Universidade do Minho
Escola de Psicologia

Carla Isabel Dias Barros

Probing auditory processing in schizophrenia: an EEG-based deep learning approach

junho de 2023

Probing auditory processing in schizophrenia: an EEG-based deep learning approah

Carla Barros

UMinho | 2023

Universidade do Minho
Escola de Psicologia

Carla Isabel Dias Barros

Probing auditory processing in schizophrenia: an EEG-based deep learning approach

Tese de Doutoramento
Psicologia Básica

Trabalho efetuado sob a orientação de
Professora Doutora Ana Patrícia Teixeira Pinheiro
Professor Doutor Carlos Alberto Batista Silva

junho de 2023

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.
Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.
Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Attribution
CC BY

https://creativecommons.org/licenses/by/4.0/

Universidade do Minho, 7 de Junho de 2023

Assinado por: **CARLA ISABEL DIAS BARROS**
Num. de Identificação: 13423621
Data: 2023.06.07 22:44:57+01'00'

Assinatura:_____

Aknowledgments

I express my gratitude to those who have contributed to completing this long research journey, marked by dedication, personal growth, and the pursuit of innovative solutions.

To my supervisor, Professor Ana Pinheiro, for offering me the opportunity to join her research team and for challenging me to undertake a multidisciplinary Ph.D.. Her scientific guidance, knowledge sharing, and provision of resources have been essential. I also express my deep gratitude to Professor Carlos Silva, who accepted the role of guiding my research in the engineering domain, a task that I recognize has been challenging. Despite the difficulties encountered, Professor Silva's commitment to my mentoring was the driving force behind the completion of this work. His dedication, extensive knowledge, and stimulating discussions contributed significantly to my academic growth and the overall quality of the research.

To the Psychological Neuroscience Laboratory (PNL) for hosting me. And to CMEMS for generously providing me with the necessary facilities, even though it was not my host institution. Their support has been crucial to the successful execution of this work.

To Judith Ford and Brian Roach for sharing data, insightful suggestions, and willingness to discuss critical aspects of the research.

To Mariana Silva, Professor Jorge Santos, Sandra Mouta, and all the Laboratory of Visualization and Perception collaborators for their welcome and guidance during my Lab Rotation.

To my friends Paulinha, Diana, Margarida, Carolina, Andreia, and Eugénia, I am grateful for their stimulating ideas and shared learning experiences. A heartfelt thank to my dear friend Paulinha, who has been by my side through the sunny and stormy chapters of this journey. We've shared countless moments of happiness, laughter, triumphs, and challenges that I will never forget.

To all my friends at *#labdaspessoasfixes*, who have not only provided an inspiring work environment but also become a second family to me. Their friendship, laughter, and silliness made this journey memorable. I special thank to Susana for her daily patience, help, knowledge sharing, and constant motivation. I also thank Adriano, Sérgio, Cocas, and Patrícia for their insightful discussions and knowledge sharing. I extend my heartfelt thanks to Sara and Denise for their support, and to all those who, directly or indirectly, contributed to this work.

Finally, I am profoundly grateful to my beloved parents and sister, who have been my pillars of strength each and every day. Their unwavering support, patience, and unconditional love have been a constant source of inspiration. I dedicate this work to them, as without their presence and encouragement, I would not have achieved this significant milestone.

This work was supported by the Grant SFRH/BD/111083/2015, funded by Fundação para a Ciência e Tecnologia (FCT) under the Programa Operacional Capital Humano (PO CH) co-funded by Portugal 2020 and European Social Fund.



STATE OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Universidade do Minho, 7 de Junho de 2023

Assinado por: **CARLA ISABEL DIAS BARROS**
Num. de Identificação: 13423621
Data: 2023.06.07 22:46:23+01'00'

Signature: _____

Resumo

Investigar o processamento auditivo em esquizofrenia: uma abordagem de deep learning com base em sinais de EEG

O diagnóstico da esquizofrenia apresenta desafios devido à sua natureza complexa e heterogénea, ao relato subjetivo de sintomas, à possível sobreposição com outras condições psiquiátricas, e ao início dos sintomas durante o período desenvolvimental de elevada vulnerabilidade para diferentes tipos de psicopatologia (final da adolescência). A compreensão da percepção auditiva na esquizofrenia é crucial para obter pistas sobre a fisiopatologia, uma vez que alucinações auditivas são sintomas proeminentes. A electroencefalografia (EEG) fornece uma medida direta da atividade elétrica do cérebro, permitindo a investigação de alterações neurais associadas à percepção auditiva em indivíduos com esquizofrenia. A utilização de *machine learning*, especialmente métodos de *deep learning*, com dados de EEG tem o potencial de aprender padrões neurais e relações complexas, predição e identificação de biomarcadores em esquizofrenia.

Neste estudo, analisamos sinais de EEG de indivíduos saudáveis e de pacientes com esquizofrenia conduzindo uma análise combinada de potenciais evocados (ERP), *trial-by-trial* e tempo-frequência para investigar alterações no processamento auditivo. Com um tamanho de amostra maior do que em estudos prévios (n>60), observamos reduções consistentes no componente ERP N100 em esquizofrenia, especialmente em elétrodos fronto-centrais, em consonância com a literatura existente. A análise trial-by-trial mostrou respostas N100 e P200 consistentes entre ensaios, mas maior variabilidade nos sinais de pacientes com esquizofrenia. Diferenças significativas nas janelas temporais N100 e P200 enfatizaram ainda mais a robustez desses componentes. A análise de tempo-frequência revelou padrões distintos de sincronia da atividade neural nas oscilações theta e alfa entre os grupos de indivíduos saudáveis e com esquizofrenia, contribuindo para diferentes assinaturas do sinal EEG em resposta a sons simples.

Tirando partido do sucesso dos métodos de deep learning e das diferenças robustas observadas nos sinais EEG, desenvolvemos um modelo de classificação binária baseado em *single-trials* usando uma arquitetura baseada em redes neuronais convolucionais (CNN). O modelo é proposto para a classificação de segmentos de EEG, registados durante uma tarefa auditiva passiva usando a repetição de sons simples. Um estudo ablativo permitiu analisar o efeito do número de elétrodos utilizados, bem como a lateralidade e distribuição da atividade elétrica no escalpe. Os resultados mostram que o modelo proposto é capaz de classificar indivíduos com esquizofrenia e sujeitos saudáveis com uma precisão média de 80% usando apenas 5 elétrodos (Fz, FCz, Cz, CPz e Pz).

Este trabalho destaca a relevância dos métodos de deep learning na exploração de possíveis biomarcadores eletrofisiológicos para a esquizofrenia, bem como as potenciais aplicações clínicas dos sistemas de classificação. Além disso, demonstra o potencial transformador da colaboração interdisciplinar entre Psicologia, Psiquiatria e Engenharia Biomédica na revolução do diagnóstico, tratamento e prevenção de doenças mentais como a esquizofrenia.

Abstract

Probing auditory processing in schizophrenia: an EEG-based deep learning approach

The diagnosis of schizophrenia poses challenges due to its complex and heterogeneous nature, subjective symptom reporting, possible overlap with other psychiatric conditions, and typical symptom onset in a developmental stage characterized by high vulnerability to distinct types of psychopathology (end of adolescence). Understanding auditory perception in schizophrenia is crucial to get clues about the disorder's pathophysiology, as auditory hallucinations are prominent symptoms. Electroencephalography (EEG) provides a direct measure of brain electrical activity, allowing investigation of neural abnormalities associated with auditory perception in individuals with schizophrenia. The use of machine learning, especially deep learning methods, with EEG data in schizophrenia research has the potential to unveil intricate neural patterns and relationships, enabling more accurate diagnosis, prediction, and biomarker identification.

In this study, we analyzed EEG signals of healthy subjects and schizophrenia patients using combined event-related potentials (ERP), trial-by-trial, and time-frequency analysis to investigate auditory processing alterations. With a sample size larger than in previous studies (n>60), we observed consistent reductions in the N100 ERP component in schizophrenia, particularly over frontal-central electrode sites, aligning with existing literature. Trial-by-trial analysis showed consistent N100 and P200 responses across trials but greater variability in schizophrenia signals. Significant differences in N100 and P200 temporal windows further emphasized the robustness of these event-related components. Time-frequency analysis revealed distinct neural activity synchrony patterns in theta and alpha oscillations between healthy and schizophrenia groups, contributing to different EEG signatures in response to simple sounds.

Building upon the remarkable achievements of deep learning-based approaches and the observed robust EEG differences, a single-trial-based binary classification model was developed using a convolutional neural network-based architecture. The model is proposed to perform the classification based on multichannel auditory-related EEG single-trials recorded during a passive listening task of simple sounds. An ablative study allowed us to analyze the effect of the number of electrodes used and the laterality and distribution of the electrical activity over the scalp. Results show that the proposed model is able to classify schizophrenia and healthy subjects with an average accuracy of 80% using only five midline channels (Fz, FCz, Cz, CPz, and Pz).

This work highlights the relevance of deep learning methods in exploring potential electrophysiological biomarkers for schizophrenia and the potential clinical applications of such classification systems. Furthermore, it showcases the transformative potential of interdisciplinary collaboration between Psychology, Psychiatry, and Biomedical Engineering in revolutionizing the diagnosis, treatment, and prevention of psychiatric disorders such as schizophrenia.

Contents

Acronyms **xi**

List of Figures **xii**

List of Tables **xiv**

1 Introduction **1**

1.1 Context and motivation . . . 2
1.2 Overview of the research . . . 4
1.2.1 Objectives . . . 4
1.2.2 Contributions to knowledge . . . 5
1.3 Thesis outline . . . 5
1.4 Publications . . . 6

2 Theoretical Concepts and Technical Background **7**

2.1 Schizophrenia . . . 8
2.1.1 Auditory Sensory Processing - Afferent Pathway . . . 9
2.1.2 Auditory Processing abnormalities in Schizophrenia . . . 11
2.2 Electrophysiology . . . 12
2.2.1 Electroencephalography - EEG technique . . . 12
2.2.2 Event-related Potentials Analysis . . . 13
2.2.3 Time-frequency Analysis . . . 19
2.3 Machine Learning . . . 25
2.3.1 Classification Algorithms . . . 27
2.3.2 Deep Learning . . . 31
2.4 Summary . . . 41

3 Advanced EEG-based learning approaches to predict schizophrenia: a review **43**

3.1 Introduction . . . 44
3.2 Search strategy and inclusion criteria . . . 45
3.3 Schizophrenia Diagnosis . . . 45
3.4 Neurobiological changes in schizophrenia . . . 46
3.5 Electrophysiological Biomarkers of Schizophrenia . . . 47
3.6 Machine Learning for Schizophrenia Detection and Prediction . . . 48
3.6.1 Classical Machine Learning Approaches . . . 48
3.6.2 Novel approaches: the prospects of Deep Learning . . . 57

3.7 Promises and Pitfalls . . . 61
3.8 Summary . . . 64

4 Auditory event-related potential abnormalities in Schizophrenia 65
4.1 Introduction . . . 66
4.1.1 Current study and hypotheses . . . 67
4.2 Methods . . . 68
4.2.1 Participants and databases . . . 68
4.2.2 Experimental Procedure . . . 70
4.2.3 Stimuli . . . 71
4.2.4 EEG acquisition and preprocessing . . . 71
4.2.5 EEG data analyses . . . 72
4.3 Results . . . 75
4.3.1 ERP Analysis . . . 75
4.3.2 Trial-by-trial Analysis . . . 76
4.3.3 Time-frequency Analysis . . . 81
4.4 Discussion . . . 85
4.4.1 ERP vs. Trial-by-trial Analysis . . . 85
4.4.2 Cluster-based Permutation Tests . . . 85
4.4.3 Event-related EEG Analysis . . . 86
4.4.4 Changes in Oscillatory Activity . . . 87
4.5 Limitations and Future directions . . . 88
4.6 Summary . . . 89

5 Decoding auditory processing in schizophrenia using deep learning 90
5.1 Introduction . . . 91
5.1.1 Related Work . . . 91
5.2 Methodology . . . 93
5.2.1 EEG Data Acquisition . . . 93
5.2.2 Problem Formulation . . . 95
5.2.3 Deep Learning Architecture . . . 95
5.2.4 Traditional Machine Learning . . . 98
5.2.5 Ensemble Strategy . . . 99
5.3 Experimental Setup . . . 99
5.3.1 Data . . . 99
5.3.2 Deep Learning Algorithm . . . 100
5.3.3 Traditional Machine Learning Algorithms . . . 102
5.3.4 Ensemble Method . . . 103
5.3.5 Evaluation Metrics . . . 104
5.3.6 Statistical Analysis . . . 105

5.3.7 Implementation Details . . . 106
5.4 Results and Discussion . . . 106
5.4.1 Kernel size for temporal convolution . . . 106
5.4.2 Ablation Study . . . 106
5.4.3 Probabilities Distribution . . . 110
5.4.4 Model interpretation . . . 111
5.4.5 Relevance of SzNet . . . 113
5.4.6 Machine learning vs. Deep learning . . . 114
5.4.7 Ensemble learning . . . 116
5.4.8 Limitations and future directions . . . 116
5.5 Summary . . . 117

6 Final Remarks **119**
6.1 Conclusions . . . 120
6.2 Limitations and Future Directions . . . 122

Appendices **124**

A EEG Databases Merging **125**

B Deep Learning Model Interpretability **130**

References **132**

Acronyms

ANN artificial neural networks.

AOT auditory oddball task.

APA American Psychiatry Association.

ASSR auditory steady-state response.

AUC-ROC area under the receiver operating characteristic curve.

CHR clinical high-risk.

CNN convolutional neural network.

DC direct current.

DSM-5 Diagnostic and Statistical Manual of Mental Disorders - 5.

DSM-IV Diagnostic and Statistical Manual of Mental Disorders IV.

DT decision tree.

EAGBR early auditory-evoked gamma-band response.

EEG electroencephalography.

EOG electrooculograms.

ERP event-related potentials.

ERSP event-related spectral perturbations.

FC Fully-Connected.

FES first episode schizophrenia.

FFNN feed-forward neural network.

FFT Fast Fourier Transform.

fMRI functional magnetic ressonance imaging.

FN False Negative.

FP False Positive.

FPR false positive rate.

GABA gamma-amino-butyric acid.

GDP Gross Domestic Product.

GPC gaussian process classification.

Grad-CAM Gradient-weighted Class Activation Mapping.

HC healthy control.

ICA independent component analysis.

ICD-10 International Statistical Classification of Diseases.

ISI interstimulus interval.

kNN k-nearest neighborhood.

LDA linear discrimination analysis.

LES Linear Eigenvalue Statistics.

LSTM long short-term memory.

MCC Matthews correlation coefficient.

MDC multi-domain connectome.

MEG magnetoencephalography.

MLP multilayer perceptron.

MMN mismatch negativity.

MNE minimum-norm estimate.

MRI magnetic ressonance imaging.

NB naïve bayes.

NIMH National Institute of Mental Health.

PANSS Positive and Negative Symptoms Scale.

PCA Principal component analysis.

PLF Phase-locking factor.

PNN probabilistic neural network.

RDoC Research Domain Criteria.

ReLU rectified linear units.

RF random forest.

RNN recurrent neural network.

ROC receiver operating characteristic.

ROI region of interest.

RSz at-risk of developing SZ.

SANS Scale for Assessment of Negative Symptoms.

SAPS Scale for Assessment of Positive Symptoms.

SCID Structured Clinical Interview for DSM.

SNN siamese neural network.

SNR signal-to-noise ratio.

SPL sound pressure level.

SPN spatial pattern network.

SVM support vector machine.

SVM-poly SVM with polynomial function kernel.

SVM-RBF SVM with radial-basis function.

SWMT Sternberg Working Memory Task.

SZ schizophrenia.

TFCE threshold-free cluster enhancement.

TN True Negative.

TP True Positive.

TPR true positive rate.

UHR ultra-high-risk.

WHO World Health Organization.

XGBoost Extreme Gradient Boosting.

List of Figures

1 Illustration of ear structures with enlarged cochlea. 10
2 Illustration of the human auditory afferent pathway. 11
3 Distribution of electrodes sites over the scalp and schematization of EEG acquisition concept. 13
4 Procedure from EEG to ERP. 14
5 Overview of the auditory ERP components. 17
6 EEG signal decomposition into frequency bands with functional meaning. 19
7 Heisenberg box representing an atom. 21
8 Heisenberg time-frequency boxes of two wavelets. 22
9 Complex Morlet wavelet. 22
10 Time-frequency decomposition of EEG signal. 25
11 Different types of learning in machine learning. 26
12 Schematic of a biological neuron and a perceptron. 31
13 Schematic of a MLP with 3 hidden layers. 32
14 Gradient descent represented for convex and non-convex loss functions. 36
15 Diagram illustrating a typical architecture of a CNN. 39
16 Outline of a convolutional layer. 40

17 PRISMA flow diagram of the systematic literature search. 45
18 Classifiers tested in binary classification of SZ. 53
19 Classification accuracies achieved with visual and resting state paradigms. 54

20 Dataset's composition in number of HC and SZ patients. 69
21 EEG experimental paradigm with three blocks: *auditory-motor*, *auditory-only*, and *motor-only* conditions. 72
22 Topographical maps for N100 and P200 amplitude distribution, and the grande average ERP waveforms. 76
23 Single-trial maps plotted for the FCz, Cz, and CPz scalp channels, comparing HC to SZ responses. 77
24 Difference waveforms between the average evoked response of all HC and SZ single-trials plotted for the 64 EEG channels. 78
25 Significant spatio-temporal clusters. 79
26 Average of the difference waveforms plotted for each electrode, masking time points for significance. 80
27 Average PLF time-frequency maps for HC and SZ groups plotted for FCz electrode. . . . 81

28 Time-frequency maps of the average values of (a) total power, (b) evoked power, and (c) induced power acquired at the FCz electrode, plotted for the HC and SZ groups. 82
29 Time-frequency permutation maps for PLF and total, induced, and evoked power. 84

30 EEG electrodes placement and experimental paradigm scheme. 94
31 ERP results over Cz electrode. 94
32 Pseudo-image created with EEG of 5 midline electrodes. 96
33 End-to-end process from signal acquisition to final classification. 97
34 10-fold cross-validation stratification. 101
35 Scheme of the feature extraction from ERP waveform. 103
36 Electrode alignment in the 2D structure and convolutional process. 108
37 Topographical distribution of electrodes used to test different classification models. . . . 108
38 ROC-curve (and corresponding AUC-score) and distribution of the probabilities of SZ and HC predicted by the SzNet-5-midline,considering a subject-based approach. 110
39 ROC-curve (and corresponding AUC-score) and distribution of the probabilities of SZ and HC predicted by the SzNet-5-midline,considering a segment-based approach. 111
40 GradCAM heatmaps for the segments most likely to belong to HC and SZ subjects. . . . 112
41 Grad-CAM heatmaps and EEG segments averages of the HC and SZ EEG segments. . . 112
42 Grad-CAM heatmaps cross-section and averaged EEG signals over Cz electrode. 113
43 Random Forest features importance. 115
44 Effects of the growing number of ensembled models on different metrics, with emphasis on accuracy. 117

45 Results of spatiotemporal cluster-based permutation test on HC subjects data between datasets *A* and *B*. 126
46 Results of spatiotemporal cluster-based permutation test on SZ subjects data between datasets *A* and *B*. 127
47 Cluster-based permutation test on temporal dimension of 5 midline electrodes to compare HC subjects data between datasets *A* and *B*. 128
48 Cluster-based permutation test on temporal dimension of 5 midline electrodes to compare SZ subjects data between datasets *A* and *B*. 129

49 GRAD-CAM heatmaps cross section and averaged EEG signals for Fz, FCz, CPz, and Pz electrodes drawn for HC and SZ groups. 131

List of Tables

1 Description of advantages and disadvantages of different supervised machine learning algorithms. Schematization of decision boundaries resulting from the application of each algorithm. 27
2 Commonly used activation functions. 33
3 Examples of proposed EEG biomarkers of SZ. 47
4 Schizophrenia classification attempts using Classical Machine Learning methods based on the EEG signal. Categorization of the selected studies according to classification task and EEG paradigm. 49
5 Schizophrenia classification attempts using Representation Learning methods based on EEG signals. 59
6 Clinical and demographic data. 69
7 Demographics of HC and SZ patients in merged dataset. 71
8 Peak amplitude, latency and location on the scalp for the N100 and P200 components found in the grand average waveforms of the HC and SZ groups. 75
9 List of hypermeters selected for the proposed method. 102
10 Confusion Matrix. 104
11 Architectures of the SzNet model adapted for the three spatial approaches tested. 107
12 Performance of the models tested (SzNet-5, SzNet-15 and SzNet-35) using different spatial ROI evaluated with five metrics: accuracy; AUC-ROC; precision; recall; and specificity. 109
13 Results of the classification performed by the RF, the proposed deep learning model, and the ensemble learning method. 116

Chapter 1

Introduction

Effective and accurate diagnosis of psychiatric disorders has significant implications for clinical management, practice, and life expectancy. Despite years of research, the mechanisms underlying the onset, symptomatology, and treatment of schizophrenia (SZ) are still elusive. The absence of appropriate analytical tools to deal with the heterogeneity and complexity of SZ may be one of the reasons behind this situation. Artificial intelligence advances have recently provided plausible methods of modeling and analyzing complex, high-dimensional, nonlinear systems. The unprecedented accuracy of such algorithms in classification and prediction tasks has revolutionized many scientific fields and is rapidly pushing SZ research forward. Deep learning, a subfield of artificial intelligence, has excellent potential to become a valuable tool for predicting, diagnosing, and treating SZ in clinical practice. In addition, deep learning can become a powerful tool for uncovering the mechanisms underlying SZ, thanks to a growing number of techniques designed to improve model interpretability.

This chapter presents the background and rationale for the research developed in the context of this Ph.D. project, the main objectives, and the main contributions to knowledge. Section 1.3 describes the structure of this document, giving a general overview of the work developed. The chapter closes with a list of the academic outputs, more precisely publications in international peer review journals that resulted from this research.

Contents

1.1 Context and motivation **2**
1.2 Overview of the research **4**
1.3 Thesis outline **5**
1.4 Publications **6**

1.1 Context and motivation

SZ is a chronic and complex brain disorder that impacts social and cognitive functioning and quality of life [1, 2]. Besides the tremendous financial burden on health systems, families, and society [3], SZ symptoms can impair daily functioning and be disabling, reducing the patients' quality of life [4].

Early diagnosis and effective treatment of SZ remain ongoing challenges for the scientific community [5, 6, 7]. Psychotic symptoms generally arise in late adolescence, often developing into the first psychotic episode in early adulthood [8, 9, 10]. In its most common form, this syndrome presents paranoid delusions and auditory hallucinations. About 70% of patients experience hallucinations in the auditory modality, most frequently as voices [11]. The high prevalence of auditory hallucinations among patients with SZ suggests alterations in brain circuity underlying auditory processing [12]. Consistently, auditory dysfunction may interfere with everyday life communication in patients with neuropsychiatric disorders [13]. Cognitive impairment is recognized as a critical component of the disorder [14]. Apart from the research focused on cognitive alterations, during the last decades, several articles were published about sensory processing disorders in SZ, focusing on auditory processing [15].

Efforts have been made to specify the neurobiological mechanisms underpinning psychotic symptoms, improving the diagnosis and prognosis of SZ [16]. A substantial body of research has been developed to understand the biological abnormalities that underlie SZ, using different techniques such as electroencephalography (EEG). The low cost and the noninvasive nature of the EEG make it a widely used technique to investigate the underlying neural mechanisms involved in perceptual, cognitive or affective processes in healthy and clinical populations [17, 18, 19]. Given its high temporal resolution, in the order of milliseconds, event-related potentials (ERP) (electrical activity elicited in response to a stimulus) provide a suitable method for studying the time course of the brain activity during the auditory processing [17, 18, 19].

Electrophysiological studies have been conducted to explore the role of the auditory cortex in processing auditory information, and consistent alterations of some auditory ERP components have been proposed as potential SZ biomarkers [20]. The extensively studied reductions in the mismatch negativity (MMN) and P300 amplitudes are the most common suggested EEG indices of SZ [21, 22]. Beyond these measures, some studies have focused on abnormalities in the N100 and P200 ERP components generation [17, 23, 24, 25].

Although the elicited ERP components appeared to indicate alterations in the auditory processing of SZ patients and are considered potential biomarkers, it has been difficult to correlate these findings with specific SZ symptoms. This difficulty is partly explained by the fact that researchers have focused more on diagnosis than symptoms [26]. Indeed, SZ treatment is dependent on an accurate and early diagnosis. Though its diagnosis mainly relies on assessing clinical symptoms' behavioral manifestations, it depends on self-reported subjective experience [27]. In addition, the complex nature of SZ itself and the problematic delimitation from other psychiatric disorders become harder for the disease recognition task [5]. Researchers are increasingly seeking sensitive and specific diagnostic tools driven by developments in the computational neuroscience field.

The future of healthcare is being shaped by advances in digital technologies, such as artificial intelli-

gence [28, 29]. Health information technologies have been going up dramatically in the last few years, primarily due to the improved computational power and a large amount of data available nowadays [30, 31]. A better grasp of complex medical and physiological issues has only been possible because of considerable increasis in the capacity of machines to understand and decode data [32]. Machine learning techniques are revolutionizing healthcare by empowering systems to learn from data and extract valuable insights [32, 33, 34]. They have contributed to resolving healthcare issues such as early diagnosis of disease, real-time patient monitoring, patient-centric care, and enhancement of treatment [35]. Given the ever-expanding availability of data concerning an individual's mental health condition, machine learning technologies offer a promising approach to enhance the understanding of psychiatric disorders and inform clinical decision-making [36]. Recent studies suggest several promising directions for applying machine learning methods in improving the diagnosis and treatment of psychiatric disorders [36]. Despite this impressive progress, only minimal knowledge has been gathered on SZ and auditory sensory processing abnormalities.

Machine learning applied to EEG data could represent a promising contribution to accurately predicting intervention response and disease trajectories [37]. Machine learning methods have the potential to significantly enhance our understanding of the brain mechanisms underlying SZ across three key domains [38]. Firstly, in diagnosis and detection, these techniques can improve accuracy by analyzing various data types such as neuroimaging, genetics, and clinical assessments. Secondly, in prognosis and onset prediction, they can leverage longitudinal data to identify early indicators and risk factors, enabling personalized interventions. Lastly, machine learning algorithms can aid in biomarker identification, revealing novel markers associated with SZ by integrating multi-modal data [39, 40].

Most of the implemented approaches in SZ are based on supervised learning models. Through advanced statistical and probabilistic techniques, these algorithms can parse the dataset (described by a set of labeled features), learn from it, and then predict the label for new data, making an informed decision [41, 42]. Machine learning has been broadly applied in schizophrenia (SZ), recognizing patterns in the data capable of successfully discriminating SZ patients from healthy subjects [38, 39]. Likewise, the classification of SZ using EEG data during auditory processing has already been attempted. Promising results were obtained using early auditory ERP component data measures as input features for machine learning models [43, 44, 45, 46, 47, 48]. In addition to successfully applying the learning methods, the results reinforce the importance of the ERP components that have been recurrently proposed as SZ biomarkers. Although these strategies have shown accurate results, feature extraction and selection require domain expertise. New challenges in data analysis boosted the evolution and improvement of machine learning algorithms throughout the years, emerging new learning paradigms. Deep learning is a relatively recent approach that can solve the dependence on the features [49]. This method has become increasingly popular because the algorithms learn high-level features directly from the raw data through hierarchical learning processes [50]. EEG-based deep learning algorithms have also seen growing interest among neuroscientists, especially in the task of brain-signal decoding [51, 52, 53]. While applying these techniques to fields such as image analysis is already well established [54, 55, 56, 57], the use of deep learning is still poorly explored in the development of EEG-based systems for SZ investigation.

1.2 Overview of the research

This section provides a comprehensive outline of the objectives and research questions the present work aims to address. Afterward, the primary contributions of the research are presented, with a particular emphasis on their broader implications for the relevant field of study.

1.2.1 Objectives

The amount and relevance of the information in the electroencephalographic signal make it an ideal candidate for applying deep learning methods, which can extract complex patterns and relationships in the data and provide valuable insights into brain function and dysfunction. Learning features from these signals may therefore represent a significant advance in studying SZ.

This thesis has two main objectives. The first is to comprehensively explore and analyze the EEG signal content in time and frequency domains during early auditory processing in response to simple stimuli. The second is to implement deep learning methods to extract SZ discriminative features from the EEG signal to create a binary classification model and address the abovementioned challenges, enabling more effective diagnosis. The specific primary objectives of this thesis are:

- To characterize and evaluate the machine and deep learning methods recently used in the SZ EEG-based research literature, discussing their potentialities and limitations;
- To study and determine which changes occur in evoked signals during early auditory processing in SZ. Since auditory impairments have been documented in SZ [23, 9, 58, 59, 60], we hypothesize that the signal's dynamic during sound processing is altered in this disorder;
- To apply appropriate EEG analysis techniques to extract information from data collected from a group of patients and controls and test for differences between groups using appropriate statistical analysis methods;
- To design and implement a deep learning architecture for the binary classification of SZ based on EEG signals;
- To assess the validity of this approach by training the developed algorithm with auditory event-related signals and testing its performance using validation metrics;
- To compare classical machine learning with deep learning architectures to classify SZ based on auditory EEG responses. Given the advantages of deep learning over machine learning methods [50, 42], we hypothesize that deep neural networks will be able to learn different patterns based on auditory EEG signals, allowing a more accurate classification of SZ subjects;
- To understand which signal features contribute more to the discrimination, and the relationship between those features with EEG signals alterations in SZ consistently suggested as biomarkers of the disorder.

1.2.2 Contributions to knowledge

Accordingly, the main contributions are the following:

- A critical review that serves as a starting point for future developments of effective EEG-based models to predict the onset of SZ, to identify subjects at risk of psychosis, or to differentiate SZ from other disorders (Chapter 3).
- The use of a passive listening task for EEG data recording to provide discriminatory information. The proposed task is of short duration, easily implemented in a clinical environment, and does not require an overt response from the subject (Chapter 4).
- An extensive analysis of brain electrical information recorded during early auditory processing from SZ and healthy subjects, using techniques such as EEG single-trials analysis, ERP, or time-frequency analysis (Chapter 4).
- The investigation of trial-by-trial spatio-temporal variability that allows the assessment of the robustness and consistency of changes found in EEG single-trials in SZ and their use for the development of classification algorithms (Chapter 4).
- A multi-channel deep learning architecture for SZ event-related EEG single-trials classification. The proposed model is based on convolutional neural networks, extracting task-related features from the time course and spatial distribution of electrical activity of the brain in response to auditory stimuli. The results demonstrate its high classification accuracy, given the heterogeneity of the disorder (Chapter 5).
- The specification of which number of electrodes and which scalp localization should be used to most accurately predict whether the signal belongs to a SZ subject or not. The Chapter 5 sheds light on which electrode location is most relevant for identifying changes in brain activity during auditory processing since most studies in this field use a different number and combination of electrodes.
- The use of ensemble methods to reduce the variance of predictions of the developed deep learning model (Chapter 5).
- A demonstration of the relevance of deep learning methods in the study of potential SZ electrophysiological biomarkers, as well as the potential of using this type of classification system for clinical purposes (Chapter 5).

1.3 Thesis outline

Structured to address and probe the auditory processing in SZ individuals, this thesis also examines the impact of electrophysiological data alterations in discriminating healthy subjects from diagnosed patients.

This chapter has supplied information on the background, motivation, goals, contributions, and outcomes and now presents the remaining structure of this thesis.

Chapter 2 describes the main theoretical and technical concepts used to conduct the research project. This chapter briefly overviews the aetiology and prevalence of SZ, besides describing how sounds are perceived in this disorder. Additionally, it describes the electrophysiological technique used in data acquisition and the endophenotypes associated with impaired auditory processing in SZ. This chapter also introduces the fundamentals of machine learning, with particular emphasis on the deep learning topic.

Chapter 3 reviews the more recent machine learning methods applied to EEG data for SZ classification. This chapter is an overview of existing research relating to the objectives of this thesis, highlighting limitations of the use of EEG data, some good practices, and promising applications of these methods.

Chapter 4 presents a combined ERP, EEG single-trials, and time-frequency approach used to investigate the mechanisms underlying auditory processing during a passive listening task in SZ. This chapter presents a detailed characterization of the EEG dataset and discusses the changes found signal components in SZ.

Chapter 5 describes in detail the application of deep learning methods for EEG signal classification. The auditory event-related EEG single-trials were used to discriminate SZ from healthy subjects. In addition, this chapter explores the contribution of features learned by the SZ classification model through a neural network interpretability technique.

Chapter 6 presents the principal conclusions of the thesis. This chapter summarizes the most relevant findings and results, points out the implications and limitations of the studies conducted, and presents the prospects for future research on these topics.

1.4 Publications

Original scientific papers resulted from the research conducted in the context of this Ph.D. project:

- **Barros, C.**, Silva, C. A., & Pinheiro, A. P. (2021). Advanced EEG-based learning approaches to predict schizophrenia: Promises and pitfalls. *Artificial intelligence in medicine, 114*, 102039.
- **Barros, C.**, Roach, B., Ford, J. M., Pinheiro, A. P., & Silva, C. A. (2022). From Sound Perception to Automatic Detection of Schizophrenia: An EEG-Based Deep Learning Approach. *Frontiers in psychiatry, 12*, 813460.

Chapter 2

Theoretical Concepts and Technical Background

This chapter introduces relevant theoretical and technical concepts at the core of the research topic addressed in this thesis. The first section defines SZ from a historical perspective. Furthermore, it focuses on alterations of the auditory processing in SZ and their relevance to the understanding of auditory hallucinations. This section also describes the afferent auditory pathway, critical for the understanding of the neural mechanisms engaged in sound decoding. The EEG was the neuroimaging technique chosen to investigate auditory processing alterations in this Ph.D. project. Therefore, the second section of this chapter focuses on this technique and presents its general principles. In addition, this section details the analysis of evoked potentials and EEG data in the time-frequency domain. Another objective of this work is the categorization of SZ. The task of classification was performed using machine learning methods. Therefore, the third section presents a broad overview of machine learning. Deep learning methods are emphasized, since they were used to classify EEG data from SZ patients in the present work. This section introduces convolutional neural networks and other key concepts.

Contents

2.1 Schizophrenia **8**
2.2 Electrophysiology **12**
2.3 Machine Learning **25**
2.4 Summary **41**

2.1 Schizophrenia

SZ is a chronic and complex neuropsychiatric disorder affecting over 21 million people worldwide [61]. Its prevalence in Portugal is estimated between 0.6% and 1%. Despite being a disabling disease, the global cost of SZ has been estimated at EUR 436.3 million, about 0.24% of the Portuguese Gross Domestic Product (GDP) in 2015 [62]. SZ is one of the most debilitating psychiatric disorders, affecting an individual's daily life and social function. [63]. SZ is considerably heterogeneous in its symptomatic manifestations, which is associated with heterogeneity in treatment response and long-term prognosis [64].

The term '*schizophrenia*' was first introduced by Eugen Bleuler in 1911 [65], although it was first described by Emil Kraepelin fifteen years earlier as *dementia praecox* [66]. The concept and nosology of SZ evolved over the years since then [67]. Nevertheless, three historical perspectives have contributed to our current understanding of SZ [68]:

- Kraepelin adopted the concept of degeneration as a diagnosis criterion, characterized by the progressive and gradual loss of mental abilities [66].
- Bleurer focused on some symptoms that would be fundamental for the diagnosis of SZ. He described the abnormal associations of ideas and thought as the first elementary function or basic symptoms. Alterations in affective processing constituted other basic symptoms of SZ. Bleuler included the following formal alterations: affective weakening, affective indifference, euphoria, depression, anguish, rigidity, and bluntness. The secondary symptoms, although not relevant for the diagnosis of the disease, could be associated with the fundamental symptoms, namely hallucinations, delusions, disorders of memory, identity, language and somatic symptoms [65].
- Years later, Kurt Schneider was interested in identifying groups of symptoms found in SZ but rarely in other disorders [69]. He established a hierarchy of symptoms according to their importance for SZ diagnosis: first- and second-order symptoms. The first-order symptoms included delusional perception, auditory hallucinations (mainly in the form of voices), sonorization of thought, experiences of influence on thought, and diffusion of thought. Delusional intuition, perplexity, depressive or manic moods, affective weakening, and other sensory disorders were part of the second-order symptoms [11].

Current diagnostic criteria have incorporated the chronicity proposed by Kraepelin, the negative symptoms proposed by Bleuler, and the positive symptoms proposed by Schneider, using different combinations and variable interpretations of these elements [68]. The positive and negative dimensions of the symptoms have been examined in the pathogenesis and course of SZ. Whereas positive symptoms are a reflection of excess or distortion of normal function (e.g., delusions, hallucinations, disorganized behavior), negative symptoms represent a decrease or absence of normal behavior associated with motivation and interest (e.g., avolition, anhedonia, asociality) or expression (e.g., blunted affect, alogia) [70, 71, 72, 73]. The SZ symptoms are assessed through well-established scales which were developed to monitor the severity of positive and negative symptoms and treatment responses. Examples of widely used scales are the

Positive and Negative Symptoms Scale (PANSS), the Scale for Assessment of Positive Symptoms (SAPS), or the Scale for Assessment of Negative Symptoms (SANS) [74]. While traditional diagnostic criteria for SZ focus on positive symptoms and negative symptoms, cognitive deficits are increasingly recognized as a key aspect of the disorder [75, 76, 77, 78, 79]. They include poor executive functioning, lack of attention, and working memory deficits, and are already observed at disease-onset [75, 78, 79].

Inadequate reception, modulation, integration, discrimination or organization of external sensory information can affect higher-order cognitive processes [80]. Information processing can occur in two directions [81]: bottom-up and top-down. In bottom-up processing, perception starts at the sensory input (data-driven), and the information is carried from the sensory organ to the brain. The brain then analyzes the sensory input and creates a perception based on the physical features of the stimulus. In top-down processing, the stimulus is analyzed on the basis of available contextual information (knowledge-driven) [81]. In this process, the brain first uses its existing knowledge and expectations to analyze the sensory input and make a prediction, which is then compared to the actual sensory input to confirm or revise the prediction [82, 83]. Current notions of bottom-up and top-down processes are not strict [84]. According to the predictive coding theory, there is an interdependence and a constant interaction between bottom-up and top-down processes [84].

Auditory verbal hallucinations are a cardinal symptom of SZ [11, 85, 86]. Auditory verbal hallucinations are experienced by 60-80% of SZ patients [87, 88, 89]. Hallucinations are defined as sensory experiences that occur without external stimulation of the sensory organ [90]. The auditory nature of this psychotic symptom raises the possibility of associated auditory processing deficits [91]. The desynchronization between the networks involved in sensory processing (bottom-up) and top-down control can play a key role in the phenomenology of auditory verbal hallucinations [92, 93, 94]. Models focusing on top-down processes suggest that the assignment of aberrant salience to external sounds or external misattribution of inner-speech account for auditory hallucinations [93]. From another perspective, the association between sensory processing abnormalities (deficits in bottom-up process) and auditory hallucinations has already been reported in the literature, despite normal hearing acuity in SZ patients [95]. An imbalance between top-down and bottom-up processing has also been linked to auditory verbal hallucinations in SZ [96]. According to this perspective, the sensory impairment caused by bottom-up deficits or hyperexcitability at rest, followed by diminished sensory input, leads to overreliance on top-down processes. This increased dependence on top-down expectations, combined with decreased top-down inhibitory control, may lead to hallucinations [96].

2.1.1 Auditory Sensory Processing - Afferent Pathway

The auditory system is fundamental to human communication. Sounds are mechanical waves created by vibrational energy propagating longitudinally in the air [97]. The human hearing system can detect atmospheric pressure variations caused by the propagation of sound waves. When the sound hits the ear, the wave propagates through the external auditory canal, impacting the tympanic membrane. The vibration caused in the tympanic membrane is translated into movement and vibration of the three ossicles

in the middle ear [98, 99].

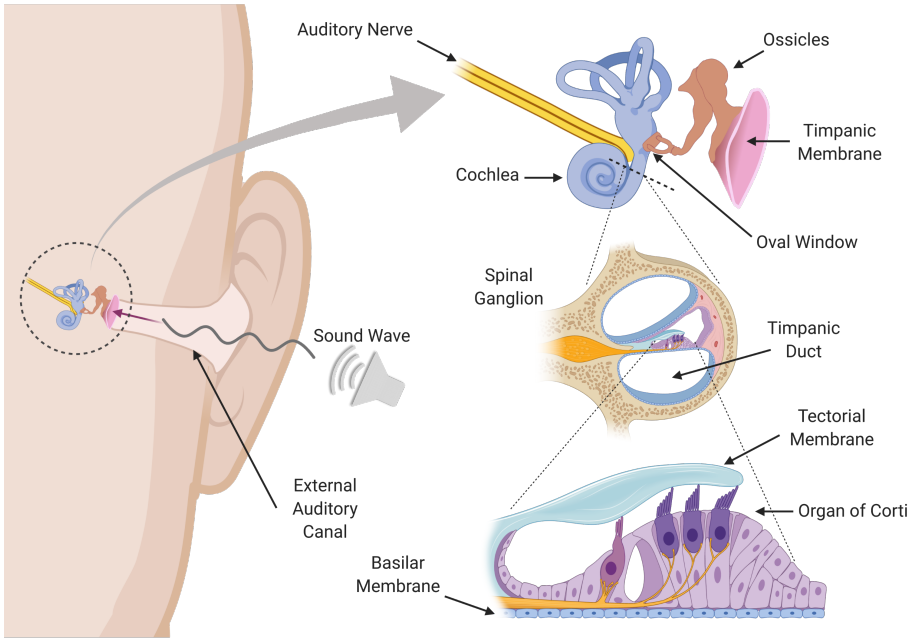


Figure 1: Illustration of the external ear (auricle, ear canal, and tympanic membrane), middle ear (ossicles), and inner ear (cochlea), with enlargement of the different structures (Created with BioRender.com).

Then, vibrations of the corresponding frequency are transmitted to the oval window of the inner ear. The oscillating movement is then propagated to the cochlear structure, causing the vibration of the basilar membrane. The organ of Corti is the sensory receptor in the cochlea responsible for transmitting information about sound to the brain. The ion channels of the inner and outer hair cells that compose the organ of Corti are susceptible to mechanical vibrations. The vibrations of the basilar membrane lead to the opening of ion channels, generating electrical energy that is propagated along the auditory nerve enervating the cochlear structure (Figure 1) [98, 99]. The cochlear nerve cell axons constitute the bilateral afferent sensory input to the central auditory nervous system [100]. The electrical signal travels along the auditory nerve up to a series of nuclei through crossed and uncrossed projections. The signal synapses within the cochlear nucleus in the brainstem. The information is then transmitted to the superior olivary nucleus, where binaural processing occurs. Through the lateral lemniscus pathway, the signal ascends to the midbrain's inferior colliculus and onto the thalamus's medial geniculate nucleus. The thalamus sends projections to the primary auditory cortex on the transverse temporal (or Heschl's) gyrus, where perception and interpretation of sounds is thought to occur (Figure 2) [99, 101, 98]. Heschl's gyrus is part of the primary auditory cortex, which is surrounded by the secondary auditory processing areas in the superior temporal gyrus [98].

Each afferent fiber is responsible for encoding the presence, amplitude, and timing of the sound energy within a frequency band [100]. The tonotopic arrangement of central projections and nuclei from the brainstem to the cortex preserves the frequency information of the sound [104, 101]. Different frequencies are detected by different primary auditory cortex neurons located on the superior temporal gyrus in the temporal lobe, forming a precise tonotopic map [105]. Each brain hemisphere is capable of lo-

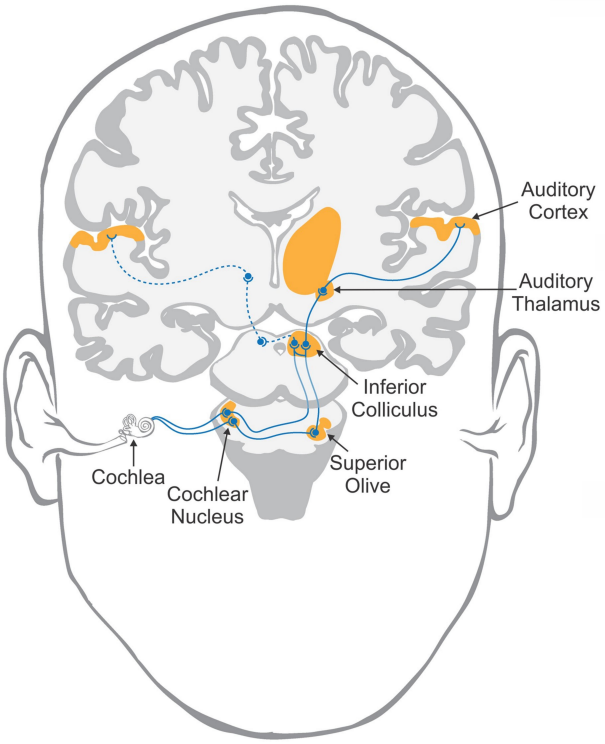


Figure 2: Illustration of the human auditory system highlighting the afferent pathway of auditory processing. Reproduced from [102]. Reprinted from Publication [103], Copyright 2020, with permission from Elsevier.

calizing sound sources. [99]. Fibers crossing defines both ipsilateral and contralateral input. Most of the projections take a contralateral pathway from each ear, which aids sound localization and interpretation [98]. Besides source localization and frequency information, temporal processing is at the core of the auditory system, especially for speech perception in which temporal sequences of sounds have to be interpreted [100]. Simple auditory tasks, such as detecting the presence of a pure tone, require relatively less neural resources when compared to more complex tasks, such as discriminating different acoustic stimuli in noise [100].

2.1.2 Auditory Processing abnormalities in Schizophrenia

Impairments in early sensory information processing have been consistently identified in SZ patients and individuals at clinical risk for developing psychosis [106, 107, 23, 108, 109]. Sensory deficits are well documented in the auditory system [110, 23, 95, 106, 15]. Studies based on different auditory processing tasks have shown that the neural responses to sounds are diminished in SZ patients compared to healthy subjects, which suggests abnormalities in the central auditory nervous system in this disorder [111]. These impairments appear even before the first psychotic episode, and may provide important insights into the pathogenesis of SZ [112]. Early auditory processing deficits in SZ manifest in difficulties in the discrimination of sound length, intensity, and pitch [109]. In addition, SZ patients also present aberrant attentional modulation of neural responses to simple sounds [106, 113].

Neuroanatomical changes have been associated with SZ, reflected in volume loss in cortical and

subcortical brain areas [23]. Reductions of gray matter volume and cortical thickness in the Heschl's gyrus and the superior temporal gyrus have been reported [114, 115, 23, 116]. Specifically, morphological alterations of the auditory cortex have been negatively correlated with auditory hallucinations severity [115, 117, 118]. However, a recent study demonstrated that the decreased volume of the right superior temporal gyrus is inversely correlated with psychosis severity but not with auditory hallucinations [119]. These brain changes can rather be explained as the cumulative effect of neurodevelopmental abnormalities, such as changes in neuroplasticity and alterations in neuronal maturation [23, 112].

Abnormal auditory processing in SZ has been examined in neuroimaging studies. fMRI studies have shown a lower activation of the auditory cortex during external stimulation in SZ patients with hallucinations when compared to non-hallucinating healthy subjects [114]. Other studies reported increased activation of the auditory cortex during auditory hallucinations in SZ patients [120, 121, 122]. This evidence confirms altered auditory sensory processing in SZ. However, other neuroimaging techniques, such as electrophysiology, can benefit from their excellent temporal resolution to shed light on the time course of auditory processing in SZ, differentiating sensory-based and higher-order operations relevant to sound decoding.

2.2 Electrophysiology

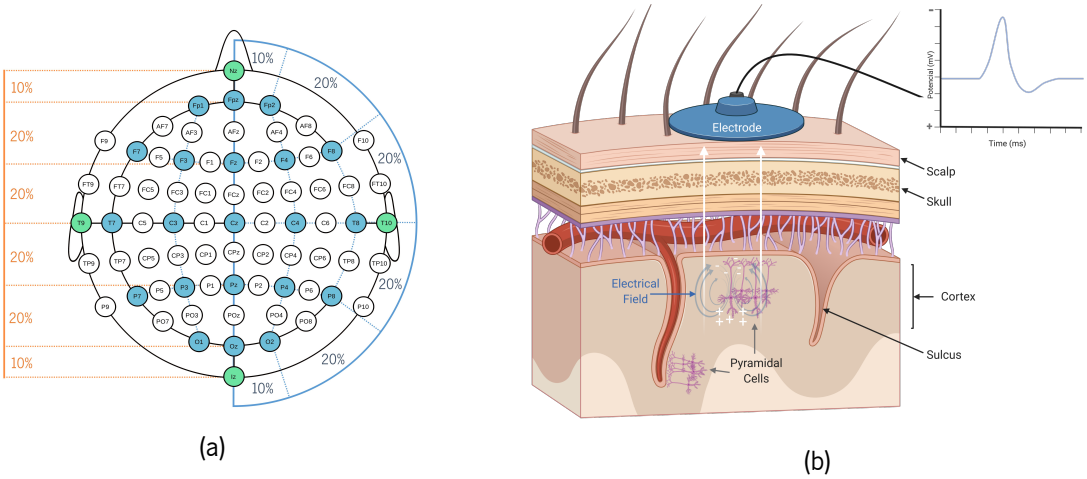
Electrophysiological studies have contributed to progress in psychology or neurosciences, enabling researchers to uncover the neural mechanisms underlying perceptual, motor, and cognitive processes. Functions such as perception or memory are evaluated and studied by examining the electrical activity generated by a population of neurons [19]. Non-invasive electrophysiology techniques, such as EEG, have been used in the investigation of neurocognitive processes for a long time. However, more sophisticated analyses have been developed in recent years to extract different types of information from the acquired data. These advances have enabled exploring the dynamics of brain electrical activity and studying its relationship with behavior and cognition [19].

2.2.1 Electroencephalography - EEG technique

MRI and fMRI are techniques characterized by their high spatial resolution, which allows specifying the morphology and volume of brain structures. Additionally, fMRI allows for the detection of changes in blood flow and oxygenation of different brain regions during stimulus processing, providing insight into the neural networks that underlie perception, higher-order cognition, and behavior [123]. Unlike these neuroimaging techniques, EEG stands out for its excellent temporal resolution, capturing the fast and sequential dynamics of events that occur in temporal windows of the order of hundreds of milliseconds [124].

The EEG technique allows recording electrical brain activity using electrodes placed on the scalp. Conventional methods for electrode positioning have relied on the international 10/20 system or its expanded systems (10/10 and 10/5). These methods first divide the head's surface in percentages (10% and 20%) of circumferential measurements from several anatomical landmarks, used as fiducial points [125, 126]. Figure 3a illustrate the positioning of 64 electrodes following the expanded 10/10 international system,

with the blue-shaded electrodes placed according to the 10/20 system.



(a) (b)

Figure 3: a) Distribution of electrodes sites over scalp using a 64-channels EEG following the expanded 10/10 international system; b) Fundamental concept of EEG acquisition: capture of pyramidal cell electrical fields via scalp electrodes (Created with BioRender.com).

Each electrode over the scalp captures neuronal activity mainly from superficial cortical layers [127]. Post-synaptic potentials in pyramidal cells, which have perpendicular orientations relative to the surface of the cortex, provide the most significant contribution to EEG signals [128]. When post-synaptic potentials release excitatory or inhibitory neurotransmitters into pyramidal cells, an ionic flow between the neuron and the extracellular space is triggered [129, 130]. Potential differences along the cell's membrane are caused by the ionic current, generating an electrical field around the neuron (Figure 3b), which is too weak to be captured by the electrode [129]. However, when the activity of thousands of pyramidal neurons is synchronized, and the ion currents flow in the same direction, the sum of the electrical fields generated by each neuron becomes powerful enough to be measured at the scalp [131, 132].

2.2.2 Event-related Potentials Analysis

Spontaneous EEG reflects the brain's activity generated without an explicit task, frequently correlated with resting-state or default-network activity [133]. Nevertheless, in everyday life, people perceive and respond to many different types of stimuli [134]. ERP are evoked or event-related responses in the brain generated in response to specific internal or external events (e.g., stimuli, responses, decisions) [18]. Controlled, long or short, and repeated or unique stimuli can elicit these potentials, which represent signatures of specific perceptual, motor, or cognitive processes [134]. Figure 4 illustrates how ERPs can be obtained from the EEG signal.

ERP waveforms are obtained after performing several preprocessing steps on the EEG signals. A general EEG acquisition and processing pipeline is followed for each subject, meeting the standardized method proposed by [135]:

1. During the experiment, continuous EEG data are recorded.

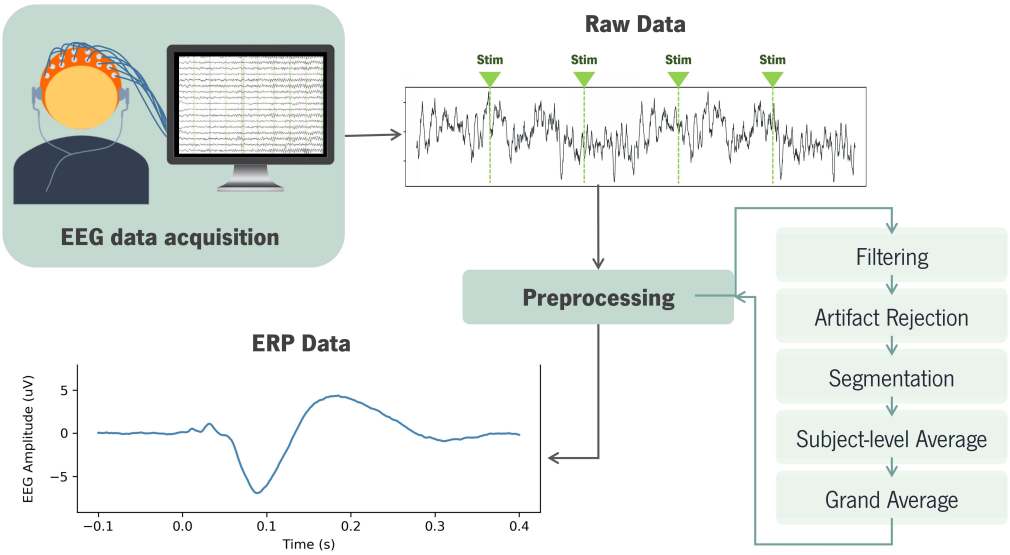


Figure 4: From EEG to ERP. Procedure steps to obtain a typical ERP waveform through the averaging process.

2. Offline data are then digitally processed. The signals are ideally referenced to an electrode or a set of electrodes that do not capture brain-related activity, that is more distant from the sources of the EEG signal, and/or do not cancel out the effects of nearby brain activity, such as earlobes or mastoids.

3. During the EEG acquisition, some channels can present some artifacts. In this case, bad channels can be removed from the analysis or replaced by interpolated values from the surrounding electrodes.

4. After that, the EEG signal is filtered. Filtering data can help remove high-frequency artifacts (low-pass filters) and low-frequency drifts (high-pass filters). Notch filters at 50 Hz can help attenuate electrical line noise. Despite some advantages in applying filters to the EEG data, when filters are poorly designed, they can introduce ripples in the time domain and cause signal distortion. A few recent studies carried out by different research groups tried to showed how inappropriate filters can lead to incorrect conclusions [136, 137, 138, 139]. Tanner *et al.* [139] demonstrated, through some simulation tests, that a high-pass filter with a maximum cutoff frequency of 0.1 Hz can be applied, ensuring that the signal is not distorted.

5. Then, continuous EEG data are segmented into epochs (single-trials), typically time-locked to the event of interest (e.g., stimulus onset), with a pre-stimulus baseline period.

6. Another important preprocessing step is artifact rejection. EEG data contain signal and noise. It is expected that appropriate preprocessing will attenuate the noise in the data. However, this may be difficult: in some cases, noise can clearly be dissociated from the signal, but in others, removing putative noise is also likely to remove some signal. It is necessary to find a trade-off between signal and noise [19]. The artifact detection can be performed using different methods: visual

inspection of single trials; semi-automatic or automatic inspection using advanced algorithms, such as the independent component analysis (ICA). After their identification, artefactual single-trials are rejected. Correlation methods are used in semi-automatic and automatic procedures to identify, select, and reject muscle artifacts, focal components (bad channels or muscle activity), focal trial activity, low signal-to-noise ratio (SNR), dipole fit residual variance, and EOG activity [140, 141].

7. The remaining single trials are normalized through the baseline correction method. The baseline correction involves subtracting the voltage measured during a prestimulus window (e.g., -500 ms to 0 ms relative to stimulus onset) from the entire waveform so that the waveform reflects the voltage relative to the average prestimulus voltage. It works as a DC offset removal [127].
8. Segments with high amplitude events (e.g., peaks above 100 µV or below -100 µV [127]) are rejected.
9. If the number of remaining EEG segments does not reach a substantial threshold value (e.g., if lower than 75% of the total number of segments), the subject/condition data are removed from the analysis. Otherwise, single trials for the same type of stimuli (same condition) are averaged for each electrode to produce ERP waveforms.

Once all subjects and all conditions are processed, the averaging process across subjects is conducted for each channel and for each condition. These are called grand averages. The final result is one segment per condition, per channel. ERP waveforms contain some peaks (components) whose variations are interpreted to reflect the speed (latency) or extent (amplitude) of processing. The ERP peak amplitude and latency can be measured for further statistical analysis.

2.2.2.1 ERP Components

The averaged ERP waveform is a sequence of peaks that vary in polarity, amplitude, and duration over time [17]. ERP components (peaks) are time-locked to a stimulus or action. The polarity (positive or negative) of ERP components does not have a straightforward relationship with neuronal discharges into thalamocortical cells or cell inhibition [142]. Instead, their polarity reflects the net electrical activity generated by underlying neural populations at a specific time point, influenced by a complex interplay of excitatory and inhibitory synaptic activity across neural networks [143, 142]. The evoked potentials are identified according to their polarity traditionally using the letters '*P*' and '*N*' in their nomenclature to indicate positive-going or negative-going peaks, respectively [127]. These letters are followed by a number indicating the position in the waveform or the latency of the peak in milliseconds. Examples of ERP components are the *N100*, which identifies the negative peak that typically occurs 100 ms after stimulus onset or the *P2* that refers to the second positive peak in the waveform [127]. Alternatively, the component names can include the nomenclature of the paradigm used to elicit it (e.g., no-go N2) or can correspond to functions that they assess (e.g., error-related negativity - ERN) [127]. The succession of peaks in ERP waveform reflects the flow of information in the brain over time, and the amplitude measures the strength of the components in terms of electrical energy [17, 127]. The scalp distribution of an ERP component

corresponds to the location over the scalp where its amplitude is highest. The brain can be roughly divided into two hemispheres, each with frontal, temporal, parietal, and occipital areas that are covered in the scalp by different EEG electrodes [144]. Most ERP components, including those in response to sound, have a wide scalp distribution [127].

While the exogenous sensory activity encompasses the transduction of stimulus in the sensory organ and the neural transmission along the sensory pathways to the brain, the endogenous activity manifests how the brain evaluates the stimulus, i.e., its cognitive processing [17]. However, exogenous ERP components are not solely driven by the physical properties of external stimuli. Individual differences in sensory processing can impact the ERP [145]. Moreover, top-down factors can also influence these components in certain conditions [146]. Therefore, while the term '*exogenous*' can be a helpful shorthand for describing specific ERP components elicited by sensory stimuli outside of the individual's attentional focus, it is important to recognize that these components are not purely exogenous [147]. Studying the neurodynamics of sound processing, from early to late procesing stages, can therefore clarify whether dysfunction occurs in sensory-driven and/or later cognitive operations underlying sound decoding in subjects diagnosed with neuropsychiatric diseases [17].

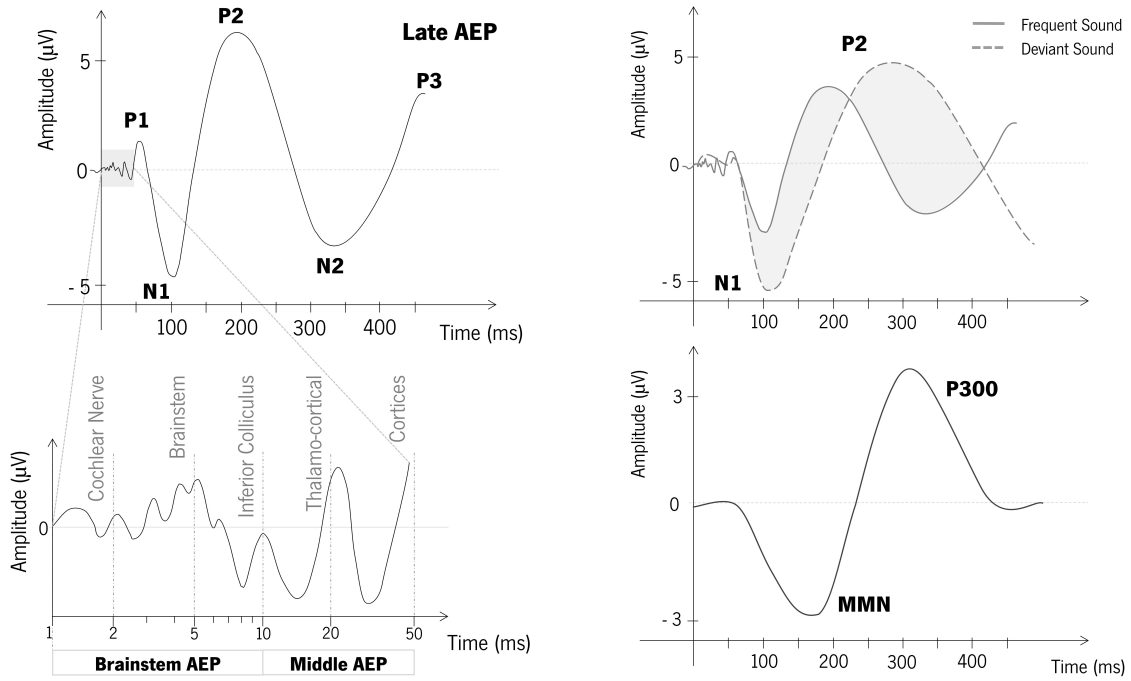
2.2.2.2 Auditory Evoked Potentials

Auditory ERP research has enhanced our knowledge of sound processing in the brain. Moreover, it has also led to a better comprehension of the neurophysiological mechanisms underlying several psychiatric disorders, such as SZ. Auditory evoked potentials can be obtained in response to various sounds and through experimental paradigms relying on passive or active listening [148]. The possibility of conducting passive auditory experiments is an advantage of the ERP technique, which can detect cortical brain responses to auditory stimuli even in the absence of behavioral responses [149].

Auditory evoked potentials may reflect the activity of the auditory system, including the auditory cortex [150]. These ERP responses can be divided into three groups (Figure 5a): the brain auditory evoked potentials, short latency components that occur in the first 10 ms after sound onset; the middle-latency components that emerge during 10-50 ms post-stimulus time interval; and the late cortical potentials with peak latencies beyond 50 ms [151].

Early auditory-related potentials are generated by the activation of brain stem nuclei of the auditory pathway during sound processing. These exogenous components of the ERP occur within the first 100 ms after stimulus onset. The longer latency (endogenous) components represent the cortical processing stages, in which the physical features of the auditory stimulus are less determinant [149].

P50 The P50 is a positive deflection that typically peaks at around 50 ms after sound onset, and it is the first middle latency potential, identified as the P1 peak in Figure 5a. This component is robustly elicited by any sounds [152]. It putatively reflects pre-attentive arousal caused by the appearance of a new auditory event [153]. Some studies have shown that the P50 is generated bilaterally in the primary auditory cortex and indexes the regulation and early cortical processing of incoming auditory information [154, 155, 156,



(a) (b)

Figure 5: a) Overview of brainstem, middle, and late auditory evoked potentials in the time domain; b) Above, ERP waveforms resulting from frequent and deviant sounds presentation; and below, the difference waveform showing the MMN and P300 components.

153, 157]. For instance, using a paradigm in which two identical sounds (S1 and S2) are presented with an interstimulus interval (ISI) of 500 ms, the amplitude of the P50 in response to the second sound decreases [158]. The P50 amplitude suppression, usually expressed as the S2/S1 ratio of the amplitudes of the two responses, is considered an index of sensory gating [158, 159]. Sensory gating refers to the brain's ability to filter out irrelevant sensory information and prioritize important or salient stimuli [160]. A reduced response to the repeated stimulus, known as sensory gating, indicates that the brain is filtering out the redundant information [161]. Abnormal P50 suppression have been reported in SZ patients, who show reduced suppression to the second sound of the pair (S2) when compared to healthy subjects [162, 163, 159]. Moreover, deficits in sensory gating have shown a relationship with auditory verbal hallucinations in psychosis [164]. Abnormalities in P50 suppression have not only been found in patients with SZ but also in bipolar affective [165], posttraumatic stress [166], or panic disorders [167].

N100 The N100 (N1 - Figure 5a) component is a relatively large, negative potential, which peaks between 80 and 120 ms after stimulus onset [148, 168]. The N100 is elicited by any perceptible auditory input regardless of task demands [168, 169]. Its amplitude is affected by various factors, such as ISI, stimulus features, and the subject's attention [169]. The auditory N100 component has been described as reflecting basic perceptual processing, being likely generated by auditory and frontal cortices [170, 171]. Anomalies in this component have been examined in various psychiatric disorders [172, 24, 173, 174, 175, 176]. Studies that used auditory oddball paradigms, or other listening tasks, have consistently found reductions in the amplitude of the auditory N100 evoked potential in SZ patients [24, 177, 159].

These robust alterations suggest deficits in mechanisms underlying the initial sensory processing and early automatic attention to sounds [175].

P200 The P200 (P2 - Figure 5a) is a positive deflection occurring approximately 200 ms after sound onset, being more prominent at central electrodes [148, 178]. Like the N100, the auditory P200 reliably occurs in response to sound presentation, and its latency and amplitude vary with aspects of selective attention or the acoustical properties of sounds [148, 179, 180]. P200 modulations are thought to reflect early stimulus categorization, the processing of task-relevant features of a stimulus, and the allocation of attention to those features [173, 159]. Little is known about the role of P200 alterations in SZ, and contradictory evidence has been found [181, 25]. However, the reduced P200 amplitude for repeated standard tones in SZ constitutes additional evidence for altered auditory processing in this disorder [25].

MMN MMN is described as *"a physiological mismatch process caused by a sensory input deviating from the memory trace ('template') formed by a frequent 'background' stimulus"* (p.324) [182]. This component is a negative deflection that typically occurs between 150-250 ms after stimulus onset [182, 183]. Oddball paradigms are usually used to elicit this component: a repetitive standard stimuli emission creates an acoustical regularity that is violated by deviant stimuli, which are seldom presented, while the participant is focused on a concurrent task [134]. The MMN is obtained from difference waves, as exemplified in Figure 5b. MMN reflects early processing at the level of the auditory sensory cortex, it is evoked pre-attentively, and indexes an echoic memory comparison of a stimulus to the neural trace of previous stimuli [183, 148, 184]. Its polarity is negative at frontal, central, and parietal scalp areas, with a larger peak amplitude over frontal electrode sites [183]. Reduced MMN amplitude has been consistently reported in SZ patients [185, 184] and at-risk subjects [186, 187]. Due to the consistency of these alterations, the MMN has been proposed as a biomarker of SZ [186, 46, 188].

P300 P300 is an index of endogenous cognitive processes, being a positive deflection that emerges approximately 300 ms after stimulus onset [189]. The P300 reflects attention operations, and it represents the activation of higher-level cortical (frontoparietal) circuits for determining the higher-level significance of the occurrence of an unexpected event [190]. This component is elicited during an oddball paradigm in which participants are instructed to pay attention to the auditory stimuli, therefore reflecting the engagement of attentional resources [191]. It can be obtained from difference waveforms, as illustrated in Figure 5b. P300 is disrupted in various neuropsychiatric disorders associated with cognitive impairment [192, 189]. Decreased P300 amplitude has been consistently reported in SZ, reflecting deficits in attention allocation to task-relevant sounds and updating of an internal model of the sensory environment [192, 193, 194].

2.2.3 Time-frequency Analysis

The EEG contains rhythmic activity, which reflects neural oscillations [195, 130]. Therefore, the direct measure of neural activity, through voltage fluctuations, is a great advantage of this technique. Moreover, EEG data are inherently multidimensional, encompassing measures of time, space, frequency, and power and phase (the magnitude and synchrony of frequency-band-specific activity, respectively) [134]. Brain activity is composed of multiple frequencies, which can be decomposed through signal processing methods. Traditionally, neural oscillations are clustered into frequency bands [19]: *delta* (δ - 1–4 Hz), *theta* (θ - 4–8 Hz), *alpha* (α - 8–12 Hz), *beta* (β - 13–30 Hz), and *gamma* (γ - >30 Hz), as exemplified in Figure 6.

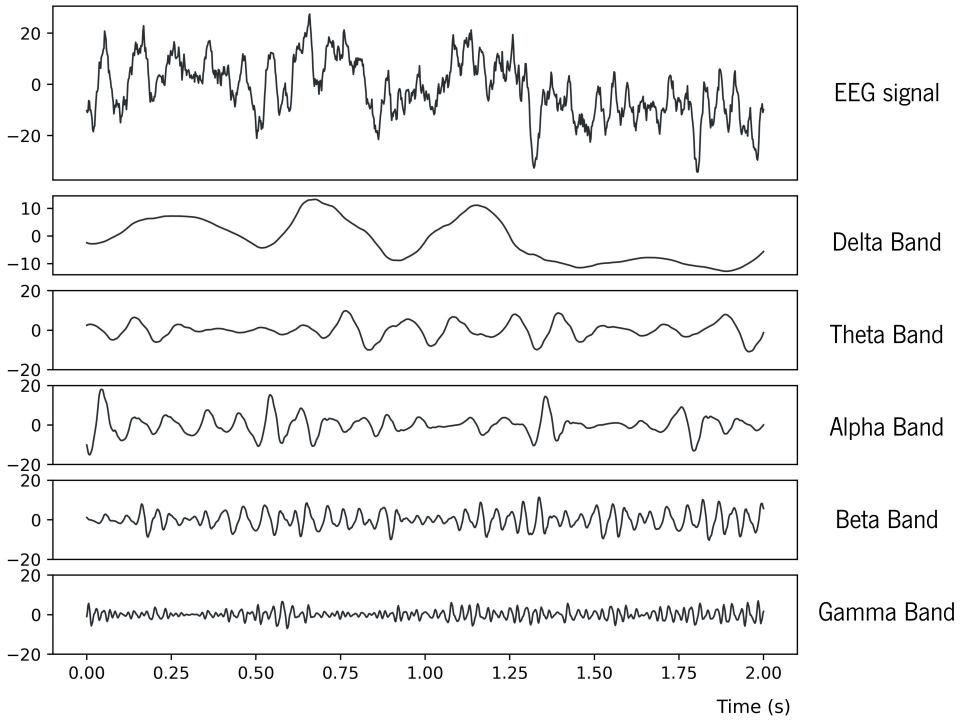


Figure 6: EEG signal decomposition into frequency bands with functional meaning.

Modulations of these frequency bands present functional significance:

- **delta**: these oscillations have been associated with cognitive processes demanding attention and are moderated by motivation [196]. Furthermore, some studies suggest that delta waves may act as inhibitors of interference, reducing the disruptive effects of irrelevant information on task performance. [196, 197].

- **theta**: this rhythmic component has been commonly associated with memory processes and cortico-hippocampal interactions [198, 199]. Studies have demonstrated that theta activity amplitude is increased during encoding and retention of information [198, 199, 197].

- **alpha**: this frequency band has been associated with relaxed awareness and sleepiness states, also emerging when the eyes are closed. These oscillations have also been associated with information

processing, inhibition control, and attention allocation [200, 197].

- **beta**: modulation of beta oscillations has been observed during active thinking and active attention. This rhythmic activity has been linked to the coordination among different representations in the cortex, motor tasks performance, decision-making processes, and focusing action-selection network functions [197, 201, 202].

- **gamma**: these faster oscillations have been suggested to reflect cortical activation. Gamma oscillations have been consistently linked with high-level cognitive functions such as attention, memory, and perception. In addition, it is believed that they are modulated by sensory input [197, 203, 204, 205].

The ERP methodology provides several advantages, including its computational simplicity and the fact that it is a widely used technique, which leads to abundant specialized literature on the method. Nonetheless, the analyses that can be conducted are limited: the ERPs have low SNR and low statistical power; the jittered and non-phase-locked activity is lost; and the link between ERPs and physiological mechanisms is less clear [19, 206].

In its turn, the time-frequency domain allows the analysis of EEG data as a multidimensional signal, having frequency as a prominent dimension. The method has some advantages. One is that many results can be interpreted regarding neurophysiological mechanisms of neural oscillations. The temporal correlations in different frequency ranges between neuronal responses are functionally relevant [207]. The strong evidence for a close relationship between neuronal oscillations and cognitive and behavioral responses makes time-frequency analysis an important method for SZ research. In addition, time-frequency analysis reveals more information than ERPs, especially in studies of cognitive processes. Nevertheless, this technique is not perfect. It presents two main limitations: the decrease of temporal precision (due to time-frequency decomposition) and the large number of complex analyses it enables [19].

A well-known time-frequency decomposition method is the Fast Fourier Transform (FFT), which enables a three-dimensional (frequency, power, and phase) representation of the time-domain signal. This technique is computationally fast and with no loss of information. Thus, the time series data can be reconstructed perfectly from the result of its FFT. However, the result of FFT is easily interpretable for stationary data since it converts the signals' original domain (such as time or space) into frequency domain [208]. This is critical since the brain activity is highly non-stationary. Therefore, the FFT has not proved itself to be the best method for time-frequency analysis of EEG data [19, 208]. Windowed functions have been considered an appropriate method to overcome the weak temporal localization [209]. However, a trade-off between temporal and spectral precision must be found. This is formalized in the Heisenberg uncertainty principle, which states that it is impossible to simultaneously have good precision both in time and in frequency [209, 134, 210]. According to this principle, when temporal resolution increases, the frequency resolution decreases, and vice versa.

A time-frequency dictionary is a set of elementary pieces of information (ϕ), called atoms. The dictionary can be represented by $D = \{\phi_n\}$ where n is the number of elements, with $n \in \mathbb{R}$. D is composed

of waveforms of unit norm $\|\phi_n\| = 1$, which are well localized in time (t) and in frequency (ω). The time localization u of ϕ_n and its spread $\sigma_{n,t}^2$, are defined by:

$$u = \int t|\phi_n(t)|^2 dt \tag{2.1}$$

$$\sigma_{n,t}^2 = \int |t-u|^2 |\phi_n(t)|^2 dt \tag{2.2}$$

Similarly, the frequency (ξ) localization and spread of the transform $\widehat{\phi}_n$ are defined by:

$$\xi = \frac{1}{2\pi} \int \omega |\widehat{\phi}_n(\omega)|^2 d\omega \tag{2.3}$$

$$\sigma_{n,\omega}^2 = \frac{1}{2\pi} \int |\omega-\xi|^2 |\widehat{\phi}_n(\omega)|^2 d\omega \tag{2.4}$$

Each piece of information can be represented in the time-frequency plane $\langle t, \omega \rangle$ by a rectangle having the position (u, ξ) and the size that depends on its spread in time σ_t^2 and in frequency σ_ω^2. This concept is illustrated in Figure 7.

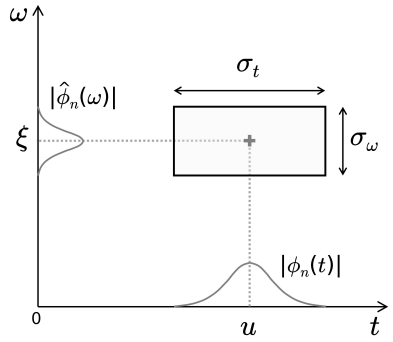


Figure 7: Heisenberg box representing an atom ϕ, adapted from [211].

The quantum information represented by the rectangle in Figure 7 has a minimum surface which limits the joint time-frequency resolution, defined by $\sigma_{n,t}\sigma_{n,\omega} \geq 1/2$.

A windowed Fourier transform decomposes signals over waveforms with the same time and frequency resolution. It is thus effective as long as the signal does not include structures having different time-frequency resolutions, some being very localized in time and others very localized in frequency.

Wavelets address this issue by changing the time and frequency resolution. A wavelet dictionary is constructed from a mother wavelet $\psi(t)$ of zero average, $\int_{-\infty}^{+\infty} \psi(t)dt = 0$, which is dilated with a scale parameter s, and translated by u:

$$D = \left\{ \psi_{s,u}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) \right\}, u \in \mathbb{R}, s > 0 \tag{2.5}$$

The continuous wavelet transform of a signal f at any scale s and position u is the projection of f on the corresponding wavelet atom:

$$Wf(s,u) = \langle f, \psi_{s,u} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt \tag{2.6}$$

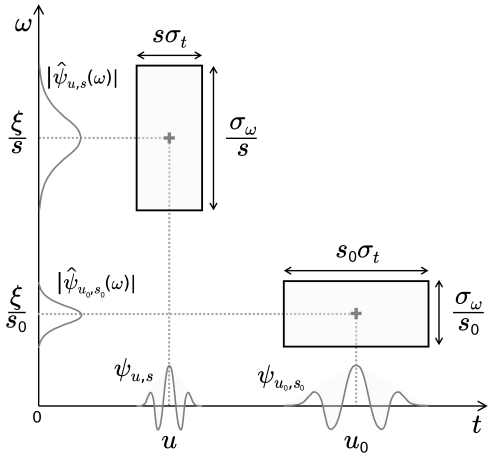


Figure 8: Heisenberg time-frequency boxes of two wavelets $\psi_{s,u}$ and ψ_{s_0,u_0}, adapted from [211].

2.2.3.1 Complex Morlet Wavelets

The non-stationarity of the signals is the primary motivation for time-frequency analysis. Magnitude and phase values can be extracted from signals using wavelet functions if they are complex-valued and well localized in time and frequency [212]. Complex Morlet wavelets have been regularly used to decompose non-stationary signals such as the EEG [210]. A Morlet wavelet is defined as a sine wave tapered by a Gaussian window [210, 212]. In turn, the complex Morlet wavelet is a complex sinusoid modulated by a Gaussian envelope as illustrated in Figure 9 [210, 212].

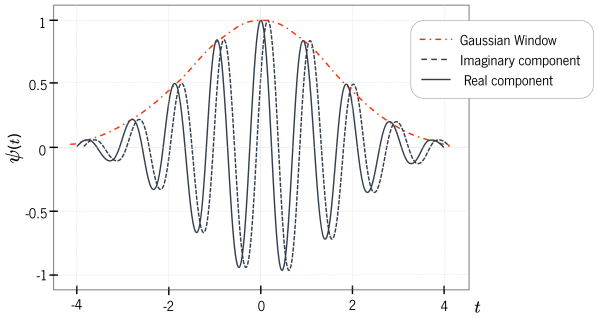


Figure 9: Complex Morlet wavelet.

The spectrotemporal representation of a signal is obtained by convolving the signal $x(t)$ with a family of wavelets $\tilde{\psi}_f(t)$ for a range of frequencies (f), that is defined by the following equation:

$$X(t,f) = x(t) * \tilde{\psi}_f(t) \tag{2.7}$$

The mother wavelet for the complex Morlet wavelet family is defined as:

$$\tilde{\psi}_{f_0}(t) = Ae^{-t^2/2\sigma_t{}^2}e^{i2\pi f_0 t} \tag{2.8}$$

where $e^{-t^2/2\sigma_t{}^2}$ is the Gaussian window with a temporal standard deviation σ_t, $e^{i2\pi f_0 t}$ is the complex oscillatory component with the center frequency f_0 Hz, and $A(\sigma_t)$ is a frequency band-specific scaling parameter. The A factor is defined by:

$$A = \frac{1}{\sqrt{\sigma_t\sqrt{\pi}}} \tag{2.9}$$

with the temporal standard deviation determined by:

$$\sigma_t = \frac{n_c}{2\pi f_0} \tag{2.10}$$

where n_c is the number of cycles of the wavelet defined for f_0 frequency inside the Gaussian window [19, 212]. The temporal standard deviation is inversely proportional to the spectral standard deviation (σ_f), as described previously by the uncertainty principle [209]. The spectral bandwidth can be calculated through:

$$\sigma_f = \frac{1}{2\pi\sigma_t} \tag{2.11}$$

A constant ratio c between the center frequency f_0 and the spectral bandwidth σ_f ($c = f_0/\sigma_f$) is a feature of the wavelet function, and therefore both temporal window and spectral bandwidth vary with f_0 [209]. Other factor m is considered when defining the number of cycles n_c of the wavelet function:

$$n_c = m\sigma_t f_0 \tag{2.12}$$

Considering this factor, the temporal window and the spectral bandwidth of the complex Morlet wavelet are scaled by m for any given frequency f_0, so that they are defined by $m\sigma_t$ and $m\sigma_f$, respectively [209]. In this way, the number of cycles can be further calculated as a function of the f_0 and values of c and m:

$$n_c = \frac{mc}{2\pi f_0} \tag{2.13}$$

The time-frequency decomposition using a complex Morlet wavelet allows the extraction and the quantification of the magnitude and phase characteristics of the signal, which shed light on the brain's dynamic responses to events.

- **Power**: it has been used to measure neural activity, and it reflects the amplitude of oscillations at specific frequency bands, independently of phase. Whereas the EEG rhythms are the result of a synchronized activity, the power changes can reveal underlying synchrony changes. The event-related oscillatory power comprises three types of sources, which may be measured: evoked, induced, and background power.

$$POW_{total} = POW_{evoked} + POW_{induced} + POW_{background} \tag{2.14}$$

The time-frequency decomposition of each segment/trial signal using the wavelet transform can be defined as:

$$S_n(t,f) = \begin{bmatrix} \psi_{f_1}(t) * x_n(t) \\ \dots \\ \psi_{f_F}(t) * x_n(t) \end{bmatrix} \tag{2.15}$$

S_n is the spectral estimate of the trial n at the frequency range $[1, F]$ of interest over time (t). $\psi_f(t)$ is the complex Morlet wavelet for frequency f and x_n is the n trial data. Being N the total number of trials, the total power is determined by:

$$POW_{total} = \frac{1}{N}\sum_{n=1}^{N} |S_n(t,f)|^2 \tag{2.16}$$

- *Evoked power*: it represents the event-related power phase-locked to the event onset. This measure reflects the synchronized oscillations across trials and is computed by the time-domain averaging of EEG epochs [209]. In order to calculate the evoked power, the time-frequency decomposition of the averaged signal is considered:

$$Se(t,f) = \begin{bmatrix} \psi_{f_1}(t) * x_N(t) \\ \dots \\ \psi_{f_F}(t) * x_N(t) \end{bmatrix} \tag{2.17}$$

  The Se is the spectral estimate of the evoked response $x_N(t)$ (averaged across N trials EEG data). Then, the evoked power can be determined by:

$$POW_{evoked} = |Se(t,f)|^2 \tag{2.18}$$

- *Induced power*: it is also event-related but non-phase-locked to stimulus onset. The underexploited analysis of induced power can be used to examine if the stimulus-related modulation of brain oscillations occurs independently of evoked responses [209]. These oscillations can appear from structural changes as well as from dynamic responses by inputs that vary over trials. The induced responses have had a role in feature-binding and top-down mechanisms of perceptual synthesis [213]. This will inform if the modulation of brain oscillations related to a stimulus has functional relevance to auditory processing, for example. The spectral estimate of the induced signal (Si_n) can be calculated through:

$$Si_n(t,f) = \begin{bmatrix} \psi_{f_1}(t) * (x_n(t) - x_N(t)) \\ \dots \\ \psi_{f_F}(t) * (x_n(t) - x_N(t)) \end{bmatrix} \tag{2.19}$$

The induced signal is obtained by subtracting the evoked signal $x_N(t)$ to each trial data $x_n(t)$. The induced power can be calculated based on the spectral estimate:

$$POW_{induced} = \frac{1}{N}\sum_{n=1}^{N}|Si_n(t,f)|^2 \tag{2.20}$$

- *Background power*: despite not being event-related, the background and spontaneous EEG oscillations contribute to the total power. This measure has become of great interest, and an increase in spontaneous oscillatory activity during the resting state has been associated with psychosis [214].

- **Phase-locking factor (PLF)**: it reflects the variance of EEG phase, or its consistency, across single trials independently of amplitude. This measure complements the total power measure and allows the study of phase synchrony in a specific frequency band. The computation of the PLF is based on the spectral estimate of each trial ($S_n(t,f)$) used to calculate the total power.

$$PLF = \frac{1}{N}\sum_{n=1}^{N}\frac{S_n(t,f)}{|S_n(t,f)|} \tag{2.21}$$

Figure 10 presents an example of the time-frequency decomposition of an event-related signal, where the color range corresponds to the magnitude of EEG activity in each frequency over time.

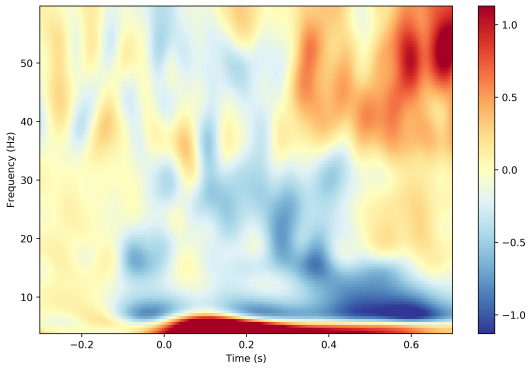


Figure 10: Time-frequency decomposition of EEG signal.

2.3 Machine Learning

Machine learning methods have been increasingly adopted to optimize classification tasks and to develop predictive models using neurobiological data. Through the combination of computational science and statistics, machine learning methods assess and automatically identify data patterns and use those recognized patterns to make predictions [215, 216]. Machine learning algorithms also learn from data like the human brain learns from experience. Data examples are represented as vectors of features [42],

which can be manually or automatically extracted and selected. They are expected to contain relevant information to effectively train the model as input to a learning phase [41].

The learning process can be classified as supervised, unsupervised, and semi-supervised, based on the data and analyses to be conducted [217]. Figure 11 outlines the different types of learning, as well as the problems addressed by each one.

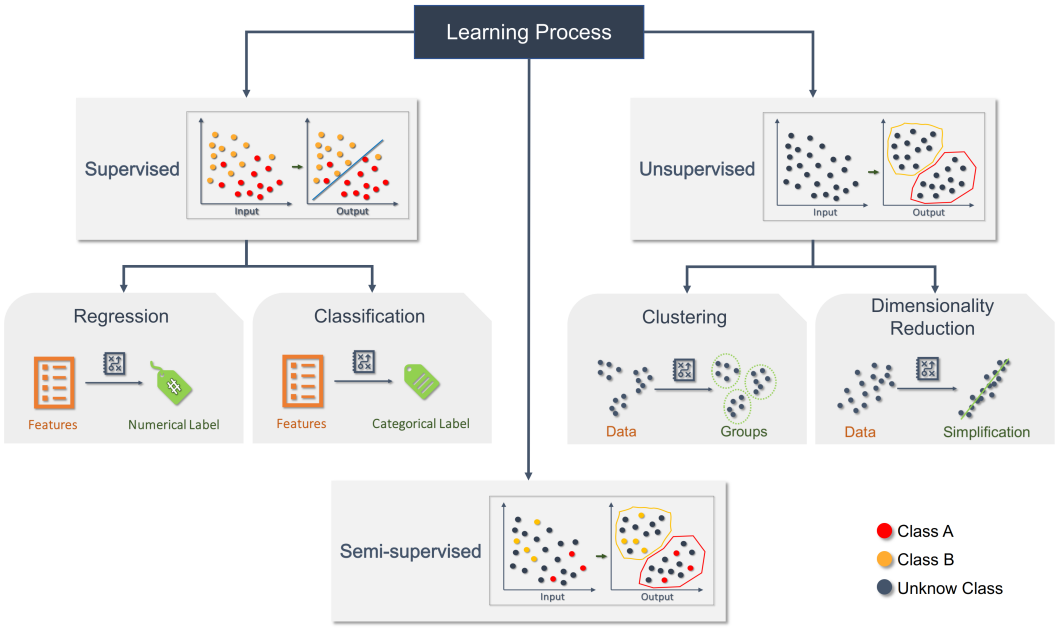


Figure 11: Different types of learning in machine learning, with problems addressed by each learning process.

Some common problems are built on classification and regression, addressed by supervised learning techniques (Figure 11). In supervised learning, a target vector (label) is provided for each data example expressing its category [42, 217]. During the learning phase, a relationship between a label and the input features is determined, allowing the model to accurately categorize or predict classes for new data, i.e., to generalize [41].

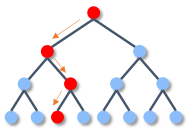
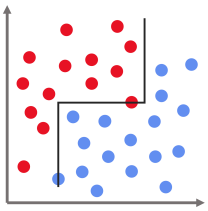
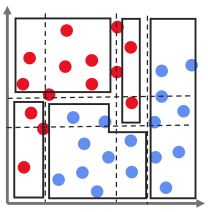
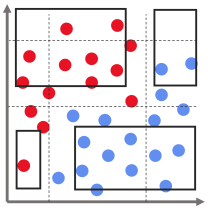
Contrarily, unsupervised learning can infer patterns from non-categorized data based on their underlying structure [215, 217]. Clustering techniques are widely employed as a common approach in unsupervised machine learning [218]. Clustering tasks involve the collection of data points based on the similarity and dissimilarity between them, determining the intrinsic grouping among the unlabeled data presented [219]. *K*-means clustering is an example of unsupervised machine learning algorithm [218]. Another unsupervised approach is the reduction of dimensionality, which allows decomplexing high-dimensional functions by reducing input variables or features [216]. Principal component analysis (PCA) is an example of an unsupervised method [42, 218, 216].

Semi-supervised methods aim to overcome the huge amount of data needed in supervised learning and the cost of labeled data acquisition. Semi-supervised algorithms are applied to a small amount of labeled data augmented by larger unlabeled datasets [220, 221].

2.3.1 Classification Algorithms

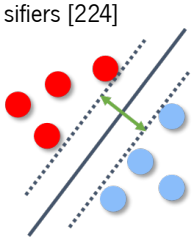
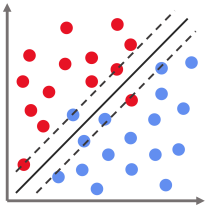
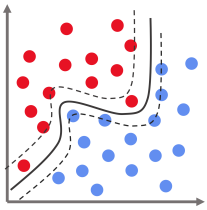
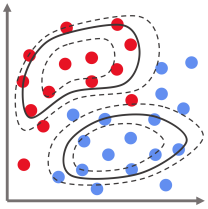
Classification aims to determine to which class the dependent variable belongs to, based on one or more independent variables [216]. Supervised classification is one of the tasks most frequently carried out. There is a wide variety of classification algorithms (or classifiers) [222, 223]. Table 1 presents some widely used ones, pointing out their advantages and disadvantages and illustrating the decision boundaries resulting from their application to a dataset.

Table 1: Description of advantages and disadvantages of different supervised machine learning algorithms. Schematization of decision boundaries resulting from the application of each algorithm.

Methods	Brief Description	Algorithms	Decision Boundaries
Logic-based learning (tree-based) [224] 	These methods involve the segmentation of the predictor space into a number of simple regions. The algorithm identifies ways to split a dataset based on different conditions. **Advantages**: simple and useful for interpretation; inexpensive to develop; excludes unimportant features. **Disadvantages**: decision boundary restricted to being parallel to attribute axes; decision tree models are often biased toward splits on features having a large number of levels.	**Decision Tree (DT)** and **Rule-based** methods	
		Random Forest (RF) - Ensemble method using bagging, boosting, or random feature subsetting to create multiple decision trees that are combined for predictions.	
		Extreme Gradient Boosting (XGBoost) - Ensemble method of boosting: like RF, it creates multiple copies of the training set, with a separate decision tree for each one; however, trees grow sequentially, using information from previously grown trees.	

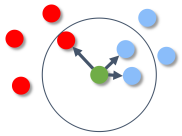
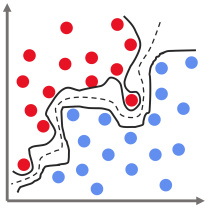
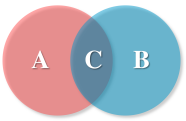
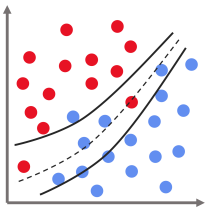
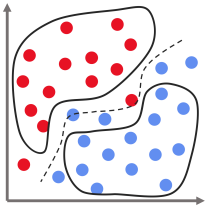
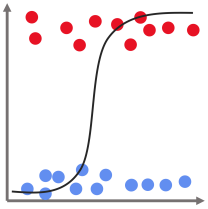
Continued on next page

Table 1 – *Continued from previous page*

Methods	Brief Description	Algorithms	Decision Boundaries
Support vector classifiers [224] 	The aim of these algorithms is to find a hyperplane in an N-dimensional space (N - the number of features) capable of classifying data points, maximizing the distance between points of both classes. **Advantages**: fairly robust against overfitting; effective in low-to-moderate dimensional spaces; relatively memory-efficient. **Disadvantages**: not suitable for large datasets; underperform when target classes are overlapping.	**Support Vector Classifier (SVC)** - This classifier results from using a linear kernel.	
		Support Vector Machine (SVM) – Polynomial - SVM is an extension of the support vector classifier that results from enlarging the feature space using other kernels, in this case the polynomial kernel; the decision boundary is much more flexible.	
		SVM – RBF - SVM using a radial-basis function (RBF) kernel.	

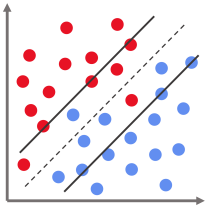
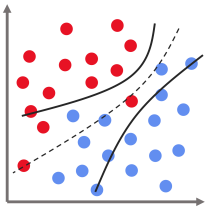
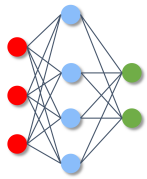
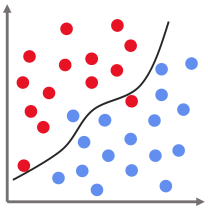
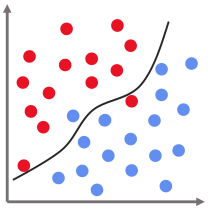
Continued on next page

Table 1 – *Continued from previous page*

Methods	Brief Description	Algorithms	Decision Boundaries
Instance-based learning [224] 	These methods produce a class label/prediction based on the similarity of the query to its nearest neighbor(s) in the training set, and do not create an abstraction from specific instances. **Advantages**: Ability to adapt to previously unseen data. **Disadvantages**: Memory-intensive; poor performance with high-dimensional data; require a meaningful distance function to calculate similarity.	**k-Nearest Neighbour (kNN)** - The classifications of the *k* most similar previous cases are used for predicting the classification of the current example.	
Statistics-based learning [224, 217] 	These methods utilize an explicit probability model that assigns probabilities to instances belonging to each class instead of solely providing a classification. **Advantages**: Fast to train; insensitive to irrelevant features. **Disadvantages**: Assume feature independence; assume distribution type for continuous features.	**Naïves Bayes (NB)** - Very simple algorithm based on conditional probability and counting	
		Gaussian process regression (GPR) - It assumes that the regression function is a nonlinear Gaussian process.	
		Logistic Regression Analysis (LRA) - Direct modeling using logistic function.	

Continued on next page

Table 1 – *Continued from previous page*

Methods	Brief Description	Algorithms	Decision Boundaries
		Linear Discriminant Analysis (LDA) - It retrieves the linear combination of features that best separate two or more classes.	
		Quadratic Discriminant Analysis (QDA) - Similar to LDA, but it assumes that each class has its own covariance matrix	
Neural Networks [216] 	These algorithms are inspired by the human brain. The artificial neurons are interconnected and communicate with each other. Each connection is weighted by previous learning events; with each new data input further learning takes place. **Advantages**: ability to learn from examples and to model non-linear and complex relationships; the model predicts based on unseen data. **Disadvantages**: Reliance on large training datasets; computationally expensive.	**Feed-forward neural network (FFNN)/ Multi-layer Perceptron (MLP)** - Series of logistic regression models stacked on top of each other, with the final layer being either another logistic regression or a linear regression model.	
		Probabilistic (PNN) - The class probability of a new input data is estimated, and Bayes' rule is then employed to allocate the class with the highest posterior probability to new input data.	

2.3.2 Deep Learning

The performance of machine learning methods depends significantly on the choice of data representation (or features) to which they are applied [202]. Although significant, this feature engineering is demanding and requires domain expertise [49]. Moreover, the need for features engineering brings out the weakness of classical machine learning algorithms, which cannot extract and organize discriminatory information from data [50]. Other methods have been developed in order to make the algorithms less dependent on the engineering of features, allowing, consequently, a breakthrough in the development of innovative applications [49, 225]. This is the case of representation learning, which not only grasps the mapping of data representation onto the output but also learns the representation itself [49, 42].

Deep learning is a typical approach for representation learning [42, 216]. This sophisticated technique can extract abstract, high-level features from raw data, allowing computers to build complex concepts from simpler representations [42]. Based on artificial neural networks (ANN), deep learning methods attempt to reproduce the neocortex's layers of neurons and their activity [226]. Deep learning depth relies on a multilayer model architecture, which allows executing more instructions and increases the algorithm power [42].

2.3.2.1 Deep Neural Networks

An ANN is composed of a large number of highly interconnected processing elements (neurons or nodes) that work together to solve specific, complex, undefined problems [227]. The structural unit of ANN is the perceptron, an abstraction of the biological neuron [42]. Figure 12 presents a typical biological neuron and the schematic of a perceptron.

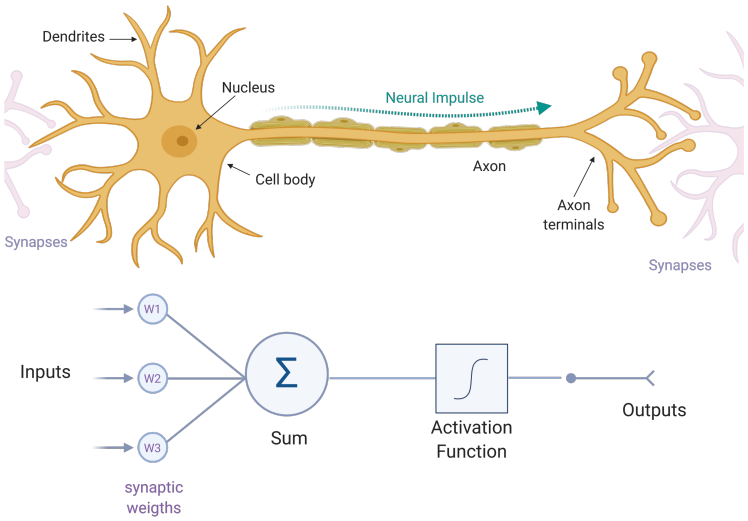


Figure 12: Schematic of a biological neuron and a perceptron (Created with BioRender.com).

In biological neurons, external impulses or impulses from other neurons are received through synapses on the dendrites. These impulses are screened and merged. When a specific threshold is reached, the neuron activates, and a response signal is emitted through the axon. The signal is sent to other synapses

and might activate other neurons. Synapse efficiency (strength) is influenced by chemical or electrical factors [228, 229]. This basic process is mimicked by the perceptron shown in Figure 12. The input signals are transmitted to the neuron body, being weighted and summed. These signals are then transformed, by passing through the activation function, into the output signal [228]. The propagation of the signal is determined by the connections between the neurons and their associated weights, which compares to the synaptic strengths in the biological system [228, 229]. Thus, the artificial neuron takes a vector of input features $x_i = \{x_1, x_2, ..., x_n\}$, and each is multiplied by a specific weight $w_1, w_2, ..., w_n$. The weighted inputs are summed together, and a constant value called bias (b) is added to them to produce the net input of the neuron. The activation function g is attached to the perceptron and determines whether it should be activated based on whether the perceptron's input is relevant for the model's prediction [42]. The output a is defined by:

$$a = g\left(\sum_{i=1}^{n} x_i w_i + b\right) \tag{2.22}$$

where n is the total number of features. This output is then transmitted to other artificial neurons.

A basic neural network is the grouping of several perceptrons that operate together. A single-layer neural network is obtained by placing perceptrons side by side [230]. When several one-layer networks are stacked, a multi-layer network is formed, usually called a multilayer perceptron (MLP) [231, 42]. What distinguishes a deep learning algorithm from a basic neural network is the number of node layers, i.e., the depth of the neural network [218]. Deep learning algorithms use ANNs, multiple-layer neural networks based on fully connected nodes (artificial neurons) [232]. Figure 13 shows an example of a three hidden layers MLP.

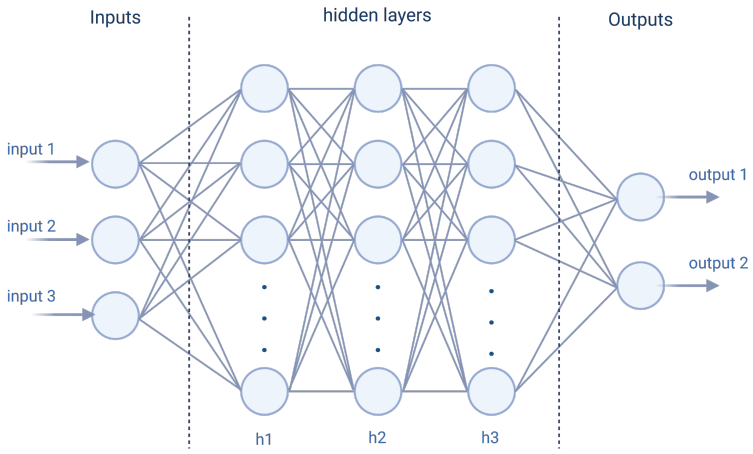


Figure 13: Schematic of a MLP with 3 hidden layers (Created with BioRender.com).

The MLP is composed of input and output layers. In addition to these two layers, it usually has one or more hidden layers of neurons [233]. This network is also designated as feed-forward neural network (FFNN), since the information flows from the input layer, through the hidden layers, to the output layer [42]. A FFNN aims to learn a non-linear function f that maps input data observations into a desirable output

$z = f(x;\theta)$ such that $z = y$ is intended, where $x \in \Re^n$ is the input, θ the learned parameters that result in the best function approximation, and y the target/desired output [42]. A sequence of layers describes this type of network. A set of neurons characterizes each layer f_l, and each neuron is a linear composition of values obtained from the previous layer or input. The activations from the previous layer serve as input for the next layer until the values reach the output [42]. Thus, FFNN can be seen as a chain of functions, which can be defined as:

$$F(x) = f_L(f_{L-1}(...(f_1(x)))) \tag{2.23}$$

where L is the total number of layers in the network, representing the neural network depth. When a FFNN includes feedback connections, it is called a recurrent neural network (RNN) [42].

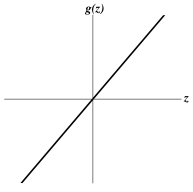
Hidden Layers Hidden layers represent an inner block of computation, without direct interaction with any outside variable [233, 42]. These hidden layers extract increasingly abstract features from the data [42, 232]. All the layers are fully connected, i.e., a neuron in one layer is connected to all neurons in the previous and the next layers [42]. The output activation of a hidden layer can be defined as:

$$h_r^l = g\left(\sum_q^Q h_q^{l-1} * w_{qr}^l + b_q^l\right) \tag{2.24}$$

where l indexes the hidden layer, g is the activation function, and h_r^l is the r^{th} feature of the layer l. The w_{qr}^l denotes the connecting weight between the r^{th} feature of the layer l and q^{th} feature of the previous layer ($l-1$), while b_q^l is the offset of the r^{th} feature of the layer l. In equation 2.24, Q represents the total number of features of the layer $l-1$.

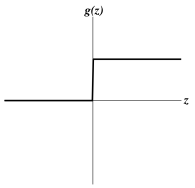
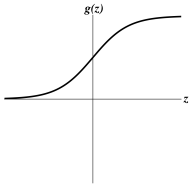
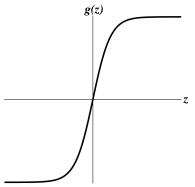
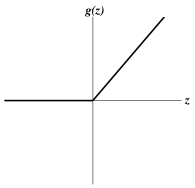
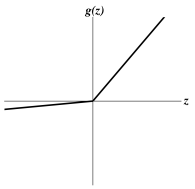
Activation Function The nonlinear activation functions are desirable for multilayer networks to detect nonlinear features [216]. Also, these functions define how each neuron in a network gets activated, propagating the output of one layer to the next layer [234]. Some of the commonly used activation functions are presented in Table 2.

Table 2: Commonly used activation functions.

Function	Equation	Graphical Representation
Linear	$g(z) = z$	

Continued on next page

Table 2 – *Continued from previous page*

Function	Equation	Graphical Representation
Step	$g(z) = max\{0, 1\}$	
Sigmoid	$g(z) = \frac{1}{1+e^{-z}}$	
Hiperbolic Tangent	$g(z) = tanh\{z\}$	
Rectified Linear Unit (ReLU)	$g(z) = max\{0, z\}$	
Leaky ReLU	$g(z) = max\{\alpha z, z\}$	

The rectified linear units (ReLU) activation function effectively achieves a non-linear transformation of the data. One major advantage of ReLU compared to other functions, such as the sigmoid or the hyperbolic tangent, is the reduced likelihood of the gradient to vanish, which arises when $z > 0$ [235]. This is particularly relevant because FFNN are trained by looking for a zero gradient (as explained next in Section 2.3.2.2). For those z values, the gradient has a constant value, resulting in a faster learning process. In contrast, the gradient of the sigmoid function, for example, becomes increasingly small as the z value increases. Other benefit of ReLU is sparsity when $z \leq 0$ [42]. Since the ReLU of all negative inputs is zero, any given unit may not be activated. This means that some of the neurons of the model output zero, which does not affect the next layers [235]. Conversely, sigmoids are always likely to generate some non-zero value resulting in dense representations. Sparsity leads to concise models, which usually have better predictive power and less overfitting. In sparse networks, neurons are more likely to be actually dealing with meaningful aspects of the problem [42]. Moreover, ReLU is less computationally expensive

than other functions because it involves simpler mathematical operations [236].

The softmax is an activation function used for classification, usually used in the output layer. The output features $z(i)$ are transformed by the softmax function whose output is the probability distribution over labels, allowing the categorization of the input [42]. The softmax function can be defined as:

$$p(i) = softmax(z(i)) = \frac{exp(z(i))}{\sum_k exp(z_k(i))} \tag{2.25}$$

where k is number of classes, and $p(i)$ the probability which the proposed model assigns to the i output label taking the input value $z(i)$.

2.3.2.2 Gradient-based learning

Deep learning is about solving huge and challenging optimization problems. A neural network is a complex function consisting of millions of parameters representing a mathematical solution to a problem. Training a neural network mainly means minimizing a loss function. The value of this loss function measures the network's performance on a given dataset [42].

The gradient descent is a first-order [1] iterative optimization algorithm used for finding the minimum of the loss function. Optimization starts by randomly initializing the network parameters (weights and biases). These parameters are then gradually updated. Repeated steps are performed in the opposite direction of the gradient of the function at a point (the direction of steepest descent) to find the parameters configuration by minimizing a loss function, which in most cases is non-convex in nature (many local minima instead of a single global minimum) (Figure 14). The convergence to a minimum (local or global) will depend on the initial parameters and the learning rate, which scales the step size of the gradient descent.

Whereas in traditional approaches, the representation and classification are separately optimized, the optimization of both phases in deep learning is performed through gradient descent in conjunction with backpropagation that determine how a machine should change its internal parameters [50]. In practice, for MLPs to work as deep neural networks, some ingredients such as backpropagation and hyperparameters tuning are needed [50].

2.3.2.3 Backpropagation

Backpropagation is the essence of neural network training. The error backpropagation algorithm is bidirectional (forward and backward directions) [50, 237]. The algorithm is used to effectively train a neural network through chain rule. The training data are presented to the network during the forward process and classified. The backward pass consists of a recursive updating of the weights in all layers based on the computed errors [237]. The output from forward propagation corresponds to the predicted value. The loss function calculates the error by comparing the predicted with the ground-truth value. The

[1] first-order optimization considers only the first derivative of the cost function

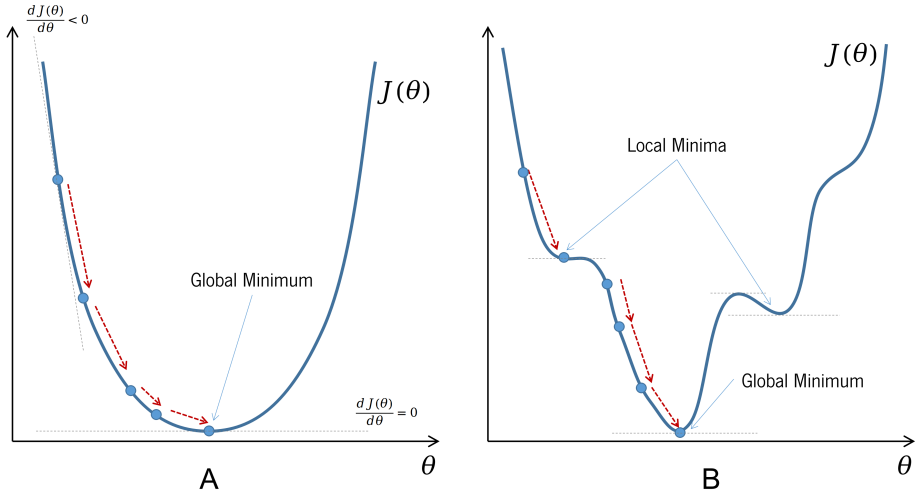


Figure 14: Gradient descent represented for *A)* a convex loss function with a global minimum, and for *B)* a highly non-convex loss function, with many local minima.

derivative of the error is then calculated concerning every weight in the neural network. The derivatives of the last layer are computed and used to determine the gradients of the second last layer. This process is repeated until the first layer. Then the gradient is subtracted from the weight value to reduce the error. Thus, backpropagation attempts to minimize the loss function, i.e., the difference between the true value and the actual model prediction. This technique makes the deep learning model reliable by increasing its generalization [50, 237].

2.3.2.4 Hyperparameters

Settings of multilayer structures, such as the number of layers, initial weights, activation function, among others, can control the networks' behavior and learning process. This challenging hyperparameter tuning is required to find the ideal model architecture, and hence to improve the model's performance [238, 239].

Initialization Initializing parameters, such as weights, is an essential initial step in neural network building. The initial point from which the iterative training begins can determine how fast learning converges to a minimum of the generalization error, with high or low cost [42]. The random choice from a Gaussian or uniform distribution has been a common strategy to initialize parameters with appropriate values [240]. The Xavier Glorot uniform [241] initializer ensures the zero-mean. It keeps the variance of activations across every layer, preventing layer activation outputs from exploding to a high value or vanishing to zero during the forward pass through the network [241, 240].

Loss Function The loss function measures how close the model's predictions are to the ground-truth label. During the backpropagation step, the loss function value must be minimized to improve the model [42]. In the classification task, the cross-entropy of the outputs can be calculated, concerning

the true labels, generating a loss, which effectively evaluates the model's performance on training data. Equation 2.26 defines the cross-entropy formula [216].

$$H(p, y) = -\sum_{k=1}^{K} y_k \log p_k \tag{2.26}$$

being K the number of classes, p the probabilistic prediction (after softmax), and y the correspondent true label.

The negative loss likelihood (NLL) function can be used to determine the loss function (or cost function) [216]. The NLL function aims to minimize the distance between input and output, and it can be described as:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} H(p(i), y(i)) \tag{2.27}$$

where H is the cross-entropy function, with $p(i)$ as the predicted output and $y(i)$ as the ground-truth label of the trial i, being m the total number of trials.

Model Optimization As mentioned before, the optimization algorithms are used to train the neural network by finding the parameters θ that minimize the cost function $J(\theta)$ [42]. The value $z(i)$ is obtained during the forward propagation step and draws on the θ parameters of the network. The optimization of the function $J(\theta)$ is achieved by updating the values of trainable parameters [42]. An iterative procedure applies changes to the parameters at each iteration of the algorithm. Being θ_t the parameters at the t iteration, the update rule can be defined as:

$$\theta_{t+1} = \theta_t + \Delta\theta_t \tag{2.28}$$

With the gradient descent algorithms, such as the stochastic gradient descent (SGD), the optimization of $J(\theta)$ follows the steepest descent direction given by the negative gradient g_t. The derivative described by equation 2.29 can be obtained by applying gradient descent approaches:

$$\Delta\theta_t = -\alpha g_t \tag{2.29}$$

where g_t is the gradient descent of parameters at iteration t, and α the learning rate. The Adaptive Moment Estimation (Adam) is a variant of SGD designed for high-dimensional parameters and suitable for optimizing CNNs many parameters. This algorithm computes adaptive learning rates for each parameter [242].

Learning Rate The learning rate α is a hyperparameter that defines the step size in the direction of the negative gradient, i.e., that controls how adjusted the weights should be each time the model is updated to converge to an optimal solution [216]. The learning rate selection is challenging, often performed by trial and error or by monitoring the loss curve as a function of time [42].

Regularization A trade-off between bias and variance is expected to reduce the model generalization error [216]. Bias measures how far the predicted values are from the real ones, whereas variance measures the variability of model prediction for a given data point, showing the spread of the data [42]. When the variance increases, the model sensitivity also increases, and the algorithm is more likely to capture random noise rather than desired data (overfitting). Therefore, the model becomes more flexible, performing better in training than in test data. Overfitting reflects the model's poor generalization from the training to the new dataset [243, 244]. Aside from training with more data, overfitting can be avoided by regularization. Regularization is a technique that modifies the learning algorithm to reduce its generalization error [42, 245]. Dropout is an excellent regularizer and easily implementable. This method prevents the model from overfitting by temporarily removing units from the network with a specific probability in each training step [246].

Other regularization approaches, such as L^2 regularization, intend to limit the models' capacity by adding a parameter norm penalty to the cost function [42]. The L^2 term, defined as $\|\theta\|^2$, only penalizes the weights leaving the biases unregularized, known as weight decay. The regularized cost function is then defined by:

$$\tilde{J}(\theta) = J(\theta) + \lambda.\|\theta\|^2 \tag{2.30}$$

where $\|\bullet\|^2$ is the L^2 norm, and $\lambda \in [0, \infty]$ the penalty factor that weights the relative contribution of the regularization.

2.3.2.5 Batch Normalization

Instead of training all input data at a time, the batch mode allows using fewer samples, which will be propagated through the network. The batch size b is a hyperparameter that defines the number of training examples in one forward/backward pass. This method enables networks to train faster, requiring less memory.

Normalizing the input of a neural network improves the convergence of its training process and enhances its overall performance by mitigating the impact of different scales [247]. Extending this insight, batch normalization limits the covariate shift (change of input values distribution), ensuring the normalization of activations in each layer. This technique significantly impacts the hidden layers, where the optimal parameters are learned from unnormalized activation distributions from previous layers [248].

Considering z as the activation, the batch B of size b can be described as $B = \{z_1, ..., z_b\}$. The mean and variance of activations over the batch can be determined through equations 2.31 and 2.32, respectively.

$$\mu_B = \frac{1}{b}\sum_{\rho=1}^{b} z_\rho \tag{2.31}$$

$$\sigma_B^2 = \frac{1}{b}\sum_{\rho=1}^{b}(z_\rho - \mu_B)^2 \tag{2.32}$$

Using these values, it is possible to normalize the activations through the equation:

$$z_{norm}^{(\rho)} = \frac{z^{(\rho)} - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{2.33}$$

being z_{norm} the normalized activations and ε a very small number to prevent the chance of a divide by zero error.

This normalization allows z to have zero mean and unit variance. However, this could perform poorly for some activation functions (e.g., sigmoid function). Thus, batch normalization enables the model to learn the optimal distribution by scaling the normalized values by γ and shifting by β.

$$\tilde{z}^{(\rho)} = \gamma z_{norm}^{(\rho)} + \beta \tag{2.34}$$

where $\tilde{z}$ are the transformed activations values [248].

Different variants of deep architectures have been developed [232]. Recent deep learning architectures have relied on three classes of ANN: MLP, RNN, and convolutional neural network (CNN). MLPs learn a representation from all data input, whereas RNNs process sequential data introducing the concept of memory [42, 232]. In turn, CNN exploits the spatial structure of the data by computing filter-selective feature maps [42]. The current work is based on CNN, to which special emphasis will be given.

2.3.2.6 Convolutional Neural Networks

CNN represents the most relevant component in the architecture of EEG classification solutions [225]. Biologically inspired in the organization of the primary visual cortex, the CNN represents a successful application of knowledge on brain function to machine learning implementations [249, 250].

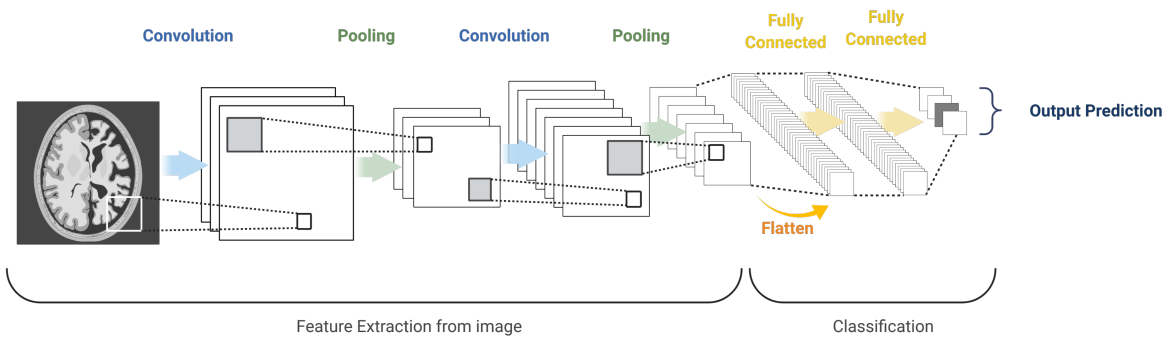


Figure 15: Diagram illustrating a typical architecture of a CNN (Created with BioRender.com).

CNN architectures vary in number and type of layers implemented, depending on the application [251, 42]. CNN is specially designed to automatically learn spatial hierarchies of features using different types of neuron layers. A typical architecture (Figure 15) consists of repetitions of several convolution layers and

a stacked pooling layer, followed by one or more fully connected layers at the end of the network [252, 42].

Convolutional layers The CNN structure is inspired by the complex arrangement of simple and complex cells of the visual cortex [253]. In the biological vision system, each neuron responds to a stimulus in a restricted region, and it is connected to other neurons, covering the entire visual field [42]. Similarly, each neuron in CNN, represented by convolutional layers, processes data in its receptive field. These layers compute convolutions of their inputs with linear filters (suggested for modeling neural responses in the visual cortex) [42]. The coefficients of each kernel are learned as weights in a neural network. The dot product between a restricted portion of the input and the kernel (set of learnable parameters) is computed [254]. Then, the kernel is convolved across all input widths and heights, producing the output. The output units are organized as feature maps, and each unit is connected to the input through a set of weights [50]. This process is illustrated in Figure 16.

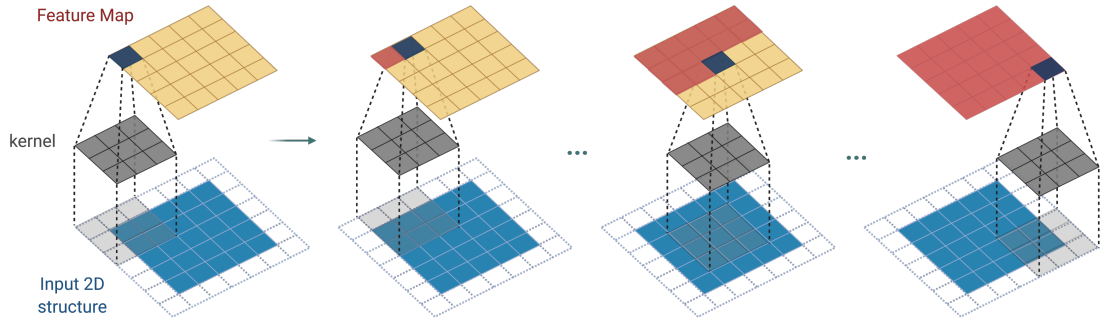


Figure 16: Outline of a convolutional layer (Created with BioRender.com).

A nonlinear (activation) function is commonly applied to each output unit after each convolution. All units in a feature map share the same weights, exploring statistical correlations. The weight sharing in a convolutional layer reduces the number of parameters in the network, which makes CNNs faster to train and less prone to overfitting [255, 50, 42]. The role of the convolutional layer is to detect local combinations of features from the previous layer, and to automatically extract the deep and abstract information of input signals [50]. CNN features are defined as spatial 2D maps. The CNN layers are hierarchically arranged, allowing the detection of more complex patterns after the detection of simpler features [42, 254].

Pooling Layers The function of pooling layers is to reduce the spatial size of the representation, thereby reducing the network parameters and controlling overfitting. This is achieved by merging semantically similar features into one [50]. For instance, the max-pooling computes the maximum output within a local rectangular patch of units in the feature map [42].

Fully connected Layers Fully connected (FC) layers are commonly used at the end of a CNN architecture. Every neuron in an FC layer has a connection to all activations in the previous layer. Thus, it can

aggregate information from high-level features extracted in preceding layers and determine which features strongly correlate to a particular class, generating the classification result [254].

2.3.2.7 Generalization, Accuracy, and Overfitting

A neural network generalizes when it correctly classifies vectors that were not used in the training dataset. The generalizability is measured by the accuracy of these classifications [256]. A neural network computes a complex, non-linear function of its inputs. If the trained model fits very well for a large number of inputs, then most likely the training process has captured an underlying functional relationship between inputs and classification. In that case, generalization to new inputs is expected to be good [256]. The generalization error measures the accuracy of the model classifying previously unseen data. The most straightforward technique for estimating this error rate is to divide the training dataset into three subsets: the training set, the validation set, and the testing set. The training set is used to train the model, adjusting its parameters based on the performance on the validation set. The validation set helps in selecting the best hyperparameters and monitoring the model's performance during training [256]. Finally, the testing set is reserved for evaluating the model's performance after training is complete. This testing set should be completely independent of the training and validation sets. The purpose of the testing set is to provide an unbiased assessment of the model's generalization ability [42].

Another method for estimating generalization accuracy is called cross-validation. It divides the vectors available for training into k disjoint subsets named '*folds*'. One of the folds is used as the validation dataset, while $k-1$ combined folds as the training set. The training and validation sets must cross over in successive k rounds to validate each vector. The generalization error is estimated by the average of the errors computed for each validation set [256].

When the model is too intricate or trains too long on sample data, it can learn the noise or irrelevant information within the data set. When the model memorizes the noise and fits too closely to the training set, the model becomes "overfitted" and cannot generalize properly to new data. Low error rates and high variance are good indicators of overfitting. Conversely, underfitting occurs when the model does not train sufficiently or when the input variables are not significant enough to determine a meaningful relationship between the input and output variables [216].

2.4 Summary

SZ is a difficult-to-diagnose psychotic disorder characterized by hallucinations, delusions, and disturbances in thought, perception, and behavior. EEG is currently the gold standard for specifying the dynamics of sensory and cognitive processing alterations in SZ. In addition to event-related potentials, the frequency domain analysis of EEG data can help uncover which neuronal processes are associated with disease symptoms, such as auditory hallucinations. The low cost and high temporal resolution of EEG make it an ideal tool to examine the neural mechanisms underpinning the characteristic symptoms of SZ. By taking advantage of advances in machine learning, especially in deep learning, it could be used

to facilitate diagnosis, prognosis, and tracking of responses to treatment. Machine Learning methods are used to learn patterns from data, which involves identifying discriminative features. This task is challenging due to EEG signals' non-linear, non-stationary, and time-varying characteristics. Deep learning unburdens the efforts for manual feature engineering through end-to-end pattern recognition, potentially presenting a promising solution for EEG signal classification, with relevant clinical implications.

Chapter 3

Advanced EEG-based learning approaches to predict schizophrenia: a review

This chapter presents an original integrative synthesis of recent machine learning-based methods for SZ classification using EEG data, discussing their potentialities and limitations. This review can be considered as a starting point for the development of more robust and reliable machine learning-based approaches for accurate SZ recognition and prediction using EEG information.

Contents

3.1 Introduction **44**
3.2 Search strategy and inclusion criteria **45**
3.3 Schizophrenia Diagnosis **45**
3.4 Neurobiological changes in schizophrenia **46**
3.5 Electrophysiological Biomarkers of Schizophrenia **47**
3.6 Machine Learning for Schizophrenia Detection and Prediction **48**
3.7 Promises and Pitfalls **61**
3.8 Summary **64**

This chapter is based on the publication:

Barros, C., Silva, C. A., & Pinheiro, A. P. (2021). Advanced EEG-based learning approaches to predict schizophrenia: Promises and pitfalls. *Artificial Intelligence in Medicine*, 102039. doi.org/10.1016/j.artmed.2021.102039

3.1 Introduction

SZ is a chronic and complex neuropsychiatric disorder, with an estimated prevalence of 1% worldwide [257]. Its onset typically occurs by the end of adolescence and the beginning of early adulthood [8]. SZ symptoms are heterogeneous and associated with functional impairments and reduced quality of life [4], being one of the top 25 leading causes of disability worldwide [258, 3]. SZ is characterized by the presence of positive (e.g., hallucinations and delusions) and negative symptoms (e.g., blunted affect), as well as cognitive deficits [9]. Functional impairment has been consistently correlated with changes in cognitive functions [9], including attention, memory, language and executive functions [259, 260], which are already observed before illness onset [10, 261]. Psychotic symptoms such as hallucinations and delusions are not exclusive of psychotic disorders [262] but might be present in other psychiatric (e.g., bipolar disorder [263]) or neurological (e.g., temporal lobe epilepsy [264]) disorders. Additionally, psychotic-like experiences have been reported in healthy subjects without a need for care [265], supporting the notion that psychotic symptoms exist on a continuum in the general population [266, 267, 268, 269, 270]. Evidence for a psychosis continuum challenges the binary categorization of psychotic disorders, with implications for clinical diagnosis. A dimensional approach to psychosis may provide critical insights on risk factors for psychosis, disease progression and prognosis [268]. In particular, identification of risk factors is imperative to prevent or delay disease onset [271]. For example, early interventions could reduce the negative impact of SZ [272]. However, the lack of reliable methods to objectively define the onset of psychiatric disorders hampers the precise diagnosis of SZ [273]. Several research teams worldwide are committed to shed light on the neurobiological mechanisms underpinning the clinical manifestations of SZ, which may lead to substantial improvements in diagnosis and prognosis [16]. Advances in neuroscience methodologies have fostered the study of this complex disorder. Studies using fMRI, MEG or EEG have identified changes in brain structure and function associated with sensory and cognitive abnormalities in SZ [274, 275, 169]. Given its high temporal resolution (in the order of milliseconds), the EEG technique can capture the brain dynamics underpinning sensory, cognitive, affective and/or motor processes in response to a stimulus [19, 18]. Furthermore, it is an affordable method to directly measure cortical brain activity [19].

Machine learning applied to EEG data might represent a promising contribution to the accurate prediction of intervention response and disease trajectories [37]. Machine learning methods could enhance our understanding of the brain mechanisms underlying SZ by contributing to three key domains: diagnosis and detection; prognosis and onset prediction; biomarker identification.

This chapter provides a systematic review of studies that relied on advanced learning methods to detect SZ based on EEG signals. Specifically, it aims to provide an overview of proposed neurophysiological biomarkers in SZ research, as well as a critical analysis of machine learning methods applied to EEG data for SZ classification since 2016, further specifying deep learning methods used for the same purpose. Moreover, it aims to identify and evaluate the potentialities and limitations of both types of learning methods.

3.2 Search strategy and inclusion criteria

A literature search of studies that applied machine learning methods on EEG data for SZ classification, published from January 2016 until April 2020, was conducted using *SCOPUS*, *PubMed*, *Google Scholar*, *IEEE Xplore*, and *ArXiv* databases, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (*PRISMA*) Statement [276]. Advanced search terms included the following keywords: *EEG* or *electroencephalography*, *schizophrenia*, and *machine learning*. Studies were included if (1) they aimed to discriminate SZ and/or at-risk patients from healthy control subjects or from patients with other psychiatric disorders; (2) they used machine learning techniques applied to EEG data. Only studies published after 2016 were considered to allow a clearer view on the trends and improvements observed in this research field in the last 5 years. A total of 25 studies were included in the current review.

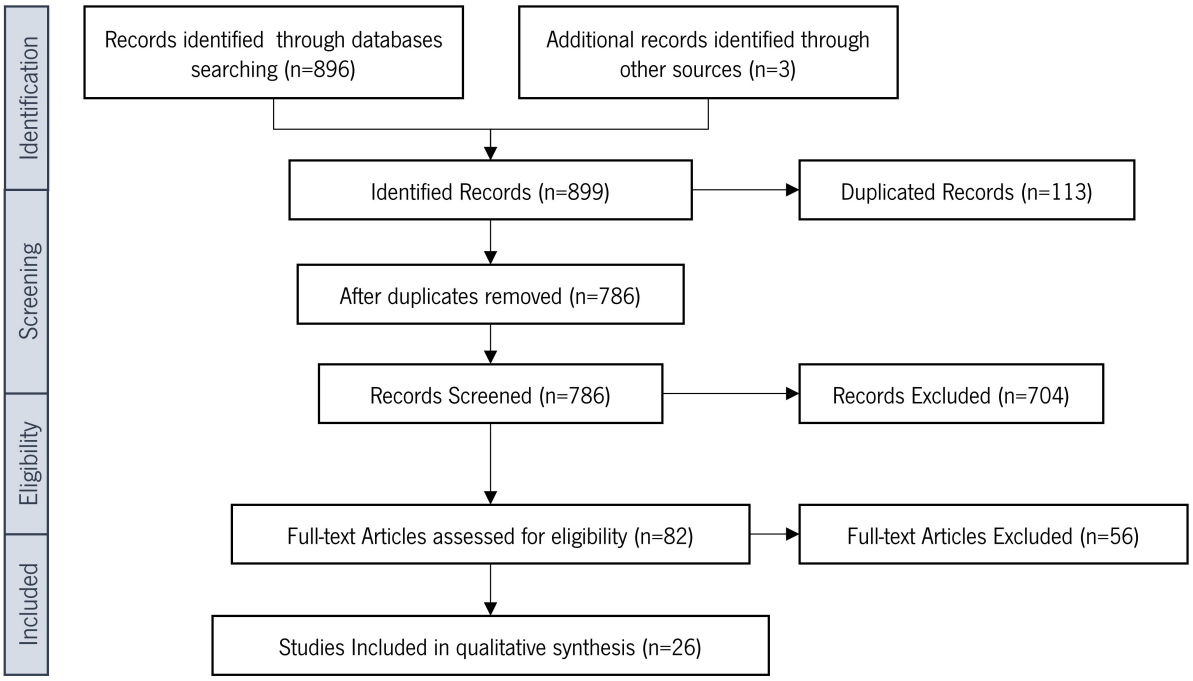


Figure 17: PRISMA flow diagram of the systematic literature search.

3.3 Schizophrenia Diagnosis

The complex nature of SZ and the controversy concerning its delimitation from other psychiatric disorders represent challenges to an objective clinical diagnosis [5]. Currently, SZ diagnosis mostly relies on the assessment of behavioral manifestations of clinical symptoms and on self-reports of subjective experience. Since SZ shares symptomatic features with other disorders, diagnosis may not be straightforward [277]. Specific tools have been developed to aid diagnosis. The Diagnostic and Statistical Manual of Mental Disorders - 5 (DSM-5), published by the American Psychiatry Association (APA) [278], and the International Statistical Classification of Diseases (ICD-10), published by the World Health Organization (WHO) [279],

list criteria for psychosis classification. Adjustments in both systems over time have been based on robust phenomenological studies [277]. Nevertheless, DSM-5 and ICD-10 have not been able to merge neurobiological measurements with clinical diagnostic criteria, mainly because brain and behavioral data do not properly map onto current symptom-based disorders [280, 281]. The Research Domain Criteria (RDoC), a National Institute of Mental Health (NIMH) project [282, 283], aims to integrate basic behavioral and neuroscience research to understand the core mechanisms underpinning symptoms that may span multiple disorders [37, 284]. Despite the improved reliability, both DSM-5 and ICD-10 diagnostic systems are limited by human subjectivity [284].

3.4 Neurobiological changes in schizophrenia

The neurobiological mechanisms underpinning SZ symptoms have been studied for decades with neuroimaging techniques, such as MRI, fMRI, MEG or EEG. These methods directly or indirectly provide information on structural or functional brain changes in the disorder. Each has advantages and disadvantages, mainly associated with its capacity to find the source of a particular brain signal and to distinguish two brain structures next to each other in space (spatial resolution), as well as with its capacity to discriminate two sequential events in time (temporal resolution) [285]. Whereas MRI and fMRI are characterized by higher spatial resolution, MEG and EEG have higher temporal resolution [286]. Structural and functional alterations have been consistently documented in patients diagnosed with SZ [287], which include a reduction of gray matter volumes in medial temporal, superior temporal, and prefrontal areas [288, 289, 290]. These brain regions are engaged in episodic memory, auditory processing, and short-term memory/decision making, respectively [287, 288].

Although the clinical symptoms and course of disorder are heterogeneous, cognitive deficits are a central feature of SZ and include changes in executive functioning, attention, and working memory [75, 76, 77, 78]. Cognitive dysfunction may precede the onset of psychosis, is related to functional outcome, and demonstrates limited responsiveness to antipsychotic treatment [75, 291]. Inadequate reception, discrimination, integration, or modulation of external sensory information can affect higher-order cognitive processes [80, 110]. fMRI provides a powerful method to explore the neural circuitry underlying cognitive deficits in SZ. These abnormalities are manifested in dysconnectivity between brain regions, as well as in changes in task-related brain activation [287, 290, 292, 293, 294, 295]. Despite its high spatial resolution, the fMRI technique has low temporal resolution (in the order of seconds). The exquisite temporal resolution of EEG, in the order of milliseconds, and its relatively low cost compared to the MEG technique are ideal to investigate how changes in information processing contribute to perceptual, cognitive, and social impairments in SZ [19, 296].

3.5 Electrophysiological Biomarkers of Schizophrenia

A great challenge for clinical research is the identification of biomarkers to aid diagnosis, prognosis, risk stratification or treatment monitoring of SZ [297, 298, 21, 299, 300, 301]. Specifically, the identification of biomarkers may shed light on the biological underpinnings of different categorically defined disorders [281]. Technological advances have enabled the identification of psychosis-related neurobiological changes that can be considered markers of psychopathology. Many of the proposed SZ biomarkers are EEG-related indices. The temporal resolution of EEG is ideal to investigate how changes in information processing contribute to sensory, cognitive, and social impairments in SZ [19, 18]. Several experimental paradigms have been designed to probe differences between SZ patients and healthy control (HC) subjects [302]. Consistent changes in specific ERP components have been proposed as potential trait markers of SZ (Table 3) [20, 303].

Table 3: Examples of proposed EEG biomarkers of SZ.

Suggested Biomarker	Type of feature	Functional Significance	Task	Amplitude	Latency	Power	Phase	References
Mismatch Negativity (MMN)	ERP component	Pre-attentive deviance detection	Auditory oddball paradigm	SZ<HC	-	-	-	[304, 134, 305, 306]
P300	ERP component	Attention allocation	Auditory oddball paradigm	SZ<HC	SZ>HC	-	-	[307, 308, 309, 310, 311, 306, 312]
P50	ERP component	Sensory gating	Paired-click paradigm	SZ<HC	-	-	-	[163, 306]
N100 (N1)	ERP component	Early auditory response, sensory and perceptual processing	Listening paradigms	SZ<HC	-	-	-	[24, 313, 314, 177]
Gamma Band (30-80Hz)	Brain oscillation	Selective attention; complex auditory information processing	Auditory oddball paradigm	-	-	SZ<HC	SZ<HC	[209, 315, 316, 317, 318]
40 Hz Auditory Steady-State Response (ASSR)	Brain oscillation	Sensory response; integrity of auditory circuits; synchronization of neural activity with external stimuli	Listening paradigm	-	-	SZ<HC	SZ<HC	[319, 320, 321, 322]

The predictive value of a biomarker depends on its generalizability across clinical populations, as well as on its sensitivity and specificity for a target population [323]. Recently, the ERP Biomarker Qualification

Consortium (LLC [324]) was launched to identify ERP biomarkers of SZ. Accordingly, a broad range of classification and predictive models may be developed [298].

3.6 Machine Learning for Schizophrenia Detection and Prediction

Machine learning methods have been increasingly adopted to optimize the use of neurobiological data in classification tasks and development of predictive models. Given the complexity and heterogeneity of neurobiological data in SZ research, machine learning methods have been increasingly adopted to optimize the use of this data in classification tasks and development of predictive models, offering a promising avenue for improving the understanding, diagnosis, and treatment of this disorder.

3.6.1 Classical Machine Learning Approaches

Machine learning methods used in psychiatry research and, specifically, in SZ have mostly relied on supervised learning models, trained with quantitative features selected and extracted from specific input data [325, 326, 327]. Following the training phase, a classification task is performed, in which test data are categorized [42, 326, 35]. SZ classification using features extracted from EEG data has been tested for several years [328, 329, 330, 331, 332, 333, 334, 335]. Table 4 summarizes the latest machine learning-based classification studies conducted since 2016, grouped according to classification goal and EEG task (see Appendix A for further details). Most recent studies relied on auditory, visual or working memory tasks, which tap into perceptual and cognitive processes that are altered in SZ, aiming to differentiate SZ from HC and at-risk subjects, as well as from other psychiatric disorders. Additionally, some approaches focused on resting-state EEG data, allowing the evaluation of intrinsic (i.e., task-unrelated) neural activity.

Table 4: Schizophrenia classification attempts using Classical Machine Learning methods based on the EEG signal. Categorization of the selected studies according to classification task and EEG paradigm.

Classification Task	Paradigm	First Author [Reference]	Year	EEG Channels	Sample Size	EEG Task	Classifier	Features	Evaluation Methods (Train/Validation/Test)	Test Set Size	Accuracy*
SZ vs. HC	Working memory	Johannesen [43]	2016	64	12 HC 40 SZ	SWMT	SVM	Event-related spectral perturbations - θ,α,β,γ frequency bands	3-fold cross-validation (66% / - / 33%)	12 HC 40 SZ	87%
	Auditory Processing	Shim [44]	2016	62	34 HC 34 SZ	Simple auditory oodball	SVM	P300 peak amplitude and latency; cortical current density values	Leave-one-out cross-validation	34 HC 34 SZ	88.24%
		Santos-Mayo [45]	2017	17	16 HC 31 SZ	Auditory oddball (standard, target, distractor)	SVM, MPL	P3b time-domain features; frequency-domain features	(60% / 10% / 30%)	47 subjects (not balanced)	93.42%
		Taylor [46]	2017	32	22 HC 21 SZ	Auditory oddball (3 paradigms, 3 different deviants)	SVM, GPC	Spatiotemporal images of MMN, deviant and standard responses	10-fold cross-validation (90% / - / 10%)	22 HC 21 SZ	80.48%
		Hsieh [48]	2018	32	12 HC 12 SZ	Passive Listening task	FLDA	N1 and P2 peak amplitudes	Leave-one-out cross-validation	12 HC 12 SZ	83.30%

Continued on next page

Table 4 – *Continued from previous page*

Classification Task	Paradigm	First Author [Reference]	Year	EEG Channels	Sample Size	EEG Task	Classifier	Features	Evaluation Methods (Train/Validation/Test)	Test Set Size	Accuracy*
SZ vs. HC	Auditory Processing	Zhang [336]	2019	64	32 HC 49 SZ	Basic sensory tasks with button press and auditory tone	RF	Demographic data; amplitude and latency of N1 and P2 peak.	10-fold cross-validation (90% / - / 10%)	32 HC 49 SZ	81%
	Visual Processing	Devia [337]	2019	32	9 HC 11 SZ	Simple visual-free task	Rule-based, LDA	Amplitude of ERP signals in 400-600ms interval	5-fold cross-validation (80% / - / 20%)	9 HC 11 SZ	71%
	Visual Processing+ Resting-state	Thilakavathi [338]	2017	23	23 HC 55 SZ	Modified visual oddball; Resting-state	FFNN, SVM	Entropy and complexity measures	(80% / - / 20%)	3 HC 15 SZ	88.50%
		Li [339]	2019	16	25 HC 23 SZ	Resting-state; Visual P300 task	LDA, SVM	P300 amplitudes SPN features	Leave-one-out cross-validation	25 HC 23 SZ	90.48%
	Resting-state	Piryatinska [340]	2017	16	39 HC 45 SZ	Resting-state	RF, SVM	-complexity	10-fold cross-validation (90% / - / 10%)	39 HC 45 SZ	85.30%

Continued on next page

Table 4 – *Continued from previous page*

Classification Task	Paradigm	First Author [Reference]	Year	EEG Channels	Sample Size	EEG Task	Classifier	Features	Evaluation Methods (Train/Validation/Test)	Test Set Size	Accuracy*
SZ vs. HC	Resting-state	Sabeti [47]	2018	24	20 HC 20 SZ	Resting-state	FLDA	Average power of $\theta,\alpha,\beta,\gamma$ frequency bands	10-fold cross-validation (90% / 10% / -)**	10 HC 10 SZ	83.74%
		Jahmunah [341]	2019	19	14 HC 14 SZ	Resting-state	SVM-RBF, SVM-poly, DT, LDA, kNN, PNN	Entropy and complexity measures	N.A.	14 HC 14 SZ	92.91%
		Buettner [342]	2019	19	14 HC 14 SZ	Resting-state	RF	Frequency bands power	10-fold cross-validation (90% / 10% / -)**	2 HC 5 SZ	71.43%
		Goshvarpour [343]	2020	19	14 HC 14 SZ	Resting-state	PNN	Entropy and complexity measures	5-fold cross-validation (80% / - / 20%	14 HC 14 SZ	100%
HC vs. 'at-risk' vs. SZ	Resting-state	Ye [344]	2017	64	40 HC 40 FES 40 CHR	Resting-state	SVM, kNN	Statistics and Linear Eigenvalue Statistics	(75% / - / 25%)	10 HC 10 FES 10 CHR	97.50%[a]; 77.33%[b]
		Liu [345]	2018	64	40 HC 40 FES 40 CHR	Resting-state	SVM, RF, NB, DT	Statistics and Linear Eigenvalue Statistics	5-fold cross-validation (80% / - / 20%)	40 HC 40 FES 40 CHR	91.16%[a] 73.31%[b]

Continued on next page

Table 4 – *Continued from previous page*

Classification Task	Paradigm	First Author [Reference]	Year	EEG Channels	Sample Size	EEG Task	Classifier	Features	Evaluation Methods (Train/Validation/Test)	Test Set Size	Accuracy*
HC vs. 'at-risk' vs. SZ	Auditory Processing	Luo [346]	2019	128	44 FES 34 UHR 19 CHR 39 HC	Somatosensory P50 supression	XGBoost	P50 amplitudes and P50 suppression ratio; behavioural data	(70% / - / 30%)	40 subjects (not balanced)	90.20%[c]
		Chang [347]	2019	64	25 FES 23 UHR 19 HC	P50 paradigm	DT	ERP measures, MCCB, brain network properties, demographic data	5-fold cross-validation (80% / - / 20%)	25 FES 23 UHR 19 HC	77.80%[b]
		Ahmedt-Aristizabal [348]	2020	30	40 HC 65 RSz	Passive auditory oddball	DT,KNN, SVM	Early and mid-latency ERP amplitudes	5-fold cross-validation (64% / 16% / 20%)	40 HC 65 RSz	44.75%
SZ vs. Other disorders	Visual Processing	Alimardani [349]	2018	21	23 BD 23 SZ	SSVEP	LDA, QDA, SVM, kNN, LRA	SNR mean, skewness, and kurtosis - SSVEPs	Leave-one-out cross-validation	23 BD 23 SZ	91.30%

* Concerning the method with best performance (underlined in the Classifier column).

** Used an independent test set

a HC vs. FES

b Three groups comparison.

c Four groups comparison

Abbreviations: EEG – Electroencephalography; HC – Healthy Control; SZ – Schizophrenia; BD – Bipolar Disorder; FES – First-Episode in Schizophrenia; CHR – Clinical high-risk; UHR – Ultra-high-risk; SWMT – Sternberg Working Memory Task; SSVEP – Steady State Visually Evoked Potential; SVM – Support Machine Vector; SVM-poly – SVM with a polynomial function kernel; SVM-RBF – SVM with radial basis function kernel; MLP – Multilayer Perceptron; GPC – Gaussian Process Classifier; RF – Random Forest; FFNN – Feed-Forward Neural Network; FLDA – Fisher Linear Discriminant Analysis; LDA – Linear Discriminant Analysis; QDA – Quadratic Discriminant Analysis; kNN – k-Nearest Neighbourhood; PNN – Probabilistic neural network; LRA – Logistic Regression Analysis; NB – Naïve Bayes; DT – Decision Tree; SPN – Spatial Pattern Analysis; XGBoost – Extreme Gradient Boosting; MMN – Mismatch Negativity; SNR – Signal-to-Noise Ratio; N.A. – information not available.

3.6.1.1 Classification of schizophrenia and healthy subjects

Figure 18 graphically summarizes the methods tested for the binary classification of SZ and the best performance of the proposed methods for specific tasks and groups of features.

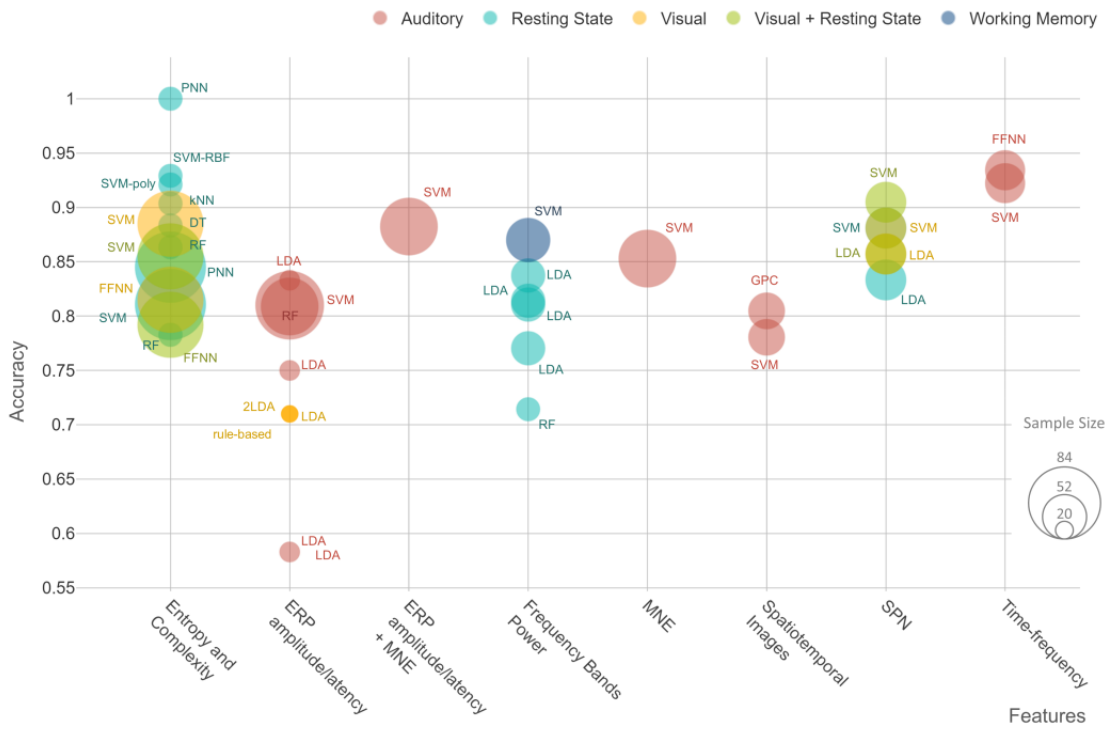


Figure 18: Accuracies achieved with machine learning classifiers (identified in dot labels) using extracted features (x-axis) for SZ binary classification performed in different EEG datasets. Dot color represents the EEG paradigm used for data collection (auditory, visual, and working memory tasks, resting-state, and combined data from visual and resting-state paradigms), whereas dot size reflects the total number of subjects. It worth noting that some results may overlap, which is reflected in the color combination of the corresponding conditions. **Abbreviations:** MNE – cortical density determined by minimum-norm estimate; SPN – spatial pattern network.

Visual processing Recent studies have focused on ERPs elicited during visual tasks. Devia *et al.* [337] examined differences between SZ and HC subjects in a simple visual task. Mean ERP amplitude in the 400-600 ms temporal window post-stimulus onset, averaged across four regions of interest (frontal, central, parietal, and occipital), was used as machine learning feature. Three classifiers were tested: linear discrimination analysis (LDA), a rule-based, and a combination of the posterior probability of two LDAs. The rule-based and the LDA classifiers using occipital ERPs in response to landscape images obtained the best performances (71% accuracy) [337]. Li and colleagues [339] examined features of signals recorded during task (P300) and rest conditions. Additionally, the functional network of both rest and task brain states was examined by computing the phase-locking value, reflecting phase synchronization between electrodes [209]. Using spatial pattern network (SPN), a supervised learning algorithm, properties of both networks were extracted. The features from both states were combined allowing a highly accurate (90.48%)

identification of SZ patients with a support vector machine (SVM) classifier [339]. Similarly, Thilakvathi *et al.* [338] combined features of visual oddball tasks and resting-state EEG signals. The classification was performed using FFNN and SVM algorithms using six entropy and complexity-related features from each condition. Unlike the findings of Li *et al.*, the combination did not improve the accuracy achieved by the visual task features alone (88.5%), as shown in Figure 19.

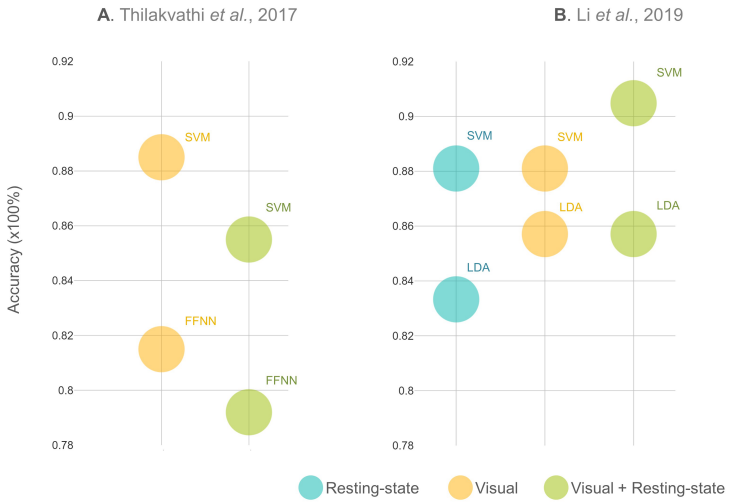


Figure 19: Classification accuracies achieved by Thilakvathi *et al.* [338] and Li *et al.* [339] using features extracted from EEG signals collected during visual and resting state conditions. The representation of these results follows the same configuration as Figure 18.

Auditory change detection Auditory tasks have also been the focus of SZ binary classification with machine learning. Shim *et al.* [44] relied on an auditory oddball task (AOT), combining sensor- (i.e., P300 peak over 62 electrodes) and source-level (i.e., cortical current density) approaches. The SVM classifier with 15 features combining eight sensor-level (in frontal electrodes) and seven source-level features (over the left temporal cortex) achieved the best accuracy (88.24%). This result is consistent with reported P300 changes in SZ [44]. Santos-Mayo *et al.* [45] used an AOT with three types of auditory stimuli (standard, target, and distractor) to elicit the P300. Features in both time and frequency domains were extracted and examined across different electrode clusters. The SVM algorithm resulted in a highly accurate SZ classification (92.23%), considering right hemisphere electrodes only; a MLP was tested using all electrodes achieving 93.42% accuracy. Taylor *et al.* [46] probed the MMN using three AOT with different types of deviants (stimulus duration, presence of right or left monaural gap, interaural time). ERP data were converted into spatiotemporal images (two-dimensional surface of the scalp over the timepoints) for each deviant and fed into two classifiers: SVM and GPC. The best accuracy achieved was 80.48% using gaussian process classification (GPC), with no significant differences between both classifiers [46]. Probing the processing of musical sounds, Hsieh *et al.* [48] reported decreased N1 and P2 amplitudes at frontal electrodes in SZ subjects, used as features in a LDA classifier. This classifier discriminated SZ against HC subjects with an 83.3% accuracy [96]. Zhang *et al.* [336] also tested SZ classification based on N1 and P2 features associated with three conditions: a button-press task to generate a tone (*press+audio*);

the passively listening of the same tone (*audio*); the button-press without sound (*press*). Using random forest (RF) algorithms, the best accuracy (81.10%) was achieved when N1 amplitude and latency over Fz from press+audio and audio conditions were combined with non-ERP related features (demographic data).

Despite the variety of features extracted, the reviewed studies show that AOTs contribute to an accurate identification of SZ patients. Specifically, the high accuracy in tasks eliciting the P300 or MMN supports the reliability of changes in these components in SZ classification. Of note, the combination of both time and frequency domains features led to improved accuracy. Data from small groups of electrodes enhanced the classifiers' performance when compared to EEG activity distributed over the entire scalp.

Working memory EEG signal decomposition into specific frequency bands allows studying neural oscillatory activity. Johannesen *et al.* [43] examined EEG time-frequency measures using a Sternberg Working Memory Task (SWMT), focusing on event-related spectral perturbations (ERSP) in theta (3-8 Hz), alpha (8-14 Hz), beta (14-23 Hz), and gamma (30-50 Hz) bands. Performance was specified using those bands over different time intervals (pre-stimulus baseline, encoding, retention, and retrieval stages) and three scalp locations (frontal-Fz; central-Cz; and occipital-Oz). SZ patients were identified with an accuracy of 87% using the SVM classifier. Specifically, gamma activity during the encoding interval was the main feature contributing to this outcome [43].

Resting-state Considering the documented resting-state EEG alterations in SZ [350], classification based on task-unrelated features has been also examined. Boostani and Sabeti [47] decomposed the resting-state signal into five frequency bands (delta, theta, alpha, beta, and gamma) across 24 electrodes, using discrete wavelet transforms, and computed the average power of each band. The most discriminative frequency band was selected and its features fed into a LDA classifier, achieving an accuracy of 83.74% [47]. Piryatinska *et al.* [340] extracted ϵ-complexity coefficients (indexing signal's complexity [351]). The model was tested on simulated data and applied to the binary classification of EEG recordings, using SVM and RF algorithms. The best performance was a mean accuracy of 85.3% using the RF classifier [340].

Three other studies also probed SZ classification utilizing a common database for their analysis [341, 342, 343]. Jahmunah *et al.* [341] and Goshvapour and Goshvarpour [343] selected entropy and complexity measures as classifier features. Jahmunah *et al.* [341] achieved the best accuracy of 92.91% using the SVM with radial-basis function (SVM-RBF) with 12 features. SVM with polynomial function kernel (SVM-poly), probabilistic neural network (PNN), decision tree (DT), k-nearest neighborhood (kNN), RF, and LDA were also tested, with accuracies, sensitivities, and specificities above 78% [103]. Goshvapour and Goshvapour [343] implemented a PNN model using only three features and testing different combination rules. The perfect classification (100% accuracy) was obtained using a feature-level fusion method [343]. Buettner *et al.* [342] decomposed the EEG data into 99 frequency values (from 0.5 to 50 Hz), and accurately (71.43%) classified SZ using a RF; 16 Hz, 41 Hz, and 49 Hz frequencies contributed more significantly to the performance.

The lack of task-related elements in the resting-state signal may hinder feature engineering. Notwithstanding, accurate classification of SZ was achieved using frequency decomposition and signal complexity

approaches. Performance was improved when a combination of task-related and resting-state features was used [339].

From risk to disease progression Changes in brain structure and function observed before SZ onset support the role of neurodevelopmental abnormalities in this disorder [272]. Studying different SZ stages (e.g., prodromal, first episode, chronic stage) may specify the pathophysiological mechanisms underpinning the disorder and inform about prognosis. Some (although few) studies relied on machine learning algorithms to discriminate stages of SZ progression. Luo *et al.* [346] and Chang *et al.* [347] examined the role of the auditory P50 as a potential biomarker for detection of psychosis risk. Both studies used the amplitude and suppression ratio as input features. Luo *et al.* [346] applied the Extreme Gradient Boosting (XGBoost) to classify HC, clinical high-risk (CHR), ultra-high-risk (UHR), and first episode schizophrenia (FES) patients. Groups were classified with an average accuracy of 90.2% based on EEG and behavioral data [346]. Chang *et al.* [347] used DT algorithms to discriminate HC, UHR, and FES. Demographic data, cognitive performance data, and brain network parameters were combined with P50 measures, resulting in an accuracy of 77.80%. Liu *et al.* [345] recorded data from HC, CHR, and FES, and tested different classifiers, relying on statistical measures of the signal (e.g., standard deviation, kurtosis, root mean square) and Linear Eigenvalue Statistics (LES). The best accuracy (approximately 91%) was achieved in the binary discrimination of HC and FES, using LES features with SVM. Discrimination of the three groups was achieved with a 73.31% accuracy using the same approach [345]. Likewise, Ye *et al.* [344] probed the role of different frequency bands in the discrimination of the same groups. Delta, theta, and low gamma frequencies played a more relevant role in the binary classification of HC and FES (97.50% accuracy), whereas alpha, beta and mid-range gamma-bands were more relevant to the discrimination of the three groups (77.3% accuracy). The classification focused on LES features extracted from those specific bands.

Compared to the high accuracy obtained in FES-HC discrimination, machine learning algorithms were less accurate when performing multi-groups classification using resting-state signals. Studies focusing on the SZ prodromal phase may provide insights into the EEG features that differ between control, at-risk, and diagnosed subjects. Of note, the use of auditory task-related signals showed improved accuracy, though combined with features such as demographic or behavioral data.

Ahmedt-Aristizabal *et al.* [348] implemented machine learning methods to identify children at-risk of developing SZ (RSz), comparing 65 at-risk subjects with a control group of 40 participants, with ages between 9 and 12 years. At-risk children presented psychotic-like experiences and/or family history of SZ spectrum illness. The mean amplitude of early (80-220 ms) and mid-latency (160-290 ms) ERP components, captured from 5 midline electrodes (Fz, FCz, Cz, CPz, Pz) during a passive auditory oddball task, were used as machine learning features. RSz classification was performed using kNN, DT, and SVM classifiers. Both classifiers achievied accuracies of approximately 44%, suggesting that the ERP features used were not sufficiently discriminative. The reduced accuracy may be accounted for by developmental factors. Maturational changes may lead to increased variability in the brain's response to stimuli in the environment, resulting in morphological alterations of the EEG signal and, consequently, rendering difficult the direct comparison of ERP responses [352].

Distinguishing schizophrenia from other disorders EEG-based machine learning algorithms have also been used to distinguish SZ from other psychiatric disorders. Alimardani *et al.* [349] tested the discrimination of bipolar disorder and SZ patients, who share symptomatic features [353], using an EEG feature-based classifier framework. Patients with SZ were identified with a 91.30% maximum accuracy, when applying the kNN classifier on EEG SNR features [349]. Despite its relevance, the potential of machine learning methods to discriminate SZ from other psychiatric disorders using EEG data has received little attention.

3.6.2 Novel approaches: the prospects of Deep Learning

The accurate results obtained with classical machine learning methods corroborate the potential of detecting and classifying SZ based on specific features of the EEG signal. However, both feature extraction and selection require domain expertise, and specifying which are the most relevant has proven difficult. Deep learning, a machine learning branch, has greatly improved learning performance in computer vision and image recognition [50]. Deep learning has the capacity to extract meaningful patterns from available data, without human bias. Deep learning methods have surpassed machine learning based on feature engineering in most fields [354, 355], and recently their potential in the realm of SZ diagnosis has also been explored.

3.6.2.1 Deep Learning algorithms

EEG-based deep algorithms have seen a growing interest in brain-signal decoding [51, 356, 52, 357]. Few studies attempted SZ classification using deep learning based on EEG signals. Table 3.6.2.2 lists the studies that explored the prospects of EEG-based deep learning. These studies mostly relied on resting-state data and achieved higher accuracy using CNNs.

3.6.2.2 Deep learning for Schizophrenia binary classification

Phang *et al.* [358] proposed a framework for SZ binary classification based on CNN. A combination of time-domain, frequency-domain, and topological measures of brain connectivity were used as model input, named multi-domain connectome (MDC). Each domain measure was fed into a 1D- (for the topology-based complex network) and 2D-CNN (for time and frequency domain measures) models. The MDC obtained accuracies above 90%. The combination of measures improved classification performance when compared to single-domain features [358]. Calhas *et al.* [359] also tested a CNN-based model on resting-state EEG to identify discriminative features of SZ. The model's architecture relied on a siamese neural network (SNN), composed of two "twin" convolutional sub-networks: each sub-network receives an input sample, which corresponds to a Discrete Short-time Fourier Transform representation of a one-minute long signal, extracted from one channel of each subject. Samples from different subjects were paired by channel. The output feature maps of each sub-network were compared, assessing samples' similarity. The network learned the desired behavior through the pairwise distance metric. Subjects' classification was performed

using the resulting feature maps fed into SVM, RF, XGBoost, NB, and kNN classifiers. RF classifier achieved an accuracy of 84% [359]. Naira and José [360] used the Pearson Correlation Coefficient to determine the relationship between 16 electrodes, hypothesizing differences in brain connectivity between SZ and controls. Images of the correlation matrices, created from 10-second resting-state EEG segments (six per subject), were used as input to the proposed 7-layer CNN-based model. SZ subjects were correctly recognized with an accuracy of 90% [360]. Oh *et al.* [361] proposed an 11-layered CNN architecture for a subject-based classification of SZ. Resting-state EEG segments (25 seconds) from 19 electrodes were used as input data. The model was able to correctly discriminate SZ from HC subjects with an accuracy of 81.26% [361].

Table 5: Schizophrenia classification attempts using Representation Learning methods based on EEG signals.

Classification Task	First Author [Reference]	Year	EEG Channels	Sample Size	EEG Task	Input Data	Architecture	Classifier	Evaluation Methods (Train/Validation/Test)	Test Set Size	Accuracy *
SZ vs. HC	Oh [361]	2019	19	14 HC 14 SZ	Resting-state	Time-domain and frequency-domain	CNN	Softmax	10-fold cross-validation (72% / 18% / 10%)	14 HC 14 SZ	81.26%
	Phang [358]	2019	16	39 HC 45 SZ	Resting-state	Time, frequency, topological connectivity	MDC-CNN (1D-CNN + 2D-CNN)	Softmax	5-fold cross-validation (80% / - / 20%)	39 HC 45 SZ	92.87%
	Calhas [359]	2019	16	39 HC 45 SZ	Resting-state	Time-frequency EEG representation	SNN	SVM; XGBoost; RF; NB; kNN	Leave-one-out cross-validation	39 HC 45 SZ	83%
	Naira [360]	2019	16	39 HC 45 SZ	Resting-state	Matrices of correlation between channels	CNN	Softmax	(70% / 30% / -)	4% of input segment	90%

Continued on next page

* Concerning the method with best performance (underlined in the *Architecture* and *Classifier* column).

a Three groups comparison.

Abbreviations: EEG – Electroencephalography; HC – Healthy Control; SZ – Schizophrenia; BD – Bipolar Disorder; FES – First-Episode in Schizophrenia; CHR – Clinical High-Risk; ANN – Artificial Neural Network; CNN – Convolutional Neural Network; RNN – Recurrent Neural Network; SNN – Siamese Neural Network; SVM – Support Machine Vector; RF – Random Forest; kNN – k-Nearest Neighbourhood; NB – Naïve Bayes; XGBoost – XGBoost algorithm; Se – Sensitivity; Sp – Specificity; n.d. – non-defined.

Table 5 – *Continued from previous page*

Classification Task	First Author [Reference]	Year	EEG Channels	Sample Size	EEG Task	Input Data	Architecture	Classifier	Evaluation Methods (Train/Validation/Test)	Test Set Size	Accuracy*
HC vs. 'at-risk' vs. SZ	Chu [362]	2017	64	40 HC 40 FES 40 CHR	Resting-state	Time-domain and frequency-domain	ANN; CNN; RNN	Softmax; RF; SVM	Cross-validation (n.d.)	40 HC 40 FES 40 CHR	92.5%[a]
	Ahmedt-Aristizabal [348]	2020	30	40 HC 65 RSz	Passive auditory oddball	Spatio-temporal EEG representation (EEG time course across electrodes)	CNN+LSTM	Sigmoid Activation	5-fold cross-validation (64% / 16% / 20%)	40 HC 65 RSz	72.54%

* Concerning the method with best performance (underlined in the *Architecture* and *Classifier* column).

a Three groups comparison.

Abbreviations: EEG – Electroencephalography; HC – Healthy Control; SZ – Schizophrenia; BD – Bipolar Disorder; FES – First-Episode in Schizophrenia; CHR – Clinical High-Risk; ANN – Artificial Neural Network; CNN – Convolutional Neural Network; RNN – Recurrent Neural Network; SNN – Siamese Neural Network; SVM – Support Machine Vector; RF – Random Forest; kNN – k-Nearest Neighbourhood; NB – Naïve Bayes; XGBoost – XGBoost algorithm; Se – Sensitivity; Sp – Specificity; n.d. – non-defined.

3.6.2.3 Deep learning for at-risk state recognition

Chu *et al.* [362] proposed an individual recognition approach. Resting-state signals were recorded from HC, CHR, and FES participants. Besides testing ANN and RNN architectures, the best performance was achieved by combining a CNN model for feature extraction with an RF classifier. The highest accuracies were above 95% for FES and HC groups, and approximately 80% for CHR, using both time-domain and frequency-domain representations as inputs [362]. This is the only study exploring the potential of deep learning to differentiate prodromes from HC and FES. Accuracy was lower in at-risk subject identification than in HC or FES recognition. Notwithstanding, deep learning methods seem to improve the discrimination of the three groups relative to machine learning algorithms.

Evidence for improved performance of deep learning compared to classical machine learning methods was shown by Ahmedt-Aristizabal *et al.* [348]. They used event-related features from auditory processing components, which were insufficient to identify children at risk for SZ. Subsequently, they attempted to perform classification using a deep learning model. For the new approach, they used the EEG signal from the 300 ms post-stimulus temporal window, registered over the same 5 midline electrodes. Two-dimensional data structures were created (electrodes x timepoints) for each trial and used as input to the 2D-CNN-LSTM model, with one CNN and two stacked long short-term memory (LSTM) layers. The RSz trials were recognized with an accuracy of 72.54% that clearly surpassed the 44.75% accuracy obtained with a DT classifier [348]. However, the accuracy obtained was lower than that achieved by resting-state models. This may reflect developmental effects and, consequently, an increased variability of EEG signals, as previously discussed. The combination of convolutional and recurrent neural networks led to an accurate discrimination of EEG signals from children at-risk of SZ conversion. This may suggest that abnormalities in auditory stimuli processing may also occur in prodromal phases of SZ.

With the exception of the study of Ahmedt-Aristizabal *et al.* [348], deep learning methods for both SZ classification and at-risk discrimination have relied on resting-state paradigms and on time and frequency-domain signals as input data. Despite the lack of studies relying on event-related biosignals, the accuracies achieved show that CNN models can successfully identify discriminative features of SZ. Even though deep learning methods have not been applied to the classification of different phases of the SZ spectrum based on event-related signals, the results reported by Ahmedt-Aristizabal *et al.* [348] suggest the potential of these methods in predicting SZ.

3.7 Promises and Pitfalls

The studies reviewed in the current chapter reveal the potentialities of EEG-based machine learning algorithms in discriminating SZ from HC, at-risk, and bipolar disorder subjects (>71% of accuracy). Some ERP measures, previously proposed as SZ biomarkers (e.g., P300, MMN, or N100), and the resting-state signal complexity were found to represent reliable features, though others have been equally successful such as statistical measures or oscillatory power. The SVM was the most commonly used classifier due to its computational efficiency [224], achieving the best performance in most studies.

The utilization of raw EEG data, as opposed to pre-defined features, offers notable advantages in the study of brain signatures associated with the SZ spectrum and illness course. One key advantage is the ability to capture subtle patterns through deep learning techniques. This capacity to learn intricate patterns from EEG data enhances the potential for deep learning methods to effectively address the challenge of identifying and characterizing brain signatures related to the SZ spectrum and the trajectory of the illness. Although few, the most recent studies reflect the growing interest in deep learning methods, which allow automatic feature extraction from resting-state EEG data and discriminate SZ from HC with accuracies above 83% using CNNs. The discrimination between normal, at-risk, and/or subjects with clinical symptoms is crucial for SZ prediction and prognosis. Resting-state and auditory event-related EEG signals were used to probe the capacity of models to recognize at-risk subjects, achieving accuracies of 80% and 72.5%, respectively. The performance achieved by the deep learning architecture relying on event-related signals represented an improvement of almost 30% in accuracy compared to the use of common classifiers, such as DT or SVM, fed with amplitude measures of ERP components. Unlike the models applied to resting-state measures, Ahmedt-Aristizabal *et al.* [348] not only used CNN in their proposed architecture but combined convolutional with recurrent layers. Specifically, the use of CNN has allowed correlating temporal, frequency, and spatial information extracted from EEG data and has proven a successful strategy in classifying SZ. However, recurrent networks such as LSTM are designed to process temporal or sequential information. These networks can play an important role in the process of learning patterns from task-related EEG signals, which correspond to sequences of ERP components. These promising results encourage the applications of deep learning to the EEG, in particular to event-related signals, whose features may reveal associations with cognitive, affective, or sensory changes in SZ.

Although risk prediction is important for early detection and treatment, the revised studies focused on diagnosis. This is indicated by the preference for supervised learning techniques rather than other machine learning methods. The supervised algorithms are designed based on labeled datasets. SZ data is obtained from subjects whose diagnosis is based on standardized methods, such as DSM-5, which mostly relies on symptom evaluation and self-reports. The proposed machine learning methods should not be considered substitutes for the gold standard diagnosis instruments (DSM-5/ICD-10); instead, they align with the RDoC framework purposes. Specifically, EEG-based machine learning may improve the detection and establishment of relationships between neurobiological signals and clinical symptoms, which can be valuable to understand the pathophysiology of SZ. Nevertheless, other limitations should be noted. The complexity of deep learning architectures, in general, is reflected in a difficult interpretation of the models, being particularly challenging to identify the most influential features of the data [363]. Whereas in classical machine learning methods the relationship between the input variables and the models' output is understandable, the deep learning architectures, data sources and training conditions hamper the comprehension of their internal representations. To overcome this limitation, interpretation techniques have been developed and implemented [364]. This could be an added value in understanding which information from the time course, frequency, or topographic distribution of the EEG signal contributed the most to the prediction. Together with clinical data (e.g., validated clinical scales and patients' self-reports), this

information may reveal the functional significance of the EEG features that are altered in SZ.

Another limitation is the quantity and quality of brain data on which the performance of machine learning algorithms depends. Since the sample size is critical, it should be representative of the SZ population or other psychiatric disorders whose data will guide learning [365]. In the studies included in this review, the number of subjects tested was variable and generally small, which limits the generalization of the current findings. Based on a large number of features and small sample sizes, the algorithm predictions might be inaccurate and not generalize well, representing a case of overfitting. Thereby, a thorough validation is necessary to avoid amplification of data errors and biases. An appropriate methodological approach is the combined use of a validation dataset, allowing the fine-tuning of model hyperparameters, and an independent test set to evaluate the final model's performance and generalization error in an unbiased way [215]. Non-overlapping subjects' data in training, validation, and test sets should be ensured [365]. Most of the described studies used leave-one-out or k-fold cross-validation (5- or 10-fold, predominantly) methods, which represent resampling techniques used to enhance the model's generalizability. The size of train and test datasets, as well as the number of folds in cross-validation, depend on the sample size, number of features, and algorithm used [366]. However, cross-validation methods can also overestimate predictive accuracy and mask the true model generalizability [367]. Future studies will benefit from larger sample sizes and datasets to improve generalizability. In particular, the increase of training data allows testing more complex models and reduces overfitting [50].

Even though the usefulness of EEG as a functional neuroimaging tool is unequivocal, some challenges should also be considered when using this technique. Despite its relatively high portability and flexibility, electrically shielded and acoustically isolated rooms are recommended for EEG data collection to reduce signal's noise. A controlled experimental design is also required to minimize EEG artifacts. Variations in signal-to-noise ratio across trials and subjects, which may also arise from differences in EEG data acquisition conditions, may affect the quality of the signal that is used as input in machine learning algorithms [127]. Moreover, EEG differences are expected between different subjects and across different sessions given the non-linearity and non-stationarity of these signals [368, 369]. Longitudinal EEG data could inform on the generalization of the models across sessions and improve the capacity of those models to learn intrinsic information [370, 371]. Finally, a well-characterized public SZ EEG database is highly recommended for the direct comparison and objective evaluation of different algorithms' performances. Only then it will be possible to accurately measure the effectiveness of the developed models and assess the usefulness of machine learning both for research and clinical applications.

Some of the clinical symptoms may overlap across psychiatric disorders [372]. Indeed, comorbidity is the rule rather than the exception in psychiatric disorders, which has implications for patients' diagnosis and treatment [372, 373]. Therefore, its effects should be considered in SZ research [374, 375]. As a potential limitation, the current review did not consider comorbidity as an exclusion criterion in studies assessed for eligibility, which may limit the interpretation of results. Notwithstanding, most of the studies included in the present review did not report whether the samples have been controlled for possible comorbidities. Future studies should address these issues.

3.8 Summary

The complexity and heterogeneity of schizophrenia (SZ) symptoms challenge an objective diagnosis, which is often based solely on behavioral and clinical manifestations. Moreover, the boundaries of SZ are not precisely demarcated from other nosologic categories, such as bipolar disorder. The early detection of SZ can lead to a more effective treatment, improving patients' quality of life. Over the last decades, hundreds of studies aimed at specifying the neurobiological mechanisms that underpin clinical manifestations of SZ, using techniques such as EEG. Changes in event-related potentials of the EEG have been associated with sensory and cognitive deficits and proposed as biomarkers of SZ. Besides contributing to a more effective diagnosis, biomarkers can be crucial to SZ onset prediction and prognosis. However, any proposed biomarker requires substantial clinical research to prove its validity and cost-effectiveness. Fueled by developments in computational neuroscience, automatic classification of SZ at different stages (prodromal, first episode, chronic) has been attempted, using brain imaging pattern recognition methods to capture differences in functional brain activity. Advanced learning techniques have been studied for this purpose, with promising results. This review provides an overview of recent machine learning-based methods for SZ classification using EEG data, discussing their potentialities and limitations. It serves as a starting point for future developments of effective EEG-based models that might predict the onset of SZ, identify subjects at high-risk of psychosis conversion or differentiate SZ from other disorders, promoting more effective early interventions.

Chapter 4

Auditory event-related potential abnormalities in Schizophrenia

Alterations in auditory processing have been consistently reported in SZ, and linked to specific symptoms such as auditory verbal hallucinations. However, the neural dynamics of sound decoding in SZ remain to be clarified. This study presents the characterization of the EEG signals of healthy subjects and patients with SZ, analyzing the main differences in the time course of the response to simple auditory stimuli. This study evaluates the robustness of auditory processing deficits in SZ indexed by the N100 and P200 ERP components. Additionaly, an ERP, trial-by-trial, and time-frequency combined analysis provide a more comprehensive understanding of the neural mechanisms underlying passive sound processing in individuals with SZ.

Contents

4.1 Introduction **66**
4.2 Methods **68**
4.3 Results **75**
4.4 Discussion **85**
4.5 Limitations and Future directions **88**
4.6 Summary **89**

4.1 Introduction

Alterations in sensory processing have been consistently reported in SZ [110] and include the auditory system [376, 377, 23]. Sensory-driven dysfunction could have widespread consequences for higher-order processes such as vocal emotion recognition [378] and account for specific psychotic symptoms such as auditory hallucinations [379]. Auditory hallucinations are a hallmark symptom of schizophrenia and may suggest altered auditory dysfunction. However, they have been associated with complex speech perception changes, or with alterations in higher-level cognitive functions, such as inhibitory control [380, 58, 110, 59, 23]. Hearing voices in the absence of corresponding acoustic stimulation is a common experience in approximately 75% of SZ patients [88]. The predominance of auditory over other forms of hallucinations is commonly indicative of the disorder [88, 381]. However, auditory verbal hallucinations are not necessarily a sign of pathology [382]. Instead, psychotic-like experiences, including hallucinations, are reported in 10-15% of the general population without requiring clinical care [265, 383].

Sound perception in SZ has been extensively studied with high temporal resolution electrophysiological techniques such as the EEG [384]. EEG-derived ERP studies have explored the neurodynamics of sound perception, from early to late perceptual stages, focusing on different components [385]. The P50, N100, and P200 are early and middle latency auditory evoked potentials. The P50 reflects a pre-attentive processing stage, whereas the N100 and P200 are more susceptible to modulation by attentional mechanisms [159, 386]. P50 abnormalities, including diminished amplitude and delayed latency, are well-established endophenotypes of SZ and suggest impaired sensory gating [318, 387, 386]. Moreover, SZ patients present reduced amplitude of N100 and P200 components [386]. Alterations in N100 generation, reflecting changes in early sensory processing, are consistently found and well-established in patients with SZ [314, 388]. Additionally, N100 amplitude modulations have been used as a quantifiable measure of the modulation of auditory processing by attention [389]. For example, attention allocation does not always lead to N100 amplitude enhancement in SZ [24], which could indicate impaired attentional selection and inhibition due to deficits in cognitive control [389].The N100 is also reduced in passive listening tasks, which suggests that both endogenous and exogenous N100 subcomponents are affected by SZ [24]. The P200 follows the N100, and they often covary [390, 391]. Nonetheless, they differ in their scalp topography and functional significance. Even though the functional significance of the P200 is less clear, it has been linked to the attention-modulated processes required for auditory discrimination, as well as early categorization processes [390, 391]. The direction of P200 alterations is less consistent, with some studies reporting decreased and others increased amplitude in SZ compared to healthy controls in auditory tasks [25].

The oscillatory nature of neuronal activity has also motivated the exploration of EEG-based measures in frequency or time-frequency domains. Oscillatory rhythms can be recorded during sensory stimulation or task performance. The procedure usually involves decomposing the continuously recorded EEG signal into different frequency bands (delta, theta, alpha, beta, and gamma) or representing the signal over time and frequency [209, 392]. Cortical oscillations are widely reported to be disturbed in SZ and shed additional light on the integrity of neural network activity related to perception and cognition in this disorder [393,

392]. The oscillatory neuronal activity in the gamma-band (30–100 Hz) is involved in representing and selecting information [394]. During normal brain functioning, the effective representation of information and propagation across hierarchical networks are facilitated by synchronized neuronal responses [395]. Besides being implicated in perceptual and motor functions, abnormalities in the gamma-band are associated with core cognitive disturbances and reflect the disrupted functioning of pyramidal cells and parvalbumin interneurons[1] in SZ [394, 393, 397]. There is substantial evidence for both abnormalities of gamma-amino-butyric acid (GABA) neurotransmission and reduced synaptic connectivity in the auditory cortex of SZ patients [398, 399, 400]. Deficits of the auditory steady-state response (ASSR) and early auditory-evoked gamma-band response (EAGBR) (25–100 msec upon the presentation of an auditory stimulus [401]) were reported in SZ patients when compared to healthy individuals [317, 402, 394]. Therefore, auditory-evoked γ oscillations can inform on the integrity of synaptic connectivity in the auditory cortex of SZ patients [394, 403]. Besides the relationship between gamma-band activity and the construction of coherent object representations, it has been shown that lower frequencies (delta-, theta-, and alpha-bands) are also functionally involved in a wide range of perceptual, cognitive, and executive processes [395]. Different studies have also revealed changes in auditory-evoked neural activity in these frequency spectra in SZ patients [404, 405, 406, 23, 407, 408, 409, 410].

4.1.1 Current study and hypotheses

Neurophysiological responses to evoked stimuli may provide a window into individual differences in sensory processing. Most studies revealed alterations in sensory processing in response to more complex auditory stimuli, such as voices or sounds with emotional content [411, 412]. By investigating how the auditory system processes simple sounds, we can gain valuable knowledge that helps us unravel the underlying mechanisms of basic auditory perception. Therefore, exploring simple nonverbal sound processing, such as pure tones, may elucidate the core principles of auditory perception that remain to be fully understood [413].

The current study aims to provide a more comprehensive picture of simple sound perception in SZ patients compared to the healthy control (HC) by combining different measures, including single-trial ERP analysis and time-frequency analysis. Previous studies have mainly relied on separate ERP and time-frequency analyses [414, 415, 209, 191], or have tested small samples, often <30 patients [313, 416, 417, 418, 419]. It remains to be clarified whether alterations in early processing of sound features are consistently observed across trials, and what is their underlying neural dynamics. The single-trial analysis allows for the investigation of trial-to-trial variability, which can be useful in examining the impact of individual differences and assessing the reliability of the results. Time-frequency analysis, on the other hand, allows for investigating changes in neural oscillations over time and frequency bands, providing insights into the temporal dynamics of sound perception. Testing a larger sample (> 50 patients and HC), the cur-

[1] A subpopulation of GABAergic interneurons that manifest unique cellular and molecular properties that dramatically influence effects on signaling and ultimately alter cognitive behaviors. GABAergic interneurons are inhibitory neurons of the nervous system that play a vital role in neural circuitry and activity. They are so named because they release GABA neurotransmitters [396].

rent study aims to address the limitations of previous studies and provide a more reliable characterization of auditory processing impairments in SZ.

Based on previous studies [25, 24, 177], we expect to observe alterations in early stages of auditory processing in SZ, reflected in N100 and P200. Building on previous research that has demonstrated alterations in neural activity among individuals with SZ [399, 400, 408, 409, 410], such as reduced neural synchronization and abnormal neural oscillations, we hypothesized alterations in oscillatory neural activity in SZ patients, suggesting disruptions in neural communication. Additionally, we hypothesize that the morphology of each single-trial signature is altered in SZ, which can drastically improve the separation of the two study groups. The robustness of the changes found in SZ at the single-trials level can catalyze its use for the development of classification or decision-making algorithms.

This chapter is organized as follows. Section 4.2 describes the experimental procedure used for data acquisition, characterizes the datasets included in the study, and details the EEG data and statistical analyses performed. Sections 4.3 and 4.4 present the results of the data analyses performed and discuss the most significant changes found. The last section points out the limitations and future directions.

4.2 Methods

This section describes the materials and methods used for EEG data acquisition and analysis. It begins by providing a detailed characterization of the datasets included in the study and the experimental procedure used for data collection. Afterward, it describes the ERP, time-frequency, and statistical analyses performed on the EEG data.

4.2.1 Participants and databases

A significant challenge of this thesis was to achieve a data pool large enough to implement deep learning methods[2]. At the time of the start of this study, only one database of EEG data on auditory processing in SZ patients was publicly available (henceforth referred to as '*dataset A*') [420]. Our research team collected another dataset (henceforth called '*dataset B*') to complement the existing dataset and increase the available data pool. Thus, this study on auditory processing changes in SZ exploited the data collected by the two research teams in a merged database.

4.2.1.1 Dataset A

Ford *et al.* [421] collected dataset *A* in the context of a study funded by the NIMH. This dataset includes EEG data from patients with a DSM-IV diagnosis of SZ (N=35) and schizoaffective disorder (N=14). These syndromes are broadly indistinguishable, and their separation is questionable[422]. For this reason, both subgroups are considered and, from now on, referred to as SZ patients (49 subjects). All volunteers have given their written consent to participate. The research team recruited 17 of these subjects in an earlier

[2]The implementation of deep learning methods for SZ classification will be addressed in Chapter 5



Figure 20: Dataset's composition in number of HC and SZ patients.

Table 6: Clinical and demographic data of HC and SZ patients that compose both datasets A and B.

	Dataset A				Dataset B				A vs. B			
	HC	SZ	*p*		HC	SZ	*p*		HC (*p*)		SZ (*p*)	
N	32	49			31	16						
Age			0.598	n.s.			0.053	n.s.	3.15e-18	***	0.0001	***
mean (years)	38.38	40.02			22.10	24.64						
SD (years)	13.92	13.48			3.33	5.17						
Gender			-0.107	n.c.			-0.031	n.c.				
females	6	8			18	7						
males	26	41			13	9						
Education			4.04e-7	***			0.077	n.s.	0.060	n.s.	0.783	n.s.
mean (years)	16.02	13.55			15.03	13.33						
SD (years)	1.92	1.96			2.12	4.29						
PANSS positive scale			-	-			-	-	-	-	0.003	**
mean rating		18.31				23.67						
SD rating		5.35				5.71						
PANSS negative scale			-	-			-	-	-	-	0.486	n.s.
mean rating		17.43				16.25						
SD rating		4.87				6.54						
SAPS Total			-	-			-	-	-	-	-	-
mean rating		6.57				-						
SD rating		3.07				-						
SANS Total			-	-			-	-	-	-	-	-
mean rating		6.69				-						
SD rating		2.71				-						

Abbreviations: SD - standard deviation; N - sample size; *p* - statistical *p*-value; PANSS - Positive and Negative Syndrome Scale; SAPS - Scale for the Assessment of Positive Symptoms; SANS - Scale for the Assessment of Negative Symptoms; n.s. - not significant ($p>.05$); * $p<.05$; ** $p<.01$; *** $p<.001$; n.c. no correlation.

stage of the disorder; the other 32 were chronic patients. The diagnosis of SZ subjects was based on the Structured Clinical Interview for DSM-IV (SCID). The clinical assessment was conducted by a trained research assistant, along with a psychiatrist or clinical psychologist, who rated SZ symptoms using the PANSS, SAPS, and SANS. Dataset *A* includes 32 healthy subjects, which form the control group (HC). Table 6 presents the clinical and demographic data for the HC and SZ groups of dataset *A*. The table reports the group average values with the standard deviation for age, years of education, and clinical ratings. Gender differences were analyzed with Pearson chi-square tests. Between-group differences in age and education level were analyzed with *t*-tests.

4.2.1.2 Dataset B

Dataset *B* includes EEG data from 16 first-episode SZ patients and 31 HC subjects. Experienced psychiatrists performed the SZ diagnosis. Symptom severity in psychotic patients was assessed using the PANSS. The institutional review boards of the University of Minho, Faculty of Psychology – University of Lisbon, and Hospital Center of Lisbon North approved this study. Before participating in the study, all volunteers were informed about the procedures and provided written informed consent. Table 6 presents the clinical and demographic data for HC and SZ groups of dataset *B*. Table 6 also reports the differences in clinical and demographic data between both datasets *A* and *B* for HC and SZ groups. Group average values and the standard deviation for age, years of education, and clinical ratings are reported. Gender was analyzed with Pearson chi-square tests. Between-group differences in age and education level were analyzed with *t*-test.

4.2.1.3 Merged Dataset

The study on the feasibility of joining the two datasets was conducted considering the similarity of the paradigm used for EEG data acquisition. The Appendix A presents this study in more detail. The merging of the datasets could add value since training deep learning neuronal networks requires a large data pool. In addition, a larger sample size allows for testing the hypotheses with greater statistical power and detecting existing effects. Table 7 presents the clinical and demographic data for the HC and SZ groups of the merged dataset.

4.2.2 Experimental Procedure

The two research teams did the signal acquisition using a similar paradigm, with some differences in sound properties. The EEG paradigm consisted of three experimental blocks: *auditory-motor*, *auditory-only*, and *motor-only* conditions. In the *auditory-motor* condition, the subjects pressed a button every 1-2 seconds to elicit a tone. This block would end after the emission of 100 consecutive sounds. The generated tones were then played back in the *auditory-only* condition, in which participants listened to the sounds passively without pressing the button. The button was pressed at the same pace as the first condition but without reproducing the sound in the last condition. In this way, it was possible to isolate the

Table 7: Demographics of HC and SZ patients in merged dataset.

	Merged Dataset			
	HC	SZ	*p*	
N	63	65		
Age			0.010	**
mean (years)	30.37	36.60		
SD (years)	13.02	13.71		
Gender			-0.159	s.c.
females	24	15		
males	39	50		
Education			3.83e-6	***
mean (years)	15.48	13.71		
SD (years)	2.07	2.03		
PANSS positive scale			-	-
mean score		19.36		
SD score		5.79		
PANSS negative scale			-	-
mean score		17.20		
SD score		5.20		

Abbreviations: SD - standard deviation; N - sample size; *p* - statistical *p*-value; PANSS - Positive and Negative Syndrome Scale; n.s. - not significant ($p>.05$); * $p<.05$; ** $p<.01$; *** $p<.001$; s.c. small correlation.

brain processing referring to the motor component of the paradigm. Figure 21 schematically presents the three blocks of this experimental paradigm, with 100 trials per condition.

This paradigm allows for comparing brain activity in response to a stimulus generated by the participant when pressing the button with the response to a stimulus generated externally [423]. Notwithstanding, a simple passive listening task prevents confounding effects related to between-participant differences in cognition and motivation [390]. For this reason, we only considered data from the response to externally generated sounds (*auditory-only* task).

4.2.3 Stimuli

Stimuli were two sinusoidal tones of different frequencies: 1000 Hz, 80 dB SPL tones for dataset *A*, and 600 Hz, 70 dB SPL tones for dataset *B*.

4.2.4 EEG acquisition and preprocessing

EEG data from both datasets *A* and *B* were recorded using a 64-channel Active Two Biosemi system (Biosemi, Amsterdam, The Netherlands, [424]). Dataset *A* was acquired in a continuous mode at a digitization rate of 1024Hz and referenced offline to averaged earlobe electrodes. Dataset *B* was collected continuously but at a sampling rate of 512 Hz and re-referenced to the averaged mastoid electrodes. Both

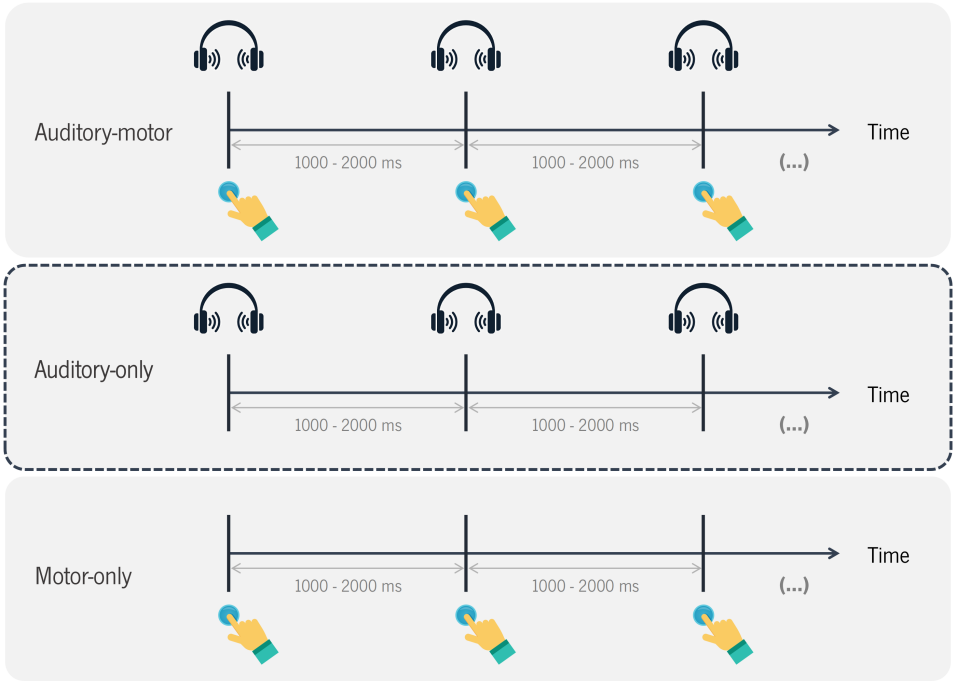


Figure 21: EEG experimental paradigm with three blocks: *auditory-motor*, *auditory-only*, and *motor-only* conditions.

datasets were preprocessed offline using EEGLab, a MatLab Toolbox [141]. Before preprocessing, dataset *A* was downsampled to 512Hz so that all data had the same sampling frequency. Then, the EEG data were digitally filtered (zero phase shift 4^{th}-order Butterworth high-pass filter; the low cutoff of 0.1 Hz), and the outlier channels were intolated. Individual segments, or epochs (one per tone delivered), were created from continuous recordings with a 3000 ms duration, time-locked to tone onset with 1500 ms pre-stimulus and post-stimulus. Segments were normalized with a baseline correction by subtracting the mean amplitude of the -100 to 0 ms pre-stimulus interval from the whole segment. Eye blinks and movements were monitored through vertical and horizontal electrooculograms (EOG). The vertical EOG was derived by subtracting the activity measured at an electrode positioned below the left eye from the electrode positioned above it (FP2). The horizontal EOG was derived by subtracting the activity measured between electrodes placed at the outer canthi of the eyes. EEG data from all electrodes were then corrected for vertical and horizontal eye movements using a regression-based algorithm [425]. Segments were also semiautomatically screened using the FASTER toolbox [426] for artifact rejection (eye and muscle movements, high-frequency noise, electrode drifting, amplifier blocking) and rejection of single trial outliers. This toolbox was also used to identify and reject ICA component outliers. After artifact rejection, at least 75% of the segments per condition per subject entered the analyses.

4.2.5 EEG data analyses

ERP technique is a commonly used approach to study cognitive and sensory processes. However, ERP can be misleading due to the averaging process since intra- and inter-subject variability might be

masked, not reflecting the subjects' actual brain dynamics [427]. In turn, analyzing EEG signals on a trial-by-trial level is likely to provide essential information for a deeper understanding of brain-behavior relationships [428]. In addition, the study of neural oscillations and their synchronization provides insight into the mechanisms underlying information processing in the brain that cannot be assessed with ERPs [209]. Time-frequency analysis provides insights into frequency-specific modulations, event-related spectral changes, and temporal dynamics of neural oscillatory activity in auditory processing [209].

EEG studies produce large spatiotemporal datasets containing crucial information for the understanding of neural mechanisms engaged in perceptual and cognitive processing, for example. However, multiple comparisons might potentiate high Type I error rates (i.e., finding differences when they do not exist) [429]. Standard approaches in ERP are characterized by statistical analyses on specific time windows and regions of interest and do not always control for Type I errors. This approach can lower the power to detect real effects. In contrast, mass univariate approaches offer an alternative analytical method [429].

The current study combined ERP, trial-by-trial, and time-frequency approaches. ERP analysis was performed only qualitatively to confirm the signal pattern in response to an auditory stimulus documented extensively in the literature. Trial-by-trial and time-frequency analyses were exploratory approaches employing cluster-based permutation tests to determine differences between HC and SZ signals. ERP, trial-by-trial, time-frequency, and statistical analyses were performed using MNE-python package [430] and custom-written python scripts.

4.2.5.1 ERP analysis

Single-trial epochs containing excessive eye blinks or motor artifacts (±100 μV criterion) were rejected. For each participant, ERP averages had at least 75% of the trials. The analyses excluded 3 HC and 4 SZ subjects, whose number of valid trials did not reach the threshold. Thus, the HC group included 60 and the SZ group 61 subjects, both groups with an average of 94±5 trials per subject. For each subject, the accepted single trials were averaged for each electrode. The averaging process across subjects was then conducted for each group, producing the HC and SZ grand averages.

Considering the N100 and P200 components, the window of interest was restricted to the time interval from -100 to 400 ms. After a careful visual inspection of grand average waveforms, the minimum (corresponding to the N100 peak) and the maximum amplitude value (relative to the P200 peak) were determined for each electrode, as well as the latency at which these values occurred. The electrodes that captured the maximum peaks for each component were determined. Topographical maps of ERP amplitude were created for each component.

4.2.5.2 Trial-by-trial Analysis

Although most auditory processing studies typically focus on the trial-average activity of EEG, the so-called evoked potentials, a single-trial analysis can provide a more powerful window into the neural processes underlying auditory perception [431]. In contrast to ERP analysis, trial-by-trial analyses consider intra-subject variability [127, 432]. The analysis described in this section aimed to verify whether: 1) the

pattern found in the grand averages is elicited clearly in each of the trials in response to the auditory stimuli; 2) there are spatiotemporal differences between groups when taking into consideration intra- and inter-subject variability.

For the visual assessment of single-trial epochs, the evoked responses of all subjects were represented as two-dimensional maps with time as abscissa, trial index as ordinate axis, and amplitudes coded by color saturation. The analysis also included the evaluation of the mean difference waveforms of the HC and SZ groups' single trials, identifying the peaks with the highest amplitude (in absolute value) and respective latency.

To examine group spatiotemporal cluster differences between HC and SZ, a non-parametric clustering-based permutation method was performed with MNE-Python software [430]. This exploratory mass univariate analysis at all sensors and time points was performed to assess differences in single-trial amplitude and topographical distribution. The multiple tests correction used the threshold-free cluster enhancement (TFCE) method, which clusters brain activity in adjacent regions [433]. For the cluster-based tests, we performed a 1000 TFCE computation (h=2.00, e=0.50). The results were mapped onto images of the EEG channels evoked responses over time, masking the statistically significant points. From a different perspective, the permutation results were visualized using topographical maps and signal time course plots, which presented the EEG channels and time intervals contributing to each significant cluster.

4.2.5.3 Time-frequency Analysis

The ERP analysis captures time-locked elements of the neural response but not phase-locked (or induced) oscillatory activity [19]. Evoked and induced responses differ in how the underlying neural activity is phase-locked to the stimulus over trials, reflecting different neurocognitive mechanisms [434, 435]. The time-frequency analysis allows the characterization of EEG spectral event-related changes over time [19]. Moreover, it allows for extracting and quantifying the magnitude and phase characteristics, informing on neuronal synchrony and, therefore, the brain's dynamic responses to events as transient perturbations of the ongoing EEG activity.

Time-frequency decomposition was performed using a Complex Morlet's wavelet transform, applied in 0.5 Hz steps from 4 to 60 Hz at each time point to yield time-frequency maps. The constant ratio (c) of central frequency (f_0) was 6 and the multiplication factor (m) was 4 [209]. The number of cycles ($n_c \approx$10,44) for the f_0 and the normalization factor ($A \approx$ 4,50) were computed from c and m parameters, as explained in section 2.2.3.

The wavelet-based spectral decomposition of single-trial EEG epochs time-locked to pure tones allowed the computation of phase-locking factor (PLF), total power (POW_{total}), and evoked power (POW_{evoked}) values. The induced power ($POW_{induced}$) was isolated by removing evoked (phase-locked) power from the single-trial-based total power estimate. Section 2.2.3 describes the operations executed to obtain each time-frequency map in detail. Baseline correction was applied to each map by subtracting the average of a baseline period (-350 to -100 ms relative to the stimulus onset) from each time point of the entire segment.

Time-frequency maps were created for the temporal window of -300 to 500 ms relative to stimulus onset. The visual analysis of the PLF was further complemented with topographic maps showing the spatial distribution of the phase synchrony values in the different frequency bands.

As in trial-by-trial analysis, an exploratory study using a clustering-based permutation method compared clusters in time-frequency PLF and non-phase and phase-locked power estimates between HC and SZ. This analysis used a non-parametric statistical procedure based on permutations and cluster-level statistics. Given the data intricacy, the spatial dimension was not considered, and permutations were performed on single-trial maps of the FCz channel. The significance level for all statistical tests was set to 0.01 two-tailed. The underlying statistical test for all permutation tests was the one-way ANOVA. We performed 1000 permutations with a significance level of 0.01 and used a threshold of 2.70 to compare single trials from HC and SZ patients for all time-frequency measurements. The contrasts between HC and SZ were visualized through time-frequency maps of PLF and power values. All cluster-based permutation tests plotted the F-statistic values as time-frequency maps, masking the statistically significant regions.

4.3 Results

The following three subsections describe and present the results obtained from the ERP, trial-by-trial, and time-frequency analyses conducted on the merged dataset. The trial-by-trial and time-frequency analysis subsections are supplemented with the visual outputs of the cluster-based permutation tests to show significant differences between the HC and SZ groups.

4.3.1 ERP Analysis

Pure tones elicited the early auditory evoked potentials observed within the first milliseconds post-stimulus onset, followed by prominent negativity around 100 ms (N100) and a significant positive deflection peaking around 200 ms (P200). Those components were broadly distributed across the scalp for HC and SZ groups. The N100 and P200 (minimum and maximum) peaks, the latencies at which they occurred, and the electrodes that recorded the highest peak values were computed from grand averages in the 0-400 ms post-stimulus time interval. Table 8 presents those values for HC and SZ groups.

Table 8: Peak amplitude, latency and location on the scalp for the N100 and P200 components found in the grand average waveforms of the HC and SZ groups.

	N100 Peak		P200 Peak	
	HC	SZ	HC	SZ
N (subjects)	60	61	60	61
Electrode	FCz	FCz	Cz	Cz
Latency (ms)	89.8	87.9	185.5	185.5
Amplitude (μV)	-6.93	-4.53	4.73	3.78

Figure 22a presents the topographical maps at 90 ms and 185 ms, showing the spatial distribution of N100 and P200 amplitude effects, respectively. These plots indicate that the N100 effect is more significant over frontocentral scalp regions, while the P200 effect was most pronounced over central electrodes.

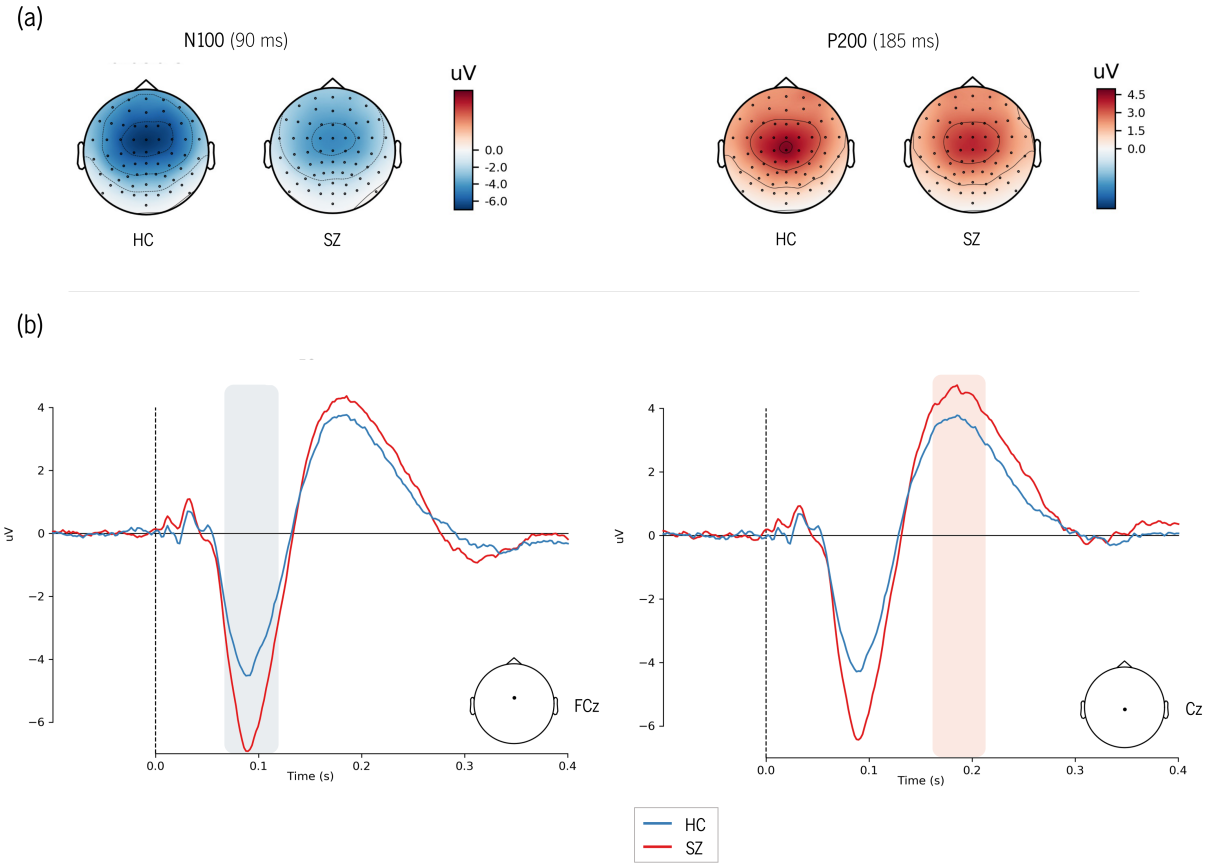


Figure 22: (a) Topographical maps for amplitude distribution at 90 ms (N100 peak - top) and 185 ms (P200 peak - bottom) for HC and SZ groups; (b) Grand average ERP waveforms for HC and SZ for FCz (top) and Cz (bottom) electrodes, with the shaded time intervals where the maximum peak values occurred for each ERP component.

Figure 22b shows the grand averages at the FCz and Cz electrodes, highlighting the time interval where the N100 and the P200 peaks occurred. Figure 22 shows no differences in the spatial distribution of N100 and P200 amplitudes between the two groups. However, differences are visible in the amplitude of the two components, with higher amplitudes (in the case of N100, more negative values) for the HC than SZ.

4.3.2 Trial-by-trial Analysis

Figure 23 shows the graphical representation of a single trial using maps, where each line (y-axis) corresponds to the plot of a trial over time (x-axis), and the color represents the magnitude of the signal. These single trials clearly depict the temporal alignment of the wave peaks corresponding to the ERP N100 and P200. This pattern of evoked responses is visible by the vertically oriented bands around the 100

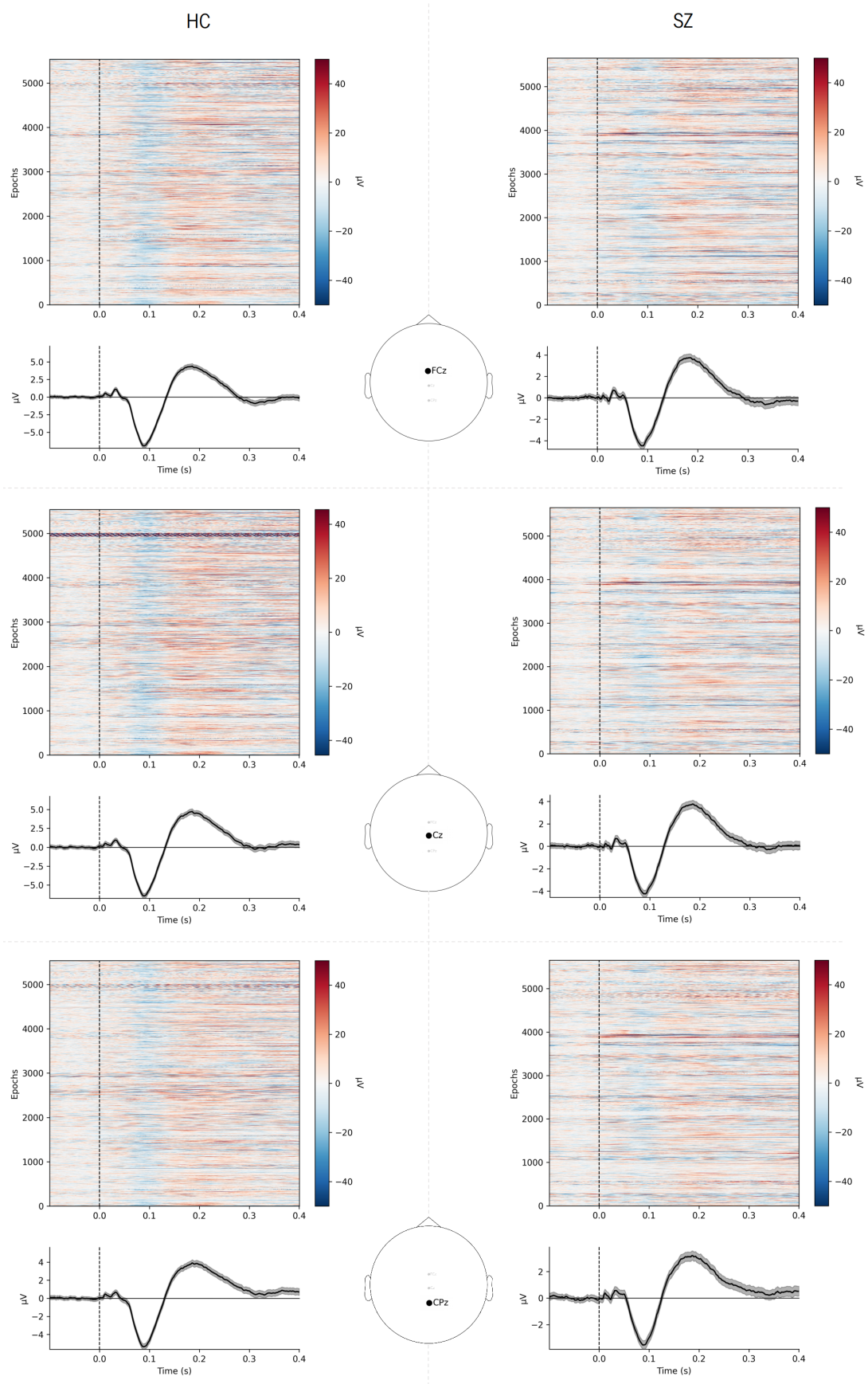


Figure 23: Single-trial maps and the corresponding average evoked response below each map, plotted for the FCz, Cz, and CPz scalp channels, comparing HC to SZ responses.

ms and 200 ms post-stimulus latencies, which correspond to larger signal amplitudes in absolute value. Although only the midline EEG channels FCz, Cz, and CPz are shown in the figure, these bands appear equally distributed over the scalp. They are well-defined in the HC group's maps, while the responses in SZ are less demarcated, with more significant amplitude fluctuations.

Difference waves were used for an overview of the contrasts between the signals obtained from HC and SZ patients. This approach consisted of point-to-point subtraction of the average amplitude values of all trials from each group, performed for all EEG channels over the scalp. Figure 24 presents the difference waves between HC and SZ averages through a butterfly plot. Figure 24 also plots the amplitude topographic maps at three local peaks latency, which is detected automatically using the MNE-python *plot_joint* function. The analysis of this figure indicates that the most significant difference occurs at 96 ms, with more negative amplitudes for HC compared to patients. This latency is close to the N100 latency in the ERP analysis (Table 8) and can reflect the amplitude differences visible in Figure 22. This difference is quite robust and consistent across all EEG channels. Other less prominent differences are displayed and occur at later latencies, coincident with the end of the P200 deflection shown in Figure 22. However, this approach does not allow practical conclusions since plots resulted from summing underlying source single-trial waveforms. Moreover, difference waves may reflect latency differences rather than amplitude differences between groups.

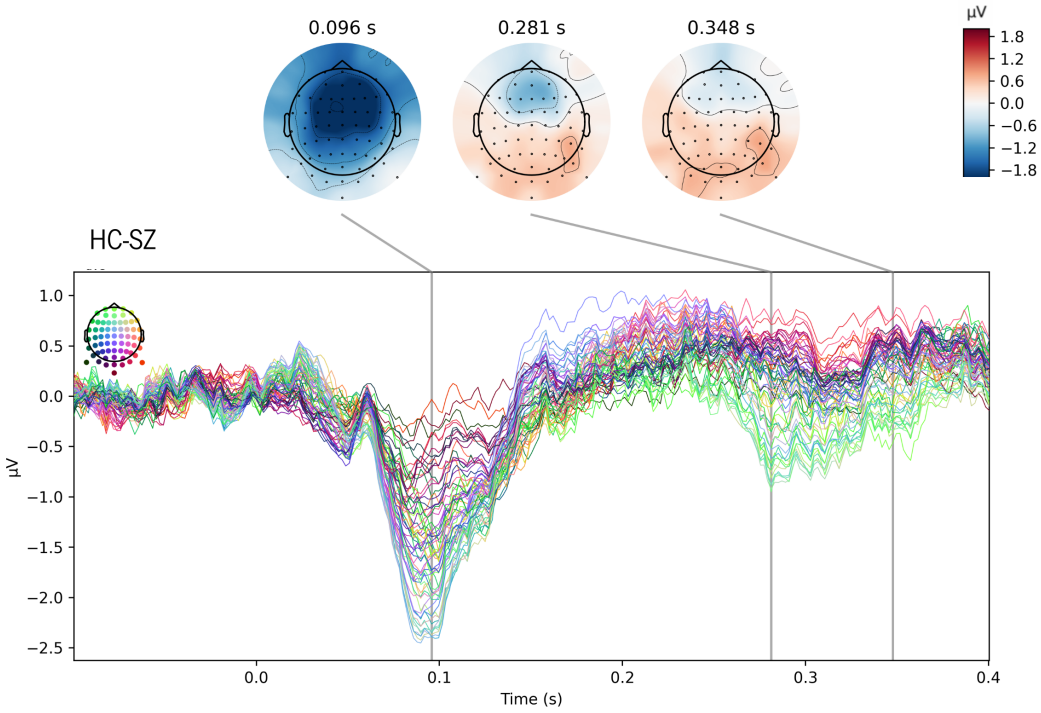


Figure 24: Butterfly plot illustrating the difference waveforms between the average evoked response of all HC and SZ single-trials plotted for the 64 EEG channels. Topographical maps complement the butterfly plot showing the amplitude distribution over the scalp at the time points of the three local peaks detected automatically.

The cluster-based permutation test detected two clusters in the single-trials EEG data: one that happens around 100 ms and is spread over all scalp regions; and another occurring between 200 and 300 ms

in the midline and right hemisphere centroparietal channels. Figure 25 and Figure 26 illustrate the spatial and temporal distribution of the two clusters, which have a probability of $p < .001$. The results allow us to conclude that the null hypothesis of the exchangeability of the data over the two groups of subjects is unlikely. Thus, the alternative hypothesis that single-trials differ between HC and SZ subjects is accepted.

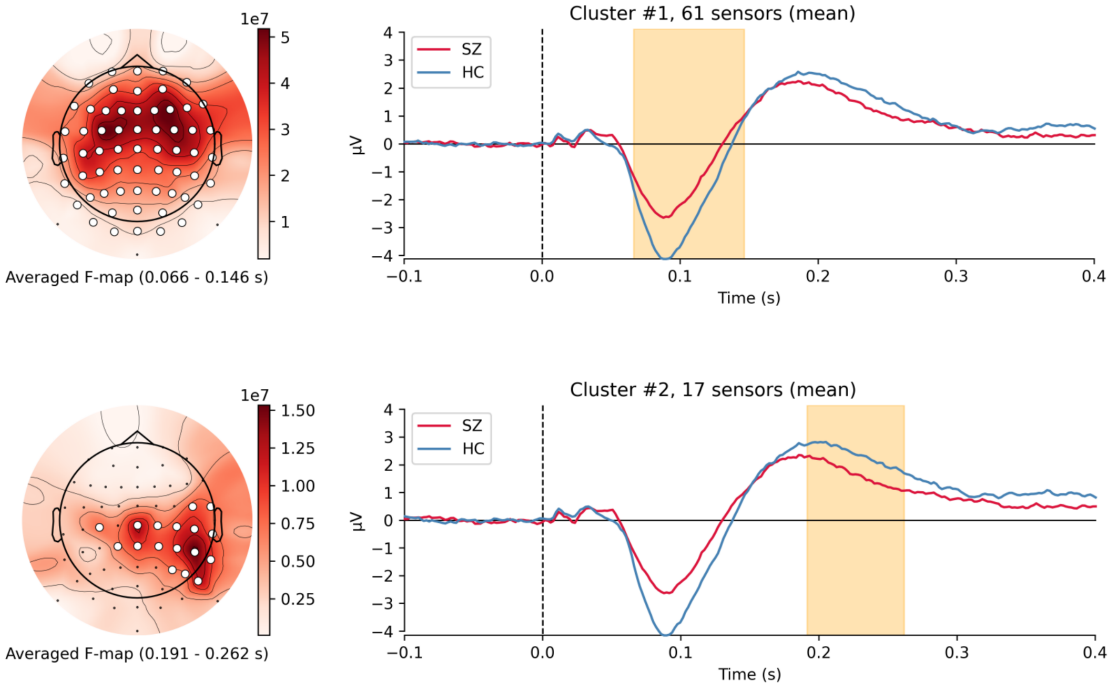


Figure 25: Visualization of the two significant spatio-temporal clusters. On the left side, the topography plot for mean F-statistic with significant sensors highlighted. On the right side, the time course of each significant spatial cluster with the extent of the temporal cluster shaded in amber tones.

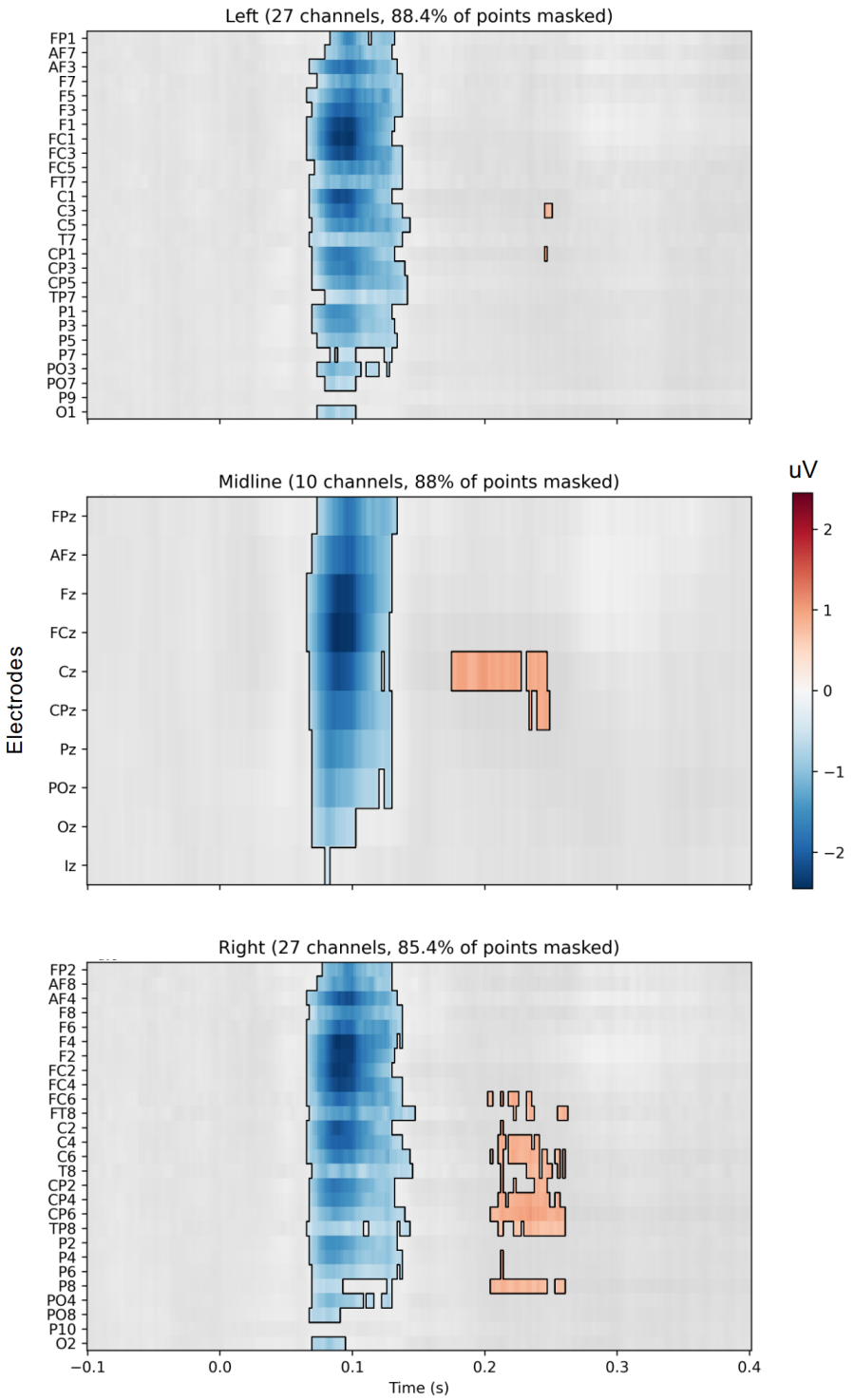


Figure 26: Average of the difference waveforms plotted for each electrode, masking time points for significance. Scalp channels are grouped by regions of interest: left hemisphere, midline region, and right hemisphere, from top to bottom, respectively.

4.3.3 Time-frequency Analysis

PLF measures the consistency of the phase relationship between trials across time and frequency, which can be interpreted as a measure of synchrony of neural activity. Figure 27 presents the time-frequency maps of PLF values averaged over all single-trials of the HC and SZ groups at the FCz electrode. Topographies of the average PLF values in time windows whose synchrony was most pronounced in each frequency band range complement the figure. Observing these results allows us to conclude that the beta- and gamma-bands are activated early in stimulus processing (in the first 100 ms). The theta- and alpha-bands have higher amplitudes in a longer temporal window (up to 300 ms post-stimulus). These responses occur in both groups, although the PLF amplitude is lower in subjects with SZ. The differences between the two groups are most evident in the theta- and alpha-bands over frontocentral electrodes.

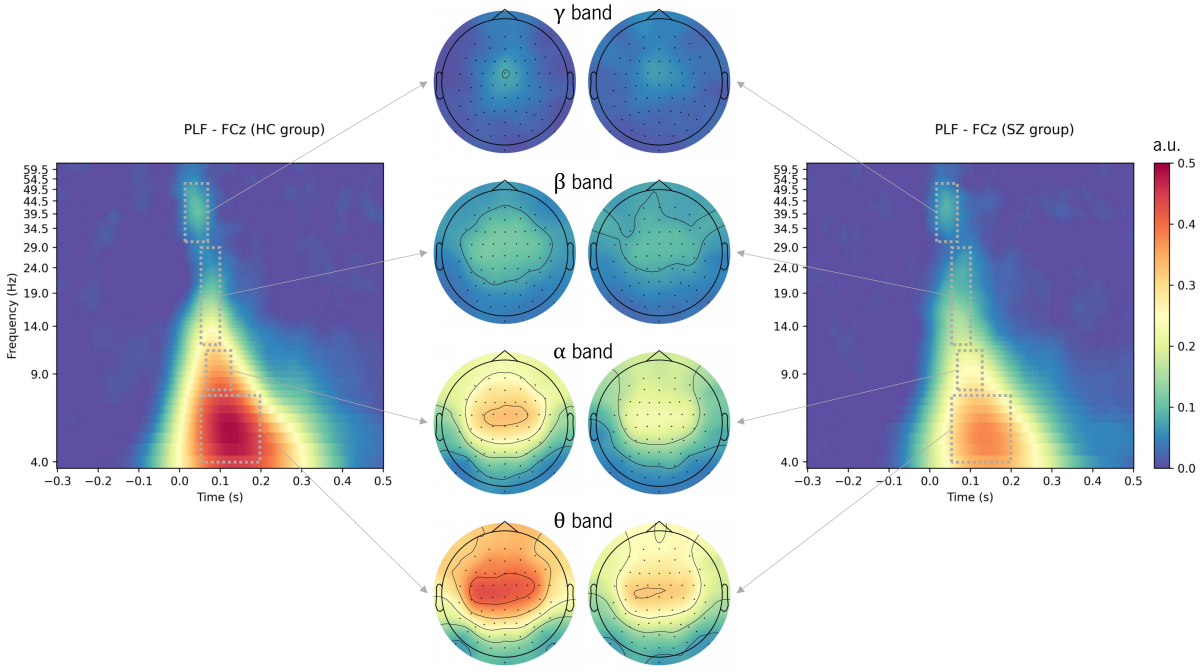
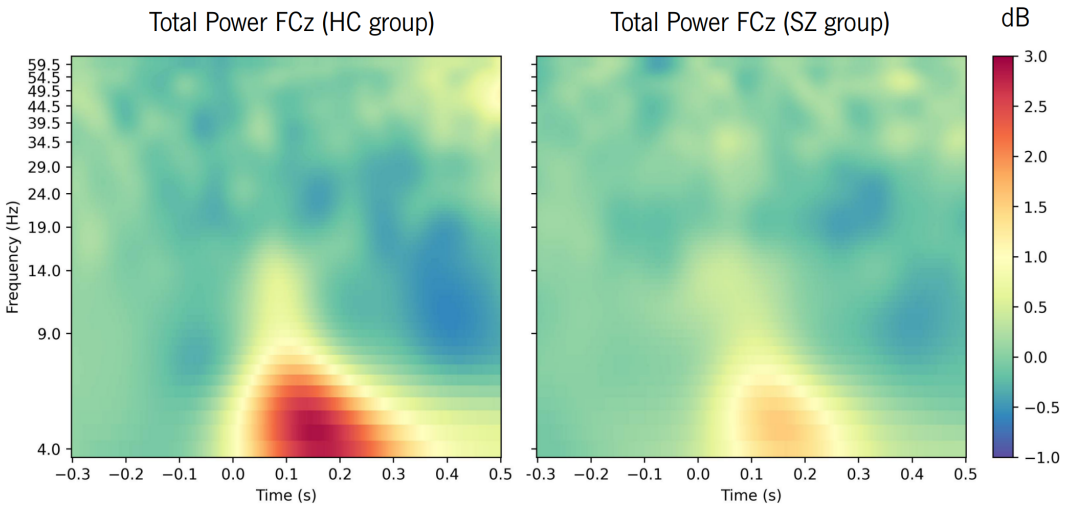


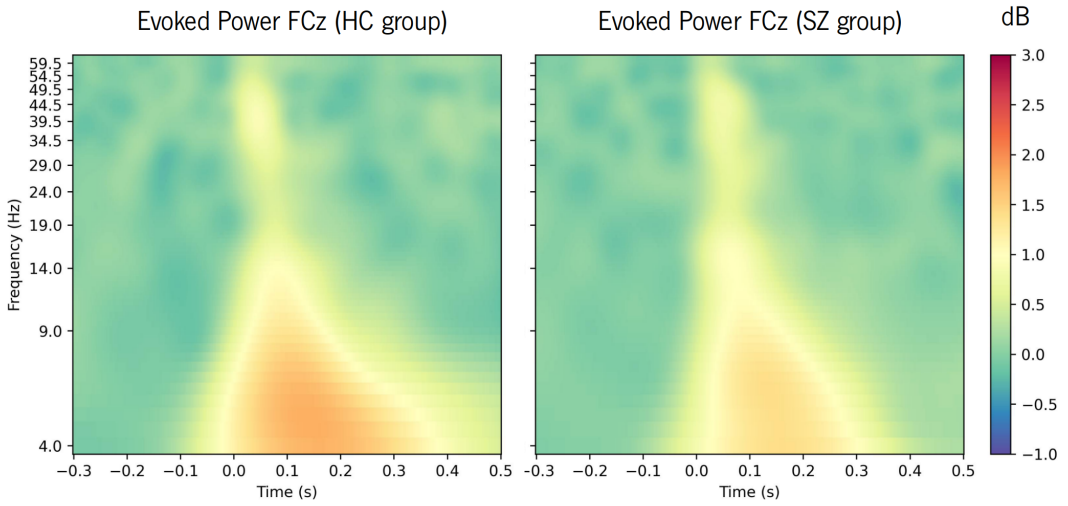
Figure 27: Average PLF time-frequency maps for HC and SZ groups plotted for FCz electrode, and topographical maps showing the spatial distribution of average PLF values for time windows (marked on each of the maps by dashes) with more prominent activity in each of the frequency bands.

Power measurement provides information about the overall strength of neural activity at different frequencies over time [434]. Changes in total power can indicate differences in neural activity in response to auditory stimuli between HC and SZ subjects. In turn, evoked power refers to the component of total power that is time-locked to a specific event (in this case, a pure tone) and reveals differences in the strength and timing of neural responses [127]. Evoked power is often contrasted with induced power, which refers to the component of total power that is not time-locked to a specific event but reflects ongoing neural activity. Figure 28a presents the total power, Figure 28b the evoked power, and Figure 28c the induced power time-frequency maps averaged over all single-trials of each group. Differences in power measures are remarkable between the two groups, with SZ signals showing weaker responses than HC.

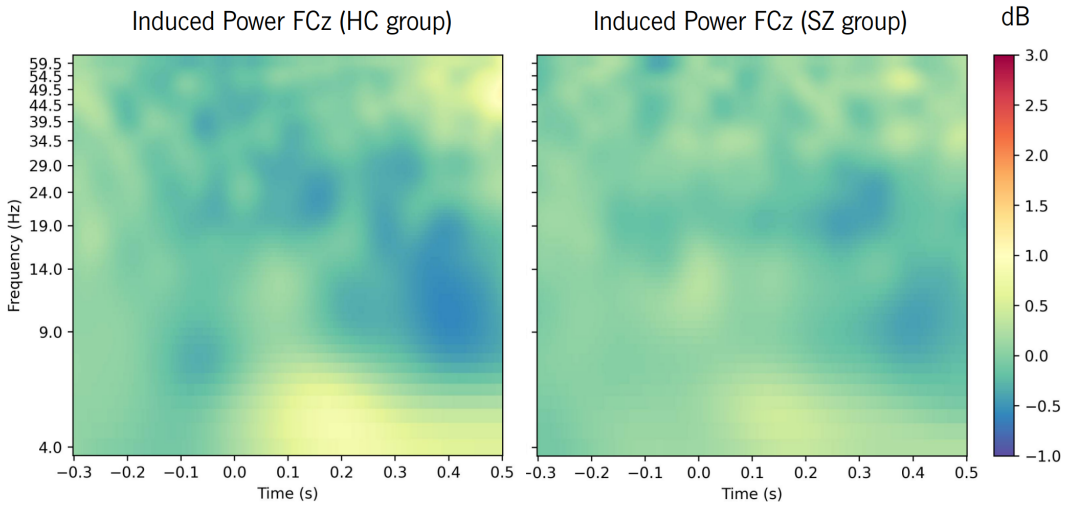
Difference time-frequency maps were computed and plotted to globally perceive the contrasts between



(a)



(b)



(c)

Figure 28: Time-frequency maps of the average values of (a) total power, (b) evoked power, and (c) induced power acquired at the FCz electrode, plotted for the HC and SZ groups.

the two groups' signals. This approach consisted of point-to-point subtraction of the PLF, total, evoked, and induced power values averaged over all trials from each group (HC - SZ). This analysis was performed for the FCz scalp channel. The difference maps are shown in the top row of each panel of Figure 29. The bottom rows of Figure 29's panels present the cluster-based permutation maps.

Analyzing the differences in PLF, Figure 1 confirms that the highest contrast happens in the 0-300 ms time interval in the frequency range from 4 to approximately 15 Hz. This result mirrors the differences in the topographic distribution of PLF amplitudes for the theta and alpha (Figure 27) frequency bands. The cluster-based permutations carried out on the F-test revealed a significant difference between the two groups ($p < .01$) for the time-frequency region masked in the PLF permutation map (bottom row of Figure 29a). We can conclude that neural activity in SZ patients is less synchronous for the mentioned frequency bands during early stages of auditory processing.

Total and evoked difference maps (Figures 29b and 29d top rows) show that the most considerable differences occur in the theta-band in the 0-300 ms time window. In this time-frequency window, power is higher for HC compared to SZ patients. However, the cluster-based permutation test did not reveal any significant difference for total and evoked power in the total spectrum and the analyzed time course. The permutation maps without any mask for significance support this finding (Figures 29b and 29d bottom rows). The differences in induced power ((Figure 29c top row) account for the small intensity fluctuations in the total power map (Figure 29b). The differences shown were not statistically significant (Figure 29c bottom row).

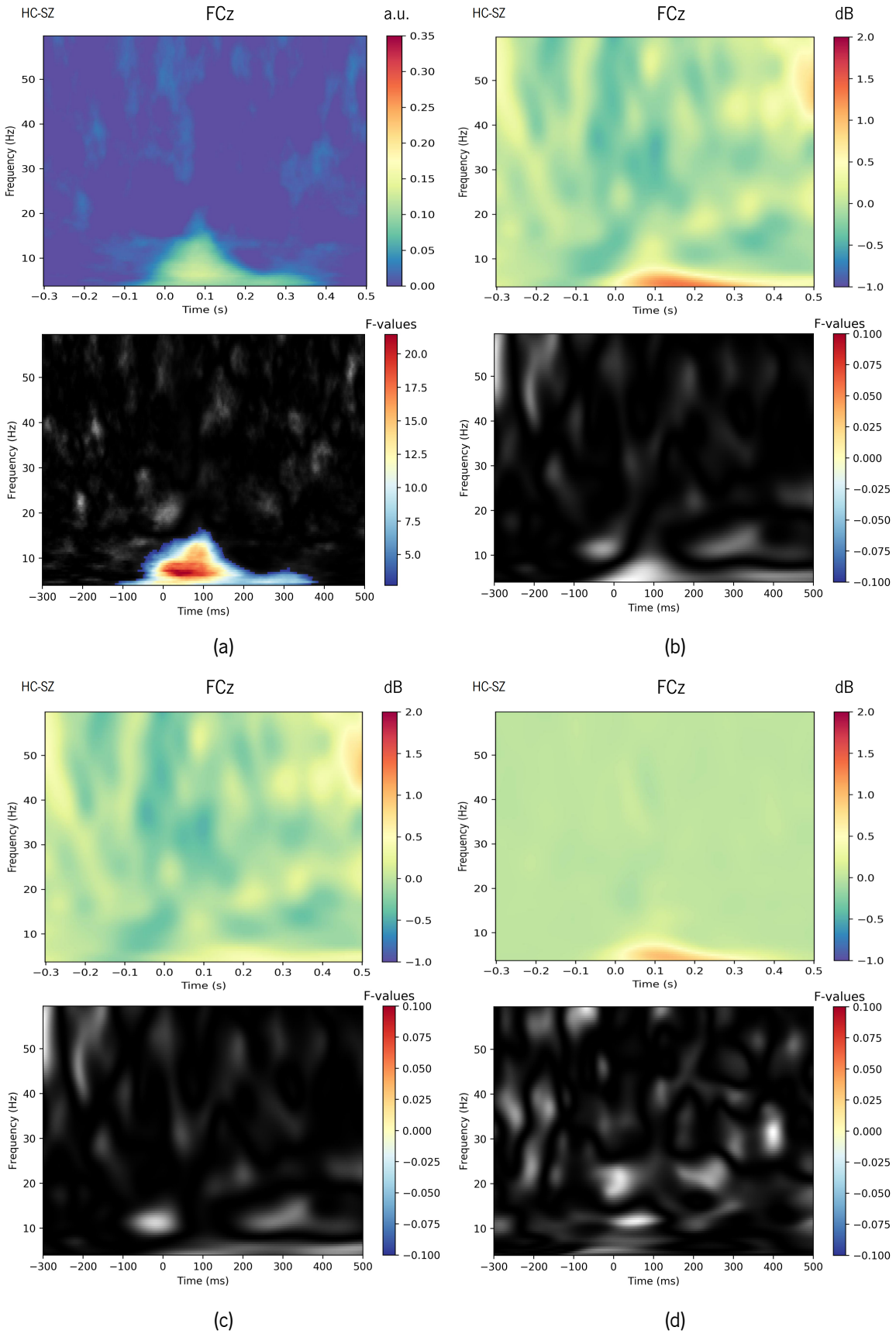


Figure 29: (a) PLF permutations; (b) Total power permutations; (c) Induced power permutations; (d) Evoked power permutations. This time-frequency plot depicts the cluster statistics which were averaged over all channels, resulting in a two-dimensional plot displaying the computed statistics as a function of time and frequency band.

4.4 Discussion

Deficits in sensory processing are now well-established in SZ, not only in auditory processing but also in visual, olfactory, and sensorimotor systems [436]. The idea that sensory dysfunction, most crudely translated as "bottom-up" process disruption, plays a role in overall cognitive impairment in SZ has received increasing support [110]. This study aimed to evaluate the processing of simple auditory stimuli and examine changes in brain electrical signal morphology and oscillatory neural activity in SZ patients compared to HC. The results revealed changes in amplitude at the N100 ERP component, which have been robustly and widely reported in SZ [24, 177]. Changes in the oscillatory dynamics of auditory processing were also observed. A more in-depth and critical analysis of the results is provided in the following subsections. Before discussing the functional significant of the current findings, we first highlight some of the potentialities of the analyses performed in the current study,

4.4.1 ERP vs. Trial-by-trial Analysis

The single-trial approach relies on trial-by-trial analysis, which outperform averaging across trials as in traditional ERP analyses [437]. On the one hand, a single-trial analysis may be more sensitive to changes in EEG activity because it can detect subtle differences that the averaging process in ERPs may obscure. This sensitivity may be necessary for studying perceptual and cognitive processes involving rapid and transient changes in neural activity [438]. On the other side, this approach allows more flexibility in the analyses to be conducted on EEG data. Researchers can examine other signal characteristics, such as spectral content, and conduct different statistical analyses, such as classification or clustering [428, 439]. EEG single-trials are particularly important for machine learning classification algorithms, as they provide a more diverse and representative dataset for training and testing. This can lead to better classification accuracy, particularly in tasks requiring high precision levels, such as detecting specific brain states or cognitive processes [437]. However, the single-trial analysis may require more sophisticated statistical methods and larger sample sizes to account for the increased variability in the data [432].

In addition to a qualitative analysis of the ERPs, we examined the data following a trial-by-trial approach in this study. This analysis allowed us to consider inter-trial variability, which is lost during the averaging process in ERP. While intertrial variability is often considered a source of noise in EEG studies, it is also a valuable source of information. Inter-trial variability can be used to investigate the stability of EEG measures over time, such as the reliability of N100 or P200 ERP components or the consistency of neural activity across repeated presentations of the same stimulus. This information can help assess the validity of EEG measures.

4.4.2 Cluster-based Permutation Tests

Neural data distributions may present non-normality and very skewed distributions or heavy tails. Therefore, using non-parametric tests for statistical inference is recommended [440]. The cluster-based

permutation test uses a permutation or randomization test to approximate the probability distribution of a non-parametric statistic, which makes them more robust to violations of assumptions such as normality [441, 442]. Cluster-based permutation tests effectively control for multiple comparisons while preserving statistical power. Since cluster-level statistics use the accumulated statistical power of all points in a cluster, the test's sensitivity is higher [443]. This approach allows the identification of large and spatially extended clusters of electrodes that show significant differences between experimental conditions [442]. Cluster-based permutation tests are a powerful and widely used method for analyzing EEG data in neuroscience research. Compared to other statistical methods, they can provide increased sensitivity, robustness, and reliability. For these reasons, we adopted a non-parametric cluster-based approach to statistically explore differences in neural activity between the two study groups.

4.4.3 Event-related EEG Analysis

Several EEG studies have shown that individuals with SZ have deficits in auditory processing. Specifically, these studies have reported differences in the N100 and P200 ERP component features [24, 177, 25].

The results obtained from evoked potential analysis (subsection 4.3.1) show that the N100 and P200 components exhibit reduced amplitudes, confirming the ERP changes and the results consistently reported for SZ patients [24, 177, 25]. Both the spatial distribution of amplitude and the ERP pattern found in the grand averages replicate prior literature [414, 185, 177].

The trial-by-trial analysis provided critical information on the consistency of brain responses to (repeated) simple sounds. Figure 23 shows the inter-trial variability and the consistency in the emergence of the N100 negativity (blue vertical band occurring around 100ms) in all trials for all subjects. The consistency and amplitude of the N100 are higher in HC than in SZ patients. In turn, the positivity corresponding to the P200 component varies in either amplitude or latency in both groups. However, it is noticeable that the amplitude of the response around 200 ms is more robust in the HC group (which is visible by the redder vertical band). Notably, the more significant variability in amplitude and latency is a major feature of single-trials in the SZ group, which may underlie the changes found at the ERP level. Indeed, enhanced trial-by-trial fluctuations of neural activity in sensory and higher-order cortical areas have been consistently observed in SZ [23].

The cluster-based permutation analysis allowed us to examine spatiotemporal differences between the two groups. Two spatiotemporal clusters showed significant differences between HC and SZ. The first cluster coincides temporally with the N100 component, and the differences are significant throughout the scalp but more prominent over frontocentral electrodes. Prior studies documented reduced N100 amplitudes in SZ patients, particularly over frontal and central scalp electrodes compared to healthy controls [177]. These alterations were observed across various auditory tasks, including simple tone discrimination [444]. Studies testing whether the N100 reduction is lateralized have reported mixed findings. Some studies suggest that SZ patients may have reduced left-lateralized N100 responses to pure tone stimuli [417, 445, 418]. However, this was not verified in the statistical analysis that we conducted. In our study, which

included a larger sample size compared to the majority of previous studies, the absence of lateralization in the N100 response to pure tone stimuli in individuals with SZ suggests that the observed findings may provide a more robust understanding of the heterogeneity and complex nature of auditory processing abnormalities in this population.

The second cluster shows differences in the amplitude of the signals in the descending phase of P200. These differences are distributed over central and right parietotemporal electrodes. The right parietotemporal region is known to play an important role in auditory processing, particularly in the processing of acoustic sound features [446]. Puschmann *et al.* suggested that the right temporoparietal junction is involved in controlling attention during selective listening [447]. While brain activity is a complex process involving multiple regions, and the P200 component may involve contributions from other brain areas [448, 25, 449], it is reasonable to speculate that the P200 amplitude recorded at the right temporoparietal electrodes captures relevant activity from the right temporoparietal region. Consequently, the evidence of altered neural activity highlighted by the second cluster suggests that the decreased P200 response in SZ may reflect a reduced ability to automatically allocate attentional resources to incoming auditory stimuli and, consequently, an increased difficulty in processing and interpreting auditory information. Regardless, further research is needed to fully understand its neural underpinnings. Both clustering outcomes indicate impaired top-down and bottom-up auditory sensory-perceptual deficits in the SZ disorder.

The results presented here support our hypothesis that single-trial analysis can greatly improve the discrimination of the two groups and suggest that temporal variability plays a significant and complex role in the auditory processing deficits observed in SZ patients.

4.4.4 Changes in Oscillatory Activity

The present study evaluated the evoked (phase-locked), induced (non-phase-locked), and total oscillatory responses to auditory stimuli in a passive listening paradigm for the wide frequency range of the EEG recordings, including theta-, alpha-, beta-, and gamma-bands, between HC and SZ groups. Due to the high computational load demanded by a multi-channel approach, this study focused only on analyzing signal changes in the FCz electrode. The frontocentral electrode was selected in line with previous research, showing that frontocentral activity reflects auditory processing in specific auditory paradigms [450, 451]. In addition, many EEG studies have highlighted atypical frontocentral activity in SZ [452, 453, 454, 455, 171], and evidence for structural alterations in frontal regions in SZ are widespread [456].

In this single-channel approach, both groups showed increases in total, evoked, and induced power in the theta-band in the first 300 ms post-stimulus. Theta-band has been described as being involved in orchestrating several cognitive processes such as working memory or allocating attentive resources [457, 458]. The phase-locked increase in theta power is typically observed broadly between 0 and 300 ms after stimulus onset and characterizes electrophysiological responses to auditory stimuli [459]. Research studies have reported altered theta power in SZ during passive listening tasks [459]. In our study, the theta power in the post-stimulus period was decreased for the SZ group, which aligns with previous studies focused on lower-frequency ranges [23]. Those studies also showed that SZ patients had increased

levels of the ongoing power of theta oscillations in auditory N100 paradigms [23, 410]. Roach *et al.* suggested that SZ N100 may derive from the less intense evoked responses and be potentially obfuscated by induced or other ongoing EEG power [410]. Although induced activity is lower for SZ than HC, the impact of this perturbation is not as evident in our results. More recently, theta activity has been reported to be impaired in SZ, possibly due to connectivity network disruption at different scales, reflected in subpar performance [460]. This approach suggests that this functional disconnection is a central pathophysiological feature of SZ, in which theta oscillations seem to be strongly implicated [460]. Although consistent with the results reported in the literature, we do not find any significant clusters from the permutation-based analysis.

On the other hand, when statistically analyzing the PLF values, we found a significant difference in the synchrony of the theta- and alpha-bands during the first 100 ms of the auditory sensory processing. This result suggests greater neural desynchronization in patients with SZ, which may lead to lower N100 amplitudes. The phase-desynchronization of alpha activity over frontal regions may reflect altered arousal or impaired attentional process via changes in thalamocortical and cortical network communication [461, 462]. Increased frontal theta activity is generally associated with information processing and error monitoring [463]. Our results suggest that decreased neural activity synchronization in theta- and alpha-bands reflects impairments in attentional resource allocation and sound processing. The alterations and variations in the strength of neural activity, reflected in fluctuations in power measurements, are responsible for the altered ERP signature in SZ, especially the modulation of the N100 ERP component. The lack of theta synchronization in a later period (200-300 ms post-stimulus) in SZ signals may mirror dysfunctional attentional networks [460], which may be responsible for changes in the P200 component. Some recent studies suggest that the intrinsic impairments in generating theta rhythms may be a marker of cognitive dysfunction in SZ [464]. Neural oscillations and their synchronization may be core pathophysiological mechanisms in SZ, as they are fundamental for coordinated neural activity during normal brain functioning [350].

Our time-frequency analysis results show no evidence of changes at the gamma-band level in SZ when performing a passive single-tone listening task. Even though prior work has proposed that alterations in gamma oscillations represent a biomarker of SZ [350, 465, 454, 466, 395], results have not been consistent [317]. Inter-trial variability may explain the discrepant results, but other factors such as prolonged medication, illness progression, and increased auditory thresholds in chronic schizophrenic populations may override EAGBR measurements [317].

4.5 Limitations and Future directions

There are some limitations to this study, which should motivate future research. The first limitation is the inclusion of patients with different subtypes and symptom profiles that can lead to variability in the results. The merged dataset includes early illness as well as chronic patient data. Moreover, we did not control for antipsychotic medication and dosages among participants, which made it impossible to dissociate the effects of the drugs on the abnormalities found [467]. Another limitation is the non-

controlled comorbidity. SZ is often associated with comorbid conditions, such as depression, anxiety, and substance abuse, which can affect ERP and other EEG measures [374].

The need for a large amount of data for developing and implementing deep learning methods for SZ classification motivated the merging of two datasets, as well as the need to balance the existing public database. However, the acoustic stimuli used to acquire both datasets have different pitch and intensity properties. These properties have been reported to affect N100 amplitude and latency [468]. This evidence motivated us to conduct a statistical analysis. However, we found no significant differences between both datasets. These findings should, nonetheless, be interpreted with caution.

ERP and EEG measures are often sensitive to task demands and task performance [127]. Individuals with SZ may have difficulties with task performance [469], which can affect and complicate the interpretation of results. We tried to mitigate this limitation by choosing a passive hearing condition with simple sound presentation. However, this condition was part of a longer and more complex paradigm, and therefore the performance of each participant may have been affected by fatigue.

Another restriction has been mentioned before. Due to computer resource limitations, performing a spatiotemporal analysis of the time-frequency domain measurements was unattainable. This limitation prevents us from understanding if there is any pattern in the distribution of the time-frequency measures across the scalp and discussing or drawing conclusions.

4.6 Summary

Sensory and perceptual processing dysfunction is important predictor of SZ or psychosis risk [392, 470]. EEG-derived ERP results have supported the hypothesis that altered sensory processes may relate to clinical symptoms and cognitive and functional outcomes in SZ [392]. This study probed auditory information processing in SZ patients and their respective HC. ERP, trial-by-trial, and time-frequency analyses were performed on EEG data from 63 HC and 65 SZ subjects. The results show that the N100 ERP component is reduced in SZ, most prominently over frontal-central electrode sites, as consistently reported in the literature [24, 177]. The trial-by-trial analysis evidenced the consistency of N100 and P200 deflections across all trials and greater variability in SZ signals. In addition, significant differences were found in N100 and P200 temporal windows, reinforcing the robustness of these event-related components. From the time-frequency analysis, we concluded that differences in neural activity synchrony of theta and alpha oscillations are responsible for different EEG signatures between HC and SZ groups in response to simple sounds. All analyses showed the robustness of the reduced responses in SZ, which improves the separation between groups and motivates the development and implementation of a single-trial-based SZ classification model.

Chapter 5

Decoding auditory processing in schizophrenia using deep learning

This chapter presents a deep learning model for SZ classification based on auditory EEG data. Auditory processing impairments have been studied using EEG-derived ERP data, and they have been associated with clinical symptoms and cognitive dysfunction in SZ. This research provides insights about altered patterns in electrical brain activity during auditory processing and their potential to discriminate SZ and healthy subjects. The proposed model using deep convolutional neural networks performs the classification based on multi-channels auditory-related EEG single-trials recorded during a passive listening task.

Contents

5.1 Introduction **91**
5.2 Methodology **93**
5.3 Experimental Setup **99**
5.4 Results and Discussion **106**
5.5 Summary **117**

This chapter is based on the publication:

Barros, C., Roach, B., Ford, J. M., Pinheiro, A. P., & Silva, C. A. (2022). From Sound Perception to Automatic Detection of Schizophrenia: An EEG-Based Deep Learning Approach. *Frontiers in Psychiatry*, 12, 813460. /doi.org/10.3389/fpsyt.2021.813460

5.1 Introduction

SZ is a chronic and complex brain disorder that affects social and cognitive functioning [2]. SZ is characterized by the presence of positive (e.g., hallucinations and delusions) and negative symptoms (e.g., blunted affect), as well as cognitive deficits [9]. About 75% of patients experience hallucinations in the auditory modality, most frequently as voices [88]. Deficits in auditory processing have frequently been reported in SZ, which may reflect auditory cortex pathology [23]. Some studies have documented larger deficits in patients with (vs. without) auditory hallucinations [12].

Several studies aimed to probe the neural mechanisms underpinning auditory processing abnormalities in SZ, using different techniques such as EEG. The ERP represent the averaged electrical activity elicited in response to a stimulus. They provide a suitable method to specify the time course of brain activity in response to auditory stimulation [18]. Alterations in ERP components have been consistently documented in SZ, which may be considered potential biomarkers of this disease [20]. The MMN and P300 amplitudes are robustly attenuated in SZ, and have been proposed as SZ biomarkers [21, 22]. In turn, sensory gating studies have shown reduced P50 suppression in SZ patients in auditory paired-stimulus paradigms [158, 188, 159, 160]. Other studies have focused on the N100 ERP component, specifically showing generalized amplitude reductions to sounds in SZ. This reduction seems to be more pronounced during the experience of auditory hallucinations [24, 471].

Deep learning applied to EEG data could represent a promising contribution to a more accurate prediction of intervention response and disease trajectories [37]. EEG-based deep learning algorithms have also seen growing interest among neuroscientists, especially in the context of brain-signal decoding. However, deep learning is still a poorly explored method in the development of EEG-based systems in SZ research. Convolutional networks have been implemented with the goal of recognizing and classifying patterns from multivariate time series [230]. CNNs have been applied to a variety of EEG analysis tasks, including ERP detection, motor imagery classification, seizure detection, and sleep stage classification, showcasing their efficacy in capturing relevant patterns from EEG signals and advancing the fields of brain-computer interfaces, clinical diagnostics, and neuroscience research [52, 472, 473, 53]. The success of this type of neural network applied to EEG data for decoding purposes has prompted us to investigate the use of CNN for SZ classification based on auditory perception.

5.1.1 Related Work

SZ classification using early auditory EEG-derived ERP components has already been attempted through classical machine learning models [46, 47, 48]. Components such as the P300, MMN, or N100 were mainly elicited with auditory oddball and passive listening paradigms and used as input features in SZ recognition with RF, SVM or LDA classifiers. The findings of these studies highlight the significant potential of auditory ERP components, which have been recurrently proposed as biomarkers for SZ, as features to discriminate patients from healthy subjects. A limitation of these approaches relates to the process of features extraction and selection, which requires domain expertise. The few studies that probed the potential

of deep learning in EEG-based classification of SZ [362, 358, 360, 361, 359] achieved the best performances with CNN-based models applied to resting-state EEG data, which is independent of cognitive or sensory processing. Despite their capacity to discriminate healthy from SZ subjects, these models do not inform about auditory processing, which is affected in SZ. Very recently, Ahmedt-Aristizabal *et al.* [348] explored both machine and deep learning techniques to identify children at risk of SZ on the basis of EEG data collected during a passive auditory oddball task. In the classical machine learning approach, the mean amplitude was extracted in the 80-220 ms and 160-290 ms latency intervals, when ERP components indexing the early stage of sensory processing (N100 and P200, respectively) are expected to emerge. The mean values were extracted for 5 midline electrodes: Fz, FCz, Cz, CPz, and Pz. Using common classifiers, such as DTs, kNN, and SVM, the discrimination between healthy and at-risk children based on those features was unsuccessful. For the deep learning approach, a 2D-CNN-LSTM was proposed, composed of one 2D convolutional layer, followed by normalization and fully-connected layers. The information from the previous block was processed with a stack of two LSTM networks, whose output was transformed with sigmoid nonlinearity for classification purposes. This model based on EEG single-trials achieved the best performance in at-risk children identification. For each trial, a spatio-temporal 2D signal was created using a 300 ms post-stimulus window for the 5 midline electrodes. The machine learning attempt illustrates the difficulty in specifying stimulus-related signal features that allow the identification of SZ risk. The results of both approaches may reflect the developmental phase of the population under study, namely ongoing developmental brain maturation processes. Notwithstanding, this study demonstrates the potential of deep learning methods in subjects' discrimination as a function of psychosis risk.

Evidence for altered auditory processing in SZ has fostered the investigation of the dynamics of electrical brain activity targeting the differentiation between patients and healthy subjects. Amplitude reduction of auditory evoked potentials such as the N100 have been reported consistently in the literature [24, 314]. Those alterations have driven the use of machine learning methods for automatic SZ recognition. Beyond the time-consuming features extraction, both machine learning models and ERP analysis exhibit a major limitation: the non-uniformity of the time windows and electrodes used for feature selection across studies. Deep learning methods alleviate the need for manual feature engineering. However, the informed decision about electrode selection and time-windows remains crucial to ensure that the model effectively captures the relevant neural activity. These choices should be driven by domain expertise and a comprehensive understanding of the research question [474]. Although deep learning architectures based on EEG signals have been proposed for SZ classification, the learning of patterns from the electrical brain response to auditory stimuli is a scarcely investigated topic. Chapter 3 provides a critical analysis of deep learning and classical machine learning methods to detect SZ based on EEG signals highlighting the potentialities of these methods in clinical research. Notwithstanding, it is evident from this analysis that further studies are needed to address the limitations of existing research and advance our understanding in this field. The current work intends to assess the potential of deep models to learn discriminatory EEG patterns in the early stages of auditory processing, which may inform about the significance of sensory changes to SZ diagnosis and prognosis. This work followed the good practices for the development and implementation of machine learning methods proposed in Chapter 3.

This chapter presents a multi-channel deep convolutional neural network for SZ and HC single-trial EEG classification across subjects. The CNN-based architecture is proposed for trial-wise decoding of EEG signals during external auditory stimulation. This chapter is organized as follows. The methodology used for data acquisition and the SzNet architecture are detailed in section 5.2. Section 5.3 describes the dataset, the evaluation procedure, and the experimental setup. In section 5.4, the results, limitions, and future directions are described and discussed.

5.2 Methodology

The *auditory-only* described in Chapter 4, a simple listening task, was used to elicit the auditory evoked potentials, which shed light on the time course of auditory processing. The signal dynamics elicited by auditory stimuli was presented also in Chapter 4, showing the expected pattern in both control and SZ subjects. This chapter begins by describing the most relevant information about data acquisition, which is retrieved from the previous chapter. Afterward, the problem is formulated and the proposed deep learning architecture is presented in detail. This architecture aims to extract information from the time course of the EEG signal and its topographical distribution. The main objective is to obtain a model capable of recognizing patterns in the electrical brain activity underlying auditory processing in SZ. Lastly, the traditional machine learning approaches commonly found in other proposals are introduced. The implementation of these algorithms provides a strong base model for a proper analysis of the specific advantages of deep learning (over machine learning) methods in predicting whether an individual suffers from SZ or not, on the basis of auditory processing alterations.

5.2.1 EEG Data Acquisition

EEG data were recorded using 64 electrodes according to the 10-10 international system configuration, while subjects performed an auditory task (see Figure 30A). During data acquisition, the EEG signals were obtained over electrodes placed on the scalp of the participant, who were asked to listen passively to the sounds. The task performed allows examining the response of the auditory system to externally generated sounds. After EEG signal acquisition, offline pre-processing was performed: the signal was filtered, segmented, and carefully inspected to remove potential artifacts.

5.2.1.1 Auditory ERP components

The N100 and the P200 components are typically elicited in response to sound onset. The N100 can be elicited by any discernible auditory stimulus irrespective of task demands [24]. This negative potential typically occurs between 80 and 120 milliseconds after sound onset, with maximal amplitude over fronto-central and central electrodes [18, 475]. The P200 is a positive deflection that occurs approximately 200 ms after sound onset [18]. This component is distributed over centro-parietal electrodes. Its amplitude is generally maximal over the vertex [18, 25]. Both N100 and P200 auditory components measured

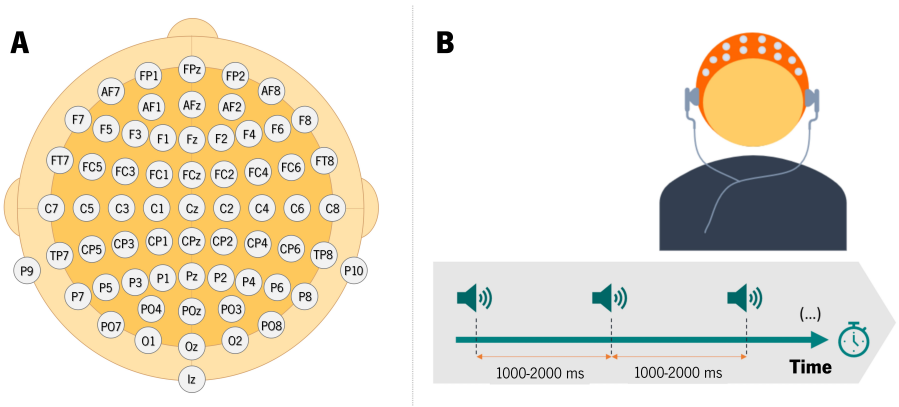


Figure 30: A) Illustration of electrodes placement over the scalp for EEG data recording; B) Schematic illustration of the auditory-only condition of the experimental paradigm.

during passive listening tasks can reflect early automatic attention allocation and stimulus categorization, respectively [173]. By focusing on the N100 and P200 components, which have consistently shown alterations in individuals with psychosis according to the studies discussed in Chapter 4 [24, 25], we employed these early auditory processing indices to accurately discriminate HC and SZ subjects.

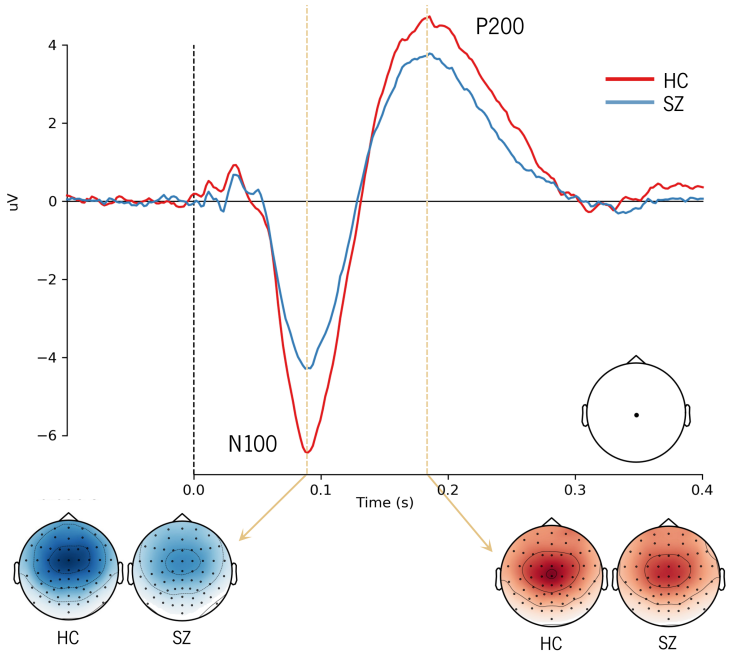


Figure 31: Grand average ERP waveforms showing the N100 and P200 responses to external sounds over Cz electrode in healthy subjects and SZ patients. The topographic distribution of the N100 and P200 components is also shown in both groups.

When conducting the traditional ERP analysis, the background activity (unrelated to the stimulus) is faded away and the N100 and P200 components emerge. The typical grand averages waveforms obtained from the ERP analysis (presented and discussed in Chapter 4) are illustrated in Figure 31, showing a reduction in the amplitude of the N100 and P200 in SZ. Although the main assumption of signal averaging is that the EEG signal in each trial has stable characteristics, such as morphology, the truth is that ERP

averaging hides the intertrial variability in latency and amplitude of the underlying components [476], which was evident in the trial-by-trial analysis also presented in Chapter 4. Therefore, the averaged ERP signals may only represent an approximate picture of the neural processes elicited by an auditory stimulus. Thus, we adopted a single-trial approach, considering the above-mentioned trial-to-trial variability.

5.2.2 Problem Formulation

The task of discriminating SZ patients from HC subjects is based on the classification of a multi-channel EEG single-trial recordings in response to an acoustic stimulus.

Let $x(i) = [[x_{11}, ..., x_{1T}], ..., [x_{C1}, ..., x_{CT}]]$ be the EEG recordings from the C scalp electrodes over T time points for the i trial. The subject dataset comprises the EEG signal of all trials, and it can be defined as $S_s = \{x_1, x_2, ..., x_n\}$, being n the number of trials accepted for subject s. The training set $X = (x(1), y(1)), ..., (x(m), y(m))$ consists of input EEG single-trial recording $x(i)$ and label $y(i)$, where $x(i) \in \mathbb{R}^{C \times T}$ and $y(i) \in \{0, 1\}$. The label $y(i) = 1$ means that the corresponding input EEG signal $x(i)$ belongs to a schizophrenic patient. The label $y(i) = 0$ corresponds to HC subject EEG recording. The index m is the number of trials from all the subjects included in the training set, whose order is randomized. The proposed model accepts $x(i)$ as input and produces $z(i)$ as output. The equation 5.1 defines the function that processes the EEG single-trial $x(i)$ from the input layer to the last fully connected layer. The equation 5.1 can be described as:

$$z(i) = f(x(i); \theta) \tag{5.1}$$

where $f(\bullet)$ corresponds to the function and θ the parameters of the proposed neural network.

5.2.3 Deep Learning Architecture

In the current work, a CNN-based deep learning architecture was developed for SZ classification. The model takes the EEG single-trial recordings as input and provides the probability that a given segment comes from a SZ or HC subject. The acronym SzNet will be used hereinafter to refer to the proposed model.

CNNs are able to find patterns by convolving a filter, or kernel, over the data. Depending on the data structure, the convolutional process can occur in different dimensions. A 1D convolution layer creates a kernel that slides over a single dimension. This can be relevant to find correlations between points in the temporal course or topographical distribution of the EEG data. On the other hand, the use of 2D convolutions allows correlating temporal and spatial information, ensuring that patterns are learned from both dimensions. The maximum signal amplitudes can change in their topographical distribution over time. This is illustrated in Figure 31 by differences in the scalp distribution of the auditory N100 (blue) and P200 (red) peaks reported in the literature, with dark colors corresponding to their maximal activity (N100 - fronto-central distribution; P200 - central distribution).

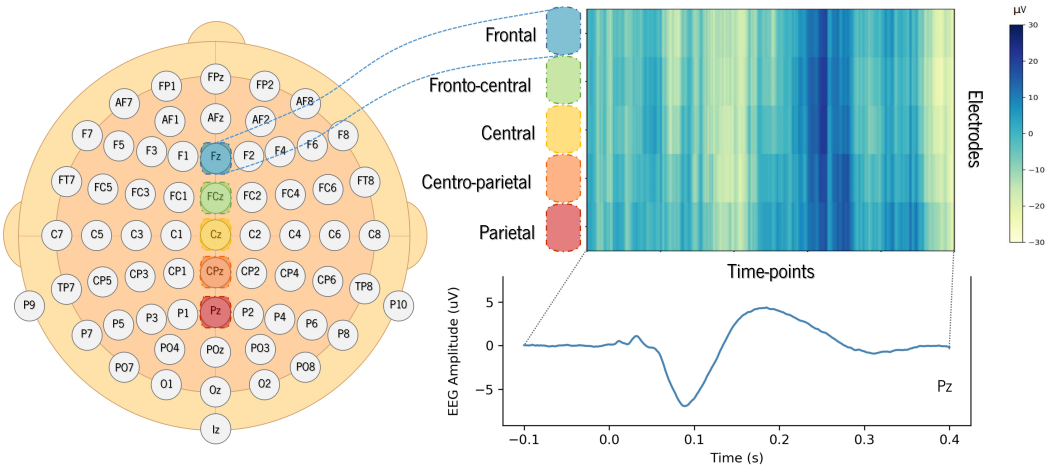


Figure 32: Example of an EEG compounded 2D signal created from one single trial using a set of 5 electrodes. The color scale shows amplitude, in microvolts, over the time points recorded for each of the selected electrodes

Therefore, a 2D convolutions strategy allows extracting information about amplitude variation over time and, simultaneously, across electrodes. In order to take advantage of bidimensional convolutions, we created a 2D structure for each trial by stacking 1D signals captured from midline electrodes over frontal (Fz), fronto-central (FCz), central (Cz), centro-parietal (CPz), and parietal (Pz) regions of the scalp. We obtained spatio-temporal 2D signals for the EEG segments in which the rows correspond to the selected electrodes, and the columns to the time points of the segment window considered, as illustrated in Figure 32. The value of each coordinate corresponds to the amplitude of the signal. The electrodes (rows) were stacked from the frontal to the parietal regions in the images created, so that the network could extract features with functional meaning.

A schematic representation of the main blocks of the proposed method is presented in Figure 33. The first block corresponds to EEG data acquisition and preprocessing, and the second one corresponds to the detailed architecture of the SzNet model. The SzNet architecture is composed of three main types of layers: convolutional, pooling, and Fully-Connected (FC) layers. Those layers are stacked to increase the network depth, enhancing the selectivity and the sensitivity to slightly relevant variations [50]. The learning process of the network is divided into two phases. The first (blue-shaded area in Figure 33) involves learning patterns from the time course of each EEG segment for each electrode. The second (orange-shaded area in Figure 33) aims to correlate those patterns between different electrodes of interest, extracting spatial information. This approach intended to mimic the traditional analysis of this type of data.

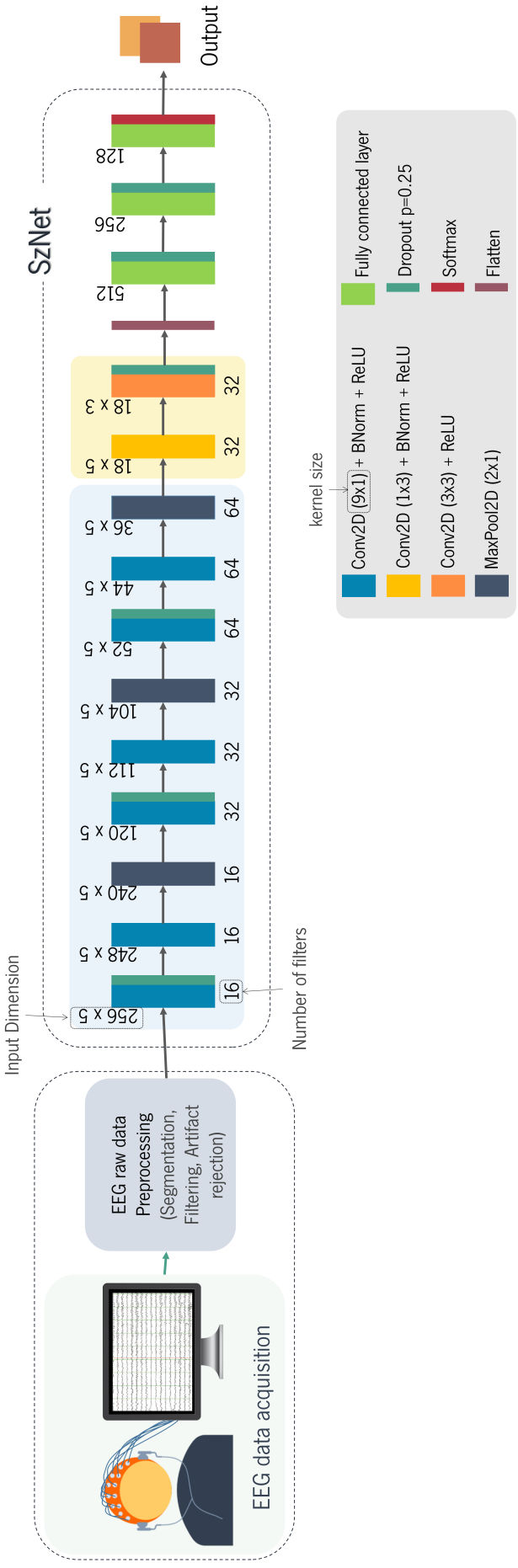


Figure 33: Schematic illustration of the end-to-end process with the first block corresponding to EEG data acquisition and pre-processing. The second block shows the architecture of the SzNet model in detail.

The proposed model consists of a stack of 14 layers of neural network connections. Layers 1, 2, 4, 5, 7, and 8 perform 9-point 1D convolutions (9x1 shaped kernels) over the time dimension. This small-size kernel allows capturing more detailed information about the signal's temporal course. After each convolutional layer, the batch normalization is conducted. In deep learning methods, the output of one layer is the input of the next, and so forth. If the parameters of one layer change, the distribution of the input values of the next layer also changes. This shift in inputs distributions, or covariance shift, can be problematic in deep learning methods with many layers [248]. Batch normalization mitigates the covariate shift ensuring the normalization of the activations of each layer [248]. The batch normalization is followed by the application of the ReLU function. Layers 3, 6, and 9 are subsampling max-pooling layers (MaxPool2D), which also aim to reduce the number of network parameters. As the network expands, the number of channels used is increased to improve network capacity. Sixteen channels are initially set for the first set of convolutional layers, being duplicated after each max-pooling layer from there onwards up to the 9th layer.

From this point onwards, the learning process is slightly different, with special attention to the scalp distribution of the EEG amplitude. The 10th layer, with 32 channels, performs 3-point 1D convolutions over the spatial dimension (1x3 kernel), searching patterns in the distribution across the different electrodes. Afterward, 2D convolutions with 3x3 kernels are applied to find correlations between temporal and spatial information. This layer also has 32 channels, and its output is flattened, becoming a 1D structure with 512 points. This transformation is followed by 3 FC layers, which aim to aggregate information from high-level features extracted in preceding layers and determine which features are more strongly correlated with a particular class. The fully connected layers are implemented through linear functions. The last layer is fully connected to all its inputs (128) and the 2 outputs nodes (one for SZ and other for HC). The first two FC layers are followed by the ReLU function. After the last FC layer, the softmax function is applied to determine the probability distribution of the two outputs.

5.2.4 Traditional Machine Learning

There are few publicly available databases of EEG signals from subjects with a diagnosis of SZ. This hampers the comparison of the distinct classical machine learning or deep learning models proposed in the literature. To assess whether our deep network architecture brings advantages over traditional machine learning methods, we tested different algorithms.

Given the versatility of RFs, these algorithms were applied to perform the SZ classification using features extracted from auditory ERP data. In fact, RF is a robust classification algorithm. This model consists of a large number of small decision trees, known as estimators. RF combines predictions of the estimators to produce a more accurate prediction. This method performs well with heterogeneous and high-dimensionality features, and its ensemble design allows to compensate for the overfitting of standard decision trees [477].

5.2.5 Ensemble Strategy

The ensemble technique consists of combining multiple models, known as base learners, in order to reduce the generalization error and variance [42]. Rather than having a single learner as the best predictor, different models are trained separately, and subsequently ensembled to generate an optimal predictive model [42]. This is the strategy behind RF, which is considered a strong classifier. RF uses bagging (Bootstrap Aggregation) to increase the diversity in training of the ensemble [478] and decision trees as base learners [477]. A large set of trees are ensembled, and then averaged [479]. Since trees are remarkably noisy and weaker classifiers, they strongly benefit from averaging [479].

In an attempt to improve the generalization ability of the proposed deep learning model, it was adopted an ensemble strategy.

5.3 Experimental Setup

This section details the EEG dataset used, describing the preprocessing and the transformations applied to prepare data to be used as the model's input. Lastly, it is described the training setup of the proposed deep learning architecture and traditional machine learning classifiers.

5.3.1 Data

As detailed in the previous chapter, the experimental EEG recordings used in this study were obtained from a publicly available Kaggle dataset [420], hereinafter referred to as *dataset A*. The use of a small pool of training data may result in an increased risk of model's overfitting, and, consequently, of poor generalization to new data. Although some methods can be implemented, such as a k-fold cross-validation technique, the increase in data quantity can help to minimize overfitting. For this reason, the *dataset A* was extended with EEG data collected by our research team, using a similar sensory task, hereinafter referred to as *dataset B*. The entire dataset used in this study encompasses EEG recordings from 63 HC and 65 SZ subjects (*dataset A*: 32 HC and 49 SZ (17 early illness, 32 chronic); *dataset B*: 31 HC and 16 SZ (all first-episode)). Each subject dataset comprises EEG segments (one segment per trial).

The EEG acquisition methods, the preprocessing pipeline and the data analysis of both datasets were described in detail in Chapter 4. Both databases were merged after statistical analyses showed that differences between the two were not significant. The preprocessed EEG recordings from dataset A were downsampled to 512 Hz to reduce the input size (matching the sampling rate of dataset B), and, partially, the complexity of the model.

5.3.2 Deep Learning Algorithm

5.3.2.1 Temporal Window

In Chapter 4, we extensively examined the alterations in the N100 and P200 amplitudes in individuals with SZ. Building upon these findings, we based the discrimination between SZ and HC on a time window corresponding to the early auditory processing indices. The time window considered as input for the proposed model was shortened to 500 ms of duration, corresponding to a total of 256 time points. EEG segments were extracted from -100 ms to 400 ms time-locked to sound onset, comprising the expected latencies of the auditory N100 and P200 peaks. Figure 32 provides an example of the 2D structure generated for one trial showing the amplitude for the selected region of interest (ROI) (set of electrodes).

5.3.2.2 Data Normalization

The amplitude range of EEG recordings varies substantially across subjects, or even within a subject. The normalization of the neural network inputs allows not only comparable measures but also the gradient descent to converge faster. Consequently, the EEG data structures created were transformed before the training phase. The min-max normalization was applied in order to rescale the data. The absolute minimum and maximum values were determined for each training segment and for each electrode. The rescaling was then performed, with the segments' amplitudes varying between the average value of the minima and the average value of the maxima determined for each electrode. A Z-score standardization was then applied to each segment by subtracting the mean and dividing by the standard deviation of the entire dataset's amplitudes. This standardization ensures that the data has a zero mean and unit variance, thereby forcing both groups of subjects to have the same distribution.

5.3.2.3 Data Partitioning

Each subject's dataset was composed of a variable number of segments after the removal of segments with artifacts (mean of segments/subject: 94,86 $\pm$ 2,27). In total, 5756 trials from 63 HC subjects and 5853 trials from 65 SZ subjects were used. A stratified 10-fold cross-validation was performed. The dataset splitting was performed by subject rather than by segment. This procedure ensured that segments of a subject contained in a fold did not leak into other folds. The subjects' datasets were shuffled and split into 10 folds, while ensuring that each fold had the same proportion of SZ and HC subjects' data. In each of the ten iterations, the model was trained using 9 folds, while the remaining fold was used for the test. From the subjects contained in the 9 training folds of each iteration, approximately 90% of them were effectively used for training, while the remaining were used as a validation set. Also, in this split we ensured that the selected subjects contained all their EEG segments (Figure 34). The validation dataset was used to tune hyperparameters and provide an unbiased evaluation of the model. This division was also stratified according to SZ and HC classes, after shuffling the subjects. As in the folds split, the procedure also ensured that segments of a subject did not leak into other folds, so that segments belonging to a training

subject were used neither for validation nor for testing. In each iteration, all data segments contained in the training and validation sets were randomized.

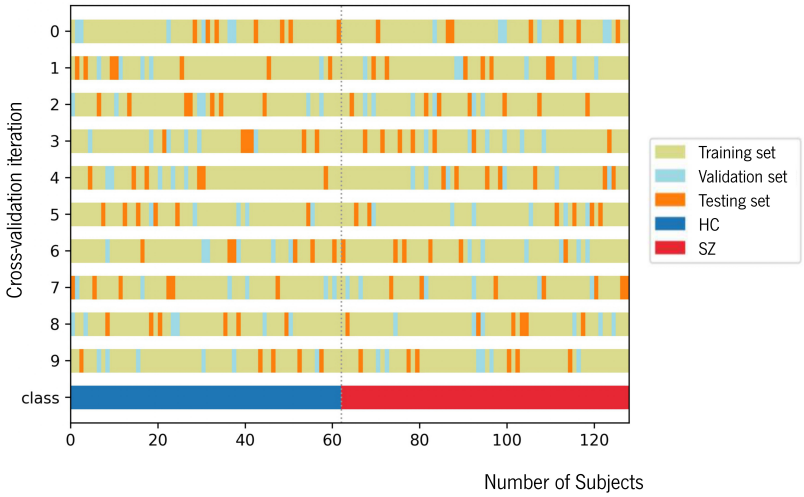


Figure 34: Stratified 10-fold cross-validation: balanced split of SZ and HC subjects data into training, validation and testing sets.

5.3.2.4 Model Training and Hyperarameters

The training process was performed on 300 epochs. The EEG images of all subjects that belonged to the training set were randomly shuffled. The network was trained with *ADAM* optimizer ($lr = 1 \times 10^{-4}$) and a mini-batch of size 4. Regularization techniques are implemented to prevent overfitting. The learning algorithm is modified in order to reduce its generalization error, but not the training error [42]. For regularization, a spatial dropout of 0.25 at some convolutions was used, as illustrated in Figure 33. To limit the model's capacity, the L^2 regularization was also implemented by adding a parameter norm penalty of 1×10^{-4} to the cost function. The *Xavier Glorot* uniform [241] was used as initializer of the model's parameters, ensuring the zero-mean and keeping the variance of activations the same across every layer. The Cross-Entropy Loss function was employed to calculate the loss by comparing the model's output probabilities with the true labels, providing a measure of dissimilarity between the predicted and actual outcomes.

As described before, the model output is a probability of an image label being assigned to the SZ or HC groups. For validation purposes, a probability threshold of 50% was considered to classify the segment as SZ or HC. The accuracy of the model was determined by all EEG images included in the analysis using cross-validation. During the optimization by backpropagation, the optimal model evaluated with the validation dataset was saved. The typical duration of model training was about 15 hours (90 minutes/fold).

Table 9: List of hypermeters selected for the proposed method.

	Hyperparameter	Value
Initialization	bias	0
	weigths	Xavier Uniform (gain=1)
Dropout	probability	0.25
Batch Normalization	ε	0.00001
Optimization	Adam	
	learning rate (α)	0.0001
	weight decay (λ)	0.0001
Training	batch size	4
	epochs	300

5.3.3 Traditional Machine Learning Algorithms

5.3.3.1 Feature Extraction

EEG feature extraction was performed based on ERP waveforms, which were obtained for each subject after averaging across all trials and presented in the previous chapter. Figure 35 provides a schematic illustration of the extracted features. Based on previous chapter results, the ERP mean amplitude was determined from the 75-105 ms latency window (post-stimulus onset), corresponding to the window in which the N100 ERP component typically emerges, and from the 150-210 ms latency window capturing the P200 component (see Table 8). Both time windows are shown with red shading in Figure 35. We also included three slopes as features: the beginning of N100 deflection, the transition from N100 to P200, and the final descending section of the P200 component. Yellow shaded areas in Figure 35 mark the latency intervals considered for calculating these slopes. All these five features were extracted from the signals captured by the electrodes Fz, FCz, Cz, CPz, and Pz. In addition, the amplitudes (circular symbol) and latencies (triangular symbol) of the N100 and P200 peaks (blue and green marks, respectively, in Figure 35) from the FCz and Cz electrodes were also extracted. A matrix of features was created, with rows corresponding to the subjects and the columns corresponding to the 33 features extracted for each one.

5.3.3.2 Data Partitioning and Normalization

A stratified 10-fold cross-validation method was applied to test machine learning models. Subjects (features matrix rows) were shuffled and split into 10 folds, ensuring that each fold had the same proportion of SZ and HC subjects data and that no subject was repeated in any fold. As implemented in the deep learning model, in each of the ten iterations, the model was trained using 9 folds, while the remaining fold was used for the test. From the subjects contained in the 9 training folds of each iteration, approximately

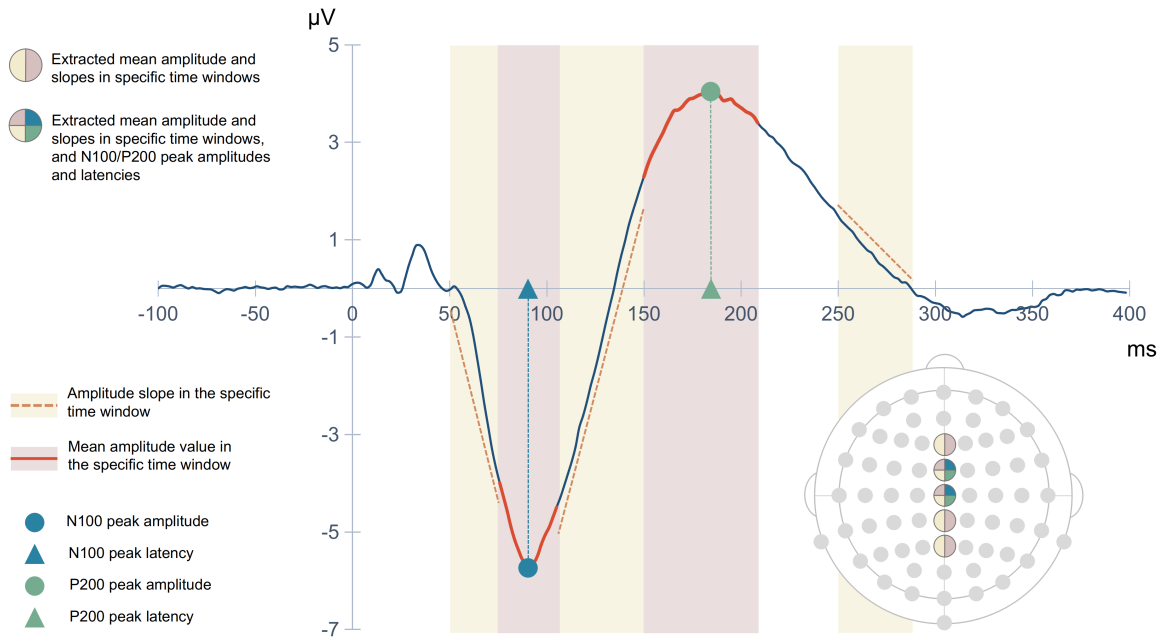


Figure 35: Schematic representation of the features extracted from each subject's ERP waveform.

10% of them were used in the validation set.

5.3.3.3 Models Training and Hyperparameters

The SZ classification was firstly performed with the RF algorithm using features extracted from the 5 electrodes. Several tests were performed to evaluate the effect of the different extracted features. The hyperparameter tuning was conducted for each one, focused on the number of estimators, maximum number of features, maximum depth of the tree, and split criterion using the validation set. Initially, only the mean amplitudes in the N100 and P200 windows were used, with ten features included per subject. RF configured with 100 estimators, a maximum number of features of 5, and a maximum depth of 2, achieved the best performance. Afterward, the N100 and P200 peak amplitudes and latencies were added to the previous data set, totaling 18 features per subject. The best model was configured with 50 estimators, a maximum number of features of 10, and a maximum depth of 2. Finally, slopes were also considered (subsection 5.4.6.1 will provide the rationale for including slopes as features). The best model performance using the 33 features per subject was achieved with 100 estimators, using a maximum of 2 features in each split and a maximum depth of 15 in each tree. In both cases, the criterion used to find the optimum split in the validation set was the *Gini* impurity measure.

5.3.4 Ensemble Method

SzNet was used as base learner and the hard voting algorithm as the ensemble method. The ensemble resulted from the combination of 5 fits of the SzNet model trained with different randomly selected seeds in model's weights initialization. This approach is justified by the fact that different initializations lead the

neural networks to converge to distinct solutions [480]. The hard voting of predictions from separately trained models is one of the simplest ensemble methods and it predicts the class with the largest sum of votes from the models.

5.3.5 Evaluation Metrics

In order to test the SzNet's performance, all input samples of every subject were fed into the model. The probability of each sample belonging to a SZ subject was computed. Unlike validation, a subject-based approach was considered to test the model. Thus, the average of each subject's input samples probabilities was determined and the same probability threshold (50%) was used to assign a label to the subject. This label can be positive (SZ) or negative (HC), and every subject fits into one of the four cases:

- True Positive (TP): subject labeled as SZ and correctly predicted as SZ;
- False Negative (FN): subject labeled as SZ and incorrectly predicted as HC;
- True Negative (TN): subject labeled as HC and correctly predicted as HC;
- False Positive (FP): subject labeled as HC and incorrectly predicted as SZ.

These four categories describe completely the classification task outcome, making up the confusion matrix (Table 10). With the binary classification, it is expected a large number of TP and TN, and less FN and FP cases. To describe the prediction quality, some metrics were derived from the confusion matrix.

Table 10: Confusion Matrix.

	Predicted SZ	Predicted HC
Actual SZ	TP	FN
Actual HC	FP	TN

The classification accuracy (Acc) is the simplest metric and is defined as the proportion of subjects labels correctly predicted by the model.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.2}$$

The precision was computed applying the equation 5.3, and is important inasmuch as it measures how consistent predictions are when tests are repeated.

$$Precision = \frac{TP}{TP + FP} \tag{5.3}$$

Sensitivity (or recall) and specificity, which are used to quantify the diagnostic ability, were also calculated. Both give the proportions of SZ and HC subjects, respectively, correctly predicted by the model.

$$Recall = \frac{TP}{TP + FN} \tag{5.4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5.5}$$

Although precision and recall are important metrics, it is not possible to maximize both simultaneously, because tipically when the FN number increases the FP reduces, and vice-versa. $F1\text{-}score$ takes into account both precision and recall (equation 5.6) and can be interpreted as a weighted average of both metrics. The model can be improved maximizing the $F1\text{-}score$, which is even more relevant when class distributions are unbalanced.

$$F1\text{-}score = 2 \times \frac{Prec \times Recall}{Prec + Recall} = \frac{2TP}{2TP + FP + FN} \tag{5.6}$$

The area under the receiver operating characteristic curve (AUC-ROC) and the Matthews correlation coefficient (MCC) were also computed, in an attempt to obtain metrics that better describe the model's ability to distinguish between the two classes.

A receiver operating characteristic (ROC) curve is obtained by plotting the recall (which also defines the true positive rate (TPR)) against the false positive rate (FPR) (equation 5.7), which corresponds to the ratio of HC subjects incorrectly predicted as SZ. ROC curves do not depend on class distribution and reflects a trade-off between TPR and FPR by changing the classification threshold. Calculating the area under the ROC curve, AUC-ROC, is equivalent to obtain the probability that a randomly chosen SZ subject is ranked higher than a randomly chosen HC subject. The closer the AUC-ROC score is to 1, the better the model performs.

$$FPR = \frac{FP}{FP + TN} \tag{5.7}$$

The MCC has been used as an alternative to accuracy and F1-score in binary classifications [481], and takes into consideration all the confusion matrix cells. This coefficient can be determined by:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5.8}$$

MCC ranges from -1 to 1, generating a high score only if the binary predictor was able to correctly predict the majority of SZ subjects and the majority of HC subjects.

5.3.6 Statistical Analysis

Repeated measures analysis of variance (ANOVA) was used to compare the different metrics between models since the same subjects were used to test each model. Before, the normality in each metric distribution was ensured through the Shapiro-Wilk's test, and the sphericity assessed with the Mauchly's test. Analyses were corrected for non-sphericity using the Greenhouse-Geisser method. Main effects were followed up with pairwise comparisons using *post-hoc t*-tests adjusted with the Bonferroni correction. The sizes of significant effects are also reported using the generalized η-squared method (η^2_g). Only significant results are presented, with a significance alpha level of 0.05. The analysis was conducted in R [482].

5.3.7 Implementation Details

Both deep and traditional machine learning models were implemented on a workstation with an Intel® Core™ CPU (i7-4770k, 3.5 GHz) and an NVIDIA® GPU (GeForce® GTX 1070). Deep learning models were implemented using the open-source framework Pytorch, based on Torch library. RF algorithms were also written in python using *Scikit Learn* [483] and *NumPy* [484] packages.

5.4 Results and Discussion

This section starts by motivating the choice of kernel for time-domain convolutions. Thereafter, it is presented and discussed the ablative study, which assesses the relevance of the amount and location of EEG spatial-temporal information used. This ablative study intended the evaluation of the number of electrodes used as well as their location over the scalp. Lastly, the developed model is compared with related work, pointing out its limitations.

5.4.1 Kernel size for temporal convolution

The kernel size determines the receptive field of a convolution and provides information about the number of input datapoints the network can look at [485]. This is a factor to take into account when considering a network's ability to encode the features, and it is associated with the learning parameters.

In this study, the kernel size used for time-domain convolution defines what information is extracted from the EEG signal. The decomposition of EEG signals can reveal oscillatory activity in specific frequency bands [209]. Activity in each band has been associated with different functions. The faster rhythms of the EEG signal, corresponding to the gamma-band (above 30 Hz), are linked to complex auditory information processing. Desynchronization of these oscillations during auditory processing has been reported in SZ [209, 418]. Thus, we search for the best kernel size using cross-validation. After testing various sizes, the 9x1 kernel was the one that reflected the best results.

5.4.2 Ablation Study

5.4.2.1 Relevance of number and location of electrodes

Different spatial ROI were analyzed in order to understand whether and which spatial information is relevant for SZ discrimination. The effects of the number of electrodes and their location were tested. Three principal approaches were considered (5, 15, and 35 electrodes over the frontal, fronto-central, central, centro-parietal, and parietal regions) with the set location varying from the left to the right side of the brain. The three models are hereinafter referred to as SzNet-5, SzNet-15, and SzNet-35 corresponding to the three sets used: 5, 15, and 35-electrodes sets, respectively. To assess the use of 15 and 35 electrodes, the SzNet architecture was slightly changed: an extra convolutional layer (with 32 filters) was incremented after layer 9, followed by a ReLU function (Table 11).

Table 11: Architectures of the SzNet model adapted for the three spatial approaches tested.

Layers	SZNet-5					SZNet-15					SZNet-35				
	type	#filters	kernel	stride	input	type	#filters	kernel	stride	input	type	#filters	kernel	stride	input
1	Conv2D	16	9x1	1x1	256x5x1	Conv2D	16	9x1	1x1	256x15x1	Conv2D	16	9x1	1x1	256x35x1
	Dropout	-	-	-	-	Dropout	-	-	-	-	Dropout	-	-	-	-
2	Conv2D	16	9x1	1x1	248x5x16	Conv2D	16	9x1	1x1	248x15x16	Conv2D	16	9x1	1x1	248x35x16
3	MaxPool2D	-	2x1	2x1	240x5x16	MaxPool2D	-	2x1	2x1	240x15x16	MaxPool2D	-	2x1	2x1	240x35x16
4	Conv2D	32	9x1	1x1	120x5x16	Conv2D	32	9x1	1x1	120x15x16	Conv2D	32	9x1	1x1	120x35x16
	Dropout	-	-	-	-	Dropout	-	-	-	-	Dropout	-	-	-	-
5	Conv2D	32	9x1	1x1	112x5x32	Conv2D	32	9x1	1x1	112x15x32	Conv2D	32	9x1	1x1	112x35x32
6	MaxPool2D	-	2x1	2x1	104x5x32	MaxPool2D	-	2x1	2x1	104x15x32	MaxPool2D	-	2x1	2x1	104x35x32
7	Conv2D	64	9x1	1x1	52x5x32	Conv2D	64	9x1	1x1	52x15x32	Conv2D	64	9x1	1x1	52x35x32
	Dropout	-	-	-	-	Dropout	-	-	-	-	Dropout	-	-	-	-
8	Conv2D	64	9x1	1x1	44x5x64	Conv2D	64	9x1	1x1	44x15x64	Conv2D	64	9x1	1x1	44x35x64
9	MaxPool2D	-	2x1	2x1	36x5x64	MaxPool	-	2x1	2x1	36x15x64	MaxPool2D	-	2x1	2x1	36x35x64
10	Conv2D	32	1x3	1x1	18x5x64	Conv2D	32	1x3	1x3	18x15x64	Conv2D	32	1x7	1x7	18x35x64
11	Conv2D	32	3x3	1x1	18x3x32	Conv2D	32	1x3	1x1	18x5x32	Conv2D	32	1x3	1x1	18x5x32
	Dropout	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	FC - 256	-	-	-	512	Conv2D	32	3x3	1x1	18x3x32	Conv2D	32	3x3	1x1	18x3x32
	Dropout	-	-	-	-	Dropout	-	-	-	-	Dropout	-	-	-	-
13	FC - 128	-	-	-	256	FC - 256	-	-	-	512	FC - 256	-	-	-	512
	Dropout	-	-	-	-	Dropout	-	-	-	-	Dropout	-	-	-	-
14	FC - 2	-	-	-	128	FC - 128	-	-	-	256	FC - 128	-	-	-	256
	-	-	-	-	-	Dropout	-	-	-	-	Dropout	-	-	-	-
15	softmax	-	-	-	2	FC - 2	-	-	-	128	FC - 2	-	-	-	128
16	-	-	-	-	-	softmax	-	-	-	2	softmax	-	-	-	2

Abbreviations: Conv - Convolutional layer; Max-pool - Max Pooling; FC - Fully Connected Layer.
Note: Input dimensions - (number of time points; number of electrodes; number of feature maps)

The electrodes in 2D structures were also aligned from the frontal to the parietal region, and from the left to the right hemisphere, as exemplified in Figure 36. *N* corresponds to the number of electrodes from each scalp region tested, which was set to 3 for SzNet-15 and to 7 for SzNet-35. The kernel size used in the extra convolutional layer is dependent on *N*. This layer aimed to gather information from each region (frontal, fronto-central, central, centro-parietal, and parietal). Convolutions with stride equal to *N* in the spatial dimension allowed obtaining one feature per region, as illustrated in Figure 36. The remaining network was unaltered.

Figure 37 schematically presents the ROI, which varies in number of electrodes and hemispheric location, used to evaluate the performance of the three models developed. The 10-fold cross-validation metrics, computed for each subset of electrodes, are presented in Table 12.

The results of SzNet-5 and SzNet-15 using different spatial subsets show differences in accuracy as a function of electrode location in the training and data sets. The training and testing with midline and right ROI data showed improved performances. Data laterality had a small effect on accuracy ($F(10, 99) = 2.298, p = 0.018, \eta_g^2 = 0.188$) and AUC-ROC ($F(10, 99) = 2.221, p = 0.022, \eta_g^2 = 0.183$) metrics, being only statistically significant in SzNet-5 when contrasting right-1 and left-1 ROIs (Acc: $p = 0.042$; AUC-ROC: $p = 0.039$). Both accuracy and AUC-ROC metrics show that SzNet-5 with midline and right-1 data, and SzNet-15-midline achieved the best performances. The lower result achieved by the SzNet-5 left-1 subset can account for the slightly lower performance obtained by SzNet-15-midline.

Although no asymmetries in auditory processing alterations have been reported in SZ patients [106],

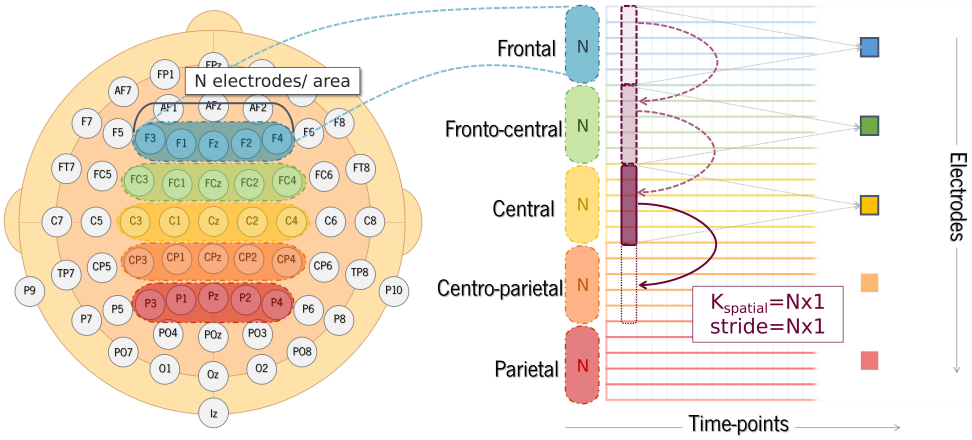


Figure 36: Illustration of the electrode alignment in the 2D structure created for each EEG segment, and schematization of the convolutional process of the extra layer added to the SzNet-15 and SzNet-35 networks. $K_{spatial}$ represents the kernel used in the convolutional layer. N corresponds to the number of electrodes of each area tested.

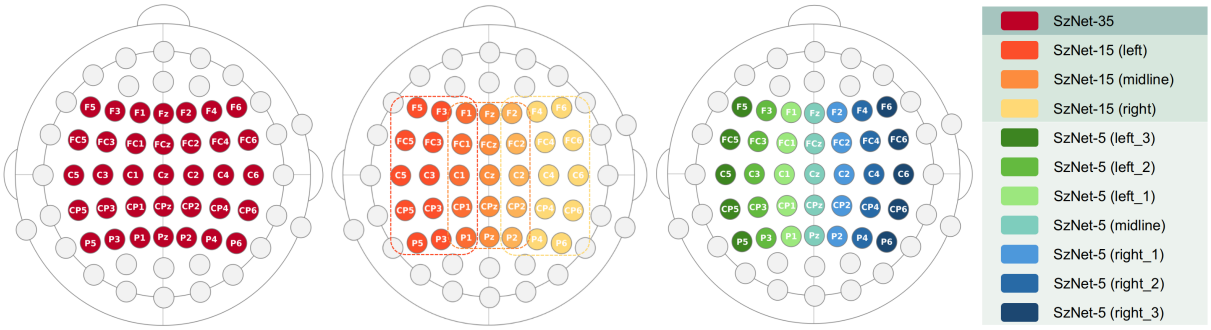


Figure 37: Schematic representation of the different ROIs showing the topographical distribution of electrodes considered to test the three models SzNet-35, SzNet-15, and SzNet-5.

these results suggest that changes in amplitude of right hemisphere EEG signals are contributing more to the discrimination of SZ and HC subjects. Nonetheless, the use of midline electrodes seems more beneficial and is corroborated by most of N100 studies in SZ, which often report a medial to midline N100 amplitude decrease [24]. The results suggest that the use of small subsets of electrodes can improve SZ classification, although no significant differences were found between SzNet-35, SzNet-15, and SzNet-5 regardless of ROI ($F(2, 107) = 0.3939, p = 0.6754$).

The results suggest that the model using 5 midline electrodes (SzNet-5-midline) more accurately predicts whether the EEG signals come from a SZ or a HC subject. Although with minor differences, this model achieved the best performance, on average, for every metric. Subtle differences in accuracy values can represent important practical improvements since a variation of only 0.8% is necessary to indicate or detect the increase or decrease of a unit in the number of correctly identified subjects.

Table 12: Performance of the models tested (SzNet-5, SzNet-15 and SzNet-35) using different spatial ROI evaluated with five metrics: accuracy; AUC-ROC; precision; recall; and specificity.

Model	ROI	Accuracy	MCC	AUC-ROC	F1-score	Precision	Recall	Specificity
SzNet-35	-	0.69±0.07	0.40±0.15	0.70±0.07	0.72±0.06	0.69±0.12	0.76±0.09	0.63±0.15
SzNet-15	left	0.67±0.09	0.37±0.18	0.67±0.09	0.71±0.06	0.67±0.14	0.79±0.15	0.55±0.26
SzNet-15	midline	0.76±0.07	0.54±0.14	0.76±0.07	0.76±0.08	0.79±0.13	0.77±0.17	0.74±0.19
SzNet-15	right	0.73±0.10	0.49±0.20	0.72±0.10	0.76±0.09	0.69±0.10	0.88±0.15	0.56±0.19
SzNet-5	left3	0.71±0.10	0.43±0.21	0.70±0.10	0.72±0.11	0.70±0.08	0.77±0.19	0.63±0.17
SzNet-5	left2	0.72±0.05	0.46±0.11	0.71±0.04	0.71±0.08	0.75±0.11	0.72±0.20	0.71±0.17
SzNet-5	left1	0.67±0.08	0.38±0.17	0.67±0.08	0.68±0.11	0.71±0.13	0.70±0.20	0.65±0.24
SzNet-5	**midline**	0.78±0.08	0.59±0.18	0.78±0.08	0.78±0.08	0.82±0.13	0.77±0.15	0.79±0.17
SzNet-5	right1	0.77±0.08	0.58±0.15	0.77±0.09	0.77±0.11	0.81±0.21	0.72±0.21	0.77±0.09
SzNet-5	right2	0.72±0.11	0.45±0.24	0.72±0.11	0.73±0.12	0.70±0.10	0.78±0.19	0.65±0.14
SzNet-5	right3	0.67±0.07	0.38±0.17	0.67±0.07	0.68±0.09	0.70±0.14	0.72±0.23	0.61±0.26
SzNet-5	midline**	0.77±0.09	0.56±0.20	0.77±0.09	0.74±0.10	0.86±0.16	0.68±0.15	0.85±0.16

Note - The highest values achieved for each metric are underlined. Similarly, the model and ROI that achieved the best results are highlighted.
** - Random electrode alignment approach.

5.4.2.2 Benefitting from electrode alignment

The strategy based on the alignment of electrodes in the 2D data structures created was adopted to feed the deep learning model. Instead of randomly stacking the EEG signal segments, it was hypothesized that the alignment according to the topographical arrangement of the electrodes could represent a more realistic way to extract spatial information from the data structures. To confirm that network performance benefits from this strategy, a comparison between random and organized arrangements of segments was conducted for the proposed model, SzNet-5, using the midline ROI. The 10-fold cross-validation metrics are presented in Table 12, with the randomly aligned ROI marked with asterisks at the bottom of the table. With the organized alignment, the accuracy, AUC-ROC, and recall metrics were improved. The most notorious difference is in the recall value, which is higher when the signal segments are not randomly stacked. The increase in this metric means that the ratio of diagnosed patients incorrectly identified as healthy subjects decreased, which would be a positive result in clinical diagnosis. This may suggest that the spatial dynamics of brain activity over time may be relevant in characterizing auditory processing in SZ.

5.4.3 Probabilities Distribution

Up to this point, a subject-based approach was used to evaluate and analyze the models' performances in order to understand which better identifies SZ patients. Figure 38 shows the distribution of the subjects' mean probabilities predicted by SzNet-5.

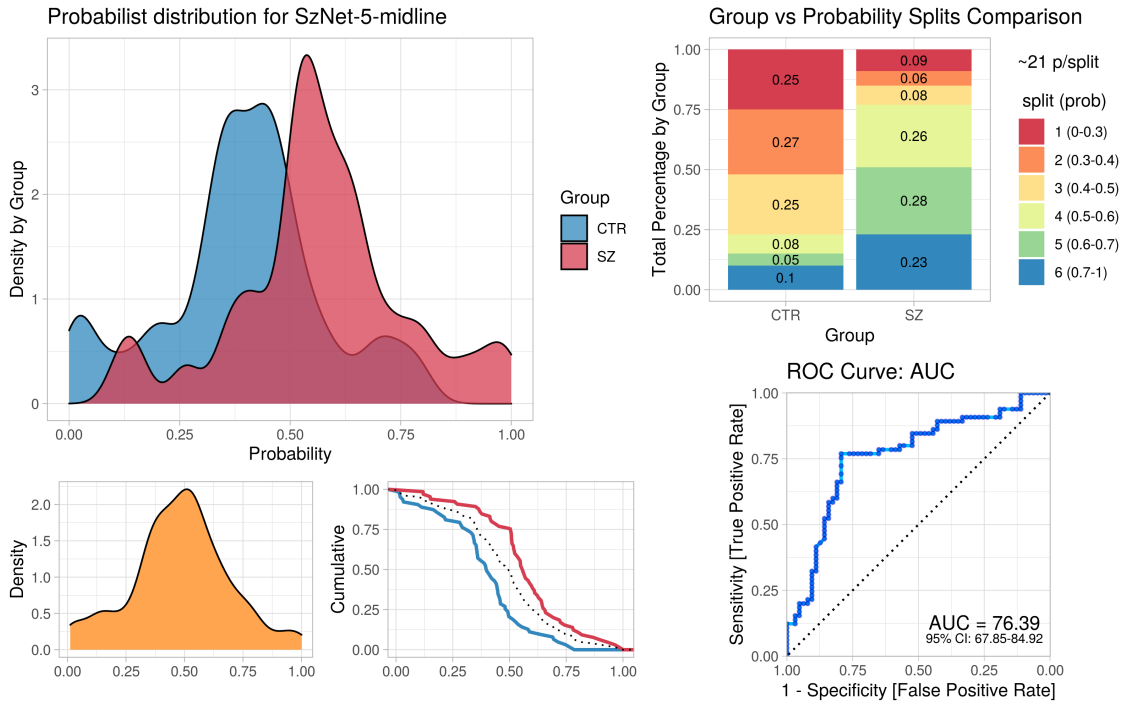


Figure 38: ROC-curve (and corresponding AUC-score) and distribution of the probabilities of SZ and HC predicted by the SzNet-5-midline,considering a subject-based approach, with average probabilities determined for each subject. The closer the probability is to zero, the greater the probability of being a HC subject. The closer to 1, the higher the probability to be a SZ patient.

A deeper separation between groups' distribution would be expected, which could, therefore, be reflected in a higher AUC-ROC value. This outcome may be due to the great intersubject variability in event-related EEG data.

Since the model was trained on EEG single-trials, a segment-based analysis can provide important insights on which features of the signal have contributed the most to distinguish SZ from HC. As for the subject-based approach, the distribution of all segments probabilities predicted by SzNet-5 was determined (Figure 39). As shown, the AUC-ROC value decreased and there is a larger overlapping of both groups' densities. A possible justification is the variability introduced by the examination of all the segments. Besides the intersubject, this result reveals the inherent intrasubject EEG data variability. Higher peaks on HC and SZ distributions at 0 and 1 probabilities, respectively, mean that a considerable percentage of segment labels are correctly predicted with absolute certainty. The features extracted from these segments were very effective in identifying the group to which they belonged. Taking advantage of this, the segments correctly identified with higher probabilities can provide clues about changes in SZ patients' auditory processing.

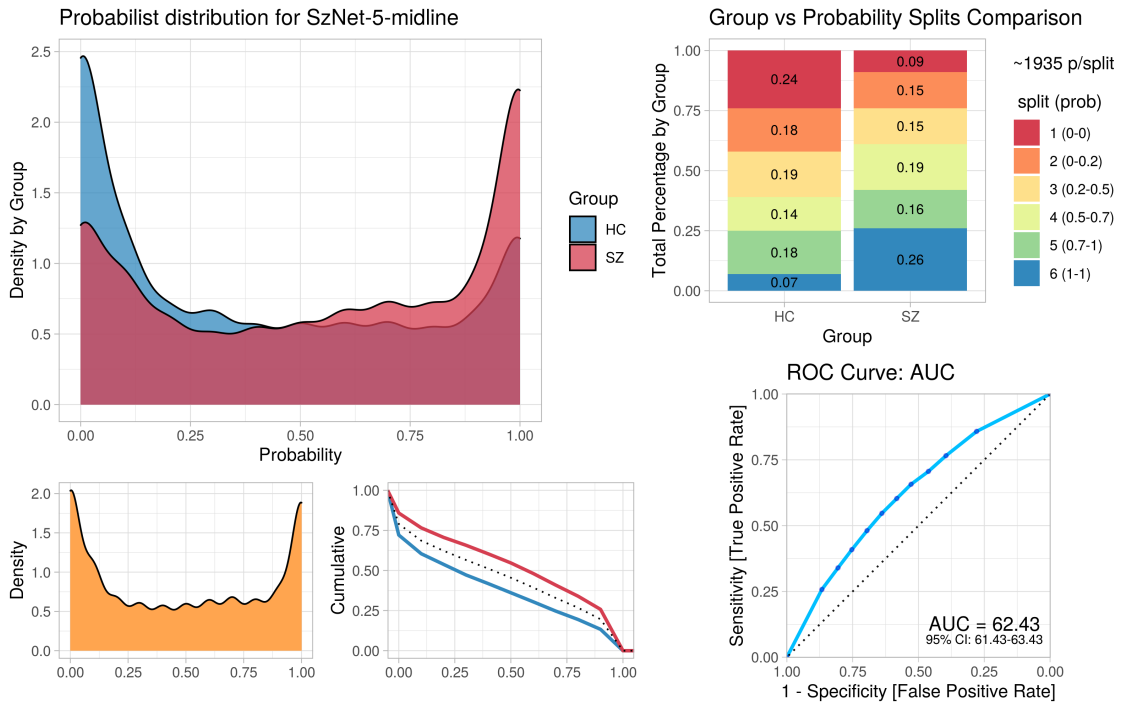


Figure 39: ROC-curve (and corresponding AUC-score) and distribution of the probabilities of SZ and HC predicted by the SzNet-5-midline,considering a segment-based approach, including all segments probabilities. The closer the probability is to zero, the more likely the EEG segment is to belong to a HC subject. The closer to 1, the higher the probability that the segment belongs to a SZ patient.

5.4.4 Model interpretation

While auditory processing abnormalities in SZ have been investigated by mean of ERP components analysis, this deep learning approach extracts features from the signal that may be directly correlated to those components ant that are relevant to the discrimination of SZ and healthy subjects. Therefore, the interpretation of the features extraction in the implemented SzNet model provides more information about their relationship with specific ERP components. Some algorithms have been developed for CNN features interpretation. An example is the Gradient-weighted Class Activation Mapping (Grad-CAM). This algorithm uses the gradients of the output with respect to the convolutional features to produce a saliency map of the important regions in the input image for the classification [486].

The importance of the features extracted by SzNet-5 model (considering the last convolutional layer features) is represented by heatmaps obtained by Grad-CAM implementation in Figure 41. The color scale shows the variation of features relevance, with dark tones representing highly discriminative features in the signal morphology.

For this purpose, it was considered the segments identified correctly with a greater degree of certainty (probabilities above 0.80 for SZ and below 0.20 for HC). An average of the heatmaps of the selected samples was computed for SZ and HC (Figure 40). An average of the EEG segments was also computed to inform on the features' location and correspondence in the spatio-temporal images. This representation of ERP-like average gives a sense of what patterns in input signals are driving classification differences.

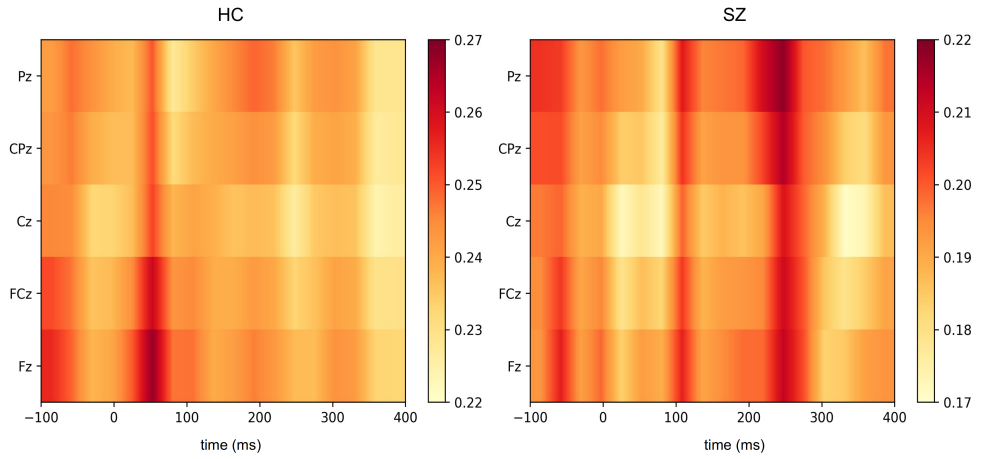


Figure 40: Heatmaps (yellow to red colormap) obtained using Grad-CAM for the segments most likely to belong to HC and SZ subjects.

Both segments and heatmaps averages were graphed in the form of surfaces, making their comparison more intuitive (Figure 41). The patterns of averaged EEG segments (surface in blue tones) is consistent to the ERP SZ and HC morphologies studied and presented in Chapter 4.

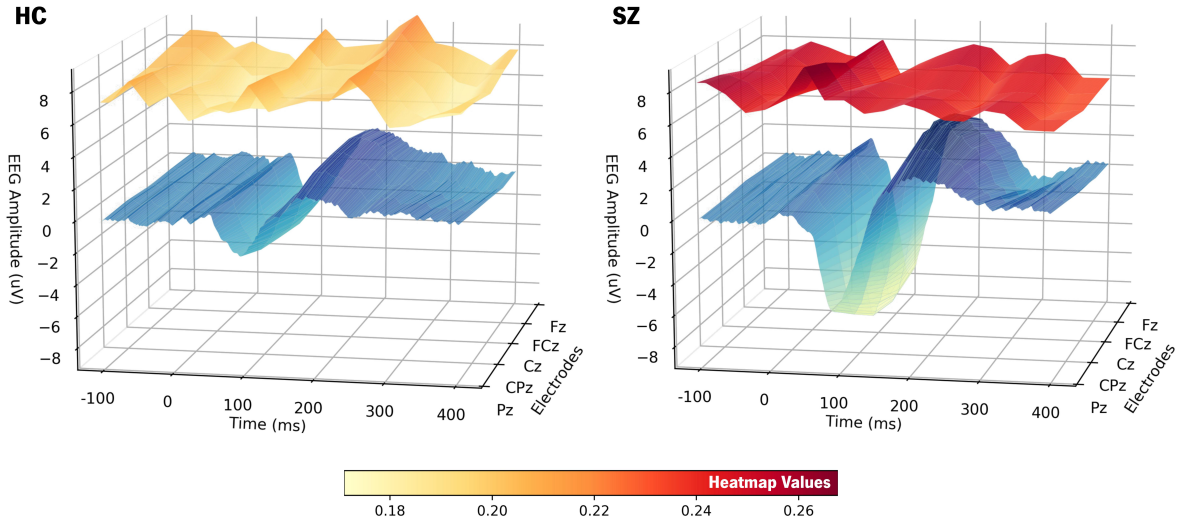


Figure 41: Surface representation of heatmaps (yellow to red colormap) obtained using Grad-CAM for the segments most likely to belong to HC and SZ subjects. In shades of blue, the surface representation of the averages of the corresponding EEG segments.

Figure 42 illustrates HC and SZ heatmaps over Cz and averaged temporal segments (see Appendix B for the other electrodes representations). Group differences around 250 ms post-stimulus onset appear to be the temporal feature receiving higher activation by the model for the classification of SZ segments. Interestingly, the transition of more pronounced negative to positive deflections around 110 ms post-stimulus onset, i.e., from the N100 to the P200 deflection, may have been pivotal for SZ identification. This transition is more abrupt in HC as a result of increased N100 and P200 amplitudes. Both salient features seem to have a more significant effect across the five electrodes (Figure 40). The feature extracted from

HC segments seem to be more prominent (higher heatmap values) for the classification by inspecting both heatmaps. It should be noted that the most important features to identify HC are located around 50 ms post-stimulus onset. Subtle differences in this time interval may correspond to the transition from a P50-like deflection, whose latency seems to vary between groups, to a N100 deflection.

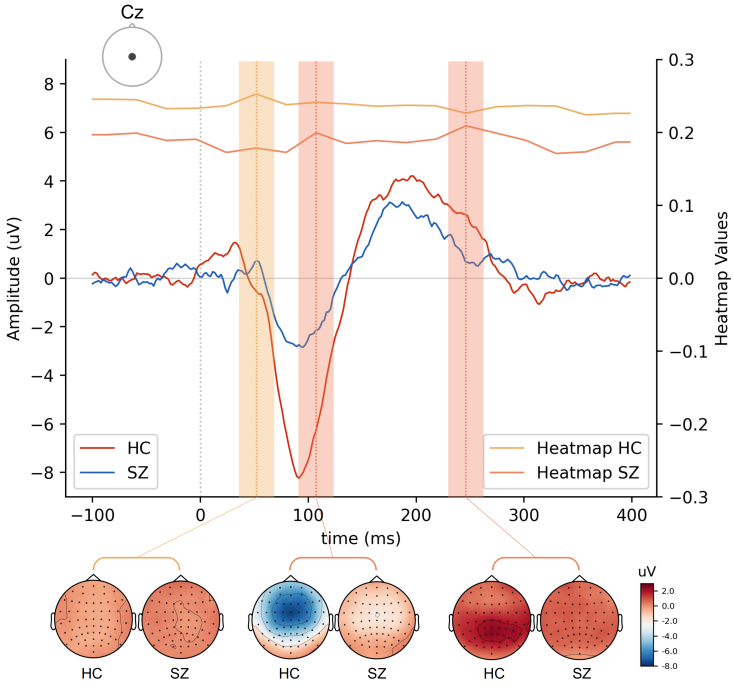


Figure 42: Grad-CAM heatmaps cross-section and averaged EEG signals over Cz electrode drawn for HC and SZ groups. The time intervals associated with the most relevant features are shaded with the color corresponding to the heatmap of each group. Topographical distribution maps of EEG signal amplitudes at those time intervals are also presented for each group.

The results suggest that this feature may have a greater effect on the most parietal electrodes. The spatial distribution of the amplitudes around the more relevant time-points is also presented in Figure 42. More pronounced spatial differences occurred around 100 ms, where HC and SZ amplitude is lower in frontal-central electrodes, and around 250 ms, showing higher parietal amplitudes of HC signals. The GradCAM results suggest that the transitions from P50 to N100, N100 to P200, and the P200 end have received higher attention from the proposed model to discriminate both groups. These results, although lacking a better optimization of the model to improve its generalization, show that ERP components consistently proposed as biomarkers may actually play an important role in characterizing auditory processing at SZ (studied in detail in Chapter 4).

5.4.5 Relevance of SzNet

The SZ classification based on classical machine learning methods in auditory tasks-related EEG data has been performed using features extracted from ERP components. Besides requiring expertise, there is a great variability in the extracted features. By contrast, the use of deep learning allowed the SzNet model to automatically learn patterns from EEG single-trials. Despite the heterogeneity of the disorder, the model

managed, through a short time window focused on the early auditory processing stage, to distinguish whether the signal belongs to SZ or HC. Although it may be considered a limitation, the fact that the SZ sample is independent of the stage of the disease (not distinguishing subgroups of the disease, first-episode, or chronic patients) may contribute to mimic the heterogeneity that will always exist in a real clinical situation. Another advantage of the SzNet approach is the reduced number of trials and the task used for EEG acquisition. The elicitation of some ERP components such as the MMN, whose features are recurrently used for classification purposes, requires a large number of stimuli and, consequently, is time-consuming. Differently, the passive listening task used in this study is simple and manageable in a clinical environment, where the diagnosis should be fast. Besides the binary classification, another aim of the current study was to understand if impairments in auditory processing may contribute to the characterization of SZ, in particular considering alterations in ERP components that have been proposed as SZ biomarkers. Only the study conducted by Ahmedt-Aristizabal [348] documented the application of deep learning methods to auditory processing signals. The aim was to identify subjects at risk of developing SZ using the time course of EEG signals in response to auditory stimuli. With a shallower architecture combining convolutional and recurrent networks and with less learnable parameters, an accuracy of 72.54% was achieved. The problem addressed as well as the database used differ from ours, so the studies are not directly comparable. Still, it should be noted that Ahmedt-Aristizabal and collaborators used the same 5 electrodes (Fz, FCz, Cz, CPz, Pz), for which our ablative study showed an improved performance.

The SzNet results indicate that changes in the auditory processing can be key features for SZ diagnosis. This shows the need to look at sensory changes that are often neglected in clinical assessment, and supports the addition of a new domain to the RDoC framework, focused on sensory impairments in psychiatric disorders [487]).

5.4.6 Machine learning vs. Deep learning

The application of classical machine learning algorithms allowed us to obtain a base-model to compare with the proposed deep learning model.

5.4.6.1 Features Selection

The selection of relevant features plays an essential role in machine learning classification. The auditory N100 and P200 components shed light on the sensory processing of sounds, and their latency and amplitude are affected by the level of subject's arousal, alertness, and attention [488]. Abnormalities in these components have been consistently reported in SZ patients, which include reduced amplitude of the N100 as well as reduced amplitude and shorter peak latencies of the P200 [174]. Moreover, both auditory ERP components have been used to examine sensory and information processing impairments in SZ [489]. In light of this, and based on Chapter 4 results, our focus turned to these components, in particular their amplitude and latency. Time windows for the extraction of the N100 and P200 mean amplitudes were based on the visual inspection of the grand average waveforms, on ERP results presented in Chapter 4, and on previous studies [18]. Since the N100 and P200 peak amplitudes are topographi-

cally distributed over frontocentral and central regions, as seen in Figure 31, the FCz and Cz signals were considered for the analysis of both peak amplitude and latency.

From the evaluation of Grad-CAM results, the transitions from P50 to N100, N100 to P200, and the P200 end seem to have relevance for the discrimination of the two groups. In order to interpret these findings, we also extracted the slopes in the time intervals corresponding to those three transitions as input features for machine learning models.

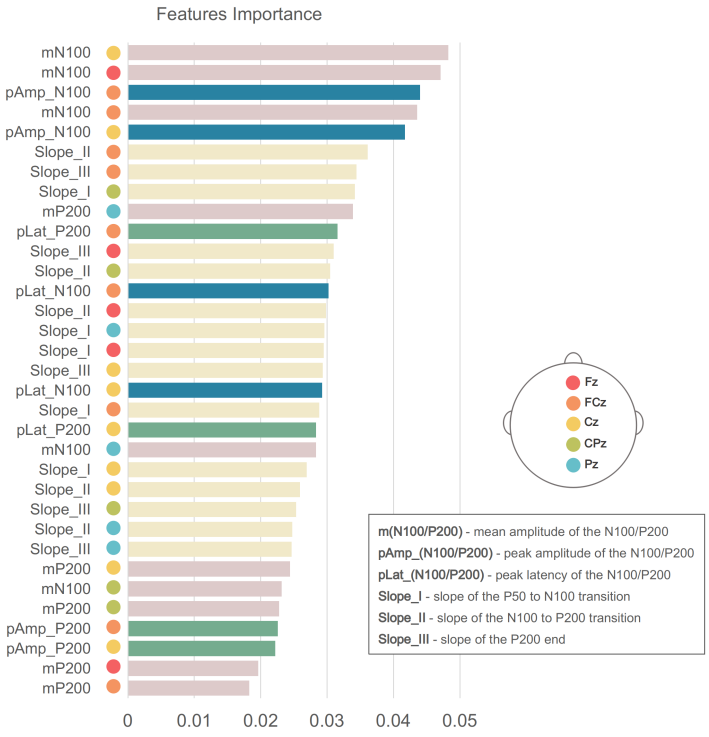


Figure 43: Importance of each feature given by its weight to final performance of RF using 33 features. Bars colors corresponds to the colors used in Figure 35.

RF can provide a measure of the importance of each feature, according to its contribution to the overall classification performance. Benefiting from the increased system interpretability, the features' weights may provide important information on the critical EEG electrodes and features for discriminating SZ. Figure 43 shows the weight of each feature for the RF result considering 33 features. N100 features from frontal, fronto-central, and central scalp regions were the most important features for classification, which is consistent with the literature [24] and the Chapter 4 studies. However, the slopes extracted from the transitions between ERP components appear to have been significant. This result suggests that the temporal dynamics of early auditory processing is altered in SZ. Hence, all these features were considered to obtain the best performance as they seem to have an impact on the differentiation of subjects with SZ.

5.4.6.2 Models Performance

Table 13 presents the RF models' performance using ERP features for SZ classification. The inclusion of amplitude and latency of the N100 and P200 peaks over FCz and Cz electrodes (RF-18) improved the

classification achieved by the RF, which only include the N100 and P200 mean amplitudes over all the 5 midline electrodes (RF-10). The best result was obtained when adding all 33 features (RF-33), which included the slopes in the three transition time intervals between ERP components suggested by the analysis of Grad-CAM. The best model distinguished HC and SZ subjects with an accuracy of 73%. Adding the slopes values as features increased the specificity, which suggests that these features may have a more significant influence on the identification of HC subjects.

Table 13: Results of the classification performed by the RF, the proposed deep learning model, and the ensemble learning method.

Model	Features	Accuracy	Precision	Recall	Specificity	AUC-ROC
RF	10	0,69±0,10	0,70±0,15	0,70±0,17	0,69±0,14	0,70±0,10
RF	18	0,70±0,11	0,71±0,14	0,70±0,17	0,71±0,15	0,70±0,11
RF	33	0,73±0,13	0,73±0,15	0,70±0,17	0,76±0,13	0,73±0,13
Proposed		0,78±0,08	0,82±0,13	0,77±0,15	0,79±0,17	0,78±0,08
Ensemble		0,80±0,08	0,82±0,14	0,82±0,14	0,78±0,18	0,80±0,08

The performance of the deep learning method surpasses that of the traditional machine learning models, and demonstrates the potential of the model we propose here for SZ classification. Unlike standard techniques in which features are extracted manually and provided to the model for classification, the deep learning model performs both feature extraction and classification. CNNs are able to detect highly specific features of the training dataset under the constraints of the specific prediction, such as the data labels [42]. The end-to-end training forces feature extraction by minimizing the loss for SZ classification. Through this optimization technique, CNNs provide improved models and, thus, more accurfig:impate results [42].

5.4.7 Ensemble learning

The result of the ensemble of SzNet models' predictions are presented in Table 13. An improved performance is observed using 5 models, surpassing the single SzNet result (see Figure 44). As expected, the voting ensemble offered a lower variance in predictions made over base learners, which improved the generalization ability.

5.4.8 Limitations and future directions

The single-trial EEG recordings contain task-related activity. However, this activity is overlaid with task-unrelated brain processes, which results in a low SNR. Variations in SNR across trials and subjects, which may also arise from differences in EEG data acquisition parameters, may affect signal quality. Moreover, fluctuations in the latency and magnitude of the EEG responses to stimuli, combined with the high dimensionality of EEG signals, increase both intra- and intersubject variability. Improved performances from raw EEG data may be achieved with larger training data sets, which also prevents overfitting: more subjects

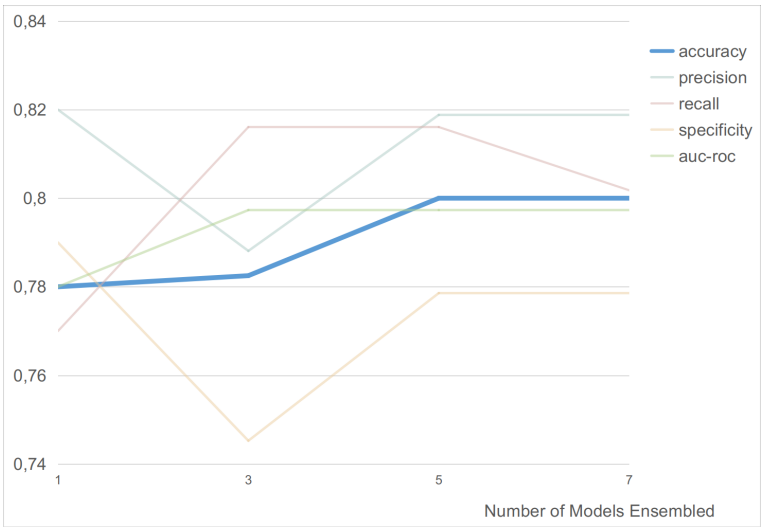


Figure 44: Effects of the growing number of ensembled models on different metrics, with emphasis on accuracy.

and more trials per subject to overcome inter and intrasubject variability, respectively. Furthermore, by encompassing diverse stages of the disorder and incorporating subjects from each stage, the analysis not only reduces variability but also enhances the robustness of the findings derived from the raw EEG data. Data augmentation techniques may increase the diversity of the data available for training the model, and thus prevent overfitting. Notwithstanding, we note that our attempts at data augmentation did not lead to improved SzNet model performance.

As mentioned before, SZ subgroups or stages (e.g., first episode vs. chronic) were not considered in this study. Future studies should address whether SzNet performance is affected by illness stage and symptom severity. However, some requirements should be taken into account: adaptation of the SzNet model for multi-class classification; inclusion of a larger amount of data from each SZ subgroup. Following the adaptation of the model to multiple classes, an important approach might be to include data from subjects at risk of developing SZ or in prodromal stages. This line of development may allow prediction of disease onset and a better understanding of the course of the underlying auditory processing alterations.

SzNet can provide a baseline model for future developments and analyses. A well-characterized public SZ EEG database is highly recommended for the direct comparison and objective evaluation of the performances of different algorithms.

5.5 Summary

Deep learning techniques have been used to decode EEG signals, with promising applications in the field of psychiatry. SZ is one of the most disabling neuropsychiatric disorders, often characterized by the presence of auditory hallucinations. Auditory processing impairments have been studied using EEG-derived event-related potentials and have been associated with clinical symptoms and cognitive dysfunction in SZ. Due to consistent changes in the amplitude of ERP components, such as the auditory N100, some

have been proposed these components as biomarkers of SZ. The work presented in this chapter examines altered patterns in electrical brain activity during auditory processing and their potential to discriminate SZ and healthy subjects. Using convolutional neural networks (CNNs), a deep learning architecture is proposed to perform the classification based on multi-channels auditory-related EEG single-trials, recorded during a passive listening task. An ablative study allowed to analyze the effect of the number of electrodes used, as well as the laterality and distribution of the electrical activity over the scalp. Results show that the proposed model is able to classify SZ and healthy subjects with an average accuracy of 80% using only 5 midline channels (Fz, FCz, Cz, CPz, and Pz). This chapter shows the potential of deep learning methods in the study of impaired auditory processing in SZ and, consequently, with implications for diagnosis. The proposed design can provide a baseline model for future developments in SZ research.

Chapter 6

Final Remarks

This thesis presents the work developed to probe auditory processing alterations in individuals with SZ and its importance in discriminating patients from healthy subjects. In each previous chapter, we presented the studies developed and the results obtained, as well as the discussion that encloses our observations and conclusions. This last chapter summarizes the main contributions and conclusions. Finally, we close this chapter with our perspectives on the limitations of the empirical studies conducted and discuss future research directions.

Contents

6.1 Conclusions **120**

6.2 Limitations and Future Directions **122**

6.1 Conclusions

The high heterogeneity of SZ and symptom overlap with other psychiatric disorders have challenged neuroscientists and engineers to develop more effective diagnostic and predictive tools. The review presented in Chapter 3 critically analyzed classical machine learning and deep learning methods to detect SZ based on EEG signals published between 2016 and 2020. Specifically, the chapter discusses the potentialities and limitations of the algorithms used and their parameterization in task-related and resting-state EEG signals. This critical review highlighted good practices for developing and implementing machine learning methodologies applied to EEG signals.

Machine learning methods are neither expected nor intended to replace the current diagnostic systems for SZ and other psychiatric disorders. Instead, advanced learning algorithms may complement and aid the discrimination of SZ and HC, improving the identification of SZ biomarkers, and establishing relationships between EEG measures and specific clinical symptoms. These techniques, especially deep learning algorithms, offer promise in providing clues to understanding the neural mechanisms of SZ because they have the potential to uncover hidden relationships and dependencies within the EEG data. Given the reduction of (time-consuming) feature engineering and the possibility of maximizing the use of unstructured data, deep learning methods show great promise in this field.

The critical review presented in Chapter 3 highlights the potentialities of advanced EEG-based learning approaches to predict SZ. The described approaches represent a good starting point for the development of EEG-based discriminative models that might predict the onset of SZ, identify subjects at high risk of converting to psychosis, or differentiate SZ from other disorders, promoting more effective early interventions. The predictive value of future models may be transferred to clinical practice, reflecting the fruitful dialogue between engineering computational sciences, neuroscience, and psychiatry.

ERPs derived from EEG signals have been studied as potential biomarkers of SZ, particularly in auditory processing. Research has consistently found that individuals with SZ often show abnormalities in auditory information processing, which can be measured using ERP components. For example, abnormalities in the N100 ERP component have been suggested as potential biomarkers and discriminators of the disorder, reflecting deficits in early sensory processing. Moreover, some researchers suggest that the combination of multiple ERP components may provide a more comprehensive understanding of the neural abnormalities associated with SZ [60]. By following this line of thinking, we hypothesized that the neural activity dynamics probed with EEG would be altered in individuals with SZ, reflecting alterations in auditory processing.

The study presented in Chapter 4 aimed to characterize the pattern of the temporal EEG response to simple sounds in a passive listening task. The analysis of early auditory processing responses analysis allowed us to assess the consistency and robustness of altered event-related signal signatures as potential discriminators between SZ and HC subjects. The different confirmatory and exploratory analyses performed led to the following conclusions:

- The morphology of the ERP grand averages agrees with the pattern of the responses to auditory stimuli consistently reported in the literature [127]. It is possible to conclude that the auditory

evoked potentials are elicited in both HC and SZ.

- The ERP analysis confirmed the N100 and P200 amplitude reductions in SZ.
- The trial-by-trial analysis allowed us to assess the robustness of the elicitation of N100 and P200 components and the inter-trial variability. The results revealed a higher amplitude and latency inter-trial variability in SZ compared to HC.
- A cluster-based permutation test in a trial-by-trial approach identified significant spatiotemporal differences in the first 400 ms of responses to auditory stimuli. The first cluster was found in 66-146 ms time interval, with decreased amplitude in SZ patients over the entire scalp, with maximal differences in frontocentral regions. The second cluster was found in the 191-262 ms in the central region and the parietotemporal area of the right hemisphere, also showing a decreased amplitude in SZ. Both clusters coincide with the N100 and P200 ERP component localization.
- Both ERP and EEG trial-by-trial analyses results suggest impaired bottom-up auditory sensory-perceptual processes in SZ, which may reflect deficits in the initial detection and processing of auditory stimuli (as reflected by the N100), as well as the processing of more complex auditory features (as reflected by the P200).
- The time-frequency analysis performed on FCz voltage showed differences in power and synchrony of HC and SZ signals, revealing weaker and less synchronized responses in SZ.
- A cluster-based permutation test allowed us to conclude if those differences were significant. Only the synchronization of neural activity in theta and alpha frequency bands, measured by PLF, was statistically different between HC and SZ groups, suggesting impaired attentional resource allocation and deficient auditory processing.

The trial-by-trial analysis results supported the hypothesis that differences in single-trial auditory responses can improve the discrimination between HC and SZ. Furthermore, our results suggest that deficits in auditory processing may be robust biomarker candidates for SZ.

The insights gained from the literature review presented in Chapter 3 and the findings from the characterization of EEG signals have driven the exploration of deep learning methods to classify SZ. Chapter 5 described the development and application of a deep convolutional neural network for SZ classification based on single-trial EEG signals recorded in the passive listening task. A CNN-based architecture is proposed, which has been designated the SZNet model for ease of identification in this document. A 2D convolutions strategy allowed extracting information about amplitude variation over time across electrodes. For this purpose, the model received as input pseudo-images (two-dimensional structures) created by stacking the 1D single-trial signals captured from different electrodes. An ablative study allowed us to study variations of the same architecture to process signals from wider or narrower regions of the scalp and to understand the effect of signal lateralization (by hemisphere) on SZ classification. Using only five midline electrodes (Fz, FCz, Cz, CPz, and Pz), the proposed model achieved an average accuracy of 78%

in the discrimination between SZ and HC subjects. By ensembling predictions of 5 fits of this model, trained with different weights initialization, SZ classification achieved an accuracy of 80%. The Grad-CAM algorithm was applied to interpret the decoding of the EEG signal by the proposed deep neural network, which produced saliency maps of the critical regions in the input images for the classification. The most relevant features to identify HC subjects corresponded to the time windows of the ascendant phase of the N100 deflection and the descendent phase of the P200 component. This result aligns with significant spatiotemporal clusters found in the trial-by-trial analysis. In turn, the most crucial time window to distinguish SZ patients corresponded to the initial phase of N100 deflection, overriding the elicitation of the P50 component, whose latency seems to vary substantially between the two groups.

In addition to the proposed deep learning architecture, classical machine learning methods were also tested to infer the advantage of using one approach or another. An RF model performed the SZ classification on 10, 18, and 33 features extracted from the time domain of the EEG signal. The features were selected based on the work described in Chapter 4 and the results of the Grad-Cam interpretation of the SZNet decoding. The features that had the most significant impact on the result of the RF classification were the magnitude of N100 in frontocentral electrodes, the slope value of the N100 to P200 transition, and the slope of the descendent phase of P200. However, the results obtained in all classification attempts with this classical machine learning method did not exceed the performance of SzNet, reaching a maximum value of 73% in accuracy. These results demonstrate the potential of the proposed deep learning model for SZ classification.

In conclusion, the deep network allowed the automatic learning of patterns from the time course and spatial distribution of EEG single-trials, capable of detecting alterations in brain indices of auditory processing in SZ, despite the significant heterogeneity of the disorder. SzNet provides a base model for future developments in SZ research, specifically (differential) diagnosis and prediction.

6.2 Limitations and Future Directions

Future research should control for the type and dose of antipsychotic medication to avoid confounding variables, isolate the disorder's effects, and ensure the results' validity since medication can affect EEG measures [392, 490]. Moreover, by controlling for medication use and dose, EEG studies can reduce medication-related variability in the data, improving the findings' accuracy and reliability and more accurately characterizing the neural abnormalities associated with SZ.

The major challenge of the work developed in this thesis was to achieve a data pool large enough to implement deep learning methods. Larger amounts of data provide the algorithm with a more diverse and representative set of examples to learn from and help prevent overfitting. The limited EEG data availability, especially well-characterized public EEG datasets with SZ patients and auditory processing tasks, posed a problem. Although the public dataset used in this study has mitigated the problem, the final dataset was very heterogeneous. The merged dataset was composed of data collected from healthy patients and patients in various disease stages (early and chronic stages). Future studies should assess whether SZNet performance is affected by illness stage and symptom severity. For that, a multi-class classification

approach should be considered, as well as the inclusion of a considerable amount of data from each SZ subgroup. Following the adaptation of the model to multiple classes, a critical approach might be to include data from subjects at risk of developing SZ or in prodromal stages. This development line may allow disease onset prediction and a better understanding of the course of the underlying auditory processing alterations.

The evolution of deep learning has been characterized by rapid advances in performance, scalability, and capabilities, leading to the development of increasingly sophisticated AI systems. New architectures can be explored in future work, taking advantage of cutting-edge methods to make SZ classification increasingly robust and effective based on responses to very simple sounds.

The interdisciplinary collaboration between Psychology, Psychiatry, and Biomedical Engineering holds great promise for the future of mental health. By harnessing the power of technology, the insights of psychological research, and the medical expertise of psychiatry, this integrated approach can potentially revolutionize the diagnosis, treatment, and prevention of psychiatric disorders, such as SZ, improving the lives of those affected by these complex disorders.

Appendices

Appendix A

EEG Databases Merging

We conducted an exploratory mass-univariate analysis at all sensors and time points to assess whether differences in EEG signals amplitude and topographical distribution of both datasets are significantly affected by variations in the stimuli frequency and intensity. To examine group spatiotemporal cluster differences between datasets *A* and *B*, a non-parametric clustering-based permutation method was performed with MNE-Python software [491]. The between-subjects study conducted required a prior averaging step, which consisted of averaging over all trials within every subject to produce evoked responses [441]. For the cluster-based tests we performed 1000 permutations, with a significance level of 0.01 and using the threshold-free cluster enhancement (TFCE) method (e=0.2;h=0; [433]). The entire temporal segments (-500 to 1500 ms related to stimulus onset) over the 64 electrodes were analyzed, comparing the data of HC and SZ subjects in the dataset *A* with those in the dataset *B*. Nevertheless, results are focused on the 35 channels used to implement the deep learning models discussed in the paper. The result of the spatiotemporal cluster-based permutation test on HC data is presented in Fig. 45. The permutation test revealed differences between HC-*A* and HC-*B*, although only in the F6 electrode at 141 and 143 ms, and in the CP6 at 150 ms after the stimulus these differences were significant ($p<.01$). The test conducted on SZ data is presented in Fig. 46. The cluster-based permutation proved that the differences between SZ-*A* and SZ-*B* were smaller than those detected in the HC data, and no difference was statistically significant ($p<.01$).

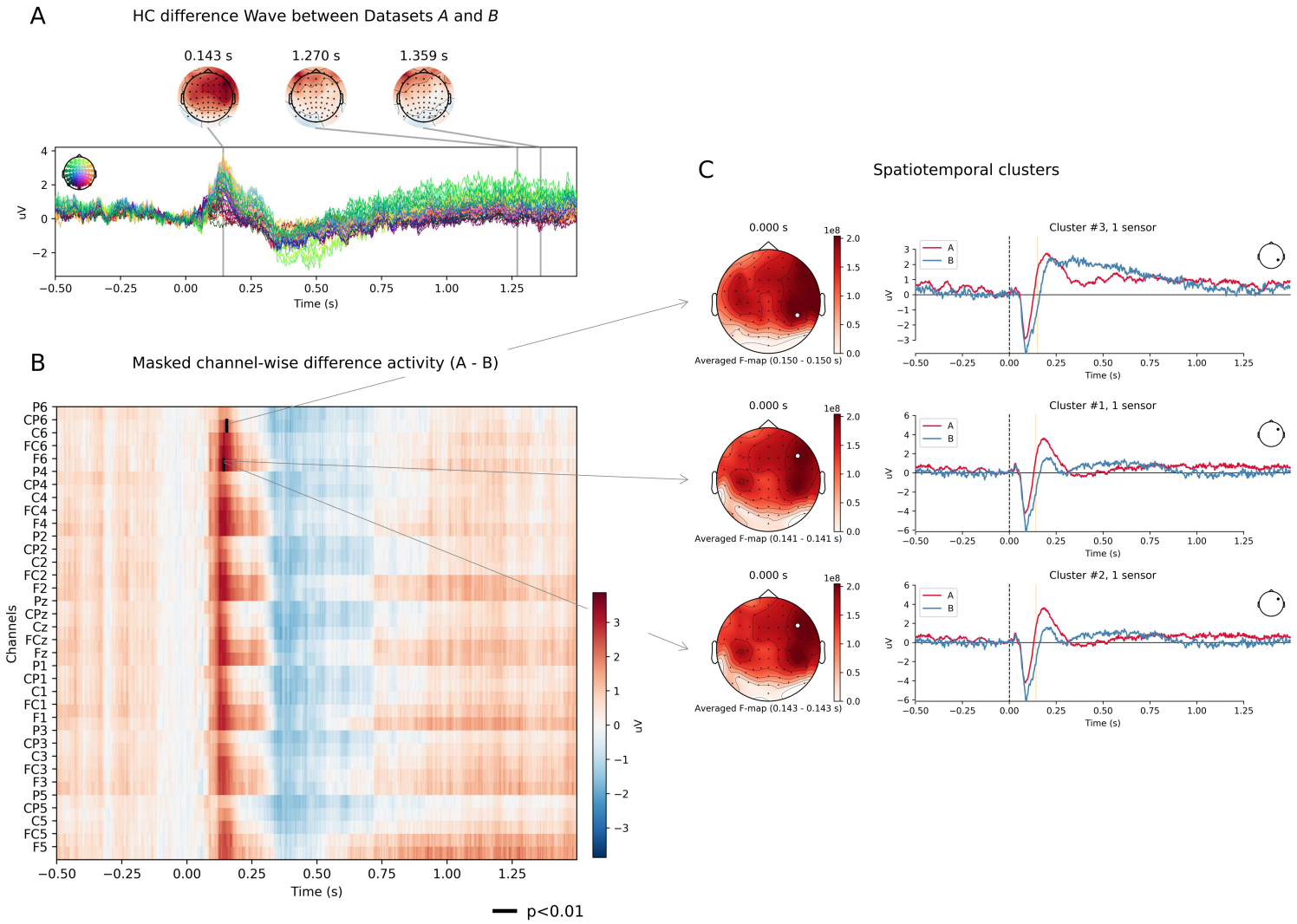


Figure 45: Results of spatiotemporal cluster-based permutation test on HC subjects data between datasets *A* and *B*. A) Difference wave contrasting dataset *A* and dataset *B*, with representative times points and respective topographical distribution of the signal amplitude. B) Channel-wise difference wave (*A-B*) masked for significance ($p < .01$). C) Topographical representation of the generated clusters that were significant, with the channels belonging to these clusters highlighted in white, and the corresponding temporal cluster shaded in orange.

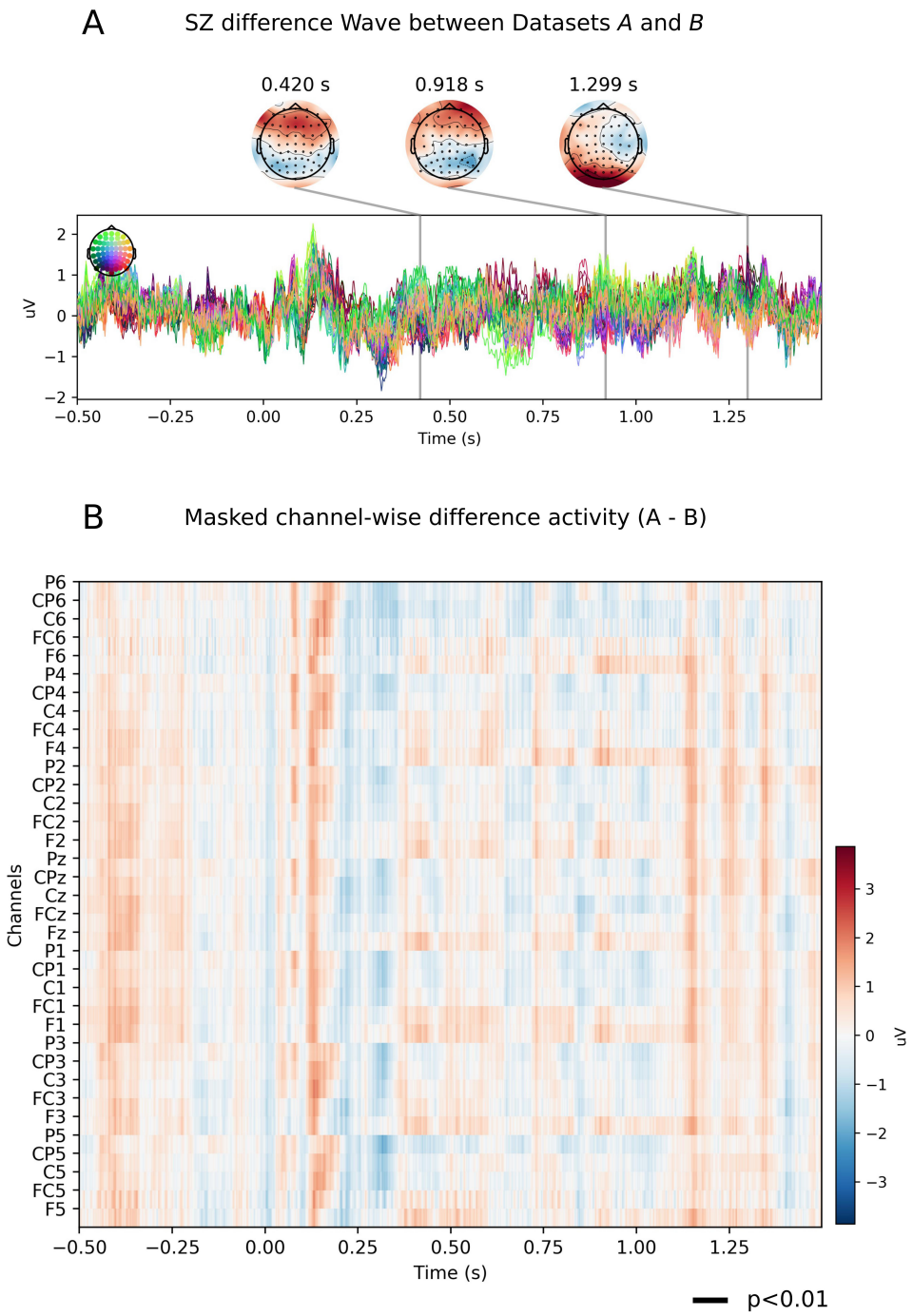


Figure 46: Results of spatiotemporal cluster-based permutation test on SZ subjects data between datasets *A* and *B*. A) Difference wave contrasting dataset *A* and dataset *B*, with representative times points and respective topographical distribution of the signal amplitude. B) Channel-wise difference wave (*A*-*B*) masked for significance ($p < .01$)

Cluster-based permutation tests were also performed only in the temporal dimension. They were conducted for each one of the 5 midline electrodes (Fz, FCz, Cz, CPz, and Pz) proposed in the SzNet model. The outputs of tests on HC and SZ data are presented in Fig. 47 and in Fig. 48, respectively, showing the differences between datasets *A* and *B*. As can be seen, more clusters were identified in the HC than in the SZ data, although the differences between datasets in these clusters were not statistically

significant. Although some clusters were generated in the time interval of interest (-100 to 400 ms relative to stimulus onset), especially in the HC data, we assume that the minor differences found do not compromise the merging of the two databases. The negligible differences between SZ subjects of both datasets may suggest that the features of the sounds emitted in the two tasks had no significant effect on the amplitude of these subjects' brain responses. Assuming this, the differences found in the HC data can be explained by high heterogeneity in this group of subjects.

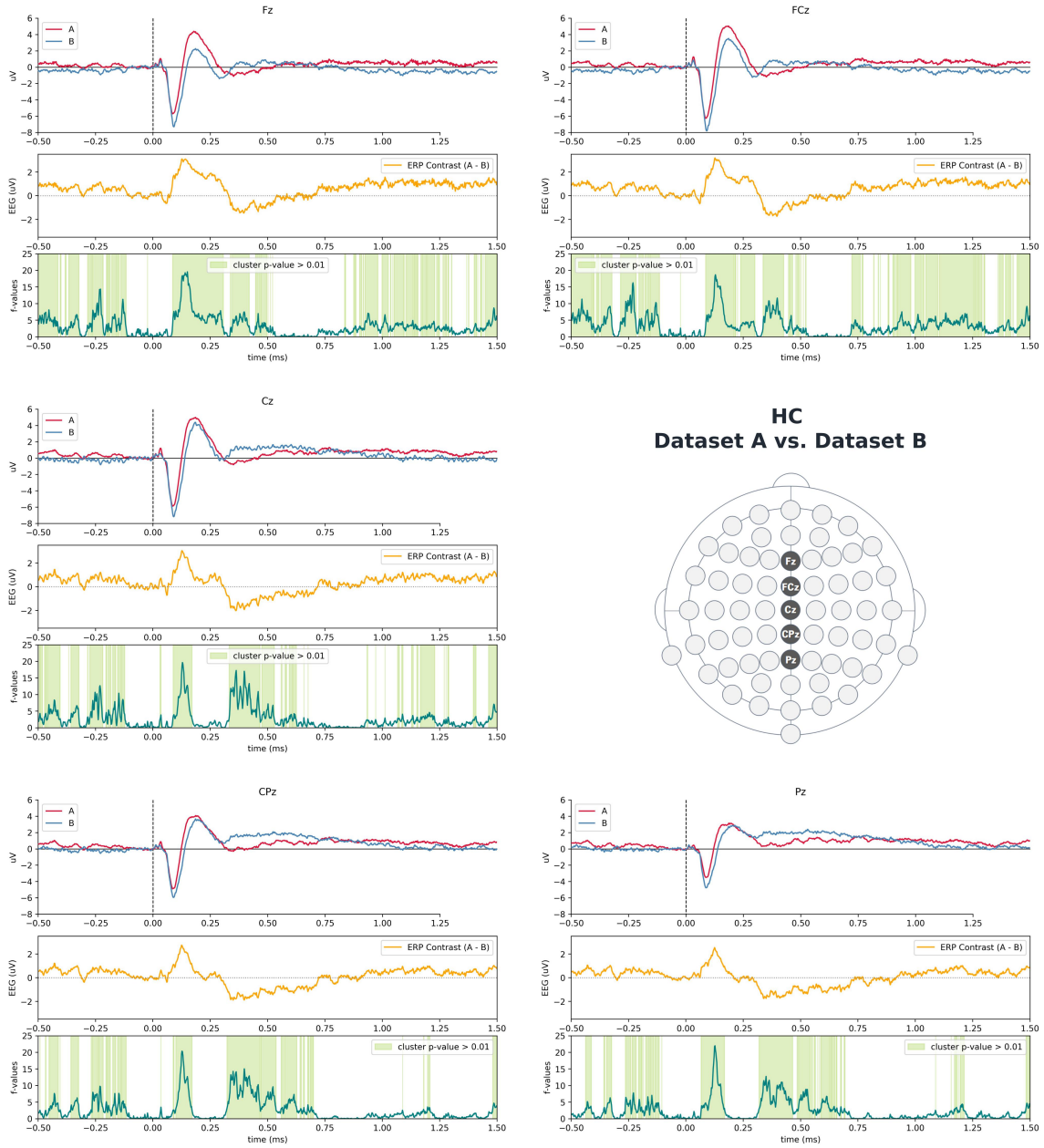


Figure 47: Cluster-based permutation test on temporal dimension of 5 midline electrodes to compare HC subjects data between datasets *A* and *B*. For each electrode analyzed, the evoked response averaged across all subjects of each dataset, the difference wave between both *A* and *B* signals, and the F-values map with generated clusters (green shaded) are presented from top to bottom, respectively.

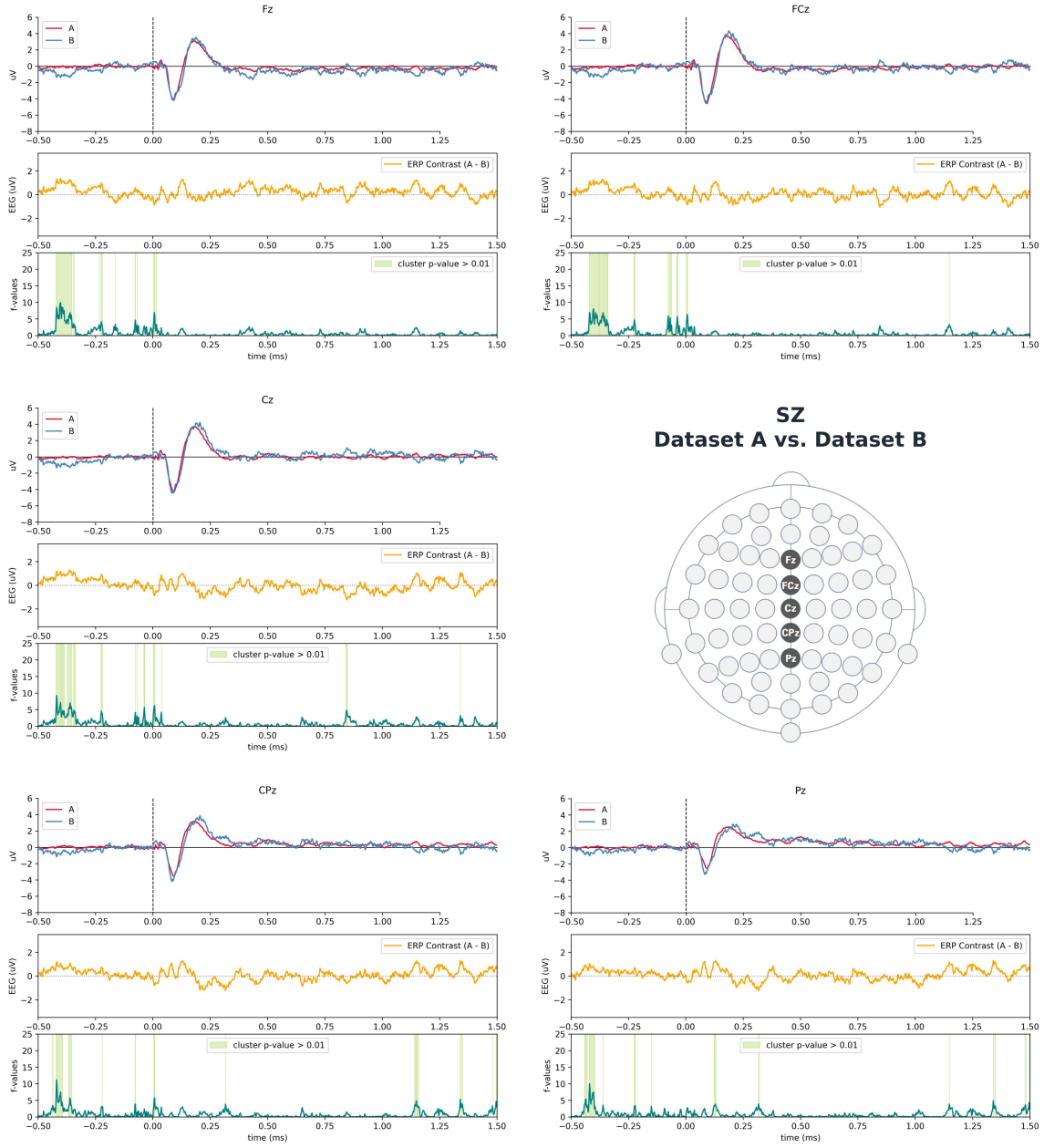


Figure 48: Cluster-based permutation test on temporal dimension of 5 midline electrodes to compare SZ subjects data between datasets *A* and *B*. For each electrode analyzed, the evoked response averaged across all subjects of each dataset, the difference wave between both *A* and *B* signals, and the F-values map with generated clusters (green shaded) are presented from top to bottom, respectively.

Appendix B

Deep Learning Model Interpretability

(In the next page is presented the figure with GRAD-CAM heatmaps.)

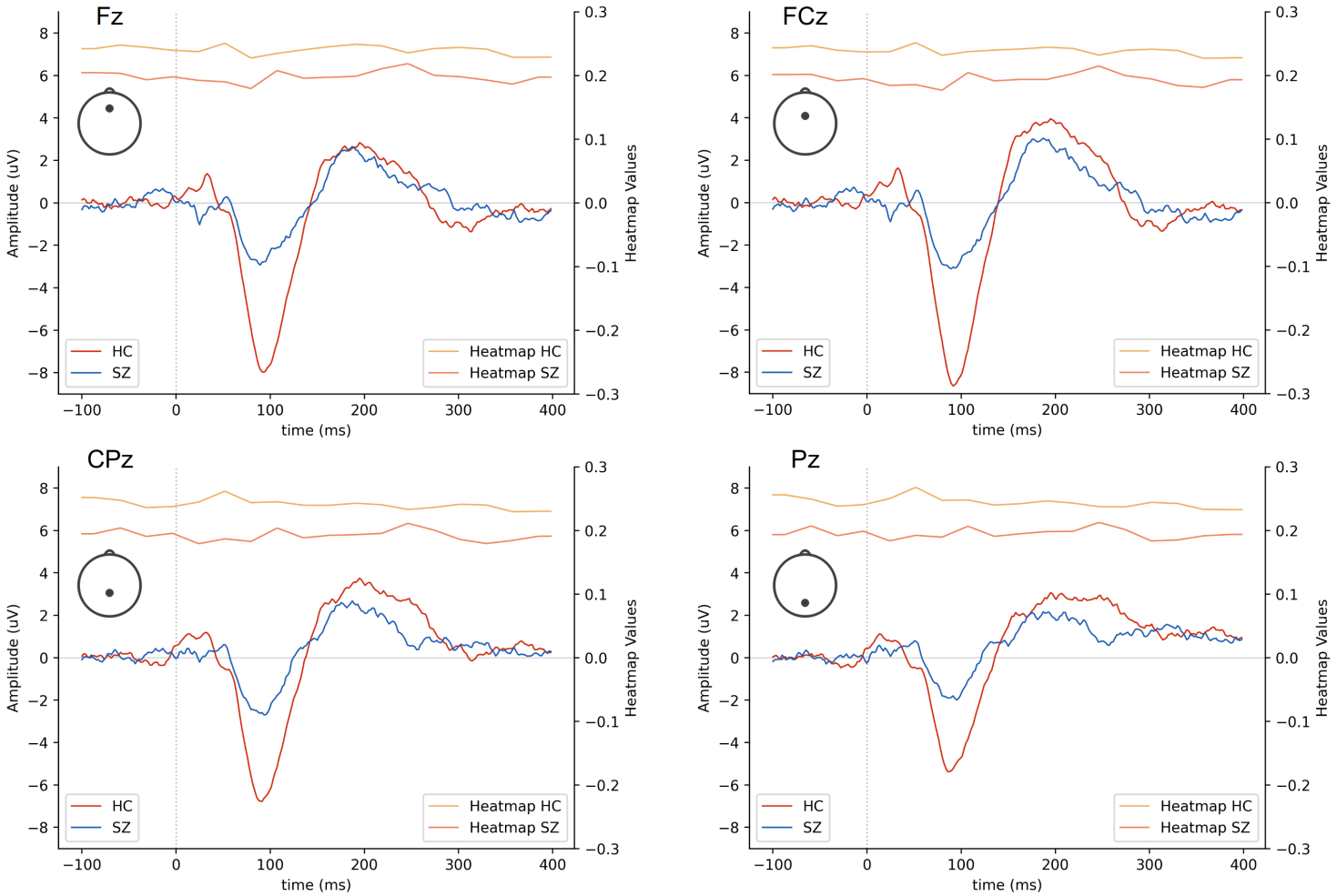


Figure 49: GRAD-CAM heatmaps cross section and averaged EEG signals for Fz, FCz, CPz, and Pz electrodes drawn for HC and SZ groups.

References

[1] O. Freudenreich. *Psychotic Disorders: A Practical Guide*. Current Clinical Psychiatry. Springer International Publishing, 2019. ISBN: 9783030294502.

[2] René S Kahn. "On the origins of schizophrenia". In: *American Journal of Psychiatry* 177.4 (2020), pp. 291–297.

[3] Huajie Jin and Iris Mosweu. "The societal cost of schizophrenia: a systematic review". In: *Pharmacoeconomics* 35.1 (2017), pp. 25–42.

[4] RS Kahn et al. "Schizophrenia. Nature Reviews Disease Primers". In: *Nov* 12.1 (2015), p. 15067.

[5] Assen Jablensky. "The diagnostic concept of schizophrenia: its history, evolution, and future prospects". In: *Dialogues in clinical neuroscience* 12.3 (2010), p. 271.

[6] W Wolfgang Fleischhacker et al. "Schizophrenia—time to commit to policy change". In: *Schizophrenia bulletin* 40.Suppl_3 (2014), S165–S194.

[7] Jeffrey A Lieberman, Scott A Small, and Ragy R Girgis. "Early detection and preventive intervention in schizophrenia: from fantasy to reality". In: *American Journal of Psychiatry* 176.10 (2019), pp. 794–810.

[8] Evelyn J Bromet and Shmuel Fennig. "Epidemiology and natural history of schizophrenia". In: *Biological Psychiatry* 46.7 (1999), pp. 871–881.

[9] Michael F Green, William P Horan, and Junghee Lee. "Social cognition in schizophrenia". In: *Nature Reviews Neuroscience* 16.10 (2015), pp. 620–631.

[10] Lara N Pantlin and Deana Davalos. "Neurophysiology for detection of high risk for psychosis". In: *Schizophrenia research and treatment* 2016 (2016).

[11] Kurt Schneider. *Clinical psychopathology.(Trans. by MW Hamilton)*. 1959.

[12] Colette M McKay, Donna M Headlam, and David L Copolov. "Central auditory processing in patients with auditory hallucinations". In: *American Journal of Psychiatry* 157.5 (2000), pp. 759–766.

[13] Wei-Xing Shi. "The auditory cortex in schizophrenia". In: *Biological psychiatry* 61.7 (2007), p. 829.

[14] Deanna M Barch and Alan Ceaser. "Cognition in schizophrenia: core psychological and neural mechanisms". In: *Trends in cognitive sciences* 16.1 (2012), pp. 27–34.

[15] Nikolaos P Moschopoulos et al. "Behavioral assessment of auditory processing deficits in schizophrenia: Literature review and suggestions for future research". In: *Scandinavian journal of psychology* 60.2 (2019), pp. 116–127.

[16] Hugo G Schnack. "Improving individual predictions: machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases)". In: *Schizophrenia research* 214 (2019), pp. 34–42.

[17] Steven J Luck and Emily S Kappenman. *The Oxford handbook of event-related potential components*. Oxford university press, 2011.

[18] Steven J Luck. "Event-related potentials". In: *APA handbook of research methods in psychology* 1 (2012), pp. 523–546.

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

[20] Mitja Bodatsch et al. "Forecasting psychosis by event-related potentials—systematic review and specific meta-analysis". In: *Biological psychiatry* 77.11 (2015), pp. 951–958.

[21] Gregory A Light et al. “Validation of mismatch negativity and P3a for use in multi-site studies of schizophrenia: characterization of demographic, clinical, cognitive, and functional correlates in COGS-2”. In: *Schizophrenia research* 163.1-3 (2015), pp. 63–72.

[22] Daniel C Javitt et al. “Neurophysiological biomarkers for drug development in schizophrenia”. In: *Nature Reviews Drug Discovery* 7.1 (2008), p. 68.

[23] Daniel C Javitt and Robert A Sweet. “Auditory dysfunction in schizophrenia: integrating clinical and basic features”. In: *Nature Reviews Neuroscience* 16.9 (2015), p. 535.

[24] Timm Rosburg, Nash N Boutros, and Judith M Ford. “Reduced auditory evoked potential component N100 in schizophrenia—a critical review”. In: *Psychiatry research* 161.3 (2008), pp. 259–274.

[25] Fernando Ferreira-Santos et al. “The auditory P200 is both increased and reduced in schizophrenia? A meta-analytic dissociation of the effect for standard and target stimuli in the oddball task”. In: *Clinical Neurophysiology* 123.7 (2012), pp. 1300–1308.

[26] Judith M Ford. “The difficulty in finding relationships between ERPs and clinical symptoms of schizophrenia”. In: *Clinical EEG and neuroscience* 49.1 (2018), pp. 6–7.

[27] Lee Anna Clark et al. “Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health’s Research Domain Criteria (RDoC)”. In: *Psychological Science in the Public Interest* 18.2 (2017), pp. 72–145.

[28] Samer Ellahham, Nour Ellahham, and Mecit Can Emre Simsekler. “Application of artificial intelligence in the health care safety context: opportunities and challenges”. In: *American Journal of Medical Quality* 35.4 (2020), pp. 341–348.

[29] J Warren Salmon and Stephen L Thompson. “Big Data: Information Technology as Control over the Profession of Medicine”. In: *The Corporatization of American Health Care*. Springer, 2021, pp. 181–254.

[30] Sabyasachi Dash et al. “Big data in healthcare: management, analysis and future prospects”. In: *Journal of Big Data* 6.1 (2019), pp. 1–25.

[31] Sascha Kraus et al. “Digital transformation in healthcare: Analyzing the current state-of-research”. In: *Journal of Business Research* 123 (2021), pp. 557–567.

[32] Andre Esteva et al. “A guide to deep learning in healthcare”. In: *Nature medicine* 25.1 (2019), pp. 24–29.

[33] Michael I Jordan and Tom M Mitchell. “Machine learning: Trends, perspectives, and prospects”. In: *Science* 349.6245 (2015), pp. 255–260.

[34] Ziad Obermeyer and Ezekiel J Emanuel. “Predicting the future—big data, machine learning, and clinical medicine”. In: *The New England journal of medicine* 375.13 (2016), p. 1216.

[35] Gyeongcheol Cho et al. “Review of machine learning algorithms for diagnosing mental illness”. In: *Psychiatry investigation* 16.4 (2019), p. 262.

[36] Chang Su et al. “Deep learning in mental health outcome research: a scoping review”. In: *Translational Psychiatry* 10.1 (2020), pp. 1–26.

[37] Danilo Bzdok and Andreas Meyer-Lindenberg. “Machine learning for precision psychiatry: opportunities and challenges”. In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 3.3 (2018), pp. 223–230.

[38] Elisa Veronese et al. “Machine learning approaches: from theory to application in schizophrenia”. In: *Computational and mathematical methods in medicine* 2013 (2013).

[39] Renato de Filippis et al. “Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: a systematic review”. In: *Neuropsychiatric Disease and Treatment* 15 (2019), pp. 1605–1627.

[40] Shradha Verma et al. “Machine learning techniques for the Schizophrenia diagnosis: A comprehensive review and future research directions”. In: *arXiv preprint arXiv:2301.07496* (2023).

[41] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[42] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[43] Jason K Johannesen et al. “Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults”. In: *Neuropsychiatric electrophysiology* 2.1 (2016), p. 3.

[44] Miseon Shim et al. “Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features”. In: *Schizophrenia research* 176.2-3 (2016), pp. 314–319.

[45] Lorenzo Santos-Mayo, Luis M San-José-Revuelta, and Juan Ignacio Arribas. “A computer-aided diagnosis system with EEG based on the P3b wave during an auditory odd-ball task in schizophrenia”. In: *IEEE transactions on biomedical engineering* 64.2 (2016), pp. 395–407.

[46] Jeremy A Taylor et al. “Auditory prediction errors as individual biomarkers of schizophrenia”. In: *NeuroImage: Clinical* 15 (2017), pp. 264–273.

[47] Reza Boostani and Malihe Sabeti. “Optimising brain map for the diagnosis of schizophrenia”. In: *International Journal of Biomedical Engineering and Technology* 28.2 (2018), pp. 105–119.

[48] Tsung-Hao Hsieh, Kuan-Yi Wu, and Sheng-Fu Liang. *Identification of Schizophrenic Patients and Healthy Controls Based on Musical Perception Using AEP Analysis*. 2018.

[49] Yoshua Bengio, Aaron Courville, and Pascal Vincent. “Representation learning: A review and new perspectives”. In: *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013), pp. 1798–1828.

[50] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *nature* 521.7553 (2015), p. 436.

[51] Minho Lee, Steven Bressler, and Robert Kozma. “Advances in Cognitive Engineering Using Neural Networks.” In: *Neural networks: the official journal of the International Neural Network Society* 92 (2017), p. 1.

[52] Robin Tibor Schirrmeister et al. “Deep learning with convolutional neural networks for EEG decoding and visualization”. In: *Human brain mapping* 38.11 (2017), pp. 5391–5420.

[53] Vernon J Lawhern et al. “EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces”. In: *Journal of neural engineering* 15.5 (2018), p. 056013.

[54] Pinkal Patel, Priya Aggarwal, and Anubha Gupta. “Classification of schizophrenia versus normal subjects using deep learning”. In: *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*. ACM. 2016, p. 28.

[55] Ling-Li Zeng et al. “Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI”. In: *EBioMedicine* 30 (2018), pp. 74–85.

[56] Muhammad Naveed Iqbal Qureshi, Jooyoung Oh, and Boreom Lee. “3D-CNN based Discrimination of Schizophrenia using Resting-state fMRI”. In: *Artificial Intelligence in Medicine* 98 (2019), pp. 10–17.

[57] Srivathsan Srinivasagopalan et al. “A deep learning approach for diagnosing schizophrenic patients”. In: *Journal of Experimental & Theoretical Artificial Intelligence* 31.6 (2019), pp. 1–14.

[58] Daniel C Javitt. “When doors of perception close: bottom-up models of disrupted cognition in schizophrenia”. In: *Annual review of clinical psychology* 5 (2009), pp. 249–275.

[59] David I Leitman et al. “The neural substrates of impaired prosodic detection in schizophrenia and its sensorial antecedents”. In: *American Journal of Psychiatry* 164.3 (2007), pp. 474–482.

[60] Bruce I. Turetsky et al. “Neurophysiological Endophenotypes of Schizophrenia: The Viability of Selected Candidate Measures”. In: *Schizophrenia Bulletin* 33.1 (Nov. 2006), pp. 69–94. ISSN: 0586-7614. DOI: 10.1093/schbul/sbl060.

[61] *Schizophrenia*. URL: http://www.who.int/mental\%5Fhealth/management/schizophrenia/en/.

[62] Miguel Gouveia et al. “O custo e a carga da esquizofrenia em Portugal em 2015”. In: *International Journal of Clinical Neurosciences and Mental Health* (2017).

[63] Piotr witaj et al. “Disability and schizophrenia: a systematic review of experienced psychosocial difficulties”. In: *BMC psychiatry* 12.1 (2012), p. 193.

[64] Dag Alnæs et al. “Brain heterogeneity in schizophrenia and its association with polygenic risk”. In: *JAMA psychiatry* 76.7 (2019), pp. 739–748.

[65] Eugen Bleuler. “Dementia praecox or the group of schizophrenias.” In: (1950).

[66] Emil Kraepelin. *Dementia praecox and paraphrenia*. Livingstone, 1919.

[67] Yekeen A Aderibigbe, D Theodoridis, and W Victor R Vieweg. “Dementia praecox to schizophrenia: the first 100 years”. In: *Psychiatry and Clinical neurosciences* 53.4 (1999), pp. 437–448.

[68] Rajiv Tandon et al. “Definition and description of schizophrenia in the DSM-5”. In: *Schizophrenia research* 150.1 (2013), pp. 3–10.

[69] P. Cowen, P. Harrison, and T. Burns. *Shorter Oxford Textbook of Psychiatry*. OUP Oxford, 2012. ISBN: 9780199605613.

[70] Krishna R Patel et al. “Schizophrenia: overview and treatment options”. In: *Pharmacy and Therapeutics* 39.9 (2014), p. 638.

[71] Christoph U Correll and Nina R Schooler. “Negative Symptoms in Schizophrenia: A Review and Clinical Guide for Recognition, Assessment, and Treatment”. In: *Neuropsychiatric Disease and Treatment* 16 (2020), p. 519.

[72] Robert A McCutcheon, Tiago Reis Marques, and Oliver D Howes. “Schizophrenia—an overview”. In: *JAMA psychiatry* 77.2 (2020), pp. 201–210.

[73] A. Crider. *Schizophrenia: A Biopsychological Perspective*. Psychology Revivals. Taylor & Francis, 2020. ISBN: 9781000089721.

[74] Suneeta Kumari et al. “An assessment of five (PANSS, SAPS, SANS, NSA-16, CGI-SCH) commonly used symptoms rating scales in schizophrenia and comparison to newer scales (CAINS, BNSS)”. In: *Journal of addiction research & therapy* 8.3 (2017).

[75] Bruce I Turetsky et al. “Profile of auditory information-processing deficits in schizophrenia”. In: *Psychiatry research* 165.1-2 (2009), pp. 27–37.

[76] James M Gold and Philip D Harvey. “Cognitive deficits in schizophrenia”. In: *Psychiatric Clinics* 16.2 (1993), pp. 295–312.

[77] Christopher R Bowie and Philip D Harvey. “Cognitive deficits and functional outcome in schizophrenia”. In: *Neuropsychiatric disease and treatment* 2.4 (2006), p. 531.

[78] Michael F Green and Philip D Harvey. “Cognition in schizophrenia: Past, present, and future”. In: *Schizophrenia Research: Cognition* 1.1 (2014), e1–e9.

[79] Yi Liu et al. “Cognitive deficits in subjects at risk for psychosis, first-episode and chronic schizophrenia patients”. In: *Psychiatry research* 274 (2019), pp. 235–242.

[80] Sophia Vinogradov and Srikantan Nagarajan. “Association of sensory processing with higher-order cognition and functioning in schizophrenia: mapping the world”. In: *JAMA psychiatry* 74.1 (2017), pp. 17–18.

[81] John Field. “An insight into listeners’ problems: Too much bottom-up or too much top-down?” In: *System* 32.3 (2004), pp. 363–377.

[82] Karl Friston. “Learning and inference in the brain”. In: *Neural Networks* 16.9 (2003), pp. 1325–1352.

[83] Moshe Bar. “The proactive brain: using analogies and associations to generate predictions”. In: *Trends in cognitive sciences* 11.7 (2007), pp. 280–289.

[84] Karsten Rauss and Gilles Pourtois. “What is bottom-up and what is top-down in predictive coding?” In: *Frontiers in psychology* 4 (2013), p. 276.

[85] Andrew Moskowitz and Dirk Corstens. “Auditory hallucinations: Psychotic symptom or dissociative experience?” In: *Journal of Psychological Trauma* 6.2-3 (2008), pp. 35–63.

[86] Janne-Elin Yttri, Annick Urfer-Parnas, and Josef Parnas. “Auditory Verbal Hallucinations in Schizophrenia: Mode of Onset and Disclosure”. In: *The Journal of Nervous and Mental Disease* 208.9 (2020), p. 689.

[87] Ralph E Hoffman. “Auditory/verbal hallucinations, speech perception neurocircuitry, and the social deafferentation hypothesis”. In: *Clinical EEG and Neuroscience* 39.2 (2008), pp. 87–90.

[88] Flavie Waters and Charles Fernyhough. “Hallucinations: A Systematic Review of Points of Similarity and Difference Across Diagnostic Classes”. In: *Schizophrenia Bulletin* 43.1 (Nov. 2016), pp. 32–43. ISSN: 0586-7614. DOI: 10.1093/schbul/sbw132.

[89] Anastasia Lim et al. “Prevalence and classification of hallucinations in multiple sensory modalities in schizophrenia spectrum disorders”. In: *Schizophrenia research* 176.2-3 (2016), pp. 493–499.

[90] Mitchell B Liester. “Toward a new definition of hallucination”. In: *American Journal of Orthopsychiatry* 68.2 (1998), pp. 305–312.

[91] Neil M McLachlan et al. “Auditory processing and hallucinations in schizophrenia”. In: *Schizophrenia research* 150.2-3 (2013), pp. 380–385.

[92] André Aleman et al. “Cognitive basis of hallucinations in schizophrenia: role of top-down information processing”. In: *Schizophrenia research* 64.2-3 (2003), pp. 175–185.

[93] Gaelle E Doucet et al. “Abnormal auditory tonotopy in patients with schizophrenia”. In: *NPJ schizophrenia* 5.1 (2019), pp. 1–6.

[94] Valentina Mancini et al. “Abnormal development and dysconnectivity of distinct thalamic nuclei in patients with 22q11. 2 deletion syndrome experiencing auditory hallucinations”. In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2020).

[95] Sarah M Haigh et al. “Abnormal auditory pattern perception in schizophrenia”. In: *Schizophrenia research* 176.2-3 (2016), pp. 473–479.

[96] Theresa M Marschall et al. “Deafferentation as a cause of hallucinations”. In: *Current Opinion in Psychiatry* 33.3 (2020), pp. 206–211.

[97] Biological Sciences Curriculum Study, National Institutes of Health, et al. “Information about Hearing, Communication, and Understanding”. In: *NIH Curriculum Supplement Series [Internet]*. National Institutes of Health (US), 2007.

[98] Sarah A Mangold and Joe M Das. “Neuroanatomy, Cortical Primary Auditory Area”. In: *StatPearls [Internet]*. StatPearls Publishing, 2020.

[99] DL Felten, MK O’Banion, and MS Maida. “14-sensory systems”. In: *Netter’s Atlas of Neuroscience* (2016), pp. 353–389.

[100] Dennis P Phillips. “Central auditory system and central auditory processing disorders: some conceptual issues”. In: *Seminars in Hearing*. Vol. 23. 04. Copyright© 2002 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New ... 2002, pp. 251–262.

[101] Diana C Peterson, Vamsi Reddy, and Renee N Hamel. “Neuroanatomy, Auditory Pathway”. In: *StatPearls [Internet]*. StatPearls Publishing, 2020.

[102] Gabriel M Swords et al. “Auditory system dysfunction in Alzheimer disease and its prodromal states: A review”. In: *Ageing research reviews* 44 (2018), pp. 49–59.

[103] Raksha A.Mudar Daniel A. Llano Gabriel M. Swords Lydia T. Nguyen. “Auditory system dysfunction in Alzheimer disease and itsprodromal states: A review”. In: *Ageing Research Reviews* 44.8 (2018), pp. 49–59.

[104] Jessica M Appler and Lisa V Goodrich. “Connecting the ear to the brain: Molecular mechanisms of auditory circuit assembly”. In: *Progress in neurobiology* 93.4 (2011), pp. 488–508.

[105] Dale Purves et al. “The Auditory Cortex”. In: *Neuroscience. 2nd edition*. Sinauer Associates, 2001.

[106] Rachel B Force, Noah C Venables, and Scott R Sponheim. “An auditory processing abnormality specific to liability for schizophrenia”. In: *Schizophrenia research* 103.1-3 (2008), pp. 298–310.

[107] Anthony J Rissling et al. “Disentangling early sensory information processing deficits in schizophrenia”. In: *Clinical Neurophysiology* 123.10 (2012), pp. 1942–1949.

[108] Michael L Thomas et al. “Modeling deficits from early auditory information processing to psychosocial functioning in schizophrenia”. In: *JAMA psychiatry* 74.1 (2017), pp. 37–46.

[109] Alice Medalia et al. “39.1 BASELINE EARLY AUDITORY PROCESSING AS A TAILORING VARIABLE TO PERSONALIZE COGNITIVE REMEDIATION FOR SCHIZOPHRENIA”. In: *Schizophrenia Bulletin* 45.Suppl 2 (2019), S151–S151.

[110] Daniel C Javitt. “Sensory processing in schizophrenia: neither simple nor intact”. In: *Schizophrenia bulletin* 35.6 (2009), pp. 1059–1064.

[111] Nikolaos Moschopoulos et al. “Auditory Processing Disorder may be present in schizophrenia and it is highly correlated with Formal Thought Disorder”. In: *Psychiatry Research* (2020), p. 113222.

[112] Adarsh Tripathi, Sujita Kumar Kar, and Rashmi Shukla. “Cognitive deficits in schizophrenia: understanding the biological correlates and remediation strategies”. In: *Clinical Psychopharmacology and Neuroscience* 16.1 (2018), p. 7.

[113] Ian S Ramsay et al. “Deficits in Auditory and Visual Sensory Discrimination Reflect a Genetic Liability for Psychosis and Predict Disruptions in Global Cognitive Functioning”. In: *Frontiers in Psychiatry* 11 (2020), p. 638.

[114] Judith M Ford et al. “Tuning in to the voices: a multisite FMRI study of auditory hallucinations”. In: *Schizophrenia bulletin* 35.1 (2009), pp. 58–66.

[115] Gemma Modinos et al. “Neuroanatomy of auditory verbal hallucinations in schizophrenia: a quantitative meta-analysis of voxel-based morphometry studies”. In: *Cortex* 49.4 (2013), pp. 1046–1055.

[116] Igor D Bandeira et al. “Analysis of the superior temporal gyrus as a possible biomarker in schizophrenia using voxel-based morphometry of the brain magnetic resonance imaging: A comprehensive review”. In: *CNS Spectr* 10 (2020), pp. 1–7.

[117] Xudong Chen et al. “Reduced cortical thickness in right Heschl’s gyrus associated with auditory verbal hallucinations severity in first-episode schizophrenia”. In: *BMC psychiatry* 15.1 (2015), p. 152.

[118] Lynn Mørch-Johnsen et al. “Auditory cortex characteristics in schizophrenia: associations with auditory hallucinations”. In: *Schizophrenia bulletin* 43.1 (2017), pp. 75–83.

[119] María Jose Escarti et al. “Auditory hallucinations in first-episode psychosis: A voxel-based morphometry study”. In: *Schizophrenia research* 209 (2019), pp. 148–155.

[120] Thomas Dierks et al. “Activation of Heschl’s gyrus during auditory hallucinations”. In: *Neuron* 22.3 (1999), pp. 615–621.

[121] Renaud Jardri et al. “Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis”. In: *American Journal of Psychiatry* 168.1 (2011), pp. 73–81.

[122] Paul Allen et al. “Neuroimaging auditory hallucinations in schizophrenia: from neuroanatomy to neurochemistry and beyond”. In: *Schizophrenia bulletin* 38.4 (2012), pp. 695–703.

[123] H Jordan Garber et al. “Use of magnetic resonance imaging in psychiatry.” In: *The American Journal of Psychiatry* (1988).

[124] Marcio Junior Sturzbecher and Draulio Barros de Araujo. “Simultaneous EEG-fMRI: Integrating spatial and temporal resolution”. In: *The relevance of the time domain to neural network models* (2012), pp. 199–217.

[125] Marc R Nuwer et al. “IFCN standards for digital recording of clinical EEG”. In: *Electroencephalography and clinical Neurophysiology* 106.3 (1998), pp. 259–261.

[126] Margitta Seeck et al. “The standardized EEG electrode array of the IFCN”. In: *Clinical Neurophysiology* 128.10 (2017), pp. 2070–2077.

[127] Steven J Luck. *An introduction to the event-related potential technique*. MIT press, 2014.

[128] Espen Hagen et al. “Multimodal modeling of neural network activity: computing LFP, ECoG, EEG, and MEG signals with LFPy 2.0”. In: *Frontiers in neuroinformatics* 12 (2018), p. 92.

[129] Timo Kirschstein and Rüdiger Köhling. “What is the source of the EEG?” In: *Clinical EEG and neuroscience* 40.3 (2009), pp. 146–149.

[130] György Buzsáki, Costas A Anastassiou, and Christof Koch. “The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes”. In: *Nature reviews neuroscience* 13.6 (2012), pp. 407–420.

[131] R.F. Burkard, J.J. Eggermont, and M. Don. *Auditory Evoked Potentials: Basic Principles and Clinical Application*. Point (Lippincott Williams and Wilkins) Series. Lippincott Williams & Wilkins, 2007. ISBN: 9780781757560.

[132] Michael X Cohen. “Where does EEG come from and what does it mean?” In: *Trends in neurosciences* 40.4 (2017), pp. 208–218.

[133] Arturo Tozzi, Marzieh Zare, and April A Benasich. “New perspectives on spontaneous brain activity: dynamic networks and energy matter”. In: *Frontiers in human neuroscience* 10 (2016), p. 247.

[134] Fengyu Cong, Tapani Ristaniemi, and Heikki Lyytinen. *Advanced signal processing on brain event-related potentials: filtering ERPs in time, frequency and space domains sequentially and simultaneously*. Vol. 13. World Scientific, 2015.

[135] Jazmín Ximena Suárez-Revelo, John Fredy Ochoa-Gómez, and Carlos Andrés Tobón-Quintero. "Validation of EEG Pre-processing Pipeline by Test-Retest Reliability". In: *Communications in Computer and Information Science* 916 (2018), pp. 290–299. ISSN: 18650929. DOI: 10.1007/978-3-030-00353-1_26.

[136] Jack B Nitschke, Gregory A Miller, and Edwin W Cook. "Digital filtering in EEG/ERP analysis: Some technical and empirical comparisons". In: *Behavior Research Methods, Instruments, & Computers* 30.1 (1998), pp. 54–67.

[137] Andreas Widmann, Erich Schröger, and Burkhard Maess. "Digital filter design for electrophysiological data–a practical approach". In: *Journal of neuroscience methods* 250 (2015), pp. 34–46.

[138] Andreas Widmann and Erich Schröger. "Filter effects and filter artifacts in the analysis of electrophysiological data". In: *Frontiers in psychology* 3 (2012), p. 233.

[139] Darren Tanner, Kara Morgan-Short, and Steven J Luck. "How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition". In: *Psychophysiology* 52.8 (2015), pp. 997–1009.

[140] Tzyy-Ping Jung et al. "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects". In: *Clinical neurophysiology* 111.10 (2000), pp. 1745–1758.

[141] Arnaud Delorme and Scott Makeig. "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis". In: *Journal of neuroscience methods* 134.1 (2004), pp. 9–21.

[142] Paul L Nunez and Ramesh Srinivasan. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006.

[143] Anton ML Coenen. *Neuronal Activities Underlying the Electroencephalogram and Evoked-Potentials of Sleeping and Waking-Implications for Information-Processing*. 1995.

[144] Terence W Picton et al. *Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria*. 2000.

[145] Kelly L Tremblay et al. "Auditory training alters the physiological detection of stimulus-specific cues in humans". In: *Clinical Neurophysiology* 120.1 (2009), pp. 128–135.

[146] Gabriella Musacchia and Charles E Schroeder. "Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex". In: *Hearing research* 258.1-2 (2009), pp. 72–79.

[147] Xiaxia Xu, Ileana L Hanganu-Opatz, and Malte Bieler. "Cross-talk of low-level sensory and high-level cognitive processing: development, mechanisms, and relevance for cross-modal abilities of the brain". In: *Frontiers in Neurorobotics* 14 (2020), p. 7.

[148] Gerard B Remijn et al. "An introduction to the measurement of auditory event-related potentials (ERPs)". In: *Acoustical Science and Technology* 35.5 (2014), pp. 229–242.

[149] Alexandra P Key. "Human Auditory Processing: Insights from Cortical Event-related Potentials". In: (2016).

[150] Michael Plöchl et al. "Oscillatory activity in auditory cortex reflects the perceptual level of audio-tactile integration". In: *Scientific reports* 6.1 (2016), pp. 1–10.

[151] Alan D Legatt. “Intraoperative monitoring of auditory evoked potentials”. In: *Disorders of peripheral and central auditory processing. Handbook of clinical neurophysiology* 10 (2013), pp. 235–252.

[152] Hillel Pratt et al. “The auditory P50 component to onset and offset of sound”. In: *Clinical Neurophysiology* 119.2 (2008), pp. 376–387.

[153] Martin Kurthen et al. “Towards a functional topography of sensory gating areas: invasive P50 recording and electrical stimulation mapping in epilepsy surgery candidates”. In: *Psychiatry Research: Neuroimaging* 155.2 (2007), pp. 121–133.

[154] C Liegeois-Chauvel et al. “Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components”. In: *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 92.3 (1994), pp. 204–214.

[155] Regina Weisser et al. “Is frontal lobe involved in the generation of auditory evoked P50?” In: *Neuroreport* 12.15 (2001), pp. 3303–3307.

[156] B Godey et al. “Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients”. In: *Clinical neurophysiology* 112.10 (2001), pp. 1850–1859.

[157] George Stothart and Nina Kazanina. “Auditory perception in the aging brain: the role of inhibition and facilitation in early processing”. In: *Neurobiology of aging* 47 (2016), pp. 23–34.

[158] JC Edgar et al. “Interpreting abnormality: an EEG and MEG study of P50 and the auditory paired-stimulus paradigm”. In: *Biological psychology* 65.1 (2003), pp. 1–20.

[159] Chen-Lan Shen et al. “P50, N100, and P200 Auditory Sensory Gating Deficits in Schizophrenia Patients”. In: *Frontiers in Psychiatry* 11 (2020), p. 868.

[160] Robert Freedman et al. “P50 inhibitory sensory gating in schizophrenia: analysis of recent studies”. In: *Schizophrenia research* 218 (2020), pp. 93–98.

[161] LA Jones et al. “Cognitive mechanisms associated with auditory sensory gating”. In: *Brain and cognition* 102 (2016), pp. 33–45.

[162] David A Smith, NASHAAT N BOUTORS, and STEVEN B SCHWARZOPF. “Reliability of P50 auditory event-related potential indices of sensory gating”. In: *Psychophysiology* 31.5 (1994), pp. 495–502.

[163] David Potter et al. “Review of clinical correlates of P50 sensory gating abnormalities in patients with schizophrenia”. In: *Schizophrenia bulletin* 32.4 (2006), pp. 692–700.

[164] Dylan M Smith et al. “Auditory verbal hallucinations in schizophrenia correlate with P50 gating”. In: *Clinical Neurophysiology* 124.7 (2013), pp. 1329–1335.

[165] Katja K Schulze et al. “P50 auditory evoked potential suppression in bipolar disorder patients with psychotic features and their unaffected relatives”. In: *Biological Psychiatry* 62.2 (2007), pp. 121–128.

[166] Thomas C Neylan et al. “Sensory gating in chronic posttraumatic stress disorder: reduced auditory P50 suppression in combat veterans”. In: *Biological psychiatry* 46.12 (1999), pp. 1656–1664.

[167] Eduardo S Ghisolfi et al. “P50 sensory gating in panic disorder”. In: *Journal of Psychiatric Research* 40.6 (2006), pp. 535–540.

[168] Jens D Rollnik. “Clinical neurophysiology of neurologic rehabilitation”. In: *Handbook of clinical neurology*. Vol. 161. Elsevier, 2019, pp. 187–194.

[169] Toshiaki Onitsuka, Naoya Oribe, and Shigenobu Kanba. “Neurophysiological findings in patients with bipolar disorder”. In: *Supplements to Clinical neurophysiology*. Vol. 62. 7. Elsevier, 2013, pp. 197–206.

[170] Rajiv Radhakrishnan et al. “Cannabis, Cannabinoids, and the Association with Psychosis”. In: *The effects of drug abuse on the human nervous system*. Elsevier, 2014, pp. 423–474.

[171] Jack S Fogarty, Robert J Barry, and Genevieve Z Steiner. “The First 250 ms of Auditory Processing: No Evidence of Early Processing Negativity in the Go/NoGo Task”. In: *Scientific reports* 10.1 (2020), pp. 1–12.

[172] BF O’donnell et al. “Auditory event-related potential abnormalities in bipolar disorder and schizophrenia”. In: *International Journal of Psychophysiology* 53.1 (2004), pp. 45–55.

[173] Marijn Lijffijt et al. “Diminished P50, N100 and P200 auditory sensory gating in bipolar I disorder”. In: *Psychiatry research* 167.3 (2009), pp. 191–201.

[174] Dean F Salisbury, KC Collins, and Robert W McCarley. “Reductions in the N1 and P2 auditory event-related potentials in first-hospitalized and chronic schizophrenia”. In: *Schizophrenia bulletin* 36.5 (2010), pp. 991–1000.

[175] Claudia JP Simons et al. “Auditory P300 and N100 components as intermediate phenotypes for psychotic disorder: familial liability and reliability”. In: *Clinical neurophysiology* 122.10 (2011), pp. 1984–1990.

[176] Marijn Lijffijt et al. “Differential relationships of impulsivity or antisocial symptoms on P50, N100, or P200 auditory sensory gating in controls and antisocial personality disorder”. In: *Journal of psychiatric research* 46.6 (2012), pp. 743–750.

[177] Timm Rosburg. “Auditory N100 gating in patients with schizophrenia: a systematic meta-analysis”. In: *Clinical Neurophysiology* 129.10 (2018), pp. 2099–2111.

[178] Kate E Crowley and Ian M Colrain. “A review of the evidence for P2 being an independent component process: age, sleep and modality”. In: *Clinical neurophysiology* 115.4 (2004), pp. 732–744.

[179] Steven A Hillyard and Terence W Picton. “Electrophysiology of cognition”. In: *Comprehensive Physiology* (2011), pp. 519–584.

[180] Silke Paulmann, Martin Bleichner, and Sonja AE Kotz. “Valence, arousal, and task effects in emotional prosody processing”. In: *Frontiers in Psychology* 4 (2013), p. 345.

[181] Geoffrey F Woodman. “A brief introduction to the use of event-related potentials in studies of perception and attention”. In: *Attention, Perception, & Psychophysics* 72.8 (2010), pp. 2031–2046.

[182] Risto Näätänen, Anthony WK Gaillard, and Sirkka Mäntysalo. “Early selective-attention effect on evoked potential reinterpreted”. In: *Acta psychologica* 42.4 (1978), pp. 313–329.

[183] Risto Näätänen et al. “The mismatch negativity (MMN)–a unique window to disturbed central auditory processing in ageing and different clinical conditions”. In: *Clinical Neurophysiology* 123.3 (2012), pp. 424–458.

[184] Guillermo Horga et al. “Deficits in predictive coding underlie hallucinations in schizophrenia”. In: *Journal of Neuroscience* 34.24 (2014), pp. 8072–8082.

[185] Risto Näätänen and Seppo Kähkönen. “Central auditory dysfunction in schizophrenia as revealed by the mismatch negativity (MMN) and its magnetic equivalent MMNm: a review”. In: *International Journal of Neuropsychopharmacology* 12.1 (2009), pp. 125–135.

[186] Mitja Bodatsch et al. “Prediction of psychosis by mismatch negativity”. In: *Biological psychiatry* 69.10 (2011), pp. 959–966.

[187] Veronica B Perez et al. “Automatic auditory processing deficits in schizophrenia and clinical high-risk patients: forecasting psychosis risk with mismatch negativity”. In: *Biological psychiatry* 75.6 (2014), pp. 459–469.

[188] Holly K Hamilton et al. “Mismatch negativity but not P300 is associated with functional disability in schizophrenia”. In: *Schizophrenia bulletin* 44.3 (2018), pp. 492–504.

[189] Yao-qin Qiu et al. “P300 aberration in first-episode schizophrenia patients: a meta-analysis”. In: *PLoS One* 9.6 (2014), e97794.

[190] John Polich. “Updating P300: an integrative theory of P3a and P3b”. In: *Clinical neurophysiology* 118.10 (2007), pp. 2128–2148.

[191] Judith M Ford et al. “Neurophysiological studies of auditory verbal hallucinations”. In: *Schizophrenia bulletin* 38.4 (2012), pp. 715–723.

[192] Bruce I Turetsky et al. “The utility of P300 as a schizophrenia endophenotype and predictive biomarker: clinical and socio-demographic modulators in COGS-2”. In: *Schizophrenia research* 163.1-3 (2015), pp. 53–62.

[193] Elvira Bramon et al. “Meta-analysis of the P300 and P50 waveforms in schizophrenia”. In: *Schizophrenia research* 70.2-3 (2004), pp. 315–329.

[194] Yang-Whan Jeon and John Polich. “Meta-analysis of P300 and schizophrenia: Patients, paradigms, and practical implications”. In: *Psychophysiology* 40.5 (2003), pp. 684–701.

[195] Gregor Thut, Carlo Miniussi, and Joachim Gross. “The functional importance of rhythmic activity in the brain”. In: *Current Biology* 22.16 (2012), R658–R663.

[196] Thalía Harmony. “The functional significance of delta oscillations in cognitive processing”. In: *Frontiers in integrative neuroscience* 7 (2013), p. 83.

[197] Christoph S Herrmann et al. “EEG oscillations: from correlation to causality”. In: *International Journal of Psychophysiology* 103 (2016), pp. 12–21.

[198] Wolfgang Klimesch. “EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis”. In: *Brain research reviews* 29.2-3 (1999), pp. 169–195.

[199] Laura Lee Colgin. “Mechanisms and functions of theta rhythms”. In: *Annual review of neuroscience* 36 (2013), pp. 295–312.

[200] Wolfgang Klimesch. “Alpha-band oscillations, attention, and controlled access to stored information”. In: *Trends in cognitive sciences* 16.12 (2012), pp. 606–617.

[201] Andreas K Engel and Pascal Fries. “Beta-band oscillations—signalling the status quo?” In: *Current opinion in neurobiology* 20.2 (2010), pp. 156–165.

[202] Robert Schmidt et al. “Beta Oscillations in Working Memory, Executive Control of Movement and Thought, and Sensorimotor Function”. In: *Journal of Neuroscience* 39.42 (2019), pp. 8231–8238.

[203] Canan Başar-Eroglu et al. “Gamma-band responses in the brain: a short review of psychophysiological correlates and functional significance”. In: *International journal of psychophysiology* 24.1-2 (1996), pp. 101–112.

[204] Xiaoxuan Jia and Adam Kohn. “Gamma rhythms in the brain”. In: *PLoS Biol* 9.4 (2011), e1001045.

[205] Supratim Ray and John HR Maunsell. “Do gamma oscillations play a role in cerebral cortex?” In: *Trends in cognitive sciences* 19.2 (2015), pp. 78–85.

[206] Imen El Karoui et al. “Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: an intracranial study in humans”. In: *Cerebral cortex* 25.11 (2015), pp. 4203–4212.

[207] Pascal Fries. “A mechanism for cognitive dynamics: neuronal communication through neuronal coherence”. In: *Trends in cognitive sciences* 9.10 (2005), pp. 474–480.

[208] Mehmet Akin. “Comparison of wavelet transform and FFT methods in the analysis of EEG signals”. In: *Journal of medical systems* 26.3 (2002), pp. 241–247.

[209] Brian J Roach and Daniel H Mathalon. “Event-related EEG time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia”. In: *Schizophrenia bulletin* 34.5 (2008), pp. 907–926.

[210] Michael X Cohen. “A better way to define and describe Morlet wavelets for time-frequency analysis”. In: *NeuroImage* 199 (2019), pp. 81–86.

[211] Mallat Stephane. “A wavelet tour of signal processing”. In: *The Sparse Way* (1999).

[212] Raymundo Cassani and Tiago H Falk. “Spectrotemporal Modeling of Biomedical Signals: Theoretical Foundation and Applications”. In: (2019).

[213] Vadim V Grubov et al. “Time-frequency analysis of epileptic EEG patterns by means of empirical modes and wavelets”. In: *Saratov Fall Meeting 2014: Optical Technologies in Biophysics and Medicine XVI; Laser Physics and Photonics XVI; and Computational Biophysics*. Vol. 9448. International Society for Optics and Photonics. 2015, 94481Q.

[214] Kevin M Spencer. “Time to be spontaneous: a renaissance of intrinsic brain activity in psychosis research?” In: *Biological psychiatry* 76.6 (2014), pp. 434–435.

[215] Christian Robert. *Machine learning, a probabilistic perspective*. 2014.

[216] Kevin P Murphy. *Machine learning: a probabilistic perspective*. Adaptive Computation and Machine Learning series. MIT press, 2012. ISBN: 9780262018029.

[217] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2014.

[218] Miroslav Kubat. *An introduction to machine learning*. Springer, 2017.

[219] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. Vol. 344. John Wiley & Sons, 2009.

[220] Xiaojin Zhu and Andrew B Goldberg. “Introduction to semi-supervised learning”. In: *Synthesis lectures on artificial intelligence and machine learning* 3.1 (2009), pp. 1–130.

[221] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-supervised Learning*. Adaptive computation and machine learning. MIT Press, 2010. ISBN: 9780262514125.

[222] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. “Supervised machine learning: A review of classification techniques”. In: *Emerging artificial intelligence applications in computer engineering* 160.1 (2007), pp. 3–24.

[223] Sotiris B Kotsiantis, Ioannis D Zaharakis, and Panayiotis E Pintelas. “Machine learning: a review of classification and combining techniques”. In: *Artificial Intelligence Review* 26.3 (2006), pp. 159–190.

[224] Gareth James et al. *An introduction to statistical learning*. Vol. 112. Springer, 2013.

[225] Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. “Deep learning for electroencephalogram (EEG) classification tasks: a review”. In: *Journal of neural engineering* 16.3 (2019), p. 031001.

[226] Robert D. Hof. “Deep Learning”. In: *Intelligent Machines, MIT Technology Review* (Apr. 2013).

[227] Maria Puig-Arnavat and Joan Carles Bruno. “Artificial neural networks for thermochemical conversion of biomass”. In: *Recent advances in thermo-chemical conversion of biomass*. Elsevier, 2015, pp. 133–156.

[228] B.G.M. Vandeginste et al. “Chapter 44 - Artificial Neural Networks”. In: *Handbook of Chemometrics and Qualimetrics: Part B*. Ed. by B.G.M. Vandeginste et al. Vol. 20. Data Handling in Science and Technology. Elsevier, 1998. DOI: https://doi.org/10.1016/S0922-3487(98)80054-3.

[229] Steven Walczak and Narciso Cerpa. “Artificial Neural Networks”. In: *Encyclopedia of Physical Science and Technology (Third Edition)*. Ed. by Robert A. Meyers. Third Edition. New York: Academic Press, 2003, pp. 631–645. ISBN: 978-0-12-227410-7. DOI: https://doi.org/10.1016/B0-12-227410-5/00837-1.

[230] Haohan Wang and Bhiksha Raj. “On the origin of deep learning”. In: *arXiv preprint arXiv:1702.07800* (2017).

[231] Tim Menzies et al. “Chapter 24 - Using Goals in Model-Based Reasoning”. In: *Sharing Data and Models in Software Engineering*. Ed. by Tim Menzies et al. Boston: Morgan Kaufmann, 2015, pp. 321–353. ISBN: 978-0-12-417295-1. DOI: https://doi.org/10.1016/B978-0-12-417295-1.00024-2.

[232] Jürgen Schmidhuber. “Deep learning in neural networks: An overview”. In: *Neural networks* 61 (2015), pp. 85–117.

[233] Simon Haykin. “Neural networks: A guided tour”. In: *Nonlinear Biomedical Signal Processing* 1 (2000), pp. 53–68.

[234] J. Patterson and A. Gibson. *Deep Learning: A Practitioner’s Approach*. O’Reilly Media, 2017. ISBN: 9781491914212.

[235] Abien Fred Agarap. “Deep Learning using Rectified Linear Units (ReLU)”. In: (2018).

[236] Lu Lu. “Dying ReLU and Initialization: Theory and Numerical Examples”. In: *Communications in Computational Physics* 28.5 (2020), pp. 1671–1706. ISSN: 1815-2406. DOI: 10.4208/cicp.OA-2020-0165. arXiv: 1903.06733v3.

[237] Timothy P Lillicrap et al. “Backpropagation and the brain”. In: *Nature Reviews Neuroscience* (2020), pp. 1–12.

[238] Yoshua Bengio. “Practical recommendations for gradient-based training of deep architectures”. In: *Neural networks: Tricks of the trade*. Springer, 2012, pp. 437–478.

[239] Tong Yu and Hong Zhu. “Hyper-Parameter Optimization: A Review of Algorithms and Applications”. In: *arXiv preprint arXiv:2003.05689* (2020).

[240] Kian Katanforoosh and Daniel Kunin. “Initializing neural networks”. In: *deeplearning.ai* (2019).

[241] Xavier Glorot and Yoshua Bengio. “Understanding the difficulty of training deep feedforward neural networks”. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. 2010, pp. 249–256.

[242] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. arXiv: 1412.6980 [cs.LG].

[243] George N Karystinos and Dimitrios A Pados. “On overfitting, generalization, and randomly expanded training sets”. In: *IEEE Transactions on Neural Networks* 11.5 (2000), pp. 1050–1057.

[244] Douglas M Hawkins. “The problem of overfitting”. In: *Journal of chemical information and computer sciences* 44.1 (2004), pp. 1–12.

[245] Jan Kukačka, Vladimir Golkov, and Daniel Cremers. *Regularization for Deep Learning: A Taxonomy*. 2017. arXiv: 1710.10686 [cs.LG].

[246] Nitish Srivastava et al. "Dropout: a simple way to prevent neural networks from overfitting". In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[247] Yashna Peerthum and Mark Stamp. "An empirical analysis of the shift and scale parameters in BatchNorm". In: *Information Sciences* 637 (2023), p. 118951.

[248] Sergey Ioffe and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *arXiv preprint arXiv:1502.03167* (2015).

[249] Nikolaus Kriegeskorte. "Deep neural networks: a new framework for modeling biological vision and brain information processing". In: *Annual review of vision science* 1 (2015), pp. 417–446.

[250] Jonghong Kim et al. "Convolutional neural network with biologically inspired retinal structure". In: *Procedia Computer Science* 88 (2016), pp. 145–154.

[251] Rikiya Yamashita et al. "Convolutional neural networks: an overview and application in radiology". In: *Insights into imaging* 9.4 (2018), pp. 611–629.

[252] Asifullah Khan et al. "A survey of the recent architectures of deep convolutional neural networks". In: *Artificial Intelligence Review* 53.8 (2020), pp. 5455–5516.

[253] David H Hubel and Torsten N Wiesel. "Receptive fields and functional architecture of monkey striate cortex". In: *The Journal of physiology* 195.1 (1968), pp. 215–243.

[254] Yann LeCun, Yoshua Bengio, et al. "Convolutional networks for images, speech, and time series". In: *The handbook of brain theory and neural networks* 3361.10 (1995), p. 1995.

[255] Ran Manor and Amir B Geva. "Convolutional neural network for multi-category rapid serial visual presentation BCI". In: *Frontiers in computational neuroscience* 9 (2015), p. 146.

[256] Nils J. Nilsson. "Neural Networks". In: *Artificial Intelligence: A New Synthesis*. Elsevier, 1998, pp. 37–57. DOI: 10.1016/B978-0-08-049945-1.50008-3.

[257] Sukanta Saha et al. "A systematic review of the prevalence of schizophrenia". In: *PLoS medicine* 2.5 (2005).

[258] Huey Yi Chong et al. "Global economic burden of schizophrenia: a systematic review". In: *Neuropsychiatric disease and treatment* 12 (2016), p. 357.

[259] Yash B Joshi and Gregory A Light. "Using EEG-Guided basket and umbrella trials in psychiatry: a precision medicine approach for cognitive impairment in schizophrenia". In: *Frontiers in psychiatry* 9 (2018), p. 554.

[260] Robert J Jirsaraie, Julia M Sheffield, and Deanna M Barch. "Neural correlates of global and specific cognitive deficits in schizophrenia". In: *Schizophrenia research* 201 (2018), pp. 237–242.

[261] Michael F Green et al. "Social cognition in schizophrenia, part 1: performance across phase of illness". In: *Schizophrenia bulletin* 38.4 (2012), pp. 854–864.

[262] Lui Unterrassner et al. "Psychotic-like experiences at the healthy end of the psychosis continuum". In: *Frontiers in psychology* 8 (2017), p. 775.

[263] Wei Lin Toh, Neil Thomas, and Susan L Rossell. "Auditory verbal hallucinations in bipolar disorder (BD) and major depressive disorder (MDD): A systematic review". In: *Journal of affective disorders* 184 (2015), pp. 18–28.

[264] Brent Elliott, Eileen Joyce, and Simon Shorvon. "Delusions, illusions and hallucinations in epilepsy: 1. Elementary phenomena". In: *Epilepsy research* 85.2-3 (2009), pp. 162–171.

[265] RJ Linscott and J Van Os. “An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders”. In: *Psychological medicine* 43.6 (2013), pp. 1133–1149. DOI: 10.1017/S0033291712001626.

[266] Hélène Verdoux and Jim van Os. “Psychotic symptoms in non-clinical populations and the continuum of psychosis”. In: *Schizophrenia research* 54.1-2 (2002), pp. 59–65.

[267] Jim Van Os et al. “A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness–persistence–impairment model of psychotic disorder”. In: *Psychological medicine* 39.2 (2009), pp. 179–195.

[268] Stephen M Lawrie et al. “The ‘continuum of psychosis’: scientifically unproven and clinically impractical”. In: *The British Journal of Psychiatry* 197.6 (2010), pp. 423–425.

[269] Jim van Os and Richard J Linscott. “Introduction: the extended psychosis phenotype—relationship with schizophrenia and with ultrahigh risk status for psychosis”. In: *Schizophrenia bulletin* 38.2 (2012), pp. 227–230.

[270] R Randeniya, LKL Oestreich, and MI Garrido. “Sensory prediction errors in the continuum of psychosis”. In: *Schizophrenia research* 191 (2018), pp. 109–122.

[271] Anna Wallace et al. “Detecting schizophrenia early: Prediagnosis healthcare utilization characteristics of patients with schizophrenia may aid early detection”. In: *Schizophrenia Research* 215 (2020), pp. 392–398.

[272] Thomas R Insel. “Rethinking schizophrenia”. In: *Nature* 468.7321 (2010), pp. 187–193.

[273] Molly K Larson, Elaine F Walker, and Michael T Compton. “Early signs, diagnosis and therapeutics of the prodromal phase of schizophrenia and related psychotic disorders”. In: *Expert review of neurotherapeutics* 10.8 (2010), pp. 1347–1359.

[274] Odin Van Der Stelt and Aysenil Belger. “Application of electroencephalography to the study of cognitive and brain functions in schizophrenia”. In: *Schizophrenia bulletin* 33.4 (2007), pp. 955–970.

[275] Daniel H Mathalon and Judith M Ford. “Neurobiology of schizophrenia: search for the elusive correlation with symptoms”. In: *Frontiers in Human Neuroscience* 6 (2012), p. 136.

[276] David Moher et al. “Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement”. In: *PLoS med* 6.7 (2009), e1000097.

[277] Wolfgang Gaebel et al. “Trends in Schizophrenia diagnosis and treatment”. In: *Advances in Psychiatry*. Springer, 2019, pp. 603–619.

[278] American Psychiatric Association et al. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[279] World Health Organization et al. “The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines”. In: *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire* 67.30 (1992), pp. 227–227.

[280] Steven E Hyman. “Can neuroscience be integrated into the DSM-V?” In: *Nature Reviews Neuroscience* 8.9 (2007), pp. 725–732.

[281] Martien J Kas et al. “A quantitative approach to neuropsychiatry: The why and the how”. In: *Neuroscience & Biobehavioral Reviews* 97 (2019), pp. 3–9.

[282] Thomas Insel et al. *Research domain criteria (RDoC): toward a new classification framework for research on mental disorders*. 2010.

[283] Bruce N Cuthbert and Thomas R Insel. “Toward the future of psychiatric diagnosis: the seven pillars of RDoC”. In: *BMC medicine* 11.1 (2013), p. 126.

[284] Julie Walsh-Messinger et al. “Relative importance of symptoms, cognition, and other multilevel variables for psychiatric disease classifications by machine learning”. In: *Psychiatry research* 278 (2019), pp. 27–34.

[285] Robin Nusslock et al. “Neurophysiological and Neuroimaging Techniques”. In: *The Encyclopedia of Clinical Psychology* (2014), pp. 1–9.

[286] K Kirk Shung, Michael Smith, and Benjamin MW Tsui. *Principles of medical imaging*. Academic Press, 2012.

[287] Vasileios Ioakeimidis et al. “A meta-analysis of structural and functional brain abnormalities in early-onset schizophrenia”. In: *Schizophrenia Bulletin Open* 1.1 (2020), sgaa016.

[288] Katherine H Karlsgodt, Daqiang Sun, and Tyrone D Cannon. “Structural and functional brain abnormalities in schizophrenia”. In: *Current directions in psychological science* 19.4 (2010), pp. 226–231.

[289] Martha E Shenton et al. “A review of MRI findings in schizophrenia”. In: *Schizophrenia research* 49.1-2 (2001), pp. 1–52.

[290] Chao Zhao et al. “Structural and functional brain abnormalities in schizophrenia: A cross-sectional study at different stages of the disease”. In: *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 83 (2018), pp. 27–32.

[291] René S Kahn and Richard SE Keefe. “Schizophrenia is a cognitive illness: time for a change in focus”. In: *JAMA psychiatry* 70.10 (2013), pp. 1107–1112.

[292] Raquel E Gur and Ruben C Gur. “Functional magnetic resonance imaging in schizophrenia”. In: *Dialogues in clinical neuroscience* 12.3 (2010), p. 333.

[293] Irene Alústiza et al. “Meta-analysis of functional magnetic resonance imaging studies of timing and cognitive control in schizophrenia and bipolar disorder: evidence of a primary time deficit”. In: *Schizophrenia research* 188 (2017), pp. 21–32.

[294] Tumbwene E Mwansisya et al. “Task and resting-state fMRI studies in first-episode schizophrenia: A systematic review”. In: *Schizophrenia research* 189 (2017), pp. 9–18.

[295] Nina V Kraguljac, Annusha Srivastava, and Adrienne C Lahti. “Memory deficits in schizophrenia: a selective review of functional magnetic resonance imaging (FMRI) studies”. In: *Behavioral Sciences* 3.3 (2013), pp. 330–347.

[296] Molly A Erickson, Emily S Kappenman, and Steven J Luck. “High temporal resolution measurement of cognitive and affective processes in psychopathology: What electroencephalography and magnetoencephalography can tell us about mental illness”. In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 3.1 (2018), pp. 4–6.

[297] Theodore P Beauchaine. “Role of biomarkers and endophenotypes in prevention and treatment of psychopathological disorders”. In: (2009).

[298] Cynthia S Weickert et al. “Biomarkers in schizophrenia: a brief conceptual consideration”. In: *Disease markers* 35.1 (2013), pp. 3–9.

[299] Matea Nikolac Perkovic et al. “Theranostic biomarkers for schizophrenia”. In: *International journal of molecular sciences* 18.4 (2017), p. 733.

[300] Daniela Rodrigues-Amorim et al. “Schizophrenia: a review of potential biomarkers”. In: *Journal of psychiatric research* 93 (2017), pp. 37–49.

[301] Martha Koukkou et al. “Neurobiology of Schizophrenia: Electrophysiological Indices”. In: *Advances in Psychiatry*. Springer, 2019, pp. 433–459.

[302] Kevin Ochsner, Stephen M Kosslyn, and Stephen Michael Kosslyn. *The Oxford Handbook of Cognitive Neuroscience, Volume 1: Core Topics*. Vol. 1. Oxford University Press, 2013.

[303] Suzie Lavoie et al. “Staging model in psychiatry: review of the evolution of electroencephalography abnormalities in major psychiatric disorders”. In: *Early intervention in psychiatry* 13.6 (2019), pp. 1319–1328.

[304] Risto Näätänen et al. “Mismatch negativity (MMN) deficiency: a break-through biomarker in predicting psychosis onset”. In: *International Journal of Psychophysiology* 95.3 (2015), pp. 338–344.

[305] Risto Näätänen, Juanita Todd, and Ulrich Schall. “Mismatch negativity (MMN) as biomarker predicting psychosis in clinically at-risk individuals”. In: *Biological psychology* 116 (2016), pp. 36–40.

[306] Holly A Earls, Tim Curran, and Vijay Mittal. “A meta-analytic review of auditory event-related potential components as endophenotypes for schizophrenia: perspectives from first-degree relatives”. In: *Schizophrenia bulletin* 42.6 (2016), pp. 1504–1516.

[307] Daniel H Mathalon et al. “P300 reduction and prolongation with illness duration in schizophrenia”. In: *Biological Psychiatry* 47.5 (2000), pp. 413–427.

[308] Elvira Bramon et al. “Is the P300 wave an endophenotype for schizophrenia? A meta-analysis and a family study”. In: *Neuroimage* 27.4 (2005), pp. 960–968.

[309] Patricia EG Bestelmeyer et al. “The P300 as a possible endophenotype for schizophrenia and bipolar disorder: Evidence from twin and patient studies”. In: *Psychiatry research* 169.3 (2009), pp. 212–219.

[310] Steven J Luck et al. “A roadmap for the development and validation of event-related potential biomarkers in schizophrenia research”. In: *Biological psychiatry* 70.1 (2011), pp. 28–34.

[311] Larry J Seidman et al. “Factor structure and heritability of endophenotypes in schizophrenia: findings from the Consortium on the Genetics of Schizophrenia (COGS-1)”. In: *Schizophrenia research* 163.1-3 (2015), pp. 73–79.

[312] Silvia Kyungjin Lho et al. “Predicting prognosis in patients with first-episode psychosis using auditory P300: a 1-year follow-up study”. In: *Clinical Neurophysiology* 130.1 (2019), pp. 46–54.

[313] Bruce I Turetsky et al. “Abnormal auditory N100 amplitude: a heritable endophenotype in first-degree relatives of schizophrenia probands”. In: *Biological psychiatry* 64.12 (2008), pp. 1051–1059.

[314] John J Foxe et al. “The N1 auditory evoked potential component as an endophenotype for schizophrenia: high-density electrical mapping in clinically unaffected first-degree relatives, first-episode, and chronic schizophrenia patients”. In: *European archives of psychiatry and clinical neuroscience* 261.5 (2011), pp. 331–339.

[315] Mei-Hua Hall et al. “The early auditory gamma-band response is heritable and a putative endophenotype of schizophrenia”. In: *Schizophrenia bulletin* 37.4 (2011), pp. 778–787.

[316] Toshiro Fujimoto et al. “Dysfunctional cortical connectivity during the auditory oddball task in patients with schizophrenia”. In: *The open neuroimaging journal* 7 (2013), p. 15.

[317] Grantley W Taylor, Robert W McCarley, and Dean F Salisbury. “Early auditory gamma band response abnormalities in first hospitalized schizophrenia”. In: *Supplements to Clinical neurophysiology*. Vol. 62. Elsevier, 2013, pp. 131–145.

[318] Emily Owens et al. “Electrophysiological endophenotypes for schizophrenia”. In: *Harvard review of psychiatry* 24.2 (2016), p. 129.

[319] Brian F O’Donnell et al. “The auditory steady-state response (ASSR): a translational biomarker for schizophrenia”. In: *Supplements to Clinical neurophysiology*. Vol. 62. Elsevier, 2013, pp. 101–112.

[320] BJ Roach et al. “Converging evidence for gamma synchrony deficits in schizophrenia”. In: *Supplements to Clinical neurophysiology*. Vol. 62. Elsevier, 2013, pp. 163–180.

[321] Hanna Thuné, Marc Recasens, and Peter J Uhlhaas. “The 40-Hz auditory steady-state response in patients with schizophrenia: a meta-analysis”. In: *JAMA psychiatry* 73.11 (2016), pp. 1145–1153.

[322] Vykinta Parciauskaite et al. “40-Hz auditory steady-state responses and the complex information processing: An exploratory study in healthy young males”. In: *Plos one* 14.10 (2019), e0223127.

[323] Donald C Goff et al. “Biomarkers for drug development in early psychosis: current issues and promising directions”. In: *European Neuropsychopharmacology* 26.6 (2016), pp. 923–937.

[324] Daniel C. Javitt et al. *ERP Biomarker Qualification Consortium*. `https://erpbiomarkers.org/` accessed 2019-09. URL: `https://erpbiomarkers.org/`.

[325] Prableen Kaur and Manik Sharma. “Diagnosis of Human Psychological Disorders using Supervised Learning and Nature-Inspired Computing Techniques: A Meta-Analysis”. In: *Journal of medical systems* 43.7 (2019), p. 204.

[326] Adrian BR Shatte, Delyse M Hutchinson, and Samantha J Teague. “Machine learning in mental health: a scoping review of methods and applications”. In: *Psychological medicine* 49.9 (2019), pp. 1426–1448.

[327] Georg Starke et al. “Computing schizophrenia: ethical challenges for machine learning in psychiatry”. In: *Psychological Medicine* (2020), pp. 1–7.

[328] M Sabeti, R Boostani, and SD Katebi. “A New approach to classify the schizophrenic and normal subjects by finding the best channels and frequency bands”. In: *2007 15th International Conference on Digital Signal Processing*. IEEE. 2007, pp. 123–126.

[329] Malihe Sabeti, Serajeddin Katebi, and Reza Boostani. “Entropy and complexity measures for EEG signal classification of schizophrenic and control participants”. In: *Artificial intelligence in medicine* 47.3 (2009), pp. 263–274.

[330] Mahsa Taghavi et al. “Usefulness of approximate entropy in the diagnosis of schizophrenia”. In: *Iranian journal of psychiatry and behavioral sciences* 5.2 (2011), p. 62.

[331] Tingting Xu, Massoud Stephane, and Keshab K Parhi. “Multidimensional analysis of the abnormal neural oscillations associated with lexical processing in schizophrenia”. In: *Clinical EEG and neuroscience* 44.2 (2013), pp. 135–143.

[332] Ming-Hsien Hiesh et al. “Classification of schizophrenia using genetic algorithm-support vector machine (ga-svm)”. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2013, pp. 6047–6050.

[333] Gary Marcel Hasey and Michael Kiang. “A review of recent literature employing electroencephalographic techniques to study the pathophysiology, phenomenology, and treatment response of schizophrenia”. In: *Current psychiatry reports* 15.9 (2013), p. 388.

[334] Jorne Laton et al. “In search of biomarkers for schizophrenia using electroencephalography”. In: *2014 International Workshop on Pattern Recognition in Neuroimaging*. IEEE. 2014, pp. 1–4.

[335] Andres H Neuhaus et al. “Critical evaluation of auditory event-related potential deficits in schizophrenia: evidence from large-scale single-subject pattern classification”. In: *Schizophrenia bulletin* 40.5 (2014), pp. 1062–1071.

[336] Lei Zhang. “EEG Signals Classification Using Machine Learning for The Identification and Diagnosis of Schizophrenia”. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 4521–4524.

[337] Christ Devia et al. “EEG Classification during Scene Free-Viewing for Schizophrenia Detection”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.6 (2019), pp. 1193–1199.

[338] B Thilakvathi et al. “EEG signal complexity analysis for schizophrenia during rest and mental activity”. In: (2017).

[339] Fali Li et al. “Differentiation of schizophrenia by combining the spatial EEG brain network patterns of rest and task P300”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.4 (2019), pp. 594–602.

[340] Alexandra Piryatinska, Boris Darkhovsky, and Alexander Kaplan. “Binary classification of multichannel-EEG records based on the -complexity of continuous vector functions”. In: *Computer methods and programs in biomedicine* 152 (2017), pp. 131–139.

[341] V Jahmunah et al. “Automated detection of schizophrenia using nonlinear signal processing methods”. In: *Artificial intelligence in medicine* 100 (2019), p. 101698.

[342] Ricardo Buettner et al. “High-performance exclusion of schizophrenia using a novel machine learning method on EEG data”. In: *IEEE Healthcom* (2019).

[343] Atefeh Goshvarpour and Ateke Goshvarpour. “Schizophrenia diagnosis using innovative EEG feature-level fusion schemes”. In: *Physical and Engineering Sciences in Medicine* (2020), pp. 1–12.

[344] Yumeng Ye et al. “Improvement of Resting-state EEG Analysis Process with Spectrum Weight-Voting based on LES”. In: *arXiv preprint arXiv:1712.07369* (2017).

[345] Haichun Liu et al. “A data driven approach for resting-state EEG signal classification of schizophrenia with control participants using random matrix theory”. In: *arXiv preprint arXiv:1712.05289* (2017).

[346] Yu Luo et al. “Discriminating schizophrenia disease progression using a P50 sensory gating task with dense-array EEG, clinical assessments, and cognitive tests”. In: *Expert review of neurotherapeutics* 19.5 (2019), pp. 459–470.

[347] Qi Chang et al. “EEG-based brain functional connectivity in first-episode schizophrenia patients, ultra-high-risk individuals, and healthy controls during P50 suppression”. In: *Frontiers in Human Neuroscience* 13 (2019), p. 379.

[348] David Ahmedt Aristizabal et al. “Identification of Children At Risk of Schizophrenia via Deep Learning and EEG Responses”. In: *IEEE Journal of Biomedical and Health Informatics* (2020).

[349] Fatemeh Alimardani et al. “Classification of bipolar disorder and schizophrenia using steady-state visual evoked potential based features”. In: *IEEE Access* 6 (2018), pp. 40379–40388.

[350] Peter J Uhlhaas and Wolf Singer. “Abnormal neural oscillations and synchrony in schizophrenia”. In: *Nature reviews neuroscience* 11.2 (2010), pp. 100–113.

[351] Boris Darkhovsky and Alexandra Pyriatinska. “Epsilon-complexity of continuous functions”. In: *arXiv preprint arXiv:1303.1777* (2013).

[352] Sarah Lippé, Natasa Kovacevic, and Randy McIntosh. “Differential maturation of brain signal complexity in the human auditory and visual system”. In: *Frontiers in Human Neuroscience* 3 (2009), p. 48.

[353] David J Schretlen et al. “Neuropsychological functioning in bipolar disorder and schizophrenia”. In: *Biological psychiatry* 62.2 (2007), pp. 179–186.

[354] Md Zahangir Alom et al. “A state-of-the-art survey on deep learning theory and architectures”. In: *Electronics* 8.3 (2019), p. 292.

[355] Shen Zhang et al. “Machine learning and deep learning algorithms for bearing fault diagnostics-a comprehensive review”. In: *arXiv preprint arXiv:1901.08247* (2019), pp. 4521–4524.

[356] Robin Tibor Schirrmeister et al. “Deep learning with convolutional neural networks for EEG decoding and visualization”. In: *Human brain mapping* 38.11 (2017), pp. 5391–5420.

[357] Oliver Faust et al. “Deep learning for healthcare applications based on physiological signals: A review”. In: *Computer methods and programs in biomedicine* 161 (2018), pp. 1–13.

[358] Chun-Ren Phang et al. “Classification of EEG-based brain connectivity networks in schizophrenia using a multi-domain connectome convolutional neural network”. In: *arXiv preprint arXiv:1903.08858* (2019).

[359] David Calhas, Enrique Romero, and Rui Henriques. “On the use of Pairwise Distance Learning for Brain Signal Classification with Limited Observations”. In: *arXiv preprint arXiv:1906.02076* (2019).

[360] Carlos Alberto Torres Naira and Cristian Jos´e L´opez Del Alamo. “Classification of People who Suffer Schizophrenia and Healthy People by EEG Signals using Deep Learning”. In: *International Journal of Advanced Computer Science and Applications* 10.10 (2019). DOI: 10.14569/IJACSA.2019.0101067.

[361] Shu Lih Oh et al. “Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals”. In: *Applied Sciences* 9.14 (2019), p. 2870.

[362] Lei Chu et al. “Individual recognition in schizophrenia using deep learning methods with random forest and voting classifiers: Insights from resting state EEG streams”. In: *arXiv preprint arXiv:1707.03467* (2017).

[363] Zachary C Lipton. “The mythos of model interpretability”. In: *Queue* 16.3 (2018), pp. 31–57.

[364] Grégoire Montavon, Wojciech Samek, and Klaus-Robert Müller. “Methods for interpreting and understanding deep neural networks”. In: *Digital Signal Processing* 73 (2018), pp. 1–15.

[365] Neeraj Tandon and Rajiv Tandon. “Using machine learning to explain the heterogeneity of schizophrenia. Realizing the promise and avoiding the hype”. In: *Schizophrenia research* 214 (2019), pp. 70–75.

[366] Dominic B Dwyer, Peter Falkai, and Nikolaos Koutsouleris. “Machine learning approaches for clinical psychology and psychiatry”. In: *Annual review of clinical psychology* 14 (2018), pp. 91–118.

[367] Gaël Varoquaux. “Cross-validation failure: small sample sizes lead to large error bars”. In: *Neuroimage* 180 (2018), pp. 68–77.

[368] Wlodzimierz Klonowski. “Everything you wanted to ask about EEG but were afraid to get the right answer”. In: *Nonlinear biomedical physics* 3.1 (2009), pp. 1–5.

[369] Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[370] Fabien Lotte et al. “A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update”. In: *Journal of neural engineering* 15.3 (2018), p. 031005.

[371] Minmin Zheng, Banghua Yang, and Yunlong Xie. “EEG classification across sessions and across subjects through transfer learning in motor imagery-based brain-machine interface system”. In: *Medical & biological engineering & computing* (2020).

[372] Deborah Dewey. “What is comorbidity and why does it matter in neurodevelopmental disorders?” In: *Current Developmental Disorders Reports* 5.4 (2018), pp. 235–242.

[373] Denise L Newman et al. “Comorbid mental disorders: implications for treatment and sample selection.” In: *Journal of abnormal psychology* 107.2 (1998), p. 305.

[374] Peter F Buckley et al. “Psychiatric comorbidities and schizophrenia”. In: *Schizophrenia bulletin* 35.2 (2009), pp. 383–402.

[375] Jack Tsai and Robert A Rosenheck. “Psychiatric comorbidity among adults with schizophrenia: A latent class analysis”. In: *Psychiatry research* 210.1 (2013), pp. 16–20.

[376] Kyung Soon Shin et al. “Aberrant auditory processing in schizophrenia and in subjects at ultra-high-risk for psychosis”. In: *Schizophrenia bulletin* 38.6 (2012), pp. 1258–1267.

[377] Dominik R Bach et al. “Evidence for impaired sound intensity processing in schizophrenia”. In: *Schizophrenia bulletin* 37.2 (2011), pp. 426–431.

[378] David I Leitman et al. “Not pitch perfect: sensory contributions to affective communication impairment in schizophrenia”. In: *Biological psychiatry* 70.7 (2011), pp. 611–618.

[379] Renaud Jardri et al. “Are hallucinations due to an imbalance between excitatory and inhibitory influences on the brain?” In: *Schizophrenia bulletin* 42.5 (2016), pp. 1124–1134.

[380] Henry Silver and Pablo Feldman. “Evidence for sustained attention and working memory in schizophrenia sharing a common mechanism”. In: *The Journal of neuropsychiatry and clinical neurosciences* 17.3 (2005), pp. 391–398.

[381] American Psychiatric Association et al. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[382] David Baumeister et al. “Auditory verbal hallucinations and continuum models of psychosis: a systematic review of the healthy voice-hearer literature”. In: *Clinical psychology review* 51 (2017), pp. 125–141.

[383] Derek K Tracy and Sukhwinder S Shergill. “Mechanisms underlying auditory hallucinations—understanding perception without stimulus”. In: *Brain sciences* 3.2 (2013), pp. 642–669.

[384] Toshiaki Onitsuka et al. “Review of neurophysiological findings in patients with schizophrenia”. In: *Psychiatry and clinical neurosciences* 67.7 (2013), pp. 461–470.

[385] M. Kubicki and M.E. Shenton. *Neuroimaging in Schizophrenia*. Springer International Publishing, 2020. ISBN: 9783030352066.

[386] Derya Şahin et al. “Early and middle latency auditory event-related potentials do not explain differences in neuropsychological performance between schizophrenia spectrum patients and matched healthy controls”. In: *Psychiatry Research* 304 (2021), p. 114162. ISSN: 18727123. DOI: 10.1016/j.psychres.2021.114162.

[387] Helena K. Kim, Daniel M. Blumberger, and Zafiris J. Daskalakis. “Neurophysiological Biomarkers in Schizophrenia—P50, Mismatch Negativity, and TMS-EMG and TMS-EEG”. In: *Frontiers in Psychiatry* 11 (2020). ISSN: 16640640. DOI: 10.3389/FPSYT.2020.00795.

[388] Giulia M. Giordano et al. “Tracing Links Between Early Auditory Information Processing and Negative Symptoms in Schizophrenia: An ERP Study”. In: *Frontiers in Psychiatry* 12 (2021), p. 790745. ISSN: 1664-0640. DOI: 10.3389/fpsyt.2021.790745.

[389] Xi Ren et al. “Deficits in attentional modulation of auditory N100 in first-episode schizophrenia”. In: *European Journal of Neuroscience* 53.8 (2021), pp. 2629–2638. ISSN: 14609568. DOI: 10.1111/EJN.15128.

[390] Judith M Ford et al. “Using concurrent EEG and fMRI to probe the state of the brain in schizophrenia”. In: *NeuroImage: Clinical* 12 (2016), pp. 429–441.

[391] C.A. Tamminga et al. *Psychotic Disorders: Comprehensive Conceptualization and Treatments*. Oxford University Press, 2020. ISBN: 9780190653286.

[392] Andrea Perrottelli et al. “EEG-Based Measures in At-Risk Mental State and Early Stages of Schizophrenia: A Systematic Review”. In: *Frontiers in Psychiatry* 12 (2021), p. 582. ISSN: 16640640. DOI: 10.3389/FPSYT.2021.653642/BIBTEX.

[393] Nicholas Murphy et al. “Intact Auditory Cortical Cross-Frequency Coupling in Early and Chronic Schizophrenia”. In: *Frontiers in Psychiatry* 11 (2020), p. 507. ISSN: 16640640. DOI: 10.3389/FPSYT.2020.00507/BIBTEX.

[394] Naoya Oribe et al. “Progressive reduction of auditory evoked gamma in first episode schizophrenia but not clinical high risk individuals”. In: *Schizophrenia Research* 208 (2019), pp. 145–152. ISSN: 0920-9964. DOI: 10.1016/J.SCHRES.2019.03.025.

[395] Yoji Hirano and Peter J. Uhlhaas. “Current findings and perspectives on aberrant neural oscillations in schizophrenia”. In: *Psychiatry and Clinical Neurosciences* 75.12 (2021), pp. 358–368. ISSN: 1323-1316. DOI: 10.1111/pcn.13300.

[396] Stefania Schiavone et al. “The Role of Parvalbumin Interneurons in Neurotransmitter Balance and Neurological Disease”. In: *Frontiers in Psychiatry | www.frontiersin.org* 1 (2021), p. 679960. DOI: 10.3389/fpsyt.2021.679960.

[397] A.C. Papanicolaou, T.P.L. Roberts, and J.W. Wheless. *Fifty Years of Magnetoencephalography: Beginnings, Technical Advances, and Applications*. Oxford University Press, 2020. ISBN: 9780190935702.

[398] Kevin M. Spencer. “The Functional Consequences of Cortical Circuit Abnormalities on Gamma Oscillations in Schizophrenia: Insights from Computational Modeling”. In: *Frontiers in Human Neuroscience* 3.OCT (2009). ISSN: 16625161. DOI: 10.3389/NEURO.09.033.2009.

[399] André Schmidt et al. “GABAergic Mechanisms in Schizophrenia: Linking Postmortem and In Vivo Studies”. In: *Article* 8 (2017), p. 1. DOI: 10.3389/fpsyt.2017.00118.

[400] Yi Liu et al. “A Selective Review of the Excitatory-Inhibitory Imbalance in Schizophrenia: Underlying Biology, Genetics, Microcircuits, and Symptoms”. In: *Frontiers in Cell and Developmental Biology* 9 (2021). ISSN: 2296634X. DOI: 10.3389/FCELL.2021.664535.

[401] C. Pantev et al. “Human auditory evoked gamma-band magnetic fields.” In: *Proceedings of the National Academy of Sciences* 88.20 (1991), pp. 8996–9000. ISSN: 0027-8424. DOI: 10.1073/pnas.88.20.8996.

[402] Gregor Leicht et al. “EEG-Informed fMRI Reveals a Disturbed Gamma-Band-Specific Network in Subjects at High Risk for Psychosis”. In: *Schizophrenia Bulletin* 42.1 (2016), pp. 239–249. DOI: 10.1093/schbul/sbv092.

[403] Baihan Wang et al. “Neurophysiology in psychosis: The quest for disease biomarkers”. In: *Translational Psychiatry 2022 12:1* 12.1 (2022), pp. 1–10. ISSN: 2158-3188. DOI: 10.1038/s41398-022-01860-x.

[404] K. Doege et al. “Reduced event-related low frequency EEG activity in schizophrenia during an auditory oddball task”. In: *Psychophysiology* 46.3 (2009), pp. 566–577. ISSN: 14698986. DOI: 10.1111/J.1469-8986.2009.00785.X.

[405] Erol Başar and Bahar Güntekin. “Review of delta, theta, alpha, beta, and gamma response oscillations in neuropsychiatric disorders”. In: 2013, pp. 303–341. DOI: 10.1016/B978-0-7020-5307-8.00019-3.

[406] Canan Başar-Eroğlu, Christina Schmiedt-Fehr, and Birgit Mathes. “Auditory-evoked alpha oscillations imply reduced anterior and increased posterior amplitudes in schizophrenia”. In: *Supplements to Clinical neurophysiology*. Vol. 62. Suppl Clin Neurophysiol, 2013, pp. 121–129. ISBN: 9780702053078. DOI: 10.1016/B978-0-7020-5307-8.00008-9.

[407] Bahar Güntekin and Erol Başar. “Review of evoked and event-related delta responses in the human brain”. In: *International Journal of Psychophysiology* 103 (2016), pp. 43–52. ISSN: 01678760. DOI: 10.1016/j.ijpsycho.2015.02.001.

[408] Yadira Roa Romero et al. “Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia”. In: *Journal of Neurophysiology* 116.3 (2016), pp. 1396–1407. ISSN: 15221598. DOI: 10.1152/JN.00096.2016/ASSET/IMAGES/LARGE/Z9K0091638040008.JPEG.

[409] William C. Hochberger et al. “Oscillatory biomarkers of early auditory information processing predict cognitive gains following targeted cognitive training in schizophrenia patients”. In: *Schizophrenia Research* 215 (2020), pp. 97–104. ISSN: 0920-9964. DOI: 10.1016/J.SCHRES.2019.11.015.

[410] Brian J. Roach et al. “Theta Phase Synchrony Is Sensitive to Corollary Discharge Abnormalities in Early Illness Schizophrenia but Not in the Psychosis Risk Syndrome”. In: *Schizophrenia Bulletin* 47.2 (2021), pp. 415–423. ISSN: 0586-7614. DOI: 10.1093/schbul/sbaa110.

[411] Ana P Pinheiro et al. “Abnormalities in the processing of emotional prosody from single words in schizophrenia”. In: *Schizophrenia research* 152.1 (2014), pp. 235–241.

[412] Michael S Kraus et al. “Basic auditory processing deficits and their association with auditory emotion recognition in schizophrenia”. In: *Schizophrenia Research* 204 (2019), pp. 155–161.

[413] Kirill V Nourski. “Auditory processing in the human cortex: An intracranial electrophysiology perspective”. In: *Laryngoscope investigative otolaryngology* 2.4 (2017), pp. 147–156.

[414] Gregory A Light and David L Braff. “Mismatch negativity deficits are associated with poor functioning in schizophrenia patients”. In: *Archives of general psychiatry* 62.2 (2005), pp. 127–136.

[415] Kevin M Spencer et al. “Left auditory cortex gamma synchronization and auditory hallucination symptoms in schizophrenia”. In: *BMC neuroscience* 10 (2009), pp. 1–13.

[416] Madiha Shaikh et al. “Reduced mismatch negativity predates the onset of psychosis”. In: *Schizophrenia research* 134.1 (2012), pp. 42–48.

[417] Anke Brockhaus-Dumke et al. “Sensory gating in schizophrenia: P50 and N100 gating in antipsychotic-free subjects at risk, first-episode, and chronic patients”. In: *Biological psychiatry* 64.5 (2008), pp. 376–384.

[418] Gregor Leicht et al. “Reduced early auditory evoked gamma-band response in patients with schizophrenia”. In: *Biological psychiatry* 67.3 (2010), pp. 224–231.

[419] Juliana Yordanova, Vasil Kolev, and John Polich. “P300 and alpha event-related desynchronization (ERD)”. In: *Psychophysiology* 38.1 (2001), pp. 143–152.

[420] Brian Roach. *EEG data from basic sensory task in Schizophrenia: Button press and auditory tone event related potentials from 81 human subjects*. =https://www.kaggle.com/broach/button-tone-sz. Dataset. 2017.

[421] Judith M Ford et al. "Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone". In: *Schizophrenia bulletin* 40.4 (2013), pp. 804–812.

[422] Leah I Hartman, R Walter Heinrichs, and Farzaneh Mashhadi. "The continuing story of schizophrenia and schizoaffective disorder: One condition or two?" In: *Schizophrenia Research: Cognition* 16 (2019), pp. 36–42.

[423] Ana P Pinheiro et al. "When temporal prediction errs: ERP responses to delayed action-feedback onset". In: *Neuropsychologia* 134 (2019), p. 107200.

[424] *Biosemi*. URL: `https://www.biosemi.com/`.

[425] Gabriele Gratton, Michael GH Coles, and Emanuel Donchin. "A new method for off-line removal of ocular artifact". In: *Electroencephalography and clinical neurophysiology* 55.4 (1983), pp. 468–484.

[426] Hugh Nolan, Robert Whelan, and Richard B Reilly. "FASTER: fully automated statistical thresholding for EEG artifact rejection". In: *Journal of neuroscience methods* 192.1 (2010), pp. 152–162.

[427] Carl M Gaspar, Guillaume A Rousselet, and Cyril R Pernet. "Reliability of ERP and single-trial analyses". In: *Neuroimage* 58.2 (2011), pp. 620–629.

[428] Stefan Debener et al. "Towards single-trial analysis in cognitive brain research." In: (2007).

[429] David M Groppe, Thomas P Urbach, and Marta Kutas. "Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review". In: *Psychophysiology* 48.12 (2011), pp. 1711–1725.

[430] Alexandre Gramfort et al. "MEG and EEG data analysis with MNE-Python". In: *Frontiers in neuroscience* 7 (2013), p. 267.

[431] Hernan Gonzalo Rey, Maryam Ahmadi, and Rodrigo Quian Quiroga. "Single trial analysis of field potentials in perception, learning and memory". In: *Current opinion in neurobiology* 31 (2015), pp. 148–155.

[432] Cyril R Pernet, Paul Sajda, and Guillaume A Rousselet. "Single-trial analyses: why bother?" In: *Frontiers in psychology* 2 (2011), p. 322.

[433] Stephen M Smith and Thomas E Nichols. "Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference". In: *Neuroimage* 44.1 (2009), pp. 83–98.

[434] Scott Makeig et al. "Mining event-related brain dynamics". In: *Trends in cognitive sciences* 8.5 (2004), pp. 204–210.

[435] T.W. Picton. *Human Auditory Evoked Potentials*. Plural Publishing, Incorporated, 2010. ISBN: 9781597566223.

[436] Clément Dondé et al. "A century of sensory processing dysfunction in schizophrenia". In: *European Psychiatry* 59 (2019), pp. 77–79. DOI: `10.1016/j.eurpsy.2019.04.006`.

[437] Carlos Trenado et al. "The potential of trial-by-trial variabilities of ongoing-EEG, evoked potentials, event related potentials and fMRI as diagnostic markers for neuropsychiatric disorders". In: *Frontiers in neuroscience* 12 (2019), p. 850.

[438] André Mouraux and Gian Domenico Iannetti. “Across-trial averaging of event-related EEG responses and beyond”. In: *Magnetic resonance imaging* 26.7 (2008), pp. 1041–1054.

[439] Emad-ul-Haq Qazi et al. “Single trial EEG patterns for the prediction of individual differences in fluid intelligence”. In: *Frontiers in human neuroscience* 10 (2017), p. 687.

[440] Diego Candia-Rivera and Gaetano Valenza. “Cluster permutation analysis for EEG series based on non-parametric Wilcoxon–Mann–Whitney statistical tests”. In: *SoftwareX* 19 (2022), p. 101170. ISSN: 2352-7110. DOI: https://doi.org/10.1016/j.softx.2022.101170.

[441] Eric Maris and Robert Oostenveld. “Nonparametric statistical testing of EEG-and MEG-data”. In: *Journal of neuroscience methods* 164.1 (2007), pp. 177–190.

[442] Marlene Meyer et al. “Enhancing reproducibility in developmental EEG research: BIDS, cluster-based permutation tests, and effect sizes”. In: *Developmental cognitive neuroscience* 52 (2021), p. 101036.

[443] Gan Huang and Zhiguo Zhang. “Improving sensitivity of cluster-based permutation test for EEG/MEG data”. In: *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE. 2017, pp. 9–12.

[444] Joseph Gonzalez-Heydrich et al. “Early auditory processing evoked potentials (N100) show a continuum of blunting from clinical high risk to psychosis in a pediatric sample”. In: *Schizophrenia research* 169.1-3 (2015), pp. 340–345.

[445] Anne M Shelley et al. “Mismatch negativity: an index of a preattentive processing deficit in schizophrenia”. In: *Biological psychiatry* 30.10 (1991), pp. 1059–1062.

[446] David Poeppel, William J Idsardi, and Virginie Van Wassenhove. “Speech perception at the interface of neurobiology and linguistics”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.1493 (2008), pp. 1071–1086.

[447] Sebastian Puschmann et al. “The right temporoparietal junction supports speech tracking during selective listening: Evidence from concurrent EEG-fMRI”. In: *Journal of Neuroscience* 37.47 (2017), pp. 11505–11516.

[448] Martha E Shenton et al. “Clinical correlations of auditory P200 topography and left temporocentral deficits in schizophrenia: a preliminary study”. In: *Journal of psychiatric research* 23.1 (1989), pp. 13–34.

[449] Ayoub Daliri and Ludo Max. “Modulation of auditory responses to speech vs. nonspeech stimuli during speech movement planning”. In: *Frontiers in Human Neuroscience* 10 (2016), p. 234.

[450] Björn Herrmann, Nadine Schlichting, and Jonas Obleser. “Dynamic range adaptation to spectral stimulus statistics in human auditory cortex”. In: *Journal of Neuroscience* 34.1 (2014), pp. 327–331.

[451] Yadira Roa Romero et al. “Reduced frontal theta oscillations indicate altered crossmodal prediction error processing in schizophrenia”. In: *Journal of Neurophysiology* 116.3 (2016), pp. 1396–1407.

[452] Daniel H Mathalon and Judith M Ford. *Divergent approaches converge on frontal lobe dysfunction in schizophrenia*. 2008.

[453] Yong-Wook Shin et al. “Gamma oscillation in schizophrenia”. In: *Psychiatry investigation* 8.4 (2011), p. 288.

[454] Daisuke Koshiyama et al. “Auditory gamma oscillations predict global symptomatic outcome in the early stages of psychosis: a longitudinal investigation”. In: *Clinical Neurophysiology* 129.11 (2018), pp. 2268–2275.

[455] Vykinta Parciauskaite, Jovana Bjekic, and Inga Griskova-Bulanova. “Gamma-range auditory steady-state responses and cognitive performance: a systematic review”. In: *Brain sciences* 11.2 (2021), p. 217.

[456] Michael J Minzenberg et al. “Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia”. In: *Archives of general psychiatry* 66.8 (2009), pp. 811–822.

[457] James F Cavanagh and Michael J Frank. “Frontal theta as a mechanism for cognitive control”. In: *Trends in cognitive sciences* 18.8 (2014), pp. 414–421.

[458] Andrea Perrottelli et al. “Unveiling the Associations between EEG Indices and Cognitive Deficits in Schizophrenia-Spectrum Disorders: A Systematic Review”. In: *Diagnostics* 12.9 (2022), p. 2193.

[459] Oren Griffiths et al. “Disrupted auditory N1, theta power and coherence suppression to willed speech in people with schizophrenia”. In: *NeuroImage: Clinical* 37 (2023), p. 103290. ISSN: 2213-1582. DOI: https://doi.org/10.1016/j.nicl.2022.103290.

[460] Rodolfo Solís-Vivanco et al. “Impairment of novelty-related theta oscillations and P3a in never medicated first-episode psychosis patients”. In: *npj Schizophrenia* 7.1 (2021), p. 15.

[461] Sayantanava Mitra et al. “Electroencephalogram alpha-to-theta ratio over left fronto-temporal region correlates with negative symptoms in schizophrenia”. In: *Asian Journal of Psychiatry* 26 (2017), pp. 70–76. ISSN: 1876-2018. DOI: https://doi.org/10.1016/j.ajp.2017.01.013.

[462] Fleur M Howells et al. “Electroencephalographic delta/alpha frequency activity differentiates psychotic disorders: a study of schizophrenia, bipolar disorder and methamphetamine-induced psychotic disorder”. In: *Translational psychiatry* 8.1 (2018), p. 75.

[463] Muzaffer Kaser et al. “Oscillatory underpinnings of mismatch negativity and their relationship with cognitive function in patients with schizophrenia”. In: *PloS one* 8.12 (2013), e83255.

[464] Shrey Grover et al. “Proximate markers of cognitive dysfunction in schizophrenia”. In: *Schizophrenia research* 233 (2021), p. 114.

[465] Brian F O’Donnell et al. “The auditory steady-state response (ASSR): a translational biomarker for schizophrenia”. In: *Supplements to Clinical neurophysiology* 62 (2013), pp. 101–112.

[466] Daisuke Koshiyama et al. “Neural network dynamics underlying gamma synchronization deficits in schizophrenia”. In: *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 107 (2021), p. 110224.

[467] Sarah K Keedy et al. “Impact of antipsychotic treatment on attention and motor learning systems in first-episode schizophrenia”. In: *Schizophrenia bulletin* 41.2 (2015), pp. 355–365.

[468] Risto Näätänen and Terence Picton. “The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure”. In: *Psychophysiology* 24.4 (1987), pp. 375–425.

[469] Katie L Nugent et al. “Distress intolerance and clinical functioning in persons with schizophrenia”. In: *Psychiatry research* 220.1-2 (2014), pp. 31–36.

[470] Juan L Molina et al. “Central auditory processing deficits in schizophrenia: Effects of auditory-based cognitive training”. In: *Schizophrenia research* 236 (2021), pp. 135–141.

[471] Remko Van Lutterveld, Iris E Sommer, and Judith M Ford. “The neurophysiology of auditory hallucinations–a historical and contemporary review”. In: *Frontiers in Psychiatry* 2 (2011), p. 28.

[472] Elham S Salama et al. “EEG-based emotion recognition using 3D convolutional neural networks”. In: *Int. J. Adv. Comput. Sci. Appl* 9.8 (2018), pp. 329–337.

[473] Hassan Ismail Fawaz et al. “Deep learning for time series classification: a review”. In: *Data Mining and Knowledge Discovery* 33.4 (2019), pp. 917–963.

[474] Pouya Bashivan et al. “Learning representations from EEG with deep recurrent-convolutional neural networks”. In: *arXiv preprint arXiv:1511.06448* (2015).

[475] Martyn Hyde. “The N1 response and its applications”. In: *Audiology and Neurotology* 2.5 (1997), pp. 281–307.

[476] Kevin M Spencer. “Averaging, detection, and classification of single-trial ERPs”. In: *Event-related potentials: A methods handbook* 209 (2005).

[477] Leo Breiman. “Random forests”. In: *Machine learning* 45.1 (2001), pp. 5–32.

[478] Yijun Bian and Huanhuan Chen. “When does diversity help generalization in classification ensembles?” In: *IEEE Transactions on Cybernetics* (2021).

[479] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[480] Stanislav Fort and Stanislaw Jastrzebski. “Large scale structure of neural network loss landscapes”. In: *arXiv preprint arXiv:1906.04724* (2019).

[481] Davide Chicco and Giuseppe Jurman. “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation”. In: *BMC genomics* 21.1 (2020), p. 6.

[482] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2017. URL: `https://www.R-project.org/`.

[483] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[484] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: `10.1038/s41586-020-2649-2`.

[485] Hung Le and Ali Borji. “What are the receptive, effective receptive, and projective fields of neurons in convolutional neural networks?” In: *arXiv preprint arXiv:1705.07049* (2017).

[486] Ramprasaath R Selvaraju et al. “Grad-cam: Visual explanations from deep networks via gradient-based localization”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626.

[487] Laura A Harrison et al. “The importance of sensory processing in mental health: a proposed addition to the Research Domain Criteria (RDoC) and suggestions for RDoC 2.0”. In: *Frontiers in psychology* 10 (2019), p. 103.

[488] Guy Lightfoot. “Summary of the N1-P2 cortical auditory evoked potential to estimate the auditory threshold in adults”. In: *Seminars in hearing*. Vol. 37. 1. Thieme Medical Publishers. 2016, p. 1.

[489] Prune Mazer et al. “Abnormal habituation of the auditory event-related potential P2 component in patients with Schizophrenia”. In: *Frontiers in psychiatry* 12 (2021), p. 297.

[490] Qiaoling Sun, Linlin Zhao, and Liwen Tan. “Abnormalities of Electroencephalography Microstates in Drug-Naïve, First-Episode Schizophrenia”. In: *Frontiers in Psychiatry* 13 (2022), p. 853602.

[491] Alexandre Gramfort et al. “MNE software for processing MEG and EEG data”. In: *Neuroimage* 86 (2014), pp. 446–460.