4214 12-core 2.2 GHz, five NVIDIA Quadro RTX 5000, five NVIDIA Quadro P5000, 768 GB of

RAM, and Linux Ubuntu 16.04 LTS. The used package was Tensorflow 2.4.1 from Python 3.8.

The seizure prediction models were developed using a computer with one AMD Ryzen Threadripper 3970X 32-core 3.7 GHz, one NVIDIA GeForce RTX 3060 Ti, 128 GB of RAM, and Windows 10. The used package was Tensorflow 2.6.0 from Python 3.7.

8.3 Results

This section presents the results obtained for standard and transfer learning approaches evaluated on the test set of each patient.

8.3.1 Individual performance of seizure prediction models

Figure 8.3 illustrates the SSs and the FPR/h values of the patient-specific models. It also presents the patients for whom models performed above chance level. The

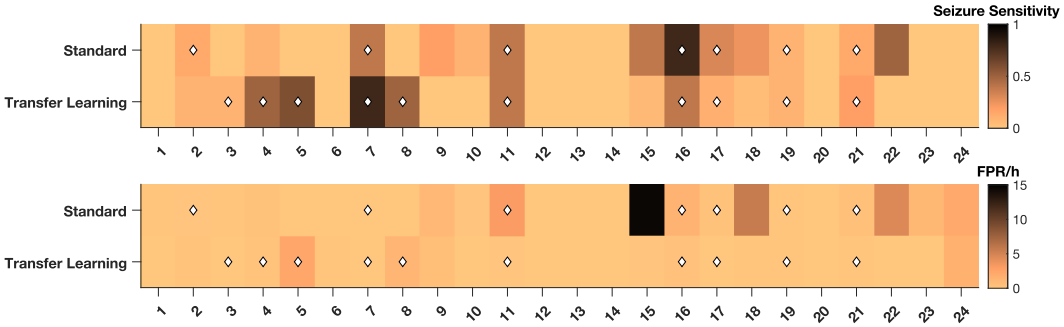


Figure 8.3: Results for each patient for standard and transfer learning approaches. The top subfigure presents the SS obtained for each patient-specific model, while the bottom figure shows the FPR/h. The diamond symbol indicates that the model performed above chance level. The scales of the colours are on the right side of the subfigures.

figure yields the following conclusions:

- The standard approach presents thirteen patients (54.2%) with an SS higher than zero, whereas the transfer learning approach obtained twelve patients (50.0%). Although the standard approach obtained more patients with an SS higher than zero, only seven (29.2%) achieved a performance above chance level. It was lower than the ten validated patients (41.7%) obtained using the transfer learning approach.
- From the eleven validated patients (45.8%), only one, that was validated by the standard approach, was not validated by the transfer learning approach.
- The usage of transfer learning weights decreased the FPR/h in seventeen patients (70.8%), especially in patients 15, 18, and 22.
- The standard approach obtained a very high FPR/h for the patient 15. It happened because the model could not correctly converge (see Figure C.1).

Detailed results are available in Table C.5.

8.3.2 Overall performance of seizure prediction models

Table 8.1 contains the average results of the standard and transfer learning approaches. Boxplots with the overall SS and FPR/h values, as well as the distributions of the performances, are displayed in Figure 8.3. Additionally, it contains the statistical comparison between both approaches for SS and FPR/h. Comparisons were made using a one-tail Wilcoxon signed test [349] considering an α value of 0.05. Furthermore, it contains a bar chart with the number of patients performing above chance level for each approach.

Table 8.1: Average results of the seizure prediction models for both approches, for all 24 patients.

Approach	Seizure Sensitivity	FPR/h	Above Chance Level (%)
Standard	0.16±0.21	1.51±3.15	7 (0.29)
Transfer Learning	0.16±0.23	0.35±0.61	10 (0.42)

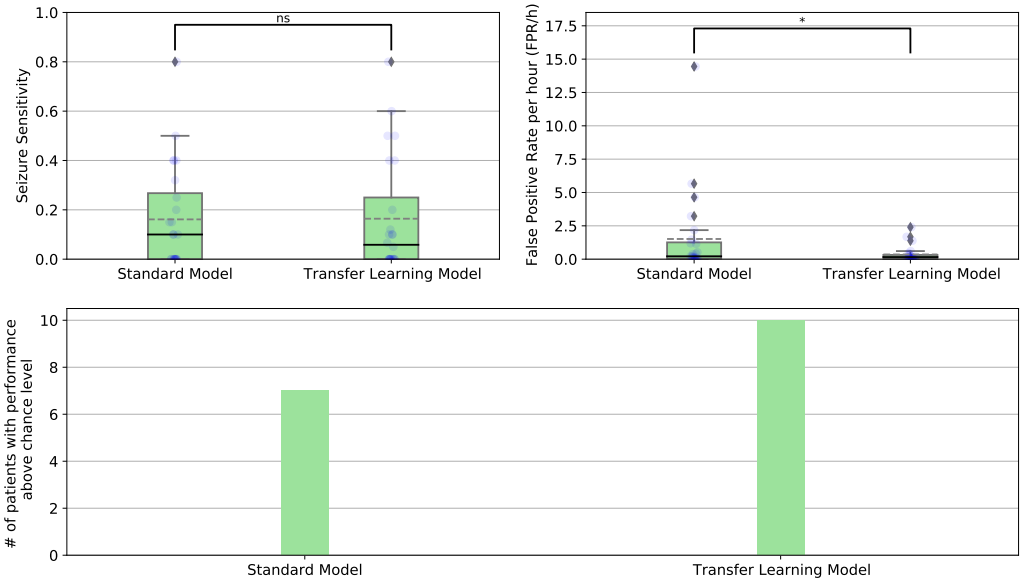


Figure 8.4: Boxplots with the overall SS and FPR/h for the standard and transfer learning approaches. Continuous black lines represent medians, dashed grey lines correspond to the averages, diamonds symbolise outliers, and the distributions of the results for each patient are presented as blue circles. Statistical significance indicators are placed above the boxplots: the ***** means that the p-value is below 0.05, and **ns** means not significant. Bar charts show the number of patients' models with performance over chance using surrogate analysis.

Both approaches obtained a similar average SS. Therefore, the comparison did not show statistically significant differences (p-value = 0.548). Regarding the FPR/h, the transfer learning approach produced an average value about four times lower than the one produced by the standard approach. The comparison showed statistically significant differences (p-value = 0.008). The transfer learning approach

obtained more three patients than the standard approach with a performance above chance level.

8.4 Discussion

Research in seizure prediction has evidenced that models should be trained following a patient-specific approach. It causes a high limitation because the seizures are usually very rare events [125]. The low number of available data may worsen the performance of deep learning models which only achieve their satisfactory performance when the number of samples is high [170]. Therefore, transfer learning may be a solid way to handle that limitation.

This study analysed the impact of using a transfer learning approach to improve patient-specific seizure prediction models. A DCAE was trained using EEG data from 41 patients available in the EPILEPSIAE database. Afterwards, the learned weights were used to develop patient-specific seizure prediction models for patients from a Personal dataset. This approach was compared to a standard approach, in which the weights were optimised from scratch.

Results evidenced that the average SS was similar in both approaches. Regarding the average FPR/h, the value obtained by the transfer learning approach was about four times lower than the standard approach. It happened due to a large number of patients with an FPR/h higher than one in the case of the standard approach, especially patient 15, whose models could not correctly converge when training from scratch. The combination of those results led to a higher number of patients with a performance above chance level in the case of the transfer learning approach. Although the DCAE was trained with data from patients from a different database with different acquisition systems, it was concluded that the models benefited from using pretrained weights.

Different algorithms have been proposed to verify whether the transfer learning approach improves the seizure prediction results [35, 181, 286, 293, 355]. The authors presenting such approaches reported better performance considering not only the seizure prediction but also the computational time, being in line with the evidence obtained in this experiment.

All studies found in the literature used only one database to develop the patient-specific models considering the leave-one-out cross-validation (CV) approach. They trained the main model using N-1 patients and optimised it for the remaining one. Based on the available evidence, this is the first study using an external database to improve patient-specific seizure prediction models. Therefore, this study is helpful in demonstrating to the scientific community that it is possible to combine data from different databases to obtain more robust seizure prediction models.

8.4.1 Study limitations

The study contains some of the already mentioned limitations presented in Section 7.4.2, such as: EEGs acquired in pre-surgical conditions; a low number of test seizures per patient on patient-specific approaches; fixed duration of the SOP; the low number of training hours used for optimising the patient-specific models; and the fixed interval used to consider a seizure correctly predicted.

The DCAE is based on the architecture presented in Chapter 7. Therefore, compared with other architectures, it is a small one. The architectures used for transfer learning in the image processing field contain dozens, if not hundreds of layers that allow a deeper learning of the task [184, 186, 357]. Thus, a search for the optimal architecture should be faced in the future.

The number of patients in the EPILEPSIAE dataset could have been larger. Only 41 patients out of 227, with scalp EEG, were used. It is the difference between training with only 4,744 hours and training with about 37,000 hours of data [269]. Increasing the number of patients could lead to a broader range of information used for optimising the model. Consequently, this may enhance the generalisation of the learned features. However, a substantial increase in available resources would be necessary to accommodate this expansion.

8.4.2 Final reflections

This chapter presents a transfer learning approach to develop patient-specific seizure prediction models based on DNNs. Therefore, it explores the ability to use data from several patients from an external database to improve the optimisation of seizure prediction models. Results showed that transfer learning allows obtaining about four times less false alarms while maintaining the same ability to predict seizures as when trained from scratch. Thus, it was concluded that the information obtained from other patients external to the database used for seizure prediction improved the models. Therefore, the DCAE will be available on a public repository so that other researchers may use it to develop their approaches. Future work should focus on the limitations presented above. The models should be tested with more seizures and over a more extended period, e.g., using ultra-long-term acquisition systems [57]. Also, the DCAE should be trained with as many patients as possible, perhaps even pooling data from other databases to increase the diversity of knowledge, e.g. CHB-MIT [254] or SeizeIT2 [258].

Chapter 9

Conclusions

This chapter summarises the main contributions of this thesis. It also discusses future directions that should be considered.

9.1 Summary of the main contributions

The main objective of this thesis is to make contributions to the advancement of adaptive seizure prediction models. Firstly, automatic electroencephalogram (EEG) denoising approaches were developed. Subsequently, the most efficient ones were used to prepare the data for the development of personalised seizure prediction models based on artificial neural networks (ANNs) and deep neural networks (DNNs). An important outcome was the consideration of a transfer learning approach to enhance the optimisation of patient-specific models using DNNs.

In Chapter 4, the dataset used to develop the EEG artefact removal models was created. EEGs were acquired from 25 patients available in the EPILEPSIAE database. Then, 20 patients were selected for the training set and 5 for the test set. The EEGs were converted into independent components (ICs) using the independent component analysis (ICA) algorithm. Finally, the ICs were visually analysed and classified as brain or noisy. Finally, a dataset of 612.68 hours of data, corresponding to 77,426 ICs, was created. Of these, 61,092 belong to the training set, whereas 16,334 are part of the test set.

In Chapter 5, an EEG artefact removal approach was proposed. It was based on deep convolutional neural networks (DCNNs), which takes as input the noisy EEGs and reconstructs them without EEG artefacts such as eye blinks. The advantage of this approach is being able to automatically remove artefacts that mask important brain information in less than a second for a 10-minute segment. Therefore, it could be used to prepare the scalp EEGs prior to their analysis, for example, before seizure prediction. The developed model showed an root mean squared error (RMSE) of 4.83 μV, an relative root mean squared error (RRMSE) of 0.52, a Pearson correlation coefficient (PCC) of 0.86, and an signal-to-noise ratio (SNR) difference of 8.81.

Compared to other models presented in the state of the art, the developed approach outperformed in terms of both reconstruction metrics and processing time.

In Chapter 6, ICs were used to develop ICs classifiers. During the visual inspection of the ICs, three different information sources were used, the time series (temporal information), the power spectrum densitys (PSDs) (spectral information), and the topographic maps (spatial information). Using these three sources, different classifiers were developed in order to verify which source was the most important. Besides that, transfer learning methods were also explored to verify whether pretraining with a large dataset improved the classification of ICs from a smaller database. The results showed that using the three sources of information is necessary to obtain the highest performance. Additionally, using transfer learning further improved the performance of classifiers for ICs from a smaller database.

In Chapter 7, the main goal was to develop seizure prediction models. Different approaches were tested, such as comparing models trained with data denoised using the model presented in Chapter 5 with noisy data, comparing the standard training approach, which consists of training the models once and testing on the following seizures with a chronological approach able to periodically adapting the models over time, and comparing models using handcrafted features with DNNs able to automatically extract information from the EEGs. Results indicated that the effect of denoising and/or retraining was more noticeable in the DNNs. It was also observed that, while models based on handcrafted features predicted more seizures than DNNs, they also generated significantly more false alarms. As a result, it was not possible to definitively determine which approach performed better, as it heavily depends on the specific application scenario. For some cases, there might be a high need to avoid missing a seizure prediction, while in other situations, the tolerability of false alarms becomes a crucial factor.

In Chapter 8, the main goal was to develop seizure prediction models using a transfer learning approach. Firstly, a deep convolutional autoencoder (DCAE) was trained using 41 patients from the EPILEPSIAE database. Then, the pretrained weights were used to initialise the parameters of patient-specific seizure prediction models developed for patients from the Epilepsy Center of the Universitätsklinikum Freiburg. Results indicated that using pretrained weights reduced the number of false alarms about four times, while keeping the seizure sensitivity (SS).

9.2 Added value of contributions and future directions

Data acquisition is a complex and expensive task. Removing artefacts from the signals is important for reducing false results obtained during the data analysis. Therefore, it is essential to continue developing more and better ways to remove artefacts while maintaining useful information [78]. In addition, creating open databases, as the one created on behalf of this thesis, with manually preprocessed data

would significantly contribute to benchmarking the current and new methodologies.

In the future, seizure prediction may involve using a wearable device that acquires EEGs in real-time, processes the data, and uses them to detect patterns leading to an upcoming seizure. This advancement could offer new hope to epilepsy patients, especially those in middle- or low-income countries, who face challenges in accessing expensive and specialised medical support. The production of such a device could go through two types of methodologies: the estimation of the probability of a seizure to happen (seizure forecasting) [140] or the raising of alarms indicating that the seizure will happen in a fixed and previously determined time horizon (seizure prediction) [133, 135]. In the case of seizure forecasting, the treatment may involve adapting the patients' lifestyle according to the seizure risk reported by the device or taking small doses of seizure suppression drugs. In the case of seizure prediction, the patient would have to take countermeasures as soon as the device indicated that a seizure would occur. The main difference between those is knowing when the seizure will happen. It is certainly an essential detail for the patients. Studies should be performed to evaluate the impact of the device's outputs on patients' lives.

In a seizure prediction device, all the steps from signal acquisition to the output must be automatic. Therefore, automatic denoising and retraining over time play a central role in the success of this pipeline. The automatic denoising should be instantaneous and robust. Otherwise, it may jeopardise the performance of the pipeline in real-time scenarios. Furthermore, in daily-life scenarios, the patients would use the device for long periods. Thus, the system must adapt to every fluctuation caused by medication intake, stress, daily activities, circadian cycles, or multidian cycles [77, 139]. One way of adaptation could be through periodic retraining. In this case, the device would adapt its parameters after a certain period or after each new change of distribution.

It is well known that the number of seizures available per patient is too small to obtain a good exploration of the complex patterns in the EEG data. Literature has reported that models should not be developed using data from several patients. Therefore, the models should be trained following a patient-specific approach. The use of transfer learning methods appears as an optimal solution. In this way, it is possible to train models using EEG data from several patients with epilepsy and then personalise them for every new patient. Such an approach combines the benefit of using information from other patients with the requirement to follow patient-specific development.

As future work, researchers should continue working on the three elements mentioned above: denoising of EEG, adaptation to new data distributions, and transfer learning using patients from different databases. The evaluation of new approaches should involve using metrics that allow the evaluation of the prediction of events instead of prediction by sample, which does not help in understanding whether the system works properly. Furthermore, the bet in ultra-long-term signals should be

made to perform more robust studies of this complex phenomenon. An improvement in seizure prediction could create a significant revolution in the world of patients with epilepsy.

References

[1] Björn Schelter, Matthias Winterhalder, Thomas Maiwald, Armin Brandt, et al. **Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction**. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **16**(1):013108, 2006. 34, 147

[2] Björn Schelter, Ralph G Andrzejak, and Florian Mormann. **Can your prediction algorithm beat a random predictor?** *Seizure prediction in epilepsy: from basic mechanisms to clinical applications*, pages 237–248, 2008. 32, 34

[3] World Health Organization. **Epilepsy - Detailed Fact Sheets**, 2019. Last accessed on 06-11-2019. 1, 10

[4] Roland D Thijs, Rainer Surges, Terence J O'Brien, and Josemir W Sander. **Epilepsy in adults**. *The Lancet*, **393**(10172):689–701, 2019. 1, 22

[5] Kenneth D Laxer, Eugen Trinka, Lawrence J Hirsch, Fernando Cendes, et al. **The consequences of refractory epilepsy and its treatment**. *Epilepsy & Behavior*, **37**:59–70, 2014. 1, 19, 20

[6] Orrin Devinsky, Annamaria Vezzani, Terence J. O'Brien, Nathalie Jette, et al. **Epilepsy**. *Nature Reviews Disease Primers*, **4**(1):18024, 2018. 1, 10, 11, 19, 20, 22, 26

[7] Bushra Sultana, Marie-Andrée Panzini, Ariane Veilleux Carpentier, Jacynthe Comtois, et al. **Incidence and prevalence of drug-resistant epilepsy: a systematic review and meta-analysis**. *Neurology*, **96**(17):805–817, 2021. 1, 19, 20

[8] Hanneke M De Boer, Marco Mula, and Josemir W Sander. **The global burden and stigma of epilepsy**. *Epilepsy & Behavior*, **12**(4):540–546, 2008. 1

[9] Jane McCagh, John E Fisk, and Gus A Baker. **Epilepsy, psychosocial and cognitive functioning**. *Epilepsy research*, **86**(1):1–14, 2009. 1

[10] Rita Nguyen and José F Téllez Zenteno. **Injuries in epilepsy: a review of its prevalence, risk factors, type of injuries and prevention**. *Neurology international*, **1**(1), 2009. 1

[11] Sonya B Dumanis, Jaqueline A French, Christophe Bernard, Gregory A Worrell, and Brandy E Fureman. **Seizure forecasting from idea to reality. Outcomes of the my seizure gauge epilepsy innovation institute workshop**. *Eneuro*, **4**(6), 2017. 1

[12] Paul R Carney, Stephen Myers, and James D Geyer. **Seizure prediction: methods**. *Epilepsy & behavior*, **22**:S94–S101, 2011. 2

[13] Juliane Klatt, Hinnerk Feldwisch-Drentrup, Matthias Ihle, Vincent Navarro, et al. **The EPILEPSIAE database: An extensive electroencephalography database of epilepsy patients**. *Epilepsia*, **53**(9):1669–1676, 2012. 2, 3, 60, 62, 88

[14] Levin Kuhlmann, Klaus Lehnertz, Mark P Richardson, Björn Schelter, and Hitten P Zaveri. **Seizure prediction—ready for a new era**. *Nature Reviews Neurology*, **14**(10):618–630, 2018. 2, 3, 69

[15] Caitlin L Grzeskowiak and Sonya B Dumanis. **Seizure forecasting: patient and caregiver perspectives**. *Frontiers in neurology*, **12**, 2021. 2

[16] Maxime O Baud, Timothée Proix, Nicholas M Gregg, Benjamin H Brinkmann, et al. **Seizure forecasting: bifurcations in the long and winding road**. *Epilepsia*, 2022. 2, 29, 69, 134, 149

[17] U Rajendra Acharya, S Vinitha Sree, G Swapna, Roshan Joy Martis, and Jasjit S Suri. **Automated EEG analysis of epilepsy: a review**. *Knowledge-Based Systems*, **45**:147–165, 2013. 2, 14, 70, 133

[18] Dean R Freestone, Philippa J Karoly, and Mark J Cook. **A forward-looking review of seizure prediction**. *Current opinion in neurology*, **30**(2):167–173, 2017. 2, 3, 27, 29, 133, 148

[19] Theoden I Netoff. **The ability to predict seizure onset**. In *Engineering in Medicine*, pages 365–378. Elsevier, 2019. 2, 69, 134, 146

[20] Khansa Rasheed, Adnan Qayyum, Junaid Qadir, Shobi Sivathamboo, et al. **Machine learning for predicting epileptic seizures using EEG signals: A review**. *IEEE Reviews in Biomedical Engineering*, **14**:139–155, 2020. 2, 3, 69, 134, 146

[21] Yang Li, Yu Liu, Yu-Zhu Guo, Xiao-Feng Liao, et al. **Spatio-temporal-spectral hierarchical graph convolutional network with**

semisupervised active learning for patient-specific seizure prediction. *IEEE transactions on cybernetics*, 2021. 2, 69, 134, 146

[22] Mark H Myers, Akshay Padmanabha, Gahangir Hossain, Amy L de Jongh Curry, and Charles D Blaha. **Seizure prediction and detection via phase and amplitude lock values**. *Frontiers in human neuroscience*, **10**:80, 2016. 2, 64, 69, 73, 85, 134

[23] Kevin T Sweeney, Tomás E Ward, and Seán F McLoone. **Artifact removal in physiological signals—Practices and possibilities**. *IEEE transactions on information technology in biomedicine*, **16**(3):488–500, 2012. 2, 17, 18, 49, 50, 51, 53, 54, 55, 97, 98

[24] Md Kafiul Islam, Amir Rastegarnia, and Zhi Yang. **Methods for artifact detection and removal from scalp EEG: A review**. *Neurophysiologie Clinique/Clinical Neurophysiology*, **46**(4-5):287–305, 2016. 2, 14, 17, 18, 50, 53, 54

[25] Md Shafiqul Islam, Ahmad M El-Hajj, Hussein Alawieh, Zaher Dawy, et al. **EEG mobility artifact removal for ambulatory epileptic seizure prediction applications**. *Biomedical Signal Processing and Control*, **55**:101638, 2020. 2, 54, 66, 69, 76, 83, 129, 134

[26] Khakon Das, Debashis Daschakladar, Partha Pratim Roy, Atri Chatterjee, and Shankar Prasad Saha. **Epileptic seizure prediction by the detection of seizure waveform from the pre-ictal phase of EEG signal**. *Biomedical Signal Processing and Control*, **57**:101720, 2020. 2, 66, 69, 76, 85, 134

[27] Syed Muhammad Usman, Shehzad Khalid, and Sadaf Bashir. **A deep learning based ensemble learning method for epileptic seizure prediction**. *Computers in Biology and Medicine*, **136**:104710, 2021. 2, 67, 69, 78, 82, 83, 134

[28] Banu Priya Prathaban and Ramachandran Balasubramanian. **Dynamic learning framework for epileptic seizure prediction using sparsity based EEG Reconstruction with Optimized CNN classifier**. *Expert Systems with Applications*, **170**:114533, 2021. 2, 66, 69, 77, 134

[29] Banu Priya Prathaban, Ramachandran Balasubramanian, and R Kalpana. **ForeSeiz: An IoMT based headband for Real-time epileptic seizure forecasting**. *Expert Systems with Applications*, **188**:116083, 2022. 2, 134

[30] Christian Meisel and Tobias Loddenkemper. **Seizure prediction and intervention**. *Neuropharmacology*, **172**:107898, 2020. 3

[31] Pierre Thodoroff, Joelle Pineau, and Andrew Lim. **Learning robust features using deep learning for automatic seizure detection**. In *Machine learning for healthcare conference*, pages 178–190, 2016. 3

[32] Christian Meisel and Kimberlyn A Bailey. **Identifying signal-dependent information about the preictal state: A comparison across ECoG, EEG and EKG using deep learning**. *EBioMedicine*, **45**:422–431, 2019. 3, 28

[33] Isabell Kiral-Kornek, Subhrajit Roy, Ewan Nurse, Benjamin Mashford, et al. **Epileptic seizure prediction using big data and deep learning: toward a mobile system**. *EBioMedicine*, **27**:103–111, 2018. 3, 60, 65, 69, 74, 83, 84, 113, 134, 146

[34] Hsiang-Han Chen, Han-Tai Shiao, and Vladimir Cherkassky. **Online Prediction of Lead Seizures from iEEG Data**. *Brain Sciences*, **11**(12):1554, 2021. 3, 66, 77, 84, 134, 146

[35] Jamal Nazari, Ali Motie Nasrabadi, Mohammad Bagher Menhaj, and Somayeh Raiesdana. **Epilepsy seizure prediction with few-shot learning method**. *Brain Informatics*, **9**(1):1–9, 2022. 3, 152, 158

[36] Ricardo Couceiro, Gonçalo Duarte, João Durães, João Castelhano, et al. **Biofeedback augmented software engineering: monitoring of programmers' mental effort**. In *2019 IEEE/ACM 41st International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER)*, pages 37–40. IEEE, 2019. 3

[37] Fábio Lopes, Adriana Leal, Júlio Medeiros, Mauro F Pinto, et al. **EPIC: Annotated epileptic EEG independent components for artifact reduction**. *Scientific Data*, **9**(1):512, 2022. 4, 87

[38] Fábio Lopes, Adriana Leal, Júlio Medeiros, Mauro F Pinto, et al. **Automatic Electroencephalogram Artifact Removal Using Deep Convolutional Neural Networks**. *IEEE Access*, **9**:149955–149970, 2021. 4, 97

[39] Fábio Lopes, Adriana Leal, Júlio Medeiros, Mauro F Pinto, et al. **Ensemble Deep Neural Network for Automatic Classification of EEG Independent Components**. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **30**:559–568, 2022. 4, 117

[40] Fábio Lopes, Adriana Leal, Mauro F Pinto, António Dourado, et al. **Removing artefacts and periodically retraining improve performance of neural network-based seizure prediction models**. *Scientific Reports*, **13**(1):5918, 2023. 5, 133

[41] Helen E Scharfman. **The neurobiology of epilepsy**. *Current neurology and neuroscience reports*, **7**(4):348–354, 2007. 9

[42] Robert S Fisher, Carlos Acevedo, Alexis Arzimanoglou, Alicia Bogacz, et al. **ILAE official report: a practical clinical definition of epilepsy**. *Epilepsia*, **55**(4):475–482, 2014. 9, 10, 12, 13

[43] Robert S Fisher, Walter Van Emde Boas, Warren Blume, Christian Elger, et al. **Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)**. *Epilepsia*, **46**(4):470–472, 2005. 9, 10

[44] Premysl Jiruska, Marco De Curtis, John GR Jefferys, Catherine A Schevon, et al. **Synchronization and desynchronization in epilepsy: controversies and hypotheses**. *The Journal of physiology*, **591**(4):787–797, 2013. 9

[45] Y Wu, D Liu, and Z Song. **Neuronal networks and energy bursts in epilepsy**. *Neuroscience*, **287**:175–186, 2015. 9

[46] Ingrid E Scheffer, Samuel Berkovic, Giuseppe Capovilla, Mary B Connolly, et al. **ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology**. *Epilepsia*, **58**(4):512–521, 2017. 10, 11, 12, 13

[47] Robert S Fisher, J Helen Cross, Jacqueline A French, Norimichi Higurashi, et al. **Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology**. *Epilepsia*, **58**(4):522–530, 2017. 10, 11, 12, 213, 229

[48] Thomas V. Kodankandath, Danny Theodore, and Debopam Samanta. **Generalized Tonic-Clonic Seizure**, 2022. 11

[49] Epilepsy Foundation. **Temporal Lobe Epilepsy**, 2022. Last accessed on 08-11-2022. 13

[50] National Institute of Neurological Disorders and Stroke. **The Epilepsies and Seizures: Hope Through Research**, 2022. Last accessed on 08-11-2022. 13

[51] Saba Jafarpour, Lawrence J. Hirsch, Marina Gaínza-Lein, Christoph Kellinghaus, and Kamil Detyniecki. **Seizure cluster: Definition, prevalence, consequences, and management**. *Seizure*, **68**:9–15, 2019. 13, 28

[52] Barry Gidal, Pavel Klein, and Lawrence J Hirsch. **Seizure clusters, rescue treatments, seizure action plans: Unmet needs and emerging formulations**. *Epilepsy & Behavior*, **112**:107391, 2020. 13

[53] Barry Gidal and Kamil Detyniecki. **Rescue therapies for seizure clusters: pharmacology and target of treatments**. *Epilepsia*, **63**:S34–S44, 2022. 13

[54] Paul L Nunez, Ramesh Srinivasan, et al. **The Physics - EEG Interface**. In *Electric fields of the brain: the neurophysics of EEG*, pages 3–55. Oxford University Press, USA, 2006. 14

[55] Saeid Sanei and Jonathon A. Chambers. **Introduction to EEG**. In *EEG Signal Processing*, chapter 1, pages 1–34. John Wiley & Sons Inc, 2007. 14

[56] Anteneh M Feyissa, Gregory A Worrell, and Terrence D Lagerlund. **EEG and Epilepsy**. *Epilepsy*, pages 77–98, 2021. 14

[57] Pedro F Viana, Jonas Duun-Henriksen, Martin Glasstëter, Matthias Dümpelmann, et al. **230 days of ultra long-term subcutaneous EEG: seizure cycle analysis and comparison to patient diary**. *Annals of clinical and translational neurology*, **8**(1):288–293, 2021. 14, 16, 55, 84, 150, 159

[58] Florian Mormann, Ralph G Andrzejak, Christian E Elger, and Klaus Lehnertz. **Seizure prediction: the long and winding road**. *Brain*, **130**(2):314–333, 2006. 14, 30

[59] Valer Jurcak, Daisuke Tsuzuki, and Ippeita Dan. **10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems**. *Neuroimage*, **34**(4):1600–1611, 2007. 14, 17, 97

[60] Robert Oostenveld and Peter Praamstra. **The five percent electrode system for high-resolution EEG and ERP measurements**. *Clinical neurophysiology*, **112**(4):713–719, 2001. 15

[61] Joseph Dien. **Issues in the application of the average reference: Review, critiques, and recommendations**. *Behavior Research Methods, Instruments, & Computers*, **30**(1):34–43, 1998. 14

[62] Shiang Hu, Yongxiu Lai, Pedro A Valdes-Sosa, Maria L Bringas-Vega, and Dezhong Yao. **How do reference montage and electrodes setup affect the measured scalp EEG potentials?** *Journal of neural engineering*, **15**(2):026013, 2018. 14

[63] Quanying Liu, Joshua H Balsters, Marc Baechinger, Onno van der Groen, et al. **Estimating a neutral reference for electroencephalographic recordings: the importance of using a high-density montage and a realistic head model**. *Journal of neural engineering*, **12**(5):056012, 2015. 14

[64] Turkey N Alotaiby, Saleh A Alshebeili, Tariq Alshawi, Ishtiaq Ahmad, and Fathi E Abd El-Samie. **EEG seizure detection and prediction algorithms: a survey**. *EURASIP Journal on Advances in Signal Processing*, **2014**(1):183, 2014. 15, 70, 133

[65] Felix Rosenow and Hans Lüders. **Presurgical evaluation of epilepsy**. *Brain*, **124**(9):1683–1700, 2001. 15

[66] Manuel R Mercier, Anne-Sophie Dubarry, François Tadel, Pietro Avanzini, et al. **Advances in human intracranial electroencephalography research, guidelines and good practices**. *NeuroImage*, page 119438, 2022. 16

[67] AboutKidsHealth. **Invasive electroencephalography (EEG) monitoring before epilepsy surgery**, 2017. Last accessed on 11-08-2020. 16

[68] S Nasehi and H Pourghassem. **Seizure detection algorithms based on analysis of EEG and ECG signals: a survey**. *Neurophysiology*, **44**(2):174–186, 2012. 16

[69] Jörg Wellmer, Ferdinand von der Groeben, Ute Klarmann, Christian Weber, et al. **Risks and benefits of invasive epilepsy surgery workup with implanted subdural and depth electrodes**. *Epilepsia*, **53**(8):1322–1332, 2012. 16

[70] Jonas Duun-Henriksen, Maxime Baud, Mark P Richardson, Mark Cook, et al. **A new era in electroencephalographic monitoring? Subscalp devices for ultra–long-term recordings**. *Epilepsia*, **61**(9):1805–1817, 2020. 16, 26

[71] Zulfi Haneef, Kaiyuan Yang, Sameer A Sheth, Fuad Z Aloor, et al. **Sub-scalp electroencephalography: A next-generation technique to study human neurophysiology**. *Clinical Neurophysiology*, 2022. 16

[72] Sigge Weisdorf, Jonas Duun-Henriksen, Marianne J Kjeldsen, Frantz R Poulsen, et al. **Ultra-long-term subcutaneous home monitoring of epilepsy—490 days of EEG from nine patients**. *Epilepsia*, **60**(11):2204–2214, 2019. 16, 26, 60, 61, 62

[73] Pedro F Viana, Line S Remvig, Jonas Duun-Henriksen, Martin Glasstetter, et al. **Signal quality and power spectrum analysis of remote ultra long-term subcutaneous EEG**. *Epilepsia*, **62**(8):1820–1828, 2021. 16

[74] Mojtaba Bandarabadi, César A Teixeira, Jalil Rasekhi, and António Dourado. **Epileptic seizure prediction using relative spectral power features**. *Clinical Neurophysiology*, **126**(2):237–248, 2015. 16, 50, 64, 72, 81, 83, 85, 134, 137, 141, 149

[75] John William Carey Medithe and Usha Rani Nelakuditi. **Study of normal and abnormal EEG**. In *2016 3rd International conference on advanced computing and communication systems (ICACCS)*, **1**, pages 1–4. IEEE, 2016. 16, 17

[76] Philippa J Karoly, Hoameng Ung, David B Grayden, Levin Kuhlmann, et al. **The circadian profile of epilepsy improves seizure forecasting**. *Brain*, **140**(8):2169–2182, 2017. 17, 28, 65, 73, 83, 85, 97, 145, 228

[77] Maxime O Baud, Jonathan K Kleen, Emily A Mirro, Jason C Andrechak, et al. **Multi-day rhythms modulate seizure risk in epilepsy**. *Nature communications*, **9**(1):1–10, 2018. 17, 36, 84, 97, 149, 163

[78] Jose Antonio Urigüen and Begoña Garcia-Zapirain. **EEG artifact removal—state-of-the-art and guidelines**. *Journal of neural engineering*, **12**(3):031001, 2015. 18, 49, 50, 51, 52, 53, 54, 98, 100, 117, 162

[79] Samuel Boudet, Laurent Peyrodie, Philippe Gallois, and Christian Vasseur. **A global approach for automatic artifact removal for standard EEG record**. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5719–5722. IEEE, 2006. 18

[80] Raphaëlle N Roy, Sylvie Charbonnier, and Stephane Bonnet. **Eye blink characterization from frontal EEG electrodes using source separation and pattern recognition algorithms**. *Biomedical Signal Processing and Control*, **14**:256–264, 2014. 18

[81] Irina I Goncharova, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw. **EMG contamination of EEG: spectral and topographical characteristics**. *Clinical neurophysiology*, **114**(9):1580–1593, 2003. 18

[82] Xiao Jiang, Gui-Bin Bian, and Zean Tian. **Removal of artifacts from EEG signals: a review**. *Sensors*, **19**(5):987, 2019. 18, 52

[83] Peter Anderer, Stephen Roberts, Alois Schlögl, Georg Gruber, et al. **Artifact processing in computerized analysis of sleep EEG–a review**. *Neuropsychobiology*, **40**(3):150–157, 1999. 18, 50

[84] Yi Wang and Zhong Chen. **An update for epilepsy research and antiepileptic drug development: Toward precise circuit therapy**. *Pharmacology & therapeutics*, **201**:77–93, 2019. 19, 20

[85] Meir Bialer and H Steve White. **Key factors in the discovery and development of new antiepileptic drugs**. *Nature reviews Drug discovery*, **9**(1):68–82, 2010. 19

[86] David G Vossler, Mindl Weingarten, Barry E Gidal, and American Epilepsy Society Treatments Committee. **Summary of antiepileptic drugs available in the United States of America**. *Epilepsy currents*, **18**(4_suppl):1–26, 2018. 19

[87] Patrick Kwan, Alexis Arzimanoglou, Anne T Berg, Martin J Brodie, et al. **Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies**. *Epilepsia*, **51**(6):1069–1077, 2010. 19, 20

[88] Michael R Sperling, Suzanne Barshow, Maromi Nei, and Ali A Asadi-Pooya. **A reappraisal of mortality after epilepsy surgery**. *Neurology*, **86**(21):1938–1944, 2016. 20

[89] Antonella Fattorusso, Sara Matricardi, Elisabetta Mencaroni, Giovanni Battista Dell'Isola, et al. **The Pharmacoresistant Epilepsy: an overview on existant and new emerging therapies**. *Frontiers in Neurology*, **12**:1030, 2021. 20

[90] Jerome Engel. **What can we do for people with drug-resistant epilepsy?: the 2016 Wartenberg lecture**. *Neurology*, **87**(23):2483–2489, 2016. 20

[91] Philippe Ryvlin, J Helen Cross, and Sylvain Rheims. **Epilepsy surgery in children and adults**. *The Lancet Neurology*, **13**(11):1114–1126, 2014. 20

[92] Chaturbhuj Rathore and Kurupath Radhakrishnan. **Concept of epilepsy surgery and presurgical evaluation**. *Epileptic disorders*, **17**(1):19–31, 2015. 20, 21, 22

[93] Jack Kirby, Veronica M Leach, Alice Brockington, Phillip Patsalos, et al. **Drug withdrawal in the epilepsy monitoring unit–The patsalos table**. *Seizure*, **75**:75–81, 2020. 20, 21

[94] Dario J Englot and Edward F Chang. **Rates and predictors of seizure freedom in resective epilepsy surgery: an update**. *Neurosurgical review*, **37**(3):389–405, 2014. 22

[95] Kayela Arrotta, Nicolas R Thompson, Ryan Honomichl, Imad Najm, et al. **Quality of life after epilepsy surgery: How domain-specific cognitive changes impact QOL within the context of seizure outcome**. *Epilepsy & Behavior*, **137**:108948, 2022. 22

[96] Shi-Yong Liu, Xiao-Lin Yang, Bing Chen, Zhi Hou, et al. **Clinical outcomes and quality of life following surgical treatment for refractory epilepsy: a systematic review and meta-analysis**. *Medicine*, **94**(6), 2015. 22

[97] Sallie Baxendale and Pamela Thompson. **Red flags in epilepsy surgery: identifying the patients who pay a high cognitive price for an unsuccessful surgical outcome**. *Epilepsy & behavior*, **78**:269–272, 2018. 22

[98] Philippe Ryvlin, Sylvain Rheims, Lawrence J Hirsch, Arseny Sokolov, and Lara Jehi. **Neuromodulation in epilepsy: state-of-the-art approved therapies**. *The Lancet Neurology*, **20**(12):1038–1047, 2021. 22, 23, 24

[99] Lahoud Touma, Bénédicte Dansereau, Alvin Y Chan, Nathalie Jetté, et al. **Neurostimulation in people with drug-resistant epilepsy: Systematic review and meta-analysis from the ILAE Surgical Therapies Commission**. *Epilepsia*, 2022. 22, 23

[100] Elinor Ben-Menachem. **Vagus-nerve stimulation for the treatment of epilepsy**. *The Lancet Neurology*, **1**(8):477–482, 2002. 22

[101] Robert S Fisher, Pegah Afra, Micheal Macken, Daniela N Minecan, et al. **Automatic vagus nerve stimulation triggered by ictal tachycardia: clinical outcomes and device performance—the US E-37 trial**. *Neuromodulation: Technology at the Neural Interface*, **19**(2):188–195, 2016. 22

[102] RS Fisher, KS Eggleston, and CW Wright. **Vagus nerve stimulation magnet activation for seizures: a critical review**. *Acta Neurologica Scandinavica*, **131**(1):1–8, 2015. 22

[103] Felice T Sun and Martha J Morrell. **The RNS System: responsive cortical stimulation for the treatment of refractory partial epilepsy**. *Expert review of medical devices*, **11**(6):563–572, 2014. 22

[104] Matthew D Bigelow and Abbas Z Kouzani. **Neural stimulation systems for the control of refractory epilepsy: a review**. *Journal of neuroengineering and rehabilitation*, **16**(1):1–17, 2019. 23

[105] George P Thomas and Barbara C Jobst. **Critical review of the responsive neurostimulator system for epilepsy**. *Medical Devices (Auckland, NZ)*, **8**:405, 2015. 23

[106] Robert Fisher, Vicenta Salanova, Thomas Witt, Robert Worth, et al. **Electrical stimulation of the anterior nucleus of thalamus for treatment of refractory epilepsy**. *Epilepsia*, **51**(5):899–908, 2010. 23

[107] Vicenta Salanova, Thomas Witt, Robert Worth, Thomas R Henry, et al. **Long-term efficacy and safety of thalamic stimulation for drug-resistant partial epilepsy**. *Neurology*, **84**(10):1017–1025, 2015. 23

[108] Natalia Rincon, Donald Barr, and Naymee Velez-Ruiz. **Neuromodulation in drug resistant epilepsy**. *Aging and disease*, **12**(4):1070, 2021. 23

[109] Andreas Schulze-Bonhage. **Long-term outcome in neurostimulation of epilepsy**. *Epilepsy & Behavior*, **91**:25–29, 2019. 23

[110] Peter Wolf, Katia Lin, and Marina Nikanorova. **Non-Pharmacological Therapy of Epilepsy**. In Simon Shorvon, Renzo Guerrini, Mark Cook, and Samden Lhatoo, editors, *Oxford Textbook of Epilepsy and Epileptic Seizures*, chapter 12, pages 135–143. Oxford University Press, 2012. 24, 25

[111] Fritz E Dreifuss, N Paul Rosman, James C Cloyd, John M Pellock, et al. **A comparison of rectal diazepam gel and placebo for acute repetitive seizures**. *New England Journal of Medicine*, **338**(26):1869–1875, 1998. 24

[112] Suresh K Agarwal, Robert L Kriel, Richard C Brundage, Vijay D Ivaturi, and James C Cloyd. **A pilot study assessing the bioavailability and pharmacokinetics of diazepam after intranasal and intravenous administration in healthy volunteers**. *Epilepsy research*, **105**(3):362–367, 2013. 25

[113] Eric B Segal, Daniel Tarquinio, Ian Miller, James W Wheless, et al. **Evaluation of diazepam nasal spray in patients with epilepsy concomitantly using maintenance benzodiazepines: An interim subgroup analysis from a phase 3, long-term, open-label safety study**. *Epilepsia*, **62**(6):1442–1450, 2021. 25

[114] European Medicines Agency. **BUCCOLAM (Midazolam) - Summary of Product Characteristics**, 2022. Last accessed on 22-11-2022. 25

[115] James Cloyd, Sheryl Haut, Enrique Carrazana, and Adrian L Rabinowicz. **Overcoming the challenges of developing an intranasal diazepam rescue therapy for the treatment of seizure clusters**. *Epilepsia*, **62**(4):846–856, 2021. 25

[116] UCB Inc. **NAYZILAM®(midazolam nasal spray). Full prescribing information**, 2021. Last accessed on 22-11-2022. 25

[117] Mackenzie C Cervenka, Bobbie J Henry, Elizabeth A Felton, Katlyn Patton, and Eric H Kossoff. **Establishing an adult epilepsy diet center: experience, efficacy and challenges**. *Epilepsy & Behavior*, **58**:61–68, 2016. 25

[118] Sándor Beniczky, Philippa Karoly, Ewan Nurse, Philippe Ryvlin, and Mark Cook. **Machine learning and wearable devices of the future**. *Epilepsia*, **62**:S116–S124, 2021. 26

[119] Levente Hadady, Péter Klivényi, Dániel Fabó, and Sándor Beniczky. **Real-world user experience with seizure detection wearable devices in the home environment**. *Epilepsia*, 2022. 26

[120] Christoph Baumgartner and Johannes P Koren. **Seizure detection using scalp-EEG**. *Epilepsia*, **59**:14–22, 2018. 26

[121] Sandor Beniczky, Tilman Polster, Troels W Kjaer, and Helle Hjalgrim. **Detection of generalized tonic–clonic seizures by a wireless wrist accelerometer: a prospective, multicenter study**. *Epilepsia*, **54**(4):e58–e61, 2013. 26

[122] Sándor Beniczky, Isa Conradsen, Oliver Henning, Martin Fabricius, and Peter Wolf. **Automated real-time detection of tonic-clonic seizures using a wearable EMG device**. *Neurology*, **90**(5):e428–e434, 2018. 26

[123] Johan Arends, Roland D Thijs, Thea Gutter, Constantin Ungureanu, et al. **Multimodal nocturnal seizure detection in a residential care setting: a long-term prospective trial**. *Neurology*, **91**(21):e2010–e2019, 2018. 26

[124] Jesper Jeppesen, Anders Fuglsang-Frederiksen, Peter Johansen, Jakob Christensen, et al. **Seizure detection based on heart rate variability using a wearable electrocardiography device**. *Epilepsia*, **60**(10):2105–2113, 2019. 26

[125] Mark J Cook, Terence J O'Brien, Samuel F Berkovic, Michael Murphy, et al. **Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study**. *The Lancet Neurology*, **12**(6):563–571, 2013. 26, 27, 28, 60, 62, 64, 71, 82, 83, 85, 158

[126] Philippa J. Karoly, Mark J. Cook, Matias Maturana, Ewan S. Nurse, et al. **Forecasting cycles of seizure likelihood**. *Epilepsia*, **61**(4):776–786, 2020. 26

[127] Rachel E Stirling, Matias I Maturana, Philippa J Karoly, Ewan S Nurse, et al. **Seizure forecasting using a novel sub-scalp ultra-long term EEG monitoring system**. *Frontiers in Neurology*, page 1445, 2021. 26, 29, 60, 61, 62

[128] Steven V Pacia, Werner K Doyle, Daniel Friedman, Daniel H Bacher, and Ruben I Kuzniecky. **Intracranial EEG validation of single-channel subgaleal EEG for seizure identification**. *Journal of Clinical Neurophysiology*, **39**(4):283–288, 2022. 26

[129] Wyss Center. **Epios.TM**, 2022. Last accessed on 02-12-2022. 26

[130] Kristina Kravalis and Andreas Schulze-Bonhage. **PIMIDES I: a pilot study to assess the feasibility of patient-controlled neurostimulation with the EASEE® system to treat medically refractory focal epilepsy**. *Neurological Research and Practice*, **2**(1):1–3, 2020. 26

[131] J Gotman and P Gloor. **Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG**. *Electroencephalography and clinical neurophysiology*, **41**(5):513–529, 1976. 27, 133

[132] Daniel F Ghougassian, Wendyl d'Souza, Mark J Cook, and Terence J O'Brien. **Evaluating the utility of inpatient video-EEG monitoring**. *Epilepsia*, **45**(8):928–932, 2004. 27

[133] Song Cui, Lijuan Duan, Yuanhua Qiao, and Ying Xiao. **Learning EEG synchronization patterns for epileptic seizure prediction using bag-of-wave features**. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–16, 2018. 27, 163

[134] Brian Litt, Rosana Esteller, Javier Echauz, Maryann D'Alessandro, et al. **Epileptic seizures may begin hours in advance of clinical onset: A report of five patients**. *Neuron*, **30**(1):51–64, 2001. 28

[135] Brian Litt and Javier Echauz. **Prediction of epileptic seizures**. *The Lancet Neurology*, **1**(1):22–30, 2002. 28, 163

[136] Florian Mormann, Ralph G. Andrzejak, and Klaus Lehnertz. **Automated prediction and assessment of seizure prediction algorithms**. In Ivan Osorio, Hitten P. Zaveri, Mark G. Frei, and Susan Arthurs, editors, *Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Engineering, and Physics*, chapter 11, pages 165–174. CRC Press, Boca Raton, FL, 1st edition, 2016. 28

[137] Elie Bou Assi, Dang K Nguyen, Sandy Rihana, and Mohamad Sawan. **Towards accurate prediction of epileptic seizures: A review**. *Biomedical Signal Processing and Control*, **34**:144–157, 2017. 28, 29, 30, 32, 34, 50, 62, 69, 81, 82, 83, 133, 134, 136

[138] Hsiang-Han Chen and Vladimir Cherkassky. **Performance metrics for online seizure prediction**. *Neural Networks*, **128**:22–32, 2020. 28

[139] M. Winterhalder, T. Maiwald, H.U. Voss, R. Aschenbrenner-Scheibe, et al. **The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods**. *Epilepsy & Behavior*, **4**(3):318–325, 2003. 28, 29, 31, 32, 36, 60, 62, 85, 163

[140] Rachel E. Stirling, David B. Grayden, Wendyl D'Souza, Mark J. Cook, et al. **Forecasting Seizure Likelihood With Wearable Technology**. *Frontiers in Neurology*, **12**:1170, 2021. 28, 61, 163

[141] Catalina Alvarado-Rojas, Mario Valderrama, A Fouad-Ahmed, H Feldwisch-Drentrup, et al. **Slow modulations of high-frequency activity (40–140 Hz) discriminate preictal changes in human focal epilepsy**. *Scientific Reports*, **4**(1):1–9, 2015. 28, 64, 72, 83, 85

[142] Mauro Pinto, Adriana Leal, Fábio Lopes, António Dourado, et al. **A personalized and evolutionary algorithm for interpretable EEG epilepsy seizure prediction**. *Scientific reports*, **11**(1):1–12, 2021. 28, 84, 89, 137, 141, 146, 147, 148, 224, 227

[143] Mauro Pinto, Tiago Coelho, Adriana Leal, Fábio Lopes, et al. **Interpretable EEG seizure prediction using a multiobjective evolutionary algorithm**. *Scientific reports*, **12**(1):1–15, 2022. 28, 84, 137, 141, 146, 147, 148, 224, 227

[144] Sriram Ramgopal, Sigride Thome-Souza, Michele Jackson, Navah Ester Kadish, et al. **Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy**. *Epilepsy & Behavior*, **37**:291–307, 2014. 29

[145] James W Wheless, Daniel Friedman, Gregory L Krauss, Vikram R Rao, et al. **Future opportunities for research in rescue treatments**. *Epilepsia*, **63**:S55–S68, 2022. 29

[146] Kais Gadhoumi, Jean-Marc Lina, Florian Mormann, and Jean Gotman. **Seizure prediction for therapeutic devices: A review**. *Journal of neuroscience methods*, **260**:270–282, 2016. 30, 133

[147] Andreas Schulze-Bonhage, Francisco Sales, Kathrin Wagner, Rute Teotonio, et al. **Views of patients with epilepsy on seizure prediction devices**. *Epilepsy & behavior*, **18**(4):388–396, 2010. 32

[148] Ralph G Andrzejak, Daniel Chicharro, Christian E Elger, and Florian Mormann. **Seizure prediction: Any better than chance?** *Clinical Neurophysiology*, **120**(8):1465–1478, 2009. 32, 33

[149] Luigi Chisci, Antonio Mavino, Guido Perferi, Marco Sciandrone, et al. **Real-Time Epileptic Seizure Prediction Using AR Models and Support Vector Machines**. *IEEE Transactions on Biomedical Engineering*, **57**(5):1124–1132, 2010. 34, 35

[150] César Teixeira, Bruno Direito, Mojtaba Bandarabadi, and António Dourado. **Output regularization of SVM seizure predictors: Kalman Filter versus the "Firing Power" method**. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6530–6533. IEEE, 2012. 34, 35, 83, 141

[151] Indrė Žliobaitė. **Learning under concept drift: an overview**. *arXiv preprint arXiv:1010.4784*, 2010. 35, 36

[152] Sofia Khan, Lino Nobili, Ramin Khatami, Tobias Loddenkemper, et al. **Circadian rhythm and epilepsy**. *The Lancet Neurology*, **17**(12):1098–1108, 2018. 36, 37

[153] Khald Ali I Aboalayon, Miad Faezipour, Wafaa S Almuhammadi, and Saeid Moslehpour. **Sleep stage classification using EEG signal analysis: a comprehensive survey and new investigation**. *Entropy*, **18**(9):272, 2016. 37

[154] Domien GM Beersma and Marijke CM Gordijn. **Circadian control of the sleep–wake cycle**. *Physiology & behavior*, **90**(2-3):190–195, 2007. 37

[155] Jim Waterhouse, Yumi Fukuda, and Takeshi Morita. **Daily rhythms of the sleep-wake cycle**. *Journal of physiological anthropology*, **31**(1):1–14, 2012. 37

[156] Patrick M Fuller, Joshua J Gooley, and Clifford B Saper. **Neurobiology of the sleep-wake cycle: sleep architecture, circadian regulation, and regulatory feedback**. *Journal of biological rhythms*, **21**(6):482–493, 2006. 37

[157] Divyani Garg, Laurel Charlesworth, and Garima Shukla. **Sleep and Temporal Lobe Epilepsy–Associations, Mechanisms and Treatment Implications**. *Frontiers in Human Neuroscience*, **16**:849899, 2022. 37

[158] Christopher P Derry and Susan Duncan. **Sleep and epilepsy**. *Epilepsy & Behavior*, **26**(3):394–404, 2013. 37

[159] Adam H Marblestone, Greg Wayne, and Konrad P Kording. **Toward an Integration of Deep Learning and Neuroscience**. *Frontiers in Computational Neuroscience*, **10**:1–41, 2016. 37

[160] Yoshua Bengio, Dong-Hyun Lee, Jorg Bornschein, Thomas Mesnard, and Zhouhan Lin. **Towards Biologically Plausible Deep Learning**. *arXiv preprint arXiv:1502.04156*, pages 1–10, 2015. 37

[161] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. **Introduction**. In *Deep Learning*, chapter 1, pages 1–26. MIT Press, 2016. 38

[162] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. **ImageNet Classification with Deep Convolutional Neural Networks**. In F Pereira, C J C Burges, L Bottou, and K Q Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. 38

[163] Clement Farabet, Camille Couprie, Laurent Najman, and Yann LeCun. **Learning Hierarchical Features for Scene Labeling**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**(8):1915–1929, 2013. 38

[164] Tomáš Mikolov, Anoop Deoras, Daniel Povey, Lukáš Burget, and Jan Černocky. **Strategies for Training Large Scale Neural Network Language Models**. In *2011 IEEE Workshop on Automatic Speech Recognition & Understanding*, pages 196–201. IEEE, 2011. 38

[165] Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, et al. **Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups**. *IEEE Signal Processing Magazine*, **29**(6):82–97, 2012. 38

[166] Michael K K Leung, Hui Yuan Xiong, Leo J Lee, and Brendan J Frey. **Deep Learning of the Tissue-Regulated Splicing Code**. *Bioinformatics*, **30**(12):i121—-i129, 2014. 38

[167] Hui Y Xiong, Babak Alipanahi, Leo J Lee, Hannes Bretschneider, et al. **The Human Splicing Code Reveals New Insights into the Genetic Determinants of Disease**. *Science*, **347**(6218):1–7, 2015. 38

[168] Wikipedia. **Neuron**, 2020. Last accessed on 17-08-2020. 39

[169] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. **Activation functions: Comparison of trends in practice and research for deep learning**. *arXiv preprint arXiv:1811.03378*, 2018. 39

[170] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. **Deep learning**. *Nature*, **521**(7553):436–444, 2015. 39, 42, 99, 118, 133, 146, 158

[171] Neuwrite. **Deep neural networks help us read your mind**, 2015. Last accessed on 17-08-2020. 40

[172] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. **Convolutional Networks**. In *Deep Learning*, chapter 9, pages 321–362. MIT Press, 2016. 40, 121

[173] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. **Practical Methodology**. In *Deep Learning*, chapter 11, pages 409–430. MIT Press, 2016. 41

[174] Min Lin, Qiang Chen, and Shuicheng Yan. **Network in network**. *arXiv preprint arXiv:1312.4400*, 2013. 42, 122

[175] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. **Sequence Modeling: Recurrent and Recursive Networks**. In *Deep Learning*, chapter 10, pages 363–408. MIT Press, 2016. 42, 44

[176] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. **Optimization for Training Deep Models**. In *Deep Learning*, chapter 8, pages 267–320. MIT Press, 2016. 42

[177] Sepp Hochreiter and Jürgen Schmidhuber. **Long Short-Term Memory**. *Neural Computation*, **9**(8):1735–1780, 1997. 42

[178] Towards Data Science. **Illustrated Guide to LSTM's and GRU's: A step by step explanation**, 2018. Last accessed on 17-08-2020. 43, 44

[179] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, et al. **Learning phrase representations using RNN encoder-decoder for statistical machine translation**. *arXiv preprint arXiv:1406.1078*, 2014. 43

[180] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. **A survey of transfer learning**. *Journal of Big data*, **3**(1):1–40, 2016. 44

[181] Theekshana Dissanayake, Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes. **Deep learning for patient-independent epileptic seizure prediction using scalp EEG signals**. *IEEE Sensors Journal*, **21**(7):9377–9388, 2021. 44, 67, 69, 77, 83, 134, 152, 158

[182] Sinno Jialin Pan and Qiang Yang. **A survey on transfer learning**. *IEEE Transactions on knowledge and data engineering*, **22**(10):1345–1359, 2009. 44

[183] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, et al. **A survey on deep transfer learning**. In *International conference on artificial neural networks*, pages 270–279. Springer, 2018. 44, 45, 46, 130

[184] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. **Deep residual learning for image recognition**. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 45, 159

[185] François Chollet. **Xception: Deep learning with depthwise separable convolutions**. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017. 45

[186] Mingxing Tan and Quoc Le. **Efficientnet: Rethinking model scaling for convolutional neural networks**. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 45, 159

[187] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, et al. **Mobilenets: Efficient convolutional neural networks for mobile vision applications**. *arXiv preprint arXiv:1704.04861*, 2017. 45

[188] Yating Jiang, Yao Lu, and Lingling Yang. **An epileptic seizure prediction model based on a time-wise attention simulation module and a pretrained ResNet**. *Methods*, **202**:117–126, 2022. 45

[189] Bahram Sarvi Zargar, Mohammad Reza Karami Mollaei, Farideh Ebrahimi, and Jalil Rasekhi. **Generalizable epileptic seizures prediction based on deep transfer learning**. *Cognitive Neurodynamics*, pages 1–13, 2022. 45, 152

[190] Rodney J Croft and Robert J Barry. **EOG correction: a new perspective**. *Electroencephalography and clinical Neurophysiology*, **107**(6):387–394, 1998. 49, 50

[191] Davide Vito Moretti, Fabio Babiloni, Filippo Carducci, Febo Cincotti, et al. **Computerized processing of EEG–EOG–EMG artifacts for multi-centric studies in EEG oscillations and event-related potentials**. *International Journal of Psychophysiology*, **47**(3):199–216, 2003. 49, 50

[192] Ben Somers, Tom Francart, and Alexander Bertrand. **A generic EEG artifact removal algorithm based on the multi-channel Wiener filter**. *Journal of neural engineering*, **15**(3):036007, 2018. 49, 51

[193] Joep JM Kierkels, Jamal Riani, Jan WM Bergmans, and Geert JM Van Boxtel. **Using an eye tracker for accurate eye movement artifact correction**. *IEEE Transactions on biomedical engineering*, **54**(7):1256–1267, 2007. 49, 51

[194] Syarifah Noor Syakiylla Sayed Daud and Rubita Sudirman. **Artifact Removal and Brain Rhythm Decomposition for EEG signal Using Wavelet Approach**. *Jurnal Teknologi*, **78**(5-7):135–143, 2016. 49, 52

[195] Murthy Keshava GN and Khan Zaved Ahmed. **Correction of ocular artifacts in EEG signal using empirical mode decomposition and cross-correlation**. *Research Journal of Biotechnology*, **9**(12):21–26, 2014. 49, 52

[196] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. **ICLabel: An automated electroencephalographic independent component classifier, dataset, and website**. *NeuroImage*, **198**:181–197, 2019. 49, 55, 57, 87, 94, 117, 118, 119, 121, 124, 126, 128, 129

[197] Weitong Sun, Yuping Su, Xia Wu, and Xiaojun Wu. **A novel end-to-end 1D-ResCNN model to remove artifact from EEG signals**. *Neurocomputing*, 2020. 49, 58, 59, 98, 102, 112, 113

[198] Nikhil Agrawal, Anil Kumar, Varun Bajaj, and Girish Kumar Singh. **Design of digital IIR filter: A research survey**. *Applied Acoustics*, **172**:107669, 2021. 50

[199] Thomas Elbert, Werner Lutzenberger, Brigitte Rockstroh, and Niels Birbaumer. **Removal of ocular artifacts from the EEG—a biophysical approach to the EOG**. *Electroencephalography and clinical neurophysiology*, **60**(5):455–463, 1985. 50

[200] T Meier, T Rosburg, M Arnold, I Kreitschmann-Andermahr, et al. **Quantification and rejection of ocular artifacts in auditory evoked fields in schizophrenics**. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, **108**(6):526–535, 1998. 50

[201] Ping He, G Wilson, and C Russell. **Removal of ocular artifacts from electro-encephalogram by adaptive filtering**. *Medical and biological engineering and computing*, **42**(3):407–412, 2004. 50

[202] Sadasivan Puthusserypady and Tharmalingam Ratnarajah. **Robust adaptive techniques for minimization of EOG artefacts from EEG signals**. *Signal processing*, **86**(9):2351–2363, 2006. 50

[203] Laurent Peyrodie, Philippe Gallois, Samuel Boudet, Hua Cao, et al. **Evaluation of the AFOP/DAFOP method for automatic filtering of EEGs of patients with epilepsy**. *Journal of Clinical Neurophysiology*, **31**(2):152–161, 2014. 50

[204] Olivier Rosanne, Isabela Albuquerque, Raymundo Cassani, Jean-François Gagnon, et al. **Adaptive filtering for improved eeg-based mental workload assessment of ambulant users**. *Frontiers in Neuroscience*, **15**:611962, 2021. 50

[205] Meltem Izzetoglu, Ajit Devaraj, Scott Bunce, and Banu Onaral. **Motion artifact cancellation in NIR spectroscopy using Wiener filtering**. *IEEE Transactions on Biomedical Engineering*, **52**(5):934–938, 2005. 51

[206] Fabio Morbidi, Andrea Garulli, Domenico Prattichizzo, Cristiano Rizzo, and Simone Rossi. **Application of Kalman filter to remove TMS-induced artifacts from EEG recordings**. *IEEE Transactions on Control Systems Technology*, **16**(6):1360–1366, 2008. 51

[207] Chee-Ming Ting, Sh-Hussain Salleh, Zaitul Marlizawati Zainuddin, and Arifah Bahar. **Artifact removal from single-trial ERPs using non-Gaussian stochastic volatility models and particle filter**. *IEEE Signal Processing Letters*, **21**(8):923–927, 2014. 51

[208] Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999. 51

[209] Doha Safieddine, Amar Kachenoura, Laurent Albera, Gwénaël Birot, et al. **Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches**. *EURASIP Journal on Advances in Signal Processing*, **2012**(1):1–15, 2012. 52

[210] Murielle Kirkove, Clémentine François, and Jacques Verly. **Comparative evaluation of existing and new methods for correcting ocular artifacts in electroencephalographic recordings**. *Signal Processing*, **98**:102–120, 2014. 52

[211] Saleha Khatun, Ruhi Mahajan, and Bashir I Morshed. **Comparative study of wavelet-based unsupervised ocular artifact removal techniques for single-channel EEG data**. *IEEE journal of translational engineering in health and medicine*, **4**:1–8, 2016. 52

[212] Mario Chavez, Fanny Grosselin, Aurore Bussalb, F De Vico Fallani, and Xavier Navarro-Sune. **Surrogate-based artifact removal from single-channel EEG**. *IEEE transactions on neural systems and rehabilitation engineering*, **26**(3):540–550, 2018. 52

[213] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, et al. **The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis**. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, **454**(1971):903–995, 1998. 52

[214] Yannis Kopsinis and Stephen McLaughlin. **Development of EMD-based denoising methods inspired by wavelet thresholding**. *IEEE Transactions on signal Processing*, **57**(4):1351–1362, 2009. 52

[215] Rajesh Patel, Madhukar PandurangRao Janawadkar, Senthilnathan Sengottuvel, Katholil Gireesan, and Thimmakudy Sambasiva Radhakrishnan. **Suppression of eye-blink associated artifact using single channel EEG data by combining cross-correlation with empirical mode decomposition**. *IEEE Sensors Journal*, **16**(18):6947–6954, 2016. 52

[216] Mario Guarascio and Sadasivan Puthusserypady. **Automatic minimization of ocular artifacts from electroencephalogram: A novel approach by combining Complete EEMD with Adaptive Noise and Renyi's Entropy**. *Biomedical Signal Processing and Control*, **36**:63–75, 2017. 52

[217] Arnaud Delorme, Jason Palmer, Julie Onton, Robert Oostenveld, and Scott Makeig. **Independent EEG sources are dipolar**. *PloS one*, **7**(2), 2012. 54, 117

[218] P LeVan, E Urrestarazu, and J Gotman. **A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and Bayesian classification**. *Clinical neurophysiology*, **117**(4):912–927, 2006. 54, 55, 56, 117

[219] Xiaoyan Wei, Lin Zhou, Zhen Zhang, Ziyi Chen, and Yi Zhou. **Early prediction of epileptic seizures using a long-term recurrent convolutional network**. *Journal of neuroscience methods*, **327**:108395, 2019. 54, 66, 69, 76, 83, 134, 137, 149

[220] Seungjin Choi, Andrzej Cichocki, Hyung-Min Park, and Soo-Young Lee. **Blind source separation and independent component analysis: A review**. *Neural Information Processing-Letters and Reviews*, **6**(1):1–57, 2005. 53, 54

[221] Anthony J Bell and Terrence J Sejnowski. **An information-maximization approach to blind separation and blind deconvolution**. *Neural computation*, **7**(6):1129–1159, 1995. 53

[222] Te-Won Lee, Mark Girolami, and Terrence J Sejnowski. **Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources**. *Neural computation*, **11**(2):417–441, 1999. 53, 93

[223] Christopher J James and Christian W Hesse. **Independent component analysis for biomedical signals**. *Physiological measurement*, **26**(1):R15, 2004. 53

[224] Arnaud Delorme, Terrence Sejnowski, and Scott Makeig. **Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis**. *Neuroimage*, **34**(4):1443–1449, 2007. 54

[225] Pascal Chevalier, Laurent Albera, Pierre Comon, and Anne Ferréol. **Comparative performance analysis of eight blind source separation methods on radiocommunications signals**. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, **1**, pages 273–278. IEEE, 2004. 54

[226] Guillermo Sahonero-Alvarez and Humberto Calderon. **A comparison of SOBI, FastICA, JADE and Infomax algorithms**. In *Proceedings*

of the 8th International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2017), pages 21–24, 2017. 54

[227] Nazareth P Castellanos and Valeri A Makarov. **Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis**. *Journal of neuroscience methods*, **158**(2):300–312, 2006. 55

[228] Mehdi Bagheri Hamaneh, Numthip Chitravas, Kitti Kaiboriboon, Samden D Lhatoo, and Kenneth A Loparo. **Automated removal of EKG artifact from EEG data using independent component analysis and continuous wavelet transformation**. *IEEE Transactions on Biomedical Engineering*, **61**(6):1634–1641, 2013. 55

[229] Ruhi Mahajan and Bashir I Morshed. **Unsupervised eye blink artifact denoising of EEG data with modified multiscale sample entropy, kurtosis, and wavelet-ICA**. *IEEE journal of Biomedical and Health Informatics*, **19**(1):158–165, 2014. 55

[230] Nadia Mammone and Francesco C Morabito. **Enhanced automatic wavelet independent component analysis for electroencephalographic artifact removal**. *Entropy*, **16**(12):6553–6572, 2014. 55

[231] Bogdan Mijović, Maarten De Vos, Ivan Gligorijević, Joachim Taelman, and Sabine Van Huffel. **Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis**. *IEEE transactions on biomedical engineering*, **57**(9):2188–2196, 2010. 55

[232] Kevin T Sweeney, Seán F McLoone, and Tomas E Ward. **The use of ensemble empirical mode decomposition with canonical correlation analysis as a novel artifact removal technique**. *IEEE transactions on biomedical engineering*, **60**(1):97–105, 2012. 55

[233] Xun Chen, Chen He, and Hu Peng. **Removal of muscle artifacts from single-channel EEG based on ensemble empirical mode decomposition and multiset canonical correlation analysis**. *Journal of Applied Mathematics*, **2014**, 2014. 55

[234] Gang Wang, Chaolin Teng, Kuo Li, Zhonglin Zhang, and Xiangguo Yan. **The removal of EOG artifacts from EEG signals using independent component analysis and multivariate empirical mode decomposition**. *IEEE journal of biomedical and health informatics*, **20**(5):1301–1308, 2015. 55

[235] Hugh Nolan, Robert Whelan, and Richard B Reilly. **FASTER: fully automated statistical thresholding for EEG artifact rejection**. *Journal of neuroscience methods*, **192**(1):152–162, 2010. 55, 56, 117

[236] Irene Winkler, Stefan Haufe, and Michael Tangermann. **Automatic classification of artifactual ICA-components for artifact removal in EEG signals**. *Behavioral and brain functions*, **7**(1):30, 2011. 55, 56, 87, 98, 102, 117, 118, 124, 129

[237] Andrea Mognon, Jorge Jovicich, Lorenzo Bruzzone, and Marco Buiatti. **ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features**. *Psychophysiology*, **48**(2):229–240, 2011. 55, 56, 117, 118, 124, 129

[238] Maximilien Chaumon, Dorothy VM Bishop, and Niko A Busch. **A practical guide to the selection of independent components of the electroencephalogram for artifact correction**. *Journal of neuroscience methods*, **250**:47–63, 2015. 55, 57, 117

[239] Thea Radüntz, Jon Scouten, Olaf Hochmuth, and Beate Meffert. **EEG artifact elimination by extraction of ICA-component features using image processing algorithms**. *Journal of neuroscience methods*, **243**:84–93, 2015. 55, 57, 101, 117

[240] Laura Frølich, Tobias S Andersen, and Morten Mørup. **Classification of independent components of EEG into multiple artifact classes**. *Psychophysiology*, **52**(1):32–45, 2015. 55, 57, 117

[241] Thea Radüntz, Jon Scouten, Olaf Hochmuth, and Beate Meffert. **Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features**. *Journal of neural engineering*, **14**(4):046004, 2017. 55, 57, 91, 117

[242] Gabriella Tamburro, Patrique Fiedler, David Stone, Jens Haueisen, and Silvia Comani. **A new ICA-based fingerprint method for the automatic removal of physiological artifacts from EEG recordings**. *PeerJ*, **6**:e4380, 2018. 55, 57, 117

[243] Pierpaolo Croce, Filippo Zappasodi, Laura Marzetti, Arcangelo Merla, et al. **Deep Convolutional Neural Networks for feature-less automatic classification of Independent Components in multi-channel electrophysiological brain recordings**. *IEEE Transactions on Biomedical Engineering*, **66**(8):2372–2380, 2018. 55, 57, 91, 117, 119, 121, 128

[244] Sangmin S Lee, Kiwon Lee, and Guiyeom Kang. **EEG artifact removal by Bayesian deep learning & ICA**. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 932–935. IEEE, 2020. 55, 58, 117

[245] Arnaud Delorme and Scott Makeig. **EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis**. *Journal of neuroscience methods*, **134**(1):9–21, 2004. 56, 93, 94

[246] Leor Shoker, Saeid Sanei, and Jonathon Chambers. **Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm**. *IEEE Signal Processing Letters*, **12**(10):721–724, 2005. 56

[247] Banghua Yang, Kaiwen Duan, Chengcheng Fan, Chenxiao Hu, and Jinlong Wang. **Automatic ocular artifacts removal in EEG using deep learning**. *Biomedical Signal Processing and Control*, **43**:148–158, 2018. 58, 59, 98, 112, 113

[248] Rajdeep Ghosh, Nidul Sinha, and Saroj Kumar Biswas. **Automated eye blink artefact removal from EEG using support vector machine and autoencoder**. *IET Signal Processing*, **13**(2):141–148, 2019. 58, 98, 112

[249] Niago Moreira Nobre Leite, Eanes Torres Pereira, Edmar Candeia Gurjao, and Luciana Ribeiro Veloso. **Deep convolutional autoencoder for EEG noise filtering**. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2605–2612. IEEE, 2018. 58, 59, 98, 112

[250] Haoming Zhang, Chen Wei, Mingqi Zhao, Quanying Liu, and Haiyan Wu. **A Novel Convolutional Neural Network Model to Remove Muscle Artifacts from EEG**. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1265–1269. IEEE, 2021. 58, 59, 98

[251] Haoming Zhang, Mingqi Zhao, Chen Wei, Dante Mantini, et al. **EEGdenoiseNet: A Benchmark Dataset for Deep Learning Solutions of EEG Denoising**. *arXiv preprint arXiv:2009.11662*, 2020. 58

[252] Sunil Gandhi, Tim Oates, Tinoosh Mohsenin, and David Hairston. **Denoising time series data using asymmetric generative adversarial networks**. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 285–296. Springer, 2018. 59

[253] Joost B Wagenaar, Gregory A Worrell, Zachary Ives, MATTHIAS Dümpelmann, et al. **Collaborating and sharing data in**

epilepsy research. *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, **32**(3):235, 2015. 60

[254] Ali Hossam Shoeb. *Application of machine learning to epileptic seizure onset detection and treatment*. PhD thesis, Massachusetts Institute of Technology, 2009. 60, 62, 159

[255] Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, et al. **Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state**. *Physical Review E*, **64**(6):061907, 2001. 60, 62

[256] Kaggle. **American Epilepsy Society Seizure Prediction Challenge - Dataset**, 2014. Last accessed on 08-12-2019. 60, 62

[257] Levin Kuhlmann, Philippa Karoly, Dean R Freestone, Benjamin H Brinkmann, et al. **Epilepsyecosystem. org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG**. *Brain*, **141**(9):2619–2630, 2018. 60, 62, 65, 69, 74, 83, 85

[258] Kaat Vandecasteele, Thomas De Cooman, Jonathan Dan, Evy Cleeren, et al. **Visual seizure annotation and automated seizure detection using behind-the-ear electroencephalographic channels**. *Epilepsia*, **61**(4):766–775, 2020. 60, 62, 159

[259] Maryann D'Alessandro, George Vachtsevanos, Rosana Esteller, Javier Echauz, et al. **A multi-feature and multi-channel univariate selection process for seizure prediction**. *Clinical neurophysiology*, **116**(3):506–516, 2005. 61, 89, 135

[260] Tal Pal Attia, Pedro F Viana, Mona Nasseri, Jonas Duun-Henriksen, et al. **Seizure forecasting using minimally invasive, ultra-long-term subcutaneous EEG: Generalizable cross-patient models**. *Epilepsia*, 2022. 61, 67, 78, 85, 134

[261] Mona Nasseri, Tal Pal Attia, Boney Joseph, Nicholas M. Gregg, et al. **Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning**. *Scientific Reports*, **11**(1):21935, 2021. 61

[262] Christian Meisel, Rima El Atrache, Michele Jackson, Sarah Schubach, et al. **Machine learning from wristband sensor data for wearable, noninvasive seizure forecasting**. *Epilepsia*, **61**(12):2653–2666, 2020. 61

[263] Florian Mormann, Thomas Kreuz, Christoph Rieke, Ralph G Andrzejak, et al. **On the predictability of epileptic seizures**. *Clinical neurophysiology*, **116**(3):569–587, 2005. 64, 69, 71, 81, 83, 133

[264] Michel Le Van Quyen, Jason Soss, Vincent Navarro, Richard Robertson, et al. **Preictal state identification by synchronization changes in long-term intracranial EEG recordings**. *Clinical Neurophysiology*, **116**(3):559–568, 2005. 64, 71, 83, 133

[265] Piotr Mirowski, Deepak Madhavan, Yann LeCun, and Ruben Kuzniecky. **Classification of patterns of EEG synchronization for seizure prediction**. *Clinical neurophysiology*, **120**(11):1927–1940, 2009. 64, 71, 82, 85, 134

[266] Yun Park, Lan Luo, Keshab K Parhi, and Theoden Netoff. **Seizure prediction with spectral power of EEG using cost-sensitive support vector machines**. *Epilepsia*, **52**(10):1761–1770, 2011. 64, 69, 71, 83, 84, 85

[267] Ahmed F Rabbi, Leila Azinfar, and Reza Fazel-Rezai. **Seizure prediction using adaptive neuro-fuzzy inference system**. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2100–2103. IEEE, 2013. 64, 72

[268] Jalil Rasekhi, Mohammad Reza Karami Mollaei, Mojtaba Bandarabadi, Cesar A. Teixeira, and Antonio Dourado. **Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods**. *Journal of Neuroscience Methods*, **217**(1-2):9–16, 2013. 64, 72, 83

[269] César Alexandre Teixeira, Bruno Direito, Mojtaba Bandarabadi, Michel Le Van Quyen, et al. **Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients**. *Computer methods and programs in biomedicine*, **114**(3):324–336, 2014. 64, 72, 83, 84, 85, 89, 134, 135, 137, 141, 149, 159

[270] Negin Moghim and David W Corne. **Predicting epileptic seizures in advance**. *PloS one*, **9**(6):e99334, 2014. 64, 72, 82

[271] Mohammad Zavid Parvez and Manoranjan Paul. **Epileptic seizure prediction by exploiting spatiotemporal relationship of EEG signals using phase correlation**. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **24**(1):158–168, 2015. 64, 69, 72, 83, 85, 134

[272] Jalil Rasekhi, MohammadReza Karami Mollaei, Mojtaba Bandarabadi, CésarA Teixeira, and António Dourado. **Epileptic sei-**

zure prediction based on ratio and differential linear univariate features. *Journal of Medical Signals & Sensors*, **5**(1):1, 2015. 64, 72, 82, 85

[273] E Bou Assi, Mohamad Sawan, DK Nguyen, and Sandy Rihana. **A hybrid mRMR-genetic based selection method for the prediction of epileptic seizures**. In *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4. IEEE, 2015. 64, 73, 82

[274] Jian Liang, Rui Lu, Changshui Zhang, and Fei Wang. **Predicting seizures from electroencephalography recordings: a knowledge transfer strategy**. In *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 184–191. IEEE, 2016. 65, 73, 84, 134

[275] Hyunho Chu, Chun Kee Chung, Woorim Jeong, and Kwang-Hyun Cho. **Predicting epileptic seizures from scalp EEG based on attractor state analysis**. *Computer methods and programs in biomedicine*, **143**:75–87, 2017. 65, 73, 85, 114

[276] Bruno Direito, César A Teixeira, Francisco Sales, Miguel Castelo-Branco, and António Dourado. **A realistic seizure prediction study based on multiclass SVM**. *International journal of neural systems*, **27**(03):1750006, 2017. 65, 73, 81, 83, 84, 85, 146

[277] Haidar Khan, Lara Marcuse, Madeline Fields, Kalina Swann, and Bülent Yener. **Focal onset seizure prediction using convolutional networks**. *IEEE Transactions on Biomedical Engineering*, **65**(9):2109–2118, 2017. 65, 73, 85, 134

[278] Mohamad Shahbazi and Hamid Aghajan. **A generalizable model for seizure prediction based on deep learning using cnn-lstm architecture**. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 469–473. IEEE, 2018. 65, 74, 81, 83, 85, 134, 137, 149

[279] Matthias Eberlein, Raphael Hildebrand, Ronald Tetzlaff, Nico Hoffmann, et al. **Convolutional Neural Networks for Epileptic Seizure Prediction**. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2577–2582. IEEE, 2018. 65, 74, 81, 134

[280] Nhan Duy Truong, Anh Duy Nguyen, Levin Kuhlmann, Mohammad Reza Bonyadi, et al. **Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram**. *Neural Networks*, **105**:104–111, 2018. 65, 69, 74, 81, 83, 84, 85, 134, 145, 228

[281] Mingrui Sun, Fuxu Wang, Tengfei Min, Tianyi Zang, and Yadong Wang. **Prediction for high risk clinical symptoms of epilepsy based**

on deep learning algorithm. *IEEE Access*, **6**:77596–77605, 2018. 65, 75, 83, 134

[282] Ahmed Abdelhameed and Magdy Bayoumi. **Semi-supervised deep learning system for epileptic seizures onset prediction**. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1186–1191. IEEE, 2018. 65, 75, 81, 83, 84, 134

[283] Kostas M Tsiouris, Vasileios C Pezoulas, Michalis Zervakis, Spiros Konitsiotis, et al. **A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals**. *Computers in biology and medicine*, **99**:24–37, 2018. 65, 75, 82, 85, 114, 133, 134

[284] Nhan Duy Truong, Levin Kuhlmann, Mohammad Reza Bonyadi, Damien Querlioz, et al. **Epileptic Seizure Forecasting With Generative Adversarial Networks**. *IEEE Access*, **7**:143999–144009, 2019. 66, 75, 81, 84, 134, 137, 146, 147, 149

[285] Petr Nejedly, Vaclav Kremen, Vladimir Sladky, Mona Nasseri, et al. **Deep-learning for seizure forecasting in canines with epilepsy**. *Journal of neural engineering*, **16**(3):036031, 2019. 66, 75, 84, 134, 146

[286] Hisham Daoud and Magdy A Bayoumi. **Efficient epileptic seizure prediction based on deep learning**. *IEEE transactions on biomedical circuits and systems*, **13**(5):804–813, 2019. 66, 75, 83, 84, 134, 152, 158

[287] Yuan Zhang, Yao Guo, Po Yang, Wei Chen, and Benny Lo. **Epilepsy seizure prediction on eeg using common spatial pattern and convolutional neural network**. *IEEE Journal of Biomedical and Health Informatics*, **24**(2):465–474, 2019. 66, 76, 84, 134, 137, 149

[288] Yankun Xu, Jie Yang, Shiqi Zhao, Hemmings Wu, and Mohamad Sawan. **An end-to-end deep learning approach for epileptic seizure prediction**. In *2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, pages 266–270. IEEE, 2020. 66, 69, 76

[289] Ratnaprabha Ravindra Borhade and Manoj S Nagmode. **Modified Atom Search Optimization-based Deep Recurrent Neural Network for epileptic seizure prediction using electroencephalogram signals**. *Biocybernetics and Biomedical Engineering*, **40**(4):1638–1653, 2020. 66, 69, 76, 85, 134

[290] Syed Muhammad Usman, Shehzad Khalid, and Zafar Bashir. **Epileptic seizure prediction using scalp electroencephalogram signals**. *Biocybernetics and Biomedical Engineering*, **41**(1):211–220, 2021. 66, 69, 76

[291] Xin Chen, Yuanjie Zheng, Changxu Dong, and Sutao Song. **Multi-dimensional enhanced seizure prediction framework based on graph convolutional network**. *Frontiers in Neuroinformatics*, **15**, 2021. 66, 69, 77

[292] Chenchen Cheng, Bo You, Yan Liu, and Yakang Dai. **Patient-specific method of sleep electroencephalography using wavelet packet transform and Bi-LSTM for epileptic seizure prediction**. *Biomedical Signal Processing and Control*, **70**:102963, 2021. 66, 77

[293] Theekshana Dissanayake, Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes. **Geometric Deep Learning for Subject Independent Epileptic Seizure Prediction Using Scalp EEG Signals**. *IEEE Journal of Biomedical and Health Informatics*, **26**(2):527–538, 2021. 67, 69, 77, 134, 158

[294] Ramy Hussein, Soojin Lee, Rabab Ward, and Martin J McKeown. **Semi-dilated convolutional neural networks for epileptic seizure prediction**. *Neural Networks*, **139**:212–222, 2021. 67, 69, 77

[295] Ziyu Wang, Jie Yang, Hemmings Wu, Junming Zhu, and Mohamad Sawan. **Power efficient refined seizure prediction algorithm based on an enhanced benchmarking**. *Scientific Reports*, **11**(1):1–9, 2021. 67, 78, 84, 134

[296] Shiqi Zhao, Jie Yang, and Mohamad Sawan. **Energy-efficient neural network for epileptic seizure prediction**. *IEEE Transactions on Biomedical Engineering*, **69**(1):401–411, 2021. 67, 78, 83

[297] Pedro F Viana, Tal Pal Attia, Mona Nasseri, Jonas Duun-Henriksen, et al. **Seizure forecasting using minimally invasive, ultra-long-term subcutaneous electroencephalography: Individualized intrapatient models**. *Epilepsia*, 2022. 67, 78, 85, 134

[298] Jens Müller, Hongliu Yang, Matthias Eberlein, Georg Leonhardt, et al. **Coherent false seizure prediction in epilepsy, coincidence or providence?** *Clinical Neurophysiology*, **133**:157–164, 2022. 67, 69, 78, 148

[299] Peizhen Peng, Yang Song, Lu Yang, and Haikun Wei. **Seizure prediction in EEG signals using STFT and domain adaptation**. *Frontiers in Neuroscience*, **15**:1880, 2022. 67, 79

[300] Abir Affes, Afef Mdhaffar, Chahnez Triki, Mohamed Jmaiel, and Bernd Freisleben. **Personalized attention-based EEG channel selection for epileptic seizure prediction**. *Expert Systems with Applications*, page 117733, 2022. 68, 79

[301] Deng Liang, Aiping Liu, Chang Li, Jun Liu, and Xun Chen. **A novel consistency-based training strategy for seizure prediction**. *Journal of Neuroscience Methods*, **372**:109557, 2022. 68, 69, 79

[302] Daniel Uvaydov, Raffaele Guida, Pedram Johari, Francesco Restuccia, and Tommaso Melodia. **AiEEG: Personalized Seizure Prediction Through Partially-Reconfigurable Deep Neural Networks**. In *2022 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 77–88. IEEE, 2022. 68, 79

[303] Imene Jemal, Neila Mezghani, Lina Abou-Abbas, and Amar Mitiche. **An interpretable deep learning classifier for epileptic seizure prediction using EEG data**. *IEEE Access*, 2022. 68, 79, 83

[304] Ines Assali, Ahmed Ghazi Blaiech, Asma Ben Abdallah, Khaled Ben Khalifa, et al. **CNN-based classification of epileptic states for seizure prediction using combined temporal and spectral features**. *Biomedical Signal Processing and Control*, **82**:104519, 2023. 68, 79

[305] Chang Li, Chenghao Shao, Rencheng Song, Guoping Xu, et al. **Spatio-temporal MLP network for seizure prediction using EEG signals**. *Measurement*, **206**:112278, 2023. 68, 69, 80, 85

[306] Xin Xu, Yin Zhang, Ruhao Zhang, and Tingting Xu. **Patient-specific method for predicting epileptic seizures based on DRSN-GRU**. *Biomedical Signal Processing and Control*, **81**:104449, 2023. 68, 80

[307] Mojtaba Bandarabadi, Jalil Rasekhi, César A Teixeira, Mohammad R Karami, and António Dourado. **On the proper selection of preictal period for seizure prediction**. *Epilepsy & Behavior*, **46**:158–166, 2015. 69, 137

[308] Visar Berisha, Chelsea Krantsevich, P Richard Hahn, Shira Hahn, et al. **Digital medicine and the curse of dimensionality**. *NPJ digital medicine*, **4**(1):153, 2021. 82

[309] François Perrin, J Pernier, O Bertrand, and JF Echallier. **Spherical splines for scalp potential and current density mapping**. *Electroencephalography and clinical neurophysiology*, **72**(2):184–187, 1989. 93

[310] Fabio Lopes, Adriana Leal, Júlio Medeiros, Mauro F Pinto, et al. **EPIC: Annotated epileptic EEG independent components for artifact reduction**, 2022. 96

[311] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. **Rectifier Nonlinearities Improve Neural Network Acoustic Models**. In *Proceedings*

of the 30th International Conference on Machine Learning, **28**, pages 1–6, Atlanta, Georgia, USA, 2013. Citeseer. 99, 121

[312] Guoqiang Zhong, Xiao Ling, and Li-Na Wang. **From Shallow Feature Learning to Deep Learning: Benefits from the Width and Depth of Deep Architectures**. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **9**(1):e1255, 2019. 100

[313] Kenji Kawaguchi, Jiaoyang Huang, and Leslie Pack Kaelbling. **Effect of Depth and Width on Local Minima in Deep Learning**. *Neural Computation*, **31**(7):1462–1498, 2019. 100

[314] Karen Simonyan and Andrew Zisserman. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. *arXiv preprint arXiv:1409.1556v6*, pages 1–14, 2014. 100

[315] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. **Empirical Evaluation of Rectified Activations in Convolution Network**. In *arXiv preprint arXiv:1505.00853v2*, pages 1–5. Deep Learning Workshop, ICML 2015, 2015. 100, 121

[316] Diederik P Kingma and Jimmy Ba. **Adam: A Method for Stochastic Optimization**. *arXiv preprint arXiv:1412.6980v9*, pages 1–15, 2015. 100, 123, 140

[317] Malik Muhammad Naeem Mannan, Muhammad Ahmad Kamran, and Myung Yung Jeong. **Identification and Removal of Physiological Artifacts from Electroencephalogram Signals: A Review**. *IEEE Access*, **6**:30630–30652, 2018. 101

[318] Hsin-Tien Chiang, Yi-Yen Hsieh, Szu-Wei Fu, Kuo-Hsuan Hung, et al. **Noise Reduction in ECG Signals using Fully Convolutional Denoising Autoencoders**. *IEEE Access*, **7**:60806–60813, 2019. 101

[319] Patrick E McKight and Julius Najab. **Kruskal-wallis test**. *The corsini encyclopedia of psychology*, pages 1–1, 2010. 111, 125

[320] Alexis Dinno. **Nonparametric pairwise multiple comparisons in independent groups using Dunn's test**. *The Stata Journal*, **15**(1):292–300, 2015. 111, 125

[321] Paul A Merolla, John V Arthur, Rodrigo Alvarez-Icaza, Andrew S Cassidy, et al. **A Million Spiking-Neuron Integrated Circuit with a Scalable Communication Network and Interface**. *Science*, **345**(6197):668–673, 2014. 113

[322] Steven K Esser, Paul A Merolla, John V Arthur, Andrew S Cassidy, et al. **Convolutional Networks for Fast, Energy-Efficient Neuromorphic Computing**. *Proceedings of the National Academy of Sciences*, **113**(41):11441–11446, 2016. 114

[323] Gianluca Borghini, Pietro Aricò, Gianluca Di Flumeri, Nicolina Sciaraffa, et al. **A New Perspective for the Training Assessment: Machine Learning-Based Neurometric for Augmented User's Evaluation**. *Frontiers in Neuroscience*, **11**:325, 2017. 114

[324] Gianluca Di Flumeri, Gianluca Borghini, Pietro Aricò, Nicolina Sciaraffa, et al. **EEG-based Mental Workload Neurometric to Evaluate the Impact of Different Traffic and Road Conditions in Real Driving Settings**. *Frontiers in Human Neuroscience*, **12**:509, 2018. 114

[325] Nir Friedman, Tomer Fekete, Kobi Gal, and Oren Shriki. **EEG-Based Prediction of Cognitive Load in Intelligence Tests**. *Frontiers in Human Neuroscience*, **13**:191, 2019. 114

[326] Pietro Aricò, Maxime Reynal, Gianluca Di Flumeri, Gianluca Borghini, et al. **How Neurophysiological Measures Can Be Used to Enhance the Evaluation of Remote Tower Solutions**. *Frontiers in Human Neuroscience*, **13**:303, 2019. 114

[327] Kin Ming Puk, Kellen C Gandy, Shouyi Wang, and Heekyeong Park. **Pattern Classification and Analysis of Memory Processing in Depression using EEG Signals**. In *International Conference on Brain Informatics*, pages 124–137. Springer, 2016. 114

[328] Kaare B Mikkelsen, Preben Kidmose, and Lars K Hansen. **On the Keyhole Hypothesis: High Mutual Information between Ear and Scalp EEG**. *Frontiers in Human Neuroscience*, **11**:341, 2017. 114

[329] Nicolas Zink, Ann-Kathrin Stock, Amirali Vahid, and Christian Beste. **On the Neurophysiological Mechanisms Underlying the Adaptability to Varying Cognitive Control Demands**. *Frontiers in Human Neuroscience*, **12**:411, 2018. 114

[330] Siddharth Kohli and Alexander J Casson. **Machine Learning Validation of EEG+ tACS Artefact Removal**. *Journal of Neural Engineering*, **17**:16034, 2020. 114

[331] Izabela Rejer and Paweł Górski. **MAICA: an ICA-based Method for Source Separation in a Low-channel EEG Recording**. *Journal of Neural Engineering*, **16**(5):056025, 2019. 114

[332] Marius Klug and Klaus Gramann. **Identifying Key Factors for Improving ICA-based Decomposition of EEG Data in Mobile and Stationary Experiments**. *European Journal of Neuroscience*, 2020. 114

[333] Ricardo Couceiro, Raul Barbosa, João Durães, Gonçalo Duarte, et al. **Spotting Problematic Code Lines using Nonintrusive Programmers' Biofeedback**. In *2019 IEEE 30th International Symposium on Software Reliability Engineering (ISSRE)*, pages 93–103. IEEE, 2019. 119

[334] Júlio Medeiros, Ricardo Couceiro, Gonçalo Duarte, João Durães, et al. **Can EEG Be Adopted as a Neuroscience Reference for Assessing Software Programmers' Cognitive Load?** *Sensors*, **21**(7):2338, 2021. 119

[335] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. **Learning deep features for discriminative localization**. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016. 122

[336] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. **Deep Feedforward Networks**. In *Deep Learning*, chapter 6, pages 163–220. MIT Press, 2016. 122

[337] Mike X Cohen. *Analyzing neural time series data: Theory and practice*. Number 1 in Issues in clinical and cognitive neuropsychology. MIT Press, 2014. 129

[338] Alina Pauline Liebisch, Thomas Eggert, Alina Shindy, Elia Valentini, et al. **A novel tool for the removal of muscle artefacts from EEG: Improving data quality in the gamma frequency range**. *Journal of Neuroscience Methods*, **358**:1–7, 2021. 129

[339] Ramy Hussein, Mohamed Osama Ahmed, Rabab Ward, Z Jane Wang, et al. **Human intracranial eeg quantitative analysis and automatic feature learning for epileptic seizure prediction**. *arXiv preprint arXiv:1904.03603*, 2019. 134

[340] Nhan Duy Truong, Yikai Yang, Christina Maher, Levin Kuhlmann, et al. **Seizure susceptibility prediction in uncontrolled epilepsy**. *Frontiers in Neurology*, **12**, 2021. 134, 137, 149

[341] Milind Natu, Mrinal Bachute, Shilpa Gite, Ketan Kotecha, and Ankit Vidyarthi. **Review on epileptic seizure prediction: machine learning and deep learning approaches**. *Computational and Mathematical Methods in Medicine*, **2022**, 2022. 134

[342] Maxime O Baud and Vikram R Rao. **Gauging seizure risk**. *Neurology*, **91**(21):967–973, 2018. 84, 149

[343] Margherita Contento, Bruno Bertaccini, Martina Biggi, Matteo Magliani, et al. **Prediction of seizure recurrence risk following discontinuation of antiepileptic drugs**. *Epilepsia*, **62**(9):2159–2170, 2021. 84

[344] Rachel E Stirling, Mark J Cook, David B Grayden, and Philippa J Karoly. **Seizure forecasting and cyclic control of seizures**. *Epilepsia*, **62**:S2–S14, 2021. 84

[345] Klaus Lehnertz, Henning Dickten, Stephan Porz, Christoph Helmstaedter, and Christian E Elger. **Predictability of uncontrollable multifocal seizures–towards new treatment options**. *Scientific reports*, **6**(1):1–9, 2016. 136

[346] Adriana Leal, Mauro F Pinto, Fábio Lopes, Anna M Bianchi, et al. **Heart rate variability analysis for the identification of the preictal interval in patients with drug-resistant epilepsy**. *Scientific reports*, **11**(1):1–11, 2021. 137, 149

[347] Adriana Leal, Juliana Curty, Fábio Lopes, Mauro F Pinto, et al. **Unsupervised EEG preictal interval identification in patients with drug-resistant epilepsy**. *Scientific Reports*, **13**(1):784, 2023. 137, 149

[348] Mauro F Pinto, Adriana Leal, Fábio Lopes, José Pais, et al. **On the clinical acceptance of black-box systems for EEG seizure prediction**. *Epilepsia Open*, 2022. 137

[349] Robert F Woolson. **Wilcoxon signed-rank test**. *Wiley encyclopedia of clinical trials*, pages 1–3, 2007. 143, 157

[350] Hayder K Fatlawi and Attila Kiss. **An Adaptive Classification Model for Predicting Epileptic Seizures Using Cloud Computing Service Architecture**. *Applied Sciences*, **12**(7):3408, 2022. 148

[351] Tal Pal Attia, Daniel Crepeau, Vaclav Kremen, Mona Nasseri, et al. **Epilepsy personal assistant device—A mobile platform for brain state, dense behavioral and physiology tracking and controlling adaptive stimulation**. *Frontiers in Neurology*, **12**:704170, 2021. 148

[352] RAB Badawy, DR Freestone, A Lai, and MJ Cook. **Epilepsy: ever-changing states of cortical excitability**. *Neuroscience*, **222**:89–99, 2012. 149

[353] Julien QM Ly, Giulia Gaggioni, Sarah L Chellappa, Soterios Papachilleos, et al. **Circadian regulation of human cortical excitability**. *Nature communications*, **7**(1):1–10, 2016. 149

[354] Zitong Wan, Rui Yang, Mengjie Huang, Nianyin Zeng, and Xiaohui Liu. **A review on transfer learning in EEG signal analysis**. *Neurocomputing*, **421**:1–14, 2021. 151

[355] Khansa Rasheed, Junaid Qadir, Terence J O'Brien, Levin Kuhlmann, and Adeel Razi. **A generative model to synthesize eeg data for epileptic seizure prediction**. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **29**:2322–2332, 2021. 152, 158

[356] CA Teixeira, B Direito, H Feldwisch-Drentrup, M Valderrama, et al. **EPILAB: A software package for studies on the prediction of epileptic seizures**. *Journal of Neuroscience Methods*, **200**(2):257–271, 2011. 155

[357] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. **Densely connected convolutional networks**. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 159

[358] Philippa J. Karoly, Vikram R. Rao, Nicholas M. Gregg, Gregory A. Worrell, et al. **Cycles in epilepsy**. *Nature Reviews Neurology*, **17**(5):267–284, 2021. 213, 229

Appendix A

Automatic electroencephalogram artefact removal using deep convolutional neural networks

A.1 Training and validation

Figure A.1 summarises the training and validation procedure presented in the Section 5.2.3. First, the training dataset was filtered by removing electroencephalogram (EEG) segments with little brain information. Then, the dataset was divided into training and validation subsets. Afterwards, the samples from both subsets were zero-padded to ensure that all samples had the same size. After that, the data were standardised and used for training the model. These procedures were repeated ten times in order to produce ten different models.

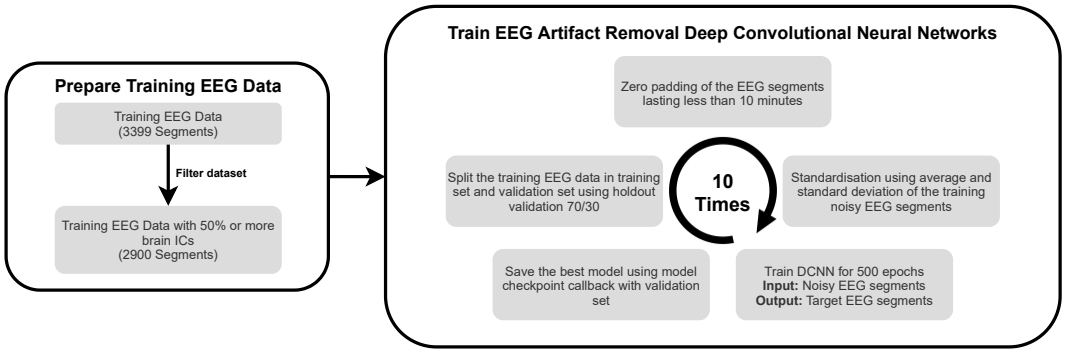


Figure A.1: Framework followed to develop the artefact removal model. It covers the following steps: preparation of training set and training and validation of the developed approach.

A.2 Evaluation procedure

Figure A.2 presents the evaluation procedure performed to assess the developed approach using the test set. The root mean squared error (RMSE), the relative root mean squared error (RRMSE), the Pearson correlation coefficient (PCC), and the signal-to-noise ratio (SNR) difference were used to evaluate the approach. The metrics were computed for the ten different models and were averaged. The performance of the developed approach was also compared with the performance of other artefact removal models.

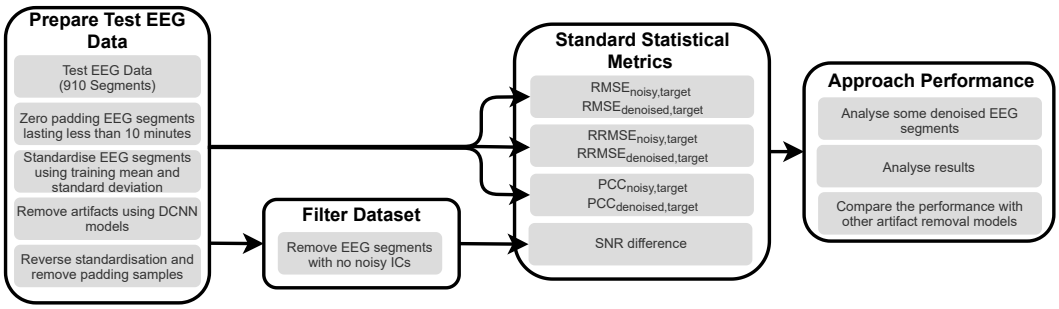


Figure A.2: Framework followed to evaluate the developed approach. It covers the following steps: preparation of the test set, evaluation metrics and performance assessment.

A.3 Results

This section presents the mean, median, standard deviation, first quartile, and third quartile for all the evaluation metrics used to assess the developed model.

Table A.1: Mean, standard deviation, first quartile, median and third quartile of the RMSE of the EEG segments, for each EEG channel, before using the DCNN model. SD: Standard deviation; 1Q: First quartile; 3Q: Third quartile.

	EEG Channel																		
	Fp1	**Fp2**	**F3**	**F4**	**C3**	**C4**	**P3**	**P4**	**O1**	**O2**	**F7**	**F8**	**T7**	**T8**	**P7**	**P8**	**Fz**	**Cz**	**Pz**
Mean	24.64	23.80	12.85	10.51	10.15	10.62	7.55	7.63	10.83	8.81	17.89	17.27	14.61	14.00	10.79	9.97	7.43	5.33	7.10
SD	19.24	16.10	16.02	10.26	18.64	14.84	10.79	8.79	29.44	12.48	15.38	14.34	15.85	14.24	16.51	11.90	21.36	9.10	11.19
1Q	8.78	8.21	3.00	2.95	2.17	2.12	1.89	2.01	2.35	2.26	6.94	6.79	3.38	3.48	2.32	2.33	1.23	1.35	1.70
Median	26.31	26.80	9.27	8.52	5.49	6.00	6.28	6.33	7.72	7.35	15.78	15.75	10.15	10.56	7.44	7.73	3.31	4.11	5.89
3Q	34.39	35.19	16.60	15.64	9.30	11.75	8.95	9.08	10.78	10.60	25.58	24.22	19.41	19.80	11.35	11.88	6.78	5.69	8.15

Table A.2: Mean, standard deviation, first quartile, median and third quartile of the RMSE of the EEG segments, for each EEG channel, after using the DCNN model. SD: Standard deviation; 1Q: First quartile; 3Q: Third quartile.

	EEG Channel																		
	Fp1	**Fp2**	**F3**	**F4**	**C3**	**C4**	**P3**	**P4**	**O1**	**O2**	**F7**	**F8**	**T7**	**T8**	**P7**	**P8**	**Fz**	**Cz**	**Pz**
Mean	5.43	5.64	5.12	4.98	4.23	5.02	3.68	3.81	4.85	4.83	6.14	5.88	6.90	6.78	5.11	5.11	3.50	2.56	3.17
SD	3.09	3.16	4.00	3.04	4.08	4.46	3.33	3.58	5.29	4.24	4.28	3.46	5.44	5.14	4.65	4.42	2.84	2.20	2.70
1Q	3.32	3.37	2.77	2.67	1.83	2.05	1.64	1.59	1.98	2.05	3.35	3.30	2.79	2.75	2.04	1.97	1.75	1.29	1.62
Median	4.62	4.67	4.30	4.77	3.12	3.54	2.77	2.77	3.39	3.34	5.24	5.25	5.40	5.35	3.78	3.81	2.87	2.06	2.47
3Q	6.50	6.86	6.04	6.49	5.16	6.43	4.29	4.45	5.87	6.33	7.81	7.67	9.20	9.54	6.36	6.77	4.25	2.91	3.72

Table A.3: Mean, standard deviation, first quartile, median and third quartile of the RRMSE of the EEG segments, for each EEG channel, before using the DCNN model. SD: Standard deviation; 1Q: First quartile; 3Q: Third quartile.

	EEG Channel																		
	Fp1	**Fp2**	**F3**	**F4**	**C3**	**C4**	**P3**	**P4**	**O1**	**O2**	**F7**	**F8**	**T7**	**T8**	**P7**	**P8**	**Fz**	**Cz**	**Pz**
Mean	4.52	4.41	2.09	1.86	2.02	1.91	1.04	1.20	1.09	0.83	2.22	2.33	2.16	2.00	1.44	1.06	1.07	0.96	1.17
SD	4.00	3.95	3.34	2.09	4.12	2.95	2.04	3.15	2.85	1.36	2.76	2.45	3.23	2.78	4.41	1.69	3.00	2.33	2.94
1Q	1.03	1.09	0.38	0.34	0.30	0.30	0.22	0.23	0.19	0.20	0.71	0.71	0.39	0.39	0.25	0.21	0.11	0.15	0.18
Median	4.06	3.94	1.06	1.03	0.80	0.86	0.84	0.80	0.68	0.62	1.84	1.92	0.87	0.96	0.73	0.68	0.39	0.60	0.84
3Q	6.95	6.72	2.81	2.92	1.36	1.91	1.20	1.20	1.08	0.93	2.90	3.26	3.23	2.79	1.03	1.03	1.09	0.93	1.18

Table A.4: Mean, standard deviation, first quartile, median and third quartile of the RRMSE of the EEG segments, for each EEG channel, after using the DCNN model. SD: Standard deviation; 1Q: First quartile; 3Q: Third quartile.

	EEG Channel																		
	Fp1	**Fp2**	**F3**	**F4**	**C3**	**C4**	**P3**	**P4**	**O1**	**O2**	**F7**	**F8**	**T7**	**T8**	**P7**	**P8**	**Fz**	**Cz**	**Pz**
Mean	0.77	0.80	0.74	0.78	0.76	0.85	0.46	0.55	0.47	0.41	0.67	0.70	0.95	0.93	0.57	0.50	0.46	0.41	0.45
SD	0.37	0.40	0.77	0.64	0.97	1.01	0.50	1.13	0.62	0.47	0.56	0.44	1.08	1.10	0.76	0.58	0.41	0.52	0.70
1Q	0.48	0.50	0.32	0.30	0.27	0.28	0.19	0.19	0.17	0.18	0.44	0.45	0.30	0.30	0.21	0.19	0.15	0.15	0.17
Median	0.74	0.78	0.63	0.65	0.45	0.47	0.33	0.34	0.28	0.31	0.59	0.63	0.46	0.45	0.38	0.36	0.37	0.28	0.31
3Q	0.98	1.05	0.89	1.08	0.71	0.94	0.53	0.51	0.48	0.45	0.74	0.82	1.47	1.26	0.53	0.53	0.65	0.45	0.51

Table A.5: Mean, standard deviation, first quartile, median and third quartile of the PCC of the EEG segments, for each EEG channel, before using the DCNN model. SD: Standard deviation; 1Q: First quartile; 3Q: Third quartile.

	EEG Channel																		
	Fp1	**Fp2**	**F3**	**F4**	**C3**	**C4**	**P3**	**P4**	**O1**	**O2**	**F7**	**F8**	**T7**	**T8**	**P7**	**P8**	**Fz**	**Cz**	**Pz**
Mean	0.41	0.41	0.62	0.63	0.71	0.68	0.77	0.76	0.79	0.81	0.55	0.54	0.64	0.65	0.76	0.77	0.82	0.81	0.77
SD	0.33	0.33	0.31	0.31	0.29	0.30	0.21	0.23	0.21	0.20	0.28	0.29	0.32	0.30	0.25	0.23	0.23	0.21	0.22
1Q	0.15	0.15	0.33	0.32	0.59	0.46	0.64	0.64	0.68	0.73	0.33	0.29	0.29	0.34	0.70	0.70	0.68	0.73	0.64
Median	0.24	0.25	0.68	0.70	0.78	0.76	0.76	0.78	0.83	0.85	0.48	0.46	0.75	0.72	0.81	0.82	0.93	0.86	0.76
3Q	0.69	0.68	0.93	0.95	0.96	0.96	0.98	0.97	0.98	0.98	0.81	0.82	0.93	0.93	0.97	0.98	0.99	0.99	0.98

Table A.6: Mean, standard deviation, first quartile, median and third quartile of the PCC of the EEG segments, for each EEG channel, after using the DCNN model. SD: Standard deviation; 1Q: First quartile; 3Q: Third quartile.

	EEG Channel																		
	Fp1	**Fp2**	**F3**	**F4**	**C3**	**C4**	**P3**	**P4**	**O1**	**O2**	**F7**	**F8**	**T7**	**T8**	**P7**	**P8**	**Fz**	**Cz**	**Pz**
Mean	0.65	0.61	0.75	0.74	0.79	0.77	0.88	0.87	0.89	0.91	0.78	0.76	0.74	0.74	0.86	0.87	0.86	0.91	0.90
SD	0.27	0.30	0.22	0.23	0.26	0.27	0.16	0.18	0.17	0.14	0.19	0.20	0.29	0.29	0.21	0.19	0.16	0.15	0.15
1Q	0.47	0.39	0.62	0.57	0.74	0.62	0.85	0.87	0.88	0.90	0.70	0.66	0.43	0.46	0.86	0.86	0.78	0.90	0.87
Median	0.70	0.69	0.80	0.79	0.91	0.89	0.95	0.94	0.96	0.96	0.83	0.80	0.91	0.90	0.93	0.94	0.94	0.96	0.95
3Q	0.88	0.88	0.95	0.96	0.97	0.97	0.98	0.98	0.99	0.99	0.91	0.90	0.96	0.96	0.98	0.98	0.99	0.99	0.99

Table A.7: Mean, standard deviation, first quartile, median and third quartile of the SNR difference of the EEG segments, for each EEG channel, for the DCNN model. SD: Standard deviation; 1Q: First quartile; 3Q: Third quartile.

	EEG Channel																		
	Fp1	**Fp2**	**F3**	**F4**	**C3**	**C4**	**P3**	**P4**	**O1**	**O2**	**F7**	**F8**	**T7**	**T8**	**P7**	**P8**	**Fz**	**Cz**	**Pz**
Mean	11.30	10.71	4.57	3.64	4.00	3.50	4.15	4.36	3.84	3.31	6.89	7.13	3.91	3.96	3.71	3.66	0.50	3.55	4.34
SD	9.58	9.69	8.22	7.43	6.20	6.21	6.94	6.32	7.34	7.11	9.20	8.26	6.64	6.55	6.72	6.09	7.81	7.02	7.23
1Q	7.38	7.22	1.24	0.83	2.10	1.54	2.16	2.44	2.04	0.93	4.64	4.85	2.59	2.63	1.86	1.65	-1.56	1.31	2.09
Median	14.24	13.31	6.74	5.90	5.37	5.22	5.71	5.64	4.96	4.88	9.92	9.90	5.70	5.60	5.04	4.86	1.69	5.60	6.39
3Q	17.50	16.84	10.46	8.58	7.15	7.12	7.94	8.12	8.27	7.65	12.21	11.99	7.50	7.43	7.68	7.38	4.68	7.41	8.72

A.4 Examples of reconstructed EEG segments

This section presents more examples of EEG segments reconstructed by the developed model.

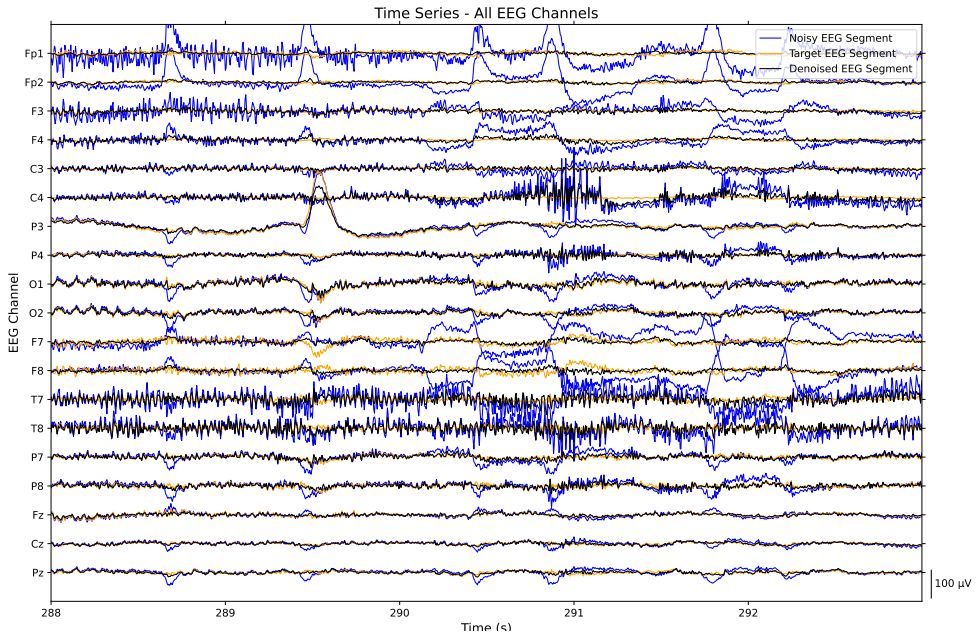


Figure A.3: Five seconds of all channels of an example EEG segment containing eye blinks, eye movements and muscle activity. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

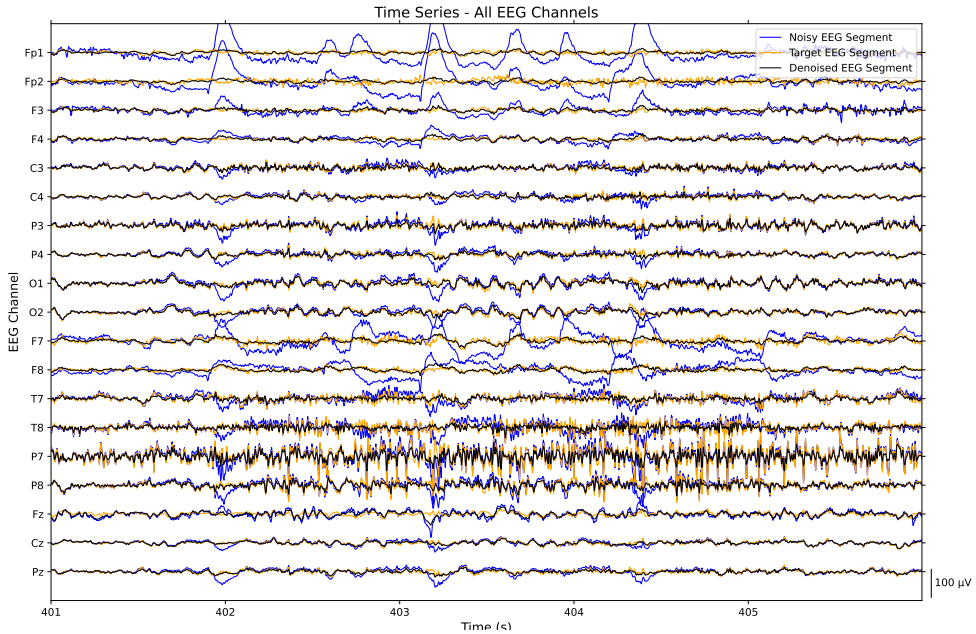


Figure A.4: Five seconds of all channels of an example EEG segment containing eye blinks, eye movements and muscle activity. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

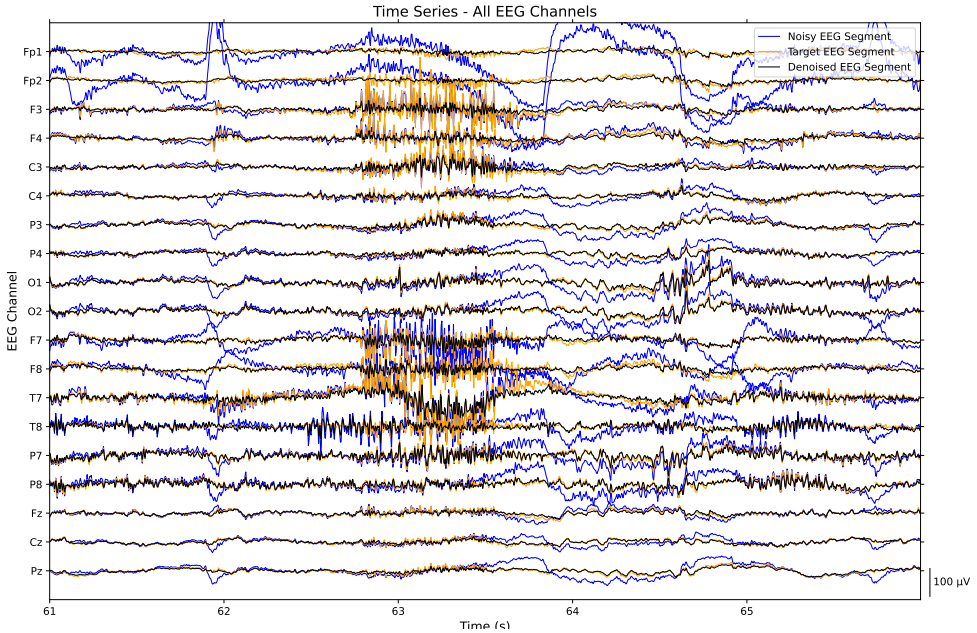


Figure A.5: Five seconds of all channels of an example EEG segment containing eye blinks, eye movements, muscle activity and electrode movements. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

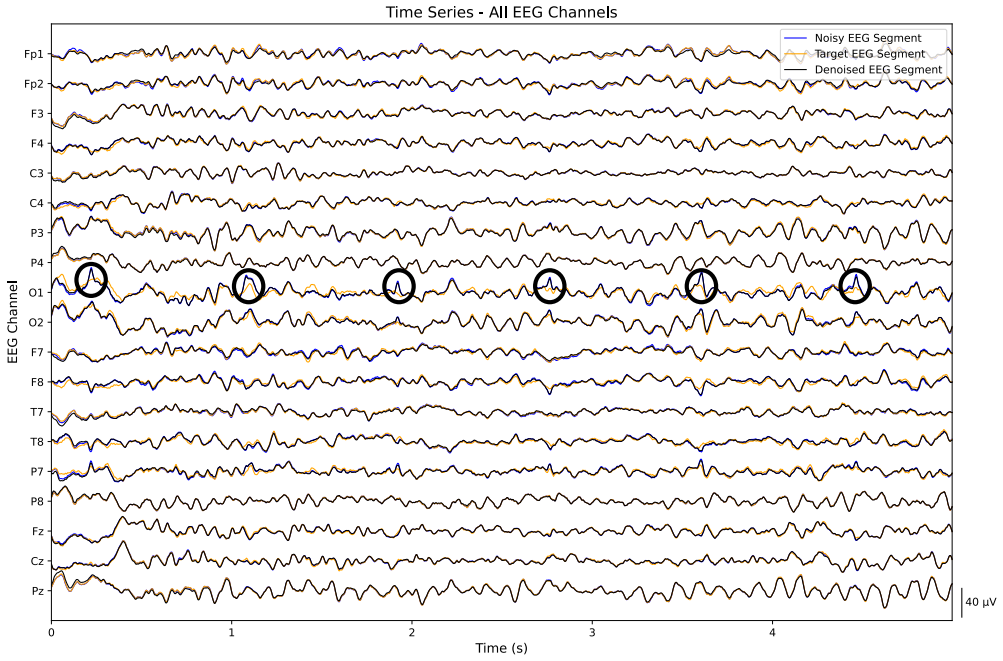


Figure A.6: Five seconds of all channels of an example EEG segment. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively. The selected portions provide the exact moments when the cardiac artefacts occur.

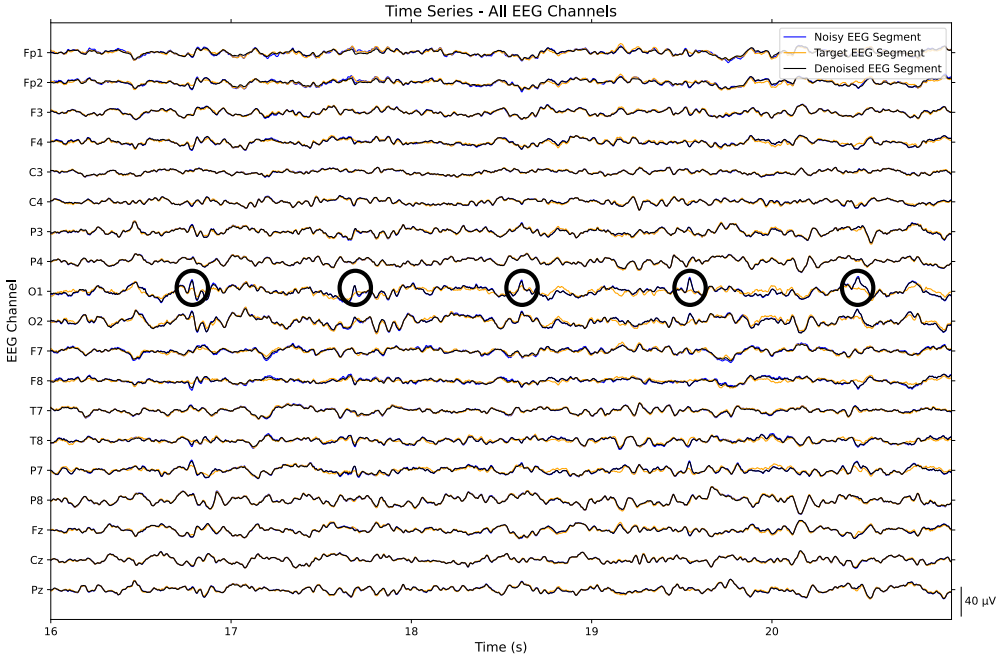


Figure A.7: Five seconds of all channels of an example EEG segment. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively. The selected portions provide the exact moments when the cardiac artefacts occur.

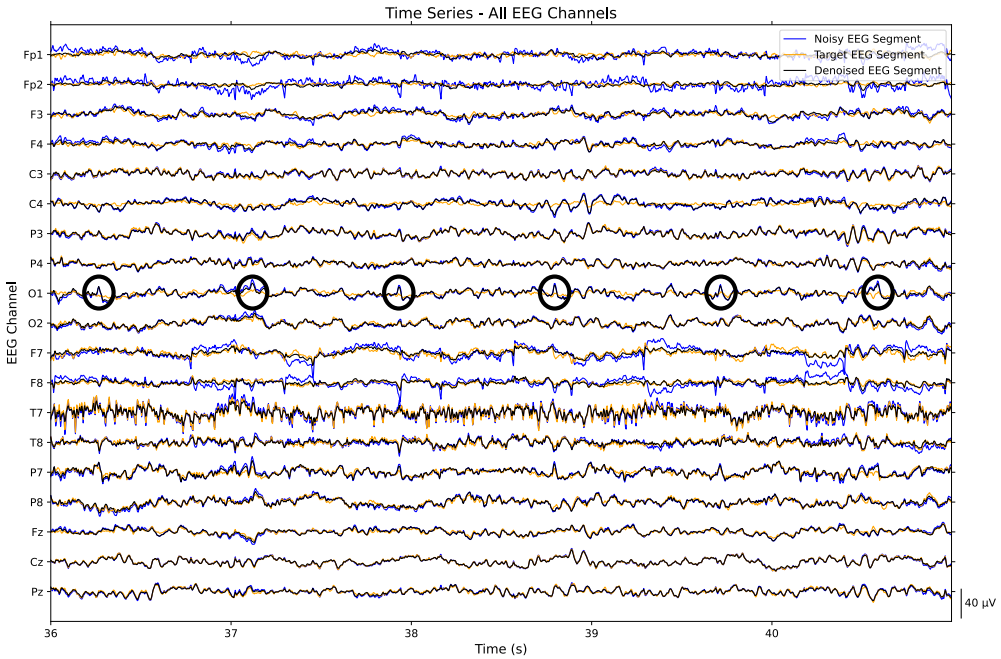


Figure A.8: Five seconds of all channels of an example EEG segment. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively. The selected portions provide the exact moments when the cardiac artefacts occur.

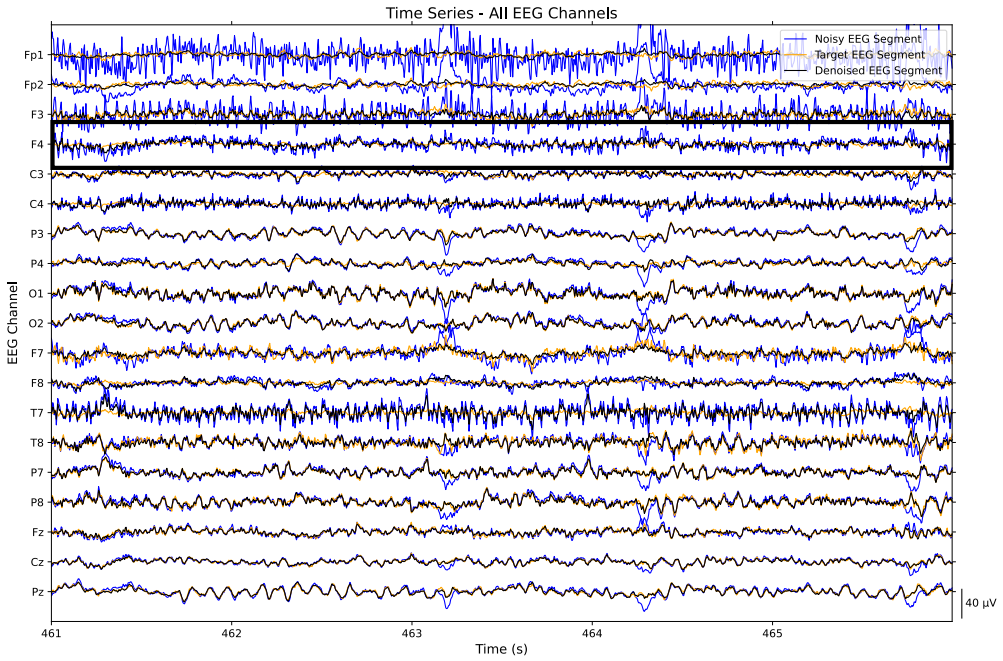


Figure A.9: Five seconds of all channels of an example EEG segment. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively. The selected portion evidences the EEG channel where the pulse artefacts occur.

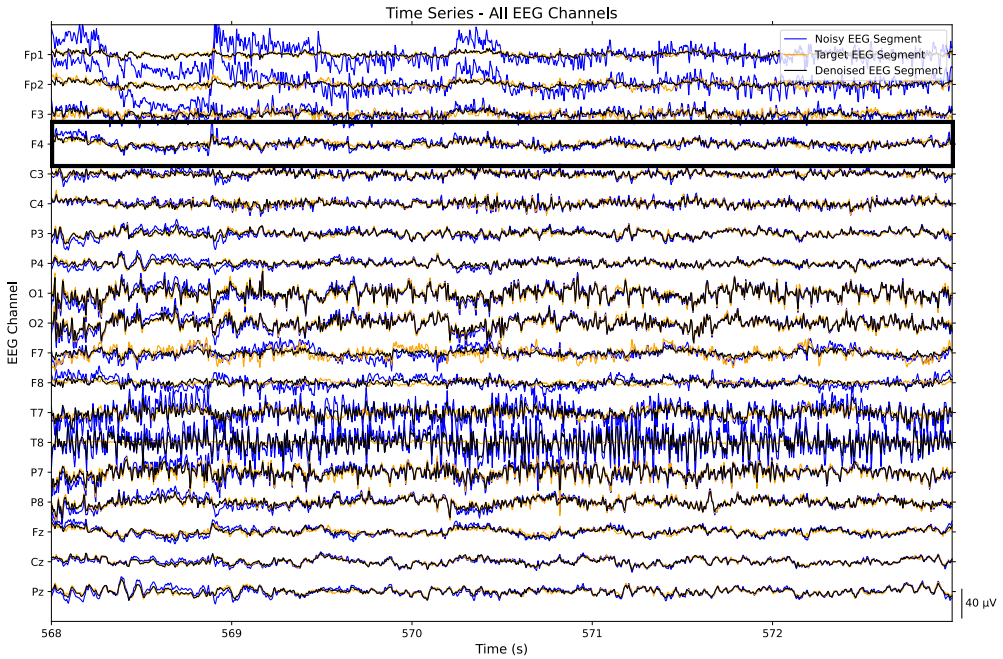


Figure A.10: Five seconds of all channels of an example EEG segment. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively. The selected portion evidences the EEG channel where the pulse artefacts occur.

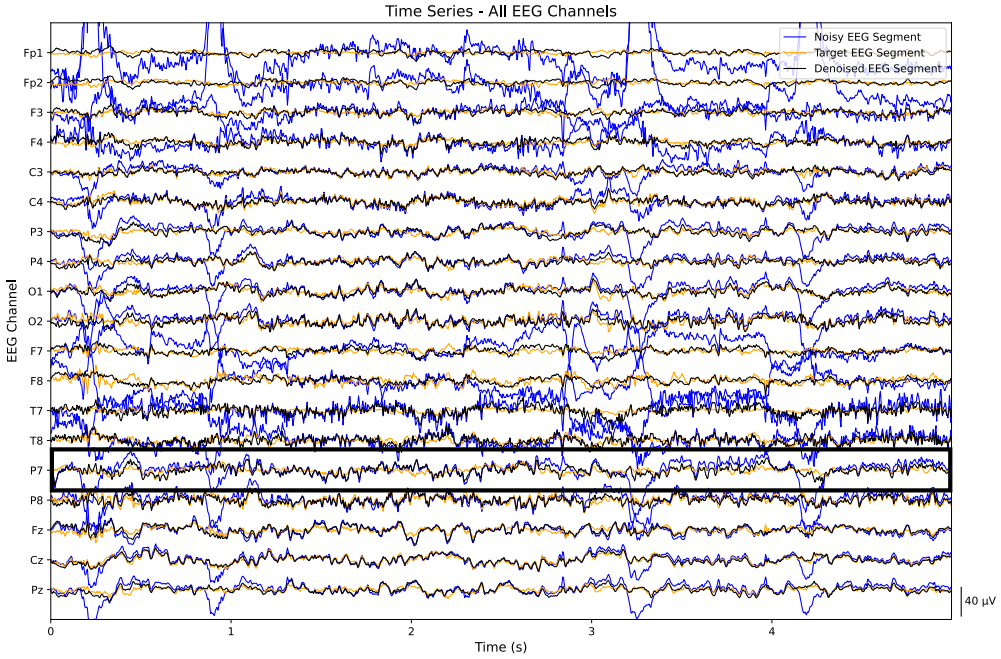


Figure A.11: Five seconds of all channels of an example EEG segment. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively. The selected portion evidences the EEG channel where the pulse artefacts occur.

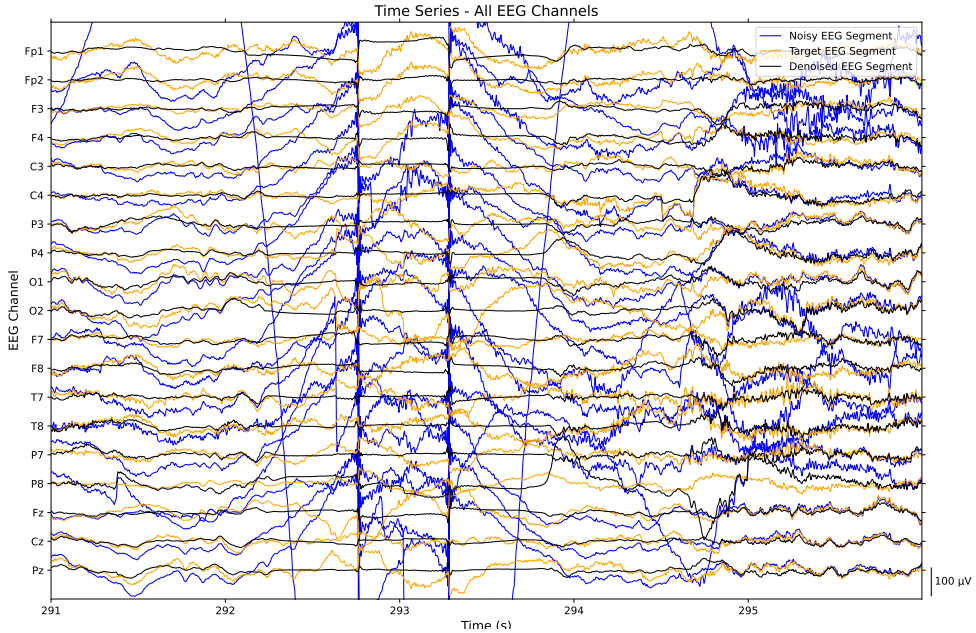


Figure A.12: Five seconds of all channels of an example EEG segment containing experimental errors which were not removed with the initial EEG preprocessing algorithm. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

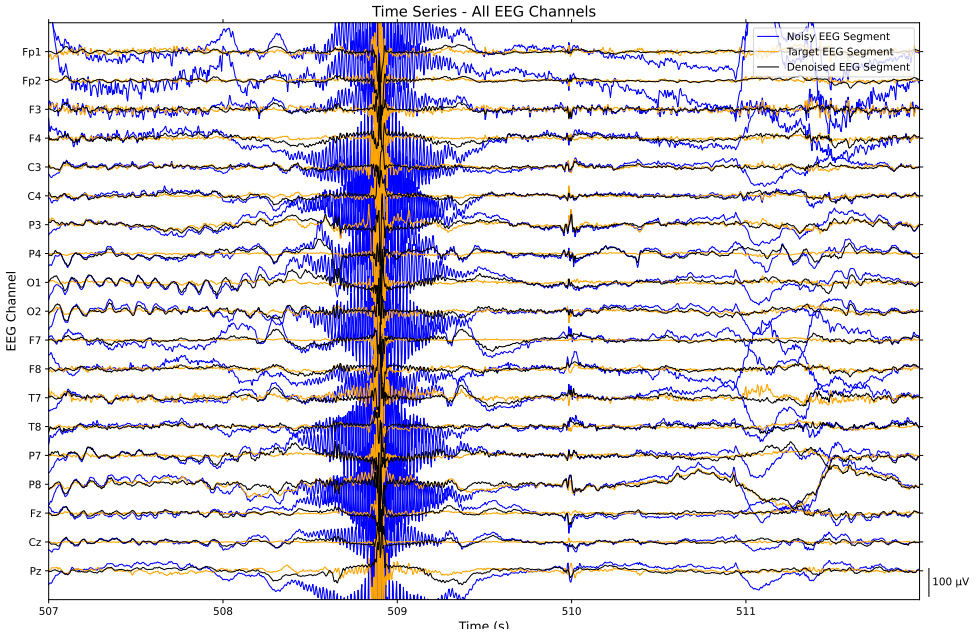


Figure A.13: Five seconds of all channels of an example EEG segment containing experimental errors which were not removed with the EEG preprocessing algorithm. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

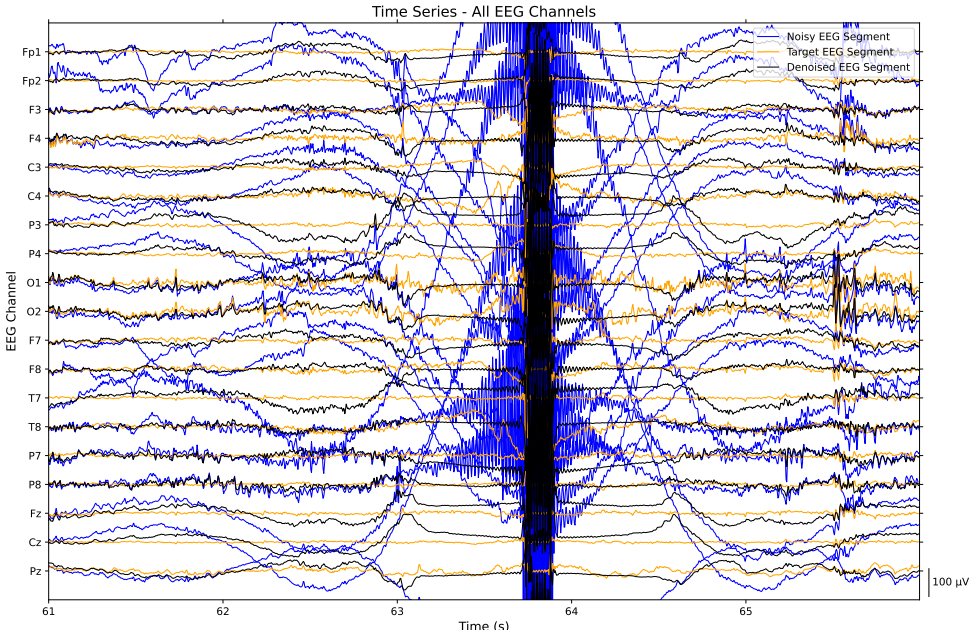


Figure A.14: Five seconds of all channels of an example EEG segment containing experimental errors which were not removed with the EEG preprocessing algorithm. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

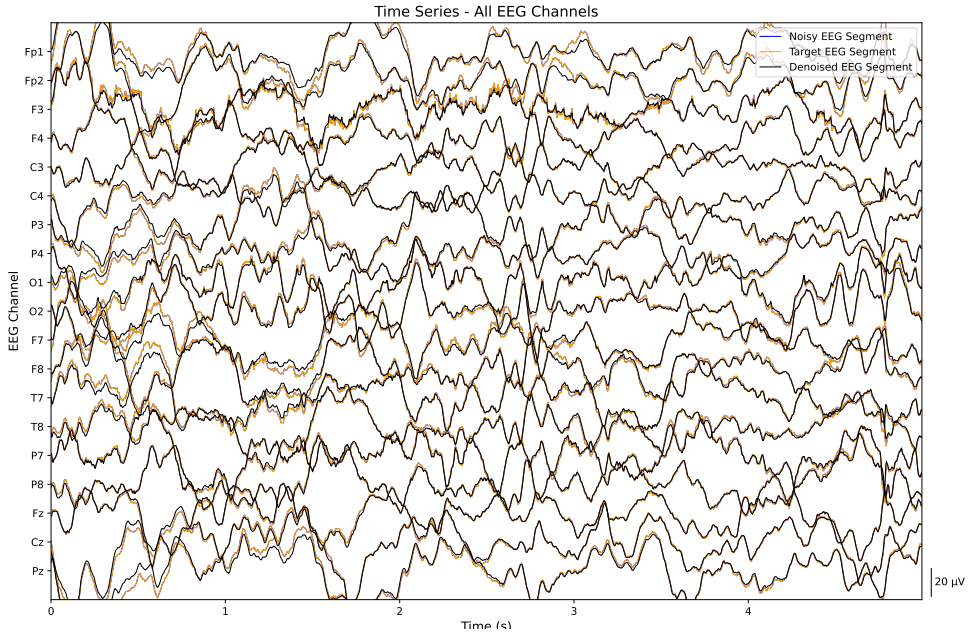


Figure A.15: Five seconds of all channels of an example EEG segment that does not contain any noisy artefact. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

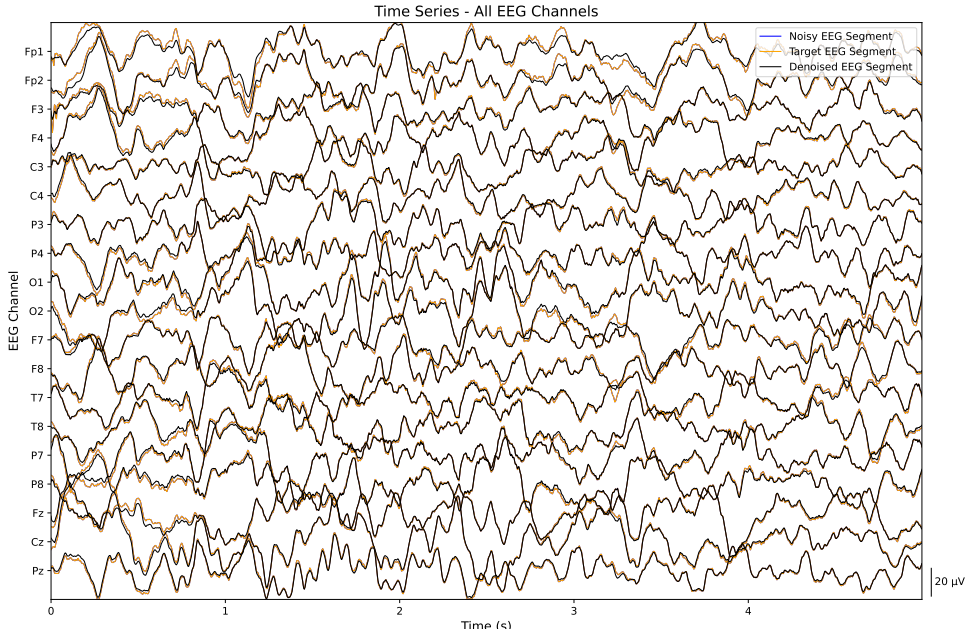


Figure A.16: Five seconds of all channels of an example EEG segment that does not contain any noisy artefact. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

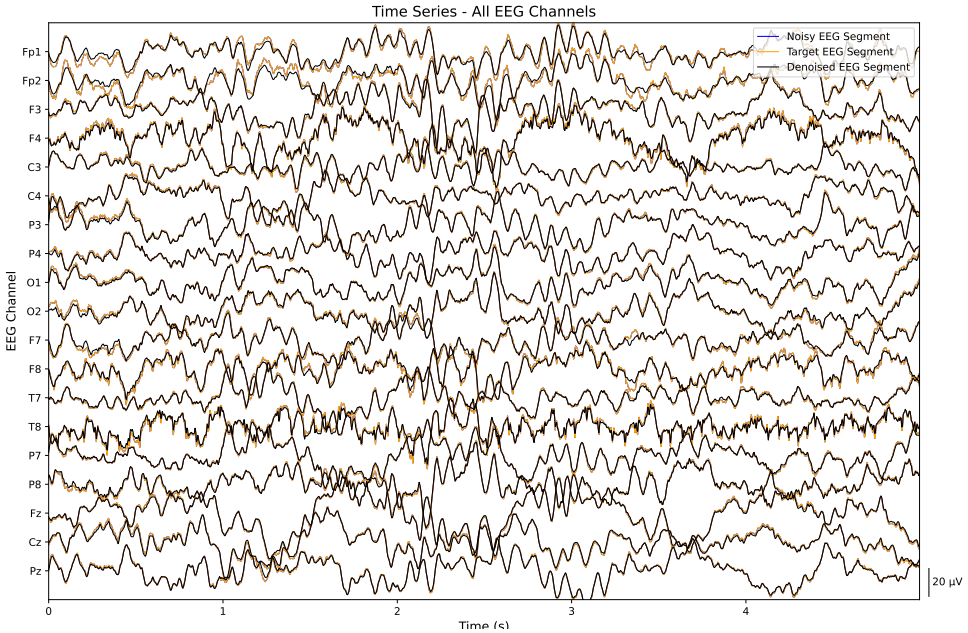


Figure A.17: Five seconds of all channels of an example EEG segment that does not contain any noisy artefact. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

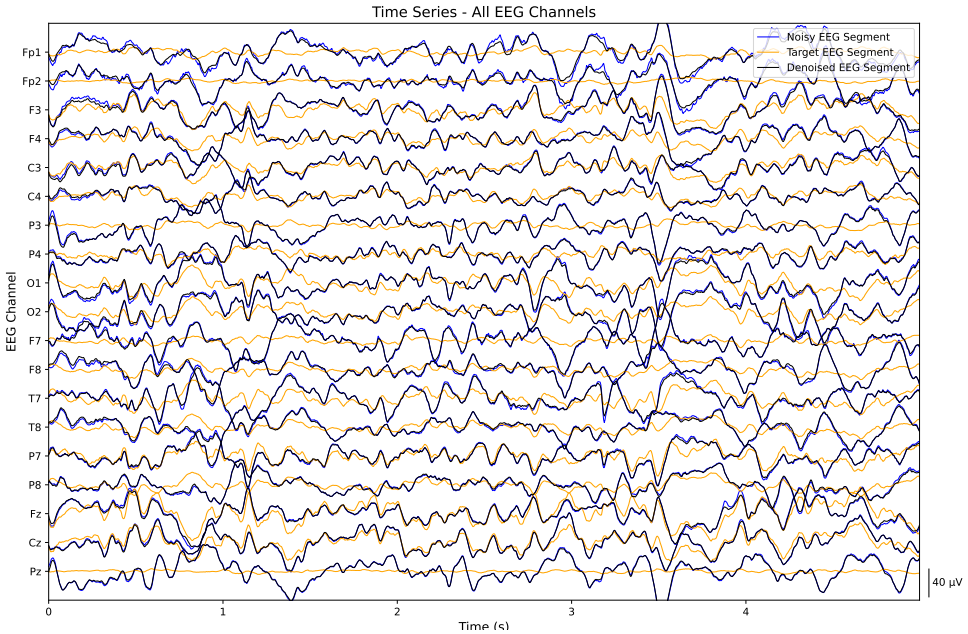


Figure A.18: Five seconds of all channels of an example EEG segment which had some brain information removed by visual inspection that was not removed by the EEG artefact removal model. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

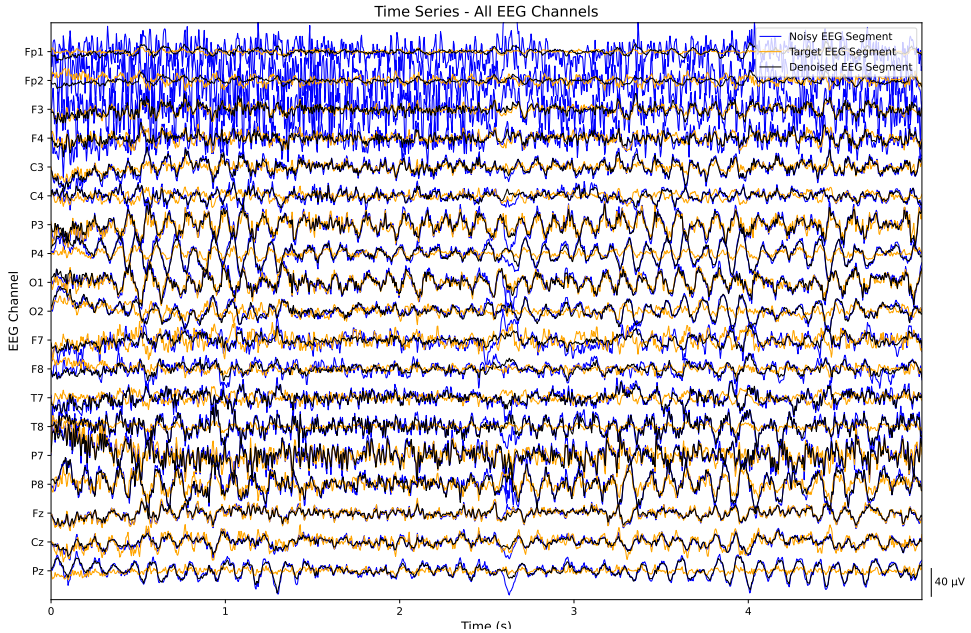


Figure A.19: Five seconds of all channels of an example EEG segment which had some brain information removed by visual inspection that was not removed by the EEG artefact removal model. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

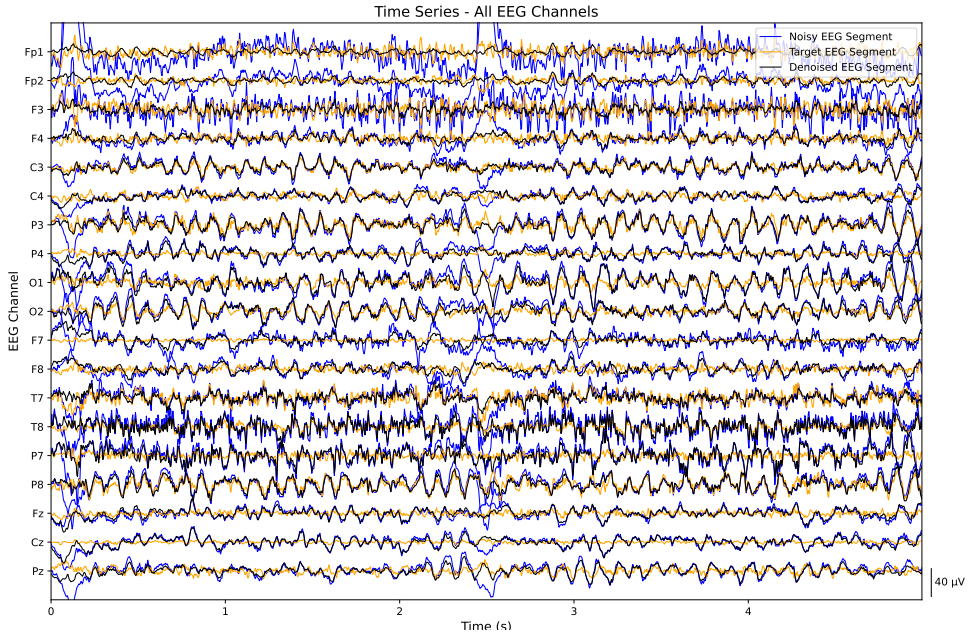


Figure A.20: Five seconds of all channels of an example EEG segment which had some brain information removed by visual inspection that was not removed by the EEG artefact removal model. The noisy segment, target segment and denoised segment are represented in blue, orange and black, respectively.

Appendix B

Removing artefacts and periodically retraining improve performance of neural network-based seizure prediction models

B.1 Patient and seizure metadata

Table B.1 contains information regarding the patients analysed in this study. The table includes information on sex, age at hospital admission and onset age (corresponding to the occurrence of the first epilepsy event), epilepsy foci lateralisation, the total number of annotated seizures and the number of lead seizures analysed for each patient, according to the considered minimum inter-seizure interval of 4.5 hours. The duration of the used electroencephalogram (EEG) data regarding training and testing seizures is also presented. It is worth noting that although it was considered lead seizures to happen at least 4.5 hours after the previous one, only 4 hours were used for training.

Table B.2 describes data collected for each analysed seizure. It includes information about the EEG onset time, used inter-seizure time, vigilance state at the time of the seizure onset, seizure classification, and seizure activity pattern. Seizures were classified according to the International League Against Epilepsy (ILAE) classification [47]. The vigilance state corresponds to one of the following states of alertness and responsiveness: wakefulness, non-rapid eye movement (NREM) sleep, further subdivided into three sleep stages (N1–3) and rapid eye movement (REM) sleep [358].

Table B.1: Dataset description regarding each patient.

P	ID	Sex	Onset Age (years)	Admission Age (years)	Lat.	#Sz	#LSz	Training time (dd hh/mm/ss)	Testing time (dd hh/mm/ss)
1	402	F	10	55	L, R	5	5	00 12:00:00	01 05:43:57
2	8902	F	23	67	L	5	5	00 12:00:00	00 22:33:21
3	11 002	M	21	41	R	8	4	00 08:00:00	00 19:51:27
4	16 202	F	43	46	L, R	8	7	00 16:00:00	00 21:20:11
5	21 902	M	44	47	L	6	4	00 08:00:00	00 17:40:40
6	23 902	M	36	36	L	5	5	00 12:00:00	01 09:59:10
7	26 102	M	15	65	L	8	4	00 08:00:00	01 21:24:37
8	30 802	M	28	28	L, R	9	8	00 20:00:00	01 19:55:36
9	32 702	F	33	62	L, R	6	5	00 12:00:00	00 22:15:49
10	45 402	F	13	41	L, R	5	4	00 08:00:00	01 04:34:30
11	46 702	F	13	15	R	5	5	00 12:00:00	00 12:39:01
12	50 802	M	2	43	L	5	5	00 12:00:00	01 11:39:33
13	52 302	F	13	61	L	7	3	00 08:00:00	00 09:33:47
14	53 402	M	0	39	L, R	8	4	00 08:00:00	02 02:30:52
15	55 202	F	3	17	R, B	9	8	00 20:00:00	01 18:54:12
16	56 402	M	18	47	L, R	7	4	00 08:00:00	06 04:28:44
17	58 602	M	17	32	L	22	6	00 16:00:00	00 10:46:12
18	59 102	M	17	47	R	7	5	00 12:00:00	03 10:15:50
19	60 002	M	47	55	L, R	8	6	00 16:00:00	05 15:37:09
20	64 702	M	3	51	R	6	5	00 12:00:00	01 07:41:43
21	75 202	M	10	13	R	8	7	00 16:00:00	01 07:21:15
22	80 702	F	14	22	B	10	6	00 16:00:00	00 19:26:12
23	81 102	F	14	22	B	13	3	00 08:00:00	00 23:26:03
24	85 202	F	4	54	L	10	5	00 12:00:00	00 20:27:13
25	93 402	M	40	67	L	7	5	00 12:00:00	02 06:06:02
26	93 902	M	43	50	R	9	6	00 16:00:00	00 13:32:53
27	94 402	F	29	37	R	11	7	00 16:00:00	00 23:44:49
28	95 202	F	13	50	L	14	7	00 16:00:00	03 06:16:09
29	96 002	M	21	58	L, R	9	7	00 16:00:00	03 04:32:26
30	98 102	M	2	36	L	5	5	00 12:00:00	01 21:48:06
31	98 202	M	3	39	R	10	8	00 20:00:00	01 07:41:52
32	101 702	M	44	52	L, R	6	5	00 12:00:00	00 23:52:54
33	102 202	M	0	17	L	28	7	00 16:00:00	01 18:32:49
34	104 602	F	8	17	L	5	5	00 12:00:00	00 15:19:10
35	109 502	M	40	50	L, R	10	4	00 08:00:00	01 23:09:24
36	110 602	M	6	56	R	8	5	00 12:00:00	01 01:58:25
37	112 802	M	47	52	L	6	6	00 16:00:00	02 00:46:59
38	113 902	F	16	29	R	25	6	00 16:00:00	00 14:47:58
39	114 702	F	31	22	R	25	8	00 20:00:00	01 00:49:43
40	114 902	F	15	16	L, R	12	7	00 16:00:00	01 16:58:57
41	123 902	F	7	25	L, R	8	5	00 12:00:00	01 06:12:37

ID: patient identifier. Sex: female (F) or male (M). Lateralisation (Lat.): L: left, R: right, B: bilateral. #Sz: total number of seizures annotated per patient. #LSz: number of leading seizures, obtained as a result of the analysis of 4.5 hours of interseizure data.

Table B.2: Dataset description regarding data preceding each seizure. The gray rows were used for training, while the other ones were used for testing.

S	ID	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
1	402	22:45:26	00 04:00:00	W	FOIA	t
2	402	21:27:34	00 04:00:00	W	FBTC	t
3	402	02:13:30	00 04:00:00	W	FOIA	t
4	402	08:53:21	00 06:09:51	W	FBTC	t
5	402	08:57:27	00 23:34:06	W	FOIA	t
6	8902	23:51:14	00 04:00:00	W	UC	a
7	8902	23:03:23	00 04:00:00	W	FOIA	b
8	8902	05:37:05	00 04:00:00	W	FOIA	a
9	8902	00:35:56	00 18:28:50	W	FOIA	m
10	8902	05:10:26	00 04:04:30	W	FOIA	a
11	11 002	00:00:10	00 04:00:00	W	UC	?
12	11 002	06:38:01	00 04:00:00	R	FOIA	s
13	11 002	15:16:42	00 08:08:40	W	FOIA	a
14	11 002	08:18:49	00 11:42:46	W	FOIA	t
15	16 202	04:34:07	00 04:00:00	W	UC	r
16	16 202	06:05:10	00 04:00:00	W	FBTC	?
17	16 202	05:07:14	00 04:00:00	W	UC	r
18	16 202	18:48:33	00 04:00:00	W	FOIA	r
19	16 202	03:34:35	00 08:16:01	W	FOIA	r
20	16 202	13:50:31	00 07:57:01	W	FOIA	?
21	16 202	19:27:39	00 05:07:08	W	FOIA	r
22	21 902	16:16:43	00 04:00:00	W	UC	t
23	21 902	08:40:51	00 04:00:00	W	FOIA	t
24	21 902	20:32:56	00 07:53:23	W	FOIA	t
25	21 902	06:50:12	00 09:47:15	R	FOIA	b
26	23 902	10:18:13	00 04:00:00	W	FOA	t
27	23 902	20:50:38	00 04:00:00	W	FOA	t
28	23 902	11:18:12	00 04:00:00	W	FOA	t
29	23 902	16:48:02	01 04:59:49	W	FOA	d

S: seizure. ID: patient identifier. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

Continued on next page

S	ID	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
30	23 902	22:17:22	00 04:59:20	W	FOA	t
31	26 102	15:31:37	00 04:00:00	W	FOIA	m
32	26 102	08:33:50	00 04:00:00	W	FOIA	t
33	26 102	07:52:54	00 22:49:04	W	FOIA	t
34	26 102	11:36:45	00 22:35:32	W	FOIA	t
35	30 802	04:33:31	00 04:00:00	R	FOA	t
36	30 802	04:52:24	00 04:00:00	W	FOA	t
37	30 802	10:58:12	00 04:00:00	N2	FOA	t
38	30 802	22:58:11	00 04:00:00	W	FOA	t
39	30 802	05:49:34	00 04:00:00	W	FOA	t
40	30 802	02:48:42	00 20:29:08	R	FOA	t
41	30 802	07:48:06	00 04:29:24	N2	FOA	t
42	30 802	03:15:10	00 18:57:04	N2	FOA	t
43	32 702	08:25:28	00 04:00:00	W	FOIA	t
44	32 702	10:22:47	00 04:00:00	W	FOIA	t
45	32 702	10:13:13	00 04:00:00	W	FOIA	t
46	32 702	17:03:16	00 06:20:03	W	FOIA	r
47	32 702	09:29:02	00 15:55:46	W	FOIA	a
48	45 402	01:48:55	00 04:00:00	W	FOIA	t
49	45 402	08:11:29	00 04:00:00	W	FOIA	t
50	45 402	14:56:37	00 06:15:07	W	FOA	t
51	45 402	15:13:34	00 22:19:21	W	FOIA	t
52	46 702	15:56:40	00 04:00:00	W	FOA	a
53	46 702	06:16:40	00 04:00:00	N2	FOIA	a
54	46 702	17:06:57	00 04:00:00	W	FOIA	t
55	46 702	02:02:23	00 08:25:26	N2	FBTC	b
56	46 702	06:45:59	00 04:13:35	W	FOIA	t
57	50 802	02:44:39	00 04:00:00	W	FOIA	t
58	50 802	06:37:35	00 04:00:00	N2	UC	t
59	50 802	12:39:04	00 04:00:00	N2	UC	t
60	50 802	22:50:41	00 09:41:37	N2	FOIA	t

S: seizure. ID: patient identifier. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

Continued on next page

S	ID	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
61	50 802	01:18:38	01 01:57:56	W	FBTC	t
62	52 302	06:29:39	00 04:00:00	W	UC	?
63	52 302	11:31:13	00 04:00:00	W	FOA	?
64	52 302	02:31:34	00 09:33:47	N1	UC	d
65	53 402	08:16:32	00 04:00:00	W	FOA	?
66	53 402	05:46:33	00 04:00:00	N2	FOA	?
67	53 402	19:02:38	01 12:46:05	W	FOA	?
68	53 402	09:17:43	00 13:44:46	W	FOIA	t
69	55 202	07:02:49	00 04:00:00	W	FOIA	t
70	55 202	09:55:11	00 04:00:00	W	FOIA	d
71	55 202	18:15:11	00 04:00:00	W	FOA	t
72	55 202	08:09:27	00 04:00:00	W	UC	t
73	55 202	17:47:47	00 04:00:00	W	UC	t
74	55 202	09:57:39	00 15:39:52	W	FOA	t
75	55 202	15:34:54	00 05:07:15	W	UC	r
76	55 202	14:11:59	00 22:07:05	W	FOIA	r
77	56 402	08:17:30	00 04:00:00	W	UC	t
78	56 402	21:11:53	00 04:00:00	W	UC	?
79	56 402	09:13:46	05 07:13:23	W	UC	?
80	56 402	06:29:39	00 20:15:19	W	FBTC	a
81	58 602	09:11:25	00 04:00:00	W	FOIA	r
82	58 602	03:29:21	00 04:00:00	R	FOIA	t
83	58 602	19:52:52	00 04:00:00	W	FOIA	t
84	58 602	09:01:07	00 04:00:00	W	FOIA	r
85	58 602	15:41:02	00 04:50:43	W	FOIA	r
86	58 602	02:31:58	00 05:55:28	N2	FOIA	t
87	59 102	08:54:51	00 04:00:00	W	FOA	?
88	59 102	15:41:55	00 04:00:00	W	FOIA	t
89	59 102	09:56:35	00 04:00:00	W	FOIA	t
90	59 102	19:51:41	00 09:25:05	W	FOIA	t
91	59 102	21:12:26	03 00:50:44	W	FOA	t

S: seizure. ID: patient identifier. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

Continued on next page

S	ID	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
92	60 002	02:45:01	00 04:00:00	N1	FOIA	d
93	60 002	02:22:55	00 04:00:00	W	FOIA	c
94	60 002	12:21:36	00 04:00:00	W	FOIA	t
95	60 002	05:40:53	00 04:00:00	R	UC	t
96	60 002	00:17:54	00 18:07:01	R	FOIA	d
97	60 002	22:18:46	04 21:30:51	N1	FOIA	d
98	64 702	13:53:39	00 04:00:00	W	FOA	?
99	64 702	04:23:21	00 04:00:00	W	FBTC	m
100	64 702	18:59:43	00 04:00:00	W	FBTC	t
101	64 702	19:50:01	01 00:20:17	W	FBTC	t
102	64 702	03:41:27	00 07:21:26	N2	FBTC	t
103	75 202	23:37:38	00 04:00:00	N2	FOA	t
104	75 202	01:10:45	00 04:00:00	N2	FOA	t
105	75 202	21:33:44	00 04:00:00	W	UC	t
106	75 202	19:27:00	00 04:00:00	W	FOA	t
107	75 202	09:46:19	00 13:49:19	W	FOA	t
108	75 202	17:43:46	00 05:20:22	W	FOA	?
109	75 202	06:25:19	00 12:11:33	W	FOA	t
110	80 702	05:03:56	00 04:00:00	W	FOIA	b
111	80 702	08:43:22	00 04:00:00	W	FOIA	b
112	80 702	20:43:38	00 04:00:00	W	UC	?
113	80 702	07:46:14	00 04:00:00	W	FOIA	c
114	80 702	17:54:17	00 04:56:33	W	FBTC	c
115	80 702	08:53:56	00 14:29:38	W	FOIA	c
116	81 102	20:48:50	00 04:00:00	W	FOIA	t
117	81 102	10:44:57	00 04:00:00	W	FOA	t
118	81 102	10:42:15	00 23:26:03	W	FOIA	t
119	85 202	23:37:05	00 04:00:00	N2	FOIA	m
120	85 202	16:51:04	00 04:00:00	W	FOIA	c
121	85 202	04:24:27	00 04:00:00	W	UC	m
122	85 202	16:08:00	00 11:13:33	W	UC	m

S: seizure. ID: patient identifier. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

Continued on next page

S	ID	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
123	85 202	01:51:40	00 09:13:39	W	UC	m
124	93 402	22:17:50	00 04:00:00	N2	FBTC	t
125	93 402	10:21:34	00 04:00:00	N2	FOIA	t
126	93 402	23:20:24	00 04:00:00	N2	FOIA	t
127	93 402	00:59:09	02 01:08:44	N2	UC	t
128	93 402	06:26:26	00 04:57:17	N2	UC	t
129	93 902	08:39:52	00 04:00:00	W	FOA	t
130	93 902	16:02:21	00 04:00:00	W	FOIA	t
131	93 902	02:31:07	00 04:00:00	N2	FBTC	d
132	93 902	18:48:40	00 04:00:00	W	FOIA	d
133	93 902	04:02:38	00 08:43:58	N2	FOIA	d
134	93 902	09:21:33	00 04:48:55	W	UC	d
135	94 402	15:29:22	00 04:00:00	W	FOA	?
136	94 402	11:02:56	00 04:00:00	W	UC	d
137	94 402	18:05:40	00 04:00:00	W	FOIA	b
138	94 402	01:36:02	00 04:00:00	N2	UC	t
139	94 402	16:10:53	00 08:39:10	W	FOA	?
140	94 402	02:48:18	00 10:07:25	N2	UC	b
141	94 402	08:16:30	00 04:58:11	W	FOA	?
142	95 202	01:28:09	00 04:00:00	N2	FBTC	b
143	95 202	15:00:18	00 04:00:00	N2	FOIA	b
144	95 202	01:35:24	00 04:00:00	N2	FOIA	b
145	95 202	14:13:22	00 04:00:00	N2	FOIA	m
146	95 202	23:30:29	00 08:47:06	N2	UC	b
147	95 202	23:55:21	00 23:54:52	N2	FOIA	b
148	95 202	00:04:20	01 21:34:10	N2	UC	t
149	96 002	17:10:35	00 04:00:00	W	FOIA	t
150	96 002	10:26:53	00 04:00:00	W	FOIA	t
151	96 002	17:46:44	00 04:00:00	W	FOIA	t
152	96 002	00:05:44	00 04:00:00	W	FOIA	d
153	96 002	00:44:10	00 23:56:34	W	UC	a

S: seizure. ID: patient identifier. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

Continued on next page

S	ID	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
154	96 002	18:57:18	00 17:43:07	W	FOIA	t
155	96 002	06:20:01	01 10:52:43	W	FOIA	a
156	98 102	07:17:49	00 04:00:00	W	FOA	?
157	98 102	18:49:53	00 04:00:00	W	UC	?
158	98 102	05:18:58	00 04:00:00	W	UC	?
159	98 102	06:11:33	01 00:22:35	W	UC	?
160	98 102	04:07:04	00 21:25:31	W	FBTC	?
161	98 202	04:50:27	00 04:00:00	W	FOIA	t
162	98 202	20:38:46	00 04:00:00	W	FOIA	a
163	98 202	07:16:40	00 04:00:00	W	FOIA	t
164	98 202	12:16:11	00 04:00:00	W	FBTC	t
165	98 202	03:37:11	00 04:00:00	W	FOIA	t
166	98 202	01:22:11	00 21:15:00	W	FOIA	t
167	98 202	07:55:06	00 06:02:55	W	FOIA	t
168	98 202	16:57:19	00 04:23:56	W	UC	t
169	101 702	07:35:40	00 04:00:00	W	FOIA	t
170	101 702	12:29:53	00 04:00:00	W	FOIA	t
171	101 702	19:33:06	00 04:00:00	W	FOIA	t
172	101 702	07:35:22	00 11:32:16	N2	FOIA	r
173	101 702	20:26:01	00 12:20:38	W	FOIA	r
174	102 202	22:50:21	00 04:00:00	N2	FOA	b
175	102 202	15:36:30	00 04:00:00	W	UC	?
176	102 202	05:47:03	00 04:00:00	N2	FOIA	t
177	102 202	22:14:59	00 04:00:00	W	UC	?
178	102 202	14:07:10	00 15:22:10	W	FOA	t
179	102 202	06:16:20	00 09:47:18	N2	FOIA	t
180	102 202	15:54:20	00 17:23:19	W	UC	t
181	104 602	15:35:45	00 04:00:00	W	FOIA	t
182	104 602	23:46:07	00 04:00:00	N2	FBTC	a
183	104 602	06:24:56	00 04:00:00	N2	FBTC	t
184	104 602	12:30:01	00 05:35:04	N2	FBTC	t

S: seizure. ID: patient identifier. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

Continued on next page

S	ID	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
185	104 602	22:44:07	00 09:44:06	N2	UC	d
186	109 502	10:00:00	00 04:00:00	W	FOIA	t
187	109 502	19:42:33	00 04:00:00	W	FOIA	t
188	109 502	07:56:09	00 05:08:49	W	UC	t
189	109 502	10:17:37	01 18:00:34	W	UC	t
190	110 602	10:20:41	00 04:00:00	W	FOIA	t
191	110 602	17:39:56	00 04:00:00	W	FOIA	t
192	110 602	08:30:09	00 04:00:00	W	FOIA	t
193	110 602	21:34:00	00 12:33:50	W	FOIA	t
194	110 602	11:28:35	00 13:24:35	W	FOA	t
195	112 802	17:05:49	00 04:00:00	W	UC	t
196	112 802	07:49:43	00 04:00:00	W	FOIA	t
197	112 802	15:36:04	00 04:00:00	W	UC	t
198	112 802	06:52:41	00 04:00:00	W	FOIA	t
199	112 802	11:54:45	00 04:32:04	W	FOIA	t
200	112 802	08:39:39	01 20:14:54	W	UC	t
201	113 902	23:32:27	00 04:00:00	W	UC	t
202	113 902	16:55:50	00 04:00:00	W	FOIA	d
203	113 902	05:17:05	00 04:00:00	N2	FOIA	t
204	113 902	13:46:12	00 04:00:00	W	FOIA	t
205	113 902	22:40:46	00 08:24:33	N2	UC	t
206	113 902	16:53:42	00 06:23:24	W	FOIA	t
207	114 702	20:52:30	00 04:00:00	W	FOIA	t
208	114 702	14:45:03	00 04:00:00	W	FOIA	t
209	114 702	04:09:15	00 04:00:00	W	UC	t
210	114 702	09:50:10	00 04:00:00	W	FOIA	t
211	114 702	14:27:45	00 04:00:00	W	FOIA	d
212	114 702	11:03:08	00 08:11:30	W	FOIA	t
213	114 702	13:27:36	00 09:30:51	W	FOIA	d
214	114 702	21:04:57	00 07:07:20	W	FOIA	t
215	114 902	08:30:29	00 04:00:00	W	FOA	s

S: seizure. ID: patient identifier. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

Continued on next page

S	ID	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
216	114 902	14:42:32	00 04:00:00	W	FOIA	b
217	114 902	19:42:40	00 04:00:00	W	FOIA	s
218	114 902	05:59:33	00 04:00:00	N2	FBTC	t
219	114 902	17:18:54	00 07:19:37	W	UC	r
220	114 902	11:52:26	00 18:03:32	W	FOIA	a
221	114 902	09:27:30	00 15:35:47	W	FOIA	t
222	123 902	02:52:47	00 04:00:00	N2	FBTC	t
223	123 902	01:38:19	00 04:00:00	N2	FBTC	t
224	123 902	02:11:22	00 04:00:00	R	FOIA	t
225	123 902	18:57:10	00 11:32:53	W	FOIA	t
226	123 902	15:22:45	00 18:39:43	W	FOA	t

S: seizure. ID: patient identifier. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

B.2 Grid search results

Table B.3 and Table B.4 contain the average geometric mean between sample sensitivity and sample specificity (G-Mean) obtained while searching for the best hyperparameters for the deep neural network (DNN) and the shallow artificial neural network (ANN), respectively. The selected hyperparameters are in bold.

B.2.1 Grid search results for deep neural network

Table B.3: Results of the grid search of the optimal hyperparameters for the deep classifier.

Number of Filters (First layer)	Filter Size	LSTM Units	Average G-Mean
32	3	32	0.291
		64	0.292
		128	0.278
	5	32	0.292
		64	0.312
		128	0.278
	7	32	0.299
		64	0.291
		128	0.275
64	3	32	0.349
		64	0.306
		128	0.300
	5	32	0.318
		64	0.315
		128	0.336
	7	32	0.295
		64	0.308
		128	0.292
128	**3**	32	0.354
		64	**0.378**
		128	0.362
	5	32	0.349
		64	0.346
		128	0.353
	7	32	0.346
		64	0.320
		128	0.316

B.2.2 Grid search results for shallow artificial neural network

Table B.4: Results of the grid search of the optimal hyperparameters for the shallow classifier.

Number of neurons in fully connected layer (FC layer)	Average G-Mean
None	**0.331**
8	0.326
16	0.320
32	0.312
64	0.311
128	0.301
256	0.307

B.3 Results obtained for all approaches

Table B.1 presents the learning curves obtained for one deep learning model applied to one patient using the chronological approach. As seen in the figure, the validation loss curves improve as the training sets become larger, indicating that the model is able to achieve better generalisation. Table B.5 and Table B.6 contain seizure sensitivitys (SSs) and false prediction rate per hour (FPR/h) values for the approaches developed using denoised EEG data and noisy EEG data, respectively. Table B.7 contains results obtained by Pinto *et al.* [142, 143].

B.3.1 Learning curves of the chronological approach

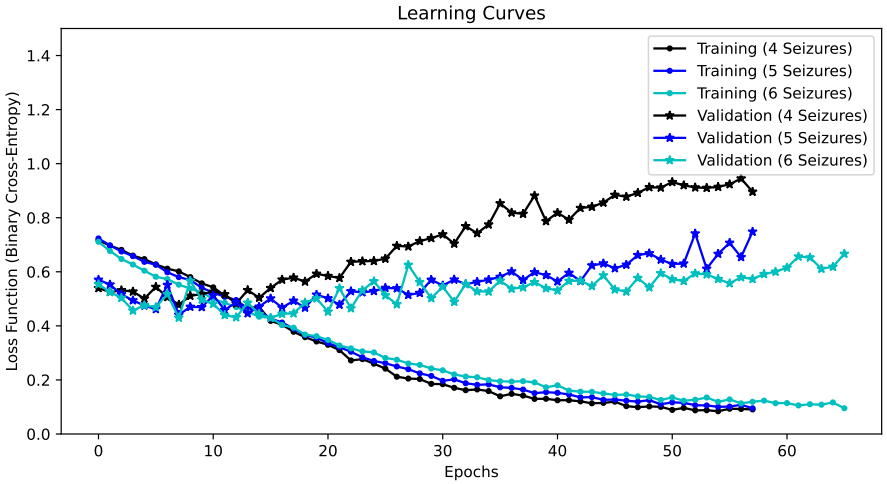


Figure B.1: Example of learning curves obtained for one model of one patient with 3 test seizures considering the chronological approach. The black lines represent the training and validation curves using 4 training seizures. The blue lines represent the training and validation curves using 5 training seizures. The cyan lines represent the training and validation curves using 6 training seizures.

B.3.2 Results obtained for approaches using denoised data

Table B.5: Results for each patient using denoised data.

	Denoised EEG$_{Standard}$			**Denoised EEG$_{Chronological}$**			**Denoised Features$_{Standard}$**			**Denoised Features$_{Chronological}$**		
Patient	**SS**	**FPR/h**	**ACL**	**SS**	**FPR/h**	**ACL**	**SS**	**FPR/h**	**ACL**	**SS**	**FPR/h**	**ACL**
402	0.00	1.316	0	0.00	0.331	0	0.00	1.474	0	0.00	0.541	0
8902	0.50	0.105	1	0.50	0.161	1	1.00	0.163	1	0.50	0.220	1
11002	0.50	0.062	1	0.50	0.130	1	0.50	0.370	1	0.50	0.370	1
16202	0.33	0.401	1	0.33	0.496	1	0.00	2.484	0	0.00	2.486	0
21902	0.50	0.068	1	0.50	0.068	1	0.50	0.880	1	0.50	1.230	1
23902	0.00	0.482	0	0.00	0.483	0	0.50	1.321	1	0.50	1.332	0
26102	0.50	1.003	1	0.50	0.828	1	0.50	0.882	1	0.50	0.988	1
30802	0.33	0.137	1	0.33	0.138	1	0.67	0.373	1	0.67	0.373	1
32702	0.00	0.000	0	0.00	0.123	0	0.00	0.192	0	0.00	0.192	0
45402	0.50	1.598	0	0.50	0.503	1	1.00	3.176	1	1.00	2.194	1
46702	0.00	1.145	0	0.00	0.342	0	0.00	3.289	0	0.00	2.588	0
50802	0.00	0.134	0	0.50	0.657	1	0.00	0.134	0	0.00	0.211	0
52302	0.00	0.000	0	0.00	0.000	0	0.00	0.000	0	0.00	0.000	0
53402	0.00	0.290	0	0.00	0.232	0	0.00	0.384	0	0.00	0.415	0
55202	0.33	0.202	1	0.33	0.344	1	0.33	1.277	0	0.00	1.701	0
56402	0.00	0.055	0	0.00	0.213	0	0.50	2.812	0	0.50	3.411	0
58602	0.00	0.121	0	0.00	0.121	0	0.00	0.567	0	0.00	0.567	0
59102	1.00	2.227	1	0.50	0.207	1	1.00	3.048	1	1.00	0.997	1
60002	0.00	0.151	0	0.00	0.124	0	0.00	2.556	0	0.50	1.614	1
64702	0.00	0.000	0	0.00	0.000	0	0.00	0.334	0	0.00	0.334	0
75202	0.00	0.417	0	0.00	0.305	0	0.00	0.207	0	0.00	0.304	0
80702	0.00	0.059	0	0.00	0.124	0	1.00	2.109	1	1.00	1.846	1
81102	0.00	0.207	0	0.00	0.207	0	1.00	0.648	1	1.00	0.648	1
85202	0.00	0.000	0	0.00	0.055	0	0.50	0.312	1	1.00	0.316	1
93402	0.00	0.179	0	0.50	0.179	1	0.50	0.256	1	1.00	0.156	1
93902	0.50	0.190	1	0.50	0.734	1	0.50	0.576	1	0.50	0.945	1
94402	0.00	0.372	0	0.00	0.372	0	0.00	1.024	0	0.00	1.069	0
95202	0.33	0.387	1	0.33	0.549	1	0.67	0.524	1	0.67	0.604	1
96002	0.00	0.051	0	0.33	0.204	1	0.67	0.586	1	0.67	0.651	1
98102	0.00	0.000	0	0.00	0.000	0	0.50	0.825	1	0.50	0.822	1
98202	0.00	0.159	0	0.00	0.159	0	0.00	0.732	0	0.00	0.732	0
101702	0.00	0.000	0	0.00	0.000	0	0.50	0.212	1	0.50	0.154	1
102202	0.33	0.311	1	0.33	0.205	1	0.00	1.071	0	0.00	0.996	0
104602	0.00	0.000	0	0.00	0.000	0	0.00	0.079	0	0.00	0.499	0
109502	0.00	0.000	0	0.00	0.213	0	0.00	0.000	0	0.50	1.060	1
110602	0.00	0.000	0	0.00	0.000	0	0.50	0.570	1	0.00	0.431	0
112802	0.50	0.434	1	0.50	0.729	1	0.00	0.868	0	0.50	1.068	1
113902	0.00	0.271	0	0.00	0.172	0	0.50	0.083	1	0.50	0.083	1
114702	0.00	0.047	1	0.33	0.000	1	0.33	0.337	1	0.33	0.746	1
114902	0.00	0.000	0	0.00	0.000	0	0.33	0.084	1	0.33	0.241	1
123902	0.00	0.000	0	0.00	0.000	0	0.00	0.000	0	0.00	0.000	0
Total	**0.150±0.244**	**0.307±0.482**	**12 (0.29)**	**0.179±0.222**	**0.237±0.225**	**17 (0.41)**	**0.341±0.350**	**0.898±0.955**	**21 (0.51)**	**0.370±0.360**	**0.857±0.770**	**22 (0.54)**

B.3.3 Results obtained for approaches using noisy data

Table B.6: Results for each patient using noisy data.

	Noisy EEG$_{Standard}$			Noisy EEG$_{Chronological}$			Noisy Features$_{Standard}$			Noisy Features$_{Chronological}$		
Patient	**SS**	**FPR/h**	**ACL**	**SS**	**FPR/h**	**ACL**	**SS**	**FPR/h**	**ACLl**	**SS**	**FPR/h**	**ACL**
402	0.00	1.655	0	0.00	0.728	0	0.00	1.461	0	0.00	0.709	0
8902	0.50	0.105	1	0.50	0.163	1	1.00	0.164	1	1.00	0.164	1
11002	0.00	0.130	0	0.50	0.130	1	0.00	0.358	0	0.00	0.272	0
16202	0.00	0.311	0	0.00	0.405	0	0.00	1.455	0	0.00	1.553	0
21902	0.00	0.403	0	0.00	0.403	0	0.50	0.866	1	0.50	1.209	1
23902	0.00	0.280	0	0.00	0.381	0	0.00	1.698	0	0.00	1.453	0
26102	0.50	0.939	1	0.50	1.000	1	0.50	0.823	1	0.50	1.200	1
30802	0.33	0.137	1	0.33	0.108	1	0.67	0.335	1	0.67	0.299	1
32702	0.00	0.059	0	0.00	0.123	0	0.00	0.192	0	0.00	0.192	0
45402	0.50	2.459	0	0.50	0.508	1	1.00	2.889	1	1.00	2.037	1
46702	0.00	1.453	0	0.00	0.842	0	0.00	2.725	0	0.00	2.241	0
50802	0.00	0.064	0	0.50	0.484	1	0.00	0.134	0	0.00	0.171	0
52302	0.00	0.000	0	0.00	0.000	0	0.00	0.000	0	0.00	0.000	0
53402	0.00	0.229	0	0.00	0.202	0	0.00	0.316	0	0.00	0.319	0
55202	0.33	0.202	1	0.33	0.141	1	1.00	0.976	1	0.67	1.253	1
56402	0.00	0.085	0	0.00	0.249	0	0.00	1.954	0	0.00	2.188	0
58602	0.00	0.121	0	0.00	0.121	0	0.50	0.261	1	0.50	0.619	1
59102	1.00	2.137	1	0.50	0.224	1	1.00	2.719	1	0.50	1.088	0
60002	0.00	0.235	0	0.50	0.151	1	0.50	1.940	0	0.50	1.081	1
64702	0.00	0.000	0	0.00	0.000	0	0.00	0.195	0	0.00	0.195	0
75202	0.00	0.307	0	0.00	0.304	0	0.33	5.549	0	0.00	1.633	0
80702	0.00	0.619	0	0.00	0.422	0	1.00	2.129	1	1.00	1.824	1
81102	0.00	0.268	0	0.00	0.268	0	1.00	0.643	1	1.00	0.643	1
85202	0.00	0.000	0	0.00	0.000	0	0.50	0.315	1	0.50	0.311	1
93402	0.00	0.132	0	0.50	0.156	1	1.00	0.540	1	1.00	0.439	1
93902	0.00	0.183	0	0.50	0.916	1	0.50	0.578	1	0.50	0.757	1
94402	0.00	0.298	0	0.00	0.452	0	0.00	0.709	0	0.00	1.209	0
95202	0.67	0.579	1	0.67	0.526	1	0.67	0.524	1	0.67	0.604	1
96002	0.00	0.000	0	0.00	0.000	0	0.67	0.583	1	0.67	0.782	1
98102	0.00	0.075	0	0.00	0.024	0	0.50	0.670	1	0.50	0.816	1
98202	0.00	0.000	0	0.00	0.000	0	0.00	0.116	0	0.00	0.116	0
101702	0.00	0.000	0	0.00	0.000	0	0.00	0.099	0	0.00	0.154	0
102202	0.33	0.239	1	0.00	0.141	0	0.00	1.070	0	0.33	1.059	1
104602	0.00	0.000	0	0.00	0.169	0	0.00	0.000	0	0.00	0.382	0
109502	0.00	0.000	0	0.00	0.154	0	0.00	0.000	0	0.50	0.997	1
110602	0.00	0.000	0	0.00	0.000	0	0.50	0.570	1	0.00	0.365	0
112802	0.50	0.509	0	0.00	0.148	0	0.00	1.242	0	0.00	1.984	0
113902	0.50	0.082	1	0.50	0.082	1	0.50	0.083	1	0.50	0.083	1
114702	0.00	0.047	0	0.33	0.151	1	0.67	0.828	1	0.33	1.177	1
114902	0.00	0.000	0	0.00	0.000	0	0.33	0.212	1	0.33	0.396	1
123902	0.00	0.000	0	0.00	0.000	0	0.00	0.000	0	0.00	0.000	0
Total	**0.126±0.244**	**0.350±0.575**	**8 (0.20)**	**0.163±0.234**	**0.251±0.261**	**14 (0.34)**	**0.362±0.384**	**0.925±1.090**	**20 (0.49)**	**0.333±0.356**	**0.829±0.647**	**21 (0.51)**

B.3.4 Results obtained in other studies using EEG data from patients from EPILEPSIAE database

Table B.7: Results for each patient obtained in studies of Pinto *et al.* [142, 143].

	Pinto *et al.* (2021) [142]			**Pinto *et al.* (2022) [143]**		
Patient	**SS**	**FPR/h**	**ACL**	**SS**	**FPR/h**	**ACL**
402	-	-	-	-	-	-
8902	-	-	-	0.20	0.16	0
11002	0.27	1.57	0	0.00	0.27	0
16202	0.30	0.49	0	-	-	-
21902	-	-	-	0.00	0.18	0
23902	-	-	-	0.08	0.22	0
26102	-	-	-	0.13	0.34	0
30802	0.58	3.77	1	0.19	0.19	0
32702	-	-	-	-	-	-
45402	-	-	-	-	-	-
46702	-	-	-	0.25	0.28	1
50802	-	-	-	0.03	0.22	0
52302	-	-	-	0.01	0.24	0
53402	0.63	0.50	1	0.30	0.26	1
55202	0.70	1.21	1	0.18	0.23	0
56402	-	-	-	0.05	0.14	0
58602	0.38	2.14	0	0.05	0.25	0
59102	-	-	-	-	-	-
60002	0.48	1.29	0	-	-	-
64702	0.02	0.92	0	0.08	0.14	0
75202	0.70	0.69	1	0.19	0.14	0
80702	0.31	0.65	0	0.10	0.14	0
81102	-	-	-	-	-	-
85202	0.47	0.36	1	0.42	0.25	1
93402	-	-	-	0.11	0.32	0
93902	-	-	-	0.37	0.23	1
94402	0.38	1.05	0	0.13	0.36	0
95202	0.14	0.66	0	0.09	0.16	0
96002	0.23	0.77	0	0.16	0.22	0
98102	-	-	-	0.32	0.11	1
98202	0.19	1.72	0	-	-	-
101702	0.23	0.44	0	0.34	0.24	1
102202	-	-	-	0.22	0.18	1
104602	-	-	-	0.33	0.26	1
109502	0.52	1.29	0	0.11	0.14	0
110602	0.47	0.33	1	0.37	0.20	1
112802	-	-	-	-	-	-
113902	-	-	-	0.28	0.07	1
114702	-	-	-	0.16	0.35	1
114902	0.31	0.25	1	0.33	0.10	1
123902	-	-	-	0.07	0.14	0
Total	**0.38±0.19**	**1.06±0.84**	**8 (0.37)**	**0.18±0.12**	**0.21±0.07**	**12 (0.38)**

B.3.5 Results obtained using other machine learning models

To ensure the robustness of the study, the effectiveness of denoised data and chronological training were evaluated on other previously proposed seizure prediction mod-

els. Specifically, a deep learning model proposed by Truong *et al.* [280] and a shallow classifier model with handcrafted features proposed by Karoly *et al.* [76] were selected. The deep learning model comprises a deep convolutional neural network (DCNN) that takes spectrograms obtained from EEG windows using short-term Fourier transform (STFT). In the case of the shallow model, some features were firstly extracted by the 19 channels, such as the energy from several frequency bands (8-16 Hz, 16-32 Hz, 32-64 Hz, 64-90 Hz) and the line length. The Kullback-Leibler distance method is then used to select the best features, which are subsequently used to develop the logistic regression.

Table B.8: Average results of the additional seizure prediction models, for all 41 patients. The models that used EEG time series were developed using a convolutional neural network with spectrograms as input as performed in Truong *et al.* [280]. The models that used features were developed using a logistic regression with the same features extracted by Karoly *et al.* [76].

Approach	SS	FPR/h	ACL (%)
Denoised EEG$_{Standard}$ - CNN (Spectrograms)	0.13±0.27	0.18±0.38	9 (0.22)
Denoised EEG$_{Chronological}$ - CNN (Spectrograms)	0.14±0.27	0.11±0.16	11 (0.27)
Denoised Features$_{Standard}$ - Logistic regression (Features)	0.44±0.37	1.91±0.96	19 (0.46)
Denoised Features$_{Chronological}$ - Logistic regression (Features)	0.42±0.36	1.97±2.51	19 (0.54)
Noisy EEG$_{Standard}$ - CNN (Spectrograms)	0.10±0.22	0.18±0.46	8 (0.20)
Noisy EEG$_{Chronological}$ - CNN (Spectrograms)	0.11±0.23	0.09±0.15	9 (0.22)
Noisy Features$_{Standard}$ - Logistic regression (Features)	0.44±0.40	2.58±3.27	16 (0.39)
Noisy Features$_{Chronological}$ - Logistic regression (Features)	0.38±0.34	2.32±2.84	13 (0.32)

Appendix C

Transfer learning on seizure prediction: Does information from several patients improve patient-specific approaches?

C.1 Patient and seizure metadata of the Personal dataset

Table C.1 contains information regarding the group of patients from the Epilepsy Center of the Universitätsklinikum Freiburg. The table includes information on sex, age at hospital admission and onset age (corresponding to the occurrence of the first epilepsy event), epilepsy foci lateralisation, the total number of annotated seizures and the number of lead seizures analysed for each patient, according to the considered minimum inter-seizure interval of 4.5 hours. The duration of the used electroencephalogram (EEG) data regarding training seizures and testing seizures is also presented. It is worth noting that although it was considered lead seizures to happen at least 4.5 hours after the previous one, only 4 hours were used for training.

Table C.2 describes data collected for each analysed seizure. It includes information about the EEG onset time, used inter-seizure time, vigilance state at the time of the seizure onset, seizure classification, and seizure activity pattern. Seizures were classified according to the International League Against Epilepsy (ILAE) classification [47]. The vigilance state corresponds to one of the following states of alertness and responsiveness: wakefulness, non-rapid eye movement (NREM) sleep, further subdivided into three sleep stages (N1–3) and rapid eye movement (REM) sleep [358].

Table C.1: Dataset description regarding each patient.

P	Sex	Onset Age (years)	Admission Age (years)	Lat.	#Sz	#LSz	Training time (dd hh/mm/ss)	Testing time (dd hh/mm/ss)
1	M	20	23	L	5	3	00 08:00:00	00 18:54:19
2	M	18	18	L	16	9	01 00:00:00	01 11:32:56
3	M	28	50	?	4	4	00 08:00:00	00 22:53:37
4	F	49	60	?	14	5	00 12:00:00	00 16:33:05
5	M	18	33	L	4	3	00 08:00:00	05 08:11:32
6	F	32	33	R	11	4	00 08:00:00	00 20:28:49
7	F	8	34	R	3	3	00 08:00:00	01 21:13:14
8	F	28	45	?	11	4	00 08:00:00	00 10:24:14
9	M	2	40	?	8	4	00 08:00:00	00 19:58:12
10	M	11	15	L	8	6	00 16:00:00	02 02:20:09
11	F	14	47	?	28	18	02 04:00:00	02 08:06:51
12	F	?	27	?	5	5	00 12:00:00	03 14:03:25
13	F	8	23	L	5	5	00 12:00:00	00 23:11:17
14	F	33	38	?	4	3	00 08:00:00	00 05:58:35
15	F	13	30	R	10	8	00 20:00:00	01 22:51:20
16	M	34	36	L	12	9	00 20:00:00	01 22:33:18
17	F	32	52	?	21	12	01 08:00:00	06 00:12:31
18	M	52	53	?	34	10	01 00:00:00	02 05:53:26
19	M	0	19	L	8	5	00 12:00:00	02 00:55:29
20	M	18	45	R	5	4	00 08:00:00	04 22:44:21
21	F	51	56	?	31	9	01 00:00:00	01 20:41:28
22	F	16	21	?	18	11	01 04:00:00	01 12:56:59
23	F	63	67	R	3	3	00 08:00:00	00 09:04:24
24	M	7	43	?	5	4	00 08:00:00	01 10:07:17

P: patient number. Sex: female (F) or male (M). Lateralisation (Lat.): L: left, R: right, B: bilateral. #Sz: total number of seizures annotated per patient. #LSz: number of leading seizures, obtained as a result of the analysis of 4.5 hours of interseizure data.

Table C.2: Dataset description regarding data preceding each seizure. The gray rows were used for training, while the other ones were used for testing.

S	P	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
1	1	16:27:06	00 04:00:00	W	FOIA	d
2	1	20:11:43	00 04:00:00	W	FOIA	d
3	1	15:37:14	00 18:54:19	W	FOIA	d
4	2	04:18:55	00 04:00:00	N2	FOIA	t
5	2	16:01:08	00 04:00:00	N2	UC	t
6	2	03:16:37	00 04:00:00	N2	UC	?
7	2	03:26:20	00 04:00:00	R	UC	d
8	2	12:56:07	00 04:00:00	W	FOIA	d
9	2	05:17:35	00 13:06:24	N2	UC	t
10	2	16:00:54	00 08:46:33	W	FOA	t
11	2	22:37:11	00 05:50:20	W	FOA	d
12	2	11:51:47	00 07:49:38	N2	FOIA	d
13	3	19:03:27	00 04:00:00	W	FOIA	d
14	3	05:18:49	00 04:00:00	W	FOIA	d
15	3	12:09:40	00 06:20:44	W	UC	t
16	3	09:01:32	00 16:32:53	W	FOIA	t
17	4	02:35:22	00 04:00:00	W	S	?
18	4	01:13:00	00 04:00:00	N2	UC	a
19	4	00:58:03	00 04:00:00	N2	UC	t
20	4	15:58:26	00 08:01:18	W	FBTC	t
21	4	01:03:09	00 08:31:47	N3	UC	a
22	5	14:21:48	00 04:00:00	N2	FOIA	d
23	5	23:42:55	00 04:00:00	N2	FOIA	d
24	5	23:57:27	05 08:11:32	N2	FBTC	t
26	6	13:00:28	00 04:00:00	W	FOA	?
27	6	02:43:01	00 04:00:00	N2	FOA	t
28	6	22:23:36	00 14:37:21	N2	S	b
29	6	07:07:34	00 05:51:27	N2	FOA	s
30	7	05:25:10	00 04:00:00	W	FOIA	?

S: seizure. P: patient number. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), subclinic (S), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

Continued on next page

S	P	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
31	7	03:30:28	00 04:00:00	N2	FOA	?
32	7	01:14:27	01 21:13:14	N2	FOIA	t
33	8	23:07:36	00 04:00:00	W	FOIA	d
34	8	14:14:36	00 04:00:00	W	FOIA	s
35	8	00:16:18	00 04:50:05	N2	FOIA	t
36	8	06:21:47	00 05:34:09	N2	FOIA	t
37	9	20:24:40	00 04:00:00	W	UC	t
38	9	06:11:12	00 04:00:00	W	UC	t
39	9	17:56:26	00 11:15:01	W	FOA	a
40	9	09:35:07	00 08:43:11	W	UC	t
41	10	06:30:28	00 04:00:00	W	UC	d
42	10	17:18:03	00 04:00:00	W	FOIA	d
43	10	08:29:33	00 04:00:00	?	FOIA	?
44	10	09:04:20	00 04:00:00	W	FOIA	t
45	10	07:31:58	00 21:56:58	W	FOIA	t
46	10	13:14:02	01 04:23:11	W	FOIA	t
47	11	20:13:59	00 04:00:00	W	UC	?
48	11	09:35:38	00 04:00:00	W	FOA	d
49	11	12:13:51	00 04:00:00	W	FOA	a
50	11	22:19:22	00 04:00:00	W	FOA	?
51	11	14:23:47	00 04:00:00	W	UC	t
52	11	19:32:42	00 04:00:00	W	FOA	b
53	11	06:20:43	00 04:00:00	W	UC	d
54	11	07:09:52	00 04:00:00	W	FOA	t
55	11	13:29:50	00 04:00:00	W	FOA	a
56	11	13:32:36	00 04:00:00	W	FOA	a
57	11	01:31:32	00 04:00:00	N2	UC	d
58	11	14:34:47	00 12:32:36	W	FOIA	t
59	11	02:26:20	00 07:33:07	N1	UC	t
60	11	07:47:03	00 04:49:55	W	FOIA	t
61	11	15:06:36	00 06:48:42	N2	UC	t

S: seizure. P: patient number. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), subclinic (S), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

Continued on next page

S	P	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
62	11	23:19:23	00 07:42:29	N2	S	?
63	11	16:07:49	00 10:49:35	W	FOA	d
64	11	22:28:40	00 05:50:27	N1	S	?
65	12	06:25:18	00 04:00:00	N2	UC	t
66	12	22:03:34	00 04:00:00	W	FOIA	d
67	12	16:50:07	00 04:00:00	W	FOIA	?
68	12	08:15:43	01 14:54:50	W	FOIA	t
69	12	07:54:59	01 23:08:35	W	FOIA	d
70	13	23:38:52	00 04:00:00	W	FOA	t
71	13	13:41:35	00 04:00:00	W	FOIA	t
72	13	12:35:42	00 04:00:00	W	FOIA	b
73	13	00:55:47	00 10:34:25	W	FOIA	d
74	13	14:04:56	00 12:36:52	W	FOIA	t
75	14	04:56:17	00 04:00:00	N2	UC	b
76	14	20:43:22	00 04:00:00	N2	UC	t
77	14	03:13:35	00 05:58:35	N2	UC	t
78	15	17:59:31	00 04:00:00	W	FOIA	?
79	15	03:49:13	00 04:00:00	N1	FOIA	t
80	15	19:16:10	00 04:00:00	W	FOIA	d
81	15	11:11:02	00 04:00:00	W	FOIA	?
82	15	22:50:41	00 04:00:00	W	FOIA	t
83	15	14:57:39	00 15:35:58	W	FOIA	d
84	15	15:59:11	01 00:30:15	W	FOIA	d
85	15	23:16:26	00 06:45:07	N1	FOIA	t
86	16	17:46:09	00 04:00:00	W	FOIA	d
87	16	17:16:00	00 04:00:00	W	FOA	d
88	16	08:51:00	00 04:00:00	W	FOA	d
89	16	21:58:23	00 04:00:00	W	FOA	t
90	16	04:48:58	00 04:00:00	W	FOA	d
91	16	16:33:44	00 07:31:49	W	FOA	t
92	16	21:46:09	00 04:43:10	?	S	?

S: seizure. P: patient number. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), subclinic (S), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

Continued on next page

S	P	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
93	16	09:00:25	00 07:30:08	W	UC	d
94	16	12:19:33	01 02:48:11	W	FOA	t
95	17	14:24:25	00 04:00:00	N2	FOIA	t
96	17	14:14:09	00 04:00:00	W	FOIA	t
97	17	02:21:13	00 04:00:00	N2	FBTC	?
98	17	08:39:21	00 04:00:00	W	FOIA	t
99	17	01:07:06	00 04:00:00	N2	FOIA	t
100	17	22:27:52	00 04:00:00	N2	FOIA	?
101	17	06:25:08	00 04:00:00	N2	FOIA	?
102	17	22:22:27	00 08:03:58	N2	FOIA	t
103	17	15:20:41	00 16:26:42	?	S	?
104	17	23:51:11	01 00:47:07	N2	FOIA	t
105	17	05:50:28	02 05:28:28	N2	UC	t
106	17	23:47:19	01 17:26:13	N2	UC	t
107	18	20:13:59	00 04:00:00	W	FOA	?
108	18	14:20:06	00 04:00:00	W	FOIA	t
109	18	09:07:46	00 04:00:00	W	FOIA	t
110	18	15:42:41	00 04:00:00	W	FOA	?
111	18	21:56:42	00 04:00:00	W	FOIA	t
112	18	14:16:54	00 04:00:00	W	FOIA	t
113	18	00:35:07	00 09:47:10	W	FOIA	t
114	18	05:10:22	00 04:03:56	W	FOA	t
115	18	01:01:43	00 19:20:11	W	FOIA	t
116	18	01:25:11	00 20:42:09	W	FOA	t
117	19	06:28:36	00 04:00:00	N1	UC	d
118	19	10:43:12	00 04:00:00	N1	UC	a
119	19	07:08:48	00 04:00:00	N1	UC	t
120	19	14:46:06	01 07:06:22	N1	UC	a
121	19	09:06:18	00 17:49:07	W	FOIA	t
122	20	17:02:48	00 04:00:00	W	FOIA	t
123	20	10:16:57	00 04:00:00	W	FOA	?

S: seizure. P: patient number. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), subclinic (S), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

Continued on next page

S	P	EEG onset time	Used interseizure time (dd hh/mm/ss)	Vigilance state	ILAE Classification	Activity pattern
124	20	04:02:18	03 16:33:10	N2	UC	t
125	20	10:44:14	01 06:11:11	W	FOIA	s
126	21	23:57:34	00 04:00:00	N2	S	b
127	21	15:19:40	00 04:00:00	W	FOA	?
128	21	16:41:35	00 04:00:00	W	FOA	b
129	21	13:56:53	00 04:00:00	W	FOA	?
130	21	05:47:22	00 04:00:00	W	FOA	t
131	21	17:49:40	00 11:31:21	W	FOA	t
132	21	12:28:32	00 18:06:58	W	FOA	t
133	21	04:45:53	00 10:58:43	N2	FOA	t
134	21	19:25:29	00 04:04:26	W	FOA	t
135	22	07:38:07	00 04:00:00	W	UC	b
136	22	14:42:44	00 04:00:00	W	FOIA	t
137	22	06:36:06	00 04:00:00	W	UC	t
138	22	13:52:01	00 04:00:00	W	UC	a
139	22	21:03:39	00 04:00:00	W	UC	a
140	22	20:27:45	00 04:00:00	W	FOIA	d
141	22	23:23:55	00 04:00:00	W	FOIA	a
142	22	04:23:00	00 04:27:41	N2	FOIA	d
143	22	08:49:20	00 20:27:14	W	FOIA	d
144	22	14:28:35	00 05:08:02	W	UC	t
145	22	05:39:10	00 06:54:02	N2	UC	a
146	23	05:35:20	00 04:00:00	W	FOIA	d
147	23	11:55:09	00 04:00:00	W	FOA	t
148	23	21:30:45	00 09:04:25	W	UC	?
149	24	14:44:25	00 04:00:00	W	FOIA	?
150	24	23:24:30	00 04:00:00	W	FOIA	r
151	24	09:19:17	00 08:56:54	W	UC	b
152	24	11:00:25	01 01:10:23	W	FOIA	?

S: seizure. ID: patient number. Seizure vigilance state: wakefulness (W), NREM sleep stage I (N1), NREM sleep stage II (N2), REM sleep stage (R). Seizure ILAE classification: focal onset aware (FOA), focal onset impaired awareness (FOIA), focal to bilateral tonic-clonic (FBTC), subclinic (S), unclassified (UC). Seizure activity pattern: rhythmic alpha waves (a), rhythmic beta waves (b), cessation of interictal activity (c), rhythmic delta waves (d), amplitude depression (m), repetitive spiking (r), rhythmic sharp waves (s), rhythmic theta waves (t), unclear (?). Training seizures are highlighted in grey.

C.2 Deep neural network architectures

Table C.3 describes the deep convolutional autoencoder (DCAE) used in the study. It contains the names of the layers, the hyperparameters and the output shape. Table C.4 contains the deep neural network (DNN) used for developing the seizure prediction models. The rows present different colours with different meanings: the ones with no colour are trained from scratch, and the yellow ones are frozen during the training.

Table C.3: DCAE architecture.

Layer	Hyperparameters	Output shape
Input	-	2560x19
Convolution 1D	Filters = 128, Size = 3, Stride = 1, Pad = 'same'	2560x128
Convolution 1D	Filters = 128, Size = 3, Stride = 2, Pad = 'same'	1280x128
Spatial dropout	Rate = 20%	1280x128
Activation	Swish function	1280x128
Batch normalisation	-	1280x128
Convolution 1D	Filters = 256, Size = 3, Stride = 1, Pad = 'same'	1280x256
Convolution 1D	Filters = 256, Size = 3, Stride = 2, Pad = 'same'	640x256
Spatial dropout	Rate = 20%	640x256
Activation	Swish function	640x256
Batch normalisation	-	640x256
Convolution 1D	Filters = 512, Size = 3, Stride = 1, Pad = 'same'	640x512
Convolution 1D	Filters = 512, Size = 3, Stride = 2, Pad = 'same'	320x512
Spatial dropout	Rate = 20%	320x512
Activation	Swish function	320x512
Batch normalisation	-	320x512
Up sampling 1D	Size = 2	640x128
Convolution 1D	Filters = 256, Size = 3, Stride = 1, Pad = 'same'	640x256
Spatial dropout	Rate = 20%	640x256
Activation	Swish function	640x256
Batch normalisation	-	640x256
Up sampling 1D	Size = 2	1280x256
Convolution 1D	Filters = 128, Size = 3, Stride = 1, Pad = 'same'	1280x128
Spatial dropout	Rate = 20%	1280x128
Activation	Swish function	1280x128
Batch normalisation	-	1280x128
Up sampling 1D	Size = 2	2560x128
Convolution 1D	Filters = 19, Size = 3, Stride = 1, Pad = 'same'	2560x19

Table C.4: DNN used for seizure prediction.

Layer	Hyperparameters	Output shape
Input	-	2560x19
Convolution 1D	Filters = 128, Size = 3, Stride = 1, Pad = 'same'	2560x128
Convolution 1D	Filters = 128, Size = 3, Stride = 2, Pad = 'same'	1280x128
Spatial dropout	Rate = 20%	1280x128
Activation	Swish function	1280x128
Batch normalisation	-	1280x128
Convolution 1D	Filters = 256, Size = 3, Stride = 1, Pad = 'same'	1280x256
Convolution 1D	Filters = 256, Size = 3, Stride = 2, Pad = 'same'	640x256
Spatial dropout	Rate = 20%	640x256
Activation	Swish function	640x256
Batch normalisation	-	640x256
Convolution 1D	Filters = 512, Size = 3, Stride = 1, Pad = 'same'	640x512
Convolution 1D	Filters = 512, Size = 3, Stride = 2, Pad = 'same'	320x512
Spatial dropout	Rate = 20%	320x512
Activation	Swish function	320x512
Batch normalisation	-	320x512
Bidirectional LSTM	Units = 64, Return sequences = False	128x1
Dropout	Dropout rate = 20%	128x1
Fully connected	Neurons = 2	2x1
Activation	Softmax function	2x1

C.3 Results obtained for all approaches

Figure C.1 presents the learning curves obtained for two deep learning models trained with data from patient 15: one using the standard approach and another using the transfer learning approach. As seen in the figure, the model that begins from scratch can not correctly optimise, i.e., the difference between the training loss and the validation loss is always large, with the validation loss also very high. Table C.5 contains seizure sensitivitys (SSs), false prediction rate per hour (FPR/h) values, and the output of the surrogate analysis for every patient for the standard approach and the transfer learning approach.

C.3.1 Learning curves of the chronological approach

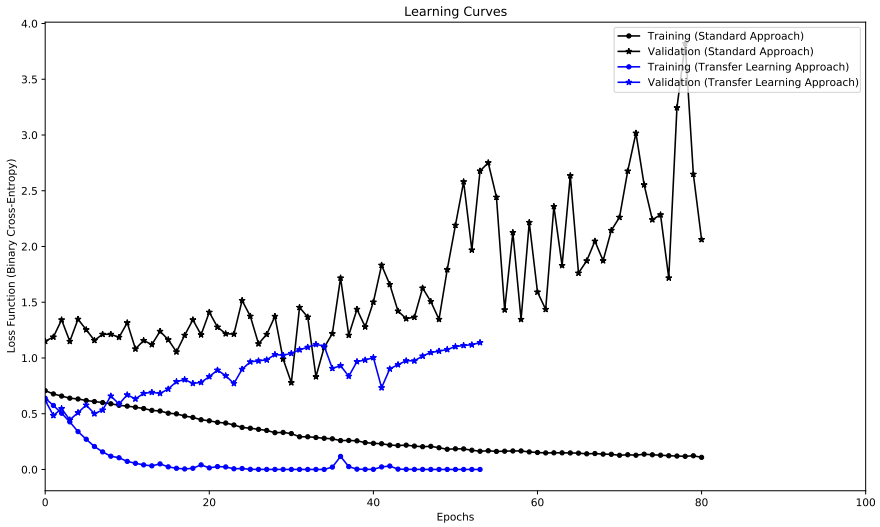


Figure C.1: Example of learning curves obtained for two models trained with data from patient 15: one using the standard approach and another following the transfer learning approach. The black lines represent the training and validation curves using the standard approach. The blue lines represent the training and validation curves using the transfer learning approach.

C.3.2 Results obtained for standard and transfer learning approaches

Table C.5: Results for each patient using standard and transfer learning approaches.

	Standard			**Transfer Learning**		
Patient	**SS**	**FPR/h**	**ACL**	**SS**	**FPR/h**	**ACL**
1	0.00	0.144	0	0.00	0.000	0
2	0.15	0.226	1	0.10	0.309	0
3	0.00	0.123	0	0.10	0.081	1
4	0.10	0.380	0	0.50	0.339	1
5	0.00	0.000	0	0.60	2.402	1
6	0.00	0.098	0	0.00	0.168	0
7	0.40	0.041	1	0.80	0.051	1
8	0.00	0.000	0	0.50	1.380	1
9	0.20	1.148	0	0.00	0.607	0
10	0.10	0.205	0	0.00	0.014	0
11	0.40	3.224	1	0.40	0.194	1
12	0.00	0.030	0	0.00	0.019	0
13	0.00	0.000	0	0.00	0.032	0
14	0.00	0.043	0	0.00	0.000	0
15	0.40	14.449	0	0.07	0.194	0
16	0.80	1.468	1	0.40	0.456	1
17	0.32	0.407	1	0.12	0.058	1
18	0.25	5.660	0	0.05	0.166	0
19	0.10	0.161	1	0.10	0.153	1
20	0.00	0.000	0	0.00	0.004	0
21	0.15	0.517	1	0.20	0.166	1
22	0.50	4.641	0	0.00	0.020	0
23	0.00	1.189	0	0.00	0.000	0
24	0.00	2.177	0	0.00	1.678	0
Total	**0.16±0.21**	**1.514±3.149**	**7 (0.29)**	**0.16±0.23**	**0.354±0.607**	**10 (0.42)**