UNIVERSITY OF IOANNINA
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS
SCHOOL OF INFORMATICS AND TELECOMMUNICATIONS

PhD Dissertation

Towards Robust EEG-based Human-Computer Interaction Applications: Exploring Machine Learning Methods in Cognitive States and Neurological Disorders

Τίτλος στα Ελληνικά

«Ανάπτυξη Εύρωστων Εφαρμογών Αλληλεπίδρασης Ανθρώπου-Μηχανής βασισμένων σε Ηλεκτροεγκεφαλογραφικές Καταγραφές: Μελέτη Μεθόδων Μηχανικής Μάθησης σε Γνωσιακές Καταστάσεις και Νευρολογικές Διαταραχές»

Andreas Miltiadous

Ανδρέας Μιλτιάδους

ΑΡΤΑ 2024

UNIVERSITY OF IOANNINA
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS
SCHOOL OF INFORMATICS AND TELECOMMUNICATIONS

PhD Dissertation

Towards Robust EEG-based Human-Computer Interaction Applications: Exploring Machine Learning Methods in Cognitive States and Neurological Disorders

Τίτλος στα Ελληνικά

«Ανάπτυξη Εύρωστων Εφαρμογών Αλληλεπίδρασης Ανθρώπου-Μηχανής βασισμένων σε Ηλεκτροεγκεφαλογραφικές Καταγραφές: Μελέτη Μεθόδων Μηχανικής Μάθησης σε Γνωσιακές Καταστάσεις και Νευρολογικές Διαταραχές»

Andreas Miltiadous

Ανδρέας Μιλτιάδους

ΑΡΤΑ 2024

Ημερομηνία αίτησης του κου Μιλτιάδους Ανδρέα: 23-03-2021

Ημερομηνία ορισμού Τριμελούς Συμβουλευτικής Επιτροπής: 54η/21-04-2021

Μέλη Τριμελούς Συμβουλευτικής Επιτροπής:

Επιβλέπων:

Τζάλλας Αλέξανδρος, Αναπληρωτής Καθηγητής του Τμήματος Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Ιωαννίνων

Μέλη:

Γλαβάς Ευριπίδης, Καθηγητής του Τμήματος Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Ιωαννίνων

Γιαννακέας Νικόλαος, Αναπληρωτής Καθηγητής του Τμήματος Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Ιωαννίνων

Ημερομηνία ορισμού θέματος διδακτορικής διατριβής: 54η/21-04-2021

«Ανάπτυξη Εύρωστων Εφαρμογών Αλληλεπίδρασης Ανθρώπου - Μηχανής βασισμένων σε Ηλεκτροεγκεφαλογραφικές Καταγραφές: Μελέτη Μεθόδων Μηχανικής Μάθησης σε Γνωσιακές Καταστάσεις και Νευρολογικές Διαταραχές»

ΟΡΙΣΜΟΣ ΕΠΤΑΜΕΛΟΥΣ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ: 123η/17-04-2023

Τζάλλας Αλέξανδρος	Αναπληρωτής Καθηγητής του Τμήματος Πληροφορικής και Τηλεπικοινωνιών της Σχολής Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Ιωαννίνων, με γνωστικό αντικείμενο «Βιοϊατρική Τεχνολογία με Έμφαση στην Επεξεργασία και Ανάλυση Βιοϊατρικών Δεδομένων»
Γλαβάς Ευριπίδης	Καθηγητής του Τμήματος Πληροφορικής και Τηλεπικοινωνιών της Σχολής Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Ιωαννίνων, με γνωστικό αντικείμενο «Αρχιτεκτονική Υπολογιστών και Λειτουργικά Συστήματα»
Γιαννακέας Νικόλαος	Αναπληρωτής Καθηγητής του Τμήματος Πληροφορικής και Τηλεπικοινωνιών της Σχολής Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Ιωαννίνων, με γνωστικό αντικείμενο «Σήματα και Συστήματα»
Γκόγκος Χρήστος	Καθηγητής του Τμήματος Πληροφορικής και Τηλεπικοινωνιών της Σχολής Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Ιωαννίνων, με γνωστικό αντικείμενο «Συνδυαστικοί Αλγόριθμοι και Προγραμματισμός με έμφαση στο Χρονοπρογραμματισμό»
Τσούλος Ιωάννης	Καθηγητής του Τμήματος Πληροφορικής και Τηλεπικοινωνιών της Σχολής Πληροφορικής και Τηλεπικοινωνιών του Πανεπιστημίου Ιωαννίνων, με γνωστικό αντικείμενο «Υπολογιστικές Μέθοδοι»
Τσίπουρας Μάρκος	Αναπληρωτής Καθηγητής του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών της Πολυτεχνικής Σχολής του Πανεπιστημίου Δυτικής Μακεδονίας με γνωστικό αντικείμενο «Ψηφιακή Επεξεργασία Σήματος»

Αστρακάς Λουκάς Αναπληρωτής Καθηγητής της Ιατρικής Σχολής του Πανεπιστημίου Ιωαννίνων, με γνωστικό αντικείμενο «Ιατρική Φυσική»

Έγκριση Διδακτορικής Διατριβής με βαθμό «ΑΡΙΣΤΑ» στις 14-05-2024

Άρτα 16-05-2024

Ο ΠΡΟΕΔΡΟΣ ΤΟΥ ΤΜΗΜΑΤΟΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Ευριπίδης Γλαβάς
Καθηγητής

Η Γραμματέας του Τμήματος

Ευαγγελία Χρήστου

Στην Γωγώ και στην Κωνσταντίνα,

την Αγγέλα και τον Γιώργη

Executive Summary

Electroencephalography is a diagnostic tool that is often underutilized despite its significant potential. It is used for monitoring electrical activity in the brain, providing a high temporal resolution that makes it particularly suitable for real-time analysis. The inherent characteristics of EEG data, such as its temporal resolution, make it an ideal candidate for application in machine learning, enhancing its capabilities for automated analysis and diagnosis.

The aim of this PhD research is to apply modern signal processing techniques and machine learning algorithms to develop robust methodologies for the automated detection of neurological conditions and cognitive states. While the primary focus is on Alzheimer's research, this study also explores applications in Brain-Computer Interfaces and epilepsy. The research findings indicate that automated methodologies, when combined with EEG data, can significantly improve screening processes for dementia and other neurological conditions. These methodologies provide valuable insights that could pave the way for future clinical applications.

Modern deep learning methodologies, particularly transformers, have shown great promise in this domain. These advanced techniques can further enhance the capability of automated systems to analyze complex EEG data, leading to more accurate and reliable diagnostic tools. By leveraging these cutting-edge approaches, this research contributes to the ongoing efforts to integrate AI-driven solutions into clinical practice, ultimately improving patient outcomes and advancing the field of neurological diagnostics.

Επιτελική Σύνοψη

Το ηλεκτροεγκεφαλογράφημα είναι ένα διαγνωστικό εργαλείο που συχνά δεν αξιοποιείται επαρκώς παρά τις σημαντικές δυνατότητές του. Χρησιμοποιείται για την παρακολούθηση της ηλεκτρικής δραστηριότητας στον εγκέφαλο, παρέχοντας υψηλή χρονική ανάλυση που το καθιστά ιδιαίτερα κατάλληλο για ανάλυση σε πραγματικό χρόνο. Τα εγγενή χαρακτηριστικά των δεδομένων ηλεκτροεγκεφαλογραφήματος, όπως η χρονική του ανάλυση, τα καθιστούν ιδανικό υποψήφιο για εφαρμογή στη μηχανική μάθηση, ενισχύοντας τις δυνατότητές τους για αυτοματοποιημένη ανάλυση και διάγνωση.

Ο στόχος αυτής της διδακτορικής έρευνας είναι η εφαρμογή σύγχρονων τεχνικών επεξεργασίας σήματος και αλγορίθμων μηχανικής μάθησης για την ανάπτυξη αξιόπιστων μεθοδολογιών για την αυτοματοποιημένη ανίχνευση νευρολογικών καταστάσεων και γνωστικών καταστάσεων. Ενώ η κύρια εστίαση είναι στην έρευνα για τη νόσο Αλτσχάιμερ, η μελέτη αυτή διερευνά επίσης εφαρμογές στις διεπαφές εγκεφάλου-υπολογιστή και στην επιληψία. Τα ευρήματα της έρευνας δείχνουν ότι οι αυτοματοποιημένες μεθοδολογίες, όταν συνδυάζονται με δεδομένα ηλεκτροεγκεφαλογραφήματος, μπορούν να βελτιώσουν σημαντικά τις διαδικασίες ελέγχου για την άνοια και άλλες νευρολογικές καταστάσεις. Αυτές οι μεθοδολογίες παρέχουν πολύτιμες γνώσεις που θα μπορούσαν να ανοίξουν το δρόμο για μελλοντικές κλινικές εφαρμογές.

Οι σύγχρονες μεθοδολογίες βαθιάς μάθησης, ιδίως τα δίκτυα transformers, φαίνεται να είναι πολλά υποσχόμενες σε αυτόν τον τομέα. Αυτές οι προηγμένες τεχνικές μπορούν να βελτιώσουν περαιτέρω την ικανότητα των αυτοματοποιημένων συστημάτων να αναλύουν σύνθετα δεδομένα ηλεκτροεγκεφαλογραφήματος, οδηγώντας σε πιο ακριβή και αξιόπιστα διαγνωστικά εργαλεία. Αξιοποιώντας αυτές τις προσεγγίσεις αιχμής, η παρούσα έρευνα συμβάλλει στις συνεχιζόμενες προσπάθειες για την ενσωμάτωση λύσεων που βασίζονται στην Τεχνητή Νοημοσύνη στην κλινική πρακτική, βελτιώνοντας τελικά τα αποτελέσματα των ασθενών και προωθώντας τον τομέα της νευρολογικής διάγνωσης.

Table of Contents

Executive Summary 1
Επιτελική Σύνοψη 3
Table of Contents 5
List of Tables 11
List of Figures 13
1.1. Introduction 18
1.2. Research Objectives 19
1.3. Dissertation Structure 21
2.1. Human Brain Anatomy 24
2.2. Electroencephalography 26
2.2.1. Basic functionality of neurons 26
2.2.2. Traditional usage of EEG 27
2.2.3. Automated Methods in EEG Analysis 28
2.2.3.1. Historical overview 28
2.2.3.2. Overview 29
2.2.4. Frequency bands of Interest 30
2.3. Conditions investigated through EEG 31
2.3.1. Epilepsy 32
2.3.2. Alzheimer's Disease and other Dementia 34
2.3.3. Other Conditions 36
2.4. Brain-Computer Interfaces 36
2.4.1. EEG-based BCI 36
2.5. Clinical devices and Wearable EEG 37
3.1. Preprocessing 40
3.1.1. Basic preprocessing techniques in EEG Analysis 40
3.1.1.1. Filtering 40
3.1.1.2. Rereferencing 42
3.1.1.3. Electrode Interpolation 43
3.1.1.4. Epoching 43

3.1.2. Artifact removal techniques 44
3.2. Frequency Domain Analysis 45
3.2.1. Fourier Transform 46
3.2.1.1. Fast Fourier Transform 47
3.2.2. Power Spectral Density 47
3.2.2.1. Welch Method 48
3.2.2.2. Burg's Method 48
3.2.2.3. Multitaper Method 49
3.2.3. Limitations of Frequency Domain analysis 50
3.3. Time-Frequency Domain Analysis 50
3.3.1. Short-Time Fourier Transform 51
3.3.2. Wavelet Transform 51
3.3.2.1. Continuous Wavelet Transform 53
3.3.2.2. Discrete Wavelet Transform 53
3.3.3. Hilbert-Huang Transform 54
3.3.3.1. Empirical Mode Decomposition 54
3.3.4. Wigner-Ville Distribution 56
3.3.5. S-Transform 56
3.3.6. Limitations of Time-Frequency Analysis 57
3.4. Characteristics extracted from EEG 58
3.4.1. Statistical Features 58
3.4.2. Frequency Features 59
3.4.3. Complexity Features 60
3.4.3.1. Entropy-based measures 60
3.4.3.2. Fractal-Dimension 61
3.4.4. Aperiodic components – 1/f slope 62
3.5. Synchronization Features 63
4.1. Introduction and Importance 67
4.1.1. Problems that are addressed 67
4.2. Machine Learning Pipeline 67
4.2.1. Preprocessing & Feature Extraction 69
4.2.2. Feature Reduction 69

4.2.2.1. Principal Component Analysis 70
4.2.2.2. Feature Selection 71
4.2.3. Training, Testing & Validation methodologies 72
4.2.4. Evaluation Metrics 73
4.3. Machine Learning 75
4.3.1. Traditional Classifiers 75
4.3.1.1. Decision Trees 75
4.3.1.2. Naïve Bayes 76
4.3.1.3. Logistic Regression 76
4.3.1.4. Support Vector Machines (SVM) 77
4.3.1.5. k-Nearest Neighbors 78
4.3.1.6. Linear Discriminant Analysis (LDA) 78
4.3.2. Ensemble Classifiers 79
4.3.2.1. Random Forests 79
4.3.2.2. Extra Trees 79
4.3.2.3. Gradient Boosting 80
4.3.3. Deep Learning 80
4.3.3.1. Feed Forward Neural Networks (FFNs) 81
4.3.3.2. Training Neural Networks: Backpropagation 82
4.3.3.3. Convolutional Neural Networks (CNNs) 83
4.3.3.4. Recurrent Neural Networks (RNNs) 85
4.3.3.5. Long Short-Term Memory Networks (LSTMs) 86
4.3.3.6. Transformers 87
4.3.4. Hyperparameter optimization 88
4.4. Clinical Applications 89
4.4.1. Challenges, Limitations and Problems 89
5.1. Alzheimer's Disease and Dementia Research 93
5.1.1. Alzheimer's Disease and Frontotemporal Dementia: A Robust Classification Method of EEG Signals and a Comparison of Validation Methods 95
5.1.2. Enhanced Alzheimer's disease and Frontotemporal Dementia EEG Detection: Combining lightGBM Gradient Boosting with Complexity Features 101
5.1.2.1. Preprocessing 102
5.1.2.2. Feature Extraction 103

5.1.2.3. Classification 103
5.1.2.4. Results 106
5.1.3. A dataset of scalp EEG recordings of Alzheimer's disease, Frontotemporal Dementia and Healthy subjects from routine EEG 107
5.1.3.1. Dataset description 109
5.1.3.2. Dataset structure 111
5.1.3.3. Preprocessing 113
5.1.3.4. Benchmark experiments 113
5.1.4. DICE-net: A Novel Convolution-Transformer Architecture for Alzheimer Detection in EEG Signals 116
5.1.4.1. Methodology 118
5.1.4.2. Results 127
5.1.4.3. Discussion 132
5.2. Epilepsy Detection with EEG based Machine Learning Research 136
5.2.1. Evaluating the Window Size's Role in Automatic EEG Epilepsy Detection . 138
5.2.2. Methodology 138
5.2.2.1. The BFGS method 139
5.2.2.2. The multistart method 140
5.2.3. Results 141
5.2.4. Machine Learning Algorithms for Epilepsy Detection based on published EEG databases: A Systematic Review 144
5.2.4.1. Methodology 144
5.2.4.2. Results 147
5.2.4.3. Discussion and Statistics 157
5.2.4.4. Related work comparison 163
5.3. Brain-Computer Interfaces for analyzing cognitive states 165
5.3.1. Assessing Electroencephalography as a Stress Indicator: A VR High-Altitude Scenario monitored through EEG and ECG 165
5.3.1.1. Methodology 168
5.3.1.2. Results 172
5.3.1.3. Discussion 176
5.3.2. Classification of EEG signals from Young Dyslexic Adults combining a Brain Computer Interface device and an Interactive Linguistic Software Tool 179
5.3.3. Methodology 180

5.3.3.2. Results 183
5.3.3.3. Discussion 185
5.3.4. An ensemble method for EEG-based texture discrimination during open eyes active touch 187
5.3.4.1. Introduction 187
5.3.4.2. Methodology 188
5.3.4.3. Results 191
5.3.4.4. Discussion 194
Regarding Future Insights 197
References 198
Abbreviations list 214

List of Tables

Table 1. Statistical characteristics (values in brackets represent standard deviation) of the participants. 95
Table 2. Accuracy, Sensitivity and Specificity results of classification algorithms using the 10-fold validation method. 97
Table 3. Accuracy results with the LOSO validation 97
Table 4. Sensitivity and Specificity of Decision Trees and Random Forest with LOSO validation 97
Table 5. Related research of AD/FTD/CN classification studies, as found at 2021. 99
Table 6. Statistical characteristics (values in brackets represent standard deviation) of the participants 102
Table 7. Classification Results of the lightGBM algorithm 106
Table 8 Participants Description. A indicates AD patient, F indicates FTD patient and C indicates a healthy subject. F indicates female and M indicates male. 110
Table 9 Classification performance metrics of LOSO validation for AD-CN problem. 115
Table 10 Classification performance metrics of LOSO validation for FTD-CN problem. 115
Table 11 The architecture and hyperparameters of DICE-net model. 123
Table 12 Computational complexity of DICE-net and of its ablation experiments ... 126
Table 13 Performance of the DICE-net methodology and the ablation studies for the AD-CN problem. 128
Table 14 Performance of the DICE-net methodology and the comparison algorithms for the AD-CN problem. 128
Table 15 Performance of the DICE-net methodology and the comparison algorithms for the FTD-CN problem. 129
Table 16 Performance metrics of the state-of-the-art EEG based deep learning methodologies on the AD/CN and FTD/CN problems. 131
Table 17 Comparison of related works, as performed in DICE-net publication. 134
Table 18 Performance results in terms of accuracy for the 4 comparison algorithms, in epoch length 1-24 second 141
Table 19 Performance results in terms of Area under RO, area under PRC and k-statistic for the 4 comparison algorithms, in epoch length 1-24 second 142
Table 20 Studies that used Bonn database published in 2017-2019 150
Table 21 Studies that used Bonn database, published in 2020-2022 152
Table 23 Studies that used CHB-MIT database, published in 2020-2022 154
Table 22 Studies that used CHB-MIT database, published in 2017-2019 154
Table 24 Studies that used other databases 155
Table 25 Studies that used multiple databases 156

Table 26 Each cell represents the percentage of studies that used axis x & axis y combination of databases, from the total of multiple database studies. 162
Table 27 Related review papers comparison, regarding EEG-based epilepsy detection. 164
Table 28 Paired T-tests regarding the alteration of each measure between the states calm and stressed. The * symbol indicates statistical significance with p value < 0.05 173
Table 29 Paired T-test in asymmetry scores for groups. The symbol * indicates statistical significance with p value <0.05 176
Table 30 Spearman Correlation of the alteration of each power band with the alteration of BPM 176
Table 31 Classification results for each Region of Interest for the 1st condition 184
Table 32 Classification results for each Region of Interest for the 2nd condition 185
Table 33 Classification results for each Region of Interest for the 3rd condition 185
Table 34 Performance results for the N-R-S-W problem 191
Table 35 Performance results for the N-S-W problem 191
Table 36 Performance results for the N-R-W problem 192
Table 37 Performance results for the N-R-S problem 192
Table 38 Correlation of errors between the weak classifiers and the voting system 193

List of Figures

Figure 1 The three main parts of the brain 24
Figure 2 The location of the 4 brain lobes. 25
Figure 3. The 5 frequency bands. 32
Figure 4 a) The brain of a heathly subject, b) the brain of an AD patient. Neuritic plaques and neurofibrillary tangles accumulation as a result of amyloid-beta peptide's (Aβ) accumulation in the most affected area of the brain, the medial temporal lobe and neocortical structures. 35
Figure 5 The 10-20 international system electrode topography. Blue: Frontal lobe, Yellow: Parietal lobe, Green: Temporal lobe, Red: Occipital lobe. 42
Figure 6 Left: EEG signal before ICA processing. Right: EEG after removing blink artifacts. Middle: Different types of artifacts expressed as signals (top) and as heatmaps (bottom). Adapted from "Extended ICA Removes Artifacts from Electroencephalographic Recordings", by Jung et al., 1998. 44
Figure 7 Signal in red is contaminated with artifacts before ASR. Signal in blue is cleared after applying ASR..... 45
Figure 8 Top: A signal in the time domain which is a convolution of 3 sinusoidal signals. Botttom: The Fourier Transform of the signal. Adapted from "Key Concepts: Fourier Transforms and Signal Processing", by Tom White, 2020 (blog article)..... 47
Figure 9 A selection of different wavelet functions. Adapted from "Efficient Deep Neural Networks for Classification of Alzheimer's Disease and Mild Cognitive Impairment from Scalp EEG Recordings", by Fouladi, 2022 52
Figure 10 Time-Frequency distribution in a spectrogram of the DTF (a), the STFT (b) and the DWT (c) 54
Figure 11 EMD of an EEG signal of the FC3 electrode location..... 56
Figure 12 A general pipeline of an EEG based machine learning classification study 68
Figure 13 A confusion matrix for a 2-class problem. 73
Figure 14 A typical Feed Forward Neural Network architecture 81
Figure 15 A Convolution Neural Network architecture. 85
Figure 16 The Long Short-Term Memory Network Architecture..... 87
Figure 17 OpenViBE xml diagram of feature extraction 96
Figure 18 Flowchart of the methodology as described in the study. 101
Figure 19 Different EEG ICA components, based on their origin. Adapted from "EEG artifact elimination by extraction of ICA-component features using image processing algorithms", by T. Radüntz et al., 2015..... 102
Figure 20 Feature Importance as calculated by the lightGBM algorithm 105
Figure 21. ROC curves of the three classification problems for the lightGBM implementation 106
Figure 22 The cover of the issue in which the data descriptor paper was published. It features our study as the highlight study of the issue. 108

Figure 23 The dataset structured in the BIDS format, as uploaded in OpenNeuro 112
Figure 24 A snapshot of an unprocessed signal, and the same signal after preprocessing 113
Figure 25 Scalp heatmaps of the Power Spectral Density for the 5 rhythms, averaged across the 3 different groups. 114
Figure 26 DICE-net Architecture 120
Figure 27 Different ablation configurations that were employed during DICE-net testing. 126
Figure 28 a) (left) PSD of a severe AD case (bottom), a mild AD case (middle) and a healthy subject (top). b) (right) scalp heatmaps of PSD across 5 frequency bands, averaged across groups AD, CN, FTD. 127
Figure 29 a) (left) Spectral Coherence Correlation (SCC) heatmaps for frequency band of each group (AD, CN, FTD). Each cell (X,Y) represents the spectral correlation of electrode X with electrode Y, averaged across each group. b) (right) SCC averaged across electrodes 130
Figure 30 ROC curves of DICE-net and of the rest of the comparison algorithms for the AD-CN and FTD-CN problems 130
Figure 31 Violin plots representing the distribution of the accuracies for each individual prediction. The width of the plot indicates the density of the scores at each value, while individual dots represent a single classification performance for one subject. 131
Figure 32 Normalized Absolute Magnitude of Convolution layer weights. (top): AD-CN. (bottom): FTD-CN classification. 132
Figure 33 The flowchart of the proposed methodology in “Evaluating the Window Size’s Role in Automatic EEG Epilepsy Detection” 141
Figure 34 Flowchart of the study selection and elimination process for the systematic review. 147
Figure 35 The basic structure of a methodology that utilizes Machine Learning for the EEG epilepsy detection 157
Figure 36 Separation of the experimental studies based on the database used. 157
Figure 38 Machine Learning methodologies that are applied in the studies of the review. 159
Figure 37 Signal processing techniques and Transformations that are applied in the studies of the review. 159
Figure 39 Comparison of Machine Learning methodologies used in the 2017-2019 and the 2020-2022 period. 160
Figure 40 Average accuracy of the different classifiers reported in the studies for each database. Studies on the Bonn database are divided into 3 categories namely Healthy-Ictal-Interictal, Seizure Detection and Healthy-Interictal. In this figure CNNs are considered a different category from Neural Networks due to their increased importance. 161

Figure 41 Popularity of each database on the reviewed studies. The barchart on the right is concerning the Multiple Databases section, in which each bar represents the percentage of the studies that examine a certain database. 162
Figure 42 Chord diagram modeling the connection of databases co-existing in studies that employ multiple databases. 163
Figure 43 Flowchart of the experiment 168
Figure 44 Experimental design of the EEG+VR stress protocol 169
Figure 45 BPM of a participant throughout the experiment. 172
Figure 46 Calm and Stress state comparison of BPM and Alpha, Beta and Gamma Bands. 173
Figure 47 Comparison of brain heatmaps between subjects from the two groups, focusing on the Theta, Alpha, Beta, and Gamma energy bands. 175
Figure 48 Alterations in Occipital activity, OAA and FAA between the 2 groups (calm group, stressed group) 175
Figure 49 Accuracy scores of each classifier for the Dyslexia-Control problem. 183
Figure 50 Accuracy, Sensitivity, Specificity scores for every RoI in every condition 183
Figure 51 ROC curves in different brain regions 184
Figure 52 Comparison of the Accuracy scores for the different RoIs 184
Figure 53 Flowchart of the methodology presented in the active touch classification study 188
Figure 54 The materials used and the experimental position 189
Figure 55 The PCA decomposition in the feature vector 190
Figure 56 Overview of performance results for the Active Touch experiment. 193

Chapter 1

Introduction

"9 out of 10 researchers recommend pursuing a PhD. The 10th lives happily."

-Unknown

1.1. Introduction

The focus of this doctoral research lies on the signal processing and machine learning aspects of the Electroencephalographic (EEG) signal and its applications on automatic detection, prediction or severity assessment of different neurological malfunctions or cognitive conditions. It was conducted in the Human-Computer Interaction Laboratory (HCI-lab) of the Department of Informatics and Telecommunications, University of Ioannina, under the constant supervision and assistance of my supervisor, Dr. Alexandros Tzallas, who is currently holding the position of Associate Professor in this Department, and was funded through the Research Programme "Immersive Virtual, Augmented and Mixed Reality Center of Epirus" (MIS 5047221) which is implemented under the Action "Reinforcement of the Research and Innovation Infrastructure", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

I am profoundly grateful for the support, guidance, and encouragement I have received throughout the journey of this doctoral thesis. First and foremost, I extend my deepest appreciation to my supervisor, Alexandros, who believed in me, supported me both intellectually and financially, and whose expertise, mentorship and insightful feedback were instrumental for shaping this work.

Furthermore, a special word of appreciation goes to Dr. Katerina Tzimourta, whose contributions to my research have been indispensable. A good part of my research was a continuation of her research and collaborations, thus without her, my work would be significantly poorer.

Moreover, I owe a dept of gratitude to the members of my advisory committee, Dr. Nikolaos Giannakeas and Dr. Euripidis Glavas for their support and valuable assistance and guidance throughout this journey.

To be fair, a short introduction is not enough to thank everyone that helped me achieve this result half as well as they deserve, as Bilbo said. This work is the least a personal achievement and more a teamwork performance. The collaboration with the 2nd Neurology Department of AHEPA Hospital in Thessaloniki, and especially the contributions of neurologist Dr. Theodora Afrantou, made the Alzheimer's research and the publication of our open-access dataset possible. My dear friend and fellow student, Emmanouil Giannidis, greatly assisted me with his Deep Learning experience, and together we created our signature work, the DICE-net architecture. Moreover, I worked together with my friend Vasileios Aspiotis, who was also pursuing his PhD, and published several articles together; thus, he was a great part of this journey and requires an honorable mention. In one way or another, I

got assistance from every member of my lab; and I sincerely thank them all. Last but not least, Big thanks to my family, my parents and my sister(s) the unsung heroes of this PhD saga, who taught me that home is behind me, the world is ahead (even though I live in our home, and they left to leave abroad), and who gave me the light of Eärendil, our most beloved star to be a light for me in dark places, when all other lights go out.

1.2. Research Objectives

The advancement of digital signal processing techniques has revolutionized various scientific domains, opening new vistas for research and applications. One such impactful area is EEG, a non-invasive technique for monitoring electrical activity in the brain. In medical settings, EEG has been indispensable for diagnosing and monitoring neurological disorders like epilepsy or other seizure disorders. EEG might also prove useful for diagnosing or treating other medical conditions such as brain tumors, brain damage from head injury, brain disfunction due to a variety of causes (encephalopathy), sleep disorders, inflammation of the brain and the Creutzfeldt-Jakob disease. EEG can also be used as a confirmation tool for brain death in someone in a persistent coma. Furthermore, although not the main diagnostic tool, it can be used complementary for neurodegenerative diseases such as Alzheimer's disease (AD) or other dementia related conditions.

Beyond healthcare, EEG has also found the way into commercial sectors, serving as the foundation for Brain-Computer Interface (BCI) technologies that facilitate direct communication between the human brain and external devices. These BCIs have applications in a wide range of industries, from entertainment and gaming to assistive technologies for individuals with mobility impairments. EEG-based controllers are being developed for virtual and augmented reality experiences, providing a more immersive and intuitive way of interacting with digital worlds. In the automotive sector, EEG is being explored for monitoring driver alertness to prevent accidents caused by drowsiness or inattention. Additionally, wearables and smart home technologies are integrating EEG sensors to offer features like meditation guidance, sleep quality tracking, and even smart lighting adjustments based on the user's mood or cognitive state. Moreover, commercial EEG devices are becoming increasingly accessible and user-friendly, allowing for consumer-level adoption. This democratization of EEG technology is not only expanding its application but also fueling public interest and investment in brain-related research and development.

Despite its widespread utility, traditional EEG analysis often involves manual interpretation by skilled clinicians, making it a time-consuming and expensive process. Manual analysis is not only resource-intensive but also subject to human error, which poses challenges not only in clinical but also commercial applications. This presents a pressing need for automated methods that can offer both speed and accuracy in EEG analysis. Such methodologies should assist in timely screening of huge amounts of recordings, thus allowing the medical practitioners to focus not on analyzing the EEG waveforms but interpreting its unique characteristics.

Over the past decade, the adoption of advanced machine learning (ML) methodologies for EEG analysis has transitioned from a speculative concept to a tangible reality. This transformation has been fueled by remarkable strides in computational capabilities and groundbreaking advancements in deep learning and signal processing techniques. ML algorithms, particularly deep learning models, have shown exceptional promise in automating complex tasks such as the classification and interpretation of EEG signals. These automated systems are capable of rapidly analyzing vast datasets, thereby significantly reducing the time and resources required for manual interpretation. Furthermore, ML models can be fine-tuned to achieve high levels of accuracy, making them invaluable tools in both clinical diagnosis and real-world applications. As computational power continues to grow and ML g algorithms become increasingly sophisticated, the automatic interpretation of EEG signals is poised to revolutionize the fields of neuroscience, healthcare, and human-computer interaction.

Furthermore, the introduction of Transformer neural networks in 2017 and their advancements since then have opened new avenues for the automatic analysis of EEG signals. Originally designed for natural language processing tasks, Transformers have proven to be highly adaptable and effective in handling sequential data, a feature highly relevant to EEG time series. Their ability to capture long-range dependencies in the data and parallelize computations makes them well-suited for EEG analysis, which often involves interpreting complex temporal patterns over extended periods. Transformers offer a significant advantage in tasks like anomaly detection, signal classification, and even real-time monitoring of cognitive states. Their self-attention mechanisms can focus on salient portions of the EEG signal, enabling more accurate identification of neurological events or states. As the architecture continues to be refined and specialized variants are developed, Transformer neural networks have the potential to significantly enhance the accuracy and efficiency of automated EEG methodologies, thereby contributing to advances in both clinical and commercial applications.

This dissertation aims to explore the intersection of ML and EEG to create robust Human-Computer Interaction (HCI) applications focused on cognitive states and neurological disorders. The research objectives of this doctoral research are divided into 3 distinct categories, that have been researched in parallel. These categories have been:

1) **Alzheimer's disease or other dementia detection using EEG.** The main body of the research that has taken place in the duration of the PhD has been dedicated to creating new methodologies for dementia detection. To achieve this, a collaboration with the 2nd Neurological Department of AHEPA University Hospital of Thessaloniki has been established, where the EEG recordings as well as the medical expertise has been obtained from.
2) **Automatic Epilepsy detection using EEG.** Although EEG-based epilepsy detection has been extensively studied over the past decade and may be considered a well-established field, there remains room for improvement and consolidation of existing knowledge. A comprehensive analysis of existing EEG databases and

methodological literature has been conducted, accompanied by research to determine the optimal window size for such studies.

3) **Brain-Computer Interfaces for analyzing cognitive states.** EEG has been integrated with Virtual Reality to create immersive experiences designed to induce stress, and correlations have been observed between stress induction and changes in EEG patterns. Additionally, specialized linguistic software tools have been developed and utilized in conjunction with EEG recordings to study alterations in brain activity among patients with dyslexia. This multidisciplinary approach has facilitated the development of ML algorithms that can effectively capture these EEG alterations. Finally, a BCI application for gamification of brain activity has been developed and presented at a panhellenic neurological meeting. For these to be possible, a collaboration with XRCenter (Virtual, Augmented and Mixed Reality Center) of the Epirus Region has been established and funding, equipment as well as Virtual Reality expertise has been obtained from.

1.3. Dissertation Structure

This dissertation presents the theoretical background for the EEG analysis and ML as well as the whole length of the methodologies that have been proposed in regard to the research axes that have been analyzed in chapter 1.1. The dissertation is separated in the general section and original research section.

Regarding the general section, chapter 2 is the background and context description. There, the EEG is meticulously analyzed, the need for ML is explained and domains in which it is employed (Alzheimer's disease, Epilepsy, BCI) are analyzed. Furthermore, an analysis on clinical and wearable devices is being established. Chapter 3 analyzes all the theoretical background on EEG signal analysis, covering preprocessing and artifact rejection, frequency and time-frequency transformations of the signal, and the typical characteristics extracted from EEG. It also discusses characteristics derived from the synchronization of different brain regions or electrodes. Finally, the chapter reviews software developed by other research teams that makes EEG analysis available. Chapter 4 outlines the theoretical background of ML as applied to EEG analysis. It defines the types of problems that can be addressed and presents an in-depth look at a typical ML pipeline. Particular focus is placed on the analysis of Transformer neural networks, which represent the state of the art at the time this dissertation was written.

Regarding the original research section, chapter 5 provides detailed description of all the original research that has been published throughout the course of the doctoral research. It is divided into 4 subchapters, each describing a separate area of research. 5.1. presents the main part of the research which is dedicated to dementia and specifically AD research. It consists of methodologies for automatic EEG classification as well as the publication of an open access dementia database acquired from AHEPA hospital, which is the first open access EEG database of Alzheimer's recordings. Next, 5.2. is about the research on epilepsy, and 5.3 consists of the research about HCI applications on EEG (and conditions like

dyslexia). Chapter 6 discusses the theoretical contribution and the practical applications of the research that has taken place as well as the limitations and the future work. Chapter 7 concludes the dissertation.

Chapter 2

Background and Context

"If the brain were so simple we could understand it, we would be so simple we couldn't.

-Lyall Watson

2.1. Human Brain Anatomy

The human brain, being the control center of the nervous system has a fundamental role to our consciousness, emotions, thoughts, and memory. It is comprised of billions of neurons and is responsible for a lot of bodily functions and cognitive processes. It is divided into several regions and each of them has a specific role, ranging from the regulation of basic life-sustaining functions like breathing and heartbeat to higher cognitive functions such as reasoning, language, and abstract thinking. The core of the human-world interaction is the ability of the brain to process sensory information, integrate it and respond to it accordingly. Moreover, its ability to adapt and rewire itself highlights its dynamic nature.

The three major parts of the brain are the Cerebrum, the Cerebellum, and the Brainstem

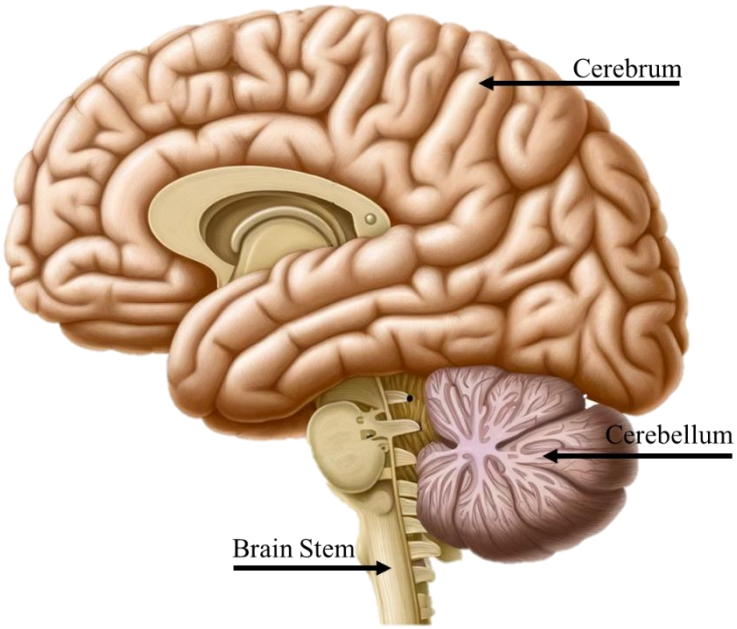


Figure 1 The three main parts of the brain

[1]. The Cerebrum is the largest part of the brain and is divided into two hemispheres (left and right). Each hemisphere is further divided into four lobes: frontal, parietal, temporal, and occipital. The Cerebrum is responsible for higher-order brain functions, including thought, emotion, sensory processing, speaking-language, and voluntary muscle movement. The Cerebellum is much smaller in size and is located under the cerebrum at the back of the brain. Its primary function is to coordinate and regulate muscular activity, balance, and posture. It does not initiate movement but contributes to coordination, precision, and accurate timing of movements. Last but not least, the Brainstem connects the brain with the

spinal cord and consists of the midbrain, pons, and medulla oblongata. It is responsible for controlling many vital life functions, such as breathing, heart rate, blood pressure, and consciousness. The brainstem also plays a role in relaying information between the brain and the rest of the body and in controlling basic attention, arousal, and some autonomic functions.

The EEG primarily focuses on the Cerebrum. As mentioned before, it is divided into 2 hemispheres and 4 lobes. A brief description of each lobe's functionality is described below.

Frontal Lobe: Located at the front of the brain, it is associated with reasoning, motor skills, higher-level cognition, and expressive language. The prefrontal cortex, part of the frontal lobe, is crucial for decision making, social behavior, and personality expression.

Parietal Lobe: Positioned in the middle section of the brain, it is involved in processing tactile sensory information such as pressure, touch, and pain. The parietal lobe integrates sensory information to form a single perception (cognition).

Temporal Lobe: It is located at the bottom section of the brain and involved in the processing of auditory stimuli, essential for the formation of explicit long-term memory. The temporal lobe also has a role in the interpretation of language and emotions.

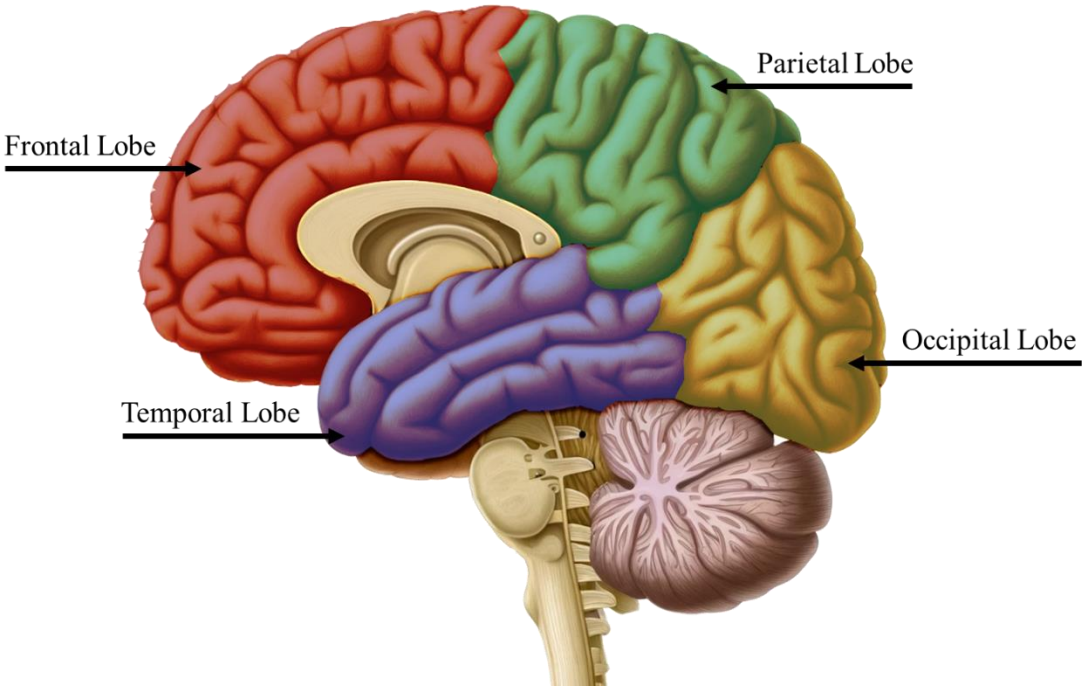


Figure 2 The location of the 4 brain lobes**.**

Occipital Lobe: It is located at the back of the brain, and is primarily responsible for visual processing. It interprets the visual information, including color, light, and movement and plays a role in visual perception such as recognizing and understanding what is seen.

2.2. Electroencephalography

EEG is a non-invasive technique used to record electrical activity generated by the brain [2]. With applications ranging from medical diagnosis to neuroscience research, EEG is very important in understanding brain function and behavior. It captures the electrical activity of the brain through electrodes placed on the scalp. At a cellular level, neurons, the primary cells constituting the nervous system, produce electrical impulses as they communicate with each other. These impulses result from the flow of ions across the cell membrane, creating voltage fluctuations known as action potentials. When a large number of neurons are activated simultaneously, their electrical activity can be captured as an EEG signal.

2.2.1. Basic functionality of neurons

Neurons and neuroglial cells serve as the foundational elements of the nervous system. Not only do they execute essential anatomical, electrophysiological, and molecular functions, but neuroglial cells also assist neurons and contribute to the development of the myelin sheath. These components are crucial for a comprehensive understanding of human brain function.

The neuron serves as a specialized cell type and constitutes the fundamental cellular unit of the nervous system [3]. All neurological activities are dependent on interactions among neurons. These neurons can be categorized according to various criteria such as their size, shape, neurochemical properties, spatial location, and patterns of connectivity, all of which are crucial for understanding their functional roles within the brain. Importantly, neurons work in conjunction with supportive neuroglial cells to establish neural circuits. These circuits act as the primary building blocks of neural systems, responsible for processing specific kinds of information and forming the structural basis for brain functions.

Neural systems encompass neurons and circuits strategically located throughout the brain, and they serve one of three general roles. Sensory systems, made up of sensory neurons, relay information concerning both the internal state of the organism and its external environment. Motor systems, featuring motor neurons, organize and transmit information to the body's external parts. Finally, intermediate systems bridge the sensory and motor aspects of the nervous system and lay the groundwork for higher-order functions, such as perception, attention, cognition, emotions, logical reasoning, and other complex activities of the human brain.

When a neuron receives a signal from its neighboring neurons, it undergoes a process called depolarization, by which positively charged ions flow into the cell. This shift in ion concentration generates an electrical impulse known as an action potential, which travels

along the neuron's axon to its terminals. Here, it triggers the release of neurotransmitters, which in turn stimulate adjacent neurons, perpetuating the electrical signal [4].

2.2.2. Traditional usage of EEG

The EEG was first developed as a clinical and research tool in the early 20th century. Hans Berger, a German psychiatrist, recorded the first human EEG in 1924 [5]. However, his findings were not immediately accepted by the scientific community. It was not until his work was validated by British scientists, including Edgar Adrian and B.H.C. Matthews, in the late 1920s and early 1930s that the EEG began to gain credibility. Berger's original observations included what he called "alpha" and "beta" waves, which represented the rhythmic activity he observed in the brain. The alpha rhythm, in particular, was significant because it was the first time that regular, ongoing electrical activity had been detected from the human brain without external stimulation. The first clinical applications of EEG were primarily in epilepsy diagnosis and research [6]. Through the years, EEG has been used for a wide range of clinical purposes, including the diagnosis of other neurological disorders, sleep disorders and research in cognitive neuroscience. By the 1950s and 1960s, EEG had become a routine part of neurological evaluation, and technological advances over the subsequent decades have only expanded its utility. Today, EEG is an essential tool in clinical neurology, as well as in various research settings.

The combined activity of millions of neurons generates electrical fields that can be sensed outside the skull. An EEG records electrical fields by using strategically positioned electrodes on the head. Each electrode measures electrical activity in relation to a reference, which can be another electrode or an averaged output from numerous electrodes. The voltage variations are magnified and shown as waveforms, offering a real-time view of brain activity. EEG signals are typically faint, frequently measuring in the microvolts range, and can be affected by noise from internal and external sources. Internal noise may originate from physiological processes such as muscle contractions or eye movements, whereas external noise may result from electrical interference from neighboring devices. Sophisticated filtering and signal processing techniques are often used to separate the genuine brain impulses from the noise.

EEG has a high temporal resolution, detecting electrical activity variations in milliseconds. This makes it an excellent tool for studying rapid cognitive processes like perception, decision-making, and motor reactions. The spatial resolution is restricted, which complicates the precise identification of the specific source of neuronal activity in the intricate structures of the brain. Progress in computational approaches and source localization techniques is slowly reducing this constraint, providing more comprehensive spatial information.

EEG is mostly utilized as a clinical instrument for diagnosing conditions such as epilepsy, sleep problems, encephalopathies, and neonatal monitoring. An EEG recording typically includes waves of electrical activity, with one waveform corresponding to each

electrode. EEG is utilized in clinical and research environments to identify different types of observations that can provide understanding of brain function and possible irregularities. Key discoveries that are commonly sought include:

1) **Frequency domain distribution of the signals:** Different frequency activities illustrate different functionalities on the brain. The EEG signal is typically separated in 5 frequency bands of interest and is analyzed as such.
2) **Epileptiform Abnormalities:** Spikes, sharp waves, or spike-and-wave discharges can indicate seizure activity or an increased risk for seizures.
3) **Focal Slowing:** Slowed wave activity in a specific area may suggest a localized brain lesion, such as a tumor or stroke.
4) **Generalized Slowing:** Often indicative of diffuse brain dysfunction and is commonly seen in encephalopathies or severe systemic illnesses affecting brain function.
5) **Asymmetry:** Asymmetric electrical activity between comparable areas on both hemispheres may indicate a lesion or dysfunction.
6) **Periodic Lateralized Epileptiform Discharges (PLEDs):** This is often a sign of acute brain dysfunction, commonly seen after acute brain injuries.

However, manually examining EEG results is a tedious and time-consuming process that is susceptible to human mistakes. Interpreting EEG signals is challenging due to their complexity and variability, often necessitating expert analysis. However, subjective interpretations might result in conflicting diagnoses. This not only prolongs the diagnostic process but also raises the likelihood of missing small yet clinically important anomalies. Additionally, the large amount of data produced during a standard EEG session, which can sometimes span hours of recordings, can render thorough manual analysis unfeasible. The problems highlight the pressing requirement for automated or semi-automatic techniques, such ML algorithms, to aid in the precise and efficient analysis of EEG data. Progress in these fields could transform the industry by offering more dependable, quick, and convenient assessments for various neurological disorders and research purposes.

2.2.3. Automated Methods in EEG Analysis

As already mentioned, the laborious nature of manual examination of EEG recordings created the need for automated methods of analysis and categorization.

2.2.3.1. Historical overview

Automated EEG analysis methods have been developed over several decades, although progress has increased notably due to advancements in processing capacity and machine learning algorithms. Automating EEG analysis began in the 1960s and 1970s with the development of basic algorithms to identify certain waveform patterns such spikes or sharp

waves associated with epilepsy. These initial techniques mostly consisted of rule-based systems and depended on simple statistical metrics.

In the 1980s and 1990s, advanced techniques such as digital signal processing and fundamental machine learning algorithms like Support Vector Machines (SVM) were created. During this time, there was an increasing interest in automated sleep stage classification, seizure identification, and artifact removal. Nevertheless, these techniques were not sufficiently strong for broad clinical use and were mostly utilized in research environments.

The 2000s saw a resurgence of major developments in machine learning and data science, leading to a considerable impact on automated EEG analysis. Support Vector Machines (SVM), Random Forests, and sophisticated neural network designs were utilized for analyzing EEG data. The advent of deep learning in the late 2010s has expanded the potential for intricate pattern identification in EEG data, leading to more precise and dependable automated assessments.

Modern techniques for analyzing EEG data are advancing rapidly, utilizing complex machine learning algorithms such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer models. These methods are utilized in both research and are starting to be integrated into clinical practice for purposes such as real-time monitoring and diagnostic assistance. Automated EEG analysis has advanced significantly in the previous two decades due to improvements in computational techniques and machine learning algorithms, while not being a new concept.

2.2.3.2. Overview

In the 2020s, automated methods are transforming EEG analysis. Modern computers' increased capabilities have made computationally expensive denoising and decomposition techniques accessible in any research environment. Previously, techniques such as Independent Component Analysis (ICA) and Automatic Artifact Subspace Reconstruction (ASR) needed dedicated high-performance computing clusters. Independent Component Analysis (ICA) was commonly used for blind source separation in order to separate individual brain and artifact sources. However, it could require many hours to days to analyze a single EEG dataset on computers from the early 2000s. ASR, a technique used to detect and eliminate abnormalities in EEG data, encountered computing constraints since it required real-time data processing and intricate matrix calculations. On previous systems, ASR could require 30 minutes to an hour to analyze only a few minutes of EEG data.

Nowadays, thanks to multi-core CPUs, Graphics Processing Units (GPUs), and efficient ML libraries, these techniques can be performed in just minutes or even seconds. Modern systems can efficiently do these calculations using tens to hundreds of channels in real-time, making them suitable for advanced research facilities, clinical settings, and smaller research groups. The democratization of computing is greatly increasing the potential for advanced, automated EEG analysis.

The principles discussed about EEG denoising techniques and advanced computational systems are equally relevant to all other aspects of EEG analysis, including signal processing and machine learning. It is important to note that EEG analysis can be conducted in two ways, each providing distinct insights into brain activity. Event-Related Potentials (ERPs) are commonly utilized to investigate the brain's reaction to particular stimuli or occurrences. ERPs offer a precise measure of brain activity synchronized with time, which is useful for studying cognitive functions like as attention, memory, and decision-making under strict control. EEG analysis can also examine oscillatory activity or 'ongoing' brain waves that are not directly linked to specific external stimuli. This method is frequently used to investigate continuous phenomena that are not tied to specific events, such as awareness states, emotional control, or baseline brain function. Both methodologies provide essential viewpoints and are frequently complementary, offering a more thorough comprehension of brain function and malfunction.

During this doctoral research, no studies employing ERPs were conducted; instead, the focus was on oscillatory activity. ML served as the main axis of the research, so significant attention was given to different approaches for EEG signal transformation, decomposition, and feature extraction. Creating a feature matrix from the EEG signal can be achieved by extracting features in various ways. Statistical features can be extracted directly from the signals, while frequency domain metrics such as Relative Band Power (RBP) can be obtained through frequency or time-frequency transforms. Synchronization features can be extracted using methodologies that consider the topology of the electrodes. In such studies, the next step involves proposing a novel ML pipeline (or a combination of well-established methodologies) that uses these feature matrices as input and attempts to classify EEG signals based on a neurological condition—either as healthy or not healthy, severe or not severe. A standard ML pipeline would include all the conventional steps found in any other classification problem, such as feature extraction, feature reduction, hyperparameter optimization, machine learning model selection, and evaluation metrics. Furthermore, state-of-the-art advancements in ML, including developments in deep learning and especially transformer neural networks, are also being incorporated into EEG research. All these methodologies are meticulously described in Chapter 4.

2.2.4. Frequency bands of Interest

One of the fundamental aspects of EEG analysis is the frequency bands of interest. Traditionally, EEG signals are divided into five primary frequency bands, each corresponding to different brain activities and serving specific purposes in both clinical and research settings. These bands are:

- **Delta Waves (0.5–4 Hz):** This frequency band is predominant in deep sleep stages and is also associated with unconscious bodily functions such as digestion. Clinically, increased delta activity can be indicative of brain injuries or abnormalities.

- **Theta Waves (4–8 Hz):** Commonly observed in drowsiness and light sleep, theta waves are also linked to creativity, intuition, and emotional responses. Elevated theta activity is often seen in meditative states and in some neurological disorders.
- **Alpha Waves (8–13 Hz):** These waves are usually detected when a person is relaxed but alert, often with eyes closed. Alpha activity is associated with wakeful relaxation and can serve as a baseline EEG rhythm.
- **Beta Waves (13–30 Hz):** This frequency band is predominant during problem-solving, focused attention, and analytical thinking. Beta waves are a sign of active cognitive engagement. In a clinical context, altered beta activity can signal stress or anxiety.
- **Gamma Waves (30–100 Hz):** Though less commonly studied due to their lower amplitude, gamma waves are associated with higher-order cognitive functions such as perception, consciousness, and rapid information processing.

The EEG signals are split into these frequency bands either by using Finite Impulse Response (FIR) filters directly, or by applying a frequency transformation such as Power Spectral Density (PSD) in the whole signals and then calculating the energy of these frequency bands by only keeping the points corresponding to each range. Figure 3. The 5 frequency bands. [7] displays the waveforms of the 5 different frequency bands.

2.3. Conditions investigated through EEG

EEG is used to investigate a wide range of psychological and neurological conditions as well as the brain function in general. Some of the key conditions commonly studied, quantified or diagnosed through EEG are epilepsy [8], [9], sleep disorders, encephalopathy [10], strokes and dementias[11], [12], [13], [14]. Mental health conditions such as depression and anxiety [15] or Attention-Deficit/Hyperactivity Disorder [16] have, also, been studied through EEG. Moreover, cognitive and behavioral research regarding cognitive load, BCI [17], emotional processing [18] or sensory processing [19] have employed EEG as a quantification tool. In this section, the main conditions that were investigated throughout this doctoral study will be analyzed.

2.3.1. Epilepsy

Epilepsy is the most studied neurological condition through EEG. This neurological disfunction is concerning since it affects more than 70 million people worldwide according to the World Health Organization. Epilepsy originates from seizures, which are abnormal electrical discharges in the cortical neurons.

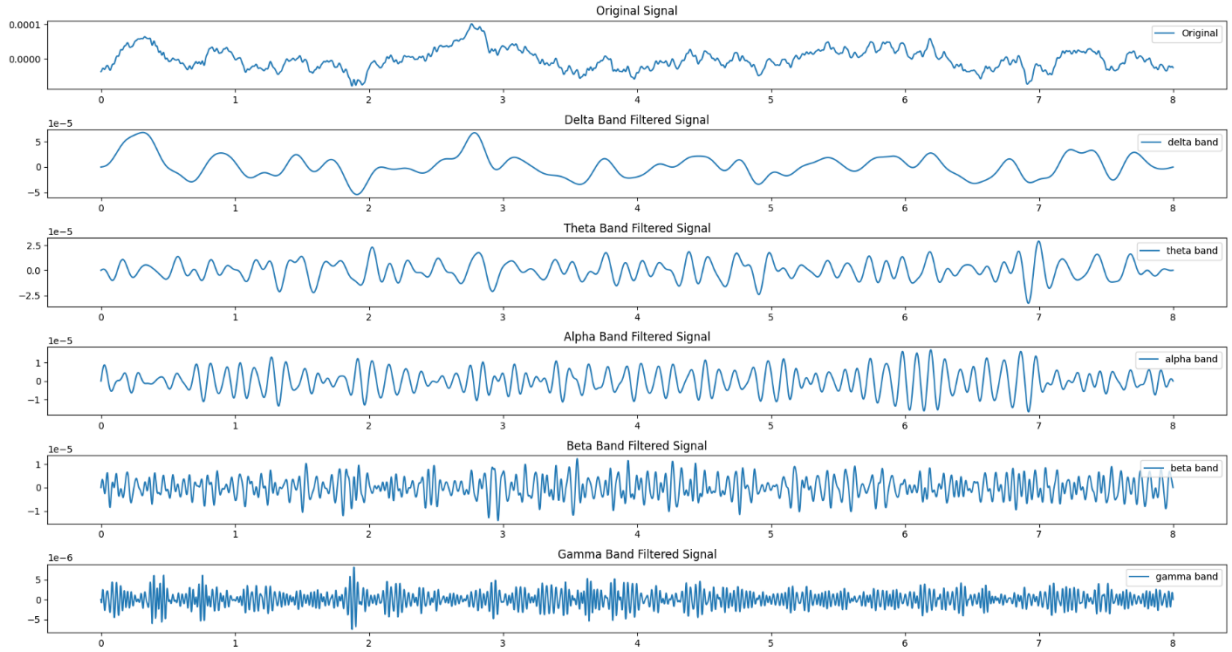


Figure 3. The 5 frequency bands.

In the majority of epilepsy cases, pharmaceutical treatments, such as antiepileptic drugs (AEDs), are effective in managing the symptoms by reducing the frequency and severity of seizures. However, for approximately 25% of patients, known as drug-resistant or refractory epileptics, conventional medication doesn't provide adequate control. For these individuals, surgical intervention, which may involve removing the portion of the brain responsible for the seizures, becomes a consideration. The understanding of epilepsy has a rich historical context. While Hippocrates, the ancient Greek physician, was among the first to suggest that epilepsy originated in the brain, it was the British neurologist John Hughlings Jackson in the 19th century who profoundly advanced our knowledge. Jackson posited that epileptic seizures arose from the cerebral cortex and provided in-depth clinical descriptions of partial (focal) epileptic seizures, laying the groundwork for modern epileptology. Epileptic seizures, although not typically fatal, pose risks such as injury during an event. They can also significantly disrupt a patient's daily activities, from driving to work commitments, leading to feelings of isolation or depression. An epileptic seizure is scientifically defined as a transient brain dysfunction wherein a group of neurons, often in a specific region of the brain, exhibits synchronous and hyperactive activity, resulting in abnormal electrical signals that manifest as observable clinical symptoms. Even though a lot of imaging tools have been used for epilepsy detection, like Magnetic Resonance Imaging (MRI), Functional-MRI,

PET-scan and others, EEG is recognized as the main diagnostic tool, due to the fact that the epileptiform discharges can be observed on an EEG recording and distinguished from normal neural activity.

Extended EEG time series, which can last for several hours, often require expert neurologist review for the identification of epileptic waveforms, making it a demanding and time-consuming task. Given the intricate nature of seizures and the large number of people affected by epilepsy, there's a strong push in research to continually create new methods for automated seizure detection to aid medical professionals. Global research initiatives have leveraged signal processing and machine learning approaches to explore and identify abnormal spikes, spike waves, and spike-wave complexes in interictal EEG data. Efforts have also been made to recognize early indicators of seizure onset for predictive purposes. The refinement of Time-Frequency analysis methods, coupled with the rise of Deep Learning, has brought scientific interest in the automated diagnosis of epilepsy.

While research on automatic epilepsy detection using EEG dates back to the 1980s, the past decade has seen a rapid increase in the number of published papers. Most of these studies gather EEG recordings from individuals with epilepsy and healthy controls, and then propose a method that uses various signal processing and feature extraction techniques, often combined with traditional Machine Learning algorithms or newly developed ones, aiming to train a system that can automatically classify an EEG time window as epileptic or not. Research groups focused on processing, algorithmic, or computational aspects often find it challenging to contribute to this field independently, as clinical recordings necessitate collaboration with medical organizations. Consequently, several epilepsy EEG databases have been made publicly available to facilitate the development of automated seizure detection methods. Therefore, many experimental studies tend to rely on these published databases rather than conducting their own data collection.

The most famous of these databases are: The database of Bonn University, which is consisted of 5 subsets of EEG recordings named Z,O,N,F,S taken from 5 healthy subjects and 5 epileptic subjects. Each group contains 100 single-channel recordings, each lasting 23.6 seconds (with a signal length of 4096 samples). The data has a sampling frequency of 173.61 Hz. The Childrens Hospital of Boston – Massachusetts Institute of Technology (CHB-MIT) database, which includes long, continuous, multi-channel recordings from 24 people with drug-resistant seizures, aged between 1.5 to 22 years. The Epilepsy Center of University of Freiburg database which includes long-term EEG recordings taken from 21 patients aged 13-47 years suffering from epilepsy, recorded from 6 intracranial channels. Apart from these, there are more databases with EEG recordings.

A systematic review on studies that use publicly available databases to propose a scheme for epilepsy detection using EEG signals is incorporated in this dissertation at chapter 5. Thus, a more detailed description of the databases is provided there.

2.3.2. Alzheimer's Disease and other Dementia

Alzheimer's disease (AD) stands as a progressive neurodegenerative disorder, ranking among the most commonly diagnosed types of dementia in elderly. It manifests through cognitive deterioration and shifts in behavior. As the median age of the global population tends to increase, it is expected that AD as well as other age-related dementias will rise. Research indicates that AD ranks as the sixth predominant cause of death in the United States, uniquely continuing its significant upward trajectory among the top ten causes. Reports from 2020 documented over 50 million dementia cases, with projections estimating AD patient counts to touch 75 million by 2030 and soar to 131 million by 2050. The prevalence of AD remains consistent across genders, registering at 1.4% for those aged 65-70 and escalating to 24% for those surpassing 85 years of age. Regarding its symptoms, the initial indicators of the disease manifest as challenges in recollecting events of short-term memory. As it advances, it progresses to issues regarding speech, orientation, fluctuating moods, neglecting personal care, and behavioral shifts. As the disease intensifies, bodily functions gradually deteriorate, at the individuals ultimately result to death. At present, no established cure exists for AD, with the available therapeutic interventions offering only modest relief from symptoms.

For the diagnosis of probable AD, specific criteria such as postmortem confirmation of neuropathological changes (accumulation of neuritic plaques and neurofibrillary tangles containing hyperphosphorylated tau proteins) should be met by the patient. Figure 4 a) The brain of a heathly subject, b) the brain of an AD patient. Neuritic plaques and neurofibrillary tangles accumulation as a result of amyloid-beta peptide's (Aβ) accumulation in the most affected area of the brain, the medial temporal lobe and neocortical structures. presents the structural differences of an AD diseased brain and a healthy brain [20]. The initial diagnosis of the AD, however, needs several criteria as provided by the Diagnostic and Statistical Manual (DSM-IV) [21] and the National Institute of Neurological, Communicative Disorders and Stroke, Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [22]. To diagnose probable AD, other possible causes of cognitive decline are eliminated. According to the protocol, this procedure involves a combination of:

1) Evaluation of patient's medical history and physical examination for overall health assessment
2) Cognitive & Neurophysiological testing for memory, language, attention, problem-solving assessment to identify any cognitive impairments or changes that may be indicative of Alzheimer's Disease
3) Imaging techniques, such as MRI or PET scans, used to detect structural or functional changes in the brain, to identify brain shrinkage or atrophy (MRI) and to detect beta-amyloid plaques and measure glucose metabolism in the brain through radioactive tracers (PET).
4) Laboratory tests, such as blood tests, to rule out other possible causes of cognitive decline such as vitamin deficiencies, thyroid problems or infections

5) Cerebrospinal fluid analysis to measure specific biomarkers associated with AD such as beta-amyloid and tau proteins.

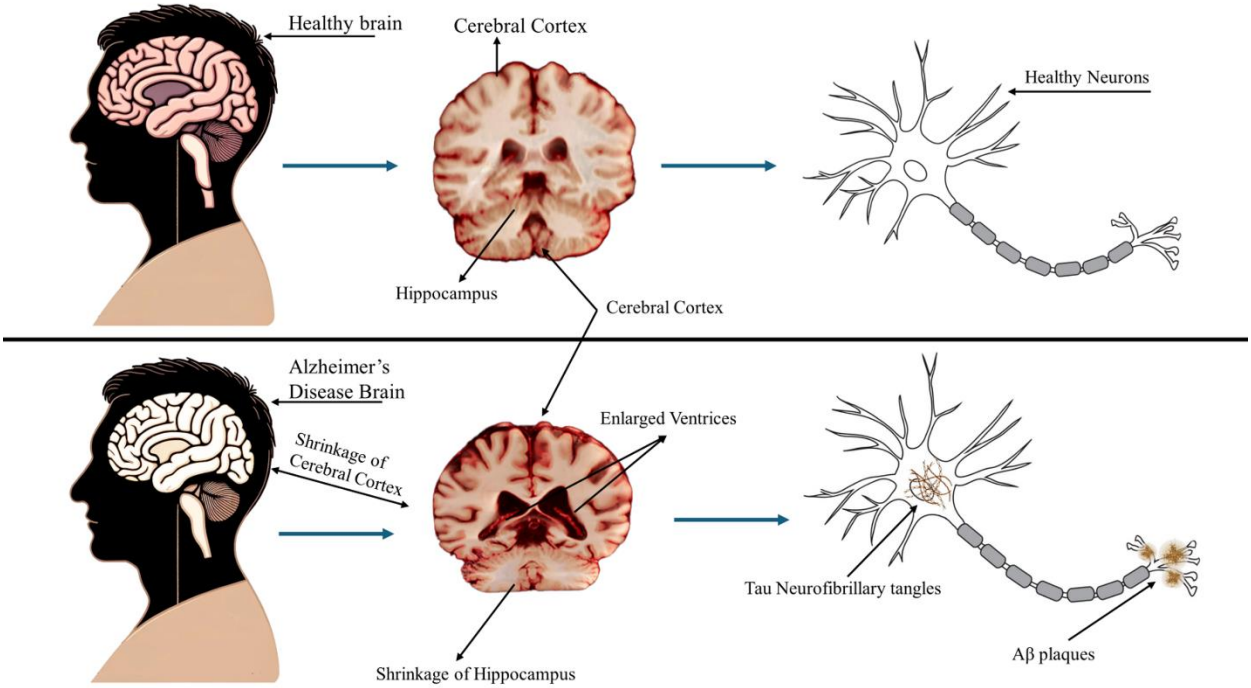


Figure 4 a) The brain of a heathly subject, b) the brain of an AD patient. Neuritic plaques and neurofibrillary tangles accumulation as a result of amyloid-beta peptide's (Aβ) accumulation in the most affected area of the brain, the medial temporal lobe and neocortical structures.

The timely nature of the protocol leads to diagnosis that is often too late and not available to everyone. Thus, creating new methods for automated and timely diagnosis is important, in order for the patient to receive better medical support and prolong their lifespan, and for the relatives or caregivers to be better prepared on how they should treat them.

Alterations in brain activity may be key findings of neurodegenerative disorders such as AD. There are various methodologies for measuring brain activity (such as MRI or PET-scan, as already mentioned), each with different spatiotemporal resolution and applicability. EEG is a tool that has great temporal resolution, yet it is not widely employed due to having low spatial resolution and being prone to noise. Nevertheless, recent advancements in computational methods such as Low-Resolution Electromagnetic Tomography (LORETA) promise increased spatial resolution by providing estimation capabilities of the underlying brain generators location. Furthermore, techniques such as Independent Component Analysis (ICA) and Artifact Subspace Reconstruction (ASR) (which perform artifact reduction or removal) have now become computationally available thus making the EEG a prominent tool in neurodegenerative disease diagnosis.

2.3.3. Other Conditions

EEG is a versatile clinical tool with applications beyond epilepsy or dementia, as mentioned above. For example, it is deployed in various research studying cognitive and emotional states, including stress and anxiety, mood disorders, cognitive workload, relaxation, motivation and emotion recognition. Various research works suggest that a functional lateralization in the frontal cortex is related to affective processing and may be an indicator for physiological stress. Specifically, Frontal Alpha Asymmetry is a measure used in a variety of protocols designed to evaluate mental stress. Also, there is evidence that such Frontal Alpha Asymmetry is related to depression [23]. Moreover, concepts such as attention and concentration have been analyzed with formulas such as beta/alpha power [24]. These applications highlight the versatility of EEG in capturing a wide array of both pathological and non-pathological brain states and can find use in neuromarketing, game design and other areas.

2.4. Brain-Computer Interfaces

Brain-Computer Interfaces (BCI) are systems that enable the direct and bidirectional communication between the brain and external devices, such as a PC or a Virtual Reality system, by bypassing the typical pathways of peripheral nerves and muscles. Although this concept seems futuristic, its foundations date back several decades, and specifically to the early 1970s where Vidal [25] introduced and explored the feasibility of direct communication between the brain and a computer and is often credited as the foundational work in the field of BCI. He introduced the idea that brain electrical activity, specifically the EEG, could be used as a channel for sending commands directly to a computer, bypassing the usual channels of communication.

In the later years, during late 1990s and early 2000s, advances in neuroscience led to the creation of increasingly advanced BCI systems, allowing for more refined interactions between the brain and external equipment. The fast expansion of computer computational capabilities, machine learning, and signal processing techniques over the last two decades has catapulted BCI from a theoretical notion to practical applications in medicine, rehabilitation, and beyond. Medical rehabilitation applications involve assisting patients with paralysis or other motor disorders by allowing them to communicate or control prosthetic limbs. Neurofeedback training applications involve providing feedback on certain brain activities, aiding in conditions like Attention Deficit Hyperactivity Disorder (ADHD) or even for enhancing cognitive performance.

2.4.1. EEG-based BCI

EEG is one of the most commonly used modalities for BCIs due to its non-invasiveness and relatively easy setup. Mostly, EEG-BCI applications involve Motor Imagery tasks, Steady-State Visual Evoked Potentials (SSVEPs), OR Event Related Desynchronization/ Synchronization (ERD/ERS). Motor Imagery is describing the applications where the

participant is imagining that they perform a motor movement and the distinct EEG patterns that are generated are used to detect or classify it. Most Motor Imagery applications regard rehabilitation purposes. SSVEPs describe how repetitive visual stimuli, like flashing lights at different frequencies, can be used to induce brain responses that are detectable and usable in BCIs.

BCI applications are usually developed to achieve one of the following goals. First and foremost, the main goal is assisting individuals with severe motor disabilities to communicate or control external devices such as wheelchairs. Secondly, neurofeedback applications are developed for therapeutic purposes, where individuals are provided with real-time feedback on their brain activity to achieve certain cognitive states such as better relaxation or focus. Last but not least, a huge effort is given on developing BCI applications for gaming and entertainment, that being integrating EEG devices in the industry for enhanced gaming experience and control.

Despite their apparent interest, there are significant restrictions impeding the progress of these BCI devices. Although BCI technology has advanced significantly in clinical and consumer applications, a notable drawback is evident when comparing wearable BCI devices to clinical versions. Consumer-oriented wearable gadgets frequently show significantly diminished signal quality. These devices often have a low signal-to-noise ratio (SNR), which increases the chances of collecting noise, artifacts, or undesired interferences. The reduced Signal-to-Noise Ratio (SNR) can greatly affect the quality and dependability of EEG data analysis, hindering the ability to attain the same degree of precision and responsiveness found in clinical-grade EEG systems. Additionally, the requirement for user training and the presence of unique variations in EEG patterns also restrict the capability of these systems.

2.5. Clinical devices and Wearable EEG

Within the field of EEG, technology has evolved into two separate categories: clinical EEG devices and wearable EEG systems. Clinical EEGs have always been crucial in medical diagnoses and neuroscience research for their sensitivity and precision. However, the development of wearable EEG technology has introduced a new level of ease and flexibility, broadening the range of EEG applications.

Clinical EEG machines offer excellent sensitivity and precision, making them essential for thorough neurological diagnosis. These devices usually use numerous electrodes positioned based on the conventional 10-20 system and need conductive gels for the best signal transmission. Their advanced technical skills enable a thorough examination of brain function, essential for detecting illnesses such as epilepsy, sleep problems, and brain traumas. Yet, the large size of these configurations and the requirement for proficient specialists for electrode positioning restrict their application to regulated settings such as hospitals and clinics. Although limited by these restrictions, the exceptional level of detail and precision in clinical EEGs remains crucial for medical diagnosis and brain research.

Wearable EEG systems are becoming increasingly popular due to their portability and user-friendly nature. These systems often have fewer electrodes, with many being dry, which eliminates the requirement for conductive gels and makes the setup process simpler. Combining wireless connectivity and battery-powered operation improves their portability, enabling a variety of applications outside clinical environments. Wearable EEG devices are not just used for research and clinical purposes but have also been integrated into consumer electronics, meditation, neurofeedback, and gaming. These devices have a wide price range, from budget-friendly alternatives for the average user to high-end systems tailored for advanced study.

Choosing between clinical and wearable EEG devices typically involves balancing precision and practicality. Clinical EEGs provide intricate and precise data but are not as convenient or accessible as wearable systems. Despite their wider range of uses and user-friendly interface, wearable EEGs often offer poorer spatial resolution and are more susceptible to noise and artifacts, which may limit their suitability for specific medical or research purposes.

Chapter 3

EEG Signal Analysis

"The brain struggling to understand the brain is society trying to explain itself.

-Colin Blakemore

3.1. Preprocessing

The EEG data is likely to be corrupted by noise from several sources including the power line or artifacts caused by muscle, eye, or jaw movements when recorded. Preprocessing in EEG analysis is crucial for reducing difficulties and improving the quality of the signals. Preprocessing of EEG data corrects distortions to enhance accuracy and reliability for various analyses such as clinical diagnostics, cognitive investigations, or brain-computer interfaces. Efficient preprocessing can separate the brain signals of interest, enhancing the signal-to-noise ratio and allowing researchers and clinicians to make more accurate interpretations from the data. Hence, comprehending and utilizing suitable preprocessing methods are crucial for those involved in EEG analysis.

3.1.1. Basic preprocessing techniques in EEG Analysis

The most basic and straightforward preprocessing techniques used in EEG analysis are: Bandpass filtering, Rereferencing, Downsampling, Channel Interpolation and Epoching. A study may use one or more of these techniques depending on the study objectives and on the methodologies that will be used for the analysis.

3.1.1.1. Filtering

Filtering is a fundamental process in analog and digital signals. A filter operate on different frequencies, allowing some of them to pass, while attenuating others. EEG filters are usually designed to allow only the frequencies of interest in the brain activity, as defined in section 2.2.4. So the main line is either using a Band Pass filter to allow all frequencies of the bands that are required (for example 0.4 – 45 Hz, for Delta-Gamma rhythms), or using a different Band Pass filter for every rhythm and creating a number of signals equal to the number of rhythms under examination (usually all five). Finally, Notch filter (a filter that removes a certain frequency) might be used for eliminating the power line noise (either on 50 Hz or 60 Hz, depending on the location of the recording). All these filters can be implemented either as Finite Impulse Response (FIR) filters or as Infinite Impulse Response (IIR) filters [26].

FIR filters

A FIR filter is a filter with no feedback in its equation, making it inherently stable and capable of producing linear phase responses. This makes them the go-to option for applications that require linear phase characteristics. However, the absence of feedback means that FIR filters often require more coefficients to achieve the same performance as their IIR counterparts. Each additional coefficient increases the computational load and memory needs for the digital signal processor. In resource-intensive systems, these

requirements could render FIR filters impractical. Thus, a FIR filter is most usually used for offline EEG signal processing. For a causal discrete-time FIR filter of order N, each value of the output sequence is a weighted sum of the most recent input values, as described in eq. 3.1.:

$$y[n] = b_0 x[n] + b_1 x[n-1] + \cdots + b_N x[n-N] = \sum_{i=0}^{N} b_i \cdot x[n-i] \quad (3.1)$$

Where:

- $x[n]$ is the input signal
- $y[n]$ is the output signal
- N is the filter order, an N[th] order filter has $N + 1$ terms on the right-hand side.
- b_i is the value of the impulse response at the i_{th} instant for $i \in [0, N]$ of an N^{th} order FIR filter.

The filter's effect on the sequence $x[n]$ is described in the frequency domain by the convolution theorem:

$$\underbrace{\mathcal{F}\{x * h\}}_{Y(\omega)} = \underbrace{\mathcal{F}\{x\}}_{X(\omega)} \cdot \underbrace{\mathcal{F}\{h\}}_{H(\omega)} \text{ and } y[n] = x[n] * h[n] = \mathcal{F}^{-1}\{X(\omega) \cdot H(\omega)\}, \quad (3.2)$$

Where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the discrete-time Fourier Transform (DTFT) and its inverse, as described in chapter 3.3.

IIR filters

Contrary to the FIR filters, an IIR filter gets the past outputs as feedback in its equation, allowing it to operate much more efficiently. However, the linear phase is almost impossible to maintain. An IIR filter is generally used for online processing. The difference equation that defines how the output signal is related to the input signal is described in eq. 3.3.:

$$\sum_{j=0}^{Q} a_j y[n-j] = \sum_{i=0}^{P} b_i x[n-i] \quad (3.3)$$

Where:

- P is the feedforward filter order
- b_i are the feedforward filter coefficients
- Q is the feedback filter order
- a_j are the feedback filter coefficients
- $x[n]$ is the input signal
- $y[n]$ is the output signal

3.1.1.2. Rereferencing

An EEG recording essentially measures the potential difference between an active electrode placed on the scalp and a reference electrode. The reference electrode can be placed at various locations, such as the earlobe, the nose tip, or another part of the scalp, depending on the re-referencing technique or montage being used. An electrode montage is essentially a specific arrangement of EEG channels, determined by the reference electrode and the placement of other electrodes on the scalp. Generally, the electrodes are placed on the scalp on pre-determined locations following the international 10-20 system [27]. Then, to reduce noise or enhance specific characteristics a reference technique or montage is applied. Some of the most used referencing techniques are described below.

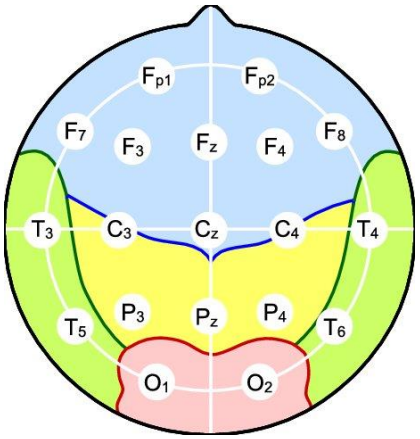


Figure 5 The 10-20 international system electrode topography. Blue: Frontal lobe, Yellow: Parietal lobe, Green: Temporal lobe, Red: Occipital lobe.

Monopolar or Referential Montage

This technique involves connecting each electrode to a single reference point. This point is usually the sum of the 2 reference electrodes placed on the mastoids, A1 and A2. For example, if every channel is recorded as the difference with Cz, after rereferencing it is now expressed as difference with (A1+A2)/2

$$(F3 - Cz) - \frac{(A1 - Cz) + (A2 - Cz)}{2} = F3 - (\frac{A1 + A2}{2}) \qquad (3.4)$$

The main advantage of using a referential montage is its sensitivity to local features (monopolar montages are generally more sensitive to local electrical fields than bipolar montages, making them useful for detecting focal abnormalities or features) and its flexibility in post-processing (data in referential montage can easily be later re-referenced to any other montage).

Average Reference Montage

Referencing to the average, commonly known as Average Reference, involves taking the mean of all the active electrodes and using that as the common reference for each electrode. This method aims to minimize the influence of a single, potentially noisy, reference electrode by distributing the reference across all electrodes.

Bipolar Montage

In bipolar referencing, each channel is formed by taking the difference in voltage between adjacent electrodes. Unlike the referential montage, where all electrodes are referenced to a single point, bipolar referencing provides a more localized view of brain activity. This is particularly useful in applications where precise spatial resolution is required, such as in epilepsy studies.

Banana Montage

Also known as the longitudinal bipolar montage, the banana montage involves comparing the voltage of an electrode to that of its adjacent electrode. This forms a chain of electrode pairs that usually follows the curve of the scalp, resembling the shape of a banana. This montage is often used to localize the source of abnormal electrical activity in the brain.

3.1.1.3. Electrode Interpolation

Electrode Interpolation is a technique used when the signal of an electrode is too corrupted to be saved. In these cases, the certain electrode is discarded, and it is reconstructed using as input the data of neighboring electrodes, based on geometric proximity, topological features or both. Various interpolation algorithms, such as spherical splines or simple weighted averaging, can be used to estimate the voltage at the bad electrode based on its neighbors. However, interpolating an electrode does not solve the underlying issue, but is rather a corrective measure and a last resort when no other correction technique can be used.

3.1.1.4. Epoching

A fundamental procedure of EEG analysis when in the context of training a machine learning algorithm is epoching or time-windowing. To train a classifier for distinguishing signals of different mental states or different neurological conditions, a properly sized dataset is required. However, enlisting an adequate number of participants is often too difficult due to bureaucratic or other reasons. Thus, a common technique is splitting the EEG signal of a participant into multiple different EEG signals of fixed duration (for example 10 seconds) and treating each EEG signal as a different recording. By doing so, a machine learning algorithm can be trained with dramatically less participants.

Epoching can be performed using either overlapping or non-overlapping intervals. Regardless of the approach, special attention must be given to the choice of validation technique in machine learning. Utilizing K-fold validation may lead to having data from the

same participant in both the training and test sets, which could artificially inflate performance metrics. Therefore, when employing epoching, the Leave-One-Subject-Out (LOSO) validation technique is generally the preferred choice.

3.1.2. Artifact removal techniques

Besides the basic methodologies already mentioned, more sophisticated techniques for artifact removal have been developed during the last decades. In this section, the 2 most used techniques, being Independent Component Analysis (ICA) [28] and Artifact Subspace Reconstruction (ASR) [29] will be discussed.

Independent Component Analysis

ICA is a computational method employed for separating multivariate signals into independent non-Gaussian signals, also known as independent components. In the context of EEG, ICA is used to separate brain signals from non-brain signals (artifacts) such as eye movements, muscle activity, and line noise. The ICA technique gets as input N number of channels and finds a linear representation of non-Gaussian data of N components that are statistically independent. In other words, ICA separated independent sources linearly mixed in several sensors. Then, by classifying these components based on their characteristics as

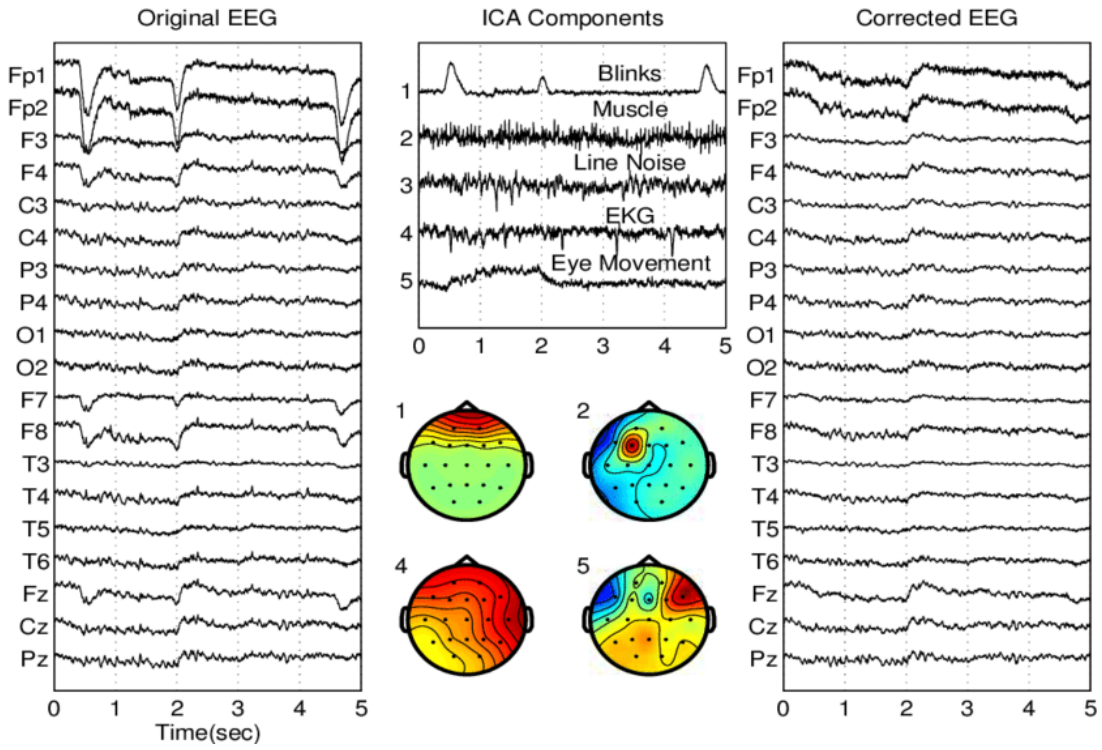


Figure 6 Left: EEG signal before ICA processing. Right: EEG after removing blink artifacts. Middle: Different types of artifacts expressed as signals (top) and as heatmaps (bottom). Adapted from "Extended ICA Removes Artifacts from Electroencephalographic Recordings", by Jung et al., 1998.

artifact or brain components, removing the not necessary and reconstructing the signal, a new set of non-infected EEG is produced. There are ICA algorithms used in EEG data analysis, with the most used being FastICA[30] and Infomax ICA [31]. Figure 6 Left: EEG signal before ICA processing. Right: EEG after removing blink artifacts. Middle: Different types of artifacts expressed as signals (top) and as heatmaps (bottom). Adapted from "Extended ICA Removes Artifacts from Electroencephalographic Recordings", by Jung et al., 1998. represents EEG signals before and after ICA reconstruction [32].

Artifact Subspace Reconstruction

ASR focuses on identifying and reconstructing artifact components directly in the data space by defining a "clean" subspace of the data and then reconstructing the original data based on this clean subspace. A clean subspace is usually defined using a calibration dataset that is devoid of artifacts, or where artifacts have been manually removed. The EEG data are then projected onto this clean subspace. This operation tends to remove artifact components that lie outside of the clean subspace. It is designed to handle non-stationary and non-repeating artifacts. Figure 7 Signal in red is contaminated with artifacts before ASR. Signal in blue is cleared after applying ASR.shows an EEG signal before and after ASR reconstruction.

However, ASR has some prerequisites [33]. Data should be already high-pass filtered, be full rank (meaning no common reference montage or bipolar montage can be use beforehand) and miscellaneous channels should have been removed before.

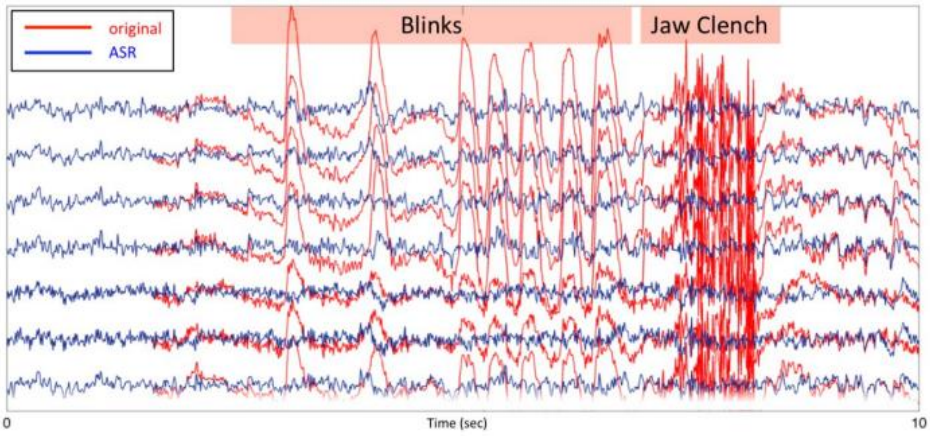


Figure 7 Signal in red is contaminated with artifacts before ASR. Signal in blue is cleared after applying ASR.

3.2. Frequency Domain Analysis

While time-domain analysis of EEG signals provides valuable insights into brain dynamics, it often falls short in revealing the intricate oscillatory patterns that underlie various cognitive and pathological states. When examining EEG waveforms for various neurological conditions, the key characteristics usually come from the frequency domain.

Examples include increased alpha activity during relaxed or drowsy states, and theta and low-beta frequencies associated with rapid eye movement (REM) sleep. In Alzheimer's disease, a change in the theta/beta ratio has been observed. Epileptic episodes often present spike-and-wave discharges in the frequency domain, and delta waves can be indicative of brain injury or pathology. Gamma frequency activity has also been linked to cognitive functions such as attention and working memory. Because of these strong correlations between frequency bands and specific neurological states or conditions, EEG signals are commonly converted to their frequency domain representation, facilitating more targeted analysis for both diagnostic and therapeutic purposes.

The fundamental aspect of frequency domain analysis in EEG signal processing is the Fourier Transform. It breaks down a time-domain signal into its individual frequencies, offering a detailed perspective of the signal's frequency components. Most techniques used for analyzing frequency in EEG, like Short-Time Fourier Transform (STFT), Wavelet Transform, and several types of spectrum analysis, are based on or connected to Fourier Transform principles. Utilizing Fourier-based methods can provide useful information about the neurological disorders or cognitive processes that exhibit distinct frequency patterns in EEG signals.

3.2.1. Fourier Transform

The Fourier Transform (FT) is a mathematical procedure that transforms a $x(t), t = time$ function to $X(\omega), \omega = frequency$. It was first proposed by Joseph Fourier in the 19th century and has been widely applied to all signal processing related applications since then. The equation of the Fourier transform is given in eq. 3.5.

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt \tag{3.5}$$

In order to apply the Fourier Transform to real-life, digital signals an implementation such as Discrete Fourier Transform (DFT) should be employed. DFT is the most fundamental implementation of FT. DFT deals with a discrete set of samples. It transforms a finite, discrete-time signal into a finite, discrete-frequency representation. Eq. 3.6 describes the DFT:

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{i2\pi}{N} kn} \tag{3.6}$$

Where a sequence $x[n]$ of length N is transformed to another sequence $X[k]$.

Figure 8 Top: A signal in the time domain which is a convolution of 3 sinusoidal signals. Botttom: The Fourier Transform of the signal. Adapted from “Key Concepts: Fourier Transforms and Signal Processing”, by Tom White, 2020 (blog article).presents an illustration of the Fourier Transform.

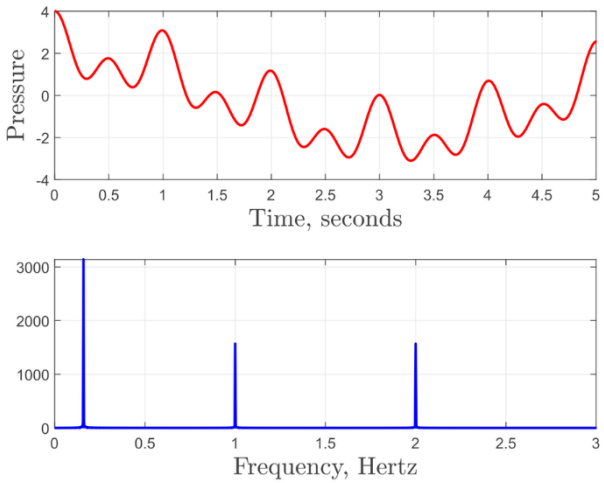


Figure 8 Top: A signal in the time domain which is a convolution of 3 sinusoidal signals. Botttom: The Fourier Transform of the signal. Adapted from "Key Concepts: Fourier Transforms and Signal Processing", by Tom White, 2020 (blog article).

3.2.1.1. Fast Fourier Transform

The Fast Fourier Transform (FFT) is an algorithm used to efficiently compute the DFT and its inverse. In the context of EEG analysis, FFT is a tool for transforming raw time-domain data into the frequency domain, enabling the identification of specific frequency bands and their contributions to the recorded signals. It was popularized by Cooley et al. at 1967 [34], when they proposed an algorithm that reduced the complexity of computing the DFT from $O(N^2)$ to $O(N \log N)$, making it much more efficient for large datasets and enabling real-time signal processing and analysis. The basic idea behind FFT is to decompose a DFT of any composite size $N = N_1 N_2$ into smaller DFTs of sizes N_1 and N_2, recursively, to reduce the computation time.

3.2.2. Power Spectral Density

The FFT is a potent tool for frequency domain analysis, however it comes with limitations such as the assumption of signal stationarity and sensitivity to the choice of window length. Power Spectral Density (PSD) is a method that alleviates these challenges providing information of how the power of a signal is distributed over different frequency components. The average power P of a signal $x(t)$ over all its time is given by the following equation, where T the period and $\hat{x}_T(f)$ the Fourier Transform of the time convolution of $x_T^*(-t)$ and $x_T(t)$ where * represents the complex conjugate.

$$S_{xx}(f) = \lim_{T\to\infty} \frac{1}{T} |\hat{x}_T(f)|^2 \quad (3.7)$$

Or for a discrete-time signal, PSD can be estimated using the FFT as:

$$\text{PSD}\ (f) = |FFT(x[n])|^2 \quad (3.8)$$

Regarding EEG analysis, PSD is used for identification of the dominant frequency bands (Delta, Theta, Alpha, Beta, Gamma) and their power contributions, which are then correlated with different cognitive or neurological states.

There are a handful of algorithms for PSD calculation, with the most common used being Welch's Method, Burg's Method and Multitaper Method.

3.2.2.1. Welch Method

The most common method for PSD calculation is the Welch method [35], first proposed by Welch in 1967. This method is consisted of 4 steps:

1) **Segmentation:** The long data sequence is divided into overlapping or non-overlapping segments, by a window function (Hamming, Hanning, etc.)
2) **FFT:** The FFT is computed for each segment.
3) **Averaging:** The squared magnitudes of each FFT for each segment are averaged, resulting in an estimate of the PSD. Averaging the spectra of these shorter segments reduces the variance of the PSD estimate.
4) **Normalization:** The average is then normalized by the length of the data segment and the power of the window function to produce the PSD.

In summary, for $x[n]$ signal, the PSD $P(f)$ via the Welch method is expressed as:

$$P(f) = \frac{1}{K} \sum_{k=1}^{K} |X_k(f)|^2 \quad (3.8)$$

Where K is the number of segments, $X_k(f)$ is the FFT of the k^{th} segment.

3.2.2.2. Burg's Method

Burg's method is a technique that estimates the PSD by fitting an Autoregressive (AR) model to the observed data and estimating the PSD based on the fitted model [36]. The basic idea is to find an AR model that best represents the data in the least squares sense. It is efficient in terms of computational complexity and does not require the data to be windowed, which avoids the data loss that occurs in Welch's method due to windowing. However, its limitations are that a) assumes that the data can be well-modeled by an AR process, which may not be the case for all types of signals (however EEG usually tend to be), and b) The model order needs to be chosen wisely. An incorrect order can lead to either an oversimplified or overfitted model, both of which can produce inaccurate spectral estimates.

3.2.2.3. Multitaper Method

A taper is a window function that modifies a signal by scaling its amplitude at various points across its length. Multitaper method applies multiple orthogonal tapers known as Slepian sequences or Discrete Prolate Spheriodal Sequences (DPSS) to the time-domain signal and then, similarly to the Welch method, it applies the FT across each tapered signal and averages the spectral estimates to generate the final PSD. It is optimized for reducing spectral leakage and bias and provides a more reliable estimate, especially when the signal-to-noise ratio is low, or when the data have non-stationarities.

When comparing with the Welch method, both the Welch and Multitaper methods begin by segmenting the data into time-segments. However, what distinguishes the Multitaper method is its use of multiple tapers (window functions) on each of those time-segments. The Multitaper method adds an additional layer of complexity by using multiple tapers for each time-segment. This is designed to produce a more reliable PSD estimate, particularly when the data may contain spectral leakage or other distortions.

So, in summary, let $x(t)$ the signal that will be analyzed with the Multitaper method. It is divided into N overlapping or non-overlapping segments, each of length T. Thus, for each segment the data $x_n(t)$ is given by:

$$x_n(t) = x(t + nT), for\ t = 0,1, \dots, T - 1 \tag{3.9}$$

We apply K different orthogonal tapers $w_k(t), k = 1,2, \dots, K$ to the n-th segment. The tapered signals $y_{n,k}(t) = x_n(t) * w_k(t)$, for each segment the FFT is computed:

$$Y_{n,k}(f) = F[y_{n,k}(t)] \tag{3.10}$$

And the PSD $S(f)$ is estimated as the average of the FFTs over all tapers and segments:

$$S(f) = \frac{1}{NK} \sum_{n=1}^{N} \sum_{k=1}^{K} \left|Y_{n,k}(f)\right|^2 \tag{3.11}$$

In summary, Welch's, Burg's, and Multitaper methods are all used to estimate the Power Spectral Density (PSD) of EEG data, each employing distinct approaches. Welch's approach is commonly used for stationary signals due to its simplicity and resilience, making it a dependable choice. Its limited frequency resolution may not be ideal for recording quick changes in non-stationary signals. Burg's technique is most appropriate for applications that necessitate a high-resolution Power Spectral Density (PSD) estimation. Its parametric character enhances efficiency but may not be suitable for all types of EEG signals. An autoregressive model assumption is necessary, which may not be applicable to all EEG applications. The Multitaper approach is particularly effective for analyzing non-stationary or noisy signals. Utilizing numerous tapers reduces spectrum leakage and provides improved management of frequency resolution. However, its complexity makes it computationally intensive. Thus, the selection of the appropriate methodology depends on the specific requirements of each specific EEG analysis in regards to these limitations.

3.2.3. Limitations of Frequency Domain analysis

Frequency Domain analysis, despite being widely used throughout EEG signal processing, has its own limitations. Specifically, frequency domain analysis is ideal when the stationarity assumption is correct, however this is not usually the case in EEG signals, since their characteristics tend to change throughout the recording. This temporal information loss that occurs when transforming the signal to the frequency domain is particularly problematic when studying ERP's or transient phenomena in EEG data.

Even when using windowing (epoching) techniques, there is a resolution trade-off between time resolution and frequency resolution. By using a larger window size the frequency resolution is increased but time resolution is decreased, and vice versa. This effect may not be important in stationary EEG such as sleep EEG or resting state EEG (for example when performing EEG recording with closed eyes to examine whether certain Alzheimer's disease characteristics may come out [12]), but it is significant when examining EEG's related to a certain activity.

Another significant disadvantage of frequency domain analysis is the lack of phase information. Traditional PSD estimates do not capture phase information, which could be critical for understanding the synchronization between different brain regions.

To address the problems of stationarity and temporal resolution, one can use time-frequency domain analytic approaches. These methods provide a comprehensive analysis of the signal by enabling the concurrent assessment of its temporal and frequency attributes. Techniques like the Short-Time Fourier Transform (STFT), Wavelet Transform, and Hilbert-Huang Transform (HHT) are highly useful for this purpose. They can capture dynamic and temporary characteristics in the data, which are important for comprehending intricate neurological disorders and cognitive processes. Time-frequency techniques are advantageous for analyzing EEG signals due to their non-stationary nature and the presence of crucial information in both time and frequency domains.

Moreover, to address the drawbacks of missing phase information, various features including Phase Locking Value (PLV), Coherence, Granger Causality, and Mutual Information are utilized to directly quantify the interactions among many signals or electrodes. Ultimately, frequency domain transformations offer important insights when employed within their acknowledged constraints.

3.3. Time-Frequency Domain Analysis

Time-frequency analysis approaches allow for the simultaneous recording of both temporal and spectral characteristics of an EEG signal. These methods aid in comprehending the intricate, ever-changing characteristics of EEG data, where certain frequency components may hold importance at distinct time periods. Frequency domain techniques such as FFT and PSD are limited because they do not account for time-localized variations in frequency components, which makes them less effective for evaluating EEG data that may

show quick changes in state or condition. Nevertheless, these methods have drawbacks, primarily due to their computational complexity.

3.3.1. Short-Time Fourier Transform

The Short-Time Fourier Transform (STFT) [37] is a widely used time-frequency analysis technique that is based on the Fourier Transform and provides a compromise between time and frequency resolution. It operates by dividing the signal into short and overlapping segments and computing the Fourier Transform (via FFT) of each segment. The STFT of a continuous-time signal $x(t)$ is given by:

$$\text{STFT}\ (x)(\tau, f) = \int_{-\infty}^{\infty} x(t) \cdot g^*(t-\tau) \cdot e^{-j2\pi f t} dt \tag{3.9}$$

Where τ is the time center of the window, f is the frequency and $g(t)$ is the window function.

Or if $g[k]$ an L-point window function and $x[k]$ the discrete valued signal then the STFT pair is given by:

$$\begin{cases} X_{\text{STFT}}\,[m,n] = \sum_{k=0}^{L-1} x[k]g[k-m]e^{-j2\pi nk/L} \\ x[k] = \sum_{m}\sum_{n} X_{\text{STFT}}\,[m,n]g[k-m]e^{j2\pi nk/L} \end{cases} \tag{3.10}$$

The STFT transform produces a 2-dimensional output where each point in time-frequency is associated with a complex number that represents the amplitude and the phase of the signal for a particular frequency and time. Thus, the output of a STFT is a spectrogram where The x-axis represents time, the y-axis represents frequency and the color or intensity at each point in the graph represents the amplitude of a particular frequency at a particular time.

3.3.2. Wavelet Transform

A wavelet is a mathematical function used to decompose a signal into different frequency components and then study each component with a resolution that matches its scale. Unlike sinusoidal functions like the sine and cosine used in Fourier transforms, wavelets can be localized in both time and frequency, making them handy for analyzing signals with non-stationary or time-varying characteristics. A wavelet is a waveform of effectively limited duration that has an average value of zero. All the wavelet functions are usually generated from a single prototype wavelet, called a "mother wavelet," by translations and dilations. Every wavelet should satisfy this condition in equation 3.11, that ensures that the wavelet has zero mean and is therefore capable of analyzing fluctuations around a baseline in a signal.

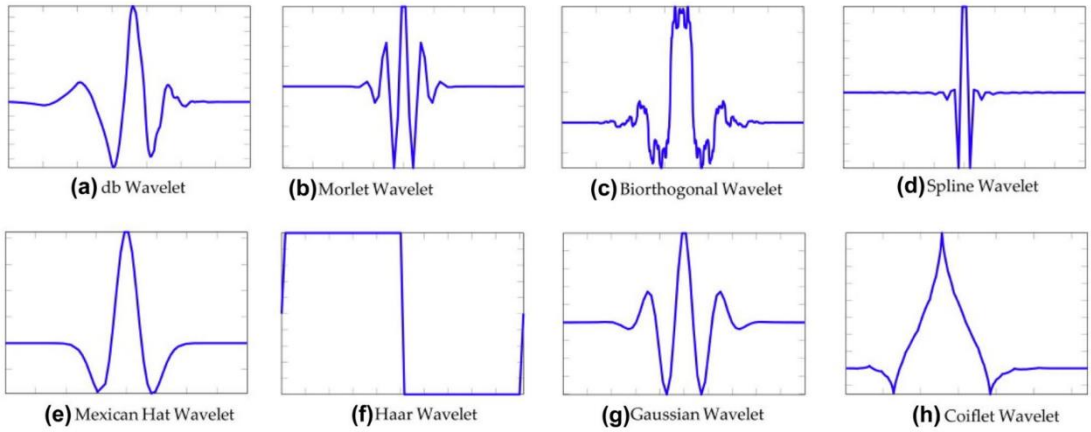


Figure 9 A selection of different wavelet functions. Adapted from "Efficient Deep Neural Networks for Classification of Alzheimer's Disease and Mild Cognitive Impairment from Scalp EEG Recordings", by Fouladi, 2022

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{3.11}$$

There are multiple different wavelet functions, each with its own set of characteristics that make it suitable for specific types of signal analysis [38]. For example, Haar wavelets are often used for their simplicity and computational efficiency, while Daubechies wavelets are preferred for their ability to capture higher-order details in signals. Morlet wavelets are commonly used in time-frequency analysis due to their sinusoidal shape, and Mexican Hat wavelets are favored for their localization properties in both time and frequency domains. The choice of wavelet function depends on the requirements of each task, such as the need for time resolution, frequency localization, or computational efficiency. Figure 9 A selection of different wavelet functions. Adapted from "Efficient Deep Neural Networks for Classification of Alzheimer's Disease and Mild Cognitive Impairment from Scalp EEG Recordings", by Fouladi, 2022 illustrates some of the most used wavelet functions.

The main limitation of STFT is the fixed time-frequency resolution where one must sacrifice time resolution to gain frequency resolution and vice versa. In contrast to STFT having equally spaced time-frequency localization, wavelet transform provides high frequency resolution at low frequencies and high time resolution at high frequencies. This adaptability makes Wavelet Transform a superior choice for analyzing non-stationary signals like EEG, where the frequency content may change over time.

Two primary methods for wavelet transformation are predominantly used: the Continuous Wavelet Transform (CWT) and the Discrete Wavelet Transform (DWT), each with its own set of advantages and limitations. While CWT provides a highly flexible and detailed analysis by employing continuous scaling and translation, it tends to be more

computationally intensive. On the other hand, DWT offers a more efficient and non-redundant multi-resolution analysis, particularly useful for digital signal processing tasks.

3.3.2.1. Continuous Wavelet Transform

The CWT of a function $x(t)$ is given by:

$$\mathrm{CWT}_x\,(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t)\psi^*\left(\frac{t-\tau}{s}\right) dt \tag{3.12}$$

Where τ is the translation parameter (a variable that shifts the wavelet function along the time axis), s is the scale parameter (a variable that scales the wavelet function) and $\psi(t)$ is the wavelet function, which must satisfy the admissibility condition (equation 3.13) which ensures that it has zero mean and is localized in both time and frequency domains. Commonly used wavelet functions in EEG analysis include the Morlet wavelet and the Haar wavelet.

$$C_\psi = \int \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \tag{3.13}$$

CWT offers high-frequency resolution for fast-changing activities and high time resolution for slower activities, but is very computationally expensive, especially for long EEG recordings. Thus, more often than not, its digital counterpart DWT is used for EEG analysis.

3.3.2.2. Discrete Wavelet Transform

The DWT serves as the digital counterpart of the Continuous Wavelet Transform (CWT) and is designed to overcome some of the computational complexities associated with the CWT, because it provides a sampled representation, reducing the amount of computational power and storage that is required. The DWT decomposes a signal $f(t)$ into a group of basis functions formed by scaling a mother wavelet $\psi(t)$ as described in equation 3.14

$$f(t) = \sum_k c_k \phi_{j,k}(t) + \sum_{j,k} d_{j,k}\psi_{j,k}(t) \tag{3.14}$$

c_k is the approximation coefficient and $d_{i,k}$ is the detail coefficient. $\varphi_{j,k}(t)$ and $\psi_{j,k}(t)$ are the scaled and shifted versions of the mother wavelet, respectively.

Figure 10 Time-Frequency distribution in a spectrogram of the DTF (a), the STFT (b) and the DWT (c) provides a comparison of how the spectrogram of the different analysis

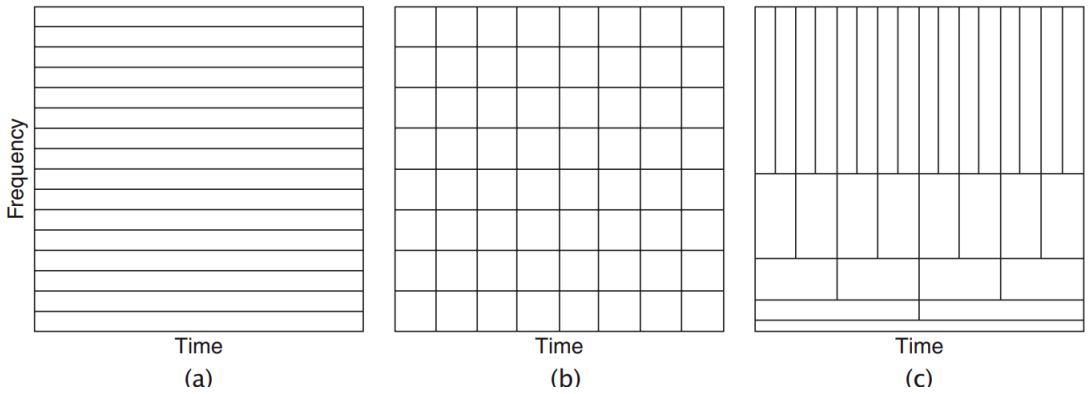


Figure 10 Time-Frequency distribution in a spectrogram of the DTF (a), the STFT (b) and the DWT (c)

methodologies are distributed across time and frequency, in regard to their resolution.

3.3.3. Hilbert-Huang Transform

Hilbert-Huang Transform (HHT)[39] is a time-frequency alternative that offers advantages when dealing with non-linear and non-stationary data at the cost of higher computational complexity and sensitivity to noise. It combines Empirical Mode Decomposition (EMD) with the Hilbert Spectral Analysis and was first proposed by Huang et al. for the analysis of hydrodynamic and meteorological data, but it is widely applied in the EEG analysis.

The HHT is consisted of 2 steps:

1) Empirical Mode Decomposition: Decompose the original signal into a finite set of Intristic Mode Functions through EMD, as analyzed in the next subchapter.
2) Hilbert Spectral Analysis: Once the IMFs are obtained, each IMF is then transformed using the Hilbert transform to obtain its instantaneous frequency, thereby providing a time-frequency representation of the signal. The Hilbert transform H of a given IMF $c_i(t)$ is defined as

$$H[c_i(t)] = \frac{1}{\pi} P \cdot V \cdot \int \frac{c_i(\tau)}{t-\tau} d\tau \tag{3.15}$$

3.3.3.1. Empirical Mode Decomposition

EMD [40] is a data-driven, adaptive method for decomposing a signal into a sum of oscillatory components called Intrinsic Mode Functions (IMFs). It does not employ predefined basis functions but rather adapts to the signal's characteristics, thus being useful for analyzing non-linear signals. Figure 11 EMD of an EEG signal of the FC3 electrode locationrepresents an EMD of an EEG signal

In simple terms, an IMF is a smooth oscillatory mode that captures a specific characteristic or frequency component of the original signal. An IMF is a function that satisfies these two conditions:

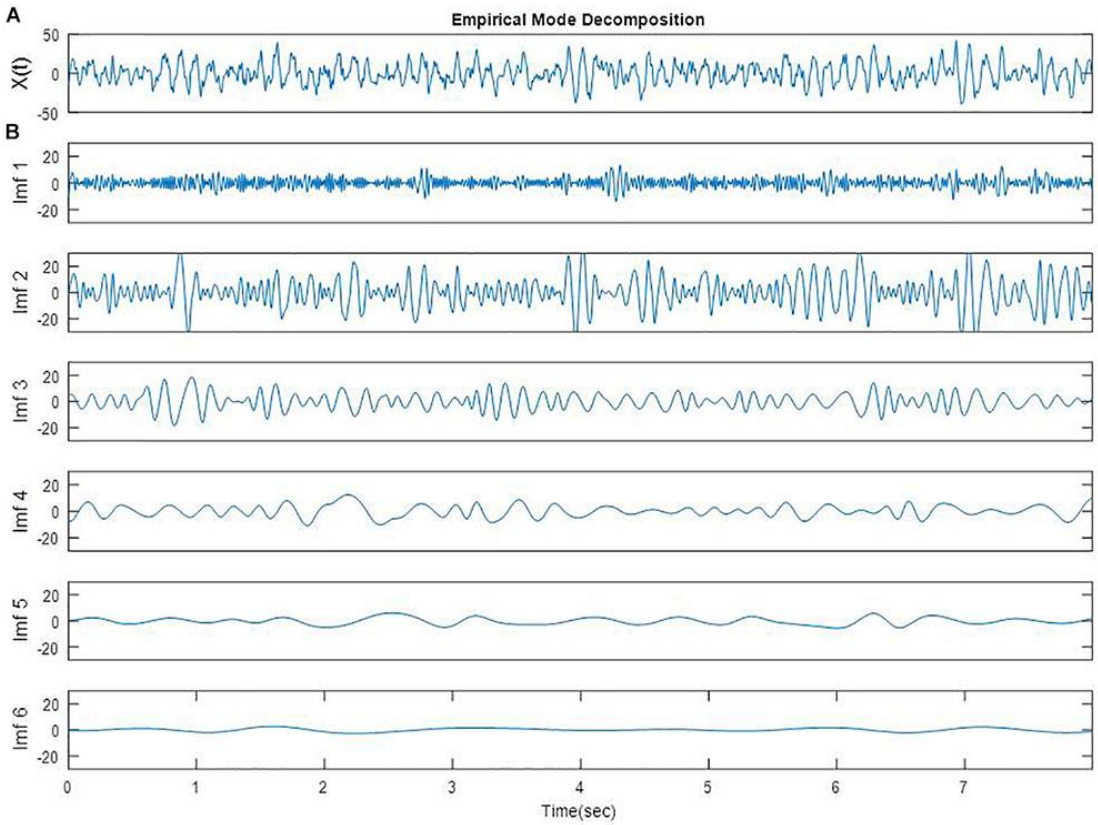
1. The number of extrema (maxima and minima) and the number of zero crossings must either be equal or differ at most by one.
2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

Given a 1D signal $x(t)$, EMD decomposes it into N IMFs $c_n(t)$ and a residue $r(t)$ such that:

$$x(t) = \sum_{n=1}^{N} c_n(t) + r(t) \tag{3.16}$$

The algorithmic steps for EMD are:

1) Extract IMF by identifying local maxima and minima of $x(t)$ and interpolate between them to create upper and lower envelops $e_{max}(t)$ and $e_{min}(t)$
2) Compute the mean $m(t) = \frac{e_{max}(t)+e_{min}(t)}{2}$
3) Extract the candidate IMF $c(t) = x(t) - m(t)$
4) Check IMF conditions. If $c(t)$ does not satisfies the IMF conditions return to step 1
5) Subtract IMF
6) Repeat. If the residue still contains information, go back to step 1.



3.3.4. Wigner-Ville Distribution

The Wigner-Ville Distribution (WVD) is a time-frequency representation method that provides a high-resolution analysis of non-stationary signals. Unlike the Short-Time Fourier Transform (STFT) and Wavelet Transform (WT), which offer fixed or multi-scale time-frequency resolutions, WVD aims to provide the best possible resolution in both time and frequency domains simultaneously.

The WVD $W_x(t,f)$ of a signal $x(t)$ is defined as:

$$W_x(t,f) = \int_{-\infty}^{\infty} x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau \qquad (3.17)$$

Where $x^*(t)$ is the complex conjugate of $x(t)$, τ is the time lag, and f is the frequency.

3.3.5. S-Transform

The S-Transform, also known as the Stockwell Transform, is a time-frequency representation that combines the best attributes of both the Short-Time Fourier Transform (STFT) and the Wavelet Transform (WT). Born by R. G. Stockwell in 1996 (just like me), the S-Transform provides a frequency-dependent resolution of the time-frequency space, offering a more flexible and intuitive representation of non-stationary signals.

Figure 11 EMD of an EEG signal of the FC3 electrode location

The S-Transform $S_x(t,f)$ of a signal $x(t)$ is defined as:

$$S_x(t,f) = \int_{-\infty}^{\infty} x(\tau) \text{ window } (t-\tau,f) e^{-j2\pi f\tau} d\tau \quad (3.18)$$

Here, window $(t-\tau,f)$ is a frequency-dependent windowing function, often a Gaussian window modulated by a sinusoid.

Unlike STFT, which has a fixed time-frequency resolution, the S-Transform allows for a resolution that varies with frequency. Moreover, it maintains phase information, which is beneficial for various applications, including EEG analysis. Compared to the Wigner-Ville Distribution (WVD), the S-Transform has fewer cross-terms, making it more suitable for multi-component signal analysis. The S-transform produces a time-frequency representation that is easy to interpret, especially in comparison to wavelet-based methods. However, it can be computationally intensive due to the double integration involved in its calculation.

3.3.6. Limitations of Time-Frequency Analysis

Time-Frequency methodologies offer the most comprehensive view of EEG signals across both time and frequency domain and are being increasingly popular across the researchers, especially in situations where the experimental protocol is not static. However, these methodologies do not come without limitations. Except from the individual limitations of each specific methodology discussed, all of them suffer, more or less, from computational complexity issues, making them unavailable or partially unavailable for real-time monitoring and online applications. Furthermore, they usually have a great number of parameters to be calibrated, and the effectiveness of the analysis severely depends on the choice of the parameters, thus they require a lot of studying to be used properly. Most of them are sensitive to noise and signal, thus appropriate preprocessing is required beforehand. Last but not least, a fundamental limitation in time-frequency analysis is the Heisenberg Uncertainty Principle, which states that it's not possible to simultaneously localize a signal in both time and frequency with arbitrary precision.

Heisenberg Uncertainty Principle

The Heisenberg Uncertainty Principle is a fundamental concept originating from quantum mechanics, but it has analogs in signal processing, particularly in the time-frequency domain. In the context of signal analysis, the principle states that it is impossible to know both the exact time and frequency content of a signal with arbitrary precision. In other words, the more accurately you know the frequency content of a signal, the less accurately you can know its time localization, and vice versa.

Mathematically, the time-frequency uncertainty principle is often expressed as:

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi} \quad (3.19)$$

Where Δt is the uncertainty in time and Δf is the uncertainty in frequency. This equation shows that the product of the uncertainties is bounded from below, indicating that you can never make both of them arbitrarily small.

3.4. Characteristics extracted from EEG

EEG signal is difficult to read and hard to interpret. Especially in its time domain representation, there are only a few cases in which particular waveforms may be easily readable and useful for creating insights, and these cases usually are limited to epileptic waveform analysis (in which spikes are easily observable). In the most cases, either for medical research, or biofeedback applications, the frequency domain representation provide a more comprehensive and explainable structure of the EEG signal.

When discussing machine learning applications on EEG signals, extracting features (characteristics) from time domain, frequency domain, time-frequency domain or other signal representation is crucial for the implementation of an automatic diagnosis pipeline. Although, the recent developments of Recurrent Neural Networks (RNN) and other NN configurations have allowed raw signal to be given as input in a classifier, still the feature extraction step remains a crucial part of these pipelines.

The selection of the features to be extracted are usually dependent to the specific problem that needs to be issued. For example, to create a methodology for Alzheimer's detection, a literature inspection could help the researcher notice that there is evidence of alterations in the theta/alpha and theta/beta ratio in Alzheimer diseased patients [41], thus these could be some indicative features to be extracted. In the following paragraphs, different characteristics that are extracted from the EEG signals are briefly mentioned.

3.4.1. Statistical Features

The simplest features to be extracted from EEG signals are the statistical features, which are primarily obtained from the time-domain signal. They provide a summary of the central tendency, dispersion and other statistical properties of EEG data. The most common statistical features include the mean (the average voltage of the EEG waveform), the standard deviation, the time of zero-crossings, the median and others. Furthermore, skewness and kurtosis are also widely employed.

Skewness measures the asymmetry of a distribution around its mean. A skewness value of 0 indicates a perfectly symmetrical distribution. A positive skewness value implies that the distribution has a longer or fatter tail on the right side of the distribution, indicating that most values are concentrated on the left side of the mean. A negative skewness value implies the opposite: a longer or fatter tail on the left side, indicating that most values are concentrated on the right side of the mean. Mathematically, skewness γ_1 is calculated as:

$$\gamma_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3 \quad (3.20)$$

Kurtosis measures the "tailedness" of a distribution, which can be useful for identifying outliers or extreme values in EEG signals: A kurtosis value of 0 indicates a distribution with tails similar to a normal distribution. A positive kurtosis value implies a distribution with heavier tails, indicating a higher likelihood of outliers. A negative kurtosis value implies a distribution with lighter tails, indicating a lower likelihood of outliers.

Mathematically, kurtosis γ_2 is calculated as:

$$\gamma_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 - 3 \quad (3.21)$$

Other statistical features that are usually extracted are:

1. **Root Mean Square** (RMS) which is defined as the square root of the arithmetic mean of the squares of the values.

$$\text{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \quad (3.22)$$

2. **Number of zero crossings**.
3. **Entropy features**: (these will be explained in the complexity features section)
4. **Autocorrelation**.
5. **Higher Order Moments**: Beyond skewness (3rd order) and kurtosis (4th order), higher-order moments can also be calculated to describe more complex characteristics of the distribution.

3.4.2. Frequency Features

Frequency based features in EEG analysis are those features that are derived from the transformation of EEG time-series into their frequency components and are useful for analyzing oscillatory activities and rhythmic patterns found in EEG data. These features are extracted after a FFT or a PSD transform, with the most commonly used one being the Relative Band Power or Absolute Band Power, calculated as:

$$\text{Band Power} = \int_{f_1}^{f_2} \text{PSD}\,(f)df \quad (3.23)$$

Where f_1 and f_2 being the lower and upper boundary of the frequency band of interest.

Besides band power, which usually refers to the average power of the signal within a specific frequency band over a given period of time, a related metric is band energy which refers to the total energy of the signal in that frequency band over a finite time interval. Band Energy is calculated as:

$$\text{Band Energy } = \int_{t_1}^{t_2} |x(t)|^2 dt \tag{3.24}$$

Other frequency domain metrics that are usually extracted are:

Peak frequency: The frequency at which the maximum power occurs.

Spectral Edge Frequency: This is the frequency below which a certain percentage (e.g., 95%) of the total power in the spectrum lies and it is used for summarizing the distribution of power over frequencies.

Cross-Spectral Density and Coherence: It measures the relationship between different EEG channels in the frequency domain. It can be also examined as a synchronization feature.

Frequency Ratios: As mentioned before, it is often useful to measure specific ratios of power in different frequency bands, such as theta/beta ratio.

Spectral Entropy: It measures the complexity of the frequency content of a certain band and is used to evaluate the randomness in the frequency domain. It is the equivalent of entropy in the time domain and will be further explained in the Complexity Features section.

3.4.3. Complexity Features

Complexity Features are used to quantify the irregularity of signals. These measures capture different aspects of the data than traditional time and frequency domain features, and they may be particularly useful for understanding more complex physiological states or conditions. Such measures are the Entropy-based and Fractal dimension measures.

3.4.3.1. Entropy-based measures

There are several entropy-based measures. All of them have a common goal: to quantify the uncertainty or complexity of a signal. They do so by evaluating the distribution of values in different domains—time, frequency, or phase space—to capture varying aspects of signal complexity. These measures are particularly useful in EEG analysis for identifying irregularities or changes in brain activity, which can be indicative of different mental states or neurological conditions. The most common and the most noteworthy are mentioned below:

Shannon Entropy: Shannon entropy [42] quantifies the amount of information contained in a signal and is the foundational measure of entropy in information theory. It is calculated as:

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x) \tag{3.25}$$

Where $p(x)$ is the probability of occurrence of the value x in the variable X. It is important to note that in the context of EEG signals, Shannon entropy is often used to quantify the complexity or unpredictability of the signal. Generally, for continuous signals like EEG, Shannon entropy can't be directly calculated as it is for discrete random variables. However,

the signal can be discretized into bins, essentially creating a histogram, to make it possible to calculate a form of Shannon entropy. Thus, the choice of number of bins can influence the resulting entropy value. The same concept applies for all entropy measures.

Spectral Entropy: Spectral entropy measures the complexity of a signal in the frequency domain, effectively capturing how the power is distributed across various frequency bands. It is calculated like Shannon entropy but in the frequency domain (usually after an FFT)

$$S = -\sum_{f} P(f)\log_2 P(f) \tag{3.26}$$

Where $P(f)$ is the normalized power of a frequency bin.

Sample Entropy: It is an extension to the **Approximate Entropy** and assesses the likelihood that runs of patterns that are close to each other will remain close in the next incremental comparisons. A lower Sample Entropy value indicates more self-similarity or regularity in the data, whereas a higher value indicates a more complex signal. Also, a widely used modification of sample entropy is **Fuzzy Entropy**.

Permutation Entropy: Permutation entropy [43] is a non-parametric complexity measure that considers the order relations between values of a time series. It quantifies the amount of order or disorder within a given time series. $p(\pi)$ represents the frequency of occurrence of each unique permutation π.

$$H = -\sum_{\pi} p(\pi)\log p(\pi) \tag{3.27}$$

Dispersion Entropy: Dispersion entropy [44] quantifies the irregularity or complexity of a time-series signal by mapping it into a sequence of dispersion patterns. Each pattern represents the relative ordering of amplitude values within a short segment of the time-series. $p(x_i)$ is the probability of encountering each unique dispersion pattern x_i.

$$\text{DispEn} = -\sum_{i=1}^{N} p(x_i)\log p(x_i) \tag{3.28}$$

3.4.3.2. Fractal-Dimension

Fractal Dimension (FD) [45] is a measure to characterize the complexity of a signal and quantifies how a metric changes with the scale at which it is measured. A higher FD indicates a more complex and irregular signal, and a lower FD indicates regular or periodic activity. The term "fractal" was coined by Benoit Mandelbrot in 1975 to describe objects that are self-similar at different scales. A fractal dimension is a statistical quantity that provides an index describing how the number of scaled measuring units changes with the scale factor. For a straight line, the fractal dimension is 1; for a plane, it is 2; and so on. There are multiple methods for calculating the fractal dimension of an EEG signal that are used in the literature, with some of the most common being briefly described below.

Higuchi's Fractal Dimension

$$F = \lim_{k\to\infty} \frac{\log(\text{ mean length } L(k))}{\log(1/k)} \tag{3.29}$$

Where $L(k)$ is the length of the curve for a given segment of length k

Katz's Fractal Dimension

$$D = \log_{10}(n)/(\log_{10}(d/L) + \log_{10}(n)) \tag{3.30}$$

Where n is the total number of points, d is the diameter, and L is the total length of the curve.

Detrended Fluctuation Analysis

This methodology [46] involves calculating the root mean square fluctuation of the integrated and detrended time series and evaluating its scaling behavior. The first step is to integrate the time series $y(t)$ as:

$$Y(k) = \sum_{i=1}^{k} [y(i) - \text{mean}(y)], k = 1,2,\dots,N \tag{3.31}$$

Where N is the length of the time series. The integrated time series $Y(k)$ is then divided into N_s non-overlapping segments of equal length n. In each segment ν, a least squares fit is performed to the data to obtain a local trend $y_\nu(k)$. The variance $F^2(\nu, n)$ of the detrended time series is then computed as:

$$F^2(\nu, n) = \frac{1}{n}\sum_{k=1}^{n} [Y((\nu-1)\cdot n + k) - y_\nu(k)]^2 \tag{3.32}$$

Then the detrended fluctuation $F(n)$ is calculated as:

$$F(n) = \sqrt{\frac{1}{2N_s}\sum_{\nu=1}^{2N_s} F^2(\nu, n)} \tag{3.33}$$

Renyi Dimension

$$D_q = \frac{1}{1-q}\lim_{r\to 0}\log\left(\sum_i p_i(r)^q\right)/\log(r) \tag{3.34}$$

Where $p_i(r)$ are the probabilities of points falling within a radius r.

3.4.4. Aperiodic components – 1/f slope

A fascinating concept in terms of signal analysis is the 1/f slope or the 1/f noise, also known as pink noise or flicker noise, and occurs in many complex systems from electronic components and financial markets to musical compositions and even the distribution of prime numbers. Unlike white noise, where the power spectral density (PSD) is constant

across frequencies, in 1/f noise the power decreases as the frequency increases, typically in a logarithmic manner. This results in a power-law relationship, expressed as:

$$\text{PSD} \propto \frac{1}{f^{\alpha}} \tag{3.35}$$

Regarding EEG, the PSD of EEG signals often follows a 1/f distribution. The slope of the 1/f distribution is quantified and is used to study various cognitive states or neurological conditions, meaning that deviations from the typical 1/f slope can be indicative of abnormal neural activity. The most common method to analyze the 1/f characteristics of EEG is through a tool called FOOOF (Fitting Oscillations and One Over F) [47] which separates the oscillatory components from the 1/f background noise in the power spectrum.

3.5. Synchronization Features

Synchronization features aim to quantify the degree of coordination of different channels (or brain areas) over time. These features also provide spatial information regarding the EEG. Such features are coherence, phase synchronization, cross-correlation, granger causality, mutual information and others.

Coherence: It is a frequency-domain measure that quantifies the linear relationship between two signals as a function of frequency. It can range from 0 (no coherence) to 1 (perfect coherence).

$$\text{Coherence } = \frac{\left|P_{xy}\right|^2}{P_{xx} \times P_{yy}} \tag{3.36}$$

Where P_{xy} is the cross-spectral density between signals x and y.

Phase Synchronization: Such methods look for the phase relationship between 2 signals. A common feature named Phase Locking Value (PLV) is calculated as:

$$\text{PLV} = \left| \frac{1}{N} \sum_{n=1}^{N} e^{i(\phi_{x,n} - \phi_{y,n})} \right| \tag{3.37}$$

Where $\phi_{x,n}$ is the phase of the signal x at time n.

Cross Correlation: This time-domain measure gives an indication of the similarity between two signals at different time lags.

$$R(\tau) = \frac{1}{T} \int_0^T x(t) y(t+\tau) dt \tag{3.38}$$

Granger Causality: Granger causality is a statistical measure used to determine if one time series can predict another time series.

Mutual Information: Mutual information measures the amount of information that knowing the state of one variable provides about the state of another variable. It is a non-linear measure of dependency.

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \tag{3.39}$$

Chapter 4

Machine Learning used in EEG analysis.

"I am so clever that sometimes I don't understand a single word of what I am saying."

-Oscar Wilde

4.1. Introduction and Importance

Machine Learning methodologies for automatic diagnosis of neurological conditions were first developed during the 1970s, focusing on developing algorithms for feature extraction and pattern recognition for epilepsy. In fact, epilepsy, the most commonly diagnosed neurological disorder by EEG was the initial focus of automated methodologies. This is due to the fact that, despite the versatility of EEG, traditional analysis methods had the need of manual feature extraction, time-costly visual inspection and limitations regarding the ability to recognize brain activity patterns.

In the last decades, automatic methodologies have been continuously developed for a wide range of medical and research contexts, from diagnosing neurological disorders to studying cognitive processes. These methodologies allowed a revolution on how EEG data is analyzed and re-established it as a formidable diagnostic tool. By incorporating machine learning in the medical procedures, screening of patients becomes a lot easier; there is no need for a manual inspector over every multi-hour recording, anymore.

4.1.1. Problems that are addressed

A wide range of problems can be addressed through the usage of EEG and machine learning. The most common one is epilepsy, but other neurological, cognitive, physiological or pathological conditions have also been studied. Clinical applications in this concept include, but are not limited to: epilepsy diagnosis and prediction, dementia detection, BCI for motor disabilities, sleep disorder diagnosis, concussion and traumatic brain injury assessment, Attention-Deficit/Hyperactivity Disorder assessment and anesthesia monitoring. Other non-clinical applications in this concept include but are not limited to: cognitive load and stress measurement, emotion recognition, neural correlates of learning and memory, auditory and visual perception, brain aging, drug response prediction and real-time neurofeedback. Regarding the research that has taken place during this PhD, particular focus has been given to Alzheimer's disease prediction, stress assessment and epilepsy detection.

4.2. Machine Learning Pipeline

There are some standard steps that comprise the standard of an EEG based machine learning methodology. More often than not, these steps are included and each of them contributes to the overall efficacy of the model. A general outline of how a typical machine learning EEG pipeline functions is given in Figure 12 A general pipeline of an EEG based machine learning classification study

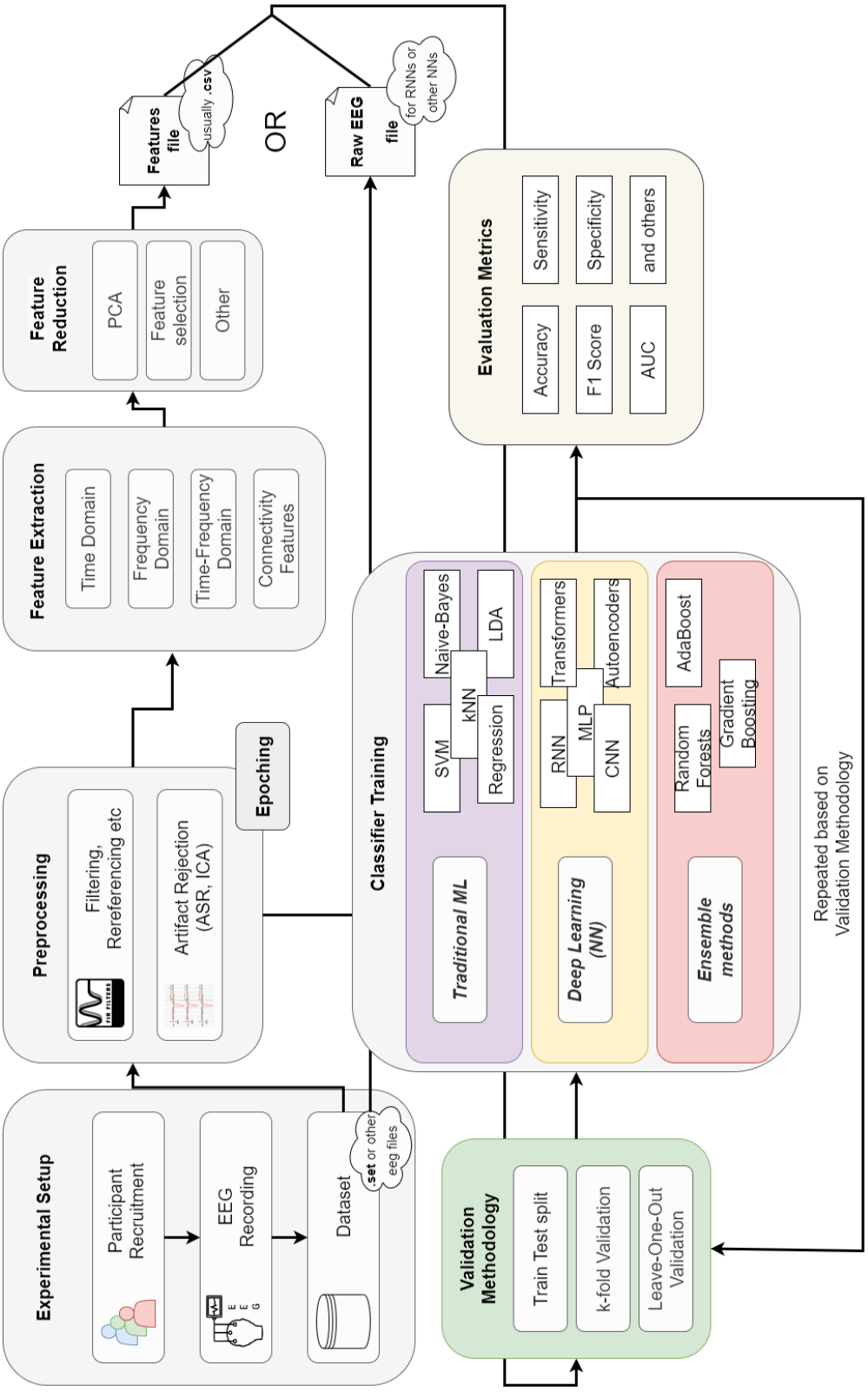


Figure 12 A general pipeline of an EEG based machine learning classification study

Generally, after an experimental protocol is created and recordings have been obtained, these recordings are preprocessed in order to eliminate any noise or not needed frequency

components. Methodologies such as ICA and ASR may also be used. Then, a feature vector may or may not be created, after epoching the data. In traditional methodologies a feature vector is consisted of 2-dimensional data, where each row is an individual observation and each column is a feature. However, deep learning architectures allow the observations being multi-dimensional data (rather than rows). The next step is feature reduction, in which only the most important features are kept. Finally, one or more algorithms are trained using a de-facto validation methodology and the performance metrics are reported.

4.2.1. Preprocessing & Feature Extraction

Previously, a detailed analysis of EEG preprocessing, as well as a description of the features that can be extracted from the time, frequency, time-frequency and non-linear domain has taken place in Chapter 3. When creating a machine learning methodology, these techniques are employed in order to transform the EEG data and obtain useful information that can be used to train a classifier. However, the number of participants is more often than not, limiting the capabilities of the methodology. Hundreds or thousands of participants would be required to train a classifier, if we considered each participant as a single instance of data, or in other words as a single row of the feature table. Thus, a technique called **epoching** is usually employed, which splits the recording of one participant into multiple overlapping or non-overlapping segments of fixed duration, called epochs. These epochs are then considered as individual instances and each epoch represents a row in the feature table. It is important to note that, when using epoched data, attention should be given to the validation methodology that is employed, as discussed in a next paragraph.

4.2.2. Feature Reduction

In most machine learning scenarios; not only EEG, feature reduction is a crucial procedure in order to ensure high classifier performance. Feature Reduction is considered any technique that reduces the number of dimensions of the dataset and it is important for several considerations such as:

Curse of Dimensionality

High-dimensional data can be problematic because as the number of dimensions increases, the volume of the space increases exponentially. In high dimensions, points that may be considered "close" in low dimensions end up very far apart, making the data sparse. This sparsity is problematic for any method that requires statistical significance. Algorithms that suffer from the curse of dimensionality are traditional algorithms such as k-NN, SVM, linear regression etc. Algorithms that are not so affected from the curse of dimensionality are the ensemble methods such as Random Forests and Gradient Boosting. Also, Naïve Bayes is often surprisingly effective in high-dimensional spaces, especially when the number of categories (classes) is also high.

Overfitting

When the number of features is close or exceeds the number of observations, the learning algorithm can fit the training set very closely but may fail to generalize to new, unseen data. Feature reduction techniques can help in avoiding overfitting by simplifying the model.

Computational complexity

Reducing the number of features can significantly speed up the training process, making it computationally more efficient. This is particularly relevant in real-time applications or when computational resources are limited.

Interpretability

A model with fewer features is easier to understand and interpret. This is crucial in medical applications like EEG analysis, where understanding the model's decision-making process can be as important as the decision itself. However, it should be noted that, when employing dimensionality reduction techniques like Principal Component Analysis (PCA) the interpretability is lost, anyways.

Noise Reduction

Feature reduction can also serve to remove irrelevant or redundant features, which can be considered as 'noise' that may hamper the learning process.

Easier Data Visualization

Reducing the dimensionality of the data can make it easier to visualize, which can be helpful for understanding the structure of the data and the relationships between features.

Improved Accuracy

In some cases, removing irrelevant or redundant features can actually improve the model's performance. Some algorithms are sensitive to such irrelevant input features, which can cause poor performance and make the model harder to optimize.

4.2.2.1. Principal Component Analysis

Principal Component Analysis (PCA) is a technique used for dimensionality reduction and is commonly used in machine learning applications to transform the feature vector into a lower-dimensional space, in order to avoid the drawbacks of high-dimensionality that were mentioned above. The main objective of this methodology is to transform the feature vector in components based on the maximization of the variance in the data. This way, the first component retains the maximum variance in the data, and the last component the least. Then, the used can discard as many components as they like from the last, to reduce the dimensionality.

So, if X the feature matrix of shape $n \times d$, where n is the number of samples and d the number of features, the first step of PCA is to center the data by subtracting the mean of each feature from the data points:

$$X_{\text{centered}} = X - \mu \tag{4.1}$$

And then, the covariance matrix Σ of the centered data is computed

$$\Sigma = \frac{1}{n} X_{\text{centered}}^{T} X_{\text{centered}} \tag{4.2}$$

Then, the eigenvalues and eigenvectors of Σ are computed. The eigenvectors corresponding to the largest eigenvalues are the principal components that capture most of the variance in the data. The original data can then be projected onto these principal components to produce a reduced set of features: $X_{\text{PCA}} = X_{\text{centered}} \times W$. Then, features of X_{PCA} are discarded based on a variance threshold.

4.2.2.2. Feature Selection

Feature Selection methods can be categorized into filter-based methods and wrapper-based methods. Their difference is how they evaluate the importance of the features. Wrapper based methods depend on the predictive model that we are training and the feature selection is based on the algorithm's performance. In the other hand, filter-based methods are algorithm independent and evaluate the features based on statistical characteristics. Wrapper methods are more computationally expensive than filter methods, however they can yield better performance for a specific classifier.

Some of well-established wrapper-based methods for feature selection are the Recursive Feature Elimination, that starts with all features and recursively removes the least important ones and the genetic algorithms that use evolutionary techniques to select features. Moreover, in the context of classification algorithms that consider feature importance (e.g., Random Forest Importance), this feature importance can serve as a criterion for feature selection.

Some of the well-established filter-based methods are the Chi-Squared test, the Fisher score, the correlation coefficient and others. The Chi-Squared (x^2) test measures the dependence between the feature and the output by assessing if there is a significant association between these variables in a contingency table, where the chi-squared statistic is calculated as:

$$\chi^2 = \sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \tag{4.3}$$

Where O_{ij} is the observed frequency of in cell (i,j) in the contingency table and E_{ij} is the expected frequency in this cell, calculated based on the total frequencies and the row and column totals in the contingency table as:

$$E_{ij} = \frac{(\text{ Total for Row } i) \times (\text{ Total for Column } j)}{\text{Grand Total}} \quad (4.4)$$

Next, the Fisher score is used to evaluate the discriminatory power of a feature with respect to the class labels. It calculates the Between-Class Variance (BCV) and the Within-Class Variance (WCV) for a feature and then finding the ratio of BCV/WCV

$$BCV(j) = \sum_{i=1}^{k} n_i(\mu_{ij} - \mu_j)^2 \quad (4.5)$$

$$WCV(j) = \sum_{i=1}^{k} \sum_{x \in C_i} (x_j - \mu_{ij})^2 \quad (4.6)$$

Where k is the number of classes, n the number of samples, $\mu_{i,j}$ the mean of feature j in class i and C_i the set of samples in class i.

Finally, the correlation coefficient can be measured in a variety of ways, depending if the goal is to measure linear relationships (Pearson correlation) or non-parametric relationships (Spearman correlation). Pearson correlation, denoted as r, measures the linear association between two variables. $r = 1$ indicates a perfect positive linear relationship, and $r = -1$ indicates a perfect negative linear relationship. It is calculated as:

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}} \quad (4.7)$$

Where X, Y the individual data points of two variables and $\bar{X}, \bar{Y}$ the means of them. On the other hand, Spearman correlation calculates the rank of both variables independently by arranging them from the smallest to the largest and then calculates the correlation as:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)} \quad (4.8)$$

Where d is the difference between the ranks of the values in the two variables and n the number of data points.

4.2.3. Training, Testing & Validation methodologies

A machine learning algorithm is trained to correctly categorize unknown signals into classes. To validate the classifier's performance, a validation methodology is employed, classifying data with known labels. The classifier's performance is then estimated based on the correctly classified samples. In this section three common validation methodologies will be described, the train-test split, the k-fold validation and the Leave-One-Subject-Out (LOSO) validation.

The train-test split is the simplest and computationally cheap methodology. It involves dividing the dataset into a train set (for example the 80% of the dataset) and a test set (the other 20%). The problem of this methodology is that it may lead to high variance in the evaluation metrics, as the test set's performance heavily depends on the specific data points

chosen for testing. Also, it is not suitable for too small datasets, because a significant portion of the data should be used for testing.

The alternative solution that alleviates these issues is the k-fold validation methodology. This methodology iteratively uses a portion of the data as test set and the rest as training set. For example, k-fold validation with k=5 uses 20% of the data as test set, for 5 times. This way, every data point will eventually be used in the testing phase, and the average performance metrics are reported. The advantage of this method is that it provides a more robust estimate of a model's performance, especially when working with limited datasets.

However, in the case of epoched signals, such as EEG data, using k-fold validation can result in the same subject's epochs appearing in both the test set and the training set. This can lead to overly optimistic performance results, as the individual characteristics of the person's signal may influence the classifier training. To alleviate this issue, LOSO validation is employed. This method ensures that all the epochs of one subject are included in the test set while all the other epochs are in the training set. This is performed iteratively for each subject, until everyone has been used as test set. In cases of too many subjects, LOSO validation can be modified to use more than one subject at the same time as test set. In this doctoral research, every study that involved epoched data employed the LOSO validation.

4.2.4. Evaluation Metrics

To assess the performance of a classifier, different well-established evaluation metrics are extracted during the test phase. The assessment of machine learning models is crucial for understanding their performance and applicability. Various metrics have been established to evaluate different aspects of these models. All of these metrics make use of the confusion matrix, which is a matrix presenting the number of instances that were correctly or incorrectly classified, as shown in Figure 13. For example, let's take a diseased or not diseased classification task. A True Positive (TP) is an instance that is diseased and was classified as diseased. A False Positive (FP) is an instance that is healthy but was incorrectly classified as diseased. A True Negative (TN) is a healthy instance that was correctly

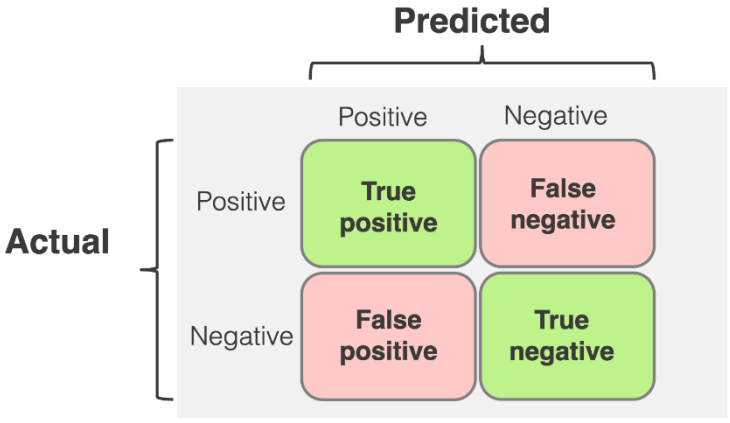


Figure 13 A confusion matrix for a 2-class problem.

classified as healthy and finally a False Positive (FP) instance is a healthy instance that was incorrectly classified as diseased. The key metrics that are employed are described here.

(a) Accuracy

Accuracy is one of the most intuitive performance measures. It is simply the ratio of correctly predicted observations to the total observations.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

(b) Sensitivity

Sensitivity measures the proportion of actual positives correctly identified. It is crucial for models where missing positive cases (e.g., diseases in medical diagnostics) are costly.

$$Sensitivity = \frac{TP}{TP + FN}$$

(c) Specificity

Specificity measures the proportion of actual negatives correctly identified and is important in ensuring the model is not raising false alarms.

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

(d) Precision

Precision measures the proportion of positive identifications that were actually correct. It is a critical metric when the cost of a false positive is high.

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

(e) F1 score

The F1 Score is the harmonic mean of Precision and Sensitivity. It is particularly useful when the class distribution is imbalanced.

$$F1 = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity}$$

(f) Area Under Curve (AUC)

AUC refers to the area under the ROC (Receiver Operating Characteristic) curve. It is used to measure the ability of a classifier to distinguish between classes and is useful for

imbalanced datasets. The ROC curve is a plot of Sensitivity against (1-Specificity) at various threshold settings. AUC ranges from 0 to 1, with higher values indicating better classification performance.

4.3. Machine Learning

Machine Learning algorithms are proving very useful in EEG research due to their ability to process massive amounts of data and make the screening process easier for the medical practitioners. An empirical categorization of the machine learning algorithms could be that of traditional, ensemble or deep learning methods.

4.3.1. Traditional Classifiers

"Traditional classifiers" in machine learning are a set of methods that are fundamental to the field and have been widely utilized and researched for many years. These classifiers are frequently compared to more sophisticated or modern methods, such deep learning and neural networks. Conventional classifiers are recognized for their straightforwardness, clarity, and effectiveness in managing organized data. They are especially preferred in situations with constrained computational resources or where a precise comprehension of the decision-making process is necessary. Traditional classifiers include:

4.3.1.1. Decision Trees

Decision Trees is a widely used and effective algorithm in machine learning and data mining for both classification and regression applications. They are recognized for their clarity and straightforwardness. A decision tree illustrates decisions and their potential outcomes, encompassing chance event results, resource expenses, and utility. A decision tree is a structured diagram where internal nodes represent tests on attributes, branches represent test outcomes, and leaf nodes represent class labels determined after evaluating all attributes. The topmost node in a tree is the root node. A leaf node represents a classification or decision. The paths from root to leaf represent classification rules.

At each internal node, the tree splits based on a decision rule. This rule is derived from one of the features that best separate the classes. Different algorithms use different metrics for this, such as Gini impurity, Information Gain, or Chi-square.

To avoid overfitting, techniques such as pruning are used. This is done by setting a maximum depth for the tree or removing sections of the tree that don’t meet a certain threshold on the decision rule.

Although they are easy to understand and interpret, can handle both numerical and categorical data and requires little data preprocessing, they have severe limitations such as being prone to overfitting, especially with noisy data, or create biased trees if some classes dominate.

4.3.1.2. Naïve Bayes

Naive Bayes classifiers are a set of effective probabilistic classifiers that utilize Bayes' theorem while assuming substantial independence between features. They are renowned for their efficacy and proficiency in handling extensive datasets and are commonly utilized in diverse applications like spam filtering, text classification, and medical diagnostics. Naive Bayes is based on Bayes' theorem, which calculates the likelihood of an event given prior knowledge of associated conditions. Mathematically, it is expressed as:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)} \tag{4.9}$$

Where:

- $P(A)$ is the prior probability of class
- $P(B)$ is the prior probability of predictor
- $P(A \mid B)$ is the posterior probability of class A given predictor B
- $P(B \mid A)$ is the likelihood which is the probability of predictor given class

Naive Bayes classifiers assume that the effect of a particular feature in a class is independent of other features. This assumption is called class conditional independence. There are multiple types, including Gaussian (for normally distributed data), Multinomial (for text data where we consider word frequencies), and Bernoulli (for binary/boolean features). During training, the model calculates the probability of each class and the conditional probability of each feature given each class. Prediction involves applying Bayes' theorem to compute the conditional probability for each class and selecting the class with the highest probability.

Naïve Bayes is not widely used because the strong feature independence assumption is often a downfall as it rarely holds true in real-world scenarios, and also suffers from the "zero frequency" pitfall in which if a categorical variable has a category in the test dataset, which was not observed in the training dataset, the model will assign it a 0 probability and will be unable to make a prediction. However, it can outperform more complex models and is particularly useful in text classification tasks like spam filtering, sentiment analysis, and document categorization.

4.3.1.3. Logistic Regression

Logistic Regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. Despite its name, logistic regression is used for binary classification problems rather than regression problems. It predicts the probability that a given instance belongs to a particular class. Logistic regression is typically used for a binary outcome (e.g., pass/fail, win/lose, healthy/sick). The outcome is modeled with a binary variable where 1 represents the presence (success) and 0 represents the absence (failure) of a characteristic or outcome. The core of logistic regression is the

sigmoid function, which maps any real-valued number into a value between 0 and 1, making it suitable for modeling a probability. Sigmoid function is defined as

$$\sigma(z) = \frac{1}{1 + e^{-z}}, where\ z = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n,\ \ b\ the\ coefficients \tag{4.10}$$

Logistic regression estimates the probability that a given input point belongs to a certain class. If the estimated probability is greater than 0.5, the model predicts that the instance belongs to the positive class (labeled as '1'); otherwise, it predicts that it belongs to the negative class (labeled as '0').

The disadvantage of logistic regression is that it assumes a linear relationship between the independent variables and the logarithm of odds, is not suitable for complex relationships in data and can be vulnerable to overfitting if there are too many features.

4.3.1.4. Support Vector Machines (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm primarily used for classification, but it can also be employed for regression tasks. SVM is well-known for its ability to handle high-dimensional data and its effectiveness in cases where the number of dimensions exceeds the number of samples. It's particularly popular in applications such as image recognition, text categorization, and bioinformatics. The core idea of SVM is to find a hyperplane in an N-dimensional space (N being the number of features) that distinctly classifies the data points. The best choice is the hyperplane that represents the largest separation, or margin, between two classes.

Mathematically, SVM tries to solve an optimization problem to find this hyperplane. The equation of a hyperplane can be written as:

$$w \cdot x + b = 0 \tag{4.11}$$

where:

- w is the weight vector
- x represents the input features
- b is the bias term

The decision function for classification is:

$$f(\boldsymbol{x}) = \text{sign}\ (\boldsymbol{w} \cdot \boldsymbol{x} + \boldsymbol{b}) \tag{4.12}$$

In SVMs, we solve the following optimization problem:

Minimize $\frac{1}{2} \parallel \boldsymbol{w} \parallel^2$, subject to $y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq 1 \forall i$, y_i the class labels

In non-linear SVM, kernel functions are used to transform the input space into a higher-dimensional space. Some common kernels include:

- Linear kernel, equivalent to no kernel
- Polynomial kernel, $K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d$
- Radial Basis Function (RBF) or Gaussian Kernel, $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2\right)$, γ is a parameter that defines the spread of the Gaussian distribution
- Sigmoid kernel, $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(\gamma \boldsymbol{x}_i \cdot \boldsymbol{x}_j + r)$

When there is a kernel, the decision function becomes:

$$f(\boldsymbol{x}) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}) + b\right) \tag{4.13}$$

4.3.1.5. k-Nearest Neighbors

k-Nearest Neighbors (kNN) is a simple and intuitive, algorithm used in both classification and regression in machine learning. Unlike many learning algorithms, kNN does not involve a model training phase; rather, it makes predictions using the entire dataset in a way that relies on data proximity. In kNN, an input point is classified based on the majority class of its 'k' nearest neighbors. The algorithm calculates the distance between the input sample and every training instance, selects the nearest 'k' instances, and performs a majority vote for classification.

The kNN algorithm calculates the distance between data points to identify the nearest neighbors. Common distance metrics include the Euclidean distance, the Manhattan distance or the Minkowski distance.

The selection of the number of neighbors (k) is crucial, since a small k means higher sensitivity to noise while a big k may cause leaking from wrong classes.

4.3.1.6. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a well-established statistical approach used both for classification and dimensionality reduction. It is particularly useful when the data exhibits gaussian-like distributions. LDA seeks to find a linear combination of features that best separate two or more classes of objects or events. LDA operates by projecting the data onto a lower-dimensional space with good class separability in order to avoid overfitting ("curse of dimensionality"). The core idea is to find a linear combination of variables that best separates the classes. LDA computes the directions ("linear discriminants") that represent the axes that maximize the separation between multiple classes.

One of its core limitations is that it assumes that the variables are normally distributed and assumes that there is equal covariance matrices for the different classes.

4.3.2. Ensemble Classifiers

When discussing ensemble classifiers, we refer to architectures that use multiple weak classifiers to create a well-educated prediction. The idea behind ensemble methods is that by combining the predictions from multiple models, you can often produce better results than any single model alone. Most ensemble classifiers are tree based, but this is not always the case, since it is possible to combine conceptually different machine learning classifiers (for example Random Forests, SVM and Neural Network, as presented in [48]). The most common ensemble classifiers are described below.

4.3.2.1. Random Forests

Random Forests is the most common ensemble methodology. It operates by constructing a multitude of decision trees during training and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random Forests combines the simplicity of decision trees with flexibility, resulting in a vast improvement in accuracy.

Random Forests build upon the concept of bagging (bootstrap aggregating) and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. The key conceptual characteristics of the Random Forest are:

Bagging: Each tree in a Random Forest is built from a sample drawn with replacement (i.e., a bootstrap sample) from the training set. This means that some samples will be used multiple times in a single tree.

Feature Randomness: When splitting a node during the construction of the tree, the split that is chosen is no longer the best split among all features. Instead, the split that is picked is the best split among a random subset of the features. This results in diversity that ensures the individual trees to capture various aspects of the data.

Random Forests has excellent predictive performance and can handle large datasets and higher dimensionality. However, it can become slow in large datasets and is less effective on very sparce data.

4.3.2.2. Extra Trees

Extra Trees, short for Extremely Randomized Trees, is an ensemble learning method, particularly used for classification and regression tasks. It is similar to Random Forests but introduces more randomness in the way the trees in the ensemble are constructed. The fundamental idea behind Extra Trees is to use the entire training dataset to build each tree and introduce randomness in the choice of the split points at each node of the tree, which adds extra diversity to the model. This randomness differentiates it from other tree-based ensemble methods, like Random Forests.

Unlike Random Forests, which uses a subset of data (bootstrap sample), Extra Trees uses the entire dataset to build each tree. At each node, instead of searching for the most discriminative thresholds (as in classical decision trees), thresholds are drawn at random for each candidate feature, and the best of these randomly generated thresholds is picked as the splitting rule. This randomness in selecting the thresholds can be mathematically represented, but it is more conceptual than formulaic. It's about selecting split points randomly and then choosing the best split among these.

Their advantages lie on the reduced variance and robustness against overfitting, their fast training (in comparison with the Random Forests) and their limitations are the increased bias due to the randomness factor and the complexity of the model (which is common to all ensemble methods)

4.3.2.3. Gradient Boosting

Gradient boosting builds on the principle of boosting, where weak learners (typically decision trees) are sequentially added to correct the errors made by previous learners, with each new learner focusing on the errors of its predecessor. This method has gained immense popularity due to its effectiveness in a wide range of applications and competitions, with implementations like LightGBM [49] and XGBoost [50] being particularly renowned.

Gradient Boosting involves three key elements: a loss function to be optimized, a weak learner to make predictions, and an additive model to add weak learners to minimize the loss function. Regarding the additive model, it is based on the concept of trees being added one at a time, and existing trees in the model are not changed. Each new tree helps in correcting errors made by the previously trained trees. The model's predictions are made by summing the predictions of these weak learners.

At each stage n, a tree $h_n(x)$ is fit on the negative gradient of the loss function $L(y, F(x))$, where y is the true value and $F(x)$ the prediction. The update formula is:

$$F_{n+1}(x) = F_n(x) + \alpha_n h_n(x) \tag{4.14}$$

Where α_n is the learning rate and $h_n(x)$ the weak learner (individual tree).

4.3.3. Deep Learning

Deep Learning is a subset of machine learning that has revolutionized the field of artificial intelligence by providing a more complex and sophisticated approach to data analysis. At its core, deep learning involves the use of neural networks with several hidden layers (hence 'deep'), to model complex and non-linear patterns and relationships in the data. Deep learning methodologies have been increasingly used during the last couple of decades due to the advancements in computational power and particularly in GPUs, making use of their ability to model such networks. It excels in handling unstructured data such as images, text, and audio. The multilayer structure of deep neural networks enables them to capture hierarchical patterns within data, making deep learning particularly adept at tasks like image

and speech recognition, natural language processing, and even playing complex games like Go and chess.

Deep learning algorithms, designed in the likeness of the human brain, consist of neural networks comprising numerous interconnected artificial neurons. Just as the human brain is made up of millions of neurons that collaborate to learn and interpret information, deep learning neural networks employ layers of these artificial neurons to perform tasks within a computer. These artificial neurons, essentially software modules known as nodes, execute mathematical computations to handle data. Forming the core of deep learning algorithms, these networks of nodes leverage their collective computational power to tackle intricate problems.

Deep Learning is used for more than classification and regression tasks. Other applications include Natural Language Processing, image and video recognition, Generative Models, Anomaly Detection, Speech recognition and generation, Recommendation systems, Autonomous systems and more. Some of the most worth-mentioning deep learning architectures are:

4.3.3.1. Feed Forward Neural Networks (FFNs)

Them being the most common neural network architecture, they are present in almost every other architecture. Feedforward Neural Networks, also known as Multilayer Perceptrons (MLPs), are the simplest type of artificial neural network architecture. In these networks, information moves in only one direction—forward—from the input nodes, through the hidden nodes (if any), and finally to the output nodes. There are no cycles or loops in the network, hence the name "feedforward."

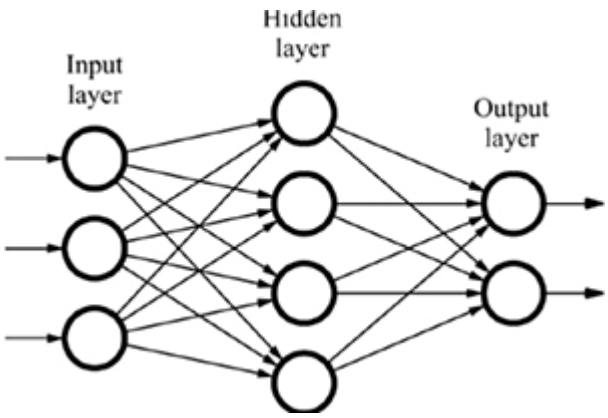


Figure 14 A typical Feed Forward Neural Network architecture

These networks are structured in layers, as seen in Figure 14. Each layer is constructed of neurons which are simple computational units that perform individual calculations. These neurons take multiple input values, apply a set of weights to them (which are learned during the training process), and then pass the weighted sum through an activation function. The output of this activation function, which introduces non-linearity into the model, serves as

the input for the next layer in the network. This process allows the network to learn complex patterns and relationships within the input data.

The input layers consist of that many input neurons as the number of features in the data. Then each hidden layer consists of one or more neurons which perform mathematical transformations on the data, and finally the output layer produces the final output of the data. For a binary class classification problem, one neuron in the output layer is enough. For a multiclass problem, usually the number of neurons in the output layer of a neural network should match the number of classes we are trying to predict. Each neuron in the output layer corresponds to a specific class, and the network is trained to output a probability distribution across these classes for each input.

The operation of a neuron can be mathematically described as:

$y = f(w \cdot x + b)$, y the output, f the activation function, w the vector of weights, x the input vector, b the bias (4.15)

The activation function determines the output of a node given a set of inputs, by calculating a weighted sum of the inputs and adding bias. The most common types of activation functions are:

Sigmoid: $\sigma(\boldsymbol{x}) = \frac{1}{1+e^{-(\boldsymbol{w}\cdot\boldsymbol{x}+b)}}$ (4.16)

ReLU: $\text{ReLU}\ (\boldsymbol{x}) = max(0, \boldsymbol{w}\cdot\boldsymbol{x}+\boldsymbol{b})$ (4.17)

Softmax: $\text{Softmax}\ (\boldsymbol{x})_i = \frac{e^{(w_i\cdot x+b_i)}}{\sum_j e^{(w_j\cdot x+b_j)}}$ (4.18)

Tanh: $\text{Tanh}\ (x) = \frac{e^{(w\cdot x+b)}-e^{-(w\cdot x+b)}}{e^{(w-x+b)}+e^{-(w\cdot x+b)}}$

Each activation function has its own benefits and is better suited for different data or problems.

4.3.3.2. Training Neural Networks: Backpropagation

Training a neural network involves adjusting its weights and biases to minimize the difference between the predicted output and the actual output (the error). This is achieved through an algorithm known as backpropagation. The steps for the training are:

1. **Initialize weights and biases:** Assign random initial values to the weights and biases of all neurons in the network.
2. **Forward pass for each training instance:** For each input in the training dataset, perform the following steps:
    i. Calculate the weighted sum of the inputs for each neuron.

   ii. Add the bias to the weighted sum.
   iii. Apply the activation function to the result.
3. **Calculate the loss:** Evaluate the performance of the model using a loss function that measures the difference between the network's prediction and the actual target values.
4. **Backpropagation:** Calculate the gradient of the loss function with respect to each weight and bias in the network. This involves:
   i. Compute the derivative of the loss function with respect to the output of the network.
   ii. Use the chain rule to recursively calculate the gradient for each layer, moving from the output layer back to the input layer.
   iii. Update the gradients of the weights and biases for each layer based on the derivative calculations.
5. **Update Weights and Biases:** Adjust the weights and biases in the opposite direction of the gradient to minimize the loss. This is typically done using an optimization algorithm like Gradient Descent. The updates are usually scaled by a learning rate.

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \cdot \nabla_{\boldsymbol{w}} L$$
$$b \leftarrow b - \eta \cdot \nabla_{b} L$$

6. **Repeat:** Repeat steps 2-5 for a predefined number of iterations or until the loss converges to a minimum.
7. **Evaluate the Model:** After training is complete, evaluate the performance of the model on a separate validation dataset not used during training to check for overfitting.
8. **Hyperparameter Tuning:** Based on the performance on the validation set, you may need to go back and tweak the hyperparameters like the learning rate, number of epochs, or even the architecture of the network (number of layers, number of neurons per layer, etc.).
9. **Final Testing:** Once satisfied with the model's performance on the validation set, perform a final test on a test set to ensure the model generalizes well to new, unseen data.

4.3.3.3. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a specialized type of neural network, primarily used for processing data with a grid-like topology, such as images. CNNs have been instrumental in the field of computer vision, setting new performance benchmarks in tasks like image classification, object detection, and facial recognition.

A number of powerful Convolutional architectures have been deployed, pretrained and are openly available for fine-tuning and individual usage. Such architectures are the AlexNet, the ResNet and others [51].

CNNs usually consist of 3 types of layers, which are: convolutional layer, pooling layer and fully connected layer. These layers are described below.

(a) Convolutional layer

The convolutional layer serves as the foundational element of a CNN and is where the bulk of the calculations are executed. It utilizes several key components: the input data, a filter, and the resultant feature map. Let's consider our input as a color image comprised of a 3D matrix of pixels, signifying the image's height, width, and depth dimensions, corresponding to the RGB channels.

Within this structure operates the feature detector, also termed as a kernel or filter. This entity traverses the image's receptive fields to identify the presence of specific features, a process known as convolution.

Functionally, the feature detector is a matrix of weights, effectively a slice of the image, conventionally fashioned in a 3x3 grid, which dictates the receptive field's size. The filter aligns with a segment of the image and computes a dot product between the filter's weights and the input pixels. This computed value populates an output matrix. The filter then proceeds, moving by a certain stride to repeat the operation, sliding across the full expanse of the image. The collection of these dot products, derived from the interactions between the input and the filter, constitutes what is termed a feature map, activation map, or convolved feature.

It's important to note that as the filter scans the image, its weights remain constant—a principle known as parameter sharing. While these weights are fine-tuned during training via backpropagation and gradient descent, there are also three pivotal hyperparameters that influence the output's volume size, and these must be pre-set prior to the neural network's training phase. These hyperparameters include:

- The **number of filters**, affecting the depth of the output
- The **stride**, being the distance or pixels that the kernel moves over the input matrix.
- The **padding**, which dictates how the filters will behave when they do not fit the input image. There are 4 types of paddings: **Zero-padding, Valid-padding, Same-padding** and **Full-padding**.

Multiple convolution layers may follow the initial convolution layer. As the convolutional process unfolds, the architecture of the CNN inherently develops a hierarchical nature, wherein the subsequent layers gain visibility into the pixels encompassed by the receptive fields of the preceding layers.

(b) Pooling layer

Pooling layers conduct dimensionality reduction, reducing the number of parameters in the input. It works similarly to the convolution layer by performing filter sweeps across the input, however it does not contains any weights, but instead it applies an aggregation function. There are two types of pooling:

In max pooling, the filter scans the input and transmits the largest pixel value in its view to the output array, a method more commonly utilized than average pooling. In contrast, average pooling computes and passes the mean pixel value within the receptive field to the output array as the filter traverses the input.

(c) Fully connected layer

This is nothing more than a Feed-Forward neural network attached to the end of the Convolutional Neural Network which ultimately performs the classification.

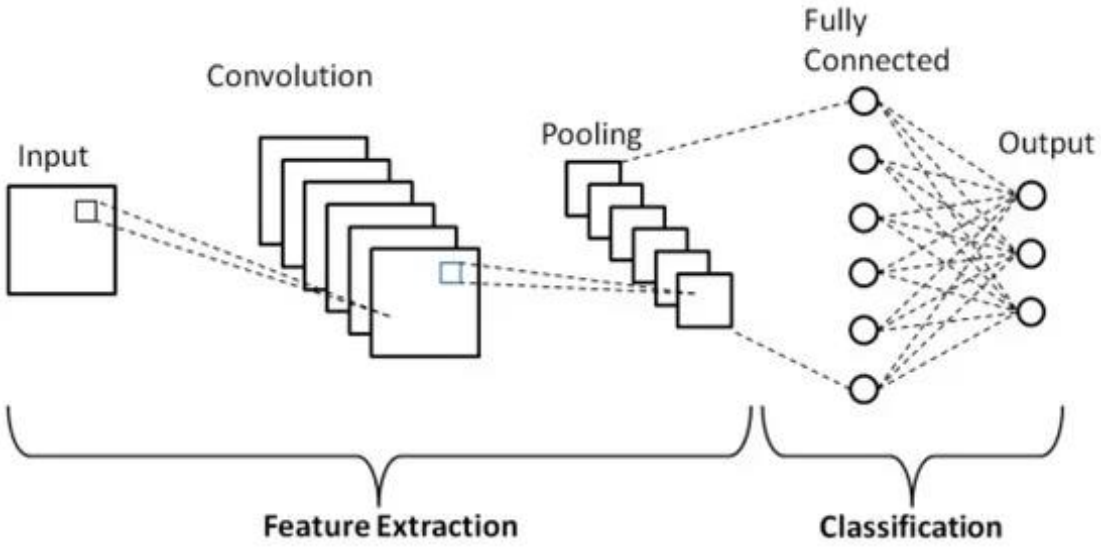


Figure 15 A Convolution Neural Network architecture.

4.3.3.4. Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) are a class of neural networks designed to handle sequential data. Unlike feedforward neural networks, RNNs possess a form of memory that allows them to process sequences of inputs. This makes them ideal for tasks where the temporal dimension is crucial, such as time-series prediction, natural language processing, and speech recognition. The key concept of the RNNs is their internal state. At each time step, the hidden layer output is influenced both by the current input and the previous hidden layer's output. An RNN can be mathematically described by:

Hidden state update:

$$\boldsymbol{h}_t = f(\boldsymbol{W}_h \cdot \boldsymbol{h}_{t-1} + \boldsymbol{W}_x \cdot \boldsymbol{x}_t + \boldsymbol{b}_h) \tag{4.19}$$

Output:

$$\boldsymbol{y}_t = \boldsymbol{W}_y \cdot \boldsymbol{h}_t + \boldsymbol{b}_y \tag{4.20}$$

Where

- h_t the hidden state at time t

- x_t the input at time t
- W_h, W_x, W_y the weight matrixes
- $\boldsymbol{b}_h$ and $\boldsymbol{b}_y$ the biases.
- f a non-linear activation function, typically a ReLU
- $\boldsymbol{y}_t$ the output at time t

Training an RNN involves backpropagation through time. The basic limitations of the RNNs are the well-known **Vanishing Gradient Problem** and **Exploding Gradient Problem**. In the first one, gradients tend to vanish or explode as they propagate over many time steps. making it hard for RNNs to capture long-term dependencies. In the second one, gradients can also explode, leading to numerical instability.

4.3.3.5. Long Short-Term Memory Networks (LSTMs)

LSTMs are a subtype of RNNs that are specifically designed to overcome the limitations of them, particularly the aforementioned problems of Vanishing and Exploding gradient. Their key concept is that they use a unique memory cell which can maintain information for long periods and is regulated by structures called gates: the input gate, the output gate and the forget gate. These gates control the flow of information into and out of the cell, making LSTMs capable of learning which data in a sequence is important to keep or discard. Its architecture is described in Figure 16.

(a) Forget gate

It decides what information to discard from the cell state, by looking at h_{t-1} (the previous hidden state) and at x_t and outputs a number between 0 and 1 for each number in the cell state C_{t-1}.

$$\boldsymbol{f}_t = \sigma\big(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_f\big) \tag{4.21}$$

(b) Input gate

It decides what new information to store in the cell state, and also creates a vector of new candidate values $\widetilde{\boldsymbol{C}}_t$ that could be added to the state.

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i) \tag{4.22}$$

(c) Cell State update

$$\boldsymbol{C}_t = \boldsymbol{f}_t * \boldsymbol{C}_{t-1} + \boldsymbol{i}_t * \widetilde{\boldsymbol{C}}_t \tag{4.23}$$

(d) Output gate

It decides what the next hidden state should be. It is used for predictions and is also passed to the next time step.

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_o \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_o) \tag{4.24}$$

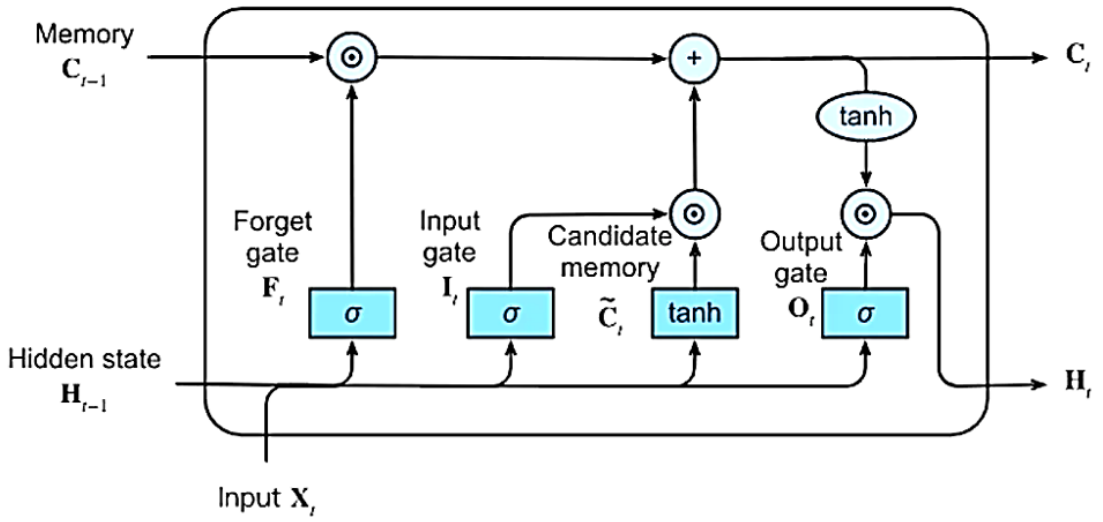


Figure 16 The Long Short-Term Memory Network Architecture

4.3.3.6. Transformers

Transformers is a revolutionary neural network architecture, first introduced in the conference paper "Attention Is All You Need"[52]. It was first developed in Natural Language Processing (NLP) and has outperformed conventional models such as RNNs and LSTMs in NLP tasks. Their key innovation is achieved by incorporating self-attention mechanisms, enabling models to assess the significance of various sections of the input data without being constrained by the sequential processing constraints of RNNs.

Transformers have expanded beyond NLP into other disciplines since they were introduced. They are becoming more commonly used in areas such as computer vision for tasks like image identification and segmentation, where their capability to manage global relationships in data is beneficial. Transformers are advancing in fields such as audio signal processing and genomics, where the sequential character of data is crucial. Transformers' ability to comprehend multiple tasks simultaneously and effectively manage complex relationships over great distances enhances their adaptability. They are versatile and can be used in a wide range of applications, going beyond their original use in NLP. Their increasing use in many areas demonstrates their adaptability and efficiency, establishing them as a major advancement in the field of deep learning. Transformer's architecture fundamentally relies on self-attention mechanisms and a unique encoder-decoder structure. A detailed analysis of the mathematical background of the transformers is skipped and can be found in paragraph 5.1.4.1(c).

4.3.4. Hyperparameter optimization

In deep learning, hyperparameter optimization is a critical process that involves fine-tuning the parameters that govern the training of neural network models. Unlike model parameters that are learned during training, hyperparameters are set before the learning process begins and have a significant impact on the performance of the model.

Optimizing hyperparameters is essential because it directly influences the behavior of the training algorithm and the performance of the model. Properly tuned hyperparameters can lead to faster convergence during training, improved model accuracy, and prevention of issues like overfitting or underfitting.

Key parameters that may be optimized in deep learning architectures are:

1. **Learning Rate:** Arguably the most crucial hyperparameter, it determines the step size at each iteration while moving toward a minimum of the loss function.
2. **Number of Epochs:** Refers to the number of times the entire training dataset is passed forward and backward through the neural network.
3. **Batch Size:** The number of training samples used in one forward/backward pass. It affects the speed and stability of the learning process.
4. **Network Architecture Parameters:** Including the number of layers, the number of neurons in each layer, and types of layers (e.g., convolutional, recurrent, etc.).
5. **Activation Functions:** Choices like ReLU, Sigmoid, or Tanh that affect the non-linear transformation applied to the input data.
6. **Regularization Techniques**: Such as dropout and L2 regularization, which help prevent overfitting.
7. **Optimizer:** Algorithms like SGD, Adam, RMSprop, each with their own strengths and weaknesses.

Some techniques for hyperparameter optimization are:

1. **Grid Search:** Systematically working through multiple combinations of parameter values. It's exhaustive and computationally expensive.
2. **Random Search:** Randomly selecting combinations of hyperparameter values. It's less computationally intensive but might miss optimal values.
3. **Bayesian Optimization:** Uses a probabilistic model to predict which hyperparameter values are most likely to yield good results.

However, it should be noted that hyperparameter optimization in deep learning architectures can be time-consuming and resource intensive, the optimal set of parameters may vary from a dataset to another, and also suboptimal training results may come from hyperparameter optimization pipelines due to hitting to local minima and plateaus.

4.4. Clinical Applications

Integrating machine learning into clinical practices, especially in EEG analysis, represents a major leap in medical technology. Machine learning is proving to be extremely effective in identifying neurological problems by analyzing complex, multi-dimensional EEG data. Deep learning models like Convolutional and Recurrent Neural Networks are highly proficient in identifying patterns in EEG signals. These models have been effectively utilized in crucial domains such as detecting and forecasting seizures, providing prompt therapies for patients with epilepsy. Moreover, they are crucial in sleep research by helping classify different sleep stages based on EEG measurements, thereby improving our comprehension of sleep disorders.

In addition to diagnostic uses, these methods are leading advancements in Brain-Computer Interfaces (BCIs), utilizing EEG signals to facilitate the operation of external devices based on cerebral activity. This innovation creates new opportunities for patient engagement and rehabilitation. Machine learning is used in the field of neurodegenerative illnesses such as Alzheimer's to identify early signs by analyzing small changes in EEG patterns.

The incorporation of deep learning in EEG analysis offers the potential for individualized treatment strategies and real-time analysis, particularly important in intensive care and surgical environments. Nevertheless, this groundbreaking intersection faces several difficulties. Ensuring the confidentiality and protection of sensitive medical information, as well as improving the clarity of deep learning models in clinical decision-making, are critical priorities. Moreover, the diversity in EEG data collection techniques necessitates a cohesive strategy to standardize data for wider clinical application. The changing role of machine learning in EEG analysis is paving the way for a future in neurological healthcare that is more precise, efficient, and personalized.

4.4.1. Challenges, Limitations and Problems

While machine learning offers great potential for EEG analysis, it also pauses several challenges and limitations. The obstacles encompass technical and computational limitations as well as medical-specific issues that must be resolved. Possible constraints include:

Absence of a Standard Protocol: A major obstacle in using machine learning with EEG data is the lack of a universally accepted procedure for collecting and analyzing the data. EEG data can change greatly because of variations in electrode location, recording technology, and patient conditions. This diversity hinders the development of models that can be applied universally across many environments and populations.

Computational needs: Machine learning models, particularly deep learning structures such as CNNs and RNNs, need significant computer resources throughout the training process. The demand for high processing power can be problematic, especially in clinical

environments where these resources may not be easily accessible. Moreover, handling substantial amounts of EEG data can be time-intensive, affecting the practicality of conducting real-time analysis.

Challenges related to the quality and accessibility of data: Acquiring extensive sets of top-notch, labeled EEG data for training machine learning models is difficult. Privacy and consent concerns in medical data add complexity to data gathering and sharing.

Challenges with interpretability and reliability: The opaque nature of numerous machine learning models is a considerable obstacle in healthcare settings. Medical professionals frequently need clear and understandable explanations of model conclusions, which can be challenging with complicated models such as deep neural networks.

Challenges with integrating into clinical workflows: Incorporating machine learning techniques into current therapeutic processes poses logistical difficulties. It is crucial for the successful adoption of these tools that they are user-friendly and cater to the needs and practices of medical practitioners

Chapter 5

Original Research

"My congratulations to you, sir. Your manuscript is both good and original; but the part that is good is not original, and the part that is original is not good."

-Samuel Johnson

In Chapter 5, the original research that has been conducted and published in peer reviewed journals and international conferences will be analyzed. Specifically, 8 studies have been published in journals and 2 in conferences. The research can be divided into three research domains, namely Alzheimer and Dementia Research, Epilepsy Research, and BCI Research. This chapter will be organized accordingly.

The papers, sorted by publication date are:

1) Alzheimer's disease and frontotemporal dementia: A robust classification method of EEG signals and a comparison of validation methods, published in Diagnostics MDPI, 2021.
2) Classification of EEG signals from young adults with dyslexia combining a Brain Computer Interface device and an Interactive Linguistic Software Tool, published in Biomedical Signal Processing and Control Elsevier, 2021.
3) An experimental protocol for exploration of stress in an immersive VR scenario with EEG, presented at SEEDA 2022 conference, held at Ioannina Greece.
4) Assessing Electroencephalography as a Stress Indicator: A VR High-Altitude Scenario Monitored through EEG and ECG, published in Sensors MDPI, 2022.
5) Evaluating the Window Size's Role in Automatic EEG Epilepsy Detection, published in Sensors MDPI, 2022.
6) Machine Learning Algorithms for Epilepsy Detection Based on Published EEG Databases: A Systematic Review, published in IEEE Access, 2022.
7) A Dataset of Scalp EEG Recordings of Alzheimer's Disease, Frontotemporal Dementia and Healthy Subjects from Routine EEG, published in Data MDPI, 2023, being the cover of the journal vol 8, issue 6, describing the dataset upload in OpenNeuro database with name: A dataset of 88 EEG recordings from: Alzheimer's disease, Frontotemporal dementia and Healthy subjects.
8) Enhanced Alzheimer's disease and Frontotemporal Dementia EEG Detection: Combining lightGBM Gradient Boosting with Complexity Features, presented at CBMS 2023 conference, held at L'Aquila Italy.
9) DICE-net: a novel convolution-transformer architecture for Alzheimer detection in EEG signals, published in IEEE Access, 2023.
10) An Ensemble Method for EEG-based Texture Discrimination during Open-eyes Active Touch, published in Engineering, Technology & Applied Science Research, 2024

5.1. Alzheimer's Disease and Dementia Research

Alzheimer's disease (AD) is a progressive neurodegenerative condition, ranking among the most prevalent forms of dementia affecting the elderly. It manifests primarily through

cognitive decline and behavioral alterations, and its incidence is estimated to escalate due to the increase of the population age. It is currently holding the sixth position in the leading causes of mortality in the United States, and notably, it is the only cause within the top 10 still experiencing a significant increase in prevalence. Over 50 million cases of dementia were reported in 2020, and it is estimated that the number of AD patients will reach 75 million by 2030 and 131 million by 2050. The prevalence of AD remains consistent between genders, with a rate of 1.4% for individuals aged 65-70, reaching up to 24% for those over 85. Symptomatically, AD's onset is characterized by difficulties in short-term memory recall, progressing into more complex challenges, including speech and orientation issues, mood fluctuations, self-care deficiencies, and behavioral changes. Ultimately, the disease leads to the deterioration of various bodily functions, culminating in the patient's demise. Currently, there does not exist a definitive cure for AD, and the available treatments merely offer limited relief from its symptoms.

For the diagnosis of probable AD, patients must meet specific clinical criteria, often compelling postmortem confirmation of distinctive neuropathological changes characterized by the accumulation of neuritic plaques and neurofibrillary tangles containing hyperphosphorylated tau proteins. However, emphasis has been given to early diagnosis and intervention due to the rising number of individuals grappling with dementia. To align with the urgency of timely AD diagnosis, biomarkers such as structural MRI, molecular Positron Emission Tomography (PET) neuroimaging, and cerebrospinal fluid analyses have been used in clinical practice. Nonetheless, these imaging techniques are resource-intensive and time-consuming, resulting in diagnoses that often occur after significant neurodegeneration has taken place. Hence, there is urgent need for methodologies that achieve prediction of future AD onset, which could expedite the identification of high-risk individuals and help with treatment planning. Thus, other cheaper and faster biomarkers such as EEG should be explored. As explained in Chapter 2.3.2 EEG can be a reliable and effective biomarker for the detection and screening of AD, if proper Machine Learning methodologies are employed.

5.1.1. Alzheimer's Disease and Frontotemporal Dementia: A Robust Classification Method of EEG Signals and a Comparison of Validation Methods

Andreas Miltiadous[1], Katerina D. Tzimourta[2], Nikolaos Giannakeas[1], Markos G. Tsipouras[2], Theodora Afrantou[3], Panagiotis Ioannidis[3], Alexandros T. Tzallas[1]

The first research [13] focusing on AD detection can be considered an explorative and preliminary research that laid the foundations for the future development of DICE-net. It was published in Diagnostics MDPI during August 2021. In this study, EEG recordings from 28 participants were used, that were obtained from the 2nd Department of Neurology of AHEPA General University Hospital of Thessaloniki. Ten of them were from AD patients, ten from FTD patients, and eight were from healthy age-matched adults that formed the control group (CN). The mean ages were 70.5, 67.5, and 68.5 for the AD, FTD, and CN groups, respectively. The Mini Mental State Examination (MMSE) score was also acquired, which is used to measure the cognitive decline and functional performance of dementia patients. Furthermore, the Clinical Dementia Rating (CDR) and the duration of the disease were also acquired. All these statistical characteristics are presented in Table 1.

Table 1. Statistical characteristics (values in brackets represent standard deviation) of the participants.

	Gender (Male/Female)	Age	MMSE	CDR	Disease Duration in Months
AD	6/4	70.5 (7.1)	19.7 (2.76)	1 (0.54)	24 (9.88)
FTD	6/4	67.5 (4.5)	21.5 (1.83)	0.75 (0.26)	26 (9.24)
CN	4/4	68.5 (7.2)	30 (0)	-	-

The author list ordered as published was: **Andreas Miltiadous**, Katerina D. Tzimourta, Nikolaos Giannakeas, Markos G. Tsipouras, Theodora Afrantou, Panagiotis Ioannidis, Alexandros T. Tzallas

The EEG device used for the recording of the signals was the Nihon Kohden EEG 2100, using 19 scalp electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2) and 2 reference electrodes on the subjects earlobes (A1, A2), placed according to the 10–20 international system. Impedance value below 5 kΩ was ensured before and throughout the duration of the experiments. The recordings were obtained according to the clinical protocol, which was in a sitting position with eyes-closed. The online filter setting was 0.5-60 Hz and the sampling rate was 500 Hz with a resolution of 10μV/mm. The mean duration of the EEG recordings for the dementia cases was 11-17 minutes (mean 13), while the duration of the control trials was 20-23 minutes (mean 21).

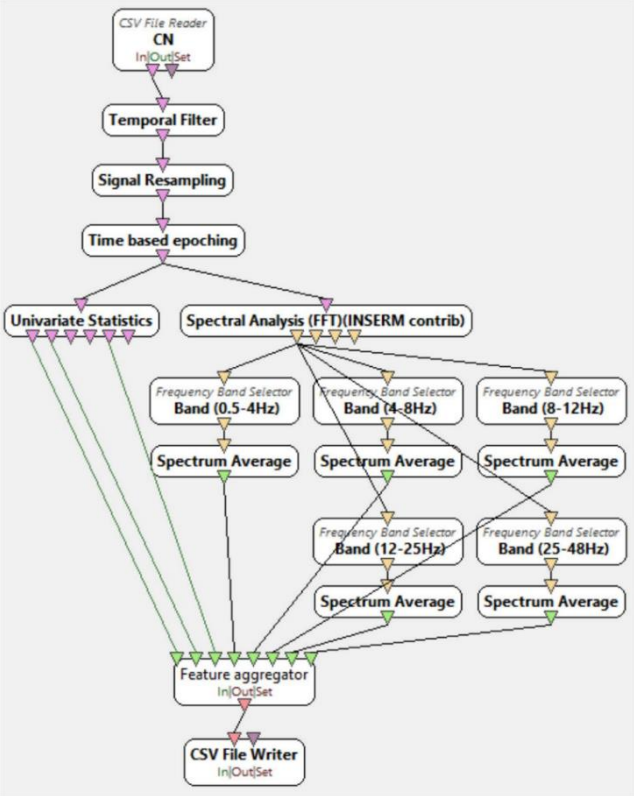


Figure 17 OpenViBE xml diagram of feature extraction

After manual removal of artifacts of severe electrode movement, signals were down-sampled from 500 Hz to 250 Hz and a Butterworth band-pass filter 0.5–48 Hz was applied, removing power line noise interference at 50 Hz. The next step was the epoching, in which the signal was time-windowed to partitions of 5 seconds with 2.5 second overlap. Finally, time domain metrics, namely mean, variance, and IQR and frequency domain metrics (after a PSD transform), namely the energies of the five EEG rhythms (delta, theta, alpha, beta, gamma) were extracted. For the manual preprocessing, the EDFBrowser was used. For the feature extraction, the OpenViBE [53] suite was used. The OpenViBE pipeline that was used for the feature extraction can be examined in Figure 17.

In this preliminary study, multiple classifiers were examined for their capability to classify the AD/CN, the FTD/CN and the AD/FTD problems, based on the extracted dataset

of features. Performance results were acquired using the k-fold validation method, k=10 (which, as discussed in paragraph 4.2.3 poses certain limitations due to the previous epoching of the data) and the LOSO validation (which alleviates these limitations). The performance results of the k-fold methodology are presented in Table 2, and the performance results of LOSO validation are presented in Table 3 and Table 4.

Table 2. Accuracy, Sensitivity and Specificity results of classification algorithms using the 10-fold validation method.

Accuracy of Classification Problem	**Decision Trees**	**Random Forests**	**ANN**	**SVM**	**Naïve Bayes**	**kNN**
AD/CN	96%	99.1%	95%	96.2%	80%	96%
FTD/CN	94.2%	98%	98%	97%	77%	97%
AD/FTD	93.8%	97.7%	90%	91%	69%	95%
Sensitivity of Classification Problem	**Decision Trees**	**Random Forests**	**ANN**	**SVM**	**Naïve Bayes**	**kNN**
AD/CN	96.6%	98.6%	96%	98%	94%	96%
FTD/CN	94.1%	98%	98.5%	97%	98%	98%
AD/FTD	95.6%	97.8%	91%	96%	80%	96%
Specifity of Classification Problem	**Decision Trees**	**Random Forests**	**ANN**	**SVM**	**Naïve Bayes**	**kNN**
AD/CN	95%	99%	94%	94.4%	58%	96%
FTD/CN	94.4%	98%	95%	97%	62%	99%
AD/FTD	91.3%	97.5%	89.1%	86%	54%	94%

Table 3. Accuracy results with the LOSO validation

Classification	AD-CN		FTD-CN		AD-FTD	
Algorithm	**MEAN**	**SD**	**MEAN**	**SD**	**MEAN**	**SD**
Decision Trees	78,50%	5,8	79,60%	11,2	73%	11
Random Forests	77,07%	7,1	86,30%	7,1	64%	12,6
ANN	73%	9,4	69,20%	14,1	61%	15,89
SVM	68,00%	11,1	75%	12,6	68%	18
Naïve Bayes	63%	14	73,80%	25	52%	21,3
kNN	60%	11,3	67,30%	9,8	51%	18,2

Table 4. Sensitivity and Specificity of Decision Trees and Random Forest with LOSO validation

	AD-CN		FTD-CN	
	Sensitivity	**Specifity**	**Sensitivity**	**Specifity**
Decision Trees	82,40%	74%	82,20%	77,45%
Random Forests	78,70%	76%	87%	83%

This study consists of two parts. Firstly, a robust method of classifying EEG signals of AD, FTD and CN participants with Decision Trees and Random Forests classification algorithms was presented and the results were validated using the most trustworthy leave-one-patient-out cross validation strategy. Secondly, the trustworthiness of k-fold cross validation method was validated (which is vastly used in previous EEG classification studies) comparing it to the LOSO validation method.

Previous studies have made observations regarding the comparison of k-fold validation and the LOSO validation on other medical engineering problems. For instance, Isler Y. in 2015, observed that the employment of k-fold cross validation in congestive heart failure diagnosis falsely affected the results of the study by enhancing them. The LOSO validation method was, as observed, the more robust method [54]. In 2012, Hafner M. suggested that validation methods employing samples from the same participant for both the training and test sets could lead to inflated accuracy. This bias arises from the fact that the classifier is trained with samples highly resembling those it will be tested on. Hafner referred to the LOSO validation method as the most realistic approach, as it prevents classifier training using patches from the same subject. [55]. This was also the case in this study, where the 10-fold validation performance results (accuracy, sensitivity and specificity) were above 90%, being significantly better than those of LOSO validation. This is happening due to the presence of epochs from the same subject in both the test and training sets. Consequently, the classifier was exposed to training and testing using epochs from the same subject, causing the classification to rely on individual characteristics rather than features that distinguish one class from another.

Regarding LOSO validation, the best classifier for the AD/CN problem was found to be Decision Trees with accuracy 78.5 with the Random Forests being second best with 77% accuracy. Naïve Bayes and kNN had the worst accuracy with 63% and 60% respectively. The best performing classifier for the FTD/CN problem was Random Forests with 86.3% accuracy while the second best was Decision Tree with 79.6% accuracy. Worst accuracy was, in both cases, achieved by kNN algorithm at 60% and 69.7%, respectively.

Finally, the best performing algorithm for the AD/FTD problem was again Decision Trees with 73% accuracy. However, all classifiers at this problem achieved low accuracies with high SD's. Also, all algorithms misclassified most FTD cases as AD, so further examination, bigger dataset, or other performance metrics such as F1 score are required to be explored.

It should be noted that the small participant number of this dataset is a limitation that may affect the performance of the classifiers. However, this limitation is partially alleviated in the following research, where more subjects were recruited.

Regarding previous research that has explored EEG detection of AD or FTD using machine learning, a variety of characteristics have been extracted, such as the time-domain and frequency domain features extracted in this study or other time-frequency domain, spatial reconstruction characteristics extracted from sLORETA [56], or other features that focus capturing deceleration, reduced complexity, reduced synchronization and neuromodulatory deficit. Also, a number of studies chooses to further divide the basic EEG rhythms and thus the rhythm alpha is found as α1 (8 - 10 Hz) and α2 (10 - 12 Hz)[57] and the rhythm beta which has the largest amplitude is divided into β1 (12.5 - 18 Hz), β2 (18.5 - 21 Hz) and β3 (21.5 - 30 Hz). An overview of the literature review of such studies, as found during the writing of this paper during 2021 can be found in Table 5.

Table 5. Related research of AD/FTD/CN classification studies, as found at 2021.

Writers	**Year**	**Sample (AD-FTD-CN)**	**Methodology**	**Classification Problem**	**Results**		
					ACC	SENS	SPEC
Lindau *et al.*[58]	2003	16 – 19 – 0	Power spectrum of EEG rhythms, cohesion, dominant rhythm	AD/FTD	93,30	-	-
Nishida *et al.* [59]	2011	19 – 19 – 22	EEG rhythms energy, sLORETTA, kNN	FTD/CN	85,80	55,00	84,00
				AD/CN	92,80	74,00	73,00
				FTD/AD	89,80	74,00	63,00
Caso *et al.* [60]	2012	39 – 39 – 39	Relative power of EEG rhythms, sLORETTA, ANOVA analysis	AD+FTD/CN	-	44.87	85
				AD/FTD	-	48.72	85
Dottori *et al.* [61]	2017	13 – 13 – 25	Connectivity features, SVM	AD+FTD/CN	54,00	-	-
				AD/FTD	73,00	-	-
				AD/CN	73,00	-	-
Proposed Methodology	2021	10 – 10 – 8	Energy, Mean, Variance, IQR, Random Forests, Decision Trees	AD/CN	78.5	82.4	74
				FTD/CN	86.3	87	83

The results of this preliminary research are promising in the context of distinguishing most common forms of dementia using quantitative EEG. However, more elaborate methodologies should be created in order to capture higher order relationships between the signals (such methodologies can be Deep Learning methodologies) and bigger participant pool should be considered in order to validate the robustness of the results. Thus, to the extent of the dementia research in this doctoral study, ways to alleviate these limitations will be considered.

Furthermore, regarding future work that can be done, but has not taken place during this doctoral research, is the study whether other dementia types such as vascular dementia and Lewy body dementia [41] can also be classified with machine learning approaches. Also, due t to the fact that Alzheimer's disease initially affects a certain part of the brain and then progresses to lesions throughout the cerebral cortex, detailed focus on specific parts of the brain should also be considered.

So, in conclusion, an easy to implement method for dementia detection using machine learning in EEG signals was introduced and evaluated in this study. The advantages and disadvantages of both k-fold and LOSO validation were examined and multiple features were evaluated. The promising results of over 78.5% accuracy for the AD/CN problem and 86.3% accuracy for the FTD/CN problem made the further development of such automatic methodologies necessary, leading to the next 3 studies on this chapter.

5.1.2. Enhanced Alzheimer's disease and Frontotemporal Dementia EEG Detection: Combining lightGBM Gradient Boosting with Complexity Features

The next research was presented in the CBMS 2023 conference in L'Aquila, Italy at June 2023. In this study, the same dataset was used again, but more elaborate machine learning approaches were tested. The basic goals for this study were the incorporation of the widest possible range of features for the classifier, as well as the employment, fine-tuning and monitoring of a well-established and highly accurate classifier that has been popular the last 5 years, being the lightGBM, a gradient boosting algorithm proposed by Microsoft Research team [49]. Moreover, the preprocessing step has been refined, incorporating sophisticated methodologies for artifact detection and rejection, such as ICA. The purpose of this work, was to maximize the discrimination capabilities for the AD/CN and FTD/CN problem, showcase our work in the research community and introduce the final part of the dementia research, which was the DICE-net and the published dataset. An overview of the methodology of this study can be found in Figure 18.

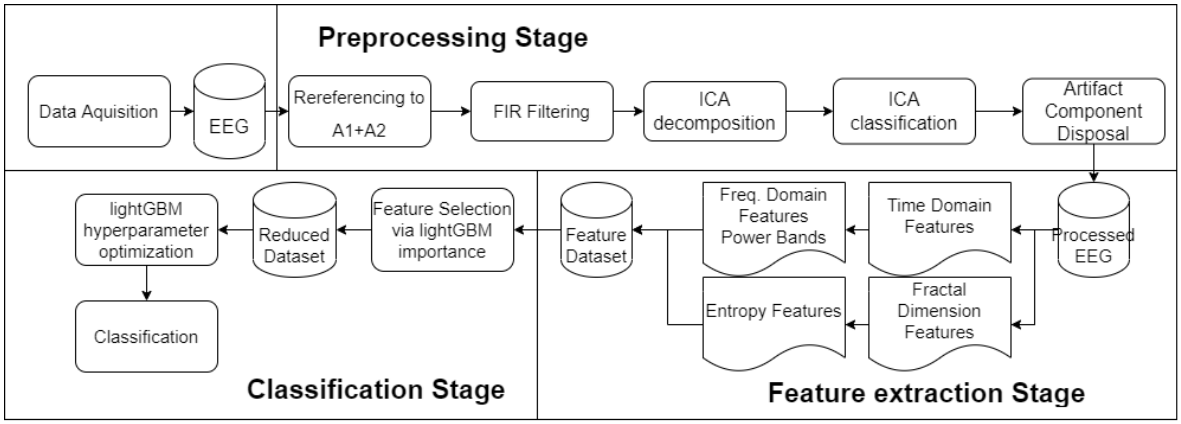


Figure 18 Flowchart of the methodology as described in the study.

Regarding the dataset used, 4 more patients were introduced, making the total number of subjects 32. Again, a detailed description of the participants statistics can be found in Table 6. The recording settings and the configuration of the clinical EEG device is the same and can be examined in the previous study.

The author list ordered as published was: **Andreas Miltiadous**, Katerina D. Tzimourta, Vasileios Aspiotis, Theodora Afrantou, Markos G. Tsipouras, Nikolaos Giannakeas, Euripidis Glavas, Alexandros T. Tzallas

Table 6. Statistical characteristics (values in brackets represent standard deviation) of the participants

	Gender (Male/Female)	Age	MMSE	CDR	Disease Duration in Months
AD	8/6	70.5 (7.1)	19.7 (2.76)	1 (0.54)	24 (9.88)
FTD	6/4	67.5 (4.5)	21.5 (1.83)	0.75 (0.26)	26 (9.24)
CN	4/4	68.5 (7.2)	30 (0)	-	-

5.1.2.1. Preprocessing

For the preprocessing, the EEG signals were first re-referenced to the A1, A2 electrodes which are located on the earlobes. A 4th order Butterworth filter was then applied to ensure that no frequency information below 0.4 or over 47 Hz was included, nor power line noise at 50 Hz. Then, the ICA decomposition, which is a robust methodology for artifact rejection in neurophysiological datasets was performed using the runICA algorithm in the EEGLAB [62] Matlab environment. The assumption which ICA functionality is based on is that the electric dipoles in the cortex can be modelled as independent sources [63]. In that sense, artifacts are classified as components based on the special properties. For example, all eye blinking artifacts are transformed into one component and thus, removing this component from the signal also removes the eye blinking artifacts. This labeling of the components as artifacts or as brain components can be done in a manual or in an automated manner. When

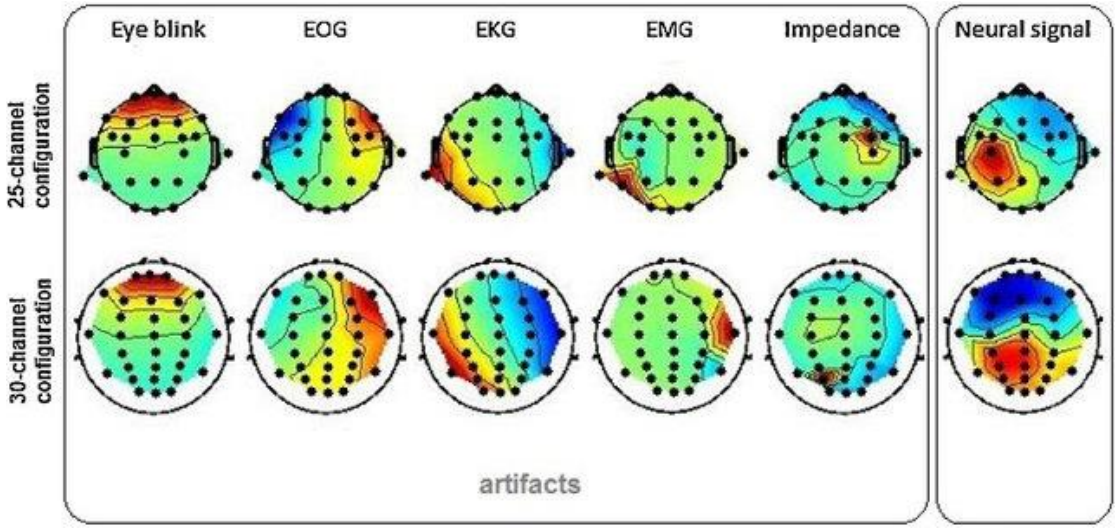


Figure 19 Different EEG ICA components, based on their origin. Adapted from "EEG artifact elimination by extraction of ICA-component features using image processing algorithms", by T. Radüntz et al., 2015.

aiming for a manual inspection, components are inspected based on their waveform or based on their PSD scalp heatmap representation, as presented in Figure 19[64].

However, in this study, a pretrained automatic methodology for ICA component classification was employed, namely the ICLabel [65], which is pre-build in the EEGLAB environment and classifies the components as brain, eye artifacts, muscle artifacts and others. The components that were classified as artifacts with a probability >0.9 were marked and automatically rejected. Even though the setting was eyes closed, still eye movement artifacts can be found, since patients with cognitive impairment have difficulties keeping their eye movement minimal. Thus, the ICA preprocessing stage is crucial to make sure that the high classification accuracy is not achieved due to the underlying eye artifacts but rather due to the brain activity.

5.1.2.2. Feature Extraction

For the feature extraction step, the MNE Python library [66], which is an EEG focused library, was used. First, the EEG signals were epoched in 4 second windows with 50% overlap. Then frequency domain, time domain, entropy and fractal dimension features were calculated for each epoch. For the frequency domain characteristics, the PSD was estimated (by eq. 5.1) by the Welch method that segments the signal in non-rectangular Hamming windows.

$$P_{x_m,M}(\omega_k) = \frac{1}{M}\left|FFT_{N,k}(x_m)\right|^2 \triangleq \frac{1}{M}\left|\sum_{n=0}^{N-1} x_m(n)e^{-j2\pi nk/N}\right|^2 \quad (5.1)$$

Then, the relative band power of the 5 frequency bands of interest was calculated for each epoch and each electrode.

Furthermore, to complete the feature matrix, the time domain characteristics, namely mean, standard deviation, skewness and kurtosis, the entropy characteristics, namely Permutation, Spectral, Singular Value Decomposition, Approximate and Sample entropy, the Hjorth features namely Mobility and Complexity and the fractal dimension characteristics namely Petrosyan, Katz, Higuchi fractal dimensions and Detrended Fluctuation were extracted. The entropy and fractal dimension characteristics were extracted using the well-established python library AntroPy. In total, 380 features were extracted, 20 (5 frequency domain + 4 time domain + 7 entropy + 4 fractal dimension) * 19 electrodes. Although all these features have been used in the EEG automated analysis, their combination is not common.

5.1.2.3. Classification

The classification pipeline consists of feature selection, hyperparameter optimization, training of the classifier and validation using the LOSO validation technique. The problems addressed are the AD/CN, FTD/CN and AD+FTD/CN.

(a) Classification Algorithm

The classifier employed was the LightGBM, which is a Gradient Boosting Decision Tree implemented by Microsoft Research team. In a Gradient Boosting Decision Tree model, each decision tree is sequentially trained, utilizing the residual errors (negative gradients) from the actual value (in the case of regression) or from the logarithm of the class probability, along with the pseudo-residuals generated by all previously constructed decision trees. This process continues throughout the training sequence of the model. Given a regression problem, and $F(x)$ the target approximation function, where $D = \{x_i, y_i\}_1^N$ the training dataset, gradient boosting goal is to minimize a loss function $L(y, F(x))$. $F(x)$ is an iterative weighted sum.

$$F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + \rho_m h_m(\boldsymbol{x}) \tag{5.2}$$

Where ρ_m is defined as the weight of the m_{th} $h(x)$. The approximation of the first function is normally the average of the training classes values. Each next approximation is trained on a new dataset $D = \{x_i, r_{mi}\}_{i=1}^N$where r_{mi} are the pseudo residuals of the step m calculated as:

$$r_{mi} = \left[\frac{\partial L(y_i, F(x))}{\partial F(x)}\right]_{F(x)=F_{m-1}(x)} \tag{5.3}$$

The lightGBM implementation has an innovative solution for ranking the absolute values of the training instances residual errors and discard the least informative ones, which is called Gradient based One-Side Sampling (GOSS). Furthermore, it also introduces an effective method for reducing the number of features called Exclusive Feature Bundling [49].

(b) Feature Selection

The total number of features (380) is too large to achieve good classification performance, given the size of the dataset. Thus, a feature selection step is required. To perform this step, a wrapper method named SelectFromModel, which is implemented in the scikit-learn python package and selects the best features based on the feature importance of a selected classifier (in this case lightGBM) was used. All the features that did not exceed 1.5* mean feature importance were discarded, significantly reducing the dimensionality of the dataset to 78 for the AD/CN problem, 50 for the FTD/CN problem and 56 for the AD+FTD/CN problem. The feature importance, as obtained by the feature selection step is presented in Figure 20.

(c) Hyperparameter Optimization

To optimize the hyperparameters of the classifier, a Bayesian optimization with Gaussian processes was used, which is implemented by the scikit optimize library (gp_minimize). The gp_minimize function represents an advanced technique in hyperparameter optimization, leveraging the principles of Gaussian Process within a Bayesian optimization framework. At its core, Gaussian Processes are utilized as surrogate models to probabilistically predict the performance of various hyperparameter configurations. This method is useful in scenarios

where direct evaluation of the performance function is computationally expensive or time-consuming. The Gaussian Process model, offers a prediction of performance for a given set of hyperparameters and quantifies the uncertainty of these predictions.

The selection of the next set of hyperparameters for assessment is dependent on an acquisition function in the operational mechanics of gp_minimize. This function balances

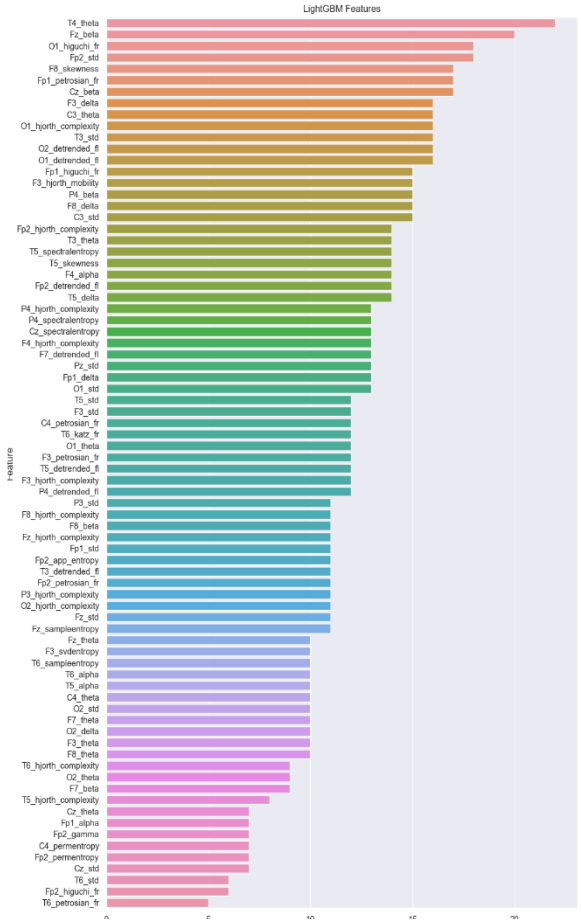


Figure 20 Feature Importance as calculated by the lightGBM algorithm.

the investigation of new hyperparameter spaces (to probe less understood areas) with the exploitation of previously found promising locations. By iteratively refining the Gaussian Process model with fresh data points, this balance provides efficient convergence to the ideal set of hyperparameters. Because of its iterative nature, gp_minimize may effectively navigate huge and complex hyperparameter spaces, typically outperforming traditional approaches such as grid or random search by lowering the number of required evaluations. This makes gp_minimize an invaluable tool for fine-tuning machine learning models, especially where hyperparameter selection precision is critical for model performance. performance.

5.1.2.4. Results

The performance of the classifier was evaluated using the LOSO validation technique. The results are presented in Table 7. The ROC curves of the three problems, are presented in Figure 21.

Table 7. Classification Results of the lightGBM algorithm

Problem	ACC	SENS	SPEC	F1	AOC
AD / CN	79.64%	83.52%	73.88%	83.06%	0.863
FTD / CN	82.67%	75.23%	89.56%	80.67%	0.925
AD+FTD/CN	89.72%	95.34%	72.75%	92.27%	0.941

This classification approach achieved similar results to the previously published performance. This indicated the need for more sophisticated extracted features that have been previously explored in the literature and found to be effective in Alzheimer's disease

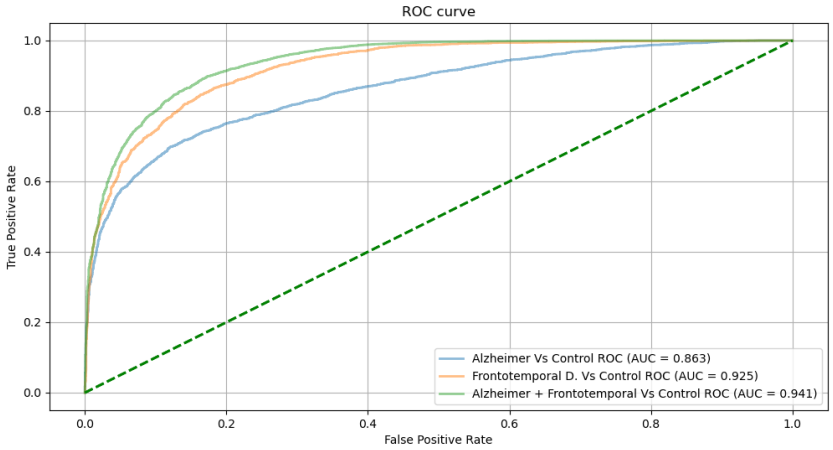


Figure 21. ROC curves of the three classification problems for the lightGBM implementation

detection. LightGBM was selected because multiple studies have highlighted the effectiveness of ensemble tree-based algorithms like Random Forests, on Alzheimer's recognition [13], [67] (along with SVM). Also, comparative analysis has shown that lightGBM is the fastest gradient boosting algorithm [68]. Therefore, this implementation was the best pick amongst all other tree ensemble classifiers for this study. However, different approaches such as deep learning Neural Networks should also be explored in order to examine any potential increase in the performance metrics. Finally, this study also had the same small dataset limitation as the previous one. This limitation was addressed in the next studies.

5.1.3. A dataset of scalp EEG recordings of Alzheimer's disease, Frontotemporal Dementia and Healthy subjects from routine EEG

This paper was published in the Data MDPI journal, during May 2023, to accompany and describe the published dataset named "A dataset of 88 EEG recordings from: Alzheimer's disease, Frontotemporal dementia and Healthy subjects" which was uploaded on OpenNeuro. The significant popularity that this dataset gained during the month of its publication (and later), led to being the cover of its issue.

The author list ordered as published was: **Andreas Miltiadous**, Katerina D. Tzimourta, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Dimitrios G. Tsalikakis, Pantelis Angelidis, Markos G. Tsipouras, Euripidis Glavas, Nikolaos Giannakeas, Alexandros T. Tzallas

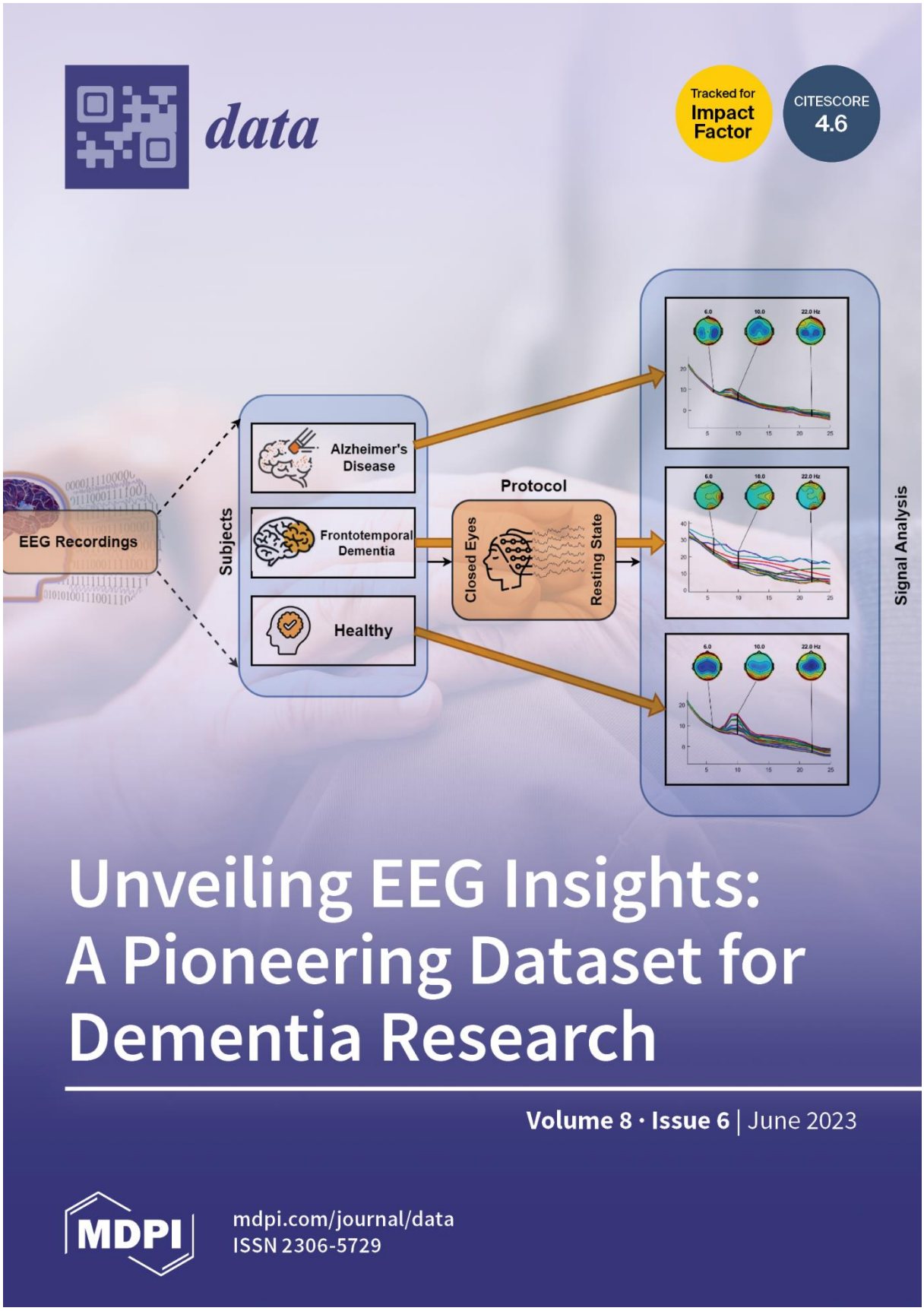


Figure 22 The cover of the issue in which the data descriptor paper was published. It features our study as the highlight study of the issue.

Alzheimer's disease and Frontotemporal Dementia are diagnosed using a mix of clinical evaluation, neurological testing, and cognitive testing. Imaging techniques such as PET-scan or MRI can also be used to diagnose certain illnesses. Both Frontotemporal Dementia and Alzheimer's disease can have moderate early symptoms that overlap with other neurodegenerative disorders or mental issues, making a diagnosis challenging in both situations. Improved detection tools are consequently required to aid in the early detection of these disorders. A quick diagnosis is critical because early treatment can help delay the appearance of symptoms that worsen and improve quality of life. A neurodegenerative disorder can be difficult to live with, but early detection allows for the implementation of safety precautions, legal and financial preparation, and emotional support services that can assist people and their families in managing. As a result, there is an urgent need for innovative detection methods that can aid in the early detection of Frontotemporal Dementia and Alzheimer's disease, ultimately leading to better outcomes for patients suffering from these conditions.

EEG is a promising biomarker for dementia detection, as established from the previous studies. A publicly accessible database could significantly amplify the research efforts in this domain, allowing the development and validation of new detection algorithms and facilitate a deeper understanding of the EEG patterns associated with dementia. Also, it could enable a broad spectrum of researchers to engage in innovative analysis of the subject and propose their findings. The insights that can be gained from such a dataset are invaluable, with the potential to lead to more accurate diagnostic tools and better patient outcomes in the fight against Alzheimer's and related dementia disorders.

During the years of this doctoral research, the previously established collaborations with the 2nd Department of Neurology at AHEPA University Hospital were leveraged to obtain approval for the data collection of EEG recordings. These recordings, clinically conducted in accordance with the established protocol, were authorized by the Scientific and Ethics Committee of AHEPA University Hospital, Aristotle University of Thessaloniki, under protocol number 142/12-04-2023. The investigations were carried out according to the rules of the Declaration of Helsinki of 1975, revised in 2008. The recordings were gathered, preprocessed and published as a publicly available dataset in the OpenNeuro database, named “A dataset of EEG recordings from: Alzheimer's disease, Frontotemporal dementia and Healthy subjects” with DOI: doi:10.18112/openneuro.ds004504.v1.0.6 (dataset and DOI version subject to change).

5.1.3.1. Dataset description

This dataset is consisted of the EEG resting state with closed eyes recordings from 88 participants in total. Out of them, 36 were diagnosed with Alzheimer's disease (AD group), 23 with Frontotemporal Dementia (FTD group) and 29 were healthy and made the Control group (CN). Evaluation of the cognitive state was performed by by the international Mini-Mental State Examination (MMSE) [69]. MMSE score ranges from 0 to 30, with lower MMSE indicating more severe cognitive decline.

Recordings were taken from 19 scalp electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2) and 2 reference electrodes on the earlobes, namely A1 and A2, placed according to the 10-20 international system. The sampling rate was 500Hz with a resolution of 10uV/mm. The duration of the EEG recordings was approximately 13.5 minutes for AD group (min=5.1, max=21.3), 12 minutes for FTD group (min=7.9, max=16.9) and 13.8 for CN group (min=12.5, max=16.5). In total, 485.5 minutes of AD, 276.5 minutes of FTD and 402 minutes of CN recordings were included in the dataset.

The recordings were performed during a period of 10 years, with the standard clinical protocol. The duration of the disease measured in months was 25 (median) with IQR range (Q1-Q3) being 24-28.5 months. No dementia related comorbidities were reported, regarding the dementia groups. For the initial diagnosis of the AD and FTD patients, the criteria provided by the Diagnostic and Statistical Manual, 3rd ed., revised (DSM-IIIR), DSM IV, ICD-10 [70] and the National Institute of Neurological, Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association (NINCDS – ADRDA) [71] were followed. Regarding the MMSE score, the average for the AD group was 17.75 (sd=4.5), for the FTD group 22.17 (sd=8.22) and for the CN group 30. The mean age of the AD group was 66.4 (sd=7.9), for the FTD group was 63.6 (sd=8.2), and for the CN group was 67.9 (sd=5.4). The description of each participant as provided in the data descriptor paper can be found in Table 8. All the participants were anonymized, and personal information has not been disclosed, due to GDPR restrictions.

Table 8 Participants Description. A indicates AD patient, F indicates FTD patient and C indicates a healthy subject. F indicates female and M indicates male.

participant_id	Gender	Age	Group	MMSE	participant_id	Gender	Age	Group	MMSE
sub-001	F	57	A	16	sub-014	M	77	A	14
sub-002	F	78	A	22	sub-015	M	61	A	18
sub-003	M	70	A	14	sub-016	F	68	A	14
sub-004	F	67	A	20	sub-017	F	61	A	6
sub-005	M	70	A	22	sub-018	F	73	A	23
sub-006	F	61	A	14	sub-019	F	62	A	14
sub-007	F	79	A	20	sub-020	M	71	A	4
sub-008	M	62	A	16	sub-021	M	79	A	22
sub-009	F	77	A	23	sub-022	F	68	A	20
sub-010	M	69	A	20	sub-023	M	60	A	16
sub-011	M	71	A	22	sub-024	F	69	A	20
sub-012	M	63	A	18	sub-025	F	79	A	20
sub-013	F	64	A	20	sub-026	F	61	A	18

participant_id	Gender	Age	Group	MMSE
sub-027	F	67	A	16
sub-028	M	49	A	20
sub-029	F	53	A	16
sub-030	F	56	A	20
sub-031	F	67	A	22
sub-032	F	59	A	20
sub-033	F	72	A	20
sub-034	F	75	A	18
sub-035	F	57	A	22
sub-036	F	58	A	9
sub-037	M	57	C	30
sub-038	M	62	C	30
sub-039	M	70	C	30
sub-040	M	61	C	30
sub-041	F	77	C	30
sub-042	M	74	C	30
sub-043	M	72	C	30
sub-044	F	64	C	30
sub-045	F	70	C	30
sub-046	M	63	C	30
sub-047	F	70	C	30
sub-048	M	65	C	30
sub-049	F	62	C	30
sub-050	M	68	C	30
sub-051	F	75	C	30
sub-052	F	73	C	30
sub-053	M	70	C	30
sub-054	M	78	C	30
sub-055	M	67	C	30
sub-056	F	64	C	30
sub-057	M	64	C	30

participant_id	Gender	Age	Group	MMSE
sub-058	M	62	C	30
sub-059	M	77	C	30
sub-060	F	71	C	30
sub-061	F	63	C	30
sub-062	M	67	C	30
sub-063	M	66	C	30
sub-064	M	66	C	30
sub-065	F	71	C	30
sub-066	M	73	F	20
sub-067	M	66	F	24
sub-068	M	78	F	25
sub-069	M	70	F	22
sub-070	F	67	F	22
sub-071	M	62	F	20
sub-072	M	65	F	18
sub-073	F	57	F	22
sub-074	F	53	F	20
sub-075	F	71	F	22
sub-076	M	44	F	24
sub-077	M	61	F	22
sub-078	M	62	F	22
sub-079	F	60	F	18
sub-080	F	71	F	20
sub-081	F	61	F	18
sub-082	M	63	F	27
sub-083	F	68	F	20
sub-084	F	71	F	24
sub-085	M	64	F	26
sub-086	M	49	F	26
sub-087	M	73	F	24
sub-088	M	55	F	24

5.1.3.2. Dataset structure

The dataset was converted and uploaded in the BIDS format, a widely recognized standard in neuroimaging. BIDS, short for Brain Imaging Data Structure that offers a uniform and clear structure for organizing brain imaging data, including EEG recordings. This format enhances the dataset's compatibility with various analysis tools such as EEGLAB, promoting easier sharing and research collaboration. Adopting BIDS also aids in research reproducibility, as it clearly outlines data and metadata organization.

The dataset is consisted of: 1) the dataset_description.json file, which contains all the necessary information about the authors of the dataset, the acknowledgements, the DOI, the BIDS version, the publication license and the ethics approval statement. 2) The

participants.json file, which contains information about each participant of the dataset, and most importantly their class, as found in Table 8, and can be accessed and used by third party software such as EEGLAB to automatically label them in a study. 3) Just like participants.json, the participants.tsv file also contains participant information in tabular separated value format. 4) a series of folders, each labeled as sub-0XX, corresponding to a unique participant ID from the participants' table. Inside every folder are three key files: The sub-0XX-task_eyesclosed_eeg.json file contains essential EEG recording details, including the electrode placement scheme (10-20), reference (A1 A2), device and amplifier model, channel count, sampling frequency, recording duration, and more. The sub-0XX_task-eyesclosed_channels.tsv file provides specifics about electrode placement.

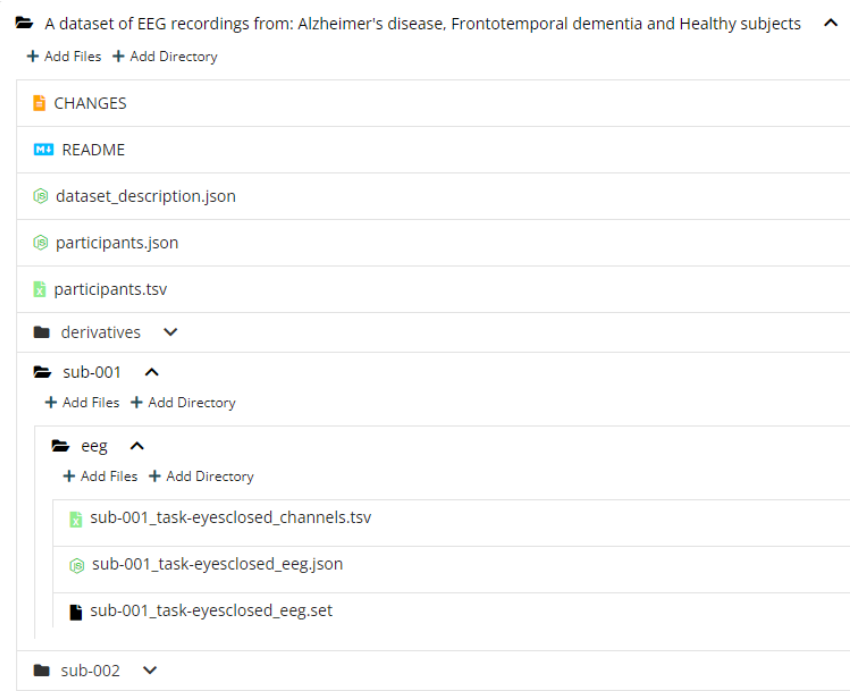


Figure 23 The dataset structured in the BIDS format, as uploaded in OpenNeuro

The sub-0XX_task-eyesclosed_eeg.set file holds the EEG recordings for the participant in .set format, which is recognized as one of the four EEG formats permitted by BIDS (the others being .edf, .vhdr or .eeg, and .bdf). Two important points to note are: Firstly, the .set files contain all necessary recording data, making them accessible even outside a BIDS framework. Secondly, the sub-0XX_task-eyesclosed_channels.tsv and sub-0XX-task_eyesclosed_eeg.json files are identical for each participant, barring recording duration, due to the uniform recording setup used. Therefore, it's unnecessary for users to review these files for every participant. 5) Finally, a 'derivatives' folder containing subfolders with similar structure, except that the EEG recordings here are preprocessed, as described in the next section. Figure 23 is a screenshot of the structure of the dataset, as found in OpenNeuro.

5.1.3.3. Preprocessing

As mentioned above, along with the original, unprocessed recordings, a folder named derivatives which contained (in BIDS format as well) the preprocessed version of the signals. Regarding these signals, the preprocessing pipeline was as follows. First, the signals were re-referenced to A1-A2 electrodes and a Butterworth band pass filter was applied, allowing only frequency content of 0.5-45 Hz. Then, the ASR routine was performed to eliminate persistent or large amplitude artifacts, removing bad data periods what exceeded the maximum acceptable 0.5 second window standard deviation of 17 (this number is considered a rather conservative value). Next, the ICA methodology was applied using the runICA algorithm, as described in previous sections, and the automatic component classification routine ICLabel was employed to label the components as brain signal or eye, muscle or other artifact. Then, the components that were labeled as artifacts with a possibility greater than 90% were automatically excluded and the signal was re-compiled cleared from artifacts. Figure 24 represents the difference between an unprocessed and a cleared signal.

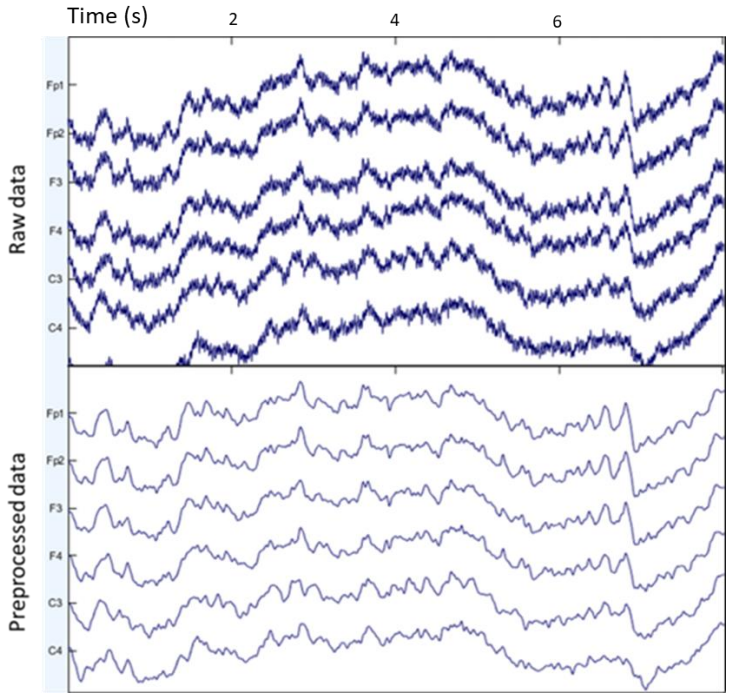


Figure 24 A snapshot of an unprocessed signal, and the same signal after preprocessing

5.1.3.4. Benchmark experiments

The next step was to benchmark this dataset based on the classification performance of some baseline classifiers, in order to provide some baseline to future researchers proposing a classification scheme. To do so, the most commonly extracted features for EEG classification tasks, being the Relative Band Power of the 5 rhythms of interest, were extracted. These rhythms were defined as:

- Delta: 0.5 – 4 Hz
- Theta: 4 – 8 Hz
- Alpha: 8 – 13 Hz
- Beta: 13-25 Hz
- Gamma: 25-45 Hz

These features were extracted after epoching the signal into 4 second overlapping epochs (50% overlap). Each row in the feature matrix was labeled as AD, FTD or CN. To obtain the RBP, the PSD was calculated using the Welch method. This method splits the signal into overlapping segments and calculates for each one the squared magnitude of the DFT, creating a final estimate by averaging the values. Scalp heatmaps of the 5 rhythms for each group are presented in Figure 25.

Some well-established machine learning algorithms were examined for the classification of AD-CN and FTD-CN for benchmarking the dataset. The validation was performed using the LOSO validation methodology. The performance metrics that were calculated were Accuracy (ACC), sensitivity (SENS), specificity (SPEC) and F1 score (F1). The classifiers were LightGBM (with Hyperopt [72] hyperparameter optimization), Multilayer Perceptron (MLP) with 1 hidden layer of 3 neurons, Random Forests, Support Vector Machines (SVM) with polynomial kernel, and k-Nearest Neighbours (kNN), k=3. The performance of the classifiers is presented in Table 9 for the AD-CN problem and Table 10 for the FTD-CN problem.

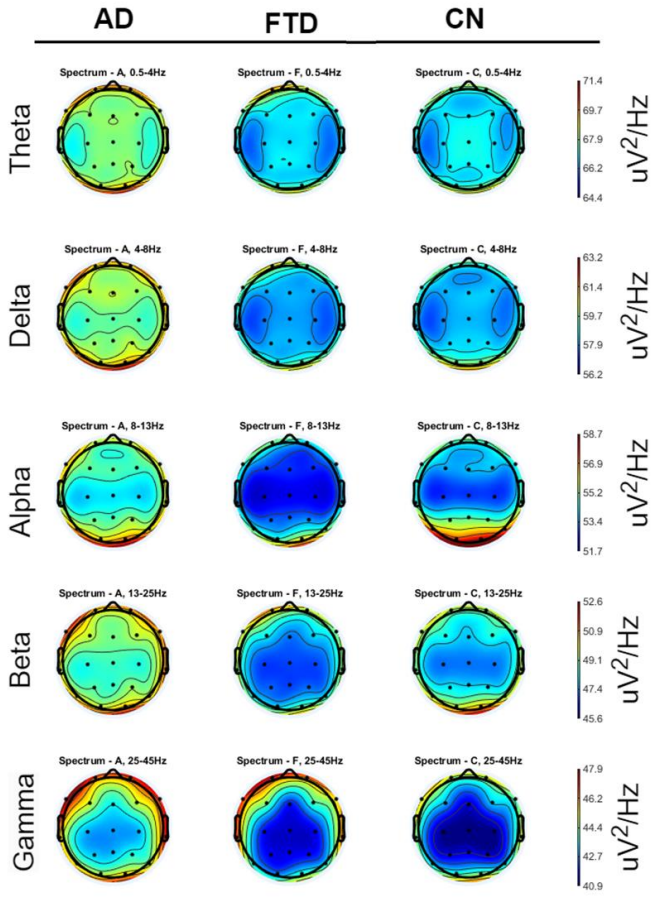


Figure 25 Scalp heatmaps of the Power Spectral Density for the 5 rhythms, averaged across the 3 different groups.

Table 9 Classification performance metrics of LOSO validation for AD-CN problem.

AD/CN	ACC	SENS	SPEC	F1
LightGBM	76.43%	76.01%	76.16%	76.12%
SVM	73.14%	71.89%	75.98%	73.74%
kNN	71.23%	69.67%	74.19%	72.81%
MLP	73.12%	73.00%	74.63%	74.82%
Random Forests	77.01%	78.32%	80.94%	75.31%

Table 10 Classification performance metrics of LOSO validation for FTD-CN problem.

FTD/CN	ACC	SENS	SPEC	F1
LightGBM	72.43%	61.13%	80.74%	67.32%
SVM	70.14%	62.41%	75.98%	68.32%
kNN	67.34%	59.67%	76.13%	70.81%
MLP	73.12%	63.00%	78.63%	72.82%
Random Forests	72.01%	72.32%	80.94%	66.31%

Thus, in both problems the lightGBM algorithm seems to be the best performing one, so its performance metrics can be considered as the baseline for this dataset. So, the expected performance of an algorithm using this dataset with LOSO validation is expected to be greater than 76.43% accuracy and 76.12% F1 score for the AD/CN problem and 72.43% accuracy and 67.32% F1 score for the FTD/CN problem.

5.1.4. DICE-net: A Novel Convolution-Transformer Architecture for Alzheimer Detection in EEG Signals

Andreas Miltiadous[1], Emmanouil Gionanidis[2], Katerina D. Tzimourta[1], Nikolaos Giannakeas[1], Alexandros T. Tzallas[1]

This article presents the peak performance algorithm for Alzheimer's disease detection, that has been proposed during doctoral research. It was published in the open access journal IEEE Access, during July 2023. The author list was: It is a robust methodology that incorporates the artifact rejection techniques that have been mentioned in previous paragraphs, namely ICA and ASR, and employs the transformers neural networks, a deep learning architecture that has been introduced by Google research during 2017 [52]. Special acknowledgement is due to Emmanouil Gionanidis

The importance and necessity of using EEG for training automated machine learning methodologies for dementia detection or severity assessment has been previously discussed. The relative band power of the 5 rhythms has been widely used as a feature and proven to be effective in the classification of different EEG related problems. However, there is also another way of analyzing the EEG signal that is increasing in popularity, that being the coherence analysis and graph theory methods, as they have powerful capabilities of investigating the organization and functional connectivity of the brain. Coherence analysis measures the degree to which different brain regions synchronize at specific frequency bands, providing information about the strength and patterns of functional connectivity. Research has shown that coherence analysis can be used as an effective biomarker for Alzheimer's disease [73]. Thus, in this study both coherence analysis and relative band power were used as biomarkers for the creation of the feature set and training of the algorithm.

Regarding the classifier, deep learning architectures have been widely used in EEG classification tasks and have proven to be effective in capturing the temporal relationships of the EEG data. Such deep learning models that have been used are CNNs [74], RNNs [14], and ARs [38]. These models have exhibit promising performance. However, a novel architecture for Natural Language Processing, namely the Transformer Neural Network have gained significant interest in various subject areas beyond their original domain, demonstrating superior performance compared to their counterparts in image classification, speech recognition and other tasks. Transformers effectiveness lies in their capacity for handling data sequences of varying lengths and scaling efficiently with large datasets. This made them a promising tool for evaluation in fields like biomedical signal processing. For instance, in the realm of EEG-based emotion recognition, Guo et

---

The author list ordered as published was: **Andreas Miltiadous**, Emmanouil Gionanidis, Katerina D. Tzimourta, Nikolaos Giannakeas, Alexandros T. Tzallas

al. [75] implemented a Transformer-based model for classifying emotional states from EEG data. Their approach achieved an accuracy of 83.03% in a three-class problem, surpassing most existing methods on the same database. Similarly, in the context of motor-imagery EEG classification, Xie et al. [76] demonstrated the effectiveness of a Transformer model on raw signals, achieving accuracies of 83.31%, 74.44%, and 64.22% in two, three, and four class problems, respectively, often outperforming alternative methods on the same dataset. These studies underscore the need for incorporating transformer encoders in EEG applications and underscore the importance of investigating their use in classifying neurodegenerative EEG data, particularly for AD and other forms of dementia.

The basic part of the Transformer architecture is the self-attention mechanism, which allows the model to attend to different parts of an input sequence and create a modified output sequence by computing a weighted sum of the input, based on the similarity between the elements in the sequence. The philosophy is that the self-attention mechanism enables the model to give more attention to the more relevant parts of an input sequence. A Transformer architecture is consisted of an encoder and a decoder block, and each of them is consisted of multiple self-attention layers and Feed Forward Neural Network layers, residual connections and layer normalizations.

The advancements in the Transformer research made it able to solve other problems such as image or text classification. Some of the most notable architectures that led to the revolution in transformers are the Bidirectional Encoder Representations from Transformers (BERT) [77] and the Vision Transformer (ViT) [78]. These approaches employ the Transformer Encoder along with the Class Token embedding, an additional sequence integrated into the input sequence. This Class Token serves as a sequence-level representation for the classification task, designed to encapsulate a contextualized representation of the entire sequence. The encoder's output, or just the CLS token, is forward into a Neural Network architecture for classification. Adaptations of these techniques have expanded the application of Transformers to classification tasks across multiple fields, extending beyond the realms of NLP and Computer Vision to include areas like speech recognition [79], protein classification [80], and time-series analysis [81].

In this study, a novel convolution – transformer hybrid architecture was fed with Spectral Coherence Connectivity (SCC) and Relative Band Power (RBP) features and achieved noteworthy performance in both AD/CN and FTD/CN problems. RBP and SCC are two of the most promising biomarkers for AD detection, as literature has shown increase in Theta/Delta ratio [82] and decrease in synchronization likelihood [83] in AD patients. These features were expressed in image-like representations (3d matrixes) and were fed into 2 parallel convolution blocks, which reduced the dimensionality of the input by extracting useful information and then the outputs were fed into 2 parallel Transformer Encoder blocks which conseptualize the sequence content by using a CLS token. Finally, a Feed-Forward Neural Network performed the final stage of the

classification. This methodology has also been evaluated for the FTD/CN problem, however it has been optimized for the AD/CN problem.

5.1.4.1. Methodology

The recordings of the published dataset that has been described in the previous chapter have been used for the evaluation of the proposed methodology. In total, 88 recordings from 36 AD, 23 FTD and 29 CN subjects were used. Their neurophysiological and cognitive state was evaluated by the MMSE score, the results of which are described in the previous chapter. The study was conducted with the approval of the Scientific and Ethics committee of AHEPA University Hospital, Aristotle University of Thessaloniki, under protocol number 142/12-04-2023, and in accordance with the rules of the Declaration of Helsinki. The recording settings are also described in the previous section.

(a) Preprocessing

For the preprocessing of the signals, a Butterworth bandpass filter of 0.5-45 Hz was first applied, then the signals were re-referenced to A1-A2 electrodes. Then, the automatic artifact rejection methodologies ICA and ASR were applied as described in the previous section.

(b) Feature Extraction

For the feature extraction, the signals were first divided into 30-second time windows with 50% overlap to create the EEG signals set that will be used for training and testing. Then, each 30 second window was further divided into 1-second epochs, and the RBP and SCC features were extracted for each epoch (T=30), for each channel (C=19), for each of the bands of interest (B=5). Thus, in total 2 3-dimensional arrays of size [T,B,C] were extracted for each 30-second window. The bounds of the frequency bands were set as:

- Delta: 0.5 – 4 Hz
- Theta: 4 – 8 Hz
- Alpha: 8 – 13 Hz
- Beta: 13-25 Hz
- Gamma: 25-45 Hz

To calculate the RBP, the PSD for each 1-second window was obtained with the Welch method, which is analyzed in chapter 3.2.2.1. The PSD of each frequency band was obtained and then the ratio of each band was extracted as the RBP of this band.

Regarding the SCC feature, it was used to quantify the synchronization of the brain signals. It is calculated by measuring the spectral coherence between each pair of signals (the measure of similarity of the frequency content between 2 signals), and then averaging the values electrode-wise. It is calculated as:

$$SCC_x = \frac{1}{C}\sum_{y=1}^{C}\frac{|S_{xy}|}{\sqrt{S_{xx} * S_{yy}}} \tag{5.1}$$

Where S_{xx} is the PSD of $x(t)$ signal and S_{yy} the PSD of the $y(t)$ signal, and $S_{xy}(f) = \lim_{T\to\infty}\frac{1}{T}\left[\widehat{x_T^*}(f)\widehat{y_T}(f)\right]$ by the Parserval's theorem. In this case, the PSD of the signals was calculated by exploiting the Morlet Wavelet Transform which was calculated as:

$$w(\omega,t) = \left(\pi^{\left(-\frac{1}{4}\right)}\right) \times \left(e^{(i*\omega*t)} - e^{\left(-\frac{1}{2}*\omega 0^2\right)}\right) \times e^{\left(-\frac{t^2}{2}\right)}, \omega \in \{2,6,10,18,35\} \tag{5.2}$$

And then the Wavelet Transform was calculated as the convolution of $w(\omega,t)$ and x(t) as:

$$C(\omega,\tau) = \langle x, \mathrm{w}_{\omega,\mathrm{t}}\rangle = \int_{\mathrm{R}} x(t)\psi^*_{\omega,\tau}(t)d\,t \tag{5.3}$$

Again, the SCC was calculated ∀ t ∈ T, ∀ channel ∈ C, ∀ band ∈ B, resulting in a 3-dimensional matrix [T,B,C], similar to the RBP matrix.

(c) DICE-net architecture

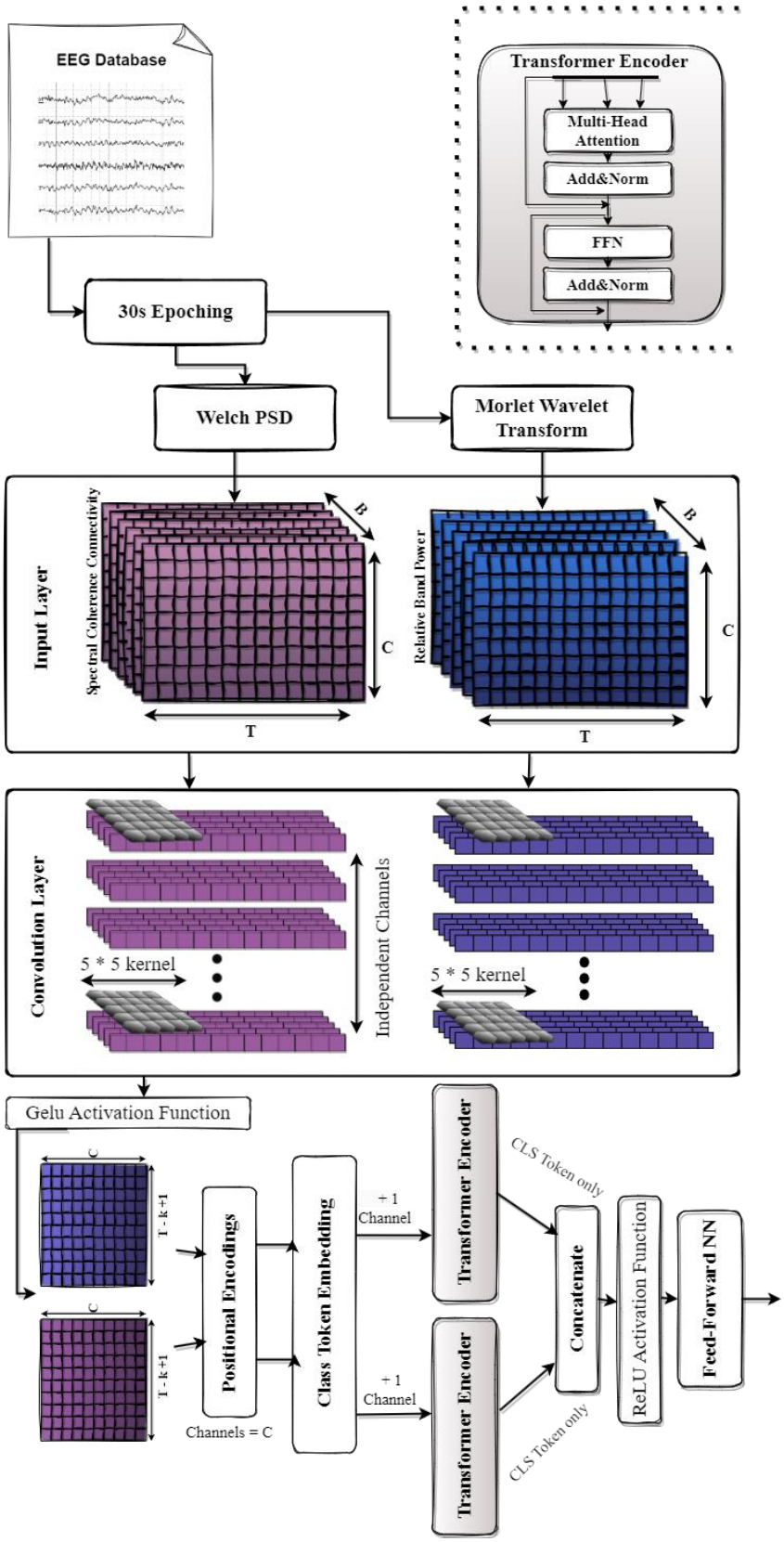


Figure 26 DICE-net Architecture

For the model architecture, two parallel blocks that receive input $X_i \in \mathbb{R}^{Ba \times T \times B \times C}$, where Ba is the batch size of the Neural Network were introduced, that got the RBP and SCC features as input. These blocks consisted of a depthwise convolution layer, a positional embedding layer a CLS token embedding and a

transformer encoder. After that, the parallel blocks are concatenated with a concatenation layer and a Feed-Forward Neural Network is used to classify and determine the class of the input. The complete architecture is described in Figure 26.

To avoid overfitting, early stopping is performed to determine the optimal number of epochs for the model. When referring to epochs in Neural Networks, we talk about the number of times that the same input will be used to fed the network and modify its weights. Dataset was splitted into test, train and validation set, and the performance of the classifier in the validation set is evaluated after each epoch. At the 20_{th} epoch of not improved performance in terms of accuracy, the training is stopped and the best model so far is kept. The validation set was created by iteratively leaving out 6 subjects. The number of epochs with the best overall performance (calculated by averaging) was then set as the number of epochs parameter. Then, the LOSO validation method was employed to calculate the performance of the model.

For maximizing the performance of the model, besides the optimization of the epoch parameter with early stopping, ablation experiments were conducted and will be mentioned in a next paragraph. It should be noted that every ablation experiment was taken place considering the performance on the AD/CN problem and not the FTD/CN problem. The FTD/CN classification was not the goal, but rather a comparison tool on how this methodology performs in a different type of dementia.

Convolution Layer

Given the dimensions of the input being $2 \times [30,5,19]$ the total number of values in the input matrix is prohibitive for training the neural network, thus a depthwise convolution layer was installed to reduce the dimensionality. A depthwise convolution layer is a convolution layer that allows data independence at a given dimension. The dimension of choice was the channel (electrode) dimension, so that the information of each channel remains independent. The depthwise convolution layer consists of C convolution kernels of size 5×5, instead of 1 kernel that is found in the canonical convolution layer. Each kernel strides in a tensor of size [T,B], with stride step = 1. No padding (zero) is added to the convolution input, thus the spatial dimensions of the convolution layer output are calculated as $Out(x,y) = (W\_in(x,y) - k(x,y) + 2P)/stride + 1$. So, the size of the output of each layer is [T-k+1,B-k+1,C]=[26,1,19] (flattened to [26,19]), k=kernel size. The kernel weights are trained with backpropagation using a Gaussian Error Linear Units (GELU) function, which is calculated as

$$GELU(x) = 0.5 \times x \times (1 + Tanh(\sqrt{\frac{2}{\pi}} \times (x + 0.044715 \times x^3)) \quad (5.4)$$

Positional Encoding Layer

In contrast to the CNN or RNN architecture, a Transformer Encoder is not aware of the positional information of the input, and this positional information needs to be modeled beforehand. To do so, a positional encoding layer is employed, which provides spatial

$$p_{k,2i} = \sin\left(\frac{k}{10000^{2i/d}}\right), p_{k,2i+1} = \cos\left(\frac{k}{10000^{2i/d}}\right) \quad (5.5)$$

information about the data's absolute or relative position. This layer usually precedes the transformer encoder in most transformer related architectures for NLP or computer vision.

For sequentially ordered data across the T axis $X \in \mathbb{R}^{T \times C}$, the Positional Encodings (PE) are calculated as:

Where $k \in \{0,1,\ldots,C-1\}$ and $i \in \{0,1,..,T/2\}$. According to the study that it was first proposed [52], it was suggested that positional encodings enable the model to understand relative positions, as any given offset P_{k+off} can be expressed as a linear function of P_k. The Positional Encoding Layer is designed without any trainable parameters, which means it does not require gradient computation during back-propagation. As a result, its weights remain unaltered throughout the training process.

Class Token Embedding

The class token or CLS token is a special token embedded to the input sequence. Its purpose is to capture the overall meaning of the sequence, to be a representation for the entire sequence for image classification or object detection tasks. It is attached to the input and attends to important information from anywhere in the sequence, during the training of the transformer. Then, all the other input is discarded, keeping only the CLS token for further classification. In this implementation, the CLS token was appended as an extra column of size [T,1] on each of the two tensors, resulting in tensors with dimensions [T,C+1]. The initialization of the CLS token was randomized from a canonical distribution.

Transformer Encoder Layer

The transformer architecture was first introduced in the NLP domain and has applications in text and image classification. The encoder is a block of the transformer architecture which reconstructs a collection of n items to another collection of n items, $f_\theta: \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^{n \times d}$, encoding the relational structure of the input as data in the reconstructed output. These items are sequences, yet the encoder is oblivious of the sequational relationships, and this is the reason why the Positional Encoding Layer is employed beforehand.

A Transformer Encoder layer may be comprised of one or more stacked Transformer Encoders, each one serving as the input for the next. Each Transformer Encoder is consisted of a Multi-Head Self-Attention (MSA) layer with residual connection around it and a FFN with residual connection. Likewise, the MSA layer is consisted of several Self-Attention heads, which calculate the relationships between different parts of an input sequence in a sentence, and in this case of the C input channels, capturing each channels individual importance in relation to the others.

A Self-Attention layer is functioning as explained in this paragraph. First, the input sequence is linearly projected to the query (Q), key (K), value (V).

$$Q^{(h)}(x_i) = W_{h,q}^T x_i, \quad K^{(h)}(x_i) = W_{h,k}^T x_i, \quad V^{(h)}(x_i) = W_{h,v}^T x_i \quad (5.6)$$

Then, the attention of each channel is determined by a score matrix calculated as:

$\alpha_{i,j}^{(h)} = softmax_j\left(\frac{\langle Q^{(h)}(x_i), K^{(h)}(x_j)\rangle}{\sqrt{k}}\right)$, k the dimension of Q and K (5.7)

And the MSA is calculated as:

$$MSA(\boldsymbol{x_i}) = \sum_{h=1}^{H} W_0 \sum_{i=1}^{C} \alpha_{i,j}^{(h)} V^{(h)}(\boldsymbol{x_i}) \quad (5.8)$$

C= input channels, H=number of self-attention heads and W_0 the trainable weights for each head. Finally, the MSA output is passed through an FFN with ReLU activation function and a dropout layer with dropout probability of 0.1. It should be noted that the input and the output of the Transformer Encoder have the same dimensions.

Feed Forward Network

The final stage of the model is an FFN. Assume that the dimensions of the Transformer Encoder outputs are [T,C+1], all channels except from the CLS token that was previously introduced were discarded and the 2 CLS tokens of T values were concatenated in an array of size 2× T which was then normalized and fed into an FFN of 1 input layer (52 neurons), 1 hidden layer of 24 neurons and an output layer. Before each layer, a dropout with probability of 0.2 was added. After each layer a Batch Normalization layer was added. The activation function for the hidden layer of the FFN was ReLU.

A complete overview of the hyperparameters and the architecture of the model can be found in Table 11.

Table 11 The architecture and hyperparameters of DICE-net model.

Layer	Type	Input	Parameters	Output
A1, A2	Input			[B,30,5,19]
C1	Conv2d	A1	kernel=[5,5], stride=[1,1], groups=19	[B,26,19]
	Gelu	C1		
P1	PositionalEncoding1D	C1	channels=19	
CLS1	Parameter(Randn)	__		[1,26,1]
	torch.expand	CLS1	(expand to batch size)	[B,26,1]
TR1	torch.concat	CLS1 P1	dim=2	[B,26,20]
	TransformerEncoderLayer	TR1	num_layers=1, dmodel=2, nhead=2	
	drop channels	TR1	[:,:,0] (only CLS1)	[B,26]
C2	Conv2d	A2	kernel=[5,5], stride=[1,1], groups=19	[B,26,19]
	Gelu	C2		
P2	PositionalEncoding1D	CG2	channels=19	
CLS2	Parameter (Randn)	__		[1,26,1]
	torch.expand	CLS2	(expand to batch size)	[B,26,1]

TR2	torch.concat	CLS2 P2	dim=2	[B,26,20]
	TransformerEncoderLayer	TR2	num_layers=1, dmodel=2, nhead=2	
	drop channels	TR2	[:,:,0] (keep only CLS2)	[B,26]
FFN	torch.concat	TR1 TR2	dim=1	[B,52]
	LayerNorm	FFN	normalized_shape=52	
	Dropout		prob=0.2	
	Linear		in_features=52, out_features=24	[B,24]
	BatchNorm1d			
	ReLU			
	Dropout		prob=0.2	
	Linear			[B,1]
	Sigmoid			
Loss	BCEWithLogitsLoss			

Ablation Experiments

In order to achieve the best possible performance for the model, different configurations were explored and different combinations of hyperparameters were tested. The different ablation experiments that were explored and are worth mentioning are:

1. NO-TRANS: Removed Transformer Encoder and Positional Embeddings. The results of CNN were directly fed to FFN.
2. E-DICE: Early concatenation. The concatenation of the inputs has taken place exactly after the CNN layers. Only one CLS token was generated.
3. 2-DICE: Two stacked encoder layers.
4. M-CLS: The CLS token is not randomly initialized but rather initialized by the mean values of each row.
5. ALL-DICE: No channels are removed. Instead, the values of all channels are fed to the FFN.
6. ALL-E-DICE: Early concatenation, no channels removed prior to FFN.
7. Some of these different ablation configurations are presented in Figure 27.

Comparison Algorithms

To evaluate the performance of the model, multiple well-established algorithms were trained and evaluated for benchmarking. These algorithms were: 1) k-Nearest Neighbors with Principal Component Analysis (PCA-kNN), 2) XGBoost, 3) LightGBM, 4) CatBoost, 5) Support Vector Machines with PCA (PCA-SVM), 6) Multilayer Perceptron (MLP). All gradient boosting algorithms have been hyperparameter optimized with

Hyperopt [84], which is a python package for machine learning hyperparameter optimization. The architecture for the MLP was 190-96-1.

The feature extraction process for these algorithms was quite different, since they do not support 3-dimensional matrix input. Thus, the same features were extracted, but instead of 30-second time windows, 15-second alternatives were splitted and the same features were extracted in a 2-dimensional, traditional manner.

Besides these algorithms, state-of-the-art deep learning architectures specifically designed for EEG raw signal classification were also examined. These architectures were EEGNet [85], EEGNetSSVEP [86], DeepConvNet, ShallowConvNet [87].

Validation Methodology

All the performance metrics, for every algorithm were calculated using the LOSO validation method.

Experimental Setup

The preprocessing of the signals has taken place in the EEGLAB Matlab [62] (2021) environment. The feature extraction along with the frequency and time-frequency transformations has taken place in python 3.10 using the MNE library, version 1.2 [66]. The deep learning model was implemented in Python 3.8 with the PyTorch library, version 1.13 [88]. The rest of the algorithms were implemented using the Scikit-Learn library. The processor that was used for the training of the algorithms was a GPU RTX 3060 Ti with CUDA 11.7.

Computational Complexity

The computational complexity of the DICE-net and the ablation experiments in terms of GFlops and parameter numbers can be found in Table 12.

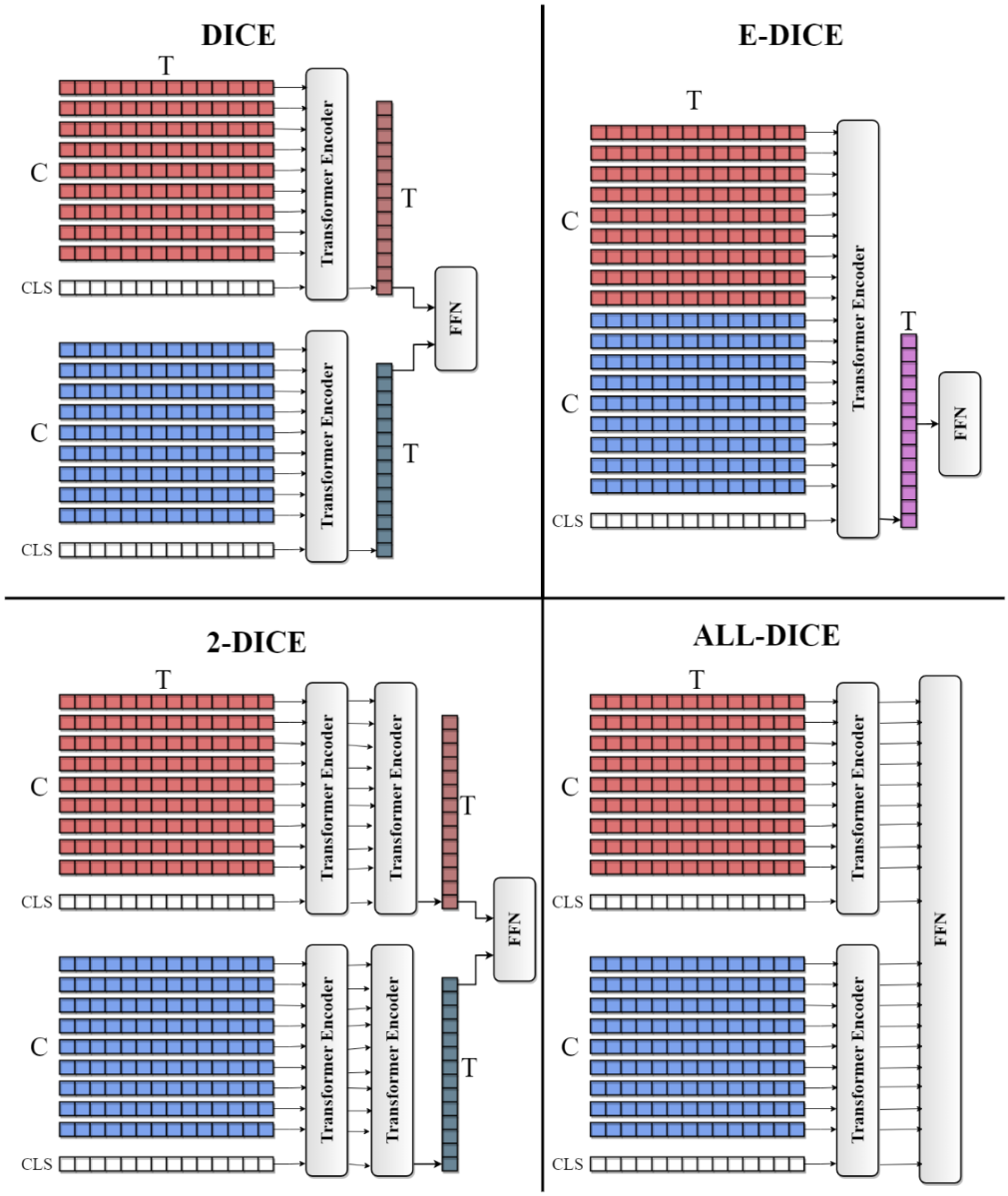


Figure 27 Different ablation configurations that were employed during DICE-net testing.

Table 12 Computational complexity of DICE-net and of its ablation experiments.

Model	N_params(M)	FLOPs(G)
DICE-net	170.5	137.4
ALL-DICE	368.6	140.6
NO-TRANS	18.8	1.42
2-DICE	338.7	274
E-DICE	163.5	133.9
M-CLS	170.5	137.4
ALL-E-DICE	357.6	130

5.1.4.2. Results

Firstly, the importance of the extracted features was evaluated by visualizing it across each group. The importance of the RBP feature is presented in Figure 28. In the left image, the PSD of 3 subjects across all electrodes is presented. It can be observed that the healthy participant exhibits increased alpha activity (top plot), whereas, as the disease progresses, the alpha activity decreases (middle and bottom plot). Furthermore, in the right image significant changes between the brain activity regarding the PSD can be observed as averaged across the different groups. All the colormaps have the same range equal to 7 uV^2/Hz. Examining the heatmaps, AD-CN discrimination seems much easier than FTD-CN discrimination.

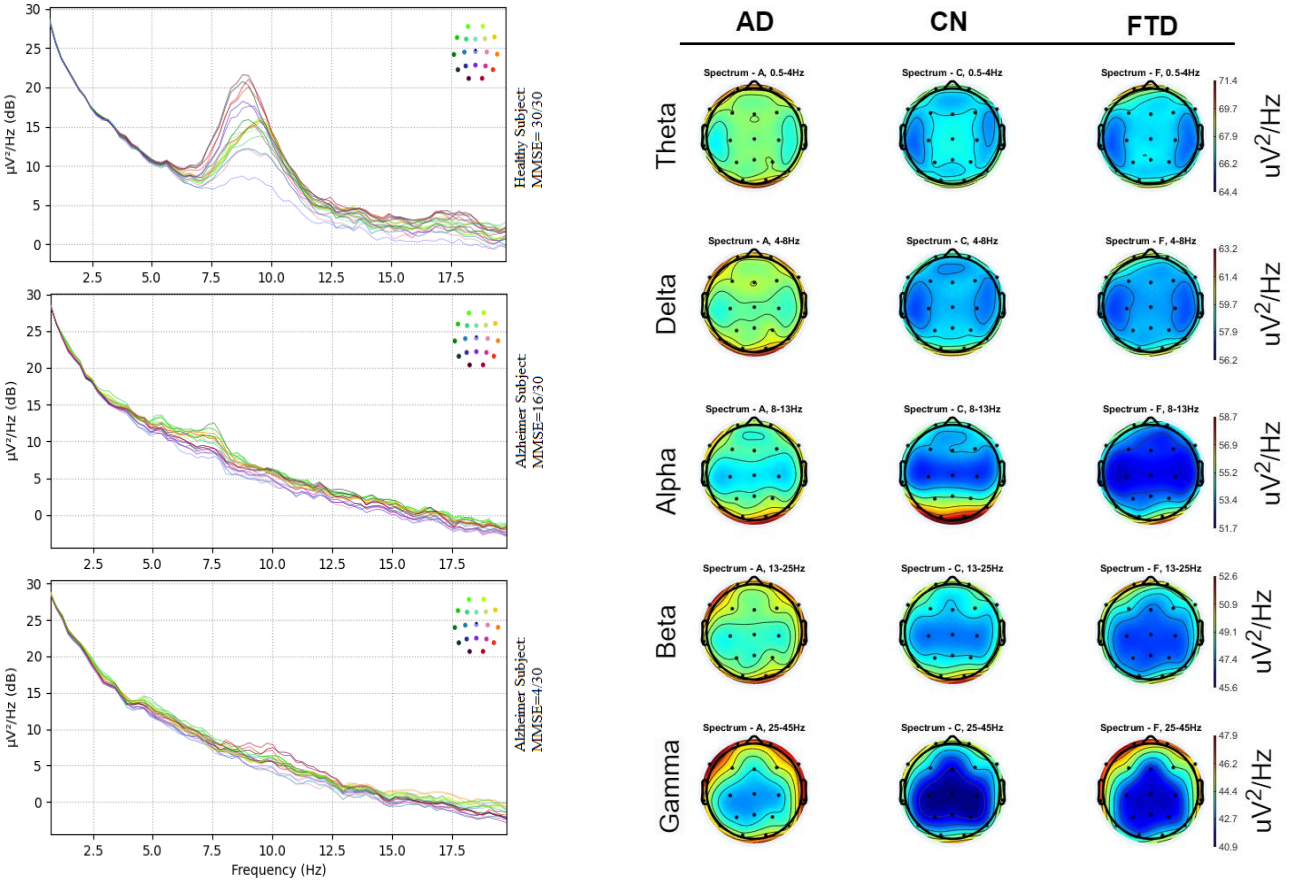


Figure 28 a) (left) PSD of a severe AD case (bottom), a mild AD case (middle) and a healthy subject (top). b) (right) scalp heatmaps of PSD across 5 frequency bands, averaged across groups AD, CN, FTD.

Regarding the SCC feature, the spectral coherence of every electrode X with every electrode Y for each frequency band is calculated and is presented in Figure 29(a). Changes in spectral coherence between electrodes are observable, especially in the Delta and Theta bands. Similarly, in Figure 29(b) the SCC feature averaged across the electrodes is presented. Each row represents the average SCC of one electrode with all the other electrodes, and each column represents the different frequency bands. Again, significant differences can be

observed in the Theta band. These findings are supported by the existing literature [89] and further signify the importance of using the SCC as feature.

Regarding the size of the dataset in terms of sets of 3-dimensional matrixes, the AD group consisted of 953, the FTD of 541 and the CN of 788.

A series of ablation studies were carried out and various hyperparameters assessed to identify the most effective methodology. The effectiveness of different ablation experiments was compared based on the LOSO accuracy. Additionally, to determine the optimal number of epochs for each experiment, a train-validation split was used in conjunction with LOSO validation. Specifically, the dataset was iteratively divided into 2 groups of participants, one group having recordings of P participants, P being an integer so that it approximates the $1/6^{th}$ of the dataset. These P participants served as the validation set, while the remaining 5/6 of the dataset were assessed using LOSO. Early stopping was implemented to prevent overfitting; For the statistical validation of the performance metrics, each model was trained 10 times. The performance of the proposed DICE-net demonstrated a statistically significant improvement (independent samples t-test, p-value < 0.05) compared to other methods across most metrics. In Table 13,Table 14,Table 15, a star symbol (*) denotes a statistically significant difference (independent samples t-test, p-value < 0.05) in a given metric relative to DICE-net.

Table 13 Performance of the DICE-net methodology and the ablation studies for the AD-CN problem.

AD/CN	ACC	SENS	SPEC	PREC	F1
NO-TRANS	79.12% *	79.87%	78.29% *	80.46% *	80.17% *
E-DICE	80.75% *	76.49% *	85.91%	86.78%	81.31% *
2-DICE	80.41% *	74.39% *	87.69%	87.35%	80.61% *
M-CLS	80.70% *	82.58%	78.42% *	82.23% *	82.40% *
ALL-DICE	78.00% *	79.32%	76.39% *	80.25% *	79.78% *
ALL-E-DICE	78.84% *	80.14%	77.25% *	81.01% *	80.22% *
DICE-net	**83.28%**	**79.81%**	**87.94%**	**88.94%**	**84.12%**

Table 14 Performance of the DICE-net methodology and the comparison algorithms for the AD-CN problem.

AD/CN	ACC	SENS	SPEC	PREC	F1
LightGBM	76.28% *	76.08% *	76.52% *	79.67% *	77.83% *
XGBoost	75.53% *	76.08% *	74.87% *	78.55% *	77.29% *
CatBoost	75.39% *	75.50% *	75.25% *	76.68% *	77.05% *
SVM+PCA	73.75% *	71.51% *	76.46% *	78.60% *	74.89% *
PCA-kNN	72.52% *	70.30% *	75.19% *	77.41% *	73.69% *

MLP	73.69% *	72.98% *	74.81% *	77.80% *	75.31% *
DICE-net	**83.28%**	**79.81%**	**87.94%**	**88.94%**	**84.12%**

Table 15 Performance of the DICE-net methodology and the comparison algorithms for the FTD-CN problem.

FTD/CN	ACC	SENS	SPEC	PREC	F1
LightGBM	69.13% *	51.57% *	81.54%	65.72%	57.79% *
XGBoost	69.22% *	52.02% *	81.73%	65.71%	57.44% *
CatBoost	68.66% *	47.41% *	83.25%	66.02%	55.19% *
SVM+PCA	70.93% *	45.85% *	86.21%	75.26%	56.98% *
PCA-kNN	67.80% *	41.50% *	85.85%	66.82%	51.20% *
MLP	69.98% *	53.60% *	81.22%	66.21%	59.24% *
DICE-net	**74.96%**	**60.62%**	**78.63%**	**64.01%**	**62.27%**

Furthermore, the ROC curves and the Area Under Curve (AOC) of the DICE-net and the comparison algorithms is presented in Figure 30. Also, Figure 31 is a visualization of the classification performance in regard to the individual predictions of the classifiers. Each dot represents the accuracy for one subject, while the color of the dot represents the class of the subject. A larger area on the upper part of the graph is observed for the DICE-net, in comparison to the other algorithms, and the misclassified subjects are less, supporting the superiority of this methodology.

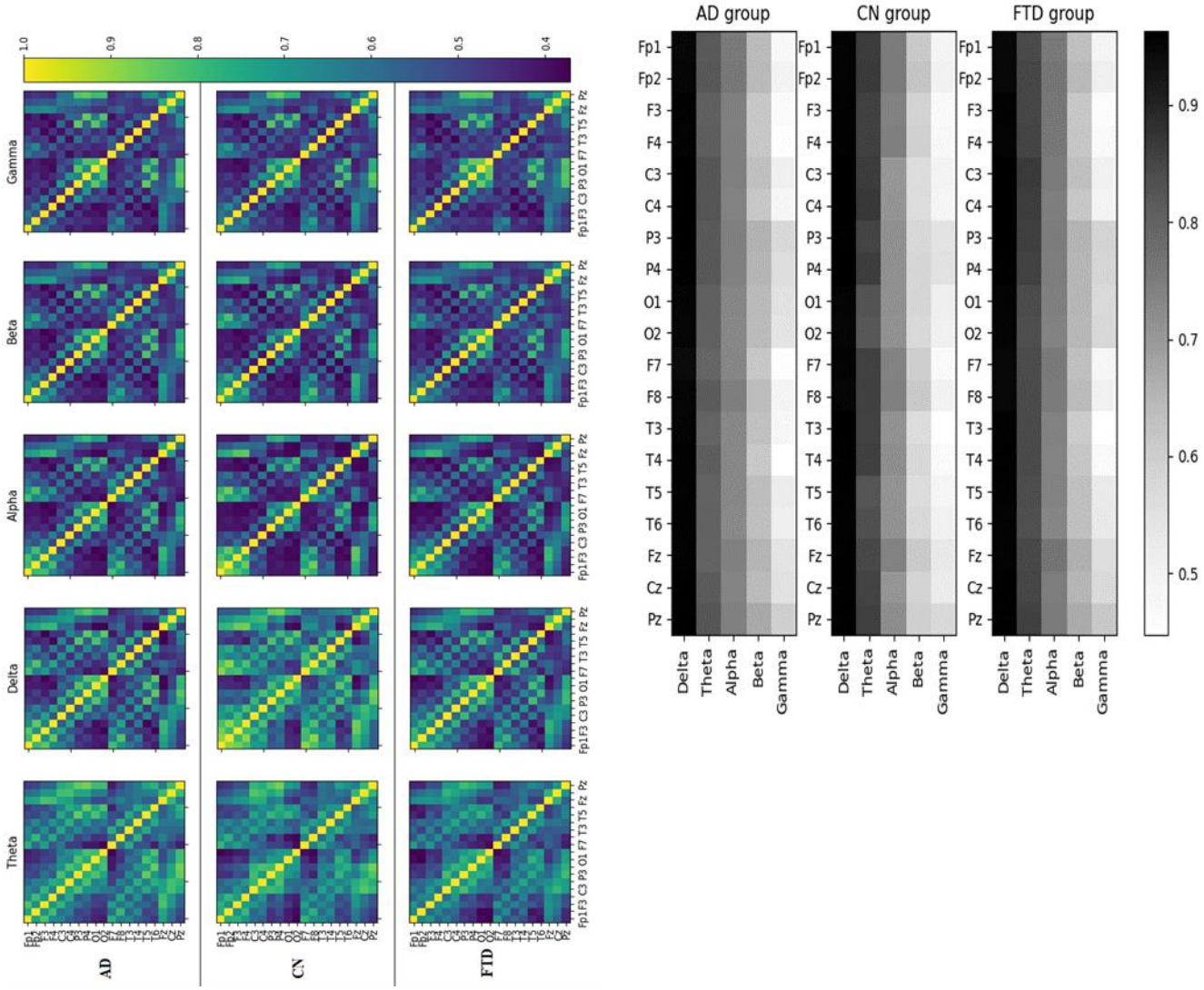


Figure 29 a) (left) Spectral Coherence Correlation (SCC) heatmaps for frequency band of each group (AD, CN, FTD). Each cell (X,Y) represents the spectral correlation of electrode X with electrode Y, averaged across each group. b) (right) SCC averaged across electrodes

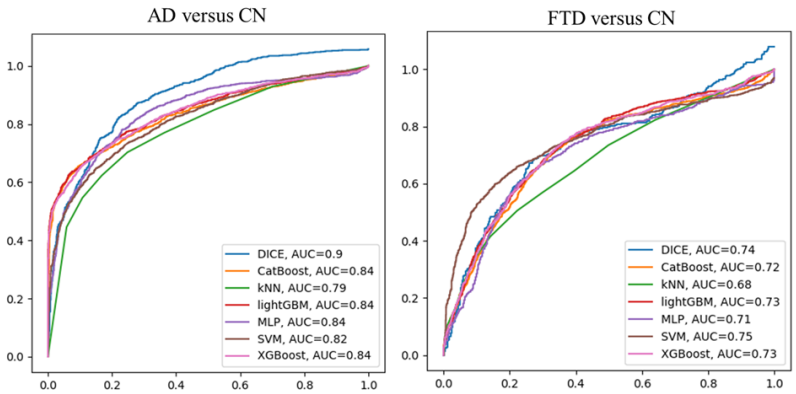


Figure 30 ROC curves of DICE-net and of the rest of the comparison algorithms for the AD-CN and FTD-CN problems

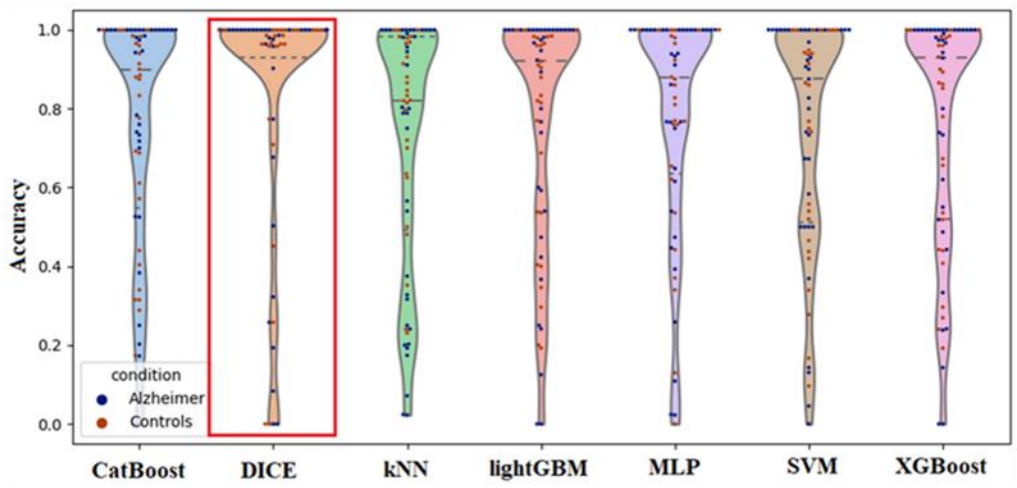


Figure 31 Violin plots representing the distribution of the accuracies for each individual prediction. The width of the plot indicates the density of the scores at each value, while individual dots represent a single classification performance for one subject.

Except from these comparison algorithms, other state-of-the-art deep learning architectures that were specifically designed for EEG signal classification were also employed, as mentioned previously. All these models take as input the raw EEG signal and do not require a feature vector. To train those, 4 second epochs with 2 second overlap were used and the sampling rate was set to 128 Hz. The smaller window length was chosen, since these algorithms input being raw signal leads to considerably larger size. Even though more than 200 epochs (in the concept of neural network epochs, meaning times that the same input is fed to the network) was used and the training accuracy was over 95%, none of these algorithms manage to classify these instances correctly in neither problem. The performance metrics of the algorithms are presented in Table 16

Table 16 Performance metrics of the state-of-the-art EEG based deep learning methodologies on the AD/CN and FTD/CN problems.

AD/CN	**ACC**	**SENS**	**SPEC**	**PREC**	**F1**
EEGNet	41%	47.20%	37.67%	37.89%	42.04%
EEGNetSSVEP	51.46%	56.78%	45.39%	47.65%	51.82%
DeepConvNet	54.21%	45.43%	57.59%	48.71%	47.01%
ShallowConvNet	42.18%	46.50%	41.11%	49.74%	48.07%
FTD/CN	**ACC**	**SENS**	**SPEC**	**PREC**	**F1**
EEGNet	46%	42.20%	57.46%	45.21%	43.65%
EEGNetSSVEP	61.46%	53.51%	75.00%	51.40%	52.43%
DeepConvNet	64.21%	62.41%	37.05%	58.14%	60.20%
ShallowConvNet	46.38%	42.58%	53.21%	42.37%	42.47%

Furthermore, this study was further elaborated by exploring which channels, and thus which brain areas were most significant for the discrimination of AD-CN and FTD-CN. In order to do so, the magnitude of the absolute value of the convolution layer weights was examined, under the assumption that larger absolute kernel weights represent higher importance in the classification, due to the increased attention given to them during the backpropagation process. To examine this, 2-dimensional heatmaps of the scalps were created, in which the color indicates the absolute magnitude of the weights. Hotter colors indicate higher absolute magnitude and thus higher importance for the classification. Values are normalized in the 0-1 range, and blue represents 0 while 1 represents red. It should be noted that blue does not represent lack of importance in the classification, it just represents that this value is at the lower end of the importance spectrum. It is easily observed that the RBP feature is more important than the SCC feature, in both classification problems. Also, regarding the AD-CN classification, the most important electrodes seem to be the T5, O1, O2, T4, F8, mainly on the temporal and occipital lobe. Respectively, for the FTD-CN classification the Fp1, Fp2, T3, T4 electrodes on the frontal and temporal lobe were given the most attention from the model. Both of these observations regarding the model's focus on different brain areas are in alignment with the established knowledge from the literature, further validating that the model indeed concentrates on the correct information. These scalp heatmaps are presented in Figure 32.

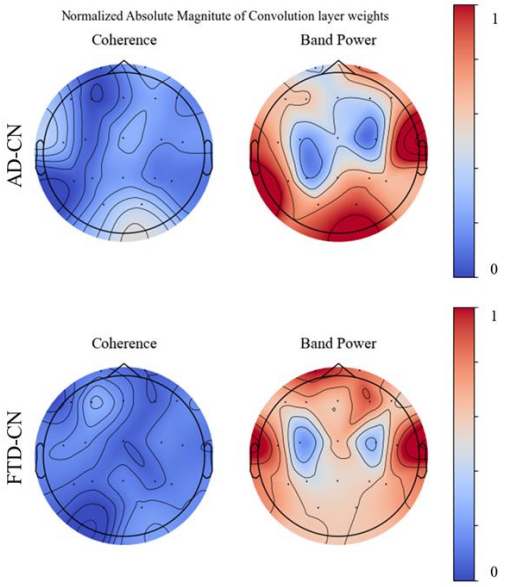


Figure 32 Normalized Absolute Magnitude of Convolution layer weights. (top): AD-CN. (bottom): FTD-CN classification.

5.1.4.3. Discussion

There have been multiple studies that address the problem of detection of different types of dementia in EEG signal using machine learning and deep learning

methodologies. The most advanced of them usually propose a Deep Learning scheme that gets as input time-frequency data through a transform like the Wavelet Transform [90] and then employ a Convolution layer for dimensionality reduction and extraction of the useful information, or they use Neural Network architectures such as autoencoders [91]. The novelty of this study is that it used a Transformer Network, which is exceptionally good in dealing with long-range dependencies and recognize patterns in sequenced data. EEG signals are often highly correlated over long time intervals, and capturing these correlations is important for high classification performance, thus Transformers possess a significant advantage over traditional CNNs. However, until the publication of this study, only one study has proposed a transformer encoder on a Raw-EEG framework for Mild Cognitive Impairment (which is the prodromic state of AD) detection [92]. Other than this, a combination of Convolution-Transformer architecture in EEG data has been proposed in an emotion recognition task with promising results [75].

Transformers offer several advantages over previously used deep learning architectures, like CNNs or RNNs, the main one being the attention mechanism, which permits the model to dynamically attend to the most relevant part of the data. Furthermore, Transformers have been found to outperform their counterparts given a large enough dataset, thus regarding EEG in the medical domain where accurate predictions are important, transformer architectures could be a solution, since they can take advantage of huge datasets. Additionally, the rationale behind choosing a transformer architecture for EEG-based Alzheimer's detection lies on two main considerations: first, the lack of existing methodologies that leverage transformers for EEG detection, presenting an untapped opportunity for groundbreaking advancements; and second, the intrinsic capability of transformers to proficiently handle long-range dependencies, a feature that is particularly suited for the complex temporal relationships found in EEG signals.

To take advantage of the Transformer Encoder’s abilities, a correct form of transformation of the EEG signal should have been made, so that it is similar to detecting dependencies of different sequences of words in a sentence, just like in Natural Language Processing. Already, modifications such as the Vision Transformer (ViT), that propose such transformations have been made for image classification tasks. In the vision transformer the image is split into patches, and a positional encoding layer adds sequence information and a CLS token is appended to learn the semantics of all the other “words” that represent the image. In the same concept, we attempted to create a 3-dimensional like input like an RGB image, instead of using the conventional “1 row – 1 sample” features. Finally, we took advantage of the abilities of a convolution layer to reduce the dimensionality of the input and transform it into a vector that looks like an NLP fed sequence of words.

Regarding the evaluation of the convolution weights to find out which brain areas were most important for the classification, we noticed that the RBP feature was more important than the SCC, and this might be the case since the lack of synchronization in

the brain is a finding only in late stages of dementia, while alpha-theta alterations are more easily detectable. Literature states that AD primarily affects the hippocampus, amygdala, and neocortex regions [93], and the focus of the model was primarily located on the occipital, temporal and frontal regions of the brain. However, the exact location of the altered brain activity is not easily detectable in EEG, even with source localization techniques. Further and different type of analysis should take place in order to determine source localization specifics.

Regarding the studying of previous work on AD detection with EEG signals, most studies perform their own signal acquisition and a minority of them uses published databases. Moreover, some use a different approach, not examining resting state recordings but rather Event Related Potentials (ERPs) or stimuli-based setups [94]. The reported accuracy of such studies lies between 70-85%. However, on FTD detection, even less studies have been published, with no one of them (except our first published study [13]) relying only on EEG signals (all used a biomarker combination such as EEG+MRI). Thus, the comparison of related works can be found in Table 17.

Table 17 Comparison of related works, as performed in DICE-net publication.

Study	Year	Cohorts	Stimuli	Methodology	Performance
Safi et al. [95]	2021	30 AD 35 CN	N/A	Entropy, Hjorth Parameters, SVM	ACC=81%, SENS=69.8%, SPEC=83.5%
Khatun et al. [94]	2019	8 MCI 15 CN	Auditory	ERP, SVM	ACC=87.9%, SENS=84.8%, SPEC=95%
Dogan et al. [96]	2022	12 AD 11 CN	Resting State	Graph-Based Feature extraction, Tunable Q-Wavelet Transform, kNN	ACC=92.01%, SENS=97.75%, SPEC= 84.03%
Miltiadous et al.[13]	2021	10 AD 8 CN	Resting State	Spectral & Temporal & Nonlinear Features, Random Forests	ACC=78.85%, SENS=82.4%, SPEC=74%
Ruiz-Gomez et al. [97]	2018	74 (AD + MCI) 37 CN	Resting State	Spectral & Nonlinear features, MLP	ACC=78.43%, SENS=82.35%, SPEC=70.59%
Araujo et al. [98]	2022	11 AD 8 MCI 11 CN	Resting State	Nonlinear features, SVM	AD-CN ACC=81%, MCI-CN ACC=79%
Lopes et al. [74]	2023	34 AD 20 CN	Resting State	Modulation Spectrum, CNN, SVM	ACC=87.3% F1=84.6%
This work	2023	36 AD 29 CN	Resting State	RBP, SCC, Dual-Input-Convolutional-Encoder	ACC= 83.28%, SENS=78.81, SPEC=87.94%, F1=84.12%

Briefly mentioning the limitations of this research, the most significant one is the size of the dataset. A larger dataset size could allow more stacked encoders to be employed and more complex input signal representations to be learned from the classifier, by creating a hierarchy of representations. Each unique encoder could attend to a different level of abstraction. However, with the present dataset size, increasing the number of encoders could worsen generalization and robustness and make the classifier prone to overfitting.

Another limitation that could be addressed in a future version is the significance of the SCC feature. It is found to be less important than the RBP feature, and even though this is not inherently negative, further and more elaborate investigation regarding

different connectivity measures that better capture AD desynchronization characteristics could be examined.

Furthermore, all the recordings were obtained from a single medical center, yet for the model to be truly applicable in medical practice in would require broader and more rigorous validation, adhering to the CLAIM criteria [99] (Credibility, Legality, Affordability, Interpretability, Maintainability) that emphasize the need for external validation using larger datasets from multiple medical centers.

Summarizing, this study was the highlight of the dementia related research in the years of this doctorate. Its novelty lies in the idea of incorporating a Convolution-Transformer combination for EEG classification and was the first study to be examining our publicly available dataset. The achieved performance metrics, with an accuracy of 83.28% and an F1 score of 84.12%, indicate that the DICE-net model is proficient in detecting spectral and spatial patterns in EEG-derived feature vectors and in extracting relevant dependencies for classification tasks.

5.2. Epilepsy Detection with EEG based Machine Learning Research

Epilepsy is a neurological dysfunction that affects more than 70 million people worldwide. Its cause is abnormal electrical activity in cortical neurons known as seizures. Epileptic patients are usually treated with anti-epileptic drugs that reduce and/or prevent seizure occurrences. Although multiple imaging techniques such as MRI, fMRI and PET scans have been employed for epilepsy detection, the EEG still stands out as the primary diagnostic tool for identifying epileptiform discharges. Nevertheless, manually examining EEG data is a laborious procedure that requires skilled and experienced medical personnel. While the clinical symptoms of seizures are often distinct, neurologists face challenges in making differential diagnoses to avert misdiagnosis and improper treatment, such as in cases resembling narcolepsy.

Since 80’s, research studies are published that focus on automatically detecting epilepsy on EEG through Machine Learning, yet an exponential growth of the number of published documents have been observed during the last decade. Although epilepsy detection through EEG is generally less complex than Alzheimer's detection, and significant strides have been made in this area, it is not entirely a solved challenge. There is a continual influx of research focused on automatic epilepsy detection, each study aiming to contribute further to the existing body of knowledge. This sustained interest stems from the ongoing need to refine detection methodologies, improve diagnostic accuracy, and adapt these techniques to a wide range of clinical scenarios. The distinct patterns of epileptic seizures in EEG data, while more identifiable compared to the subtle complexities in Alzheimer's, still present challenges in terms of variability among individuals and seizure types. Consequently, researchers remain engaged in exploring new techniques, enhancing algorithmic precision, and seeking even marginal advancements that can enrich the field and offer more robust solutions in epilepsy diagnostics.

The proliferation of publicly available EEG databases has been a driving force behind the increasing number of studies aimed at identifying epilepsy. These databases provide researchers with accessible data and drive advances in methodology and algorithm development. However, these databases, including well-known ones like the Bonn database, come with their own limitations. For instance, the Bonn database, despite its diversity, poses challenges due to its inclusion of data from both intracranial and scalp electrodes, leading to significant discrepancies in signal characteristics (Z, O, N, F, S). Moreover, other databases such as CHB-MIT, although comprehensive, can face problems related to variability in recording conditions and limited representation of seizure types. These issues highlight the need for critical evaluation of existing databases. Therefore, the systematic review we published fills this gap by carefully analyzing different EEG databases and exploring their strengths and limitations. This comprehensive review aims to help researchers select the best dataset for their studies and ensure a more informed and effective approach to EEG-based epilepsy research.

Furthermore, the need for systematic organization of the existing literature is further supported by the vast number of studies in this area that render the proposition of an unconventional methodology difficult as it would require extensive literature exploration.

During this doctorate research, 2 studies have been published regarding EEG epilepsy detection. In the first study, with the role of the second author, the objective was to identify the optimal length of time-window division for training classifiers in automatic seizure detection. In the second study, with the role of the first author, the objective was to organize all the published literature of epilepsy automated detection methodologies with EEG during the last 5 years, in a systematic review, in order to facilitate as a starting point for future researchers. The focus of the review was both technical, keeping track of the different frequency transformations, feature extraction and machine learning employed from the studies, and also database-oriented, focusing on the analysis of the used databases.

5.2.1. Evaluating the Window Size's Role in Automatic EEG Epilepsy Detection

Vasileios Christou[1], Andreas Miltiadous[1], Ioannis Tsoulos[1], Evaggelos Karvounis[1], Katerina D. Tzimourta[2], Markos G. Tsipouras[2], Nikolaos Anastasopoulos[3], Alexandros T. Tzallas[1], Nikolaos Giannakeas[1].

A huge number of studies have been proposed to automatically diagnose epilepsy from EEG recordings. Almost every of these studies splits the signals using an epoching mechanism to increase the population of the dataset. The length of these splits is usually not elaborately determined but rather arbitrary selected or selected through trial of 2-3 different window sizes. The goal of this study was to determine the optimal time-window length for epilepsy detection, so, all the integer window lengths between 1 second and 24 second were evaluated using 4 algorithms, to determine which window size performs the best. This study was published in the open access journal Sensors MDPI, during November 2022.

The algorithms used was a single-layer neural network with 10 hidden nodes, trained using 3 different approaches, and the k-NN classifier. The approaches for the training of the neural network were the Broyden-Flecher-Goldfarb-Shanno (BFGS) algorithm [100], the multistart algorithm [101], and a modified genetic algorithm [102].

5.2.2. Methodology

This research examined the impact of window size on classifying short-term epileptic EEG signals, focusing on the four machine learning methods that were mentioned before. The University of Bonn was used for the training and evaluation, since it is the most commonly used EEG database. This database consists of 5 groups of EEG recordings, namely Z, O, N, F, S. Each group has a different setting or condition. The Z and O groups are closed and open eyes recordings of non-epileptic, healthy participants. The N, F, S groups are intracranial EEG recordings from five epileptic patients during presurgical examination. Particularly, the N is interictal recordings originating from the opposite hemisphere of the epileptic zone, while F is from the epileptic zone. Finally, the S consists of 100 intracranial EEG's during epileptic activity. Each group consisted of 100 single channel recordings, each being 23.6 seconds in duration.

For the classification, all groups were used for a 5-class problem. Prior to the experiment, the signals were low pass FIR filtered to 40 Hz, and then epoched to all the different window lengths examined in this study. The Relative Band Power of each

---

The author list ordered as published was: Vasileios Christou, **Andreas Miltiadous**, Ioannis Tsoulos, Evaggelos Karvounis, Katerina D. Tzimourta, Markos G. Tsipouras, Nikolaos Anastasopoulos, Alexandros T. Tzallas, Nikolaos Giannakeas

frequency band was calculated through a Fast Fourier Transform. The bands were defined as:

1. Alpha: 8-12 Hz
2. Beta: 12-25 Hz
3. Gamma: 25-40 Hz
4. Delta: 1–4 Hz
5. Theta: 4-8 Hz

5.2.2.1. The BFGS method

The BFGS method is a quasi-Newton approach that is used to solve unconstrained optimization problems. In an unconstrained optimization problem, the objective is the minimization of a target function $f(x), x \in \mathbb{R}^n$, $\mathbb{R}^n$ denoting a n-dimensional Euclidean space, while: $f: \mathbb{R}^n \rightarrow \mathbb{R}$, is twice differentiable. The update formula is defined as:

$$\begin{aligned} s_k &\overset{\text{def}}{=} x_{k+1} - x_k \\ y_k &= g_{k+1} - g_k \end{aligned} \tag{5.9}$$

Where s_k and y_k are the step vectors, and g the gradient for minimizing $f(x)$.

The update formula takes the form:

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \tag{5.10}$$

Where B_k denotes the Hessian approximation at a point x_k, and the matrix B_{k+1} is generated so as to satisfy the following secant formula:

$$B_{k+1} s_k = y_k \tag{5.11}$$

To determine the f value, the quasi-Newton methods can be defined using the equation:

$$d_k = -B_k^{-1} g_k \tag{5.12}$$

The algorithm of the BFGS is defined as:

Algorithm: The BFGS Algorithm

1. Having a starting point x_0 and $B_0 = I_n$. Set the values for s, β, σ.
2. End when $\|g(x_{k+1})\| < 10^{-6}$
3. Find the search direction using eq. 5.12
4. Calculate the difference $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$
5. Update B_k by eq. 5.10 in order to obtain B_{k+1}
6. $k = k + 1$
7. Return to step 2

5.2.2.2. The multistart method

Genetic Algorithms are optimization methodologies that are based on Charles Darwin's evolution theory. They mimic natural selection, by beginning with a set of candidate solutions which are equivalent to the chromosomes of biological organisms. These chromosomes are changed evolutionarily in an iterative process using genetic operations such as selection, crossover and mutation. This is continued iteratively until a certain termination criterion is met, or an optimal or suboptimal solution to the problem has been found.

The Genetic Algorithm is defined as:

Algorithm: Genetic Algorithm

1. Create N random points in Ω from the uniform distribution, being set S
2. $iter = 0$
3. Evaluate each chromosome by the evaluation function
4. If termination criterion is achieved, stop
5. Select $m \leq N$ parents from S
6. Create m offspring using the parent chromosomes
7. Replace parents with offsprings
8. Create a trial point $\tilde{x}$. If $f(\tilde{x}) \leq f(x_h)$ where $x_h$is the current worst point in S, and replace x_h with $\tilde{x}$.
9. $iter = iter + 1$
10. Go to step 3

The difference of the multistart as proposed in [102] with the normal Genetic Algorithm is that it includes a new stopping rule, a new mutation operator and a local search procedure.

The flowchart of the proposed methodology can be examined in Figure 33.

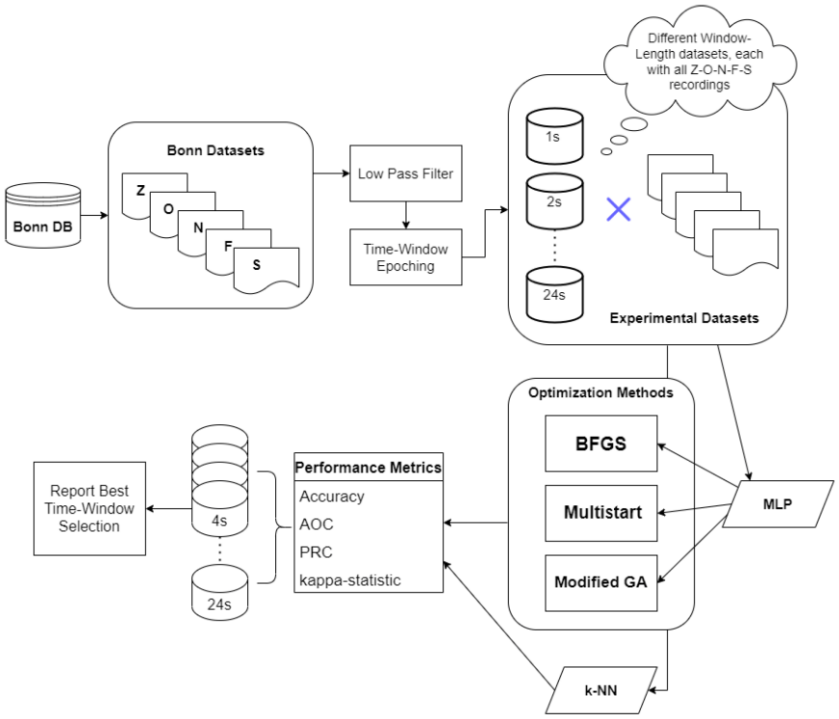


Figure 33 The flowchart of the proposed methodology in "Evaluating the Window Size's Role in Automatic EEG Epilepsy Detection"

5.2.3. Results

All the experiments for each window size were repeated 30 times, and the accuracy results that are presented are the average classification accuracy obtained from the 30 repeats. The performance results in terms of accuracy of the different algorithms are presented in Table 18. The performance results in terms of area under ROC, area under PRC and k-statistic are presented in Table 19.

Table 18 Performance results in terms of accuracy for the 4 comparison algorithms, in epoch length 1-24 second

Epoch (s)	BFGS	Multistart	GA	K-NN
1s	56.86%	57.68%	56.91%	68.9%
2 s	65.06%	65.56%	65.06%	75.14%
3 s	69.7%	69.57%	69.01%	76.66%
4 s	72.62%	70.53%	70.06%	76.99%
5 s	75.69%	73.46%	71.96%	77.89%
6 s	74.63%	76.37%	75.44%	79.53%
7 s	74.76%	75.84%	74.43%	79.1%

8 s	76.06%	75.55%	74.95%	78.41%
9 s	76.25%	77.64%	76.5%	79.88%
10 s	76.96%	77.12%	76.38%	80.05%
11 s	76.42%	79.01%	77.2%	79.08%
12 s	76.55%	78.26%	77.06%	79.84%
13 s	77.04%	78.04%	76.05%	78.56%
14 s	77.81%	78.26%	77.13%	79.01%
15 s	79.75%	78.98%	78.41%	78.68%
16 s	77.35%	80.98%	78.59%	79.52%
17 s	77.7%	78.05%	77.82%	79.92%
18 s	78.5%	79.24%	78.10%	79.92%
19 s	80.7%	79.71%	78.47%	79.49%
20 s	80.92%	81.59%	80.78%	80.00%
21 s	80.92%	81.23%	81.06%	79.25%
22 s	80.04%	80.88%	81.00%	81.17%
23 s	80.69%	80.88%	80.89%	78.88%
24 s	80.25%	80.43%	79.98%	79.04%

Table 19 Performance results in terms of Area under RO, area under PRC and k-statistic for the 4 comparison algorithms, in epoch length 1-24 second

Epoch (s)	AOC	PRC	k-Stat
1 S	78.91%	48.6%	62.21%
2 s	79.89%	50.2%	68.74%
3 S	80.68%	50.1%	75.23%
4s	86.44%	53.3%	71.95%
5 s	85.92%	56.8%	74.62%
6 s	85.45%	54.0%	76.38%
7 s	83.21%	58.1%	77.55%
8s	87.21%	60.9%	77.19%
9 s	87.17%	61.8%	80.02%
10 S	86.57%	64.3%	78.84%
11 S	90.89%	64.2%	83.40%
12 S	90.49%	64.8%	82.32%
13 S	89.04%	68.1%	82.14%
14 S	88.88%	68.3%	82.85%
15 S	86.22%	70.4%	79.94%
16 S	85.45%	70.1%	80.15%
17 S	85.92%	73.6%	82.15%
18 S	84.70%	73.0%	84.29%
19 S	86.07%	74.7%	85.42%
20 S	92.22%	78.5%	85.49%
21 S	92.51%	76.5%	83.26%
22 S	88.70%	77.3%	82.44%

23 S	82.28%	75.7%	83.51%
24 S	88.37%	73.7%	80.00%

As observed from the previous tables, the window size in epilepsy detection significantly impacts the classification performance. Larger window sizes, between 20-21 seconds are found to provide the best performance results. Selecting the right window length is essential for machine learning approaches dealing with signal data, like EEG. Time windows that are too brief might not adequately capture the distinct signal characteristics of each condition. For instance, in epilepsy detection, an excessively small time-window could miss encompassing the entire seizure waveform. Conversely, overly large time windows might encompass characteristics of two distinct states (like ictal and interictal states), which could adversely impact classification accuracy. This study's findings are valuable for future research proposing classification schemes in EEG-based epilepsy detection, providing insights into optimal window length selection.

5.2.4. Machine Learning Algorithms for Epilepsy Detection based on published EEG databases: A Systematic Review

This article was published in the open access journal IEEE Access, during January 2023.

Epilepsy is the condition with the most published studies regarding EEG detection, since the examination of prolonged timeseries of EEG data for visual detection of epileptic waveforms is a laborious endeavor. Thus, researchers are applying signal processing and machine learning algorithms attempting to detect abnormal spikes, spike waves and spike wave complexes in the interictal EEG recordings. Other than this, research efforts are being made into detecting early signs of seizure to predict future onset of seizures. Most of these studies gather EEG recordings from both epileptic and healthy individuals. They introduce methodologies that utilize a variety of signal processing and feature extraction techniques, combined with either traditional machine learning algorithms or more innovative ones. The goal is to develop a system capable of automatically determining whether an EEG time window is indicative of epilepsy. Research teams focused on processing, algorithmic, or computing problems have found it difficult to contribute to the field by themselves, as clinical records are required and thus the need to collaborate with healthcare institutions. Consequently, several epilepsy EEG databases have been made available online, aiding the creation of automated seizure detection protocols. Therefore, it is common for most experimental research to depend on these publicly accessible databases rather than conducting independent data collection. Thus, there was an imminent need for a systematic exploration of the published studies focusing both on the methodologies employed and the published database that is used for the validation of it.

Thousands of studies analyzing this topic have been published during the last decades, so the analysis of all of these would be difficult and out of scope and purpose. Thus, the systematic review that was published was focused only on the bibliographic data published on the last 5 years. The goal of this study was to summarize the EEG epilepsy detection research and provide guidance for future researchers regarding the selection of the database and the proposed machine learning pipeline.

5.2.4.1. Methodology

The entire methodology follows in detail the guidelines of the PRISMA protocol [103]. Also, it is registered on the PROSPERO and can be accessed via its PROSPERO id which is CRD42022365313. The search was limited to studies focusing exclusively

---

The author list ordered as published was: **Andreas Miltiadous**, Katerina D. Tzimourta, Nikolaos Giannakeas, Markos G. Tsipouras, Euripidis Glavas, Konstantinos Kalafatakis, Alexandros T. Tzallas

on machine learning methodologies for automatic epilepsy detection using EEG from published databases. The studies were collected from the most well-known electronic libraries/search engines, them being Elsevier's Scopus, IEEE Xplore, Elsevier's ScienceDirect and MEDLINE PubMed. For the literature review, the keywords that were used for inclusion were "EEG" and "epilepsy" or "seizure" and "detection" in the title or the summary or the keywords of the article were searched, while excluding studies that included the keywords "animal" or "mouse" or "mice" in their title or summary or keywords. The retrieval of the studies was carried out on the 1st of May 2022 and in total 3975 studies (including duplicates) were found from all 4 search engines. 1006 of them were conference papers so they were removed. From these studies, 1454 duplicates were found and 1272 papers and 43 theoretical studies were excluded using the Rayyan software, after appropriateness evaluation. At last, 190 studies published in the last 5 years were found fitting the review criteria and included in the systematic review.

The rejection criteria used during the appropriateness evaluation by 3 independent researchers, as mentioned in the published paper were:

1. Studies that study epilepsy in animals and were not excluded from the query in the first research (rats, pigs, dogs, sheep, mammals)
2. Studies that study epilepsy at a microscopic level (chromosomes/genes/proteins)
3. Studies that do not analyze merely EEG data (e.g. EEG, TMS-EEG, MEG, Positron emission tomography, MRI/CT scan, Electromyography, video-EEG)
4. Studies that study epilepsy as an association with other diseases or neurological conditions (e.g. encephalitis, schizophrenia, dementia, Autism Spectrum Disorder, brain disorders, trauma, bleeding, tumors, multiple sclerosis, psychogenic non epileptic seizures, heart conditions etc.)
5. Pharmacological studies that study the effect of antiepileptic drugs on EEG
6. Surgical treatment of epileptic seizures studies
7. Studies that study the implementation of devices and systems for the detection of seizures
8. Studies for seizure prediction
9. Case studies
10. Epilepsy studies on newborns and children
11. Studies that do not apply machine learning algorithms to EEG data (Socio-cultural aspects of epilepsy, differences between types of epilepsy, therapeutic approaches, keto diet, neurostimulation, quality of life and behavior-psychology assessment studies)
12. EEG analysis studies not in resting state (patients in a coma, hyperventilation, visual stimulus, sleep studies)
13. Studies that study the removal of noise/interferences from EEG

14. Inaccessible studies (invalid Digital Object Identifier, inability to find and/or obtain the study)
15. High Frequency Oscillations (HFO) analysis studies
16. Patient specific for seizure detection studies
17. Studies for source localization of seizures without performing detection
18. Studies on non-real clinical data (Surrogate/synthetic data)
19. Studies for EEG montage or sampling frequency in epilepsy
20. Studies for neuron connectivity in epilepsy
21. Studies that don't propose a certain methodology but focus on the comparison of existing algorithms

In order to perform the evaluation of the appropriateness of the studies, the Title, Abstract, Methodology and in some cases Discussion was examined by 3 different researchers that performed independently. If one of the mentioned criteria applied, the reviewer marked the study to be rejected. After the end of the evaluation of all studies, conflicts between the reviewers were marked automatically in the Rayyan suite and were resolved by having the reviewers come to an agreement.

When the exclusion process was finished, the 190 studies that were left to be included were extracted in .RIS format and transferred to the Mendeley Reference Manager environment. They were splitted into folders based on the published database that they were using, with the folders being: "Bonn Database", "CHB-MIT Database", "Freiburg Database", "Other Database", "Multiple Databases". Finally, the main aspects of each paper were reported in a data extraction sheet. These aspects were:

1. database used,
2. signal processing methodology,
3. feature extraction methodology,
4. classification methodology,
5. specific problem addressed (for example ictal-interictal),
6. performance metrics as reported in the study.

Next, the data synthesis of the systematic review was performed. For this step, the following metrics were extracted from the sum of the studies:

1. Popularity of each database
2. Popularity of type of signal transform (Frequency, Time-Frequency (TF), Time-domain, Non-Linear, Raw Signal)
3. Regarding only the Time-Frequency studies: Popularity of the specific TF decomposition.
4. Popularity of classification algorithm
5. Accuracy comparison for each classification problem
6. Comparison of the methodologies used in the first half of the 5 years and in the second half.

The flowchart of the study selection and elimination process, regarding the identification, screening, eligibility and inclusion evaluation can be observed in Figure 33.

5.2.4.2. Results

The creation of methodologies that differentiate the activity associated with epileptic seizures is made possible due to the availability of the several milestone databases that are publicly available.

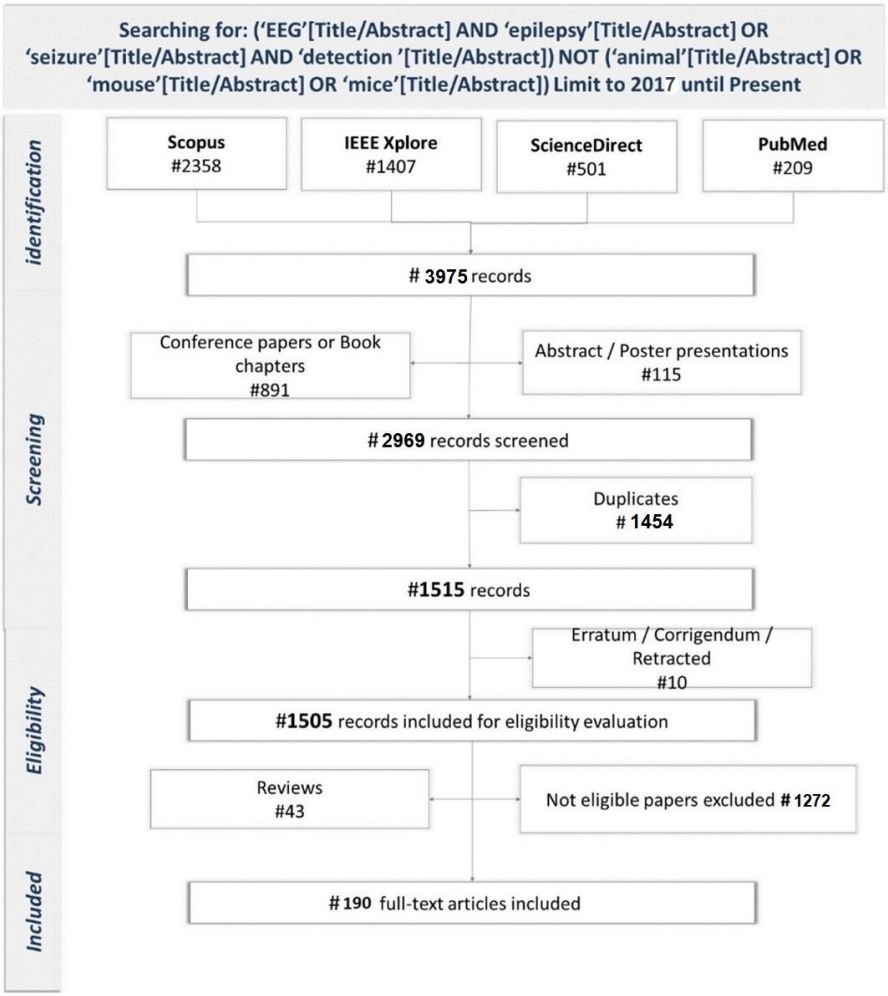


Figure 34 Flowchart of the study selection and elimination process for the systematic review.

(a) Bonn Database

The most well-known database (DB) is the Bonn DB [104], which contains short-term scalp and intracranial EEG recordings. The database consists of five subsets of EEGs which are distinguished by the capital letters A, B, C, D, E, referring to the signals starting with the letters Z, O, N, F, S. In this document, the Z, O, N, F, S annotations will be used. In this database, a total of 5 healthy subjects and 5 epileptic patients were recruited. Each group is comprised of 100 single-channel recordings, each lasting 23.6 seconds with a signal length of 4096 samples. The data has a sampling frequency of

173.61 Hz. Interferences such as muscle activity or eye movements were identified and eliminated by the database owners through visual inspection.

As already mentioned in a previous chapter, sets Z and O consist of scalp EEG from healthy volunteers, in a closed and open setting respectively. N, F and S consisted of intracranial EEG from epileptic patients during presurgical examination. Specifically, the N recordings were from interictal recordings from the opposite hemisphere of the epileptic zone, the F were from the same hemisphere and the S were during the epileptic activity.

Some crucial limitations of the Bonn database, that limits its appropriateness are the extremely limited size of it (3,5 hours in total), the limited number of electrodes and most importantly the existence of both scalp and intracranial recordings. Without a shadow of a doubt, the electrical activity of the brain captured from scalp recordings differs significantly from that recorded directly from exposed brain regions. Training algorithms on such varied data can lead to significant errors in the quality of the machine learning model and subsequently affect the development of the system.

(b) Children's Hospital of Boston Base – Massachusetts Institute of Technology (CHB-MIT)

The CHB-MIT database is the second most used EEG database [105]. It includes long continuous recordings from the scalp of 24 people with drug resistant seizures. Each participant has 9-42 consecutive .edf files with recorded interval of seizures. The sampling rate was 256 Hz with a 16-bit analog-to-digital converter. Totally, this database consists of 664 files of 140 EEG recordings and 198 seizures, and the demographic characteristics of the patients.

An issue that needs to be considered or regarded as a limitation is the fact that the database consists mostly of pediatric cases of many different age groups. It is known that epilepsy has different brain patterns during adulthood and during childhood, thus it could pose the risk of including significant bias towards performance of a classifier evaluated on these recordings.

(c) Epilepsy Center of University of Freiburg

This database includes continuous long-term EEG recordings from 21 patients (8 men, 13 women) with drug resistant focal epilepsy. Six intracranial channels are used for the recordings, 3 being focal and 3 non-focal, with a sampling rate of 256 Hz. The recordings are consisted of ictal, preictal and interictal activity, with at least a full day (24 hours) of continuous signal being recorded for each patient. Each patient may exhibit 2-5 seizures, making a total of 88 seizure events, 509 hours of non-seizure interictal activity and 199 hours of preictal or ictal activity. It is worth noticing that a clear distinction is made between ictal (including preictal) and interictal recordings by including at least 50 minutes of preictal activity before each seizure. The Freiburg Database is considered one of the most complete and accessible EEG databases and it

has been used extensively by researchers worldwide. However, it is not currently available, but it can be accessed through the EPILEPSIAE project [106].

(d) Other Databases

Except those 3 popular databases, there are many more epileptic EEG databases, either openly accessible, such as Bern-Barcelona, Temple and other databases, or not published datasets that are used in other studies.

The Bern-Barcelona database was published in 2012 partially by the same team of researchers that published Bonn database and is consisted of 7500 intracranial recordings from 5 subjects with drug-resistant temporal epilepsy. Each recording is 20 second long and is originated from 2 electrodes sampled at 512 Hz.

Another database is the Temple University Hospital database (TUH-EEG) [107], which is actively renewed since now and contains the most long-time recordings of pathological EEG, thus being the most suitable for deep learning methodologies. In total, 16.986 recordings from 10.874 subjects (epileptic and non epileptic) are in the database. All the signals are taken from 19 electrodes with 256 Hz sampling rate. Since the number of subjects in the database is subject to change, it is worth mentioning that during 2018, 315 patients with 10 different types of epileptic seizures were included in the database.

Other, not so popular databases worth mentioning are:

- Neurology and Sleep Center, New Delhi (NCS)
- Peking University People's Hospital (PUPH)
- Institute of Neuroscience, Ramaiah Memorial Hospital, India (RMCH)
- Department of Neurology, Epilepsy Center, Zhejiang University (INeuro)
- KU Leuven dataset
- All India Institutes of Medical Sciences (AIIMS)
- Department of Clinical Neurophysiology, Maastricht (MUMC)
- Karunya Institute of Technology and Sciences (KITS)
- pone_pat dataset
- European Epilepsy DB (EPILEPSIAE)

In summary, the NCS database includes EEG signals from 10 epileptic subjects, recorded using a 16-electrode system at a 200 Hz sampling rate, categorized into ictal, interictal, and preictal states. The PUPH database contains EEG recordings from 7 epileptic subjects at a 256 Hz sampling rate. The KU Leuven database features EEG data from 22 subjects with 22 electrodes, covering ictal, interictal, and preictal phases. AIIMS database comprises 20-minute EEG recordings from 13 epilepsy patients using a 32-channel system at 256 Hz. MUMC database includes 40 routine EEG recordings from an intensive care unit, utilizing a 19-electrode setup at 250 Hz. KITS database features 258 EEG signals of normal, generalized, and focal epilepsy, recorded with a 16-channel system at 256 Hz. Pone_pat database offers EEG and ECG data from 15 epileptic patients, recorded at a 512 Hz sampling rate. Lastly, the EPILEPSIAE database

holds scalp EEG recordings from 217 patients and intracranial EEG from 58 patients, collected from three epilepsy centers during long-term presurgical monitoring.

(e) Database Combination

Several studies examined their proposed methodology with more than one database, to validate its robustness and ensure that the methodology is not prone to bias due to specific recording properties. Usually, the methodologies that examine a combination of databases do examine two of the three most known databases (Bonn, CHB-MIT, Freiburg), along with the NSC database. Other than that, researchers may examine their proposed methodology to a combination of one or more published databases and a custom database (from their own clinical environment).

A detailed analysis of each study can be found in the published document [9]. However, the tables below contain all the examined studies along with the categorization analysis that has been performed for the systematic review. Specifically, each paper is analyzed by its publication year, its signal transform methodology, its feature extraction methodology, and the classification methodology. Furthermore, the classification problem examined along with the performance metrics achieved are mentioned. The studies are divided in tables based on their database of usage and based on the date of publication (2017-2019 or 2020-2022). Specifically, Table 20 and Table 21 contain the Bonn database studies from 2017-2019 and 2020-2022 respectively. Table 23 and Table 22 contain the CHB-MIT studies. Table 24 contains the studies that used other databases and Table 25 contains the studies that used a combination of databases.

Table 20 Studies that used Bonn database published in 2017-2019

Author	Year	Signal Transform	Feature Extraction	Classification	Classification Problem	Results
Akut et al.[108]	2019	Time-Frequency	DWT	CNN	ZO-NF-S	ACC=99.4%, SENS=98.5%, SPEC=99.45%
Alzami et al.[109]	2019	Time-Frequency	DWT	Adaptive Hybrid Feature Sel.-Based Classifier	ZO-NF-S	ACC=96%, SENS=96.53%, SPEC=98.93%
Attia et al.[110]	2019	Non-Linear	Autoregressive Model, Firefly optimazation	SVM	Z-N Z-F O-N	ACC=96%, ACC=95%, ACC=94%
Bandil and Wadhwani [111]	2019	Time-Frequency	DWT, Entropy features	ANN	Z-F-S	ACC=99%
Chen et al. [112]	2019	Time-Frequency	DWT, Entropy features	LS-SVM	F-S	ACC=99.5%, SENS=100%, SPEC=99.4%
Chiang et al. [113]	2019	Time-Frequency	DWT, Fuzzy entropy	Petri-Net	ZO-NFS	ACC=93.8%
Gupta and Pachori [114]	2019	Frequency	DFT, Renyi entropy	LS-SVM	ZONF-S Z-N-S	ACC=98.6%, ACC=97.3%
Hussein et al. [115]	2019	Raw Signal		LSTM	Z-O-N-F-S	ACC=100%
Li et al. [116]	2019	Time-Frequency	STFT spectogram, scalogram	SVM	ZO-NF-S	ACC=99.6%, SENS=99.33%, SPEC=100%

Mahjoub et al. [117]	2019	Time-Frequency	EMD	SVM	ZONF-S	ACC=97%, SENS=90.62%, SPEC=98.59%
Zhao et al. [118]	2019	Time-Frequency	Stationary WT, Entropy features	Back-Propagation NN	ZO-NF-S	ACC=93.3%, SENS=96.67%, SPEC=96.67%
Singh et al. [119]	2019	Time-Frequency	EMD	Ensemble classifier with grasshopper optimization	ZO-NF-S	ACC=99.2%
Tsipouras et al. [120]	2019	Time-Frequency	DWT, Entropy features	Random Forests	Z-O-N-F-S ZO-NF-S	ACC=91.2%, ACC=98.8%
Wang et al. [121]	2019	Time-Frequency	DWT	Gradient Boosting with Grid Search optimizer	ZONF-S ZO-NF-S	ACC=98.4%, ACC=96.5%
Yang et al. [122]	2019	Time-Frequency	Intrinsic Time-scale Decomposition	ANN	ZO-NF-S Z-N-S	ACC=99.5%, ACC=99.67%
Bose et al. [123]	2018	Non-Linear	multifractal detrended fluctuation analysis	SVM	ZO-NF	ACC=96.25%, SENS=95.44%, SPEC=97.06%
Choubey et al. [124]	2018	Frequency	FFT	KNN	ZF-S	ACC=97%, SENS=98%, SPEC=97%
Kabir et al. [125]	2018	Time domain	K-means clustering, statistical features	SVM	Z-S, O-S	ACC=98.13%, ACC=97.75%
Lahmiri et al. [126]	2018	Non-Linear	multifractal detrended fluctuation analysis, Hurst exponent	KNN	NF-S	ACC=100%
Li et al. [127]	2018	Non-Linear	Fuzzy entropy, Dispersion Entropy	Quadratic Discriminant Classifier	ZONF-S	ACC=92.8%, SENS=90.67%, SPEC=96%
Mert and Akan [128]		Time-Frequency	EMD	Not reported	Z-S, F-S, O-S	ACC=97.89%, ACC=93%, ACC=83.68%
Saini et al. [129]	2018	Time domain	Statistical and Entropy Features	ANN	Z-F-S	ACC=99.3%, SENS=99.02%, SPEC=99.34%
Sharaf et al. [130]	2018	Time-Frequency	TQWT, chaotic features	Firefly optimization, Random Forest	ZO-NF-S	ACC=99%, SENS=98%, SPEC=97%
Sharma et al. [131]	2018	Time-Frequency	EMD, Entropy	Random Forests	ZO-NF-S	ACC=98%, SENS=97.8%, SPEC=99%
Sharmila and Geethanjali [132]	2018	Time Domain	Slope sign changes and Statistical features	SVM	ZONF-S	ACC=95.15%, SENS=81.23%, SPEC=96.6%
Acharya et al. [133]	2017	Time Domain	Z-score normalization	CNN	O-F-S	ACC=88.67%, SENS=95%, SPEC=90%
Amorim et al. [134]	2017	Time-Frequency	DWT, Shearlet & Contourlet transforms	Random Forests	Z-O-N-F-S, Z-S	ACC=88.67%, ACC=100%
Biju et al. [135]	2017	Time-Frequency	EMD	Adaptive Neuro-Fuzzy Neural Network	ZO-S	ACC=100%
Chatterjee et al. [136], [137]	2017	Time-Frequency	Stockwell Trasform	SVM, kNN	Z-S, F-S	ACC=100%, ACC=99.25%
Jaiswal et al. [137]	2017	Time Domain	GModPCA	SVM	ZO-S NF-S ZONF-S	ACC=99.66%, ACC=95.8%, ACC=97.17%
Jaiswal et al. [138]	2018	Time Domain	SubXPCA	SVM	Z-N-F, ZO-NF-S, Z-O-N-F-S	ACC=97.2%, ACC=97.43%, ACC=94.6%
Liu et al. [139]	2017	Time-Frequency	WPD, energy, entropy, kurtosis	ELM	Z-S, Z-F-S	ACC=97.7%, ACC=96.5%
Li et al. [140]	2017	Frequency	PSD, Autoregressive Model	SVM	NF-S	ACC=98.73%, SENS=98%, SPEC=99.1%

Mohammadpoory et al. [141]	2017	Non-Linear	Weighted Visibility Graph Entropy	Decision Trees	Z-F-S	ACC=97%
Sharma et al. [142]	2017	Time-Frequency	TQWT	LS-SVM	ZO-S, ZO-NF, ZONF-S	ACC=99.67% ACC=98.5% ACC=99.6%
Sharmila et al. [143]	2018	Time-Frequency	DWT, shannon entropy	SVM	ZONF-S, NF-S	ACC=78%, ACC=88%
Sharmila et al. [144]	2017	Time-Frequency	DWT	kNN	Z-S	ACC=100%

Table 21 Studies that used Bonn database, published in 2020-2022

Authors	Year	Signal Transform	Feature Extraction	Classification	Classification Problem	Results
Al-Hadeethi et al. [145]	2020	Time domain	Statistical Features	adaboost LS-SVM	All combinations	ACC=99%
Aliyu et al. [146]	2021	Time-Frequency	DWT	LSTM	ZONF-S	ACC=99%
Amin et al. [147]	2020	Time-Frequency	DWT + Arithmetic Coding	SVM, KNN, MLP	Z-S, ZONF-S	ACC=100%
Anuragi et al. [148]	2022	Time-Frequency	EWT	Extra Trees	Z-N-S, ZO-NF-S	ACC=99.33%, ACC=97.8%,
Arı et al. [149]	2020	Time-Frequency	WPD + dispersion entropy	SVM	Z-O-N-F-S	ACC=99,53%
Ashokkumar et al. [150]	2021	Time-Frequency	DWT, Fractional S-transform, Entropy	Deep CNN	NF-S	ACC= 99.7% , SENS = 97.71%, SPEC= 98.7%
Ashokkumar et al. [151]	2020	Time-Frequency	Optimal equilateral wavelet filter bank, Fuzzy, Renyi and Kraskov entropy	Gaussian SVM	ZONF-S NF-S	ACC=99.4% ACC=98.6%
Ashokkumar et al. [152]	2020	Time-Frequency	Q-Tuned Wavelet Transform, Approximate entropy	Extreme learning adaptive neuro-Fuzzy Inference System	ZO-NF-S	ACC=99.72%
Bari et al. [153]	2020	Time-Frequency	EMD with normalized Intristic Mode Function	Quadratic discriminant analysis (QDA)	NF-S	ACC=99%, SENS=98.5%, SPEC=100%
Baykara et al. [154]	2021	Time-Frequency	Stockwell Transform, entropies and Perservals energy	ELM	ZO-NF-S	ACC=90%, SENS=95%, SPEC=82%
Brari et al. [155]	2021	Non-Linear	Higuchi Fractal Dimension	KNN	ZO-NF-S	ACC=97.28%
De La O Serna et al. [156]	2020	Frequency domain	Taylor-Fourier Filter Bank with O-Splines	SVM	ZO-NF-S	ACC=94.88%
Eltrass et al. [157]	2021	Time domain	Energy of signal	Quantized Kernel Least Mean Square	Z-O-N-F-S	ACC=97.88%, SENS=98.8%, SPEC=97.65%
Gao et al. [158]	2020	Non-Linear	Approximate Entropy, Recurrence Quantification Analysis	CNN	ZO-NFS	ACC=99.26%, SENS=98.84%, SPEC=99.26%
Goshvarpour et al. [159]	2020	Non-Linear	Lagged Poincare Plot	KNN	Z-F-S	ACC=96%, SPEC=99.48%, SENS=95.19%
Gu et al. [160]	2021	Raw Signal	-	Hierarchical discriminative sparse representation classifier	Z-O-N-F-S	ACC=98.8%
Hassan et al. [161]	2020	Time-Frequency	Complete Ensemble EMD with adaptive noise	AdaBoost	ZO-NF-S ZONF-S	ACC=97.6% ACC=99.2%
Jana et al. [162]	2021	Time-Frequency	STFT	SVM	ZONF-S	ACC=97.63%, SENS=98.38%, SPEC=94.67%
Jang et al. [163]	2020	Time-Frequency	DWT and Phase-Space Reconstruction	Neural Network with weighted fuzzy	Z-S	ACC=97.5%, SENS=95%, SPEC=100%

				membership (NEWFM)		
Jiang et al. [164]	2020	Time-Frequency	Scattering Transform, Fuzzy entropy, Log Energy entropy	SVM	ZO-NF-S	ACC=99.87%
Lee et al. [165]	2020	Time-Frequency	DWT	Hidden Markov Model	ZONF-S	ACC=99.54%, SENS=99.51% SPEC = 98.6%
Li et al. [166]	2021	Time-Frequency	STFT, WPD, KPCA	TSK fuzzy system	ZO-NFS	ACC=98.67%
Lian et al. [167]	2020	Raw Signal	_	CNN	Z-S	ACC= 99.84%, SENS= 99.5%, SPEC=99.6%
Ma et al. [168]	2021	Raw Signal	_	CNN + RCNN		ACC=100%, SPEC=100%, SENS 100%
Mandhouj et al. [169]	2021	Time-Frequency	STFT spectogram	CNN	ZO-S	ACC: 98.88%, SENS 98.33%, SPEC=100%
Mardini et al. [170]	2020	Time-Frequency	DWT	Naïve Bayes	ZONF-S	ACC=99.3%
Mathur et al. [171]	2021	Frequency domain	Ramanujan periodic subspace (RPS), energy of the projection	SVM	Z-S O-S N-S F-S	ACC= 99.5%, ACC= 98.6%, ACC= 98% ACC= 97.5%
Molla et al. [172]	2020	Time-Frequency	DWT, Graph Eigen Decomposition (GED)	Feed-Forward NN	ZONF-S	ACC=99.39%
Na et al. [173]	2021	Frequency domain	DFT	KNN	ZONF-S	ACC=99.4%
Nogay et al. [174]	2021	Time-Frequency	STFT spectogram	CNN + ALEXNET	ZO-NF-S	ACC=100%
Oinam et al. [175]	2020	Time-Frequency	DWT	MLP trained with hybrid PSO, GSA	ZF-S F-S	ACC=95.33% ACC=93%
Omidvar et al. [176]	2021	Time-Frequency	DWT, entropy	SVM, MLP with GA for feature selection	ZO-NF-S	ACC=98.7% SENS=97.5%, SPEC=100%
Pal et al. [177]	2021	Frequency domain	FFT, Bubble entropy	KNN	Z-N-S	ACC=99.73%
Polat et al. [178]	2020	Frequency domain + Time domain	FFT + time features	SVM	Z-O-N-F-S	ACC=82.5%
Radman et al. [179]	2021	Time-Frequency	DWT + FFT + time features	Random Forest	ZO-NF-S	ACC= 99.33%, Sens = 98.33%, SPEC=98.88%
Shekokar et al. [180]	2022	Raw Signal	-	LSTM	Z-S	ACC=99.5%
Singh et al. [181]	2021	Time-Frequency	Complete Ensemble EMD, dispersion entropy	MLP	ZO-NF-S	ACC=98.7%
Sujatha et al. [182]	2020	Time-Frequency	DWT, Approximate entropy, Statistical features	SVM	ZONF-S	ACC=96.5%
Sukriti et al. [183]	2021	Time-Frequency	Variational Mode Decomposition (VMD)	Random Forest	ZO-NF-S	ACC=98.2%, SENS=99.7%, SPEC=98.7%
Wang et al. [8]	2020	Frequency domain	FFT	weighted kNN	Z-F-S	ACC=99%
Woodbright et al. [184]	2021	Raw Signal		CNN	ZONF-S	ACC= 98.65%, SENS=96.29%, SPEC=99.25%
Yildiz et al. [185]	2021	Time-Frequency	STFT spectogram, scalogram	CNN, alexnet, resnet-18, googlenet	ZO-NF-S	ACC=100%
Zeng et al. [186]	2021	Time domain	Gray Recurrence Plot	DenseNet NN	ZONF-S	ACC=100%
Zeng et al. [187]	2020	Time-Frequency	Intrinsic Time-Scale Decomposition, DWT, PSR	MLP	Z-O-N-F-S	ACC=94%
Zhang et al. [188]	2020	Raw Signal		CNN with Multi-Scale Non-Local Layer	Z-O-N-F-S	ACC=94.01%, F1=89.46%
Zhang et al. [189]	2021	Time-Frequency, Non Linear	Frequency slice WT (FSWT), Fuzzy entropy, Higuchi FD	t-distributed stochastic neighbor embedding (t-SNE)	Z-O-N-F-S	ACC=93.62%

Zhao et al. [190]	2020	Raw Signal		CNN	ZO-NF-S Z-O-N-F-S	ACC=96.73% ACC=93.55%
Zhou et al. [191]	2020	Time-Frequency	DWT, entropy features	RBF NN	ZO-NFS Z-O-N-F-S	ACC=96.3% ACC=78.4%
Zhou et al. [192]	2021	Time-Frequency	WPD	Graph-Based ELM	ZONF-S	ACC=94.8%
Zubair et al. [193]	2021	Time-Frequency	DWT, Spike Rhythmicity, Relative Spike Amplitude, Spectral entropy	CatBoost	ZONF-S	ACC=97%
Dehuri et al. [194]	2022	Time-Frequency	DWT	SVD-ELM	Z-O-N-F-S	ACC=95%

Table 22 Studies that used CHB-MIT database, published in 2017-2019

Author	Year	Signal Transform	Feature Extraction	Classification	Classification Problem	Results
Cao et al.[195]	2019	Time-Frequency	STFT	CNN, ELM	a) seizure/non-seizure, b) preictal/ ictal/ interictal	a) ACC=99.33%, b) ACC=98.62%
Harpale et al.[196]	2018	Time-Frequency	Time domain features+ FFT features + Pattern Adapted WT	Fuzzy classifier	ictal-interictal	ACC=96.02% SPEC=94.5%
Mansouri et al.[197]	2019	Frequency	FFT, band power	Assosiation Network	seizure/non-seizure	SENS=83%
Tian et al.[198]	2019	Time-Frequency	WPD, FFT	TSK fuzzy system	seizure/non-seizure	ACC=98.3%, SENS=96.7%, SPEC=99.1%
Birjandtalab et al.[199]	2017	Frequency	FFT, band power	kNN	seizure/non-seizure	SENS=80.87%, F-score=56.23%
Ibrahim et al.[200]	2017	Time-Frequency	DWT, Shannon entropy	kNN	seizure/non-seizure	SENS=94.5%, FDR=1.14/h

Table 23 Studies that used CHB-MIT database, published in 2020-2022

Author	Year	Signal Transform	Feature Extraction	Classification	Classification Problem	Results
Gabr et al.[201]	2020	Time-Frequency	STFT spectogram, scalogram	CNN	ictal-preictal-interictal	ACC=97%
Mouleesh-uwarapprabu et al.[202]	2020	Time-Frequency	DWT	Nonlinear Vector Decomposed NN	ictal-interictal	ACC=95.6%, SENS=94.7%, SPEC=94.1%, Recall=89.3%, Precision=78.1%
Zhang et al.[203]	2020	Frequency Domain	DFT, band energies	Attention Network AttVGGNet	ictal-interictal	ACC=95.6%, SENS=94.7%, SPEC=94.1%, Recall=89.3%, Precision=78.1%

Khan et al.[204]	2020	Time-Frequency	DWT	LDA	ictal-interictal	ACC=99.6%, SENS=99.8%
Akbarian et al.[205]	2020	Frequency Domain	DFT, effective brain connectivity	Autoencoder NN	ictal-interictal	ACC=97.91% SENS=97.65%, SPEC=98.06%
Zeng et al.[206]	2021	Time-Frequency	EWT + DWT	kNN	ictal-interictal	ACC=99.77%, SENS=99.88%, SPEC=99.88
Zhao et al.[207]	2021	Raw Signal	Pearson Correlation	Linear graph Convolution Network	ictal-interictal	ACC=99.3%, SENS=99.43%, SPEC=98.82%, F1=98.73%
Nasiri et al.[208]	2021	Time-Frequency	STFT	CNN	ictal-interictal	ACC=91.71%, SENS=91.09%, SPEC=94.73%
Zhang et al.[209]	2020	Time-Frequency	STFT	deep CNN, ImageNet	ictal-interictal	ACC= 97.75% Recall=98.44%, Precision=97.47%
Hu et al.[210]	2020	Time Domain	Local Mean Decomposition	Bidirectional LSTM	ictal-interictal	G-mean 92.66% SENS=93.61%, SPEC=91.85%
Slimen et al.[211]	2020	Time-Frequency	EMD, DWT, dual-tree complete WT (DTCWT)	LDA	ictal-interictal	ACC=100%
Quintero-Rincón et al.[212]	2020	Time Domain	R-square value, RMS	Random Forests	ictal-interictal	ACC=94.1%, TPR=92% TNR=96%
Siddiqui et al.[213]	2020	Time Domain + Non Linear	Statistical + Entropy Features	Random Forests,	ictal-interictal	ACC=66%, Recall= 83.01%
Bhandari et al. [214]	2022	Time-Frequency	STFT+DWT	Modified Tunicate Swarm, LSTM	ictal-interictal	ACC= 96.87%, SENS=98.7%, PREC=97.98%

Table 24 Studies that used other databases

Authors	Database	Date	Category of Features	Feature Extraction	Classification	Classification Problem	Results
Zeng et al.[215]	PUPH	2017	Non-Linear	Permutation Entropy	QDA	ictal-preictal-interictal	ACC=90.3%
Gao et al. [216]	Bern-Barcelona	2018	Time-Domain	Statistical Features	Autoregressive Linear Model	Seizure-Non seizure	F1= 93.75%
Zhang & Yang et al. [217]	TUH	2018	Frequency, Non-Linear	FFT, Statistical Features, Fractal Dimension	SVM	ictal-interictal	ACC=99.4%
Sriraam et al. [218]	RMH	2018	Non-Linear	Teager Energy	MLP	ictal-interictal	SENS=94.38%, SPEC=98.25%,
Ma *et al.* [219]	Freiburg	2019	Time-Frequency	DWT	Dictionary Learning	ictal-interictal	ACC=98.5%, SENS=93.39%, SPEC=98.51%
Mahmoodian *et al.*[220]	Freiburg	2019	Frequency, Non-Linear	FFT, Cross-bispectrum Analysis	SVM	seizure- non seizure	ACC=96.8%, SENS=95.8%, SPEC=96.7%
Alhussein et al.[221]	TUH	2019	Raw Signal	-	CNN, Transfer Learning	ictal-interictal	ACC=87.96%
Zhang et al.[222]	TUH	2020	Time-Frequency	DWT	CNN	ictal-interictal	SENS=70.98%, SPEC=73.17%
Sharma et al. [223]	TUH	2020	Time-Frequency	DWT, minimized orthogonal wavelet filter bank, Fuzzy entropy, Fractal Dimension	SVM	ictal-interictal	ACC=79.34%
Iešmantas et al. [224]	TUH	2020	Frequency Domain	Phase-Locking Value	CNN	ictal-interictal	ACC=74%

Zhan et al. [225]	Freiburg	2020	Time-Frequency	DWT, Fourier Transform, Convolution Block	Fuzzy C-means (FCM)	ictal-interictal	ACC=89.75%, SENS=85.52%
Tuncer et al. [226]	TUH	2021	Time-Frequency	WPD, Chaotic one-dimensional local binary pattern (CLBP)	SVM	ictal-interictal	ACC=96.23%, PREC=96.92%, Recall=95.01%, F1=95.93%
Priyasad et al. [227]	TUH	2021	Raw-Signal	Attentive Feature Fusion	deep CNN	ictal-interictal	F1 score=96.7%
Lu et al. [228]	Bern-Barcelona	2021	Non-Linear	Sample Entropy, Higuchi Fractal Dimension	SVM	ictal-interictal	PREC 88%, Recall 79%, F1= 81%
Mu et al. [229]	Freiburg	2021	Time-Frequency	DWT + Graph-regularized non-negative matrix factorization (GNMF)	Bayesian linear discriminant analysis (BLDA)	ictal-interictal	ACC=98.16%, SENS=93.2%, SPEC=98.16%
Hadiyoso et al. [230]	NSC	2021	Non-Linear	Spectral Entropy, Katz & Sevcik Fractal Dimension	Naïve Bayes	Ictal- Preictal-Interictal	ACC= 85.3%, SPEC= 92.7% SENS=85.3%

Table 25 Studies that used multiple databases

Author	Year	Databases	Author	Year	Databases
Chen et al.[231]	2017	Bonn, PUPH	Jiang et al. [258]	2020	CHB-MIT, Bonn
Solaija et al.[232]	2018	CHB-MIT, KU Leuven	Liu et al. [259]	2020	Temple, EPILEPSIAE
Raghu et al. [233]	2018	Temple, RMCH	Lyu et al. [260]	2021	CHB-MIT, Custom
Kumar et al. [234]	2019	Bonn, Bern	Anuragi et al. [261]	2021	CHB-MIT, Bonn
Jiang et al. [235]	2019	Temple, Bonn	Supriya et al. [262]	2021	Bern, Bonn
Raghu et al. [236]	2019	Bonn, RMCH, CHB-MIT	Panda et al. [263]	2021	Bonn, NSC
Al Ghayab et al.[237]	2019	Bonn, Bern	Xiong et al. [264]	2021	CHB-MIT, Bonn
Bilal et al. [238]	2019	Bonn, NSC	Shankar et al. [265]	2021	CHB-MIT, Bonn
Pandey et al. [239]	2019	CHB-MIT, MIH Arrythmia	Shariat et al. [266]	2021	CHB-MIT, Custom
Wu et al. [240]	2019	CHB-MIT, INeuro	He et al. [267]	2021	CHB-MIT, Bonn, Pone_pat
Truong et al. [241]	2019	CHB-MIT, Freiburg, EPILEPSIAE	Liu et al. [268]	2021	CHB-MIT, Bonn
Gómez et al. [242]	2020	CHB-MIT, EPILEPSIAE	Jiang et al. [269]	2021	CHB-MIT, Bonn
Sameer et al. [243]	2020	Bonn, NSC	Glory et al. [270]	2021	CHB-MIT, Bonn
Lian et al. [244]	2020	Bonn, Freiburg	Wang et al. [271]	2021	Bonn, NSC
Zhou et al. [245]	2020	Bonn, NSC	Hu et al. [272]	2021	CHB-MIT, INeuro
Ansari et al. [246]	2020	CHB-MIT, Bonn, AIIMS	Praveena et al. [273]	2021	Bern, Bonn, Temple
Raghu et al. [247]	2020	CHB-MIT, Bonn, TUH, MUMC	Peng et al. [274]	2021	CHB-MIT, Bonn, NSC
Abiyev et al. [248]	2020	CHB-MIT, Bonn	Aayesha et al. [275]	2021	CHB-MIT, Bonn
Ayodele et al. [249]	2020	CHB, Temple	Hu et al. [276]	2021	CHB-MIT, INeuro
George et al. [250]	2020	KITS, Temple	Li et al. [277]	2021	CHB-MIT, Bonn
Li et al. [251]	2020	Bonn, CHB, Freiburg	Zarei et al. [278]	2021	CHB-MIT, Bonn, NSC
Li et al. [252]	2020	CHB-MIT, Bonn, Temple	Sahani et al. [279]	2021	CHB-MIT, Bonn
Rout et al. [253]	2020	Bonn, NSC	Rafiammal et al. [280]	2021	CHB-MIT, Bern, Bonn
Wu et al. [254]	2020	CHB-MIT, Bonn	Cao et al. [281]	2021	CHB-MIT, INeuro
Kiranmayi et al. [255]	2020	Bonn, custom	Huang et al. [282]	2021	Bonn, CHB-MIT
Zhang et al. [256]	2020	Bonn, NSC	Chen et al. [231]	2017	Bonn, CHB-MIT
Dash et al. [257]	2020	CHB-MIT, AIIMS, Custom	Jiang et al. [283]	2019	Bonn, CHB-MIT

Author	Year	Databases	Author	Year	Databases
Wang et al. [284]	2017	Bonn, CHB-MIT	Zhou et al. [290]	2018	CHB, Freiburg
Yuan et al. [285]	2018	Bonn, CHB-MIT	Yu et al. [291]	2019	CHB, Freiburg
Lu et al. [286]	2018	Bonn, CHB-MIT	Sun et al. [292]	2019	Bonn, CHB-MIT, Freiburg
Hussain et al. [287]	2019	Bonn, CHB-MIT	Fu et al. [293]	2019	Bonn, PUPH
Zhang et al. [288]	2019	Bonn, CHB-MIT			
Abdelhameed et al. [289]	2019	Bonn, CHB-MIT			

Figure 35 The basic structure of a methodology that utilizes Machine Learning for the EEG epilepsy detection.

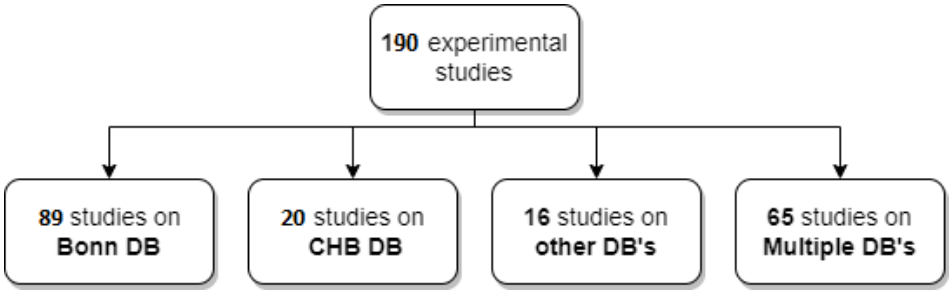


Figure 36 summarizes the categorization of the experimental studies.

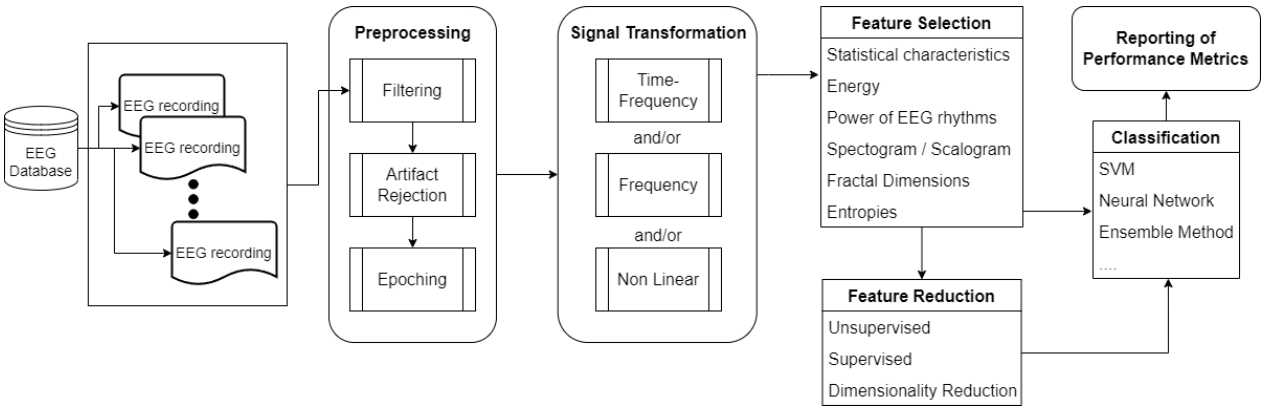


Figure 36 Separation of the experimental studies based on the database used.

5.2.4.3. Discussion and Statistics

While various methodological approaches are presented, the most of them follow a basic structure as depicted in Figure 35. Initially, a preprocessing phase is conducted (or is pre-completed in the database recordings), involving filtering of the EEG signals and application of one or more artifact rejection methodology. During this phase, the frequency content is restricted, typically allowing frequencies in the [0.5-45] Hz range, which are of scientific relevance for EEG interpretation. Additionally, in many studies, the recordings are segmented into fixed-length time windows, determined either by assessing the optimal window length or through arbitrary selection. Subsequently, a signal transformation step may occur, utilizing established methods like FFT or DWT, and the

signal is often categorized into the five EEG rhythms: Alpha, Beta, Gamma, Theta, Delta (or another division specific to the study). Then comes the feature extraction stage, where the feature vector is formed, potentially incorporating elements from multiple domains (Time, Frequency, TF, or Non-Linear). Some studies even apply multiple TF transformations (like DWT combined with STFT) and conduct feature extraction on each. Following this, a feature selection process is undertaken to refine the feature vector to its most critical attributes for classification, or it is transformed using a method such as PCA. Finally, the chosen Machine Learning algorithm is trained to tackle one or more classification tasks. The study concludes with the reporting of performance results obtained through validation techniques like k-fold cross-validation or leave-one-patient-out validation.

The goal of this study is to present a comprehensive overview of the methodologies and databases utilized in epilepsy detection research over the past five years, serving as a reference for researchers looking to develop new detection methods. Following, the selected signal transformations and classifiers are discussed, and their popularity is quantified. Additionally, regarding the studies that employ multiple databases, insights about which databases are most commonly used together are provided. Concluding this review, there will be a comparative analysis with other related systematic reviews in the field of epilepsy, and a discussion on the limitations of the current methodology.

Firstly, regarding the Signal Transformation step, it is evident that Time-Frequency methodologies are the most popular, with 59.84% of the studies employing one. As mentioned by Morales et al. [294], these transformations offer better interpretability over other EEG transformations because they provide the most direct information about the neurophysiological mechanisms of the EEG data. Furthermore, the fact that most Time-Frequency methodologies produce an image output (for example STFT spectrogram), make them suitable for usage with CNN implementations. The most used Time-Frequency transform is the DWT (43.8% of them), with second being the EMD (13.7%) followed by STFT (12.3%). A noticeable trend is seen in the use of combined Time-Frequency (TF) methods in studies. While from 2017 to 2019, only 7.1% of studies using TF methods combined different TF techniques, this percentage increased significantly to 17.8% between 2020 and 2022. This shows that researchers are increasingly trying out more complex methods to improve epilepsy detection. Last but not least, another simple but important trend to note is the increase in studies using Raw-Signal methodologies. In the earlier period from 2017 to 2019, only 4.1% of studies (2 studies) used Raw-Signal approaches. However, this number grew to 8.9% (9 studies) in the later period from 2020 to 2022. This increase is largely due to the enhanced computing power of modern systems, particularly in GPU performance. These advancements allow for the use of complex Convolutional Neural Networks that can process Raw-Signal data using 1-D convolutional layers. Figure 37 provides an overview of the popularity of each signal transformation.

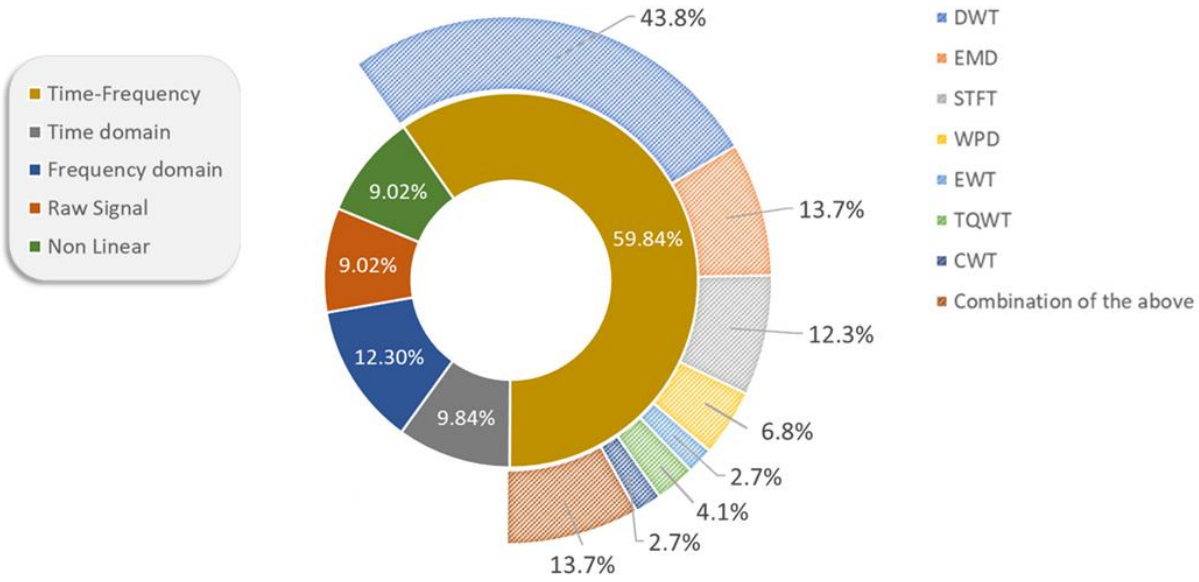


Figure 38 Signal processing techniques and Transformations that are applied in the studies of the review.

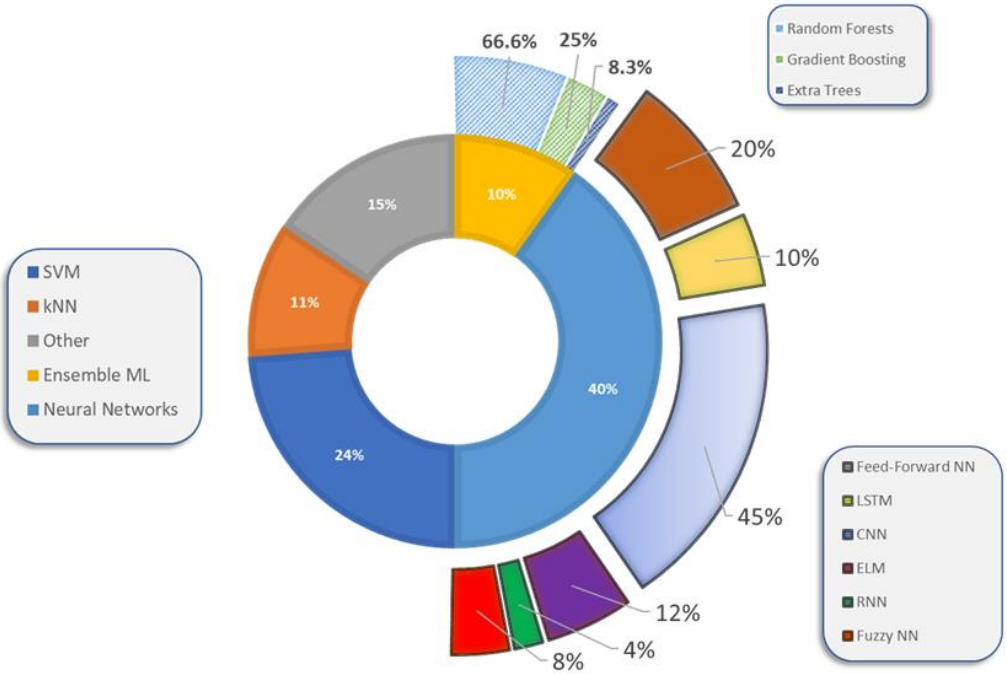


Figure 37 Machine Learning methodologies that are applied in the studies of the review.

Moreover, concerning the classifiers employed by the reviewed studies, 50% of them were traditional classifiers (like kNN, SVM or Naïve Bayes), the other 10% were ensemble methods (like Random Forests or Gradient Boosting), and the rest 40% were Neural Networks and deep learning architectures, with CNN being by far the most used scheme. Furthermore, a notable insight can be obtained from the comparison of the popularity of the type of the classifier used in the studies during the 2017-2019 and 2020-2022 periods, as depicted in Figure 39. A significant increase in the popularity of the Deep

Learning implementations can be observed, increasing 25% in popularity (25% to 50%). This is partially due to the fact that the computational power of the systems has increased. Yet more increase in their popularity is expected during the next years, due to the significant surge in the use of transformer networks, which are rapidly becoming a dominant force in various areas of machine learning and data analysis.

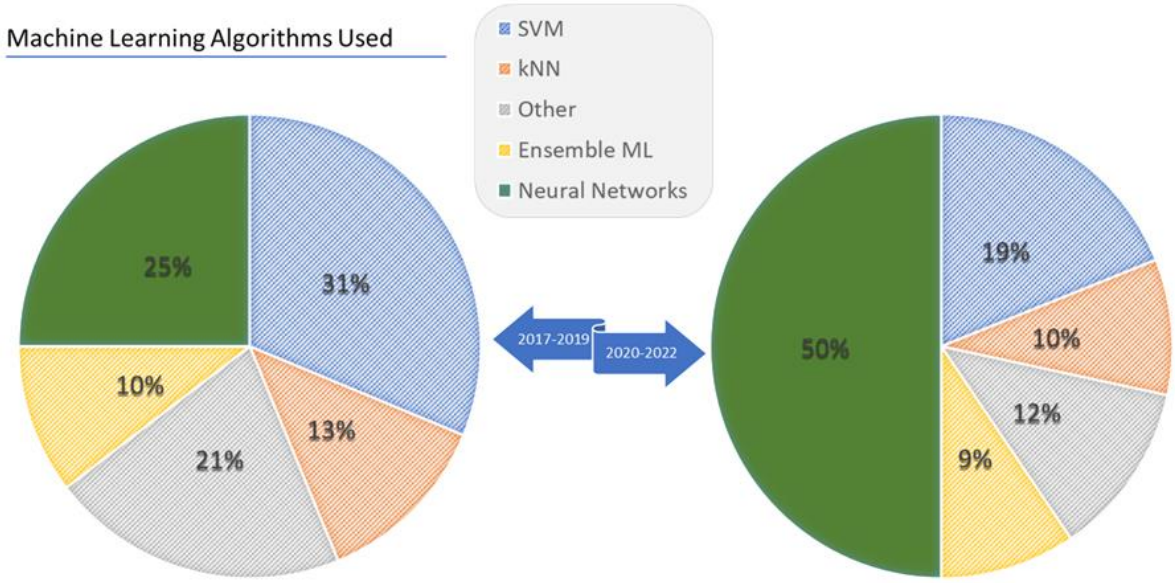


Figure 39 Comparison of Machine Learning methodologies used in the 2017-2019 and the 2020-2022 period.

Although a direct comparison of the performance of different classifiers would be useful in choosing a classification methodology for future research, it poses several limitations which make it impossible. Such limitations are:

a) Not every study uses the same dataset. Even though there are studies conducted on the same database, the selection of EEG signals or subjects differs.
b) Each study performs a different classification problem (i.e., studies on the Bonn DB evaluate any combination of Z-O-N-F-S). There is a huge number of different combinations of database used and problem addressed.
c) Expanding the previous limitation, there is a deficiency in the number of studies using a standardized classifier, database, and problem definition, which hinders the ability to conduct a thorough statistical analysis of variations in classifier performance.
d) Some studies might have methodological shortcomings, like overfitting, leading to artificially inflated performance scores.

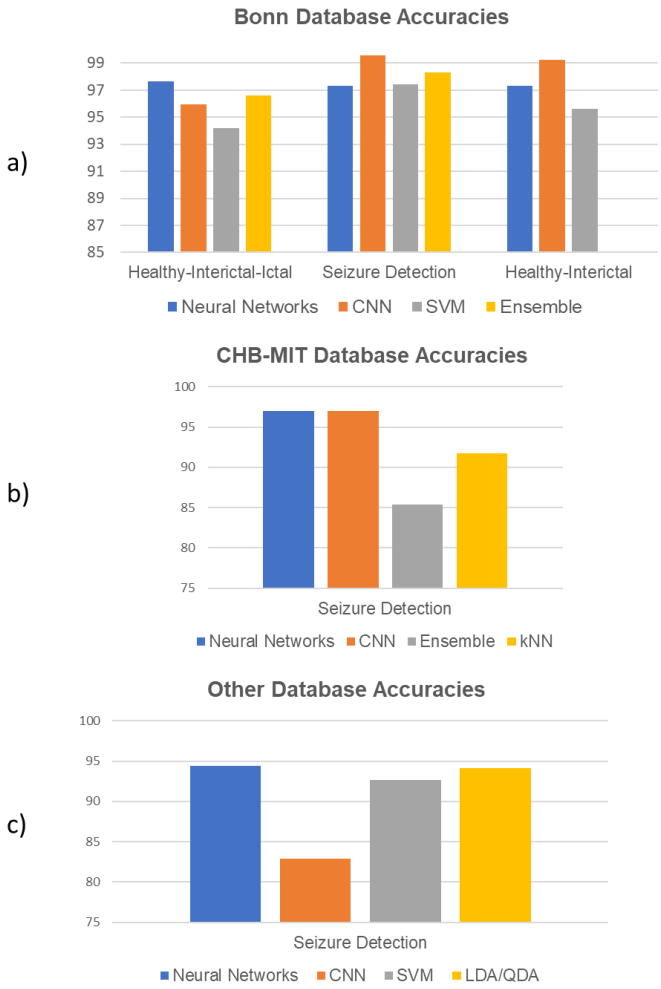


Figure 40 Average accuracy of the different classifiers reported in the studies for each database. Studies on the Bonn database are divided into 3 categories namely Healthy-Ictal-Interictal, Seizure Detection and Healthy-Interictal. In this figure CNNs are considered a different category from Neural Networks due to their increased importance.

Having established these limitations, a visualization of the performance of different classifiers on the same database is presented in Figure 40. The Bonn database studies were divided into the three problems, namely Healthy-Ictal-Interictal, seizure detection, Healthy-Interictal, while for CHB-MIT and Other databases, only the seizure detection problem has been explored. Furthermore, the CNNs are considered a different category than the rest of the Neural Networks, due to their popularity. It is observed that CNN along with the other Neural Network architectures achieved the greatest performance in all the problems.

Moving forward to the comparison of the popularity of the existing public databases, it is notable that the Bonn database is the most utilized one with almost half of the studies (46.8%) utilizing it alone. Not only that, but 72.31% of the multiple database studies make use of the Bonn recordings, despite the important limitations of the Bonn database that were discussed earlier. The CHB -MIT database is the second most used, while the

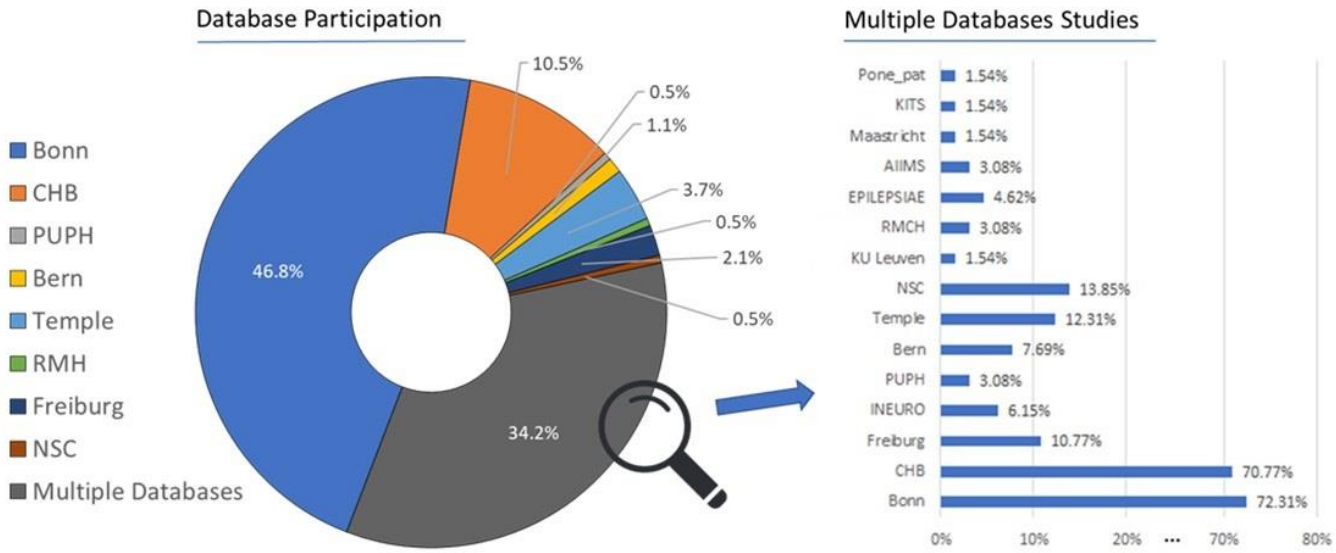


Figure 41 Popularity of each database on the reviewed studies. The barchart on the right is concerning the Multiple Databases section, in which each bar represents the percentage of the studies that examine a certain database.

Freiburg database, although once popular, is no longer preferred, due to not being freely available anymore. To illustrate the popularity of each database, as well as the employment of each database on the studies with multiple databases, Figure 41 is created. In this left side the popularity of each database, and on the right side the popularity regarding only the multiple databases studies is illustrated.

Furthermore, an intriguing aspect of the systematic review involved evaluating the relationships between databases, particularly in terms of their coexistence in studies that utilized multiple databases. A chord diagram is created, which models the connection of two databases with a chord, the width of which demonstrates the importance of the connection. The strong connection between the Bonn and the CHB-MIT database can be observed, since 47.7% of the multiple database studies employ the combination of these two. The second most important combination is that of Bonn and Neurology Sleep Center databases, coexisting on 13.8%of the studies. Another notable connection is that of CHB-MIT and Freiburg databases, which exists at the 9.2% of the multiple database studies. The quantification of these connections can be found in Table 26 while the chord diagram visualization can be found in Figure 42.

Table 26 Each cell represents the percentage of studies that used axis x & axis y combination of databases, from the total of multiple database studies.

	Bonn	CHB-MIT	Freiburg	Bern	Temple	Upenn	Pone_pat
CHB-MIT	47.7%	--	--	--	--	--	--
Freiburg	4.6%	9.2%	--	--	--	--	--

INEURO		5.9%		--	--	--	--
PUPH	3.1%			--	--	--	--
Bern	7.8%	1.5%		--	--	--	--
Temple	6.2%	4.6%		1.5%	--	--	--
NSC	13.8%	3.1%				--	--
KU Leuven		1.5%				--	--
RMCH	1.5%	1.5%			1.5%	--	--
MIH Arr.		1.5%				--	--
Mayo Clinic		1.5%	1.5%			1.5%	--
EPILEPSIAE		3.1%	1.5%		1.5%		--
AIIMS	1.5%	3.1%					--
Maastricht	1.5%	1.5%			1.5%		--
KITS					1.5%		--
Pone_pat	1.5%	1.5%					--

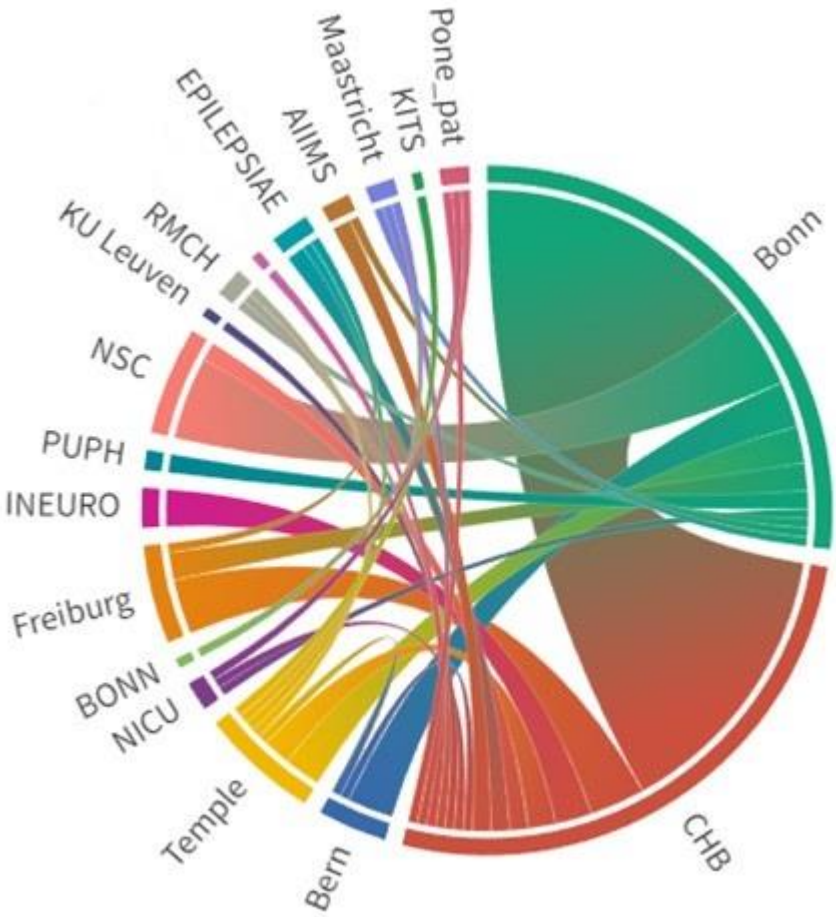


Figure 42 Chord diagram modeling the connection of databases co-existing in studies that employ multiple databases.

5.2.4.4. Related work comparison

Numerous extensive review papers have been published in 2021-2022, focusing on Machine Learning in EEG Epilepsy detection and the latest developments in this area. These studies are concisely summarized in Table 8. However, none of these reviews

adopted a PRISMA-based methodology for a systematic evaluation of EEG-based epilepsy detection studies. Our review offers a distinct advantage over other recent related reviews by methodically collecting, analyzing, and evaluating experimental studies in line with the PRISMA statement. This approach provides valuable insights at each stage of developing an EEG-based epilepsy detection methodology. Specifically, we offer an in-depth comparison of public databases in terms of their characteristics, frequency of use in other studies, and their inter-combinations. Additionally, we conduct a thorough assessment of the Signal Transformation and Feature Extraction phases, analyzing prevalent trends and their evolution post-2020. Similarly, we scrutinize the Classification Step, highlighting the rising popularity of Neural Networks in recent years through detailed descriptions and visual aids.

It is, however, crucial to acknowledge the limitations of this review. Firstly, the extensive scope of our study, covering 190 studies, precluded an individualized evaluation of each implementation. Instead, we opted for a broader analysis of distinct steps, as outlined in Tables 1-5. Moreover, a systematic comparison of classifier performances was not feasible due to various previously mentioned challenges, making the accuracy score comparisons in our study indicative rather than definitive guides for selecting the best methodology. Lastly, the vast volume of research in EEG epilepsy detection necessitated a focus on studies from a specific timeframe (2017-2022), which limited our ability to track methodological trends over a longer period. Table 27 presents a concise analysis of review studies with related analysis.

Table 27 Related review papers comparison, regarding EEG-based epilepsy detection.

Author	Year	Search Methodology	Years of Coverage	Inclusion Criteria / Focus
Saminu et al. [295]	2022	Not Reported	2016-2021	Computed Aided Devices, EEG or MRI,
Ahmad et al. [296]	2022	Not Reported	Not Reported	EEG, Epilepsy, Deep Learning/ Machine Learning, Further inclusion or exclusion criteria are not reported
Praveena et al. [297]	2021	Not Reported	Not Reported	EEG, Epilepsy, Deep Learning
Supriya et al. [298]	2021	Not Reported	Not Reported	Graph Theory, EEG, Epilepsy
Rasheed et al. [299]	2020	Not Reported	1970-2019	EEG, Epilepsy, Deep Learning/ Machine Learning, prediction, Focus on discussion of Machine Learning pitfalls
This study	**2022**	**PRISMA**	**2017-2022**	**EEG, Epilepsy, Detection, Machine Learning, Published DB Focused**

5.3. Brain-Computer Interfaces for analyzing cognitive states

The last section of the doctoral research contains studies that employ a Human-Computer Interaction protocol along with EEG recording to analyze a cognitive state or a learning disability. Specifically, three papers were published, one regarding the study of high-altitude stress through a Virtual Reality environment and EEG + electrocardiogram (ECG) combination, one regarding the evaluation of dyslexia through machine learning on EEG derived from an Interactive Linguistic Language tool specifically designed for this research, and one about exploring the haptic perception during active touch on different surfaces with open eyes.

5.3.1. Assessing Electroencephalography as a Stress Indicator: A VR High-Altitude Scenario monitored through EEG and ECG

This article was published in the open access journal MDPI Sensors, during August 2022. Humans experience stress as a fundamental part of their daily lives. When faced with stress, individuals often enter a state of hyperarousal, experiencing involuntary changes like fluctuations in respiratory rate, muscle tone, or heart rate, alongside subconscious neuroendocrine adaptations [300]. In the short term, activating the stress system can positively impact motivation, attention, and goal-oriented behavior. However, long-term stress is associated with negative effects on physical and mental health, including impairing memory and contributing to mood disorders [301]. Over the past decades, the academic world has delved into the concept of stress, focusing on its neurophysiological processes, and examining its two main components: the 'stressors' (causes of stress) and the 'stress response' (the body's reaction) [300]. Stress is classified into three types according to the literature: acute, associated with short, intense exposures; episodic, tied to repeated occurrences over limited periods; and chronic, linked to ongoing, long-lasting stressors [302].

Research into stress, employing diverse approaches, investigates various biomarkers like cortisol levels, heart rate metrics, galvanic skin response, pupil diameter, and others, following the induction of a specific stressor. Designing these experimental frameworks is challenging, particularly considering the individual variations in stress responses, coping mechanisms, and personal perceptions. Stress responses are complex, manifesting

---

The author list ordered as published was: Vasileios Aspiotis, Andreas Miltiadous, Konstantinos Kalafatakis, Katerina D. Tzimourta, Nikolaos Giannakeas, Markos G. Tsipouras, Dimitrios Peschos, Euripidis Glavas, Alexandros T. Tzallas. The first two authors shared first co-authorship of the article.

in behavioral, cognitive, psychological, or neurophysiological patterns, which must be carefully considered from the research perspective. The selection of assessment methods and stress-inducing stimuli is critical for generating results that are reliable, measurable, and replicable. A range of methods to induce stress have been used in studies, including performance-based, psychological, or social tests like the Trier Social Stress Test [303] and the Maastricht Acute Stress Test [304]. These varying approaches adopted by researchers come with their own sets of pros and cons, differing in elements such as unpredictability, controllability, consistency across different ages and demographics, and factors like cost and equipment availability.

Recent advancements in Virtual Reality (VR) and Head Mounted Displays (HMDs) are being leveraged by researchers to explore specific stressors and trigger particular stress responses. HMDs provide a unique combination of feasibility, repeatability, and control, while also paving the way for creative and tangible innovations. Highly immersive virtual environments, capable of replicating real-life scenarios or creating entirely fantastical ones, allow for significant levels of interaction and immersion. Research involving stress induction using VR HMDs has led to innovative adaptations of traditional stress tests like the Trial Social Stress Test or the Maastricht Acute Stress Test, yielding promising outcomes. Additionally, VR enables the creation of novel scenarios that would be too costly, risky, or impractical to execute in real life, such as train or roller coaster simulations, horror-themed experiences, escape from perilous situations, emergency response scenarios, and more [305]. Another method for inducing physiological stress is exposure to high altitudes. Based on utilizing specific phobias as stressors and the well-established connection between fear and the activation of the stress system, this approach has been effectively used as a research tool to elicit physiological stress responses [306].

Given that the human brain plays a pivotal role in the stress response system [307], utilizing electroencephalography (EEG) to monitor its activity can reveal aspects of acute stress responses and lead to the discovery of novel detection methods. EEG analysis has been widely applied in the study of neurodegenerative diseases [308], cognitive disorders, and Brain-Computer Interface (BCI) applications. Despite its extensive use in these areas, EEG measurements have not been as thoroughly explored in stress research, in contrast to other types of neurophysiological data. In stress-related EEG studies, common approaches include analyzing brain activity and conducting automatic classification. Brain activity analysis often involves investigating functional and effective connectivity, with the former examining time-related coherence among neural activities. While these studies have yielded intriguing results using various analysis techniques, there is a lack of specific guidelines regarding the selection of EEG features and their integration.

In this context, assessing brain region asymmetry is frequently utilized as a measure of functional connectivity. Research indicates that functional lateralization in the frontal cortex is linked to affective processing and may specifically serve as an indicator of physiological stress [309]. Frontal Alpha Asymmetry (FAA), in particular, is a metric used in numerous protocols to assess mental stress. FAA is derived from the logarithmic difference between two symmetrical frontal EEG electrodes (commonly F3-F4 or F7-F8).

A dominance in the right frontal area is associated with negative emotion regulation and tendencies towards social withdrawal, whereas left side dominance correlates with positive emotions, greater emotional adaptability, and more effective emotion regulation. Nonetheless, the use of FAA has predominantly been confined to experiments involving cognitive stress, such as mental stress induction, mathematical problem-solving, or computational tasks. Other measures of brain asymmetry have also been subjects of study, including Occipital Alpha Asymmetry (OAA), which has been linked to low responsiveness to antidepressants [310], and anxiety related to physical performance. However, these findings have not been consistently replicated in larger, multicenter studies [311].

In the past five years, there have been notable advancements in evaluating stress through Virtual Reality (VR) environments. Yet, only a few studies have incorporated multiple biomarkers, including EEG, for stress assessment [312]. Stolz et al. employed a VR environment featuring avatars with varying facial expressions and sounds in a threat-conditioned context, using EEG's Event Related Potentials (ERP) to study cortical processing [313]. Fadeev et al. conducted a small-scale, custom study using various VR scenarios to observe changes in EEG, respiration rate, and heart rate as indicators of stress [314]. Wang et al. utilized the VR Richie's Plank Experience for EEG data collection and classified subjects based on their self-reported fear of heights, analyzing the EEG signals [315]. However, to our knowledge, there are only a few studies that explore the relationship between EEG and ECG-established bioindicators like BPM within immersive VR settings [316], [317].

In our study, we seek to investigate the correlation between brain neurodynamics and a recognized cardiovascular biomarker for stress, utilizing VR HMDs to simulate a hazardous-type stressor. Our focus is on whether different brain regions activate in response to a VR high altitude scenario and how this relates to stress, using BPM as a verifying measure. Specifically, an increase in heart rate under these conditions would indicate stress, as the participants are not subjected to any physical exertion or environmental changes like temperature variations during the experiment, other than virtual high-altitude exposure. The rise in heart rate is a well-established marker of mental stress, as confirmed by various studies [318], [319], [320], [321]. Additionally, we aim to determine if frontal and occipital asymmetry is significantly impacted by this stressor and to explore potential correlations between brain area energy variations, FAA, OAA, and BPM. By dividing participants into two groups of interest and comparing EEG biomarkers between them, we seek to uncover how high-altitude stress may manifest differently among individuals. These insights could further our understanding of the brain's role in stress response and aid in developing VR Exposure Therapy methods, affective computing techniques, and diagnostics for stress-related disorders.

5.3.1.1. Methodology

The experimental protocol, the EEG and ECG data processing, the feature extraction and the statistical analysis is explained in this section. A flowchart of the experiment is presented in Figure 43.

(a) Subjects

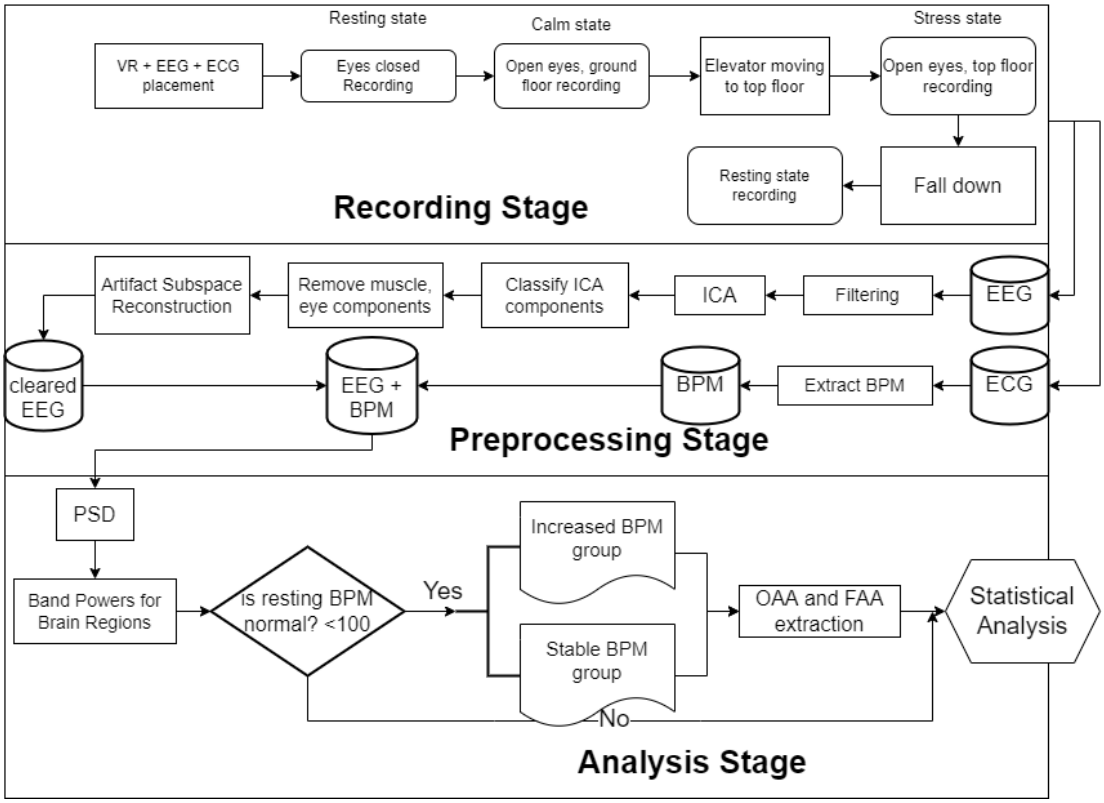


Figure 43 Flowchart of the experiment.

For this experiment, we selected twenty-one participants aged between 20 and 27 years, comprising 8 females and 13 males, all with normal or corrected-to-normal vision. None had prior familiarity with the VR scenario utilized (Richie's Plank Experience steam game, reference), and their experience with HMDs was minimal to none. While participants were aware that the experiment involved a VR experience, they were not informed about its specific nature or the research goal of assessing stress. Out of the 21 recordings obtained, only 18 were considered suitable for inclusion in the study.

(b) Experimental protocol

For the VR stimulus in this experiment, a Meta Quest 2 VR device was used, while EEG recordings were captured using a DSI-24 wearable EEG device, equipped with 21 electrodes. The Quest 2 VR headset provides a resolution of 1920x3664 with 773 PPI and offers a frame rate ranging from 60 to 120 Hz. It features a virtual grid created by inside-out tracking, delineating the user's movement boundaries. The DSI-24, manufactured by Wearable Sensing in San Diego, CA, USA, is a wireless EEG headset with dry electrodes. Electrodes Fz, F3, F4, Cz, C3, C4, T7, T8, Pz, P3, P4, P7, P8, O1, O2, A1, and A2 were

positioned as per the 10-20 international standard. However, four electrodes originally intended for F7, F8, Fp1, and Fp2 were moved upwards to accommodate both the VR and EEG headsets comfortably, leading to their exclusion from this study. Additionally, two electrodes placed under the heart on the sternum recorded ECG activity. These ECG electrodes, included in the DSI-24 bundle, connect directly to the headset, eliminating the need for EEG and ECG signal synchronization. The sampling rate was set at 300 Hz, and electrode impedances were maintained below 5KΩ throughout the study. The Cz electrode was used as the ground for the EEG recordings.

The Richie's Plank Experience steam game is a VR simulation that transports players to the top of a skyscraper, challenging them to walk and ultimately jump off a plank protruding from the building. In our experiment, a real wooden plank, matching the dimensions of the virtual one, was placed on the floor for participants to physically walk on. Initially, participants were equipped with the EEG, ECG, and HMD devices and given time to acclimate to the equipment. They were then asked to stand with their eyes closed for 1-2 minutes, during which a resting state EEG recording was taken. Subsequently, they were instructed to open their eyes, finding themselves in a virtual elevator at the ground floor of a building. Participants were encouraged to explore their surroundings from within the elevator, while the EEG and ECG devices continued recording. They then pressed a button inside the elevator, which took them to the top floor, where the door opened to reveal the plank. At this moment, they realized the existence of the actual plank in front of them. Participants were instructed to walk across the plank and then jump off.

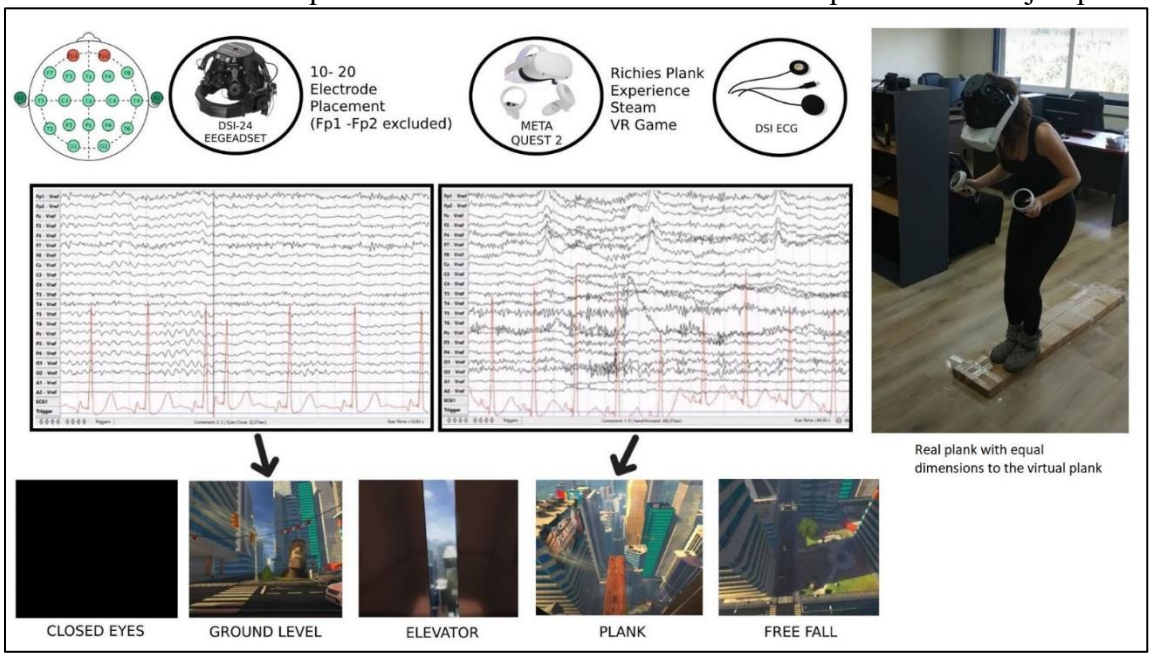


Figure 44 Experimental design of the EEG+VR stress protocol.

This sequence – elevator ascent, plank walk, and jump – constituted the stress-inducing part of the experiment. Post-experiment, participants were informed about the study's focus on stress assessment through VR HMD and were asked to complete the Perceived Stress Scale questionnaire. An overview of the experimental setup is presented in Figure 44.

(c) Preprocessing

For preprocessing, the EEGLAB Matlab Toolbox was utilized. The EEG recordings were re-referenced to the A1 and A2 electrodes positioned on the mastoids. Frequencies between 0.4-48Hz were isolated using a 4th order Butterworth band-pass filter. The EEG data were divided into three separate files: resting state, calm state, and stress state. Artifact rejection involved the use of both ASR and the ICA) method across all signals, employing the FastICA algorithm. A conservative threshold of 17 was selected for ASR as the maximum acceptable standard deviation in a 0.5-second window. In ICA, components identified as eye or muscle artifacts with a probability of 0.9 or higher were automatically excluded. Figure 3 illustrates four distinct Independent Components identified by the "ICLabel" automatic classification routine in EEGLAB. The first two components, identified as eye artifacts, and the third component, identified as a muscle artifact, were removed. The fourth scalp heatmap, representing a brain component, was not removed.

Following this, the signal was segmented into 4-second epochs, and the PSD for each frequency band at every electrode was computed using the Welch method. The frequency bands were categorized as follows:

- Delta: 0.5 – 4 Hz
- Theta: 4 – 8 Hz
- Alpha 8 – 13 Hz
- Beta: 13 – 25 Hz
- Gamma 25 – 45 Hz

In the end, each frequency band was averaged across each brain region.

The preprocessing of the ECG signal involved several steps. Initially, a FIR filter was applied to the signal. This was followed by the use of a peak enhancement function, which normalized the amplitude and accentuated the R-peak amplitude relative to other parts of the signal. For R-peak detection, an adaptive peak detection threshold was employed. Once the peaks were identified, a 6-second sliding window technique was used to calculate the Beats Per Minute (BPM) at each time point. Subsequently, the BPM signal underwent a smoothing process using 3-second averaging windows. The ECG signal was also divided into three segments: resting state, calm state, and stress state. The average BPM was calculated for each of these segments.

(d) Feature extraction

In this section, we present the computed EEG and ECG metrics for each participant. Concerning the EEG, the power of each band, calculated for every 4-second time window, was averaged across the three distinct states. Subsequently, the average power for each band was determined for specific brain regions. These brain regions are defined as:

- Occipital= {O1,O2}

- Temporal= {T3, T5, T6, T4}
- Parietal= {C3, Cz, C4, P3, Pz, P4}
- Frontal= { F7, F3, F4, F8}

The difference of band power across each region and each band was calculated as the band power at the stress state, minus the stress state at the calm state. In each state, the PSD for each band is calculated over distinct 4-second windows and subsequently averaged. The Welch method, used for estimating the PSD, segments the signal into non-rectangular windows utilizing the Hamming method. Finally, for L time windows, we define the periodogram for each window as:

$$Y_i(\omega) = \frac{1}{PQ}\left|\sum_{n=0}^{P-1} x_i(n)c(n)e^{-j\omega n}\right|^2 \tag{5.13}$$

And the average energy of each window Q is calculated as:

$$Q = \frac{1}{P}\sum_{n=0}^{P-1} c^2(n) \tag{5.14}$$

(e) Asymmetry Measures

The FAA and OAA were calculated. FAA is a widely employed metric to express the asymmetry of the frontal cortex and is calculated as [309]:

$$FAA = log(F4) - log(F3) \quad \text{or} \quad FAA = log(F8) - log(F7) \tag{5.15}$$

A positive FAA signifies increased Alpha power on the right side. Studies indicate that right-sided Alpha power correlates with heightened activation of the left hemisphere of the brain, and conversely for the left side [23]. Therefore, a positive FAA value is indicative of left-hemispheric activation. In this research, the asymmetry score is determined using the F4-F3 electrode combination. Following the calculation of the PSD for each time window, the FAA score is computed.

Like the FAA, the OAA metric was calculated as:

$$OAA = log(O2) - log(O1) \tag{5.16}$$

5.3.1.2. Results

The dissimilarities of the different states presented through a statistical approach through the different EEG, ECG and questionnaire metrics were the quantifiable results of this study. First of all, an indicative BPM diagram throughout the experiment was presented to illustrate the different stages of it as well as the alterations on the heart rate caused due to stress. Figure 45 presents this BPM diagram. This observation reveals that the participant experienced heightened heart activity during the stressful state (indicated by the light blue area), confirming that being on the top floor of the skyscraper was indeed a stressful experience.

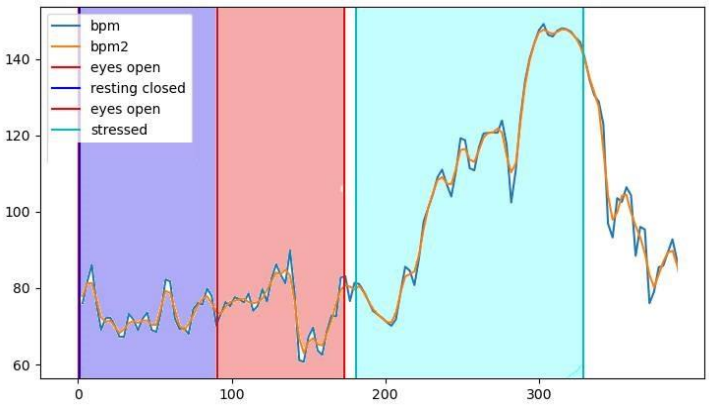


Figure 45 BPM of a participant throughout the experiment.

This study conducted a comparative analysis of Alpha, Beta, and Gamma band powers across calm and stress states in all brain regions. Figure 46 showcases a comparison of BPM across subjects between the calm state (ground floor of the elevator) and the stress state (top floor). It also presents the analysis of Alpha, Beta, and Gamma bands in each brain region. Notably, the occipital area exhibited significantly higher engagement than other regions during the stress condition. To assess the significance of changes in ECG and EEG biomarkers, statistical analyses were carried out on various parameters: Alpha, Beta, and Gamma power in the Frontal, Parietal, Temporal, and Occipital regions, along with OAA, FAA, and BPM. Initially, a Kolmogorov-Smirnov test confirmed the normal distribution of each marker. Subsequently, a paired T-test was used to determine the statistical significance of changes between calm and stress states for each marker. The paired T-test results, detailed in Table 28, indicated statistically significant changes in the Alpha, Beta, and Gamma bands of the Parietal, Temporal, and Occipital regions, as well as in BPM, with two-sided p-values of the paired T-test being less than 0.05. Changes in the Alpha and Beta bands in the Frontal region were not statistically significant, although one-sided p-values for Frontal Alpha and Beta were less than 0.05. As anticipated, BPM

changes were statistically significant. However, changes in FAA and OAA did not yield statistical significance.

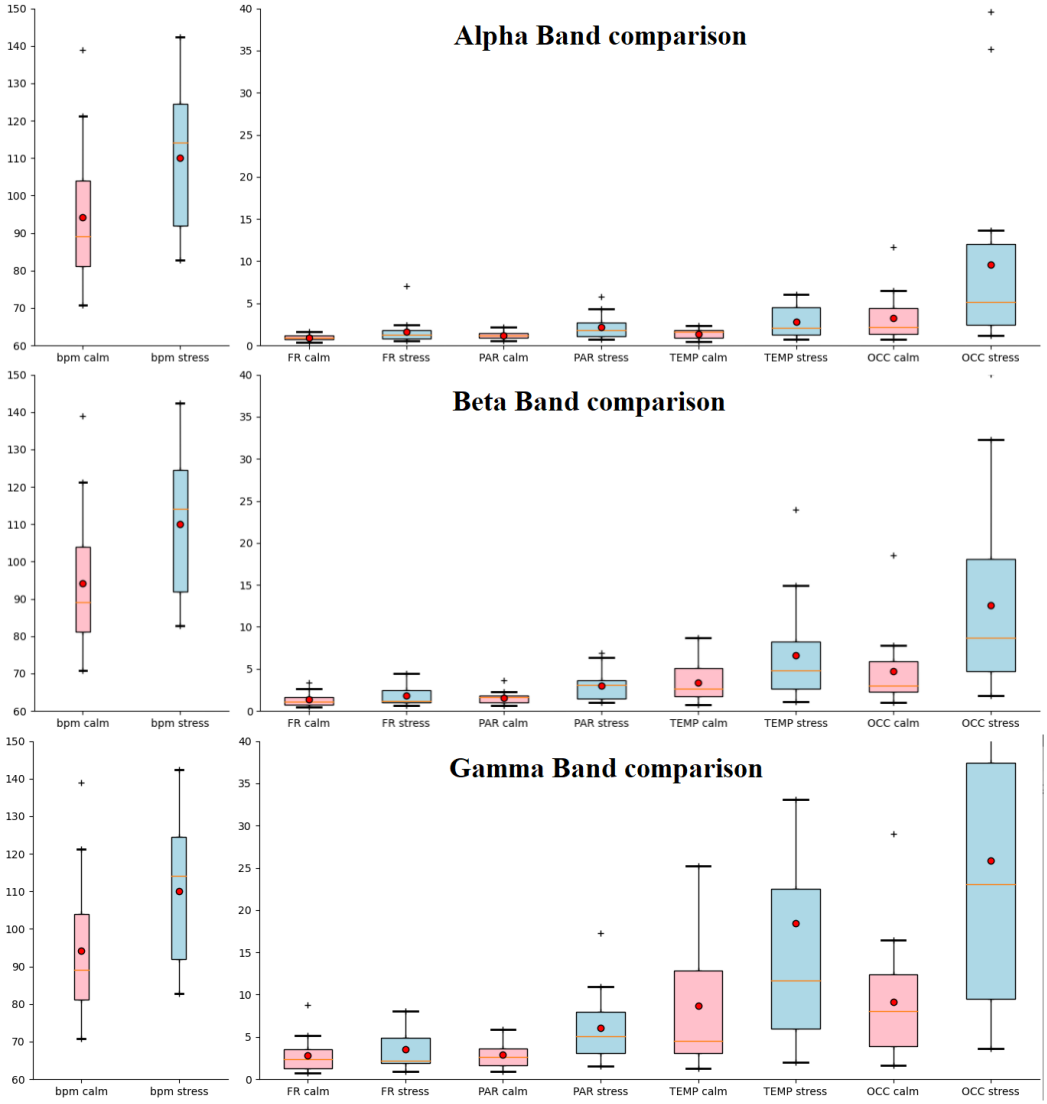


Figure 46 Calm and Stress state comparison of BPM and Alpha, Beta and Gamma Bands.

Table 28 Paired T-tests regarding the alteration of each measure between the states calm and stressed. The * symbol indicates statistical significance with p value < 0.05

	Calm-Stress	t	One-Sided p	Two-Sided p
Frontal	Alpha	-2.036	0.030*	0.060
	Beta	-1.890	0.039*	0.078

	Gamma	-1.239	0.117	0.234
Parietal	Alpha	-3.620	0.001*	0.003*
	Beta	-4.265	0.000*	0.001*
	Gamma	-4.359	0.000*	0.001*
Temporal	Alpha	-3.807	0.001*	0.002*
	Beta	-3.315	0.002*	0.005*
	Gamma	-3.039	0.004*	0.008*
Occipital	Alpha	-2.701	0.008*	0.016*
	Beta	-3.823	0.001*	0.002*
	Gamma	-4.506	0.000*	0.000*
	BPM	-4.327	0.000*	0.001*
	FAA	-0.599	0.279	0.557
	OAA	1.008	0.164	0.328

The significant variance in occipital activation highlighted in Figure 5 necessitated an investigation into whether this activation was associated with stress induced by visual stimuli. To explore this, we divided participants into two groups. The first group comprised individuals with a normal baseline heart rate (BPM<100) that did not increase by more than 13 BPM under stress. The second group included those with a normal baseline heart rate but experienced an increase exceeding 13 BPM during stress; this 13 BPM (0.22 Hz) threshold was established based on a review of relevant literature [322]. An increase in BPM beyond 13 was deemed significant, whereas an increase less than 13 was considered insignificant. Group 1 consisted of 7 participants, while Group 2 had 9. Two participants were excluded due to a high baseline BPM. Figure 47 presents a scalp heatmap contrasting brain activity between a member of Group 1 and one from Group 2, with color coding reflecting deviations from the average Power Spectral Density for each band, measured in 10*log10(μV2/Hz), ranging from -8 (deep blue) to +8 (deep red). Figure 48 showcases a comparison of occipital activity, OAA and FAA change between

the two groups, suggesting that occipital activity could serve as an indicator of stress system activation in response to visual stimuli.

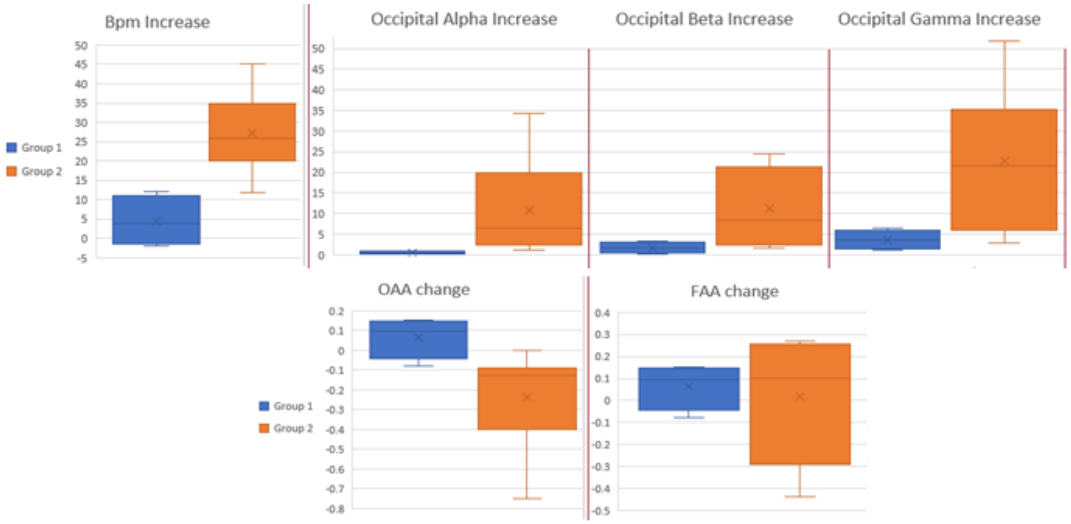


Figure 48 Alterations in Occipital activity, OAA and FAA between the 2 groups (calm group, stressed group)

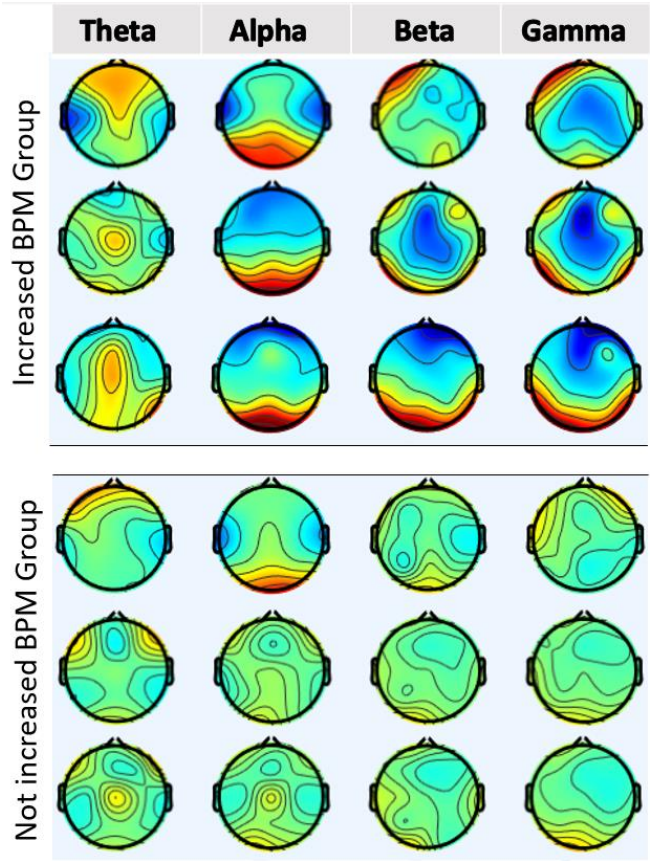


Figure 47 Comparison of brain heatmaps between subjects from the two groups, focusing on the Theta, Alpha, Beta, and Gamma energy bands.

A paired T-test was performed in the entire participant pool to examine if the change in the OAA and FAA scores was statistically significant. Then, the same paired T-tests

were performed for each group separately. The results of the tests can be found in Table 29. As observed, no statistically important alterations were found for the group 1. However, group 2 has shown statistically significant changes in OAA.

Table 29 Paired T-test in asymmetry scores for groups. The symbol * indicates statistical significance with p value <0.05

	OAA calm-Stress			FAA calm-Stress		
	t	One-Sided p	Two-Sided p	t	One-Sided p	Two-Sided p
Group 2	2.733	0.015*	0.029*	-0.203	0.422	0.845
Group 1	-1.971	0.072	0.143	-1.269	0.147	0.294

The next step was to examine whether the alteration of each power band was correlated to the alteration of the BPM between the calm and stress states, using a Spearman Correlation test. Interesting observations were made, as presented in Table 30

Table 30 Spearman Correlation of the alteration of each power band with the alteration of BPM

Frontal				
Delta	Theta	Alpha	Beta	Gamma
0.16	0.18	0.37	0.23	0.24
Temporal				
Delta	Theta	Alpha	Beta	Gamma
0	0.26	**0.47**	0.32	0.31
Parietal				
Delta	Theta	Alpha	Beta	Gamma
0.17	0.079	0.44	0.31	0.22
Occipital				
Delta	Theta	Alpha	Beta	Gamma
0.64*	**0.5***	**0.55***	0.44	0.43

5.3.1.3. Discussion

This research delved into the EEG characteristics of the human brain under stress, particularly at high altitudes, simulated within a VR setting. Heart rate measurements were utilized as a corroborative indicator of stress presence. The study focused on the power of each brain region's band and brain activation asymmetries in relation to EEG data.

In recent years, VR environments have been increasingly employed to elicit emotional responses in studies [314], [323]. Common measures for assessing stress in VR research

include eye tracking [324], questionnaires [325], respiration signals, and ECG [326], while the use of EEG has been less frequent [48]. One major challenge has been integrating EEG with VR headsets due to practical difficulties. For instance, Hu et al. conducted a study using Richie's Plank Experience software to categorize participant fear levels through EEG [327]. Similarly, Fadeev et al. analyzed the impact of different VR scenarios on subjects with intense emotional responses due to health conditions [314]. However, our study stands out by combining two crucial elements: 1) a well-documented biomarker for verifying stress, and 2) continuous EEG recording during the experiment, not just at fixed intervals.

Several studies have utilized EEG signals for mental stress assessment, as outlined by Katmah et al. [328]. Techniques like connectivity methods, spectral analysis, and asymmetry characteristics have been investigated as stress markers. Frontal Alpha Asymmetry and Frontal power have also been explored as indicators of stress [329], [330]. However, these studies often involved mathematical or social challenges as stressors [331], whereas our research associates stress with visual stimuli, anticipating increased occipital activity. We introduce an uncommon measure, occipital Alpha asymmetry, to explore its connection with stress.

Frontal Alpha EEG asymmetry, indicative of relatively stronger neural activity in the left frontal cortex compared to the right, is recognized in psychological research as a key marker in affective processing. It's often considered either a predictor or an outcome variable linked to motivation, emotion regulation, and psychopathology. However, its relation to complex neural dynamics and psychological mechanisms means that its replicability across different experimental setups and study populations is inconsistent. The connection between FAA and stress system mobilization remains a subject of ongoing research. Current evidence does not establish a link between FAA and other stress induction markers [332], [333], and resting FAA has not been consistently identified as a reliable marker for post-traumatic stress disorder. In line with these findings, our study did not observe a significant presence of FAA during the stress-inducing segment of the experiment.

In this study, we noted the occurrence of occipital Alpha EEG asymmetry during the experiment's stressful phases. The significance of this finding remains uncertain, calling for additional research, as there are limited sources discussing the physiological or psychological implications of OAA. OAA, however, has been associated with the activation of the behavior inhibition system (BIS) - a neuropsychological system that influences an individual's reaction to anxiety-inducing stimuli in their environment [334]. We speculate that the BIS may have been engaged during the experiment's high-stress moments, particularly when participants were walking on the narrow plank and facing the challenge of jumping from a great height.

Recent meta-analyses suggest that heart rate variability (HRV) in the context of stress may indicate the extent of integration between higher-level cortical systems and the brainstem nuclei regulating the heart [335]. Stress-induced variations in HRV could mirror the neural processing differences in response to stress. Accordingly, we divided

our study participants based on their stress-induced heart rate changes to examine if EEG markers varied between these groups. Our data, highlighting strong correlations between occipital EEG rhythms and heart rate and differences in occipital EEG activity and OAA between groups, further endorse the notion that these EEG changes are stress-related and possibly associated with inherent traits.

However, there are limitations to this study. The integration of the EEG device with the VR headset was complex, leading to the exclusion of the four frontal electrodes to ensure proper fitting for all participants. Additionally, individuals with a head circumference smaller than 54 cm were not included due to device size constraints. Time limitations also restricted the size of our participant pool. Other potential stress indicators, like HRV features, were not analyzed at this stage. Also, some collected data, such as the Perceived Stress Scale questionnaire responses, were not included in the experimental protocol.

Looking ahead, our research plans include further exploration of EEG as a stress indicator. We aim to integrate HRV features into our methodology to uncover potential correlations between EEG attributes and stress mechanisms. We also plan to develop a method for real-time, automatic stress detection from EEG signals using advanced Machine Learning techniques. This approach may incorporate Perceived Stress Scale questionnaire scores to predict stress levels accurately. Finally, we are looking to expand our study to include more virtual environment stressors to examine various stress manifestations.

5.3.2. Classification of EEG signals from Young Dyslexic Adults combining a Brain Computer Interface device and an Interactive Linguistic Software Tool

This article was published in the journal Elsevier Biomedical Signal Processing and Control, during March 2022. The hypothesis of a magnocellular pathway deficit as a potential cause of dyslexia offers an alternate explanation for auditory and visual processing challenges. Numerous studies have leveraged machine learning techniques with anatomical brain imaging to categorize these deficits, proving EEG graph measures to be both robust and dependable. This study aims to classify university students with and without dyslexia using a Brain Computer Interface (BCI) Device and an Interactive Linguistic Software Tool, to confirm the effectiveness of such technology in identifying dyslexia within a higher education setting. We recorded EEG signals from 12 university students diagnosed with dyslexia and 14 typically developed, age-matched peers using a wearable, sensory EEG recording device. The participants were tested under three experimental conditions: a) auditory discrimination, b) visual recognition, and c) visual recognition with background music. Spectral features from each EEG rhythm (δ, θ, α1, α2, β1, β2, γ) were utilized to train a Random Forests classifier, with the goal of pinpointing EEG features that distinctly mark dyslexia across various brain regions. The results demonstrated high accuracy, sensitivity, and specificity (exceeding 95%) throughout the brain, particularly in the left and right hemispheres. The most notable discrimination efficacy was observed in the third experimental condition, involving background music. Each experimental setup yielded high classification accuracy, effectively distinguishing between higher education students with and without dyslexia.

Dyslexia, a prevalent specific developmental learning disorder, affects an estimated 5–15% of students, though this figure varies widely due to differing criteria in reading assessment scales and diverse theoretical perspectives on dyslexia. Characterized as a specific deficit in reading acquisition not attributable to low IQ, inadequate education, or clear sensory or neurological impairments, dyslexia has been the subject of numerous theories, including the phonological and cerebral deficit hypotheses.

Despite most dyslexia research focusing on children and its continuation into adolescence and adulthood, there remains a significant gap in literature regarding the persistence of reading difficulties into later life, with adult studies constituting only about 6% of all dyslexia research. Notably, studies integrating neuroimaging techniques, particularly EEG, in adult dyslexia research are rare, suggesting that neurocognitive deficits, including connectivity issues, continue into adulthood. EEG, a non-invasive brain function measure, has been extensively used to explore brain behavior and offers further

---

The author list ordered as published was: Pavlos Christodoulides, **Andreas Miltiadous**, Katerina D. Tzimourta, Dimitrios Peschos, Georgios Ntritsos, Victoria Zakopoulou, Nikolaos Giannakeas, Loukas G. Astrakas, Markos G. Tsipouras, Konstantinos I. Tsamis, Euripidis Glavas and Alexandros T. Tzallas

understanding of cortical lateralization models. The application of Machine Learning Classifiers to identify dyslexia-specific brain patterns has been explored, and the role of musical elements perception in this context, though under-researched, could aid in differentiating individuals with and without dyslexia.

This study focuses on classifying university students with and without dyslexia. We analyze EEG recordings from both dyslexic students and age-matched, typically developed peers, utilizing a Brain Computer Interface Device to ascertain its efficacy in detecting dyslexia within a higher education setting. Time domain features are extracted to form a feature vector, which is then used to train classifiers for nine different brain Regions of Interest (RoI). The accuracy levels of these classifiers were assessed across three different experimental conditions detailed in the methodology section. To our knowledge, this is the first comprehensive study of university students with dyslexia that examines a broad range of features across various experimental settings, achieving such high classification accuracy.

5.3.3. Methodology

The methodology consists of two parts. First, the EEG signal recording and acquisition protocol is described, where a wearable device is used to capture EEG signals from 26 participants while using an interactive linguistic software, and then spectral features are extracted and used as features to train a Random Forest classifier for dyslexia detection.

(a) Participant description

This research involved 26 right-handed students from the University of Ioannina, Greece, who participated in an experiment assessing performance through a novel interactive application. This application measured audiovisual recognition and discrimination of words under three experimental conditions. The participants were divided into two groups based on the presence or absence of learning difficulties. The study included 12 students (6 females and 6 males) diagnosed with dyslexia, and a control group of 14 individuals (10 females and 4 males). All dyslexic participants had received early intervention, and no comorbidities related to dyslexia were reported. The age range was similar across both groups, with the dyslexia group averaging 21.58 years old and the control group 20.93 years old. Education levels were uniform across participants, as all were university students. Written informed consent was obtained from all participants before the commencement of the study.

(b) Linguistic software tool description

The application evaluates participants' responses across three experimental setups: (a) audio recognition, (b) visual discrimination, and (c) visual discrimination with background music. The tasks in these conditions were designed based on typical dyslexia-related errors in the Greek language, particularly those involving visually similar letters (κ, γ, χ) or acoustically similar sounds (f, v, θ, ð). In each scenario, participants had to

identify the correct word from a set of three options (one real Greek word and two pseudowords), with an average response time of 8.5 seconds.

In the audio recognition condition, participants differentiated between verbally presented words with similar sounds (e.g., fo´vame, fo´δame, fo´θame). They viewed three numbered boxes on the screen and selected the number corresponding to the correct word. The visual discrimination condition required subjects to read words displayed on the screen, paying close attention to choose the correct one. The incorrect options included common dyslexic errors like sequential (fridge - frigde), insertion (computer – compluter), omission (bicycle - bicyle), or substitution (dog – tog). The final condition paralleled the visual discrimination task, with the added element of background music. For this part of the experiment, Mozart's Sonata for Two Pianos in D major, K. 448, was played, providing an auditory backdrop to the visual task.

(c) Data acquisition and EEG processing

EEG recordings were conducted in a room designed to minimize sound and light, capturing data while each participant engaged in the evaluation test. During the test and EEG recording, subjects remained seated in an upright position, relaxed and in a restful state with their eyes open. The length of each EEG recording varied from 21 to 38 minutes, averaging 28 minutes, and was dependent on the time each subject took to complete the test. Recordings were stopped immediately if any participant experienced discomfort with the device or procedure. In total, 5 hours and 51 minutes of EEG data were collected from dyslexic subjects, and 5 hours and 47 minutes from non-dyslexic subjects, accumulating approximately 11.5 hours of recordings.

The Emotiv EPOC+, a popular commercial wearable EEG device, was employed for these recordings. This device features 14 sensors with corresponding felt pads positioned on the scalp following the International 10-20 System (AF3, F3, F7, FC5, T7, P7, O1, AF4, F4, F8, FC6, T8, P8, and O2). Two additional rubber electrodes were placed on the mastoids, serving as reference channels. With a sampling frequency of 128 Hz, electrode-scalp connectivity was maintained using a saline solution applied to the sensor pads. The setup was carried out as per EmotivPRO Software guidelines, with regular checks for connectivity quality both at the start and during the recording.

For EEG signal preprocessing, recordings were made using the linked mastoids montage. Post-recording, the EEG data was exported in ".edf" format and processed using MATLAB and the EEGLAB toolbox. A Butterworth notch filter was applied to eliminate 50 Hz power line noise oscillations, and a high-pass FIR digital filter at 0.5 Hz was used to remove low-frequency oscillations. Subsequently, seven equiripple FIR filters were designed to isolate specific frequency ranges corresponding to the 5 EEG rhythms, thereby enabling extraction of spectral features from each frequency sub-band. The EEG recordings were then segmented into 10-second non-overlapping epochs for feature extraction.

(d) Feature extraction and Classification

Energy for every band of interest from the FFT transformation and Shannon Entropy from the entire spectrum were extracted (for each 10 second epoch) to create the feature vector that was then used for the training of the classifier.

For the classification, a Random Forests classifier was employed, and the evaluation is performed on the binary classification problem Control versus Dyslectic. Furthermore, other well-established classifiers namely Naïve Bayes, kNN, Decision Trees, SVM and MLP are examined for a performance comparison. To evaluate the effectiveness of each brain region on the total classification task, a separate classification is performed using only the electrodes of each unique brain region. Thus, in total, 9 classification tasks are performed from the respective RoIs, namely:

- Entire brain (AF3, F3, FC5, T7, P7, O1, AF4, F4, FC6, T8, P8 and O2)
- Left hemisphere (AF3, F3, FC5, T7, P7, O1)
- Right hemisphere (AF4, F4, FC6, T8, P8 and O2)
- Left frontal (AF3, F3)
- Left temporal (T7, FC5)
- Left occipital (O1, P7)
- Right frontal (AF4, F4)
- Right temporal (T8, FC6)
- Right occipital (O2, P8)

5.3.3.2. Results

The performance results were reported in terms of Accuracy, Sensitivity and Specificity. The performance metrics in terms of ACC are presented in Figure 49. Moreover, the comparison of the accuracies for every RoI are presented in Figure 51. A ROC curve analysis has taken place for every RoI and for every condition and the results of it are reported in Figure 51. The last figure in this study presents an overview of all the performance metrics for all RoIs and all conditions. The numeric results of the Random Forest classifier for each RoI are separated in 3 tables based on the condition and are presented in Table 31, Table 32 and Table 33.

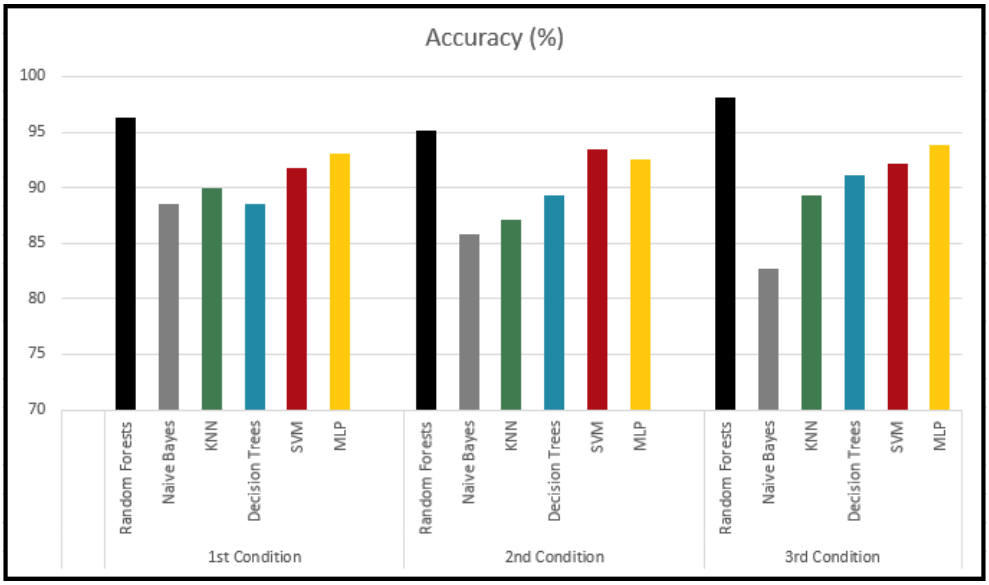


Figure 49 Accuracy scores of each classifier for the Dyslexia-Control problem.

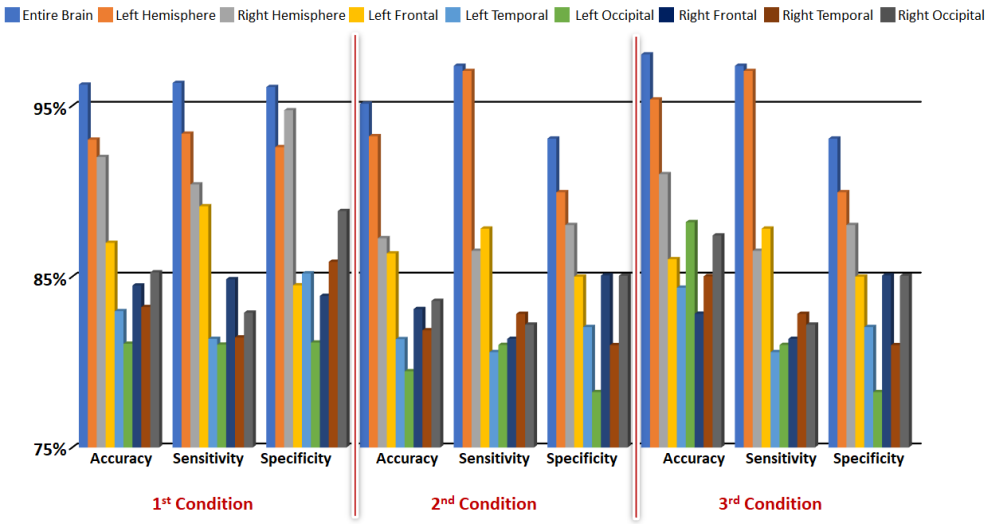


Figure 50 Accuracy, Sensitivity, Specificity scores for every RoI in every condition

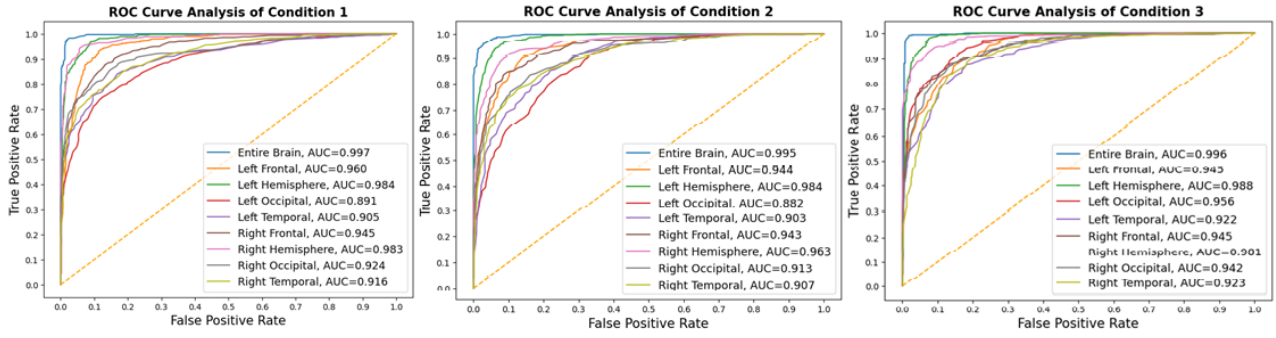


Figure 51 ROC curves in different brain regions

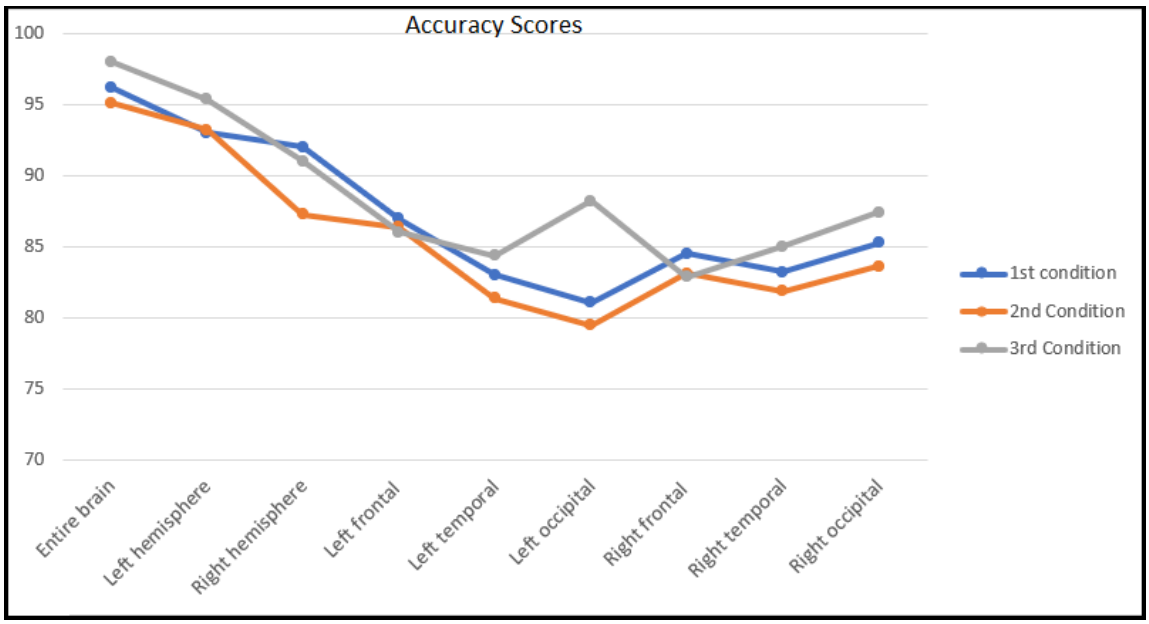


Figure 52 Comparison of the Accuracy scores for the different RoIs

Table 31 Classification results for each Region of Interest for the 1st condition

RoI	ACC (%)	SENS (%)	SPEC (%)
Entire brain	96.24 (1.66)	96.34 (1.67)	96.1 (2)
Left hemisphere	93.02 (2.7)	93.37 (2.71)	92.58 (2.82)
Right hemisphere	92.00 (2.7)	90.40 (2.73)	94.74 (3.44)
Left frontal	86.98 (3.14)	89.12 (3.14)	84.49 (3.25)
Left temporal	82.98 (3.35)	81.36 (3.36)	85.19 (3)
Left occipital	81.07 (3.6)	81.02 (3.63)	81.14 (3.71)
Right frontal	84.48 (3.27)	84.85 (3.28)	83.87 (3.78)
Right temporal	83.22 (3.79)	81.44 (3.79)	85.87 (4.09)
Right occipital	85.25 (3.58)	82.89 (3.59)	88.84 (3.89)

Table 32 Classification results for each Region of Interest for the 2nd condition

RoI	ACC (%)	SENS (%)	SPEC (%)
Entire brain	95.12 (1.92)	97.34 (1.92)	93.08 (1.91)
Left hemisphere	93.22 (2.17)	97.04 (2.18)	89.94 (2.17)
Right hemisphere	87.25 (3.15)	86.51 (3.16)	88.02 (3.17)
Left frontal	86.36 (3.69)	87.81 (2.94)	85.00 (2.94)
Left temporal	81.34 (3.69)	80.58 (3.7)	82.05 (3.64)
Left occipital	79.46 (2.88)	81.00 (2.89)	78.24 (3)
Right frontal	83.10 (3)	81.36 (3)	85.07 (3.03)
Right temporal	81.86 (3.47)	82.82 (3.47)	80.99 (3.47)
Right occipital	83.58 (3.31)	82.19 (3.32)	85.05 (3.3)

Table 33 Classification results for each Region of Interest for the 3rd condition

RoI	ACC (%)	SENS (%)	SPEC (%)
Entire brain	98.01 (1.53)	98.73 (1.53)	97.39 (1.62)
Left hemisphere	95.37 (2.18)	97.41 (2.18)	93.39 (2.14)
Right hemisphere	91.00 (2.81)	90.75 (2.81)	91.23 (2.84)
Left frontal	86.02 (3.54)	89.77 (3.54)	82.68 (3.46)
Left temporal	84.35 (3.26)	85.27 (3.26)	83.40 (3.25)
Left occipital	88.19 (2.87)	90.38 (2.88)	86.07 (2.9)
Right frontal	82.83 (3.39)	82.49 (3.4)	83.12 (3.53)
Right temporal	85.02 (3.4)	87.17 (3.4)	82.91 (3.37)
Right occipital	87.41 (2.88)	87.90 (2.89)	86.88 (2.89)

5.3.3.3. Discussion

In this study, we developed a methodology to identify EEG patterns associated with dyslexia across different brain regions. We analyzed EEG data from 12 dyslexic individuals and 14 non-dyslexic subjects during active tasks. Spectral features extracted from electrode pairs and clusters formed the basis of a feature vector used to train a Random Forests classifier, distinguishing between dyslexic and control groups.

Our EEG-based approach yielded notable results in identifying the dyslexic group. The highest classification accuracy, sensitivity, and specificity were observed across the entire brain, followed by the left and right hemispheres, particularly in the left frontal region. Additionally, the most effective classification occurred during the third condition, where participants visually discriminated word groups with background music.

This study also evaluated the Emotiv EPOC+, a commercial EEG wearable device, which demonstrated effective discrimination capabilities. This research highlights the

potential of lifestyle EEG devices to capture brain dynamic differences between healthy young adults and age-matched individuals with learning disorders, marking a pioneering use of lightweight wearable technology for dyslexia analysis.

Despite yielding several statistically significant findings with high accuracy in classifying the two subject groups, there are notable limitations to this study. Firstly, the sample size, comprising 12 dyslexic students and 14 controls, may seem small. Secondly, the study's group categorization was based on previous formal diagnoses of reading and writing difficulties. It's important to note that university students with dyslexia generally have more extensive training in language and cognitive skills than other young dyslexic individuals. Nonetheless, the classification in this study effectively identified dyslexia patterns, underscoring the persistence of their learning challenges. Last but not least, the lack of Leave One Subject Out validation almost certainly overestimates the performance of the classifier.

A significant practical issue encountered was the detachment of electrodes during recordings. Despite regular connectivity checks, we had to limit spatial information. Yet, the classification performance remained satisfactory. The results indicate that dyslexia can be detected with about 70% accuracy, even with a limited number of channels from the occipital/parietal region.

5.3.4. An ensemble method for EEG-based texture discrimination during open eyes active touch

This article was published in the open access journal Engineering, Technology & Applied Science Research, during January 2024. Touch is a crucial sensory modality enabling humans to comprehend and interact with their surroundings. Often, touch relies on visual input to consolidate and verify sensory information. Discerning various materials and surfaces through touch involves intricate neurophysiological processes. To explore these processes, neuroimaging and neurophysiological tools are used, with EEG being the most prevalent. In this study, we aim to differentiate brain responses to touching diverse natural textures (smooth, rough, and liquid). Data were gathered using a commercially available EEG wearable device. We extracted time and frequency-based features, processed them with PCA decomposition, and utilized them to train a composite classifier comprising Random Forests, Support Vector Machine, and Neural Network. We achieved high accuracy rates of 79.64% for a four-class problem and 89.34% for a three-class problem (Null-Rough-Water). These results demonstrate the effectiveness of our methodology in classifying distinct brain states elicited by various haptic stimuli.

5.3.4.1. Introduction

Touch is a fundamental human sense, critical for interacting with the environment and understanding material properties like shape, size, texture, and temperature. The sense of touch involves complex neurophysiological processes, where mechanoreceptors in the skin convert physical contact into nerve signals that reach the brain for interpretation. Haptic perception, integrating touch with proprioceptive and thermal information, allows for both discriminative and affective understanding of surroundings.

Haptic perception encompasses both active and passive touch, with ongoing research investigating the neural and perceptual differences between these forms. Its importance extends beyond basic material distinction, contributing to emotional, social, and cognitive understanding. This growing research area has significant applications in medical, industrial, and technological fields, ranging from rehabilitation and assistive robotics to affective product design and haptic interfaces.

Non-invasive methods like fMRI and EEG are key in studying the brain's response to touch. EEG's portability and high temporal resolution facilitate novel experimental scenarios in haptic research, employing techniques like PSD and ERP’s. Machine Learning advancements aid in classifying brain states under various haptic stimuli, examining aspects like roughness recognition and tactile pleasantness.

This study aims to explore the multisensory nature of haptic perception by distinguishing brain responses to different materials during active touch with constant

---

The author list ordered as published was: **Andreas Miltiadous**, Vasileios Aspiotis, Dimitrios Peschos, Katerina D. Tzimourta, Al Husein Sami Abosaleh, Nikolaos Giannakeas, Alexandros T. Tzallas

visual contact. We propose a methodology using EEG signal classification to discriminate between natural material textures. Our experimental protocol includes EEG acquisition and preprocessing, feature extraction, PCA, and the use of multiple classification algorithms. An ensemble classification method, combining various algorithms, demonstrates superior performance in classifying EEG signals under diverse haptic stimuli.

5.3.4.2. Methodology

A brief description of the proposed methodology can be extracted from the flowchart in Figure 53.

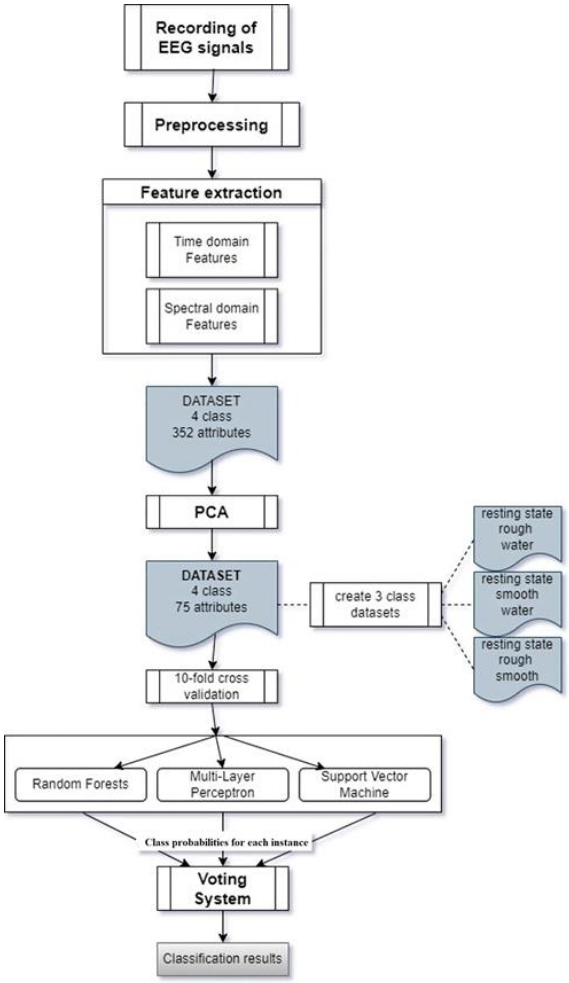


Figure 53 Flowchart of the methodology presented in the active touch classification study.

(a) Dataset

The experiment was conducted in a quiet, clinical setting. Participants, seated comfortably, were briefed on the procedure and allowed time to acclimatize to the device. Their dominant hand (confirmed via the Edinburgh Handedness Inventory [336]) was placed on an ergonomic arm support. They were instructed to gently rub different textures with their fingertips for one minute each, in a circular clockwise motion, maintaining visual contact with the materials. To ensure consistent movement and pressure, participants' arms were secured to the support with a rubber band, positioned at a height parallel to the surfaces. An experienced researcher supervised the entire process. The protocol involved:

1. 1-minute eyes-open resting-state recording
2. 1-minute recording during active texture rubbing smooth surface
3. 1-minute rest
4. 1-minute recording during active texture rubbing rough surface
5. 1-minute rest
6. 1-minute recording during active texture touching liquid surface

In total, 12 subjects participated, of whom 7 were males and 5 females, with ages between 25-27 and no history of neurological or psychiatric disorders. Two materials of different roughness and a liquid surface were used as the surfaces. For the smooth material, a satin polished stainless steel was used with Roughness average Ra < 0.5 μm. For the rough material, 120 Grit sandpaper was used (estimated Ra equivalent = 1.32 μm). Finally, for the liquid surface, room temperature water was used.

The device used was an Emotiv EPOC Flex wearable device with 32 gel Ag/AgCl sensors and sampling rate of 1024 Hz. The electrodes were positioned in the 10-20 International Reference System locations namely Cz, Fz, Fp1, F7, F3, Fc1, C3, Fc5, T9, T7, Tp9, Cp5, Cp1, P3, P7, O1, PZ, OZ, O2, P8, P4, Cp2, Cp6, Tp10, T8, Fp10, Fc6, C4, Fc2, F4, F8, Fp2, and on the earlobes A1 and A2. The impedance value was ensured to be

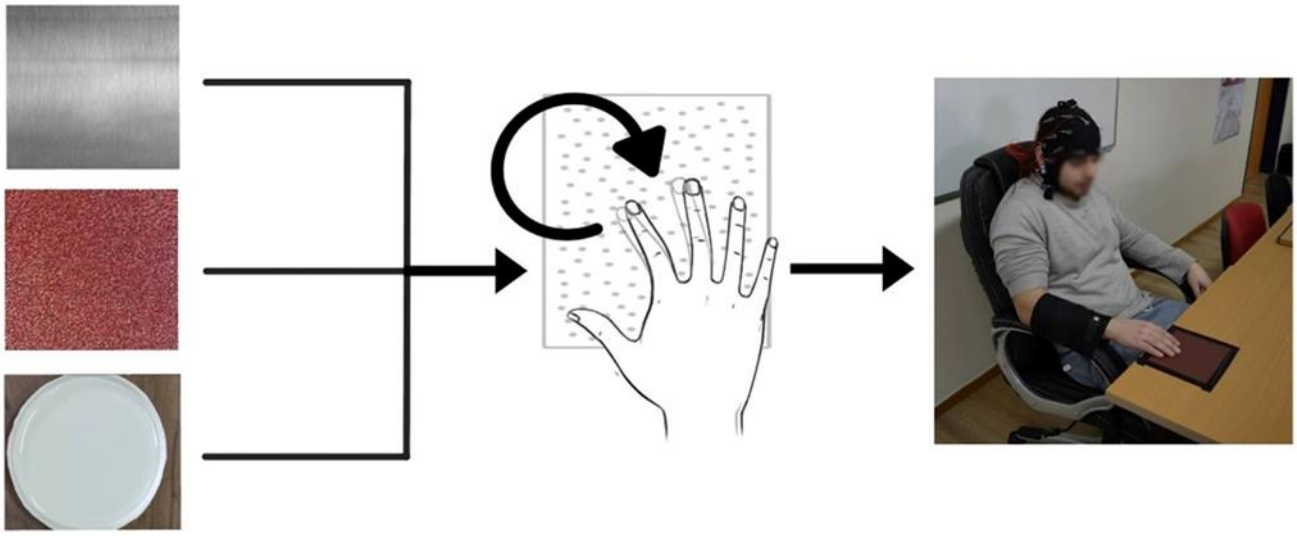


Figure 54 The materials used and the experimental position.

under 20 kΩ for the duration of the recordings. The materials and the experimental position are depicted in Figure 54.

(b) Data Preprocessing and Feature Extraction

The EEG recordings were re-referenced to A1 A2, downsampled to 128 Hz and filtered using a high-pass 4th order Butterworth filter at 0.4 Hz. Low pass filtering was not necessary, due to the pre-installed high pass filter of the Emotiv EPOC Flex firmware at 45 Hz. Finally, time periods with artifacts of electrode movement were manually removed.

Next, epoching of the EEG was performed, creating 1 second epochs with 50% overlap. The Time domain metrics Mean, Variance, Range, Median, Interquartile Range, 30% Percentiles, along with the Frequency domain features Relative Band Power of each of the 5 frequency bands were extracted. In total, 352 features were extracted (6 time domain and 5 spectral domain features for each of the 32 electrodes).

Obviously, this is a large number of features in regards to the size of the dataset, thus the PCA methodology was employed to reduce the dimensionality of the dataset. After keeping the features that account for 95% of the variance in the dataset, the initial dataset of 352 features was transformed to a linear projection of 75 features. This procedure is depicted in Figure 55.

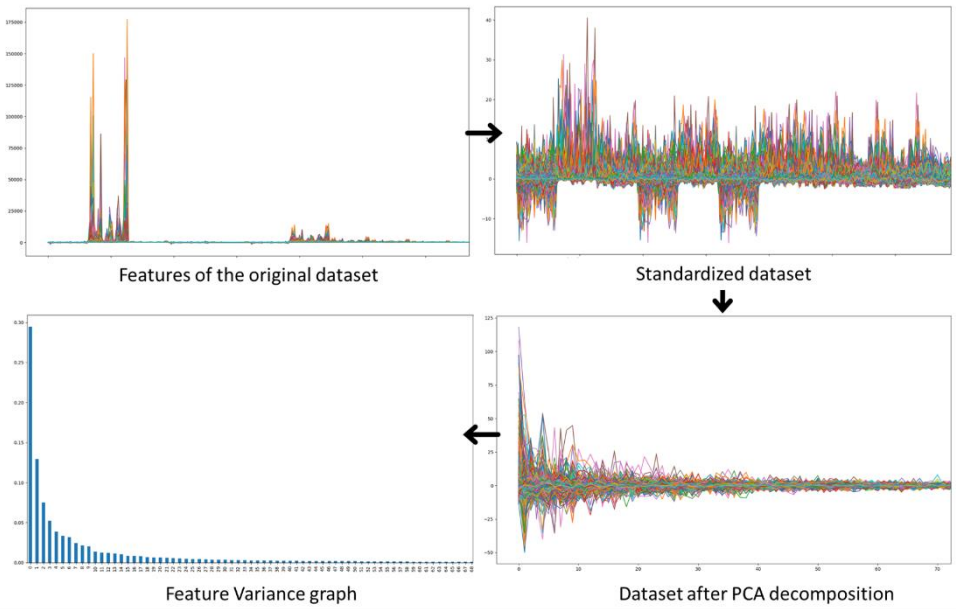


Figure 55 The PCA decomposition in the feature vector.

(c) Classification

For the classification stage, Random Forests, k-NN, MLP, C4.5 Decision Trees, LDA, and SVM were employed. The three best performing were then combined in an ensemble method for optimized performance. The performance results that were extracted were Accuracy, Sensitivity and Specificity and were extracted using a 10-fold validation

method. It should be noted that, in this particular case, k-fold validation does not pose a problem (so a LOSO validation is not required), since every participant performed every experiment, thus there are no subject specific issues that could arise.

The 4 classes were:

1. Resting state or Null (N)
2. Rough (R)
3. Smooth (S)
4. Water (W)

The classification problems examined were:

1. N-R-S-W
2. N-S-W
3. N-R-W
4. N-R-S

The algorithms that were ultimately used for the ensemble method were MLP, Random Forest and SVM, using a probability voting.

5.3.4.3. Results

The results for each of the classification tasks are presented in the Table 34,Table 35,Table 36,Table 37. An overall representation of the performance of every classifier in every problem can be found in

Table 34 Performance results for the N-R-S-W problem

Classification Algorithm	**Accuracy**	**Sd**	**Sensitivity**	**Specificity**
Voting System	79.64%	2.19	77.30%	92.40%
Random Forests	70.86%	2.32	70.80%	90.20%
Decision Trees	47.14%	2.83	47.10%	82.30%
KNN	64.90%	2.52	67.00%	89.00%
LDA	67.80%	2.35	67.80%	89.20%
SVM	76.50%	2.21	76.40%	92.10%
MLP	77.06%	2.33	77.00%	92.30%

Table 35 Performance results for the N-S-W problem

Classification Algorithm	Accuracy	Sd	Sensitivity	Specificity
Decision Tree	55.64%	3.64	57.10%	78.50%
Random Forests	79.30%	2.51	79.60%	89.80%
KNN	71.22%	2.94	72.20%	86.00%
MLP	84.70%	2.34	84.80%	92.40%
LDA	81.10%	2.39	81.10%	90.50%
SVM	84.50%	2.26	84.50%	92.20%
Voting System	87.67%	2.00	85.90%	93.00%

Table 36 Performance results for the N-R-W problem

Classification Algorithm	Accuracy	Sd	Sensitivity	Specificity
Decision Tree	60.10%	3.55	60.00%	80.00%
Random Forests	82.56%	2.56	82.50%	91.20%
KNN	74.31%	3.37	77.10%	88.50%
MLP	86.50%	2.36	86.40%	92.00%
LDA	82.25%	2.22	82.10%	91.00%
SVM	86.30%	1.99	86.30%	92.10%
Voting System	89.34%	1.42	85.90%	93.00%

Table 37 Performance results for the N-R-S problem

Classification Algorithm	Accuracy	Sd	Sensitivity	Specificity
Decision Tree	54.27%	3.48	56.00%	78.00%
Random Forests	75.00%	2.67	75.00%	87.40%

KNN	70.06%	3.27	71.40%	85.70%
MLP	78.33%	2.89	78.10%	89.00%
LDA	70.62%	2.74	70.60%	85.30%
SVM	79.13%	2.62	79.20%	89.60%
Voting System	82.10%	2.33	79.80%	89.90%

Next, to evaluate the contribution of each individual classifier in the ensemble methodology, the in-between correlation of errors, as well as the correlation of errors between each classifier and the ensemble method was investigated. In theory, a low correlation among individual classifiers combined with a higher correlation between the Ensemble Method and each individual classifier suggests that the classifiers are not making the same misclassifications. This implies an advantage in utilizing a combination of these classifiers, as evidenced by the comparative performance results. The Mathiews Correlation of them is presented in Table 38.

Table 38 Correlation of errors between the weak classifiers and the voting system

Mathiews Correlation	**SVM**	**MLP**	**RF**	**Voting System**
SVM	---	0.33	0.4	0.75
MLP	0.33	---	0.3	0.564
RF	0.4	0.3	---	0.59

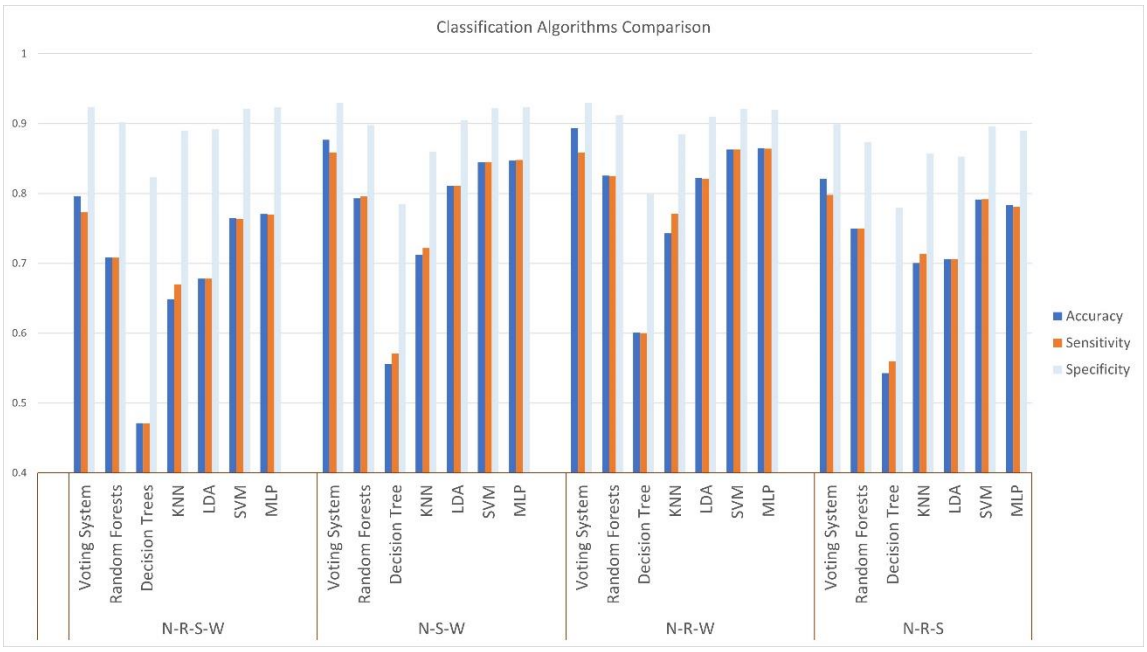


Figure 56 Overview of performance results for the Active Touch experiment.

5.3.4.4. Discussion

This study demonstrates that PCA feature reduction not only shortens training and testing times but also enhances overall accuracy. Specifically, the training time for the voting system in the 4-class classification problem was cut down to 42 seconds with PCA, a significant reduction from the 887 seconds required without PCA. This time efficiency was observed across all classification problems.

The aim of this research was to classify various active touch states during a specific hand movement while maintaining visual contact with the texture. While other studies have explored distinct brain responses to different haptic stimuli, there is no standard protocol for experimental procedures during signal acquisition. Unlike most studies where hand movement is tightly controlled or participants' eyes are kept closed, our study allowed for more natural hand movements and visual contact.

However, this methodology has its limitations. Extending EEG recording duration could lead to more accurate and generalizable results. Analyzing brain activity during the experiment might provide deeper insights into the interaction between specific brain regions, potentially enhancing feature extraction. The individual variability in finger movement, such as speed, strength, and muscle movement, could impact the classification due to the varying activation of joint and muscle mechanoreceptors. The non-randomized experimental protocol also poses a potential limitation, as the trial order might influence classification accuracy due to factors like mental fatigue. Randomizing trial order could introduce other issues, like non-repeatability or altered EEG signals due to moisture from prior water trials. Despite these considerations, we chose a non-randomized protocol as the most suitable approach for this study.

To summarize, we developed a method to classify EEG signals during active touch, achieving significant accuracy across various categorization scenarios. By effectively combining SVM, MLP, and Random Forests in an ensemble method and utilizing PCA for feature reduction, we attained high accuracies: 79.64% in the N-R-S-W problem, 87.67% for N-S-W, 89.34% for N-R-W, and 82.1% for N-R-S. This feature reduction not only expedited computations but also enhanced the accuracy and reliability of our classifications. Our method's time-independent nature underlines its durability for prolonged use, making it suitable for broader applications in fields like prosthetics, robotics, and communication restoration, as well as enriching our comprehension of EEG-based tactile discrimination due to its demonstrated stability and precision.

Conclusions

“It's the job that's never started as takes longest to finish, as my old gaffer used to say.”

-Samwise Gamgee

EEG stands as a pivotal tool in the realm of neurology and clinical neuroscience. Its versatility and non-invasive nature make it an invaluable asset in diagnosing and understanding a myriad of neurological conditions. Notably, EEG's role in epilepsy management, from seizure detection to intervention strategies, highlights its clinical significance. Furthermore, its applications extend beyond epilepsy to sleep disorder analysis, brain-computer interface development, and monitoring cognitive health, illustrating its broad utility in medical practice.

The integration of advanced EEG processing techniques with automated machine learning and deep learning pipelines has marked a significant stride forward in medical diagnostics and research. This fusion of technology and neuroscience paves the way for more sophisticated, automated methodologies that significantly enhance clinical practices. By leveraging the power of machine learning, particularly deep learning, the analysis of EEG data has become more accurate, efficient, and insightful. This progression not only aids in the early detection of neurological anomalies but also contributes to personalized medicine, tailoring treatments to individual patient profiles.

Deep learning, and notably the advent of transformer models, has profoundly transformed EEG data analysis and interpretation. The unique strength of transformers in handling sequential data, coupled with their capacity to discern intricate patterns and relationships in large datasets, has significantly advanced our understanding of brain function and dysfunction. In the context of EEG analysis, the implementation of transformer architectures has been instrumental in developing more robust and precise diagnostic models, enhancing predictive analytics, and driving innovative research in neurology. These advancements are particularly vital in areas where conventional data processing methods prove inadequate, underscoring the transformative impact of deep learning, especially transformers, in the realm of neurological study and diagnostics.

Throughout this research journey, significant contributions have been made to the field, evidenced by the publication of 8 journal articles and 2 conference papers. A notable achievement is the development of a novel deep learning transformer architecture, DICE-net, specifically tailored for Alzheimer's disease research. This innovation represents a substantial advancement in the early detection and study of Alzheimer's disease using EEG data. Additionally, the creation and release of the first open-access EEG database for Alzheimer's patients stand as a testament to our commitment to collaborative research and open science. This database has already had a substantial impact on the scientific community, fostering further research and innovation in the field.

The culmination of this PhD journey reflects a significant endeavor in pushing the frontiers of EEG research and its application in medical science. The hope is that the contributions made will serve as valuable assets in the ongoing quest to understand and treat neurological conditions more effectively. The journey has been a testament to the power of interdisciplinary collaboration, blending advanced computational techniques

with clinical insights. As we celebrate these achievements, the vision for the future is one of continued innovation and exploration, with the ultimate goal of enhancing patient care and advancing our understanding of the human brain.

Regarding Future Insights

The research field of automatic EEG analysis, whether in a medical or commercial context, is far from reaching a plateau. The innovative approaches of deep learning, especially those involving transformer neural networks, promise significant advancements in the field, with highly effective methodologies on the horizon. There is no doubt that EEG's role in the medical area, particularly for screening various conditions, will be significantly enhanced. Moreover, we should anticipate a surge in commercial products leveraging EEG technology for wellness, gaming, or other applications.

One pressing challenge that future research must address is the need for more robust and generalizable models capable of handling the inherent variability in EEG data across different individuals. This variability can stem from physiological differences, electrode placement, or even the recording environment. Developing algorithms that can adapt to or normalize these differences is crucial for both clinical and commercial applications. Furthermore, the integration of multimodal data, combining EEG with other physiological signals, could offer a more holistic understanding of the brain's activity and its implications for health and behavior. Researchers are encouraged to explore novel architectures and training paradigms, such as few-shot learning or domain adaptation techniques, to tackle these challenges effectively.

Additionally, as EEG technology becomes increasingly integrated into everyday devices, ethical considerations and societal impacts must be carefully evaluated. The potential for continuous monitoring raises privacy concerns and questions about data ownership and consent. Future work should also consider the accessibility of EEG-based applications, ensuring that the benefits of such technologies are widely available and do not exacerbate existing inequalities. Collaborative efforts between engineers, ethicists, policymakers, and user communities are essential to navigate these issues thoughtfully. By addressing these ethical and practical challenges head-on, the field can move towards responsible and equitable advancements that harness the full potential of EEG technology for society's benefit.

References

[1] B. U. Forstmann, M. C. Keuken, and A. Alkemade, “An Introduction to Human Brain Anatomy,” in *An Introduction to Model-Based Cognitive Neuroscience*, New York, NY: Springer New York, 2015, pp. 71–89. doi: 10.1007/978-1-4939-2236-9_4.

[2] M. X. Cohen, “Where Does EEG Come From and What Does It Mean?,” *Trends Neurosci*, vol. 40, no. 4, pp. 208–218, Apr. 2017, doi: 10.1016/j.tins.2017.02.004.

[3] E. Pannese, *Neurocytology: fine structure of neurons, nerve processes, and neuroglial cells*. Springer, 2015.

[4] M. Teplan, “FUNDAMENTALS OF EEG MEASUREMENT,” 2002.

[5] R. İnce, S. S. Adanır, and F. Sevmez, “The inventor of electroencephalography (EEG): Hans Berger (1873–1941),” *Child’s Nervous System*, vol. 37, no. 9, pp. 2723–2724, Sep. 2021, doi: 10.1007/s00381-020-04564-z.

[6] J. M. V. BUREN and C. AJMONE-MARSAN, “A Correlation of Autonomic and EEG Components in Temporal Lobe Epilepsy,” *Arch Neurol*, vol. 3, no. 6, pp. 683–703, Dec. 1960, doi: 10.1001/archneur.1960.00450060071009.

[7] M. Abo-Zahhad, S. M. Ahmed, and S. N. Abbas, “A New EEG Acquisition Protocol for Biometric Identification Using Eye Blinking Signals,” *International Journal of Intelligent Systems and Applications*, vol. 7, no. 6, pp. 48–54, May 2015, doi: 10.5815/ijisa.2015.06.05.

[8] Z. Wang, J. Na, and B. Zheng, “An Improved kNN Classifier for Epilepsy Diagnosis,” *IEEE Access*, vol. 8, pp. 100022–100030, 2020, doi: 10.1109/ACCESS.2020.2996946.

[9] A. Miltiadous *et al.*, “Machine Learning Algorithms for Epilepsy Detection Based on Published EEG Databases: A Systematic Review,” *IEEE Access*, vol. 11, pp. 564–594, 2023, doi: 10.1109/ACCESS.2022.3232563.

[10] P. W. Kaplan and A. O. Rossetti, “EEG Patterns and Imaging Correlations in Encephalopathy,” *Journal of Clinical Neurophysiology*, vol. 28, no. 3, pp. 233–251, Jun. 2011, doi: 10.1097/WNP.0b013e31821c33a0.

[11] M. Ahmadlou, H. Adeli, and A. Adeli, “Fractality and a Wavelet-chaos-Methodology for EEG-based Diagnosis of Alzheimer Disease,” *Alzheimer Dis Assoc Disord*, vol. 25, no. 1, pp. 85–92, Jan. 2011, doi: 10.1097/WAD.0b013e3181ed1160.

[12] A. Miltiadous, E. Gionanidis, K. D. Tzimourta, T. Afrantou, N. Giannakeas, and A. T. Tzallas, “DICE-net: A Novel Convolution-Transformer Architecture for Alzheimer Detection in EEG Signals.”

[13] A. Miltiadous *et al.*, “Alzheimer’s disease and frontotemporal dementia: A robust classification method of eeg signals and a comparison of validation methods,” *Diagnostics*, vol. 11, no. 8, 2021, doi: 10.3390/diagnostics11081437.

[14] M. Alessandrini, G. Biagetti, P. Crippa, L. Falaschetti, S. Luzzi, and C. Turchetti, “EEG-Based Alzheimer’s Disease Recognition Using Robust-PCA and LSTM Recurrent Neural Network,” *Sensors*, vol. 22, no. 10, p. 3696, May 2022, doi: 10.3390/s22103696.

[15] V. Aspiotis *et al.*, “Assessing Electroencephalography as a Stress Indicator: A VR High-Altitude Scenario Monitored through EEG and ECG,” *Sensors*, vol. 22, no. 15, p. 5792, Aug. 2022, doi: 10.3390/s22155792.

[16] A. Lenartowicz and S. K. Loo, “Use of EEG to Diagnose ADHD,” *Curr Psychiatry Rep*, vol. 16, no. 11, p. 498, Nov. 2014, doi: 10.1007/s11920-014-0498-0.

[17] A. Nijholt, “BCI for Games: A ‘State of the Art’ Survey,” 2008.

[18] S. J. Reznik and J. J. B. Allen, “Frontal asymmetry as a mediator and moderator of emotion: An updated review,” *Psychophysiology*, vol. 55, no. 1, p. e12965, Jan. 2018, doi: 10.1111/psyp.12965.

[19] M. K. Kim, J. H. Cho, and J. H. Jeong, “Classification of Tactile Perception and Attention on Natural Textures from EEG Signals,” in *9th IEEE International Winter Conference on Brain-Computer Interface, BCI 2021*, 2021. doi: 10.1109/BCI51272.2021.9385296.

[20] Z. Breijyeh and R. Karaman, “Comprehensive Review on Alzheimer’s Disease: Causes and Treatment,” *Molecules*, vol. 25, no. 24, p. 5789, Dec. 2020, doi: 10.3390/molecules25245789.

[21] C. C. Bell, “DSM-IV: Diagnostic and Statistical Manual of Mental Disorders,” *JAMA: The Journal of the American Medical Association*, vol. 272, no. 10, p. 828, Sep. 1994, doi: 10.1001/jama.1994.03520100096046.

[22] B. Dubois *et al.*, “Research criteria for the diagnosis of Alzheimer’s disease: revising the NINCDS–ADRDA criteria,” *Lancet Neurol*, vol. 6, no. 8, pp. 734–746, Aug. 2007, doi: 10.1016/S1474-4422(07)70178-3.

[23] N. van der Vinne, M. A. Vollebregt, M. J. A. M. van Putten, and M. Arns, “Frontal alpha asymmetry as a diagnostic marker in depression: Fact or fiction? A meta-analysis,” *Neuroimage Clin*, vol. 16, pp. 79–87, 2017, doi: 10.1016/j.nicl.2017.07.006.

[24] S. D. You, “Classification of Relaxation and Concentration Mental States with EEG,” *Information*, vol. 12, no. 5, p. 187, Apr. 2021, doi: 10.3390/info12050187.

[25] J. J. Vidal, “Toward direct brain-computer communication,” *Annu Rev Biophys Bioeng*, vol. 2, no. 1, pp. 157–180, 1973.

[26] R. Oshana, “Overview of DSP Algorithms,” in *DSP for Embedded and Real-Time Systems*, Elsevier, 2012, pp. 113–131. doi: 10.1016/B978-0-12-386535-9.00007-X.

[27] R. W. Homan, “The 10-20 Electrode System and Cerebral Location,” *American Journal of EEG Technology*, vol. 28, no. 4, Dec. 1988, doi: 10.1080/00029238.1988.11080272.

[28] F. Westad and M. Kermit, “Independent Component Analysis,” in *Comprehensive Chemometrics*, Elsevier, 2009, pp. 227–248. doi: 10.1016/B978-044452701-1.00045-4.

[29] M. Plechawska-Wójcik, P. Augustynowicz, M. Kaczorowska, E. Zabielska-Mendyk, and D. Zapała, “The Influence Assessment of Artifact Subspace Reconstruction on the EEG Signal Characteristics,” *Applied Sciences*, vol. 13, no. 3, p. 1605, Jan. 2023, doi: 10.3390/app13031605.

[30] E. Oja and Zhijian Yuan, “The FastICA Algorithm Revisited: Convergence Analysis,” *IEEE Trans Neural Netw*, vol. 17, no. 6, pp. 1370–1381, Nov. 2006, doi: 10.1109/TNN.2006.880980.

[31] D. Langlois, S. Chartier, and D. Gosselin, “An introduction to independent component analysis: InfoMax and FastICA algorithms,” *Tutor Quant Methods Psychol*, vol. 6, no. 1, pp. 31–38, 2010.

[32] T.-P. Jung *et al.*, “Extended ICA Removes Artifacts from Electroencephalographic Recordings Understanding the neural mechanisms and temporal dynamics of perceptual learning View project brain computer interface View project Extended ICA Removes Artifacts from Electroencephalographic Recordings,” 1998. [Online]. Available: https://www.researchgate.net/publication/2437771

[33] C. Kothe, “The artifact subspace reconstruction method,” *Accessed: Jul*, vol. 17, p. 2017, 2013.

[34] W. T. Cochran *et al.*, “What is the fast Fourier transform?,” *Proceedings of the IEEE*, vol. 55, no. 10, pp. 1664–1674, 1967, doi: 10.1109/PROC.1967.5957.

[35] P. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967, doi: 10.1109/TAU.1967.1161901.

[36] J. W. H. SWANEPOEL and J. W. J. VAN WYK, “The bootstrap applied to power spectral density function estimation,” *Biometrika*, vol. 73, no. 1, pp. 135–141, 1986, doi: 10.1093/biomet/73.1.135.

[37] N. Kehtarnavaz, “Frequency Domain Processing,” in *Digital Signal Processing System Design*, Elsevier, 2008, pp. 175–196. doi: 10.1016/B978-0-12-374490-6.00007-6.

[38] S. Fouladi, A. A. Safaei, N. Mammone, F. Ghaderi, and M. J. Ebadi, “Efficient Deep Neural Networks for Classification of Alzheimer’s Disease and Mild Cognitive Impairment from Scalp EEG Recordings,” *Cognit Comput*, vol. 14, no. 4, pp. 1247–1268, Jul. 2022, doi: 10.1007/s12559-022-10033-3.

[39] N. E. Huang, *Hilbert-Huang transform and its applications*, vol. 16. World Scientific, 2014.

[40] A. Quinn, V. Lopes-dos-Santos, D. Dupret, A. Nobre, and M. Woolrich, “EMD: Empirical Mode Decomposition and Hilbert-Huang Spectral Analyses in Python,” *J Open Source Softw*, vol. 6, no. 59, p. 2977, Mar. 2021, doi: 10.21105/joss.02977.

[41] K. Baik *et al.*, “Implication of EEG theta/alpha and theta/beta ratio in Alzheimer’s and Lewy body disease,” *Sci Rep*, vol. 12, no. 1, p. 18706, Nov. 2022, doi: 10.1038/s41598-022-21951-5.

[42] Y. Zhang, Y. Jiang, L. Qi, M. Z. A. Bhuiyan, and P. Qian, “Epilepsy Diagnosis Using Multi-view & Multi-medoid Entropy-based Clustering with Privacy Protection,” *ACM Trans Internet Technol*, vol. 21, no. 2, 2021, doi: 10.1145/3404893.

[43] E. Olofsen, J. W. Sleigh, and A. Dahan, “Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect,” *Br J Anaesth*, vol. 101, no. 6, pp. 810–821, Dec. 2008, doi: 10.1093/bja/aen290.

[44] M. Rostaghi and H. Azami, “Dispersion Entropy: A Measure for Time-Series Analysis,” *IEEE Signal Process Lett*, vol. 23, no. 5, pp. 610–614, May 2016, doi: 10.1109/LSP.2016.2542881.

[45] W. Lutzenberger, T. Elbert, N. Birbaumer, W. J. Ray, and H. Schupp, “The scalp distribution of the fractal dimension of the EEG and its variation with mental tasks,” *Brain Topogr*, vol. 5, no. 1, pp. 27–34, 1992, doi: 10.1007/BF01129967.

[46] R. M. Bryce and K. B. Sprague, “Revisiting detrended fluctuation analysis,” *Sci Rep*, vol. 2, no. 1, p. 315, Mar. 2012, doi: 10.1038/srep00315.

[47] T. Donoghue *et al.*, “Parameterizing neural power spectra into periodic and aperiodic components,” *Nat Neurosci*, vol. 23, no. 12, pp. 1655–1665, Dec. 2020, doi: 10.1038/s41593-020-00744-x.

[48] A. Miltiadous *et al.*, “An Ensemble Method for EEG-based Texture Discrimination during Open Eyes Active Touch,” *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 1, pp. 12676–12687, Feb. 2024.

[49] Ke Guolin *et al.*, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” in *Advances in Neural Information Processing Systems 30*, 2017.

[50] T. Chen and T. He, “xgboost: eXtreme Gradient Boosting,” 2023.

[51] A. Ullah, H. Elahi, Z. Sun, A. Khatoon, and I. Ahmad, “Comparative Analysis of AlexNet, ResNet18 and SqueezeNet with Diverse Modification and Arduous Implementation,” *Arab J Sci Eng*, vol. 47, no. 2, pp. 2397–2417, Feb. 2022, doi: 10.1007/s13369-021-06182-6.

[52] A. Vaswani *et al.*, “Attention is All You Need,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, in NIPS’17. Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 6000–6010.

[53] Y. Renard *et al.*, “OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain–Computer Interfaces in Real and Virtual Environments,” *Presence: Teleoperators and Virtual Environments*, vol. 19, no. 1, Feb. 2010, doi: 10.1162/pres.19.1.35.

[54] Y. Isler, A. Narin, and M. Ozer, “Comparison of the Effects of Cross-validation Methods on Determining Performances of Classifiers Used in Diagnosing Congestive Heart Failure,” *Measurement Science Review*, vol. 15, no. 4, Aug. 2015, doi: 10.1515/msr-2015-0027.

[55] M. Hafner, M. Liedlgruber, S. Maimone, A. Uhl, A. Vecsei, and F. Wrba, “Evaluation of cross-validation protocols for the classification of endoscopic images of colonic polyps,” in *2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, Jun. 2012. doi: 10.1109/CBMS.2012.6266355.

[56] R. D. Pascual-Marqui *et al.*, “Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia,” *Psychiatry Res Neuroimaging*, vol. 90, no. 3, Jun. 1999, doi: 10.1016/S0925-4927(99)00013-X.

[57] M. DV, “Relationship between EEG Alpha3/Alpha2 Ratio and the Nucleus Accumbens in Subjects with Mild Cognitive Impairment,” *J Neurol Neurophysiol*, vol. 04, no. 02, 2013, doi: 10.4172/2155-9562.1000149.

[58] M. Lindau, V. Jelic, S.-E. Johansson, C. Andersen, L.-O. Wahlund, and O. Almkvist, “Quantitative EEG Abnormalities and Cognitive Dysfunctions in Frontotemporal Dementia and Alzheimer’s Disease,” *Dement Geriatr Cogn Disord*, vol. 15, no. 2, 2003, doi: 10.1159/000067973.

[59] K. Nishida *et al.*, “Differences in quantitative EEG between frontotemporal dementia and Alzheimer’s disease as revealed by LORETA,” *Clinical Neurophysiology*, vol. 122, no. 9, Sep. 2011, doi: 10.1016/j.clinph.2011.02.011.

[60] F. Caso *et al.*, “Quantitative EEG and LORETA: valuable tools in discerning FTD from AD?,” *Neurobiol Aging*, vol. 33, no. 10, Oct. 2012, doi: 10.1016/j.neurobiolaging.2011.12.011.

[61] M. Dottori *et al.*, “Towards affordable biomarkers of frontotemporal dementia: A classification study via network’s information sharing,” *Sci Rep*, vol. 7, no. 1, Dec. 2017, doi: 10.1038/s41598-017-04204-8.

[62] A. Delorme and S. Makeig, “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” *J Neurosci Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004, doi: 10.1016/j.jneumeth.2003.10.009.

[63] C. Melissant, A. Ypma, E. E. E. Frietman, and C. J. Stam, “A method for detection of Alzheimer’s disease using ICA-enhanced EEG measurements,” *Artif Intell Med*, vol. 33, no. 3, pp. 209–222, Mar. 2005, doi: 10.1016/j.artmed.2004.07.003.

[64] T. Radüntz, J. Scouten, O. Hochmuth, and B. Meffert, “EEG artifact elimination by extraction of ICA-component features using image processing algorithms,” *J Neurosci Methods*, vol. 243, pp. 84–93, Mar. 2015, doi: 10.1016/j.jneumeth.2015.01.030.

[65] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, “ICLabel: An automated electroencephalographic independent component classifier, dataset, and website,” *Neuroimage*, vol. 198, pp. 181–197, Sep. 2019, doi: 10.1016/j.neuroimage.2019.05.026.

[66] A. Gramfort, “MEG and EEG data analysis with MNE-Python,” *Front Neurosci*, vol. 7, 2013, doi: 10.3389/fnins.2013.00267.

[67] K. D. Tzimourta *et al.*, “Machine Learning Algorithms and Statistical Approaches for Alzheimer’s Disease Analysis Based on Resting-State EEG Recordings: A Systematic Review,” *Int J Neural Syst*, vol. 31, no. 05, p. 2130002, May 2021, doi: 10.1142/S0129065721300023.

[68] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, “A comparative analysis of gradient boosting algorithms,” *Artif Intell Rev*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.

[69] L. Kurlowicz and M. Wallace, “The Mini-Mental State Examination (MMSE),” *J Gerontol Nurs*, vol. 25, no. 5, May 1999, doi: 10.3928/0098-9134-19990501-08.

[70] S. B. GUZE, “Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (DSM-IV),” *American Journal of Psychiatry*, vol. 152, no. 8, pp. 1228–1228, Aug. 1995, doi: 10.1176/ajp.152.8.1228.

[71] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, “Clinical diagnosis of Alzheimer’s disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer’s Disease,” *Neurology*, vol. 34, no. 7, pp. 939–939, Jul. 1984, doi: 10.1212/WNL.34.7.939.

[72] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox, “Hyperopt: a Python library for model selection and hyperparameter optimization,” *Comput Sci Discov*, vol. 8, no. 1, p. 014008, Jul. 2015, doi: 10.1088/1749-4699/8/1/014008.

[73] R. Wang, J. Wang, H. Yu, X. Wei, C. Yang, and B. Deng, “Power spectral density and coherence analysis of Alzheimer’s EEG,” *Cogn Neurodyn*, vol. 9, no. 3, pp. 291–304, Jun. 2015, doi: 10.1007/s11571-014-9325-x.

[74] M. Lopes, R. Cassani, and T. H. Falk, “Using CNN Saliency Maps and EEG Modulation Spectra for Improved and More Interpretable Machine Learning-Based Alzheimer’s Disease Diagnosis,” *Comput Intell Neurosci*, vol. 2023, pp. 1–17, Feb. 2023, doi: 10.1155/2023/3198066.

[75] J.-Y. Guo *et al.*, “A Transformer based neural network for emotion recognition and visualizations of crucial EEG channels,” *Physica A: Statistical Mechanics and its Applications*, vol. 603, p. 127700, Oct. 2022, doi: 10.1016/j.physa.2022.127700.

[76] J. Xie *et al.*, “A Transformer-Based Approach Combining Deep Learning Network and Spatial-Temporal Information for Raw EEG Classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2126–2136, 2022, doi: 10.1109/TNSRE.2022.3194600.

[77] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” [Online]. Available: https://github.com/tensorflow/tensor2tensor

[78] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” Oct. 2020, [Online]. Available: http://arxiv.org/abs/2010.11929

[79] Z. Tian, J. Yi, Y. Bai, J. Tao, S. Zhang, and Z. Wen, “Synchronous Transformers for end-to-end Speech Recognition,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May 2020, pp. 7884–7888. doi: 10.1109/ICASSP40776.2020.9054260.

[80] S. Rahardja, M. Wang, B. P. Nguyen, P. Fränti, and S. Rahardja, “A lightweight classification of adaptor proteins using transformer networks,” *BMC Bioinformatics*, vol. 23, no. 1, p. 461, Nov. 2022, doi: 10.1186/s12859-022-05000-6.

[81] D. Kostas, S. Aroca-Ouellette, and F. Rudzicz, “BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data,” *Front Hum Neurosci*, vol. 15, Jun. 2021, doi: 10.3389/fnhum.2021.653659.

[82] Y. Özbek, E. Fide, and G. G. Yener, “Resting-state EEG alpha/theta power ratio discriminates early-onset Alzheimer’s disease from healthy controls,” *Clinical Neurophysiology*, vol. 132, no. 9, pp. 2019–2031, Sep. 2021, doi: 10.1016/j.clinph.2021.05.012.

[83] C. J. Stam, Y. van der Made, Y. A. L. Pijnenburg, and Ph. Scheltens, “EEG synchronization in mild cognitive impairment and Alzheimer’s disease,” *Acta Neurol Scand*, vol. 108, no. 2, pp. 90–96, Aug. 2003, doi: 10.1034/j.1600-0404.2003.02067.x.

[84] J. Bergstra, D. Yamins, and D. Cox, “Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures,” in *Proceedings of the 30th International Conference on Machine Learning*, S. Dasgupta and D. McAllester, Eds., in Proceedings of Machine Learning Research, vol. 28. Atlanta, Georgia, USA: PMLR, Feb. 2013, pp. 115–123. [Online]. Available: https://proceedings.mlr.press/v28/bergstra13.html

[85] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces,” *J Neural Eng*, vol. 15, no. 5, p. 056013, Oct. 2018, doi: 10.1088/1741-2552/aace8c.

[86] N. Waytowich *et al.*, “Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials,” *J Neural Eng*, vol. 15, no. 6, p. 066031, Dec. 2018, doi: 10.1088/1741-2552/aae5d8.

[87] R. T. Schirrmeister *et al.*, “Deep learning with convolutional neural networks for EEG decoding and visualization,” *Hum Brain Mapp*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017, doi: 10.1002/hbm.23730.

[88] A. Paszke *et al.*, “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” in *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[89] M. Hata *et al.*, “Functional connectivity assessed by resting state EEG correlates with cognitive decline of Alzheimer’s disease – An eLORETA study,” *Clinical Neurophysiology*, vol. 127, no. 2, pp. 1269–1278, Feb. 2016, doi: 10.1016/j.clinph.2015.10.030.

[90] K. D. Tzimourta *et al.*, “Analysis of electroencephalographic signals complexity regarding Alzheimer’s Disease,” *Computers & Electrical Engineering*, vol. 76, Jun. 2019, doi: 10.1016/j.compeleceng.2019.03.018.

[91] S. Fouladi, A. A. Safaei, N. Mammone, F. Ghaderi, and M. J. Ebadi, “Efficient Deep Neural Networks for Classification of Alzheimer’s Disease and Mild Cognitive Impairment from Scalp EEG Recordings,” *Cognit Comput*, vol. 14, no. 4, pp. 1247–1268, Jul. 2022, doi: 10.1007/s12559-022-10033-3.

[92] E. Sibilano *et al.*, “An attention-based deep learning approach for the classification of subjective cognitive decline and mild cognitive impairment using resting-state EEG,” *J Neural Eng*, Feb. 2023, doi: 10.1088/1741-2552/acb96e.

[93] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk, “Brain imaging in Alzheimer disease,” *Cold Spring Harb Perspect Med*, vol. 2, no. 4, 2012, doi: 10.1101/cshperspect.a006213.

[94] S. Khatun, B. I. Morshed, and G. M. Bidelman, “A Single-Channel EEG-Based Approach to Detect Mild Cognitive Impairment via Speech-Evoked Brain Responses,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 1063–1070, May 2019, doi: 10.1109/TNSRE.2019.2911970.

[95] M. S. Safi and S. M. M. Safi, “Early detection of Alzheimer’s disease from EEG signals using Hjorth parameters,” *Biomed Signal Process Control*, vol. 65, p. 102338, Mar. 2021, doi: 10.1016/j.bspc.2020.102338.

[96] S. Dogan *et al.*, “Primate brain pattern-based automated Alzheimer’s disease detection model using EEG signals,” *Cogn Neurodyn*, Aug. 2022, doi: 10.1007/s11571-022-09859-2.

[97] S. Ruiz-Gómez *et al.*, “Automated Multiclass Classification of Spontaneous EEG Activity in Alzheimer’s Disease and Mild Cognitive Impairment,” *Entropy*, vol. 20, no. 1, p. 35, Jan. 2018, doi: 10.3390/e20010035.

[98] T. Araújo, J. P. Teixeira, and P. M. Rodrigues, “Smart-Data-Driven System for Alzheimer Disease Detection through Electroencephalographic Signals,” *Bioengineering*, vol. 9, no. 4, p. 141, Mar. 2022, doi: 10.3390/bioengineering9040141.

[99] J. Mongan, L. Moy, and C. E. Kahn, “Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers,” *Radiol Artif Intell*, vol. 2, no. 2, p. e200029, Mar. 2020, doi: 10.1148/ryai.2020200029.

[100] J. D. Head and M. C. Zerner, “A Broyden—Fletcher—Goldfarb—Shanno optimization procedure for molecular geometries,” *Chem Phys Lett*, vol. 122, no. 3, pp. 264–270, Dec. 1985, doi: 10.1016/0009-2614(85)80574-1.

[101] I. E. Lagaris and I. G. Tsoulos, “Stopping rules for box-constrained stochastic global optimization,” *Appl Math Comput*, vol. 197, no. 2, pp. 622–632, Apr. 2008, doi: 10.1016/j.amc.2007.08.001.

[102] I. G. Tsoulos, “Modifications of real code genetic algorithm for global optimization,” *Appl Math Comput*, vol. 203, no. 2, pp. 598–607, Sep. 2008, doi: 10.1016/j.amc.2008.05.005.

[103] D. Moher, A. Liberati, J. Tetzlaff, D. Altman, and T. P. Group, “Preferred Reporting Items for Systematic Reviews and MetaAnalyses: The PRISMA Statement,” *PLoS*, vol. 6, no. 7, p. e1000097, 2009, doi: 10.1371/journal.pmed1000097.

[104] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state,” *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, vol. 64, no. 6, p. 8, 2001, doi: 10.1103/PhysRevE.64.061907.

[105] A. L. Goldberger, L. A. N. Amaral, L. Glass, and E. Al., “PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals,” *[Online]*, vol. 101, no. 23, pp. e215–e220, 2000.

[106] M. Ihle *et al.*, “EPILEPSIAE – A European epilepsy database,” *Comput Methods Programs Biomed*, vol. 106, no. 3, pp. 127–138, 2012, doi: https://doi.org/10.1016/j.cmpb.2010.08.011.

[107] I. Obeid and J. Picone, “The Temple University Hospital EEG data corpus,” *Front Neurosci*, vol. 10, no. MAY, 2016, doi: 10.3389/fnins.2016.00196.

[108] R. Akut, "Wavelet based deep learning approach for epilepsy detection," *Health Inf Sci Syst*, vol. 7, no. 1, p. 8, 2019, doi: 10.1007/s13755-019-0069-1.

[109] F. Alzami *et al.*, "Adaptive Hybrid Feature Selection-Based Classifier Ensemble for Epileptic Seizure Classification," *IEEE Access*, vol. 6, pp. 29132–29145, 2018, doi: 10.1109/ACCESS.2018.2838559.

[110] A. Attia, A. Moussaoui, and Y. Chahir, "Epileptic seizures identification with autoregressive model and firefly optimization based classification," *Evolving Systems*, 2019, doi: 10.1007/s12530-019-09319-z.

[111] M. K. Bandil and A. K. Wadhwani, "Multi-Resolution EEG AND EEG Sub-Band Features Optimization for Epileptic Classification Using Hybrid Evolutionary Computing Technique," *Procedia Comput Sci*, vol. 152, pp. 243–251, 2019, doi: https://doi.org/10.1016/j.procs.2019.05.020.

[112] S. Chen, X. Zhang, L. Chen, and Z. Yang, "Automatic Diagnosis of Epileptic Seizure in Electroencephalography Signals Using Nonlinear Dynamics Features," *IEEE Access*, vol. 7, pp. 61046–61056, 2019, doi: 10.1109/ACCESS.2019.2915610.

[113] H. Chiang, M. Chen, and Y. Huang, "Wavelet-Based EEG Processing for Epilepsy Detection Using Fuzzy Entropy and Associative Petri Net," *IEEE Access*, vol. 7, pp. 103255–103262, 2019, doi: 10.1109/ACCESS.2019.2929266.

[114] V. Gupta and R. B. Pachori, "Epileptic seizure identification using entropy of FBSE based EEG rhythms," *Biomed Signal Process Control*, vol. 53, p. 101569, 2019, doi: https://doi.org/10.1016/j.bspc.2019.101569.

[115] R. Hussein, H. Palangi, R. K. Ward, and Z. J. Wang, "Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals," *Clinical Neurophysiology*, vol. 130, no. 1, pp. 25–37, 2019, doi: https://doi.org/10.1016/j.clinph.2018.10.010.

[116] M. Li, X. Sun, W. Chen, Y. Jiang, and T. Zhang, "Classification epileptic seizures in EEG using time-frequency image and block texture features," *IEEE Access*, vol. 8, pp. 9770–9781, 2020, doi: 10.1109/ACCESS.2019.2960848.

[117] C. Mahjoub, R. le Bouquin Jeannes, T. Lajnef, and A. Kachouri, "Epileptic seizure detection on EEG signals using machine learning techniques and advanced preprocessing methods," *Biomed Tech (Berl)*, 2019, doi: 10.1515/bmt-2019-0001.

[118] X. Zhao, R. Zhang, Z. Mei, C. Chen, and W. Chen, "Identification of epileptic seizures by characterizing instantaneous energy behavior of EEG," *IEEE Access*, vol. 7, pp. 70059–70076, 2019, doi: 10.1109/ACCESS.2019.2919158.

[119] G. Singh, B. Singh, and M. Kaur, "Grasshopper optimization algorithm–based approach for the optimization of ensemble classifier and feature selection to classify epileptic EEG signals," *Med Biol Eng Comput*, vol. 57, no. 6, pp. 1323–1339, 2019, doi: 10.1007/s11517-019-01951-w.

[120] M. G. Tsipouras, "Spectral information of EEG signals with respect to epilepsy classification," *EURASIP J Adv Signal Process*, vol. 2019, no. 1, 2019, doi: 10.1186/s13634-019-0606-8.

[121] X. Wang, G. Gong, and N. Li, "Automated recognition of epileptic EEG states using a combination of symlet wavelet processing, gradient boosting machine, and grid search optimizer," *Sensors (Switzerland)*, vol. 19, no. 2, 2019, doi: 10.3390/s19020219.

[122] L. Yang, S. Ding, H.-M. Zhou, and X. Yang, "A strategy combining intrinsic time-scale decomposition and a feedforward neural network for automatic seizure detection," *Physiol Meas*, vol. 40, no. 9, 2019, doi: 10.1088/1361-6579/ab3e2e.

[123] R. Bose, S. Pratiher, and S. Chatterjee, "Detection of epileptic seizure employing a novel set of features extracted from multifractal spectrum of electroencephalogram signals," *IET Signal Processing*, vol. 13, no. 2, pp. 157–164, 2019, doi: 10.1049/iet-spr.2018.5258.

[124] H. Choubey and A. Pandey, "A new feature extraction and classification mechanisms For EEG signal processing," *Multidimens Syst Signal Process*, vol. 30, no. 4, pp. 1793–1809, 2019, doi: 10.1007/s11045-018-0628-7.

[125] E. Kabir, Siuly, J. Cao, and H. Wang, "A computer aided analysis scheme for detecting epileptic seizure from EEG data," *International Journal of Computational Intelligence Systems*, vol. 11, no. 1, pp. 663–671, 2018, doi: 10.2991/ijcis.11.1.51.

[126] S. Lahmiri and A. Shmuel, "Accurate Classification of Seizure and Seizure-Free Intervals of Intracranial EEG Signals from Epileptic Patients," *IEEE Trans Instrum Meas*, vol. 68, no. 3, pp. 791–796, 2019, doi: 10.1109/TIM.2018.2855518.

[127] P. Li, C. Karmakar, J. Yearwood, S. Venkatesh, M. Palaniswami, and C. Liu, "Detection of epileptic seizure based on entropy analysis of short-term EEG," *PLoS One*, vol. 13, no. 3, 2018, doi: 10.1371/journal.pone.0193691.

[128] A. Mert and A. Akan, "Seizure onset detection based on frequency domain metric of empirical mode decomposition," *Signal Image Video Process*, vol. 12, no. 8, pp. 1489–1496, 2018, doi: 10.1007/s11760-018-1304-y.

[129] J. Saini and M. Dutta, "Epilepsy classification using optimized artificial neural network," *Neurol Res*, vol. 40, no. 11, pp. 982–994, 2018, doi: 10.1080/01616412.2018.1508544.

[130] A. I. Sharaf, M. A. El-Soud, and I. M. El-Henawy, "An automated approach for epilepsy detection based on tunable Q - Wavelet and firefly feature selection algorithm," *Int J Biomed Imaging*, vol. 2018, 2018, doi: 10.1155/2018/5812872.

[131] R. R. Sharma, P. Varshney, R. B. Pachori, and S. K. Vishvakarma, "Automated System for Epileptic EEG Detection Using Iterative Filtering," *IEEE Sens Lett*, vol. 2, no. 4, pp. 1–4, 2018, doi: 10.1109/LSENS.2018.2882622.

[132] A. Sharmila and P. Geethanjali, "Effect of filtering with time domain features for the detection of epileptic seizure from EEG signals," *J Med Eng Technol*, vol. 42, no. 3, pp. 217–227, 2018, doi: 10.1080/03091902.2018.1464075.

[133] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput Biol Med*, vol. 100, pp. 270–278, 2018, doi: 10.1016/j.compbiomed.2017.09.017.

[134] P. Amorim, T. Moraes, D. Fazanaro, J. Silva, and H. Pedrini, "Electroencephalogram signal classification based on shearlet and contourlet transforms," *Expert Syst Appl*, vol. 67, pp. 140–147, 2017, doi: https://doi.org/10.1016/j.eswa.2016.09.037.

[135] K. S. Biju, H. A. Hakkim, M. G. Jibukumar, B. K.S., H. A. Hakkim, and J. M.G., "Ictal EEG classification based on amplitude and frequency contours of IMFs," *Biocybern Biomed Eng*, vol. 37, no. 1, pp. 172–183, 2017.

[136] S. Chatterjee, N. Ray Choudhury, and R. Bose, "Detection of epileptic seizure and seizure-free EEG signals employing generalised S-transform," *IET Science, Measurement and Technology*, vol. 11, no. 7, pp. 847–855, 2017.

[137] A. K. Jaiswal and H. Banka, "Epileptic seizure detection in EEG signal with GModPCA and support vector machine," *Biomed Mater Eng*, vol. 28, no. 2, pp. 141–157, 2017, doi: 10.3233/BME-171663.

[138] A. K. Jaiswal and H. Banka, "Epileptic seizure detection in EEG signal using machine learning techniques," *Australas Phys Eng Sci Med*, vol. 41, no. 1, pp. 81–94, 2018, doi: 10.1007/s13246-017-0610-y.

[139] Q. Liu, X. Zhao, Z. Hou, and H. Liu, "Epileptic seizure detection based on the kernel extreme learning machine," vol. 25, no. S1, pp. S399--S409, 2017, doi: 10.3233/THC-171343.

[140] Y. Li, X.-D. Wang, M.-L. Luo, K. Li, X.-F. Yang, and Q. Guo, "Epileptic Seizure Classification of EEGs Using Time-Frequency Analysis Based Multiscale Radial Basis Functions," *IEEE J Biomed Health Inform*, vol. 22, no. 2, pp. 386–397, 2018, doi: 10.1109/JBHI.2017.2654479.

[141] Z. Mohammadpoory, M. Nasrolahzadeh, and J. Haddadnia, "Epileptic seizure detection in EEGs signals based on the weighted visibility graph entropy," *Seizure*, vol. 50, pp. 202–208, 2017, doi: https://doi.org/10.1016/j.seizure.2017.07.001.

[142] M. Sharma and R. B. Pachori, "A novel approach to detect epileptic seizures using a combination of tunable-Q wavelet transform and fractal dimension," *J Mech Med Biol*, vol. 17, no. 7, 2017, doi: 10.1142/S0219519417400036.

[143] A. Sharmila, S. Aman Raj, P. Shashank, and P. Mahalakshmi, "Epileptic seizure detection using DWT-based approximate entropy, Shannon entropy and support vector machine: a case study," *J Med Eng Technol*, vol. 42, no. 1, pp. 1–8, 2018, doi: 10.1080/03091902.2017.1394389.

[144] A. Sharmila and P. Mahalakshmi, "Wavelet-based feature extraction for classification of epileptic seizure EEG signal," *J Med Eng Technol*, vol. 41, no. 8, pp. 670–680, 2017, doi: 10.1080/03091902.2017.1394388.

[145] H. Al-Hadeethi, S. Abdulla, M. Diykh, R. C. Deo, and J. H. Green, "Adaptive boost LS-SVM classification approach for time-series signal classification in epileptic seizure diagnosis applications," *Expert Syst Appl*, vol. 161, 2020, doi: 10.1016/j.eswa.2020.113676.

[146] I. Aliyu and C. G. Lim, "Selection of optimal wavelet features for epileptic EEG signal classification with LSTM," *Neural Comput Appl*, 2021, doi: 10.1007/s00521-020-05666-0.

[147] H. U. Amin, M. Z. Yusoff, and R. F. Ahmad, "A novel approach based on wavelet analysis and arithmetic coding for automated detection and diagnosis of epileptic seizure in EEG signals using machine learning techniques," *Biomed Signal Process Control*, vol. 56, 2020, doi: 10.1016/j.bspc.2019.101707.

[148] A. Anuragi, D. Singh Sisodia, and R. B. Pachori, "Epileptic-seizure classification using phase-space representation of FBSE-EWT based EEG sub-band signals and ensemble learners," *Biomed Signal Process Control*, vol. 71, no. PA, p. 103138, 2022, doi: 10.1016/j.bspc.2021.103138.

[149] A. Arı, "Analysis of EEG signal for seizure detection based on WPT," *Electron Lett*, vol. 56, no. 25, pp. 1381–1383, 2020, doi: 10.1049/el.2020.2701.

[150] S. R. Ashokkumar, S. Anupallavi, M. Premkumar, and V. Jeevanantham, "Implementation of deep neural networks for classifying electroencephalogram signal using fractional S-transform for epileptic seizure detection," *Int J Imaging Syst Technol*, vol. 31, no. 2, pp. 895–908, 2021, doi: 10.1002/ima.22565.

[151] S. R. Ashokkumar, G. MohanBabu, and S. Anupallavi, “A novel two-band equilateral wavelet filter bank method for an automated detection of seizure from EEG signals,” *Int J Imaging Syst Technol*, vol. 30, no. 4, pp. 978–993, 2020, doi: 10.1002/ima.22441.

[152] S. R. Ashokkumar and G. Mohanbabu, “Extreme learning adaptive neuro-fuzzy inference system model for classifying the epilepsy using Q-Tuned wavelet transform,” *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 1, pp. 233–248, 2020, doi: 10.3233/JIFS-191015.

[153] M. F. Bari and S. Anowarul Fattah, “Epileptic seizure detection in EEG signals using normalized IMFs in CEEMDAN domain and quadratic discriminant classifier,” *Biomed Signal Process Control*, vol. 58, p. 101833, 2020, doi: 10.1016/j.bspc.2019.101833.

[154] M. Baykara and A. Abdulrahman, “Seizure detection based on adaptive feature extraction by applying extreme learning machines,” *Traitement du Signal*, vol. 38, no. 2, pp. 331–340, 2021, doi: 10.18280/TS.380210.

[155] Z. Brari and S. Belghith, “A Novel Machine Learning Model for the Detection of Epilepsy and Epileptic Seizures Using Electroencephalographic Signals Based on Chaos and Fractal Theories,” *Math Probl Eng*, vol. 2021, 2021, doi: 10.1155/2021/2107113.

[156] J. A. De La O Serna, M. R. A. Paternina, A. Zamora-Mendez, R. K. Tripathy, and R. B. Pachori, “EEG-Rhythm Specific Taylor-Fourier Filter Bank Implemented with O-Splines for the Detection of Epilepsy Using EEG Signals,” *IEEE Sens J*, vol. 20, no. 12, pp. 6542–6551, 2020, doi: 10.1109/JSEN.2020.2976519.

[157] A. S. Eltrass, M. B. Tayel, and A. F. EL-qady, “Automatic epileptic seizure detection approach based on multi-stage Quantized Kernel Least Mean Square filters,” *Biomed Signal Process Control*, vol. 70, no. July, p. 103031, 2021, doi: 10.1016/j.bspc.2021.103031.

[158] X. Gao, X. Yan, P. Gao, X. Gao, and S. Zhang, “Automatic detection of epileptic seizure based on approximate entropy, recurrence quantification analysis and convolutional neural networks,” *Artif Intell Med*, vol. 102, p. 101711, 2020, doi: 10.1016/j.artmed.2019.101711.

[159] A. Goshvarpour and A. Goshvarpour, “Diagnosis of epileptic EEG using a lagged Poincare plot in combination with the autocorrelation,” *Signal Image Video Process*, vol. 14, no. 7, pp. 1309–1317, 2020, doi: 10.1007/s11760-020-01672-w.

[160] X. Gu, C. Zhang, and T. Ni, “A Hierarchical Discriminative Sparse Representation Classifier for EEG Signal Detection,” *IEEE/ACM Trans Comput Biol Bioinform*, vol. 18, no. 5, pp. 1679–1687, 2021, doi: 10.1109/TCBB.2020.3006699.

[161] A. R. Hassan, A. Subasi, and Y. Zhang, “Epilepsy seizure detection using complete ensemble empirical mode decomposition with adaptive noise,” *Knowl Based Syst*, vol. 191, no. xxxx, p. 105333, 2020, doi: 10.1016/j.knosys.2019.105333.

[162] G. C. Jana, M. S. Praneeth, and A. Agrawal, “A Multi-View SVM Approach for Seizure Detection from Single Channel EEG Signals,” *IETE J Res*, 2021, doi: 10.1080/03772063.2021.1913074.

[163] S.-W. Jang and S.-H. Lee, “Detection of epileptic seizures using wavelet transform, peak extraction and PSR from EEG signals,” *Symmetry (Basel)*, vol. 12, no. 8, 2020, doi: 10.3390/SYM12081239.

[164] Y. Jiang, W. Chen, and Y. You, “Scattering transform-based features for the automatic seizure detection,” *Biocybern Biomed Eng*, vol. 40, no. 1, pp. 77–89, 2020, doi: 10.1016/j.bbe.2019.11.002.

[165] M. Lee, J. Ryu, and D. H. Kim, “Automated epileptic seizure waveform detection method based on the feature of the mean slope of wavelet coefficient counts using a hidden Markov model and EEG signals,” *ETRI Journal*, vol. 42, no. 2, pp. 217–229, 2020, doi: 10.4218/etrij.2018-0118.

[166] Y. Li, P. Qian, S. Wang, and S. Wang, “Novel multi-view Takagi–Sugeno–Kang fuzzy system for epilepsy EEG detection,” *J Ambient Intell Humaniz Comput*, 2021, doi: 10.1007/s12652-021-03189-7.

[167] J. Lian, Y. Zhang, R. Luo, G. Han, W. Jia, and C. Li, “Pair-Wise Matching of EEG Signals for Epileptic Identification via Convolutional Neural Network,” *IEEE Access*, vol. 8, pp. 40008–40017, 2020, doi: 10.1109/ACCESS.2020.2976751.

[168] M. Ma, Y. Cheng, X. Wei, Z. Chen, and Y. Zhou, “Research on epileptic EEG recognition based on improved residual networks of 1-D CNN and indRNN,” *BMC Med Inform Decis Mak*, vol. 21, 2021, doi: 10.1186/s12911-021-01438-5.

[169] B. Mandhouj, M. A. Cherni, and M. Sayadi, “An automated classification of EEG signals based on spectrogram and CNN for epilepsy diagnosis,” *Analog Integr Circuits Signal Process*, vol. 108, no. 1, pp. 101–110, 2021, doi: 10.1007/s10470-021-01805-2.

[170] W. Mardini, M. M. Bani Yassein, R. Al-Rawashdeh, S. Aljawarneh, Y. Khamayseh, and O. Meqdadi, “Enhanced detection of epileptic seizure using EEG signals in combination with machine learning classifiers,” *IEEE Access*, vol. 8, pp. 24046–24055, 2020, doi: 10.1109/ACCESS.2020.2970012.

[171] P. Mathur, V. K. Chakka, and S. B. Shah, “Ramanujan periodic subspace based epileptic EEG signals classification,” *IEEE Sens Lett*, vol. 5, no. 7, pp. 7–10, 2021, doi: 10.1109/LSENS.2021.3086755.

[172] M. K. I. Molla, K. M. Hassan, M. R. Islam, and T. Tanaka, “Graph eigen decomposition-based feature-selection method for epileptic seizure detection using electroencephalography,” *Sensors (Switzerland)*, vol. 20, no. 16, pp. 1–23, 2020, doi: 10.3390/s20164639.

[173] J. Na, Z. Wang, S. Lv, and Z. Xu, “An Extended K Nearest Neighbors-Based Classifier for Epilepsy Diagnosis,” *IEEE Access*, vol. 9, pp. 73910–73923, 2021, doi: 10.1109/ACCESS.2021.3081767.

[174] H. S. Nogay and H. Adeli, “Detection of Epileptic Seizure Using Pretrained Deep Convolutional Neural Network and Transfer Learning,” *Eur Neurol*, vol. 83, no. 6, pp. 602–614, 2021, doi: 10.1159/000512985.

[175] C. Oinam, R. Laishram, M. Romesh Singh, and P. T. Gangmei, “EEG signal classification using ANN trained with hybrid PSO and GSA,” *International Journal of Scientific and Technology Research*, vol. 9, no. 1, pp. 3589–3595, 2020, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85078738361&partnerID=40&md5=a6a76c1a6f1950b39d3bccfb94e65b02

[176] M. Omidvar, A. Zahedi, and H. Bakhshi, “EEG signal processing for epilepsy seizure detection using 5-level Db4 discrete wavelet transform, GA-based feature selection and ANN/SVM classifiers,” *J Ambient Intell Humaniz Comput*, vol. 12, no. 11, pp. 10395–10403, 2021, doi: 10.1007/s12652-020-02837-8.

[177] H. Pal and A. Kumar, “Stability Analysis of Multiscale Bubble Entropy and Power Metric based Seizure Detection Technique with MLA,” *IETE J Res*, 2021, doi: 10.1080/03772063.2021.1912650.

[178] K. Polat and M. Nour, “Epileptic Seizure Detection Based on New Hybrid Models with Electroencephalogram Signals,” *IRBM*, vol. 41, no. 6, pp. 331–353, 2020, doi: 10.1016/j.irbm.2020.06.008.

[179] M. Radman, M. Moradi, A. Chaibakhsh, M. Kordestani, and M. Saif, “Multi-Feature Fusion Approach for Epileptic Seizure Detection from EEG Signals,” *IEEE Sens J*, vol. 21, no. 3, pp. 3533–3543, 2021, doi: 10.1109/JSEN.2020.3026032.

[180] K. S. Shekokar and S. Dour, “Automatic epileptic seizure detection using LSTM networks,” *World Journal of Engineering*, vol. 19, no. 2, pp. 224–229, 2022, doi: 10.1108/WJE-06-2021-0348.

[181] G. Singh, M. Kaur, and B. Singh, “Detection of Epileptic Seizure EEG Signal Using Multiscale Entropies and Complete Ensemble Empirical Mode Decomposition,” *Wirel Pers Commun*, vol. 116, no. 1, pp. 845–864, 2021, doi: 10.1007/s11277-020-07742-z.

[182] K. Sujatha, “Automatic epilepsy detection using hybrid decomposition with multi class support vector method,” *Multimed Tools Appl*, vol. 79, no. 15–16, pp. 9871–9890, 2020, doi: 10.1007/s11042-019-08359-6.

[183] Sukriti, M. Chakraborty, and D. Mitra, “Epilepsy seizure detection using kurtosis based VMD’s parameters selection and bandwidth features,” *Biomed Signal Process Control*, vol. 64, no. October 2020, p. 102255, 2021, doi: 10.1016/j.bspc.2020.102255.

[184] M. Woodbright, B. Verma, and A. Haidar, “Autonomous deep feature extraction based method for epileptic EEG brain seizure classification,” *Neurocomputing*, vol. 444, pp. 30–37, 2021, doi: 10.1016/j.neucom.2021.02.052.

[185] A. Yildiz, H. Zan, and S. Said, “Classification and analysis of epileptic EEG recordings using convolutional neural network and class activation mapping,” *Biomed Signal Process Control*, vol. 68, no. January, p. 102720, 2021, doi: 10.1016/j.bspc.2021.102720.

[186] M. Zeng, X. Zhang, C. Zhao, X. Lu, and Q. Meng, “GRP-DNet: A gray recurrence plot-based densely connected convolutional network for classification of epileptiform EEG,” *J Neurosci Methods*, vol. 347, 2021, doi: 10.1016/j.jneumeth.2020.108953.

[187] W. Zeng, M. Li, C. Yuan, Q. Wang, F. Liu, and Y. Wang, “Identification of epileptic seizures in EEG signals using time-scale decomposition (ITD), discrete wavelet transform (DWT), phase space reconstruction (PSR) and neural networks,” *Artif Intell Rev*, vol. 53, no. 4, pp. 3059–3088, 2020, doi: 10.1007/s10462-019-09755-y.

[188] G. Zhang *et al.*, “MNL-Network: A Multi-Scale Non-local Network for Epilepsy Detection From EEG Signals,” *Front Neurosci*, vol. 14, 2020, doi: 10.3389/fnins.2020.00870.

[189] T. Zhang, Z. Han, X. Chen, and W. Chen, “Subbands and cumulative sum of subbands based nonlinear features enhance the performance of epileptic seizure detection,” *Biomed Signal Process Control*, vol. 69, no. 5988, p. 102827, 2021, doi: 10.1016/j.bspc.2021.102827.

[190] W. Zhao *et al.*, “A Novel Deep Neural Network for Robust Detection of Seizures Using EEG Signals,” *Comput Math Methods Med*, vol. 2020, 2020, doi: 10.1155/2020/9689821.

[191] D. Zhou and X. Li, “Epilepsy EEG Signal Classification Algorithm Based on Improved RBF,” *Front Neurosci*, vol. 14, 2020, doi: 10.3389/fnins.2020.00606.

[192] J. Zhou, X. Zhang, and Z. Jiang, “Recognition of Imbalanced Epileptic EEG Signals by a Graph-Based Extreme Learning Machine,” *Wirel Commun Mob Comput*, vol. 2021, 2021, doi: 10.1155/2021/5871684.

[193] M. Zubair *et al.*, “Detection of Epileptic Seizures from EEG Signals by Combining Dimensionality Reduction Algorithms with Machine Learning Models,” *IEEE Sens J*, vol. 21, no. 15, pp. 16861–16869, 2021, doi: 10.1109/JSEN.2021.3077578.

[194] S. Dehuri and N. Singh, “Epilepsy detection from electroencephalogram signal using singular value decomposition and extreme learning machine classifier,” *Int J Biomed Eng Technol*, vol. 39, no. 1, p. 22, 2022, doi: 10.1504/IJBET.2022.10047885.

[195] J. Cao, J. Zhu, W. Hu, and A. Kummert, “Epileptic Signal Classification with Deep EEG Features by Stacked CNNs,” *IEEE Trans Cogn Dev Syst*, p. 1, 2019, doi: 10.1109/TCDS.2019.2936441.

[196] V. Harpale and V. Bairagi, “An adaptive method for feature selection and extraction for classification of epileptic EEG signal in significant states,” *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 6, pp. 668–676, 2018, doi: 10.1016/j.jksuci.2018.04.014.

[197] A. Mansouri, S. P. Singh, and K. Sayood, “Online EEG seizure detection and localization,” *Algorithms*, vol. 12, no. 9, 2019, doi: 10.3390/a12090176.

[198] X. Tian *et al.*, “Deep Multi-View Feature Learning for EEG-Based Epileptic Seizure Detection,” *IEEE Trans Neural Syst Rehabil Eng*, vol. 27, no. 10, pp. 1962–1972, 2019, doi: 10.1109/TNSRE.2019.2940485.

[199] J. Birjandtalab, M. Baran Pouyan, D. Cogan, M. Nourani, and J. Harvey, “Automated seizure detection using limited-channel EEG and non-linear dimension reduction,” *Comput Biol Med*, vol. 82, pp. 49–58, 2017, doi: https://doi.org/10.1016/j.compbiomed.2017.01.011.

[200] S. W. Ibrahim and S. Majzoub, “EEG-based epileptic seizures detection with adaptive learning capability,” *International Journal on Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 813–824, 2017, doi: 10.15676/ijeei.2017.9.4.13.

[201] R. H. Gabr, A. I. Shahin, A. A. Sharawi, and M. A. Aouf, “A deep learning identification system for different epileptic seizure disease stages,” *Journal of Engineering and Applied Science*, vol. 67, no. 4, pp. 925–944, 2020, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85089162485&partnerID=40&md5=b7d7981631a9945edcfe9c1c28e851e6

[202] R. Mouleeshuwarapprabu and N. Kasthuri, “Nonlinear vector decomposed neural network based EEG signal feature extraction and detection of seizure,” *Microprocess Microsyst*, vol. 76, p. 103075, 2020, doi: 10.1016/j.micpro.2020.103075.

[203] J. Zhang, Z. Wei, J. Zou, and H. Fu, “Automatic epileptic EEG classification based on differential entropy and attention model,” *Eng Appl Artif Intell*, vol. 96, no. September, p. 103975, 2020, doi: 10.1016/j.engappai.2020.103975.

[204] K. A. Khan, S. P. P., Y. U. Khan, and O. Farooq, “A hybrid Local Binary Pattern and wavelets based approach for EEG classification for diagnosing epilepsy,” *Expert Syst Appl*, vol. 140, 2020, doi: 10.1016/j.eswa.2019.112895.

[205] B. Akbarian and A. Erfanian, “A framework for seizure detection using effective connectivity, graph theory, and multi-level modular network,” *Biomed Signal Process Control*, vol. 59, p. 101878, 2020, doi: 10.1016/j.bspc.2020.101878.

[206] J. Zeng, X. dan Tan, and C. A. Zhan, “Automatic detection of epileptic seizure events using the time-frequency features and machine learning,” *Biomed Signal Process Control*, vol. 69, no. July, p. 102916, 2021, doi: 10.1016/j.bspc.2021.102916.

[207] Y. Zhao *et al.*, “EEG-Based Seizure detection using linear graph convolution network with focal loss,” *Comput Methods Programs Biomed*, vol. 208, p. 106277, 2021, doi: 10.1016/j.cmpb.2021.106277.

[208] S. Nasiri and G. D. Clifford, “Generalizable seizure detection model using generating transferable adversarial features,” *IEEE Signal Process Lett*, vol. 28, pp. 568–572, 2021, doi: 10.1109/LSP.2021.3060967.

[209] B. Zhang *et al.*, “Cross-Subject Seizure Detection in EEGs Using Deep Transfer Learning,” *Comput Math Methods Med*, vol. 2020, 2020, doi: 10.1155/2020/7902072.

[210] X. Hu, S. Yuan, F. Xu, Y. Leng, K. Yuan, and Q. Yuan, “Scalp EEG classification using deep Bi-LSTM network for seizure detection,” *Comput Biol Med*, vol. 124, 2020, doi: 10.1016/j.compbiomed.2020.103919.

[211] I. B. Slimen, L. Boubchir, Z. Mbarki, and H. Seddik, “EEG epileptic seizure detection and classification based on dual-tree complex wavelet transform and machine learning algorithms,” *J Biomed Res*, vol. 34, no. 3, pp. 151–161, 2020, doi: 10.7555/JBR.34.20190026.

[212] A. Quintero-Rincón, C. D’Giano, and H. Batatia, “A quadratic linear-parabolic model-based EEG classification to detect epileptic seizures,” *J Biomed Res*, vol. 34, no. 3, pp. 205–212, 2020, doi: 10.7555/JBR.33.20190012.

[213] M. K. Siddiqui, X. Huang, R. Morales-Menendez, N. Hussain, and K. Khatoon, "Machine learning based novel cost-sensitive seizure detection classifier for imbalanced EEG data sets," *International Journal on Interactive Design and Manufacturing*, vol. 14, no. 4, pp. 1491–1509, 2020, doi: 10.1007/s12008-020-00715-3.

[214] V. Bhandari and Manjaiah D.H, "Improved ensemble learning model with optimal feature selection for automated epileptic seizure detection," *Comput Methods Biomech Biomed Eng Imaging Vis*, pp. 1–31, Apr. 2022, doi: 10.1080/21681163.2022.2058616.

[215] K. Zeng, G. Ouyang, H. Chen, Y. Gu, X. Liu, and X. Li, "Characterizing dynamics of absence seizure EEG with spatial-temporal permutation entropy," *Neurocomputing*, vol. 275, pp. 577–585, 2018, doi: https://doi.org/10.1016/j.neucom.2017.09.007.

[216] Z. Gao *et al.*, "Automatic change detection for real-time monitoring of EEG signals," *Front Physiol*, vol. 9, no. APR, 2018, doi: 10.3389/fphys.2018.00325.

[217] Y. Zhang, S. Yang, Y. Liu, Y. Zhang, B. Han, and F. Zhou, "Integration of 24 feature types to accurately detect and predict seizures using scalp EEG signals," *Sensors (Switzerland)*, vol. 18, no. 5, 2018, doi: 10.3390/s18051372.

[218] N. Sriraam *et al.*, "Automated epileptic seizures detection using multi-features and multilayer perceptron neural network," *Brain Inform*, vol. 5, no. 2, 2018, doi: 10.1186/s40708-018-0088-8.

[219] X. Ma, N. Yu, and W. Zhou, "Using Dictionary Pair Learning for Seizure Detection," *Int J Neural Syst*, vol. 29, no. 4, 2019, doi: 10.1142/S0129065718500053.

[220] N. Mahmoodian, A. Boese, M. Friebe, and J. Haddadnia, "Epileptic seizure detection using cross-bispectrum of electroencephalogram signal," *Seizure*, vol. 66, pp. 4–11, 2019, doi: https://doi.org/10.1016/j.seizure.2019.02.001.

[221] Alhussein Musaed, Muhammad Ghulam, and Hossain Shamim M, "EEG Pathology Detection based on Deep Learning," *Test Engineering and Management*, vol. 81, no. 11–12, pp. 5587–5591, 2019.

[222] Z. Zhang *et al.*, "DWT-net: Seizure detection system with structured EEG montage and multiple feature extractor in convolution neural network," *J Sens*, vol. 2020, 2020, doi: 10.1155/2020/3083910.

[223] M. Sharma, S. Patel, and U. R. Acharya, "Automated detection of abnormal EEG signals using localized wavelet filter banks," *Pattern Recognit Lett*, vol. 133, pp. 188–194, 2020, doi: 10.1016/j.patrec.2020.03.009.

[224] T. Iešmantas and R. Alzbutas, "Convolutional neural network for detection and classification of seizures in clinical data," *Med Biol Eng Comput*, vol. 58, no. 9, pp. 1919–1932, 2020, doi: 10.1007/s11517-020-02208-7.

[225] Q. Zhan and W. Hu, "An Epilepsy Detection Method Using Multiview Clustering Algorithm and Deep Features," *Comput Math Methods Med*, vol. 2020, 2020, doi: 10.1155/2020/5128729.

[226] T. Tuncer, S. Dogan, and U. Rajendra Acharya, "Automated EEG signal classification using chaotic local binary pattern," *Expert Syst Appl*, vol. 182, 2021, doi: 10.1016/j.eswa.2021.115175.

[227] D. Priyasad, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Interpretable Seizure Classification Using Unprocessed EEG with Multi-Channel Attentive Feature Fusion," *IEEE Sens J*, vol. 21, no. 17, pp. 19186–19197, 2021, doi: 10.1109/JSEN.2021.3090062.

[228] X. J. Lu, J. Q. Zhang, S. F. Huang, J. Lu, M. Q. Ye, and M. S. Wang, "Detection and classification of epileptic EEG signals by the methods of nonlinear dynamics," *Chaos Solitons Fractals*, vol. 151, p. 111032, 2021, doi: 10.1016/j.chaos.2021.111032.

[229] J. Mu *et al.*, "Automatic detection for epileptic seizure using graph-regularized nonnegative matrix factorization and Bayesian linear discriminate analysis," *Biocybern Biomed Eng*, vol. 41, no. 4, pp. 1258–1271, 2021, doi: 10.1016/j.bbe.2021.08.009.

[230] S. Hadiyoso, I. Wijayanto, and A. Humairani, "Signal dynamics analysis for epileptic seizure classification on EEG signals," *Traitement du Signal*, vol. 38, no. 1, pp. 73–78, 2021, doi: 10.18280/TS.380107.

[231] D. Chen, S. Wan, J. Xiang, and F. S. Bao, "A high-performance seizure detection algorithm based on Discrete Wavelet Transform (DWT) and EEG," *PLoS One*, vol. 12, no. 3, 2017, doi: 10.1371/journal.pone.0173138.

[232] M. S. J. Solaija, S. Saleem, K. Khurshid, S. A. Hassan, and A. M. Kamboh, "Dynamic mode decomposition based epileptic seizure detection from scalp EEG," *IEEE Access*, vol. 6, pp. 38683–38692, 2018, doi: 10.1109/ACCESS.2018.2853125.

[233] S. Raghu, N. Sriraam, G. P. Kumar, and A. S. Hegde, "A novel approach for real-time recognition of epileptic seizures using minimum variance modified fuzzy entropy," *IEEE Trans Biomed Eng*, vol. 65, no. 11, pp. 2612–2621, 2018, doi: 10.1109/TBME.2018.2810942.

[234] M. Ravi Kumar and Y. Srinivasa Rao, “Epileptic seizures classification in EEG signal based on semantic features and variational mode decomposition,” *Cluster Comput*, vol. 22, pp. 13521–13531, 2019, doi: 10.1007/s10586-018-1995-4.

[235] Z. Jiang, F.-L. Chung, and S. Wang, “Recognition of Multiclass Epileptic EEG Signals Based on Knowledge and Label Space Inductive Transfer,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 630–642, 2019, doi: 10.1109/TNSRE.2019.2904708.

[236] S. Raghu, N. Sriraam, Y. Temel, S. V. Rao, A. S. Hegde, and P. L. Kubben, “Performance evaluation of DWT based sigmoid entropy in time and frequency domains for automated detection of epileptic seizures using SVM classifier,” *Comput Biol Med*, vol. 110, pp. 127–143, 2019, doi: https://doi.org/10.1016/j.compbiomed.2019.05.016.

[237] H. R. al Ghayab, Y. Li, S. Siuly, and S. Abdulla, “Epileptic seizures detection in EEGs blending frequency domain with information gain technique,” *Soft comput*, vol. 23, no. 1, pp. 227–239, 2019, doi: 10.1007/s00500-018-3487-0.

[238] M. Bilal, M. Rizwan, S. Saleem, M. M. Khan, M. S. Alkatheir, and M. Alqarni, “Automatic Seizure Detection Using Multi-Resolution Dynamic Mode Decomposition,” *IEEE Access*, vol. 7, pp. 61180–61194, 2019, doi: 10.1109/ACCESS.2019.2915609.

[239] A. Pandey, R. Sequeria, P. Kumar, and S. Kumar, “A Multistage Deep Residual Network for Biomedical Cyber-Physical Systems,” *IEEE Syst J*, vol. 14, no. 2, pp. 1953–1962, 2020, doi: 10.1109/JSYST.2019.2923670.

[240] D. Wu *et al.*, “Automatic epileptic seizures joint detection algorithm based on improved multi-domain feature of cEEG and spike feature of aEEG,” *IEEE Access*, vol. 7, pp. 41551–41564, 2019, doi: 10.1109/ACCESS.2019.2904949.

[241] N. D. Truong, L. Kuhlmann, M. R. Bonyadi, D. Querlioz, L. Zhou, and O. Kavehei, “Epileptic Seizure Forecasting With Generative Adversarial Networks,” *IEEE Access*, vol. 7, pp. 143999–144009, 2019, doi: 10.1109/ACCESS.2019.2944691.

[242] C. Gómez, P. Arbeláez, M. Navarrete, C. Alvarado-Rojas, M. le Van Quyen, and M. Valderrama, “Automatic seizure detection based on imaged-EEG signals through fully convolutional networks,” *Sci Rep*, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-78784-3.

[243] M. Sameer and B. Gupta, “Detection of epileptical seizures based on alpha band statistical features,” *Wirel Pers Commun*, vol. 115, no. 2, pp. 909–925, 2020, doi: 10.1007/s11277-020-07542-5.

[244] J. Lian, Y. Shi, Y. Zhang, W. Jia, X. Fan, and Y. Zheng, “Revealing False Positive Features in Epileptic EEG Identification,” *Int J Neural Syst*, vol. 30, no. 11, 2020, doi: 10.1142/S0129065720500173.

[245] X. Zhou, B. W. K. Ling, C. Li, and K. Zhao, “Epileptic seizure detection via logarithmic normalized functional values of singular values,” *Biomed Signal Process Control*, vol. 62, p. 102086, 2020, doi: 10.1016/j.bspc.2020.102086.

[246] A. Q. Ansari, P. Sharma, and M. Tripathi, “Automatic seizure detection using neutrosophic classifier,” *Phys Eng Sci Med*, vol. 43, no. 3, pp. 1019–1028, 2020, doi: 10.1007/s13246-020-00901-3.

[247] S. Raghu *et al.*, “Cross-database evaluation of EEG based epileptic seizures detection driven by adaptive median feature baseline correction,” *Clinical Neurophysiology*, vol. 131, no. 7, pp. 1567–1578, 2020, doi: 10.1016/j.clinph.2020.03.033.

[248] R. Abiyev, M. Arslan, J. B. Idoko, B. Sekeroglu, and A. Ilhan, “Identification of epileptic eeg signals using convolutional neural networks,” *Applied Sciences (Switzerland)*, vol. 10, no. 12, 2020, doi: 10.3390/APP10124089.

[249] K. P. Ayodele, W. O. Ikezogwo, M. A. Komolafe, and P. Ogunbona, “Supervised domain generalization for integration of disparate scalp EEG datasets for automatic epileptic seizure detection,” *Comput Biol Med*, vol. 120, 2020, doi: 10.1016/j.compbiomed.2020.103757.

[250] S. T. George, M. S. P. Subathra, N. J. Sairamya, L. Susmitha, and M. Joel Premkumar, “Classification of epileptic EEG signals using PSO based artificial neural network and tunable-Q wavelet transform,” *Biocybern Biomed Eng*, vol. 40, no. 2, pp. 709–728, 2020, doi: 10.1016/j.bbe.2020.02.001.

[251] Y. Li *et al.*, “Automatic Seizure Detection using Fully Convolutional Nested LSTM,” *Int J Neural Syst*, vol. 30, no. 4, 2020, doi: 10.1142/S0129065720500197.

[252] Y. Li, Y. Liu, W. G. Cui, Y. Z. Guo, H. Huang, and Z. Y. Hu, “Epileptic Seizure Detection in EEG Signals Using a Unified Temporal-Spectral Squeeze-and-Excitation Network,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 4, pp. 782–794, 2020, doi: 10.1109/TNSRE.2020.2973434.

[253] S. K. Rout and P. K. Biswal, “An efficient error-minimized random vector functional link network for epileptic seizure classification using VMD,” *Biomed Signal Process Control*, vol. 57, p. 101787, 2020, doi: 10.1016/j.bspc.2019.101787.

[254] J. Wu, T. Zhou, and T. Li, “Detecting epileptic seizures in EEG signals with complementary ensemble empirical mode decomposition and extreme gradient boosting,” *Entropy*, vol. 22, no. 2, 2020, doi: 10.3390/e22020140.

[255] G. R. Kiranmayi and V. Udayashankara, “Detection of epilepsy using discrete cosine harmonic wavelet transform-based features and neural network classifier,” *Int J Biomed Eng Technol*, vol. 32, no. 2, pp. 109–122, 2020, doi: 10.1504/IJBET.2020.105649.

[256] T. Zhang, W. Chen, and M. Li, “Complex-valued distribution entropy and its application for seizure detection,” *Biocybern Biomed Eng*, vol. 40, no. 1, pp. 306–323, 2020, doi: 10.1016/j.bbe.2019.10.006.

[257] D. P. Dash, M. H. Kolekar, and K. Jha, “Multi-channel EEG based automatic epileptic seizure detection using iterative filtering decomposition and Hidden Markov Model,” *Comput Biol Med*, vol. 116, p. 103571, 2020, doi: 10.1016/j.compbiomed.2019.103571.

[258] Y. Jiang, W. Chen, and M. Li, “Symplectic geometry decomposition-based features for automatic epileptic seizure detection,” *Comput Biol Med*, vol. 116, 2020, doi: 10.1016/j.compbiomed.2019.103549.

[259] T. Liu, N. D. Truong, A. Nikpour, L. Zhou, and O. Kavehei, “Epileptic Seizure Classification with Symmetric and Hybrid Bilinear Models,” *IEEE J Biomed Health Inform*, vol. 24, no. 10, pp. 2844–2851, 2020, doi: 10.1109/JBHI.2020.2984128.

[260] C. Lyu, Y. Chen, Z. Chen, Y. Liu, and Z. Wang, “Automatic epilepsy detection based on generalized convolutional prototype learning,” *Measurement (Lond)*, vol. 184, 2021, doi: 10.1016/j.measurement.2021.109954.

[261] A. Anuragi, D. S. Sisodia, and R. B. Pachori, “Automated FBSE-EWT based learning framework for detection of epileptic seizures using time-segmented EEG signals,” *Comput Biol Med*, vol. 136, 2021, doi: 10.1016/j.compbiomed.2021.104708.

[262] S. Supriya, S. Siuly, H. Wang, and Y. Zhang, “New feature extraction for automated detection of epileptic seizure using complex network framework,” *Applied Acoustics*, vol. 180, 2021, doi: 10.1016/j.apacoust.2021.108098.

[263] S. Panda, A. Das, S. Mishra, and M. N. Mohanty, “Epileptic seizure detection using deep ensemble network with empirical wavelet transform,” *Measurement Science Review*, vol. 21, no. 4, pp. 110–116, 2021, doi: 10.2478/msr-2021-0016.

[264] Z. Xiong *et al.*, “A study on seizure detection of eeg signals represented in 2d,” *Sensors*, vol. 21, no. 15, 2021, doi: 10.3390/s21155145.

[265] A. Shankar, H. K. Khaing, S. Dandapat, and S. Barma, “Analysis of epileptic seizures based on EEG using recurrence plot images and deep learning,” *Biomed Signal Process Control*, vol. 69, no. March, p. 102854, 2021, doi: 10.1016/j.bspc.2021.102854.

[266] A. Shariat, A. Zarei, S. A. Karvigh, and B. M. Asl, “Automatic detection of epileptic seizures using Riemannian geometry from scalp EEG recordings,” *Med Biol Eng Comput*, vol. 59, no. 7–8, pp. 1431–1445, 2021, doi: 10.1007/s11517-021-02385-z.

[267] H. He, X. Liu, and Y. Hao, “A progressive deep wavelet cascade classification model for epilepsy detection,” *Artif Intell Med*, vol. 118, Aug. 2021, doi: 10.1016/j.artmed.2021.102117.

[268] Y. Liu, B. Jiang, J. Feng, J. Hu, and H. Zhang, “Classification of EEG Signals for Epileptic Seizures Using Feature Dimension Reduction Algorithm based on LPP,” *Multimed Tools Appl*, vol. 80, no. 20, pp. 30261–30282, 2021, doi: 10.1007/s11042-020-09135-7.

[269] Y. Jiang, W. Chen, M. Li, T. Zhang, and Y. You, “Synchroextracting chirplet transform-based epileptic seizures detection using EEG,” *Biomed Signal Process Control*, vol. 68, no. 5988, p. 102699, 2021, doi: 10.1016/j.bspc.2021.102699.

[270] H. A. Glory, C. Vigneswaran, S. S. Jagtap, R. Shruthi, G. Hariharan, and V. S. S. Sriram, “AHW-BGOA-DNN: a novel deep learning model for epileptic seizure detection,” *Neural Comput Appl*, vol. 33, no. 11, pp. 6065–6093, 2021, doi: 10.1007/s00521-020-05384-7.

[271] Q. Wang, H. L. Wei, L. Wang, and S. Xu, “A novel time-varying modeling and signal processing approach for epileptic seizure detection and classification,” *Neural Comput Appl*, vol. 33, no. 11, pp. 5525–5541, 2021, doi: 10.1007/s00521-020-05330-7.

[272] D. Hu, J. Cao, X. Lai, J. Liu, S. Wang, and Y. Ding, “Epileptic Signal Classification Based on Synthetic Minority Oversampling and Blending Algorithm,” *IEEE Trans Cogn Dev Syst*, vol. 13, no. 2, pp. 368–382, 2021, doi: 10.1109/TCDS.2020.3009020.

[273] H. D. Praveena, C. Subhas, and K. R. Naidu, “Automatic epileptic seizure recognition using reliefF feature selection and long short term memory classifier,” *J Ambient Intell Humaniz Comput*, vol. 12, no. 6, pp. 6151–6167, 2021, doi: 10.1007/s12652-020-02185-7.

[274] H. Peng *et al.*, “Automatic epileptic seizure detection via Stein kernel-based sparse representation,” *Comput Biol Med*, vol. 132, 2021, doi: 10.1016/j.compbiomed.2021.104338.

[275] Aayesha, M. B. Qureshi, M. Afzaal, M. S. Qureshi, and M. Fayaz, "Machine learning-based EEG signals classification model for epileptic seizure detection," *Multimed Tools Appl*, vol. 80, no. 12, pp. 17849–17877, 2021, doi: 10.1007/s11042-021-10597-6.

[276] D. Hu, J. Cao, X. Lai, Y. Wang, S. Wang, and Y. Ding, "Epileptic State Classification by Fusing Hand-Crafted and Deep Learning EEG Features," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 4, pp. 1542–1546, 2021, doi: 10.1109/TCSII.2020.3031399.

[277] M. Li and W. Chen, "FFT-based deep feature learning method for EEG classification," *Biomed Signal Process Control*, vol. 66, no. July 2020, p. 102492, 2021, doi: 10.1016/j.bspc.2021.102492.

[278] A. Zarei and B. M. Asl, "Automatic seizure detection using orthogonal matching pursuit, discrete wavelet transform, and entropy based features of EEG signals," *Comput Biol Med*, vol. 131, 2021, doi: 10.1016/j.compbiomed.2021.104250.

[279] M. Sahani, S. Rout, and P. K. Dash, "Epileptic Seizure Recognition Using Reduced Deep Convolutional Stack Autoencoder and Improved Kernel RVFLN from EEG Signals," *IEEE Trans Biomed Circuits Syst*, 2021, doi: 10.1109/TBCAS.2021.3090995.

[280] S. Syed Rafiammal, D. Najumnissa Jamal, and S. Kaja Mohideen, "Detection of Epilepsy Seizure in Adults Using Discrete Wavelet Transform and Cluster Nearest Neighborhood Classifier," *Iranian Journal of Science and Technology - Transactions of Electrical Engineering*, vol. 45, no. 4, pp. 1103–1115, 2021, doi: 10.1007/s40998-021-00437-6.

[281] J. Cao, D. Hu, Y. Wang, J. Wang, and B. Lei, "Epileptic Classification with Deep Transfer Learning based Feature Fusion Algorithm," *IEEE Trans Cogn Dev Syst*, 2021, doi: 10.1109/TCDS.2021.3064228.

[282] C. Huang, W. Chen, M. Chen, and B. Yuan, "A Feature Fusion Framework and Its Application to Automatic Seizure Detection," *IEEE Signal Process Lett*, vol. 28, pp. 753–757, 2021, doi: 10.1109/LSP.2021.3069344.

[283] X. Jiang, K. Xu, R. Zhang, H. Ren, and W. Chen, "A redundancy removed, dual-tree, discretewavelet transform to construct compact representations for automated seizure detection," *Applied Sciences (Switzerland)*, vol. 9, no. 23, 2019, doi: 10.3390/app9235215.

[284] Y. Wang, Z. Li, L. Feng, C. Zheng, and W. Zhang, "Automatic Detection of Epilepsy and Seizure Using Multiclass Sparse Extreme Learning Machine Classification," *Comput Math Methods Med*, vol. 2017, 2017.

[285] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-context learning approach for EEG epileptic seizure detection," *BMC Syst Biol*, vol. 12, 2018, doi: 10.1186/s12918-018-0626-2.

[286] Y. Lu, Y. Ma, C. Chen, and Y. Wang, "Classification of single-channel EEG signals for epileptic seizures detection based on hybrid features," vol. 26, no. S1, pp. S337–S346, 2018, doi: 10.3233/THC-174679.

[287] L. Hussain *et al.*, "Regression analysis for detecting epileptic seizure with different feature extracting strategies," *Biomedizinische Technik*, 2019, doi: 10.1515/bmt-2018-0012.

[288] S.-L. Zhang, B. Zhang, Y.-L. Su, and J.-L. Song, "A novel EEG-complexity-based feature and its application on the epileptic seizure detection," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 12, pp. 3339–3348, 2019, doi: 10.1007/s13042-019-00921-w.

[289] A. M. Abdelhameed and M. Bayoumi, "Semi-supervised EEG signals classification system for epileptic seizure detection," *IEEE Signal Process Lett*, vol. 26, no. 12, pp. 1922–1926, 2019, doi: 10.1109/LSP.2019.2953870.

[290] M. Zhou *et al.*, "Epileptic seizure detection based on EEG signals and CNN," *Front Neuroinform*, vol. 12, 2018, doi: 10.3389/fninf.2018.00095.

[291] Z. Yu *et al.*, "Automatic seizure detection based on kernel robust probabilistic collaborative representation," *Med Biol Eng Comput*, vol. 57, no. 1, pp. 205–219, 2019, doi: 10.1007/s11517-018-1881-5.

[292] C. Sun, H. Cui, W. Zhou, W. Nie, X. Wang, and Q. Yuan, "Epileptic Seizure Detection with EEG Textural Features and Imbalanced Classification Based on EasyEnsemble Learning," *Int J Neural Syst*, vol. 29, no. 10, 2019, doi: 10.1142/S0129065719500217.

[293] R. Fu, Y. Tian, P. Shi, and T. Bao, "Automatic Detection of Epileptic Seizures in EEG Using Sparse CSP and Fisher Linear Discrimination Analysis Algorithm," *J Med Syst*, vol. 44, no. 2, 2020, doi: 10.1007/s10916-019-1504-1.

[294] S. Morales and M. E. Bowers, "Time-frequency analysis methods and their application in developmental EEG data," *Dev Cogn Neurosci*, vol. 54, p. 101067, Apr. 2022, doi: 10.1016/j.dcn.2022.101067.

[295] S. Saminu *et al.*, "Applications of Artificial Intelligence in Automatic Detection of Epileptic Seizures Using EEG Signals: A Review", doi: 10.47852/bonviewAIA2202297.

[296] I. Ahmad *et al.*, “EEG-Based Epileptic Seizure Detection via Machine/Deep Learning Approaches: A Systematic Review,” *Computational Intelligence and Neuroscience*, vol. 2022. Hindawi Limited, 2022. doi: 10.1155/2022/6486570.

[297] D. Merlin Praveena, D. Angelin Sarah, and S. Thomas George, “Deep Learning Techniques for EEG Signal Applications–A Review,” *IETE Journal of Research*, vol. 68, no. 4. Taylor and Francis Ltd., pp. 3030–3037, 2022. doi: 10.1080/03772063.2020.1749143.

[298] S. Supriya, S. Siuly, H. Wang, and Y. Zhang, “Epilepsy Detection from EEG using Complex Network Techniques: A Review,” *IEEE Rev Biomed Eng*, 2021, doi: 10.1109/RBME.2021.3055956.

[299] K. Rasheed *et al.*, “Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review,” Feb. 2020, [Online]. Available: http://arxiv.org/abs/2002.01925

[300] J. M. Koolhaas *et al.*, “Stress revisited: A critical evaluation of the stress concept,” *Neurosci Biobehav Rev*, vol. 35, no. 5, pp. 1291–1301, Apr. 2011, doi: 10.1016/j.neubiorev.2011.02.003.

[301] H. Yaribeygi, Y. Panahi, H. Sahraei, T. P. Johnston, and A. Sahebkar, “THE IMPACT OF STRESS ON BODY FUNCTION: A REVIEW,” *EXCLI J*, vol. 16, pp. 1057–1072, 2017.

[302] R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, and H. Al-Nashash, “A review on mental stress assessment methods using eeg signals,” *Sensors*, vol. 21, no. 15, 2021, doi: 10.3390/s21155043.

[303] J. U. Frisch, J. A. HÃ¤usser, and A. Mojzisch, “The Trier Social Stress Test as a paradigm to study how people respond to threat in social interactions,” *Front Psychol*, vol. 6, Feb. 2015, doi: 10.3389/fpsyg.2015.00014.

[304] A. L. Shilton, R. Laycock, and S. G. Crewther, “The Maastricht Acute Stress Test (MAST): Physiological and Subjective Responses in Anticipation, and Post-stress,” *Front Psychol*, vol. 8, Apr. 2017, doi: 10.3389/fpsyg.2017.00567.

[305] “Evoking stress reactivity in virtual reality a systematic review and metaanalysis.”

[306] D. Krupić, B. Žuro, and P. J. Corr, “Anxiety and threat magnification in subjective and physiological responses of fear of heights induced by virtual reality,” *Pers Individ Dif*, vol. 169, Feb. 2021, doi: 10.1016/j.paid.2019.109720.

[307] E. R. de Kloet, M. Joëls, and F. Holsboer, “Stress and the brain: From adaptation to disease,” *Nat Rev Neurosci*, vol. 6, no. 6, pp. 463–475, 2005, doi: 10.1038/nrn1683.

[308] A. Miltiadous *et al.*, “Alzheimer’s Disease and Frontotemporal Dementia: A Robust Classification Method of EEG Signals and a Comparison of Validation Methods,” *Diagnostics*, vol. 11, no. 8, p. 1437, Aug. 2021, doi: 10.3390/diagnostics11081437.

[309] E. Alyan *et al.*, “Frontal Electroencephalogram Alpha Asymmetry during Mental Stress Related to Workplace Noise,” *Sensors*, vol. 21, no. 6, p. 1968, Mar. 2021, doi: 10.3390/s21061968.

[310] M. Bares, T. Novak, P. Vlcek, M. Hejzlar, and M. Brunovsky, “Early change of prefrontal theta cordance and occipital alpha asymmetry in the prediction of responses to antidepressants,” *International Journal of Psychophysiology*, vol. 143, pp. 1–8, Sep. 2019, doi: 10.1016/j.ijpsycho.2019.06.006.

[311] M. Arns *et al.*, “EEG alpha asymmetry as a gender-specific predictor of outcome to acute treatment with different antidepressant medications in the randomized iSPOT-D study,” *Clinical Neurophysiology*, vol. 127, no. 1, pp. 509–519, Jan. 2016, doi: 10.1016/j.clinph.2015.05.032.

[312] J. Marín-Morales, C. Llinares, J. Guixeres, and M. Alcañiz, “Emotion Recognition in Immersive Virtual Reality: From Statistics to Affective Computing,” *Sensors*, vol. 20, no. 18, p. 5163, Sep. 2020, doi: 10.3390/s20185163.

[313] C. Stolz, D. Endres, and E. M. Mueller, “Threat-conditioned contexts modulate the late positive potential to faces—A mobile EEG/virtual reality study,” *Psychophysiology*, p. e13308, Dec. 2018, doi: 10.1111/psyp.13308.

[314] K. A. Fadeev, A. S. Smirnov, O. P. Zhigalova, P. S. Bazhina, A. v. Tumialis, and K. S. Golokhvast, “Too Real to Be Virtual: Autonomic and EEG Responses to Extreme Stress Scenarios in Virtual Reality,” *Behavioural Neurology*, vol. 2020, pp. 1–11, Mar. 2020, doi: 10.1155/2020/5758038.

[315] H. Wang, Q. Wang, and F. Hu, “Are you afraid of heights and suitable for working at height?,” *Biomed Signal Process Control*, vol. 52, pp. 23–31, Jul. 2019, doi: 10.1016/j.bspc.2019.03.011.

[316] S. Pratiher *et al.*, “Classification of VR-Gaming Difficulty Induced Stress Levels using Physiological (EEG & ECG) Signals and Machine Learning,” *UMBC Student …*, 2021, doi: 10.36227/techrxiv.16873471.v1.

[317] M. Athif *et al.*, “Using Biosignals for Objective Measurement of Presence in Virtual Reality Environments,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, Jul. 2020, pp. 3035–3039. doi: 10.1109/EMBC44109.2020.9176022.

[318] J. M. Torpy, A. E. Burke, and R. M. Glass, “Acute Emotional Stress and the Heart,” *JAMA*, vol. 298, no. 3, p. 360, Jul. 2007, doi: 10.1001/jama.286.3.374.

[319] W. E. J. Knight and N. S. Rickard, “Relaxing Music Prevents Stress-Induced Increases in Subjective Anxiety, Systolic Blood Pressure, and Heart Rate in Healthy Males and Females,” *J Music Ther*, vol. 38, no. 4, pp. 254–272, Dec. 2001, doi: 10.1093/jmt/38.4.254.

[320] H. Steiner, E. Ryst, J. Berkowitz, M. A. Gschwendt, and C. Koopman, “Boys’ and girls’ responses to stress: affect and heart rate during a speech task,” *Journal of Adolescent Health*, vol. 30, no. 4, pp. 14–21, Apr. 2002, doi: 10.1016/S1054-139X(01)00387-1.

[321] J. Taelman, S. Vandeput, A. Spaepen, and S. van Huffel, “Influence of Mental Stress on Heart Rate and Heart Rate Variability,” 2009, pp. 1366–1369. doi: 10.1007/978-3-540-89208-3_324.

[322] S. L. Corrigan, S. Roberts, S. Warmington, J. Drain, and L. C. Main, “Monitoring stress and allostatic load in first responders and tactical operators using heart rate variability: a systematic review,” *BMC Public Health*, vol. 21, no. 1, p. 1701, Dec. 2021, doi: 10.1186/s12889-021-11595-x.

[323] K. P. Sahoo *et al.*, “Alterations in Multi-channel EEG Dynamics During a Stressful Shooting Task in Virtual Reality Systems,” in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, Nov. 2021, pp. 6207–6210. doi: 10.1109/EMBC46164.2021.9630007.

[324] M. Stoeve, M. Wirth, R. Farlock, A. Antunovic, V. Müller, and B. M. Eskofier, “Eye Tracking-Based Stress Classification of Athletes in Virtual Reality,” *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, vol. 5, no. 2, pp. 1–17, May 2022, doi: 10.1145/3530796.

[325] D. R. Sanchez, E. Weiner, and A. van Zelderen, “Virtual reality assessments (VRAs): Exploring the reliability and validity of evaluations in VR,” *International Journal of Selection and Assessment*, vol. 30, no. 1, pp. 103–125, Mar. 2022, doi: 10.1111/ijsa.12369.

[326] A. Kalatzis, L. Stanley, and V. G. Prabhu, “Affective State Classification in Virtual Reality Environments Using Electrocardiogram and Respiration Signals,” in *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, IEEE, Nov. 2021, pp. 160–167. doi: 10.1109/AIVR52153.2021.00037.

[327] F. Hu, H. Wang, J. Chen, and J. Gong, “Research on the characteristics of acrophobia in virtual altitude environment,” in *2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR)*, IEEE, Aug. 2018, pp. 238–243. doi: 10.1109/IISR.2018.8535774.

[328] R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, and H. Al-Nashash, “A Review on Mental Stress Assessment Methods Using EEG Signals,” *Sensors*, vol. 21, no. 15, p. 5043, Jul. 2021, doi: 10.3390/s21155043.

[329] G. Vanhollebeke, S. de Smet, R. de Raedt, C. Baeken, P. van Mierlo, and M.-A. Vanderhasselt, “The neural correlates of psychosocial stress: A systematic review and meta-analysis of spectral analysis EEG studies,” *Neurobiol Stress*, vol. 18, p. 100452, May 2022, doi: 10.1016/j.ynstr.2022.100452.

[330] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, “Review on Psychological Stress Detection Using Biosignals,” *IEEE Trans Affect Comput*, vol. 13, no. 1, pp. 440–460, Jan. 2022, doi: 10.1109/TAFFC.2019.2927337.

[331] S. Paszkiel, P. Dobrakowski, and A. Łysiak, “The Impact of Different Sounds on Stress Level in the Context of EEG, Cardiac Measures and Subjective Stress Level: A Pilot Study,” *Brain Sci*, vol. 10, no. 10, p. 728, Oct. 2020, doi: 10.3390/brainsci10100728.

[332] X. Zhang, P. Bachmann, T. M. Schilling, E. Naumann, H. Schächinger, and M. F. Larra, “Emotional stress regulation: The role of relative frontal alpha asymmetry in shaping the stress response,” *Biol Psychol*, vol. 138, pp. 231–239, Oct. 2018, doi: 10.1016/j.biopsycho.2018.08.007.

[333] C. W. E. M. Quaedflieg, T. Meyer, F. T. Y. Smulders, and T. Smeets, “The functional role of individual-alpha based frontal asymmetry in stress responding,” *Biol Psychol*, vol. 104, pp. 75–81, Jan. 2015, doi: 10.1016/j.biopsycho.2014.11.014.

[334] M. Schneider, L. Chau, M. Mohamadpour, N. Stephens, K. Arya, and A. Grant, “EEG asymmetry and BIS/BAS among healthy adolescents,” *Biol Psychol*, vol. 120, pp. 142–148, Oct. 2016, doi: 10.1016/j.biopsycho.2016.09.004.

[335] J. F. Thayer, F. Åhs, M. Fredrikson, J. J. Sollers, and T. D. Wager, “A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health,” *Neurosci Biobehav Rev*, vol. 36, no. 2, pp. 747–756, Feb. 2012, doi: 10.1016/j.neubiorev.2011.11.009.

[336] J. F. Veale, “Edinburgh Handedness Inventory – Short Form: A revised version based on confirmatory factor analysis,” *Laterality: Asymmetries of Body, Brain and Cognition*, vol. 19, no. 2, pp. 164–177, Mar. 2014, doi: 10.1080/1357650X.2013.783045.

Abbreviations list

Abbreviation	Full Phrase
AD	Alzheimer's Disease
ADHD	Attention Deficit Hyperactivity Disorder
AED	Anti-Epileptic Drug
AIIMS	All India Institutes of Medical Sciences
AR	AutoRegressive
ASR	Artifact Subspace Reconstruction
AUC	Area Under Curve
BCI	Brain-Computer Interfaces
BCV	Between-Class Variance
BERT	Bidirectional Encoder Representations from Transformers
BFGS	Broyden-Flecher-Goldfarb-Shanno (algorithm)
BIDS	Brain Imaging Data Structure
BIS	Behavior Inhibition System
BPM	Beats Per Minute
CDR	Clinical Dementia Rating
CHB-MIT	Childrens Hospital of Boston – Massachusetts Institute of Technology
CLAIM	Credibility, Legality, Affordability, Interpretability, Maintainability
CLS	Class token
CN	Control Group
CNN	Convolutional Neural Network
CT scan	Computed Tomography scan
DB	DataBase
DFT	Discrete Fourier Transform
DICE-net	Dual Input Convolution Transformer network
DSM-IV	Diagnostic and Statistical Manual
DWT	Discrete Wavelet Transform
ECG	ElectroCardioGram
EEG	Electroencephalogram
EMD	Empirical Mode Decomposition
ERD/ERS	Event Related Desynchronization/ Synchronization
ERP	Event-Related Potential
FAA	Frontal Alpha Asymmetry
FFN	Feed Forward Network
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
fMRI	functional-MRI
FN	False Negative
FP	False Positive
FT	Fourier Transform
FTD	FrontoTemporal Dementia

GOSS	Gradient based One-Side Sampling
GPU	Graphic Processing Unit
HHT	Hilbert-Huang Transform
HMD	Head Mounted Display
HRV	Heart Rate Variability
ICA	Independent Component Analysis
IIR	Infinite Impulse Response
KITS	Karunya Institute of Technology and Sciences
k-NN	k-Nearest Neighbors
LDA	Linear Discriminant Analysis
LORETA	Low-Resolution Electromagnetic Tomography
LOSO	Leave-One-Subject-Out
LSTM	Long Short-Term Memory
ML	Machine Learning
MLP	MultiLayer Perceptron
MMSE	Mini Mental State Examination
MRI	Magnetic Resonance Imaging
MSA	Multi-Head Self-Attention
MUMC	Department of Clinical Neurophysiology, Maastricht
NCS	Neurology and Sleep Center
NLP	Natural Language Processing
OAA	Occipital Alpha Asymmetry
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PLV	Phase Locking Value
PPI	Pixels per Inch
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSD	Power Spectral Density
PUPH	Peking Universily People's Hospital
RBP	Relative Band Power
REM	Rapid Eye Movement
RGB	Red Green Blue
RMCH	Institute of Neuroscience, Ramaiah Memorial Hospital
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
RoI	Region of Interest
SCC	Spectral Coherence Connectivity
SNR	Signal to Noise ratio
SSVEP	Steady-State Visual Evoked Potential
STFT	Short-Time Fourier Transform
SVM	Support Vector Machines
TF	Time-Frequency
TN	True Negative
TP	True Positive
TUH-EEG	Temple University Hospital EEG database

ViT	Vision Transformer
VR	Virtual Reality
WCV	Within-Class Variance