POLITECNICO DI TORINO

Master Degree in Computer Engineering

Artificial Intelligence and Data Analytics

Master Thesis

EEG-based Neurological Disease Classification using Machine Learning Techniques

Supervisor
Dr. Luigi Borzì

Candidate
Simone Scalora

Academic Year 2025-2026

Summary

Electroencephalography (EEG) is a non-invasive, relatively inexpensive and widely available technique for monitoring brain activity. Due to its high temporal resolution and ease of acquisition, EEG has become an important tool for investigating neurological disorders and is increasingly being employed in computer-aided diagnostic systems. In recent years, machine learning techniques have shown considerable potential in identifying disease-specific EEG patterns, supporting clinicians in the diagnosis and monitoring of neurological conditions. However, most existing studies are limited to single datasets acquired under homogeneous conditions and mainly focus on binary classification problems, reducing the generalizability of proposed models and their applicability to real-world clinical scenarios.

This work addresses these limitations by investigating the automatic classification of three neurological disorders — Alzheimer's disease (AD), Epilepsy (EP) and Amyotrophic Lateral Sclerosis (ALS) — using EEG recordings collected from multiple publicly available sources. To this end, six EEG datasets were integrated into a unified analysis framework, comprising a total of 525 subjects, including 312 Healthy Controls (HC), 157 patients with Alzheimer's disease, 50 patients with Epilepsy and 6 patients with Amyotrophic Lateral Sclerosis. The datasets were acquired using different experimental protocols and distributed in heterogeneous formats.

A standardized preprocessing pipeline was developed, including filtering, resampling, epoch segmentation and artifact rejection. Following preprocessing, spectral and statistical features were extracted from the canonical EEG frequency bands (delta, theta, alpha, beta, gamma) using power spectral density estimation. Additional clinically relevant spectral ratios were also computed in order to capture disease-related changes in brain rhythms.

The extracted feature vectors were used to train Support Vector Machine (SVM) classifiers. Model development was performed within a Nested Cross-Validation framework combined with GroupKFold splitting, ensuring complete subject-level separation between training and testing sets. Furthermore, feature importance analysis, feature selection strategies, class balancing techniques and Optuna-based hyperparameter optimization were investigated to maximize classification performance while minimizing overfitting.

The proposed framework was evaluated on both binary and multiclass classification tasks. Binary experiments included HC vs. AD, HC vs. EP, and HC vs. ALS comparisons. The best-performing models achieved balanced accuracies

of 92.02%, 87.05% and 91.67%, respectively. In addition, several multiclass classification strategies were explored to simultaneously discriminate among healthy subjects and all three neurological disorders. Across these configurations, subject-level balanced accuracies above 83.9% were consistently achieved, demonstrating the feasibility of multiclass EEG-based neurological screening despite the increased complexity of the task.

The obtained results show that EEG-derived spectral features, when combined with machine learning techniques, can effectively discriminate between healthy individuals and patients affected by different neurological disorders. The main contributions of this work are the integration of multiple heterogeneous EEG datasets into a single analysis framework, the investigation of both binary and multiclass diagnostic scenarios and the adoption of a rigorous validation methodology based on Nested Cross-Validation, subject-level data separation and balanced accuracy as the primary evaluation metric. These methodological choices improve the reliability and generalizability of the results, particularly in the presence of class imbalance and heterogeneous data sources.

In general, these findings support the potential of EEG and machine learning as complementary tools for the development of automated neurological screening systems. Future work may involve the inclusion of additional neurological conditions and the validation of the proposed framework on larger and more balanced clinical datasets. Furthermore, the integration of complementary modalities such as eye-tracking data could provide additional information about cognitive and behavioral processes, potentially improving classification performance. Another research direction is to extend the proposed conceptual framework towards the classification and recognition of human actions, cognitive states, and task-related activities, enabling applications in brain-computer interfaces, human-computer interaction and neurorehabilitation.

Contents

List of Tables 6

List of Figures 8

1 Introduction 10
- 1.1 Neurological Disorders and Diagnostic Challenges . . . 10
- 1.2 Electroencephalography . . . 12
  - 1.2.1 Overview of Electroencephalography . . . 12
  - 1.2.2 EEG Signal Characteristics . . . 12
  - 1.2.3 EEG Frequency Bands . . . 13
- 1.3 EEG in Neurological Disease Assessment . . . 15
  - 1.3.1 EEG and Alzheimer's Disease . . . 15
  - 1.3.2 EEG and Epilepsy . . . 15
  - 1.3.3 EEG and Amyotrophic Lateral Sclerosis . . . 16
  - 1.3.4 Comparison with Healthy Subjects . . . 16
- 1.4 Machine Learning for EEG Analysis . . . 18
- 1.5 Study Motivation and Objectives . . . 19
- 1.6 Thesis Organization . . . 20

2 Background and Related Work 21
- 2.1 Electroencephalography Fundamentals . . . 21
  - 2.1.1 EEG Acquisition Process . . . 21
  - 2.1.2 International 10-20 System . . . 22
  - 2.1.3 Artifacts and Noise . . . 23
  - 2.1.4 EEG Preprocessing Techniques . . . 24
  - 2.1.5 EEG Feature Extraction . . . 25
- 2.2 Machine Learning for EEG Analysis . . . 27
  - 2.2.1 Support Vector Machine . . . 27
  - 2.2.2 Multiclass Classification Strategies . . . 28
  - 2.2.3 Data Augmentation Techniques . . . 28
- 2.3 Related Work . . . 30

2.4 Research Gap . . . 33
2.5 Thesis Contributions . . . 34

3 Materials and Methods 35
3.1 Overview of the Proposed Framework . . . 35
3.2 EEG Datasets . . . 37
3.2.1 Alzheimer's Disease Datasets . . . 38
3.2.2 Epilepsy Dataset . . . 39
3.2.3 ALS Dataset . . . 40
3.3 EEG Preprocessing . . . 41
3.3.1 Signal Standardization . . . 41
3.3.2 Re-referencing . . . 42
3.3.3 Filtering . . . 42
3.3.4 Resampling . . . 42
3.3.5 Epoch Segmentation . . . 43
3.3.6 Artifact Rejection . . . 43
3.4 Feature Extraction . . . 45
3.5 Classification Framework . . . 48
3.5.1 Support Vector Machine Classifier . . . 48
3.5.2 Nested Cross-Validation Strategy . . . 48
3.5.3 Class Balancing Strategy . . . 49
3.5.4 Feature Importance Analysis . . . 49
3.5.5 Feature Selection . . . 50
3.5.6 Hyperparameter Optimization . . . 50
3.5.7 Evaluation Metrics . . . 51

4 Experimental Results 54
4.1 Binary Classification . . . 54
4.1.1 Alzheimer's Disease Classification . . . 54
4.1.2 Epilepsy Classification . . . 57
4.1.3 Amyotrophic Lateral Sclerosis Classification . . . 59
4.2 Multiclass Classification . . . 63
4.3 Two-Stage Classification Analysis . . . 67
4.3.1 Healthy vs Disease Classification . . . 67
4.3.2 Disease-Specific Classification . . . 69
4.4 Leakage-Free Hierarchical Cascade . . . 73
4.4.1 Class-wise Performance Analysis . . . 76
4.5 Data Augmentation using SMOTE and CTGAN . . . 78
4.6 Three-class Classification . . . 81
4.7 Comparison of the Proposed Strategies . . . 86
4.8 Comparison with State-of-the-Art Studies . . . 88

5 Discussion 91
5.1 Main Findings . . . 91
5.2 Clinical Implications . . . 93
5.3 Limitations . . . 94

6 Conclusions and Future Research 95
6.1 Future Research . . . 97

A Additional Experimental Details 103
A.1 Optuna Hyperparameter Optimization . . . 103
A.2 Binary Classification . . . 104
A.2.1 Alzheimer's Disease . . . 104
A.2.2 Epilepsy . . . 104
A.2.3 Amyotrophic Lateral Sclerosis . . . 105
A.3 Multiclass Classification . . . 105
A.3.1 Standard Multiclass SVM . . . 105
A.4 Hierarchical Classification Strategies . . . 105

B Additional Segment-Level Results 106
B.1 Standard Multiclass Classification . . . 106
B.2 Healthy vs Disease Classification . . . 107
B.3 Disease-Specific Classification . . . 107
B.4 Leakage-Free Hierarchical Cascade . . . 108
B.5 Data Augmentation using SMOTE and CTGAN . . . 108
B.6 Clinically-Guided Three-Class Classification . . . 109
B.6.1 Stage 1: Healthy vs Epilepsy vs Neurodegenerative Diseases 109
B.6.2 Stage 2: Alzheimer's Disease vs ALS . . . 109

List of Tables

1.1 EEG Frequency Bands 13

2.1 Summary of representative EEG-based machine learning studies for the classification of Alzheimer's disease, Epilepsy and Amyotrophic Lateral Sclerosis. The table focuses on methodological aspects and study limitations. Abbreviations: AD, Alzheimer's disease; EP, Epilepsy; ALS, Amyotrophic Lateral Sclerosis; FTD, Frontotemporal Dementia; HC, Healthy Controls; Ch., channels. 31

3.1 Datasets used in this work. Abbreviations: AD: Alzheimer's disease; EP: Epilepsy; ALS: Amyotrophic Lateral Sclerosis; FTD: Frontotemporal Dementia; MCI: Mild Cognitive Impairment; HC: Healthy Controls. 37

4.1 Performance obtained for Alzheimer's disease binary classification using different stages of the proposed classification pipeline. 55

4.2 Performance obtained for Epilepsy binary classification using different stages of the proposed classification pipeline. 57

4.3 Segment-level performance obtained for ALS binary classification using different stages of the proposed classification pipeline. 59

4.4 Subject-level performance obtained for ALS binary classification using different stages of the proposed classification pipeline. 60

4.5 Subject-level performance obtained for the multiclass classification problem. 63

4.6 Subject-level performance obtained for the Healthy vs Disease classification task. 68

4.7 Subject-level performance obtained for the disease-specific classification task (Alzheimer's disease, Epilepsy and ALS). 70

4.8 Subject-level performance obtained for the leakage-free hierarchical cascade. 73

4.9 Class-wise subject-level performance obtained by the final optimized leakage-free hierarchical cascade (Optuna + Top-150 features). All metrics are reported on a scale from 0 to 1. 76

4.10 Subject-level performance obtained using SMOTE and CTGAN data augmentation. . . . 78
4.11 Subject-level performance obtained for the first stage of the three-class hierarchical framework (Healthy vs Epilepsy vs Neurodegenerative Diseases). . . . 81
4.12 Subject-level performance obtained for the second stage of the three-class hierarchical framework (Alzheimer's disease vs ALS). . . . 82
4.13 Overall comparison of the best subject-level balanced accuracy obtained for the different classification strategies investigated in this work. . . . 86
4.14 Comparison between representative state-of-the-art EEG-based neurological disease classification studies and the proposed framework. Reported results are not directly comparable because of differences in datasets, datasets sizes, EEG acquisition protocols, validation strategies, task complexity and evaluation metrics. . . . 89
A.1 Best hyperparameter configurations identified by Optuna for the Alzheimer's disease classification experiments. . . . 104
A.2 Best hyperparameter configurations identified by Optuna for the epilepsy classification experiments. . . . 104
A.3 Best hyperparameter configurations identified by Optuna for the ALS classification experiments. . . . 105
A.4 Best hyperparameter configurations identified by Optuna for the standard multiclass SVM experiments. . . . 105
B.1 Segment-level performance obtained for the standard multiclass classification task. . . . 106
B.2 Segment-level performance obtained for the Healthy vs Disease classification task. . . . 107
B.3 Segment-level performance obtained for the disease-specific classification task (Alzheimer's disease, Epilepsy and ALS). . . . 107
B.4 Segment-level performance obtained for the leakage-free hierarchical cascade. . . . 108
B.5 Segment-level performance obtained using SMOTE and CTGAN data augmentation. . . . 108
B.6 Segment-level performance obtained for the first stage of the clinically-guided hierarchical framework. . . . 109
B.7 Segment-level performance obtained for the second stage of the clinically-guided hierarchical framework. . . . 109

List of Figures

2.1 Representation of the International 10–20 electrode placement system commonly adopted in EEG acquisition 23
3.1 Workflow of the proposed methodology 36
4.1 Balanced Accuracy obtained for different Top-K feature subsets during the feature selection procedure for Alzheimer's disease classification. Error bars represent the standard deviation across the outer folds of the nested cross-validation. 56
4.2 Balanced Accuracy obtained for different Top-K feature subsets during the feature selection procedure for epilepsy classification. Error bars represent the standard deviation across the outer folds of the nested cross-validation. 58
4.3 Segment-level Balanced Accuracy obtained for different Top-K feature subsets during the feature selection procedure for ALS classification. Error bars represent the standard deviation across the outer folds of the nested cross-validation. 61
4.4 Subject-level Balanced Accuracy obtained for different Top-K feature subsets during the feature selection procedure for ALS classification. 62
4.5 Multiclass: Baseline SVM Confusion Matrix. 65
4.6 Multiclass: Baseline SVM Confusion Matrix using Optuna + Top-125. 65
4.7 Healthy vs Disease: Baseline Confusion Matrix. 68
4.8 Healthy vs Disease: Confusion Matrix obtained using the Top-100 Feature subset. 69
4.9 Disease-specific Classification: Baseline Confusion Matrix. 71
4.10 Disease-specific Classification: Baseline Confusion Matrix using the Top-50 Feature subset. 71
4.11 Cascade: Baseline Confusion Matrix. 74
4.12 Cascade: Baseline Confusion Matrix using Optuna + Top-150. . . . 75
4.13 SMOTE: Confusion matrix. 79
4.14 CTGAN: Confusion Matrix. 80
4.15 Three-Class: Baseline Confusion Matrix. 83

4.16 Three-Class: Baseline Confusion Matrix using Optuna + Top-125. . 83
4.17 AD vs ALS: Baseline Confusion Matrix. 84
4.18 AD vs ALS: Baseline Confusion Matrix using Optuna + Top-40. . . 85

Chapter 1

Introduction

1.1 Neurological Disorders and Diagnostic Challenges

Neurological disorders represent one of the most significant global health challenges, affecting millions of individuals worldwide and constituting a major cause of disability and reduced quality of life [1].

Among neurological disorders, conditions such as Alzheimer's Disease (AD), Epilepsy (EP) and Amyotrophic Lateral Sclerosis (ALS) are particularly relevant for their clinical complexity, progressive nature and impact on patients' daily lives. Despite their differences in terms of symptoms and progression, these disorders share a common challenge: obtaining an accurate and timely diagnosis remains difficult, especially during early stages or atypical clinical manifestations.

Alzheimer's Disease is the most common cause of dementia [2] and is characterized by a progressive decline in cognitive functions, including memory, reasoning and decision-making abilities. Since symptoms often develop gradually and may initially overlap with normal aging effects, early detection remains challenging. Epilepsy, on the other hand, is a chronic neurological condition [3] characterized by recurrent seizures caused by abnormal neuronal activity. Although seizures can often be identified clinically, distinguishing pathological brain activity and accurately characterizing epileptic conditions may still require complex evaluations. Amyotrophic Lateral Sclerosis is a progressive neurodegenerative disease [4], leading to muscular impairment and loss of motor control. In this context, ALS diagnosis may also be difficult and time-consuming.

Traditional diagnostic approaches for neurological disorders often depend on clinical evaluations, medical history, imaging techniques and expert interpretation. While these methods remain essential in medical practice, they may present limitations related to subjectivity, accessibility, cost or delayed identification of

disease-related alterations. Consequently, research efforts have focused on capture neurological dysfunction in a non-invasive and reliable way.

In this context, physiological signals generated by brain activity and, in particular, Electroencephalography (EEG) has gained growing interest due to its non-invasive nature, relatively low acquisition cost, high temporal resolution and capability to capture real-time electrical brain dynamics [5]. These characteristics make EEG a promising candidate for the investigation of neurological alterations and for the development of computational methods finalized at automatic disease characterization and classification.

1.2 Electroencephalography

1.2.1 Overview of Electroencephalography

Electroencephalography is a non-invasive neurophysiological technique used to measure the electrical activity generated by neuronal interactions within the brain [6].

EEG signals are acquired through electrodes placed on the scalp, which detect small voltage changes caused by large groups of brain cells (neurons) working together at the same time [7]. The placement of electrodes generally follows standardized configurations, among which the International 10–20 System is one of the most commonly used approaches, ensuring consistent spatial coverage of different brain regions and reproducibility across clinical and research settings.

One of the main advantages of EEG is in its non-invasive nature, portability, relatively low acquisition cost and high temporal resolution, typically in the order of milliseconds [8]. These properties make EEG suitable for continuous monitoring, allowing its use in both clinical practice and research applications. However, EEG recordings are also characterized by high complexity and susceptibility to noise, requiring dedicated signal processing and analytical methodologies for reliable interpretation.

Due to these characteristics, EEG has been extensively used in the diagnosis and monitoring of several neurological disorders, including epilepsy, sleep disorders and neurodegenerative diseases.

1.2.2 EEG Signal Characteristics

Despite its numerous advantages, electroencephalographic signals present several intrinsic challenges that make their analysis and interpretation particularly complex.

First, brain signals are never still — they change continuously depending on what we think, feel or experience. This means that the signal keeps shifting, so finding stable patterns is difficult.

Second, EEG is easily disturbed. Small movements like eye movement and blinking, cardiac activity or head movements can create "noise" that mixes with the brain's real signals. Environmental electrical interference can also cause problems.

Third, brain activity varies a lot from person to person. Age, mood, health or even the time of day can change the signal. The same person can also show different patterns on different days.

Fourth, EEG records from many spots on the scalp at once. This creates a large, complex dataset that combines time, space and multiple brain rhythms [9].

Because of these characteristics, traditional manual inspection of EEG recordings may be time-consuming and strongly dependent on expert interpretation. Consequently, advanced signal processing techniques and computational approaches have become increasingly important for extracting relevant information and identifying subtle neural patterns that may not be immediately observable through conventional analysis.

1.2.3 EEG Frequency Bands

Electroencephalographic activity is commonly analyzed in terms of oscillatory components distributed across specific frequency ranges, commonly referred to as EEG frequency bands. These rhythms reflect different patterns of neuronal synchronization and are associated with distinct physiological and cognitive processes. The spectral decomposition of EEG signals represents one of the most widely adopted approaches for studying brain activity, as variations in specific frequency components may provide useful information about neural states and neurological dysfunctions.

Traditionally, EEG activity is divided into five main frequency bands (as shown in Table 1.1): delta, theta, alpha, beta and gamma, each characterized by a specific frequency range and functional interpretation [10].

Frequency Band	**Frequency Range**	**Typical Functional Association**
Delta	0.5–4 Hz	Deep sleep, unconscious states
Theta	4–8 Hz	Drowsiness, falling asleep, REM phase
Alpha	8–13 Hz	Relaxed wakefulness, eyes closed, resting state
Beta	13–30 Hz	Active thinking, attention, cognitive processing
Gamma	>30 Hz	Higher cognitive processing, high focus

Table 1.1. EEG Frequency Bands

Delta activity is generally dominant during deep sleep and unconscious physiological conditions. Elevated delta power in awake individuals may sometimes be associated with abnormal neurological states. Theta rhythms are commonly linked to attention mechanisms and transitions between wakefulness and sleep. Alterations in theta activity have been reported in several neurological disorders, particularly those involving cognitive impairment.

The alpha band is one of the most prominent EEG rhythms and is typically observed during relaxed wakefulness, especially with closed eyes. Changes in alpha power or synchronization have frequently been associated with neurodegenerative

processes and altered cognitive functioning. Conversely, beta activity is generally related to active cognitive engagement, motor functions, and alertness. Finally, gamma oscillations, have been associated with higher-order cognitive functions, sensory integration, and neural communication mechanisms.

The analysis of EEG frequency bands plays a central role in neurological research and clinical investigations, as abnormalities in spectral activity may reflect alterations in brain dynamics associated with pathological conditions [11]. Several studies have shown that neurological disorders such as Alzheimer's Disease, Epilepsy and Amyotrophic Lateral Sclerosis may exhibit characteristic changes in specific EEG rhythms, motivating the use of spectral information for computational analysis and disease classification.

1.3 EEG in Neurological Disease Assessment

Electroencephalography has been extensively investigated as a valuable tool for studying neurological disorders due to its capability to capture alterations in brain electrical activity associated with pathological conditions. EEG recordings may reveal characteristic patterns that differ from those observed in healthy individuals [12].

In the context of this study, particular attention is devoted to Alzheimer's Disease, Epilepsy and Amyotrophic Lateral Sclerosis, as these disorders may present measurable variations in neural activity that can potentially be exploited for computational disease discrimination.

1.3.1 EEG and Alzheimer's Disease

Alzheimer's Disease is a progressive neurodegenerative disorder primarily associated with cognitive decline, memory impairment, and loss of executive functions. From an electrophysiological perspective, AD has been associated with significant alterations in brain activity, making EEG a promising tool for investigating disease-related neural changes.

Several studies have reported a characteristic slowing of EEG activity in patients affected by Alzheimer's Disease, typically involving increased activity in lower-frequency rhythms such as delta and theta bands, accompanied by reductions in alpha and beta oscillations [13], suggesting that EEG signals may contain discriminative information useful for distinguishing healthy individuals from affected subjects.

Because of these properties, EEG-based approaches have increasingly been combined with computational techniques and machine learning models aimed at automatic disease characterization and classification.

1.3.2 EEG and Epilepsy

Epilepsy is a chronic neurological disorder characterized by recurrent and unprovoked seizures resulting from abnormal and excessive neuronal activity in the brain. Among neurological disorders, epilepsy represents one of the most prominent application domains of electroencephalography, as EEG recordings are routinely employed in clinical practice to support diagnosis, seizure monitoring and characterization of epileptic activity.

Unlike neurodegenerative diseases, where EEG alterations may emerge progressively and subtly, epilepsy is often associated with more evident electrophysiological abnormalities. In particular, EEG recordings of epileptic patients may exhibit characteristic patterns such as spikes, sharp waves, spike-and-wave complexes

and abnormal rhythmic discharges, which are considered important indicators of epileptiform brain activity [14]. However, these abnormalities are not always continuously observable and may vary substantially depending on seizure type, brain region involvement, recording conditions and patient-specific characteristics.

Despite the clinical relevance of EEG in epilepsy diagnosis, manual interpretation of EEG recordings remains a complex and time-consuming task that strongly depends on expert neurological assessment.

For these reasons, computational approaches based on signal processing and machine learning have gained increasing attention for automatic seizure detection, epileptic pattern recognition, and EEG-based classification systems.

1.3.3 EEG and Amyotrophic Lateral Sclerosis

Amyotrophic Lateral Sclerosis is a progressive neurodegenerative disorder primarily affecting upper and lower motor neurons, leading to progressive muscle weakness, loss of motor control, and severe functional impairment.

Unlike epilepsy, where abnormal EEG patterns may be directly observable, or Alzheimer's Disease, where characteristic slowing phenomena have been extensively documented, EEG alterations in ALS are generally less clearly defined.

In particular, EEG investigations have highlighted modifications in sensorimotor rhythms and alterations in spectral power distributions, especially within alpha and beta frequency bands, which are closely related to motor processing. These findings suggest that EEG may capture functional brain changes associated with ALS progression and potentially provide complementary information for disease characterization [15].

Despite the relatively limited availability of EEG-based biomarkers for ALS compared to other neurological disorders, recent advances in computational analysis and machine learning techniques have encouraged further exploration of EEG signals for automatic disease assessment and discrimination between healthy and pathological subjects. Consequently, EEG-based approaches may contribute to improving the understanding and characterization of ALS-related neural alterations.

1.3.4 Comparison with Healthy Subjects

A common objective in EEG-based neurological research is the identification of differences between healthy individuals and patients affected by neurological disorders. Since many pathological conditions involve alterations in neuronal activity, cortical synchronization, and brain connectivity, EEG recordings may exhibit measurable deviations from patterns typically observed in healthy subjects.

In healthy individuals, EEG activity generally exhibits physiological oscillatory patterns across different frequency bands, reflecting normal variations in brain

activity associated with age, cognitive state, and level of vigilance [12].

However, distinguishing healthy and pathological EEG signals remains a challenging task due to the intrinsic complexity of brain activity, signal variability, overlapping characteristics across neurological conditions and the presence of noise and artifacts. In particular, some EEG alterations may be subtle and difficult to identify through conventional visual inspection alone, especially in early disease stages or heterogeneous patient populations.

For these reasons, machine learning techniques offer promising opportunities for automatically identifying complex neural patterns and supporting the classification of healthy and diseased subjects based on EEG-derived information.

1.4 Machine Learning for EEG Analysis

In the context of electroencephalography, machine learning (ML) has emerged as a promising approach for identifying complex patterns embedded within EEG recordings and supporting the automatic characterization of neurological conditions.

As discussed in the previous sections, EEG signals are complex for their non-stationary nature, susceptibility to noise and substantial inter-subject variability. These characteristics often make traditional manual interpretation difficult, time-consuming and highly dependent on expert knowledge.

Although clinical EEG inspection remains fundamental in medical practice, ML-based systems may support classification, anomaly detection, feature selection and predictive analysis tasks in neurological applications.

In EEG-based disease assessment, machine learning pipelines commonly involve several processing stages, including signal preprocessing, dimensionality reduction, feature extraction and classification. In this context, classification algorithms can be trained to distinguish between different subject categories or neurological conditions.

Over the past years, several studies have explored the integration of EEG signals with machine learning approaches for the analysis of neurological disorders, reporting promising results in areas such as seizure detection in Epilepsy and cognitive impairment assessment in Alzheimer's Disease. However, many existing studies focus on individual diseases independently or address binary classification problems [16].

These considerations motivate the investigation of more comprehensive computational frameworks capable of exploiting EEG information for distinguishing multiple neurological conditions.

1.5 Study Motivation and Objectives

The growing availability of electroencephalographic data and the rapid development of machine learning techniques have created new opportunities for supporting the diagnosis of neurological disorders through automated computational methods. Despite the promising results reported in the literature, developing robust classification models remains challenging due to the intrinsic complexity of EEG signals, the heterogeneity of acquisition protocols and the limited availability of large and balanced datasets.

Motivated by these challenges, this work investigates the use of machine learning techniques applied to EEG signals for the classification of neurological disorders, with particular focus on Alzheimer's disease, Epilepsy and Amyotrophic Lateral Sclerosis. Binary classification experiments are first performed to distinguish healthy subjects from each neurological condition independently. Subsequently, several multiclass classification strategies are investigated to evaluate their ability to discriminate multiple neurological disorders simultaneously.

The study also explores different methodological solutions, including hierarchical classification frameworks, feature selection techniques, hyperparameter optimization and data augmentation strategies, with the objective of assessing their impact on classification performance. Overall, this work aims to evaluate the effectiveness of EEG-based machine learning approaches for neurological disease classification under increasingly complex and clinically relevant experimental scenarios.

1.6 Thesis Organization

The remainder of this thesis is organized as follows.

Chapter 2 presents the theoretical background and related work concerning electroencephalographic signal analysis, neurological disorders and machine learning approaches applied to EEG-based disease characterization.

Chapter 3 describes the experimental framework adopted in this work, including dataset characteristics, preprocessing procedures, feature extraction, classification methodologies and experimental protocols employed for both binary and multiclass scenarios.

Chapter 4 presents the obtained experimental results, reporting the performance achieved across the different classification settings and comparative evaluation strategies.

Chapter 5 discusses the obtained findings, analyzes the strengths and limitations of the proposed approaches, and compares the experimental outcomes with observations reported in the literature.

Finally, Chapter 6 concludes the thesis by summarizing the main findings of the study, discussing limitations and outlining possible directions for future research.

Chapter 2

Background and Related Work

2.1 Electroencephalography Fundamentals

Electroencephalography represents one of the most adopted neurophysiological techniques for investigating brain activity in both clinical and research contexts. Because of its non-invasive nature, high temporal resolution and relatively low acquisition cost [8], EEG has been extensively employed for supporting neurological diagnosis.

From a computational point of view, EEG recordings often require dedicated signal processing methodologies and feature extraction strategies capable of capturing meaningful neural information while reducing signal complexity [7].

This chapter provides the theoretical foundations necessary for understanding the methodologies adopted in this work. First, the main principles of EEG acquisition and recording are introduced, followed by a discussion of electrode placement systems, signal artifacts, preprocessing procedures and EEG feature extraction techniques. Subsequently, the machine learning methodologies adopted for EEG classification are described, together with a review of related studies concerning EEG-based neurological disease assessment.

2.1.1 EEG Acquisition Process

EEG recordings are acquired using electrodes positioned on the scalp, which measure voltage varations over time relative to a reference electrode [7]. The recorded electrical signals are typically characterized by very small amplitudes, generally in the range of microvolts (µV).

The recorded EEG signal can be represented as a multichannel time series,

where each channel corresponds to a specific scalp location and captures neural activity associated with underlying cortical regions. Depending on the acquisition protocol and clinical setting, EEG systems may involve different numbers of electrodes.

One of the most relevant characteristics of EEG acquisition is its high temporal resolution, typically allowing brain activity to be sampled at frequencies ranging from hundreds to thousands of samples per second. This capability enables EEG to capture rapid neural dynamics and transient brain events that may not be observable through imaging techniques characterized by lower temporal sensitivity.

Despite its advantages, EEG acquisition is sensitive to several factors that may affect signal quality, including electrode placement, environmental interference and physiological artifacts. Consequently, appropriate preprocessing and signal conditioning procedures are generally required before performing computational analyses.

2.1.2 International 10-20 System

To ensure consistency and reproducibility in EEG acquisition, standardized electrode placement systems have been developed to define the spatial position of scalp electrodes. Among them, the International 10–20 System is one of the most widely adopted standards in both clinical and research settings, providing a systematic method for electrode positioning based on anatomical landmarks of the skull.

The 10–20 system determines electrode locations according to proportional distances between specific cranial reference points. The terminology 10–20 refers to the fact that adjacent electrodes are placed at distances corresponding to either 10% or 20% of the total skull measurements, ensuring a standardized and approximately uniform spatial coverage of cortical areas.

Electrodes are conventionally identified through a combination of letters and numbers. Letters indicate the cortical region associated with electrode placement, while numbers identify the hemisphere location. In particular:

- Fp: frontopolar region
- F: frontal region
- C: central region
- T: temporal region
- P: parietal region
- O: occipital region

Odd numbers are generally associated with the left hemisphere, even numbers with the right hemisphere, while the suffix z (zero) indicates electrodes positioned along the brain midline.

The International 10–20 System enables reproducible EEG recordings across subjects and acquisition sessions, facilitating comparisons among studies and ensuring consistency in both clinical interpretation and computational analyses.

Figure 2.1 shows an example of electrode positioning according to the International 10–20 System.

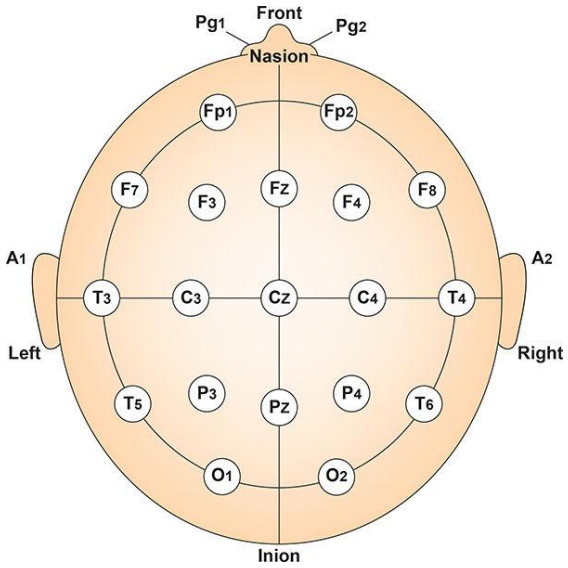

Figure 2.1. Representation of the International 10–20 electrode placement system commonly adopted in EEG acquisition

2.1.3 Artifacts and Noise

Despite its effectiveness in capturing brain electrical activity, electroencephalographic recordings are highly susceptible to various forms of noise and signal contamination that may significantly affect data quality and interpretability. Since EEG signals are characterized by relatively low amplitudes, typically measured in microvolts, even minor disturbances may interfere with the recorded neural activity.

Artifacts in EEG recordings may originate from both physiological and non-physiological sources. Physiological artifacts arise from biological processes unrelated to brain electrical activity, whereas non-physiological artifacts are typically associated with external environmental factors or acquisition-related issues.

Physiological disturbances may include cardiac activity (electrocardiographic artifacts), respiratory movements, sweating and head movements, all of which may alter signal stability and recording quality [7].

Non-physiological artifacts are commonly associated with power-line interference, typically occurring at frequencies of 50 Hz or 60 Hz depending on the electrical infrastructure, poor electrode contact and movement of recording equipment. Environmental electromagnetic interference may also introduce unwanted noise into EEG signals, particularly in non-controlled acquisition settings.

The presence of artifacts represents a major challenge in EEG analysis, as noise components may overlap with meaningful neural activity and potentially bias feature extraction and classification procedures. Consequently, preprocessing operations aimed at reducing signal contamination, such as filtering, referencing, segmentation and artifact rejection techniques, are commonly employed prior to computational analysis to improve signal quality and reliability.

2.1.4 EEG Preprocessing Techniques

Preprocessing represents a fundamental step in EEG analysis pipelines [17]. The primary objective of preprocessing is to improve signal quality, reduce unwanted interference and facilitate the extraction of meaningful information useful for subsequent computational analysis.

EEG preprocessing procedures generally aim to mitigate the impact of physiological and environmental noise while preserving relevant neural information. Depending on the acquisition protocol, application domain and analytical objectives, preprocessing pipelines may involve several operations, including signal referencing, filtering, resampling, segmentation, normalization and artifact rejection.

One commonly adopted preprocessing operation is re-referencing, which consists of redefining the electrical reference used for EEG signal acquisition. Since EEG measures voltage differences between electrodes, the choice of reference may substantially influence signal characteristics and its analyses. Common referencing strategies are linked-ear references, mastoid references and average referencing. The latter is widely employed to reduce spatial bias and improve signal interpretability.

Another essential preprocessing step is frequency filtering, which is commonly used to remove unwanted frequency components and improve signal quality. Since EEG signals typically occupy specific frequency ranges, filtering techniques may be employed to suppress slow drifts, high-frequency noise, muscular interference and electrical contamination. In particular, band-pass filters are frequently adopted to preserve neural activity within physiologically relevant frequency ranges while attenuating irrelevant signal components.

Resampling procedures are also applied to standardize acquisition frequencies

and reduce computational complexity. Since EEG recordings may originate from heterogeneous acquisition systems characterized by different sampling rates, resampling allows signals to be represented using a consistent temporal resolution, facilitating comparative analysis and machine learning applications.

Another important preprocessing operation concerns signal segmentation, often referred to as epoching. Instead of analyzing continuous EEG recordings directly, signals may be divided into shorter temporal windows or epochs, enabling the investigation of localized neural dynamics and improving the extraction of stable signal descriptors. Epoch-based analysis is particularly relevant in machine learning contexts, where fixed-length segments facilitate feature extraction and dataset standardization.

Finally, artifact rejection and quality control procedures are commonly employed to identify and remove corrupted signal segments. Threshold-based rejection, statistical criteria or dedicated artifact removal techniques may be applied to reduce the influence of noisy epochs.

Overall, EEG preprocessing plays a crucial role in ensuring signal consistency and reliability, constituting a necessary step before performing feature extraction and machine learning-based classification tasks.

2.1.5 EEG Feature Extraction

Directly employing raw EEG recordings for computational analysis may be challenging due to signal redundancy, noise susceptibility and the large volume of temporal information. For these reasons, feature extraction plays a central role in EEG-based machine learning pipelines, enabling the transformation of raw signals into compact and informative representations suitable for classification tasks [18].

The objective of feature extraction is to identify meaningful signal characteristics capable of capturing relevant neural information while reducing dimensionality and improving model interpretability. In EEG analysis, extracted features generally aim to describe temporal, spectral, statistical or connectivity-related properties of neural activity.

Time-domain features are directly computed from the temporal evolution of EEG signals and provide information about signal amplitude and waveform characteristics. Commonly adopted descriptors include statistical measures such as signal mean, variance and zero-crossing rate.

Although time-domain features may provide useful information regarding signal morphology and variability, their discriminative capability may be limited in scenarios where disease-related information is more strongly associated with alterations in oscillatory brain activity distributed across specific frequency bands.

Since EEG signals are characterized by oscillatory activity occurring at different frequency ranges, frequency-domain analysis represents one of the most widely

adopted approaches for EEG characterization. Spectral analysis enables the investigation of signal energy distributions across different EEG rhythms, facilitating the identification of alterations potentially associated with neurological disorders.

Among spectral analysis techniques, the Power Spectral Density (PSD) is commonly employed to quantify the distribution of signal power across frequency components. PSD estimation methods, such as the Welch method (the standard for the EEG analysis), are frequently adopted for their robustness and capability to provide stable spectral estimates while mitigating signal variability.

Frequency-domain features are often extracted from canonical EEG frequency bands, including delta, theta, alpha, beta, and gamma rhythms, whose functional and physiological significance was previously discussed in Section 1.2.3.

Rather than directly analyzing complete spectral representations, statistical measures are frequently employed to summarize EEG characteristics in a compact and interpretable manner

Common statistical descriptors include the mean, representing average signal behavior, the variance, quantifying signal variability, skewness and kurtosis, which describe asymmetry and distribution shape, respectively. These measures may provide complementary information regarding the statistical properties of EEG activity and help capture subtle differences between healthy and pathological neural dynamics.

In addition to absolute spectral measures, spectral ratios derived from combinations of EEG frequency bands have also attracted attention as potential indicators of neurological alterations. These ratios aim to quantify relative changes between slow and fast brain rhythms, potentially enhancing the characterization of abnormal neural activity.

Overall, EEG feature extraction represents a crucial stage in computational analysis pipelines, enabling the transformation of complex neural signals into structured representations suitable for machine learning algorithms and automated classification systems.

2.2 Machine Learning for EEG Analysis

Machine learning techniques have become increasingly relevant in biomedical signal analysis due to their capability to automatically identify complex patterns and relationships within high-dimensional data. In the context of electroencephalographic analysis, machine learning algorithms have been widely employed for tasks such as neurological disease classification.

Machine learning approaches are particularly valuable for extracting discriminative information from EEG-derived features and supporting automated classification procedures.

This section introduces the main machine learning methodologies relevant to this study, Support Vector Machines, multiclass classification strategies and data augmentation techniques employed in imbalanced EEG classification scenarios.

2.2.1 Support Vector Machine

Support Vector Machines (SVMs) are supervised machine learning algorithms widely adopted for classification and pattern recognition tasks. Due to their effectiveness in high-dimensional spaces and their capability to generalize well even in relatively small datasets, SVMs have become particularly popular in biomedical signal analysis and EEG-based classification problems [16].

The fundamental objective of an SVM classifier is to identify an optimal decision boundary, commonly called hyperplane, capable of separating samples belonging to different classes [19]. In binary classification problems, the algorithm aims to maximize the distance between the separating hyperplane and the closest training samples, known as support vectors. Maximizing this margin improves model generalization and robustness to unseen data.

In its simplest formulation, an SVM performs linear classification. However, many real-world datasets are not linearly separable. To address this limitation, SVMs can employ the so-called kernel trick, which enables the projection of input data into higher-dimensional feature spaces where linear separation becomes more feasible. Several kernel functions have been proposed, including linear, polynomial, sigmoid and Radial Basis Function (RBF) kernels.

Although SVMs were originally designed for binary classification, several strategies have been developed to extend them to multiclass problems. Common approaches include one-vs-one and one-vs-rest classification schemes, where multiple binary classifiers are combined to distinguish among several classes.

Due to their robustness, interpretability and effectiveness in small-to-medium sized datasets, SVMs remain widely employed in EEG-based neurological disorder classification and constitute the classification framework adopted in this work.

2.2.2 Multiclass Classification Strategies

Compared to binary classification, multiclass problems introduce additional complexity due to the presence of multiple decision boundaries, increased class overlap and potentially heterogeneous disease-related EEG patterns. In neurological applications, this challenge may become particularly relevant when different disorders exhibit partially overlapping electrophysiological characteristics or varying levels of inter-subject variability.

Since many machine learning algorithms, including Support Vector Machines, were originally designed for binary tasks, several strategies have been proposed to extend classification capabilities to multiclass settings.

One widely adopted approach is the One-vs-Rest (OvR) strategy, where a separate binary classifier is trained for each class against all remaining categories. During prediction, the final class assignment is typically determined according to the classifier providing the highest confidence score. Alternatively, the One-vs-One (OvO) strategy constructs classifiers for every possible pair of classes, with final predictions commonly obtained through voting procedures among binary decisions.

Other strategies have been proposed to address increasing classification complexity. Among them, hierarchical classification frameworks decompose multiclass problems into sequential decision-making stages, where simpler classification tasks are solved progressively. For example, an initial classifier may first separate broad categories before applying finer discrimination among subclasses in subsequent stages. Such approaches may improve interpretability and potentially simplify decision boundaries when dealing with heterogeneous classes.

Another strategy involves group-based classification, where clinically or biologically related conditions are initially merged into same categories before performing more detail discrimination. In neurological disease analysis, grouping disorders characterized by partially shared mechanisms or electrophysiological similarities may facilitate intermediate classification stages and potentially improve model performance.

The selection of an appropriate multiclass strategy depends on several factors, including dataset characteristics, class balance, disease heterogeneity and the complexity of underlying decision boundaries.

2.2.3 Data Augmentation Techniques

One of the major challenges in biomedical machine learning concerns the limited availability and imbalance of labeled datasets. In EEG-based neurological studies, collecting large and balanced datasets may be particularly difficult due to practical, economic and clinical constraints associated with patient recruitment, acquisition protocols,and disease prevalence. As a consequence, machine learning

models trained on imbalanced datasets may suffer from poor generalization and biased predictions toward majority classes.

Class imbalance represents a particularly relevant issue in classification tasks, as models may become disproportionately influenced by classes containing a larger number of training samples. This phenomenon may lead to reduced classification performance for less presented categories and compromise the reliability of disease discrimination systems.

To address these limitations, data augmentation techniques are commonly employed to artificially improve dataset balance and increase sample diversity. In structured biomedical datasets, augmentation strategies typically aim to generate additional synthetic samples while preserving the statistical properties of the original data distribution.

One of the data Augmentation strategies for addressing class imbalance problems is the Synthetic Minority Oversampling Technique (SMOTE). Unlike conventional random oversampling methods, which duplicate existing minority samples, SMOTE generates new synthetic instances through interpolation between neighboring samples belonging to the minority class. Synthetic samples are created by selecting neighboring observations in feature space and generating intermediate points along the line connecting them [20]. Despite its effectiveness, SMOTE may present limitations when applied to highly complex or heterogeneous datasets, as generated samples are constrained to linear interpolations among existing observations and may not fully capture intricate feature relationships.

Another strategies that can be adopted is the use of Generative Adversarial Networks (GANs). A GAN architecture generally consists of two competing neural networks: a generator, which produces synthetic samples, and a discriminator, which attempts to distinguish generated data from real observations. Through adversarial training, the generator progressively improves its ability to produce realistic samples that resemble the original dataset.

Among GAN-based approaches for tabular data generation, the Conditional Tabular Generative Adversarial Network (CTGAN) has emerged as a promising framework for handling structured datasets [21]. Unlike traditional GAN models originally designed for image generation, CTGAN is specifically tailored for tabular data.

In EEG-related machine learning applications, synthetic data generation techniques such as SMOTE and CTGAN may contribute to mitigating class imbalance, increasing training diversity and improving classifier robustness, particularly in scenarios characterized by limited sample availability.

2.3 Related Work

In recent years, several studies have explored the integration of EEG-derived features with machine learning techniques for neurological disease discrimination. In particular, Support Vector Machines have emerged as one of the most widely adopted classifiers due to their robustness in high-dimensional feature spaces and their effectiveness in biomedical applications characterized by relatively limited sample sizes.

Regarding Alzheimer's disease, several studies have demonstrated the effectiveness of EEG-based machine learning frameworks for distinguishing pathological subjects from healthy controls. Akbar et al. [22] employed EEG recordings from the ds004504 dataset and proposed an SVM-based framework for the classification of neurodegenerative conditions. In addition, representative studies selected from the systematic review conducted by Uyanik et al. [16] confirm the relevance of EEG-driven machine learning approaches for Alzheimer's disease classification. For example, Perez-Valero et al. [23] investigated EEG-based Alzheimer's disease discrimination using an SVM classifier, although the study involved a relatively limited number of subjects. Similarly, Chu et al. [24] explored EEG-based Alzheimer's disease detection in a larger population, further supporting the applicability of SVM-based approaches in this domain.

In the context of Epilepsy, EEG has long represented a primary diagnostic tool. Tran et al. [25] proposed an SVM-based classification framework for epilepsy discrimination, although the study was conducted on a very small cohort. Moreover, studies selected from the review presented in [16] highlight the increasing adoption of advanced machine learning and deep learning models for epilepsy classification. In particular, Lih et al. [26] investigated transformer-based architectures for EEG classification, while Liu et al. [27] employed convolutional neural networks for EEG-based classification.

Compared to Alzheimer's disease and epilepsy, ALS-related EEG classification remains considerably less explored in the literature. Ngo et al. [28] investigated EEG-based ALS analysis through an SVM framework using the EEGET-ALS dataset, highlighting the potential of EEG-driven machine learning techniques for supporting neurological assessment in ALS patients. However, ALS datasets are often characterized by a very limited number of pathological subjects, making class imbalance a particularly relevant issue.

Table 2.1 summarizes representative EEG-based classification studies selected from the literature. The included works were chosen according to their relevance to the neurological disorders investigated in this thesis, methodological similarity to the proposed framework, and adoption of machine learning techniques suitable for EEG analysis. Rather than providing a direct performance comparison, the table emphasizes methodological aspects such as the number of subjects, EEG

channels, adopted classifier, validation strategy and main relevance or limitation of each study.

Study	**Subjects**	**Ch.**	**Classifier**	**Validation**	**Relevance Limitations**
Akbar et al. (2026) [22]	36 AD/ 29 HC/ 23 FTD	19	SVM	Nested CV	Multiclass setting, but limited to dementia-related conditions.
Perez-Valero et al. (2023) [23]	8 AD/ 8 HC	16	SVM	LOSO CV	Subject-level AD classification on a very small dataset.
Chu et al. (2023) [24]	112 AD/ 93 HC	19	SVM	Cross-validation	Larger AD dataset, but limited to binary classification.
Tran et al. (2022) [25]	5 EP/ 5 HC	100	SVM	Cross-validation	High-density EEG epilepsy study with very few subjects.
Lih et al. (2023) [26]	71 EP/ 50 HC	35	Transformer	10-fold CV	Deep learning-based epilepsy classification in a binary setting.
Liu et al. (2024) [27]	24	23	Pseudo-3D CNN	Cross-validation	CNN-based EEG classification on a small dataset.
Ngo et al. (2024) [28]	6 ALS/ 170 HC	32	SVM	Cross-validation	ALS binary classification with strong class imbalance.

Table 2.1. Summary of representative EEG-based machine learning studies for the classification of Alzheimer's disease, Epilepsy and Amyotrophic Lateral Sclerosis. The table focuses on methodological aspects and study limitations. Abbreviations: AD, Alzheimer's disease; EP, Epilepsy; ALS, Amyotrophic Lateral Sclerosis; FTD, Frontotemporal Dementia; HC, Healthy Controls; Ch., channels.

The methodological characteristics summarized in Table 2.1 are essential for interpreting the current state of the literature. Differences in datasets size, EEG montage, validation protocol, and classification task can substantially influence the reliability and generalizability of reported results. In particular, several studies are based on relatively small number of subjects, single-disease settings or binary classification tasks. Moreover, some works rely on homogeneous datasets acquired under specific experimental conditions, which may limit the applicability of the resulting models to more heterogeneous real-world scenarios.

Overall, these studies confirm the potential of EEG-based machine learning for neurological disease classification, while also showing that differences in datasets size, validation protocol, task formulation and dataset heterogeneity must be carefully considered when interpreting reported results.

2.4 Research Gap

The literature reviewed in this chapter shows the growing relevance of electroencephalography combined with machine learning techniques for the automated characterization of neurological disorders. Existing studies demonstrate that EEG-derived features can provide discriminative information for conditions such as Alzheimer's disease, Epilepsy and Amyotrophic Lateral Sclerosis.

However, the current literature remains fragmented. Most studies address individual neurological disorders separately, typically formulating the problem as a binary classification task between healthy controls and one pathological group. Although this approach is useful for evaluating disease-specific EEG alterations, it does not fully reflect more complex diagnostic scenarios, where different neurological conditions may need to be distinguished within the same computational framework.

A further limitation concerns the limited heterogeneity of the data used in many previous works. Several studies rely on small datasets or recordings acquired under homogeneous experimental protocols, which may reduce the robustness and generalizability of the resulting models when applied to data collected from different sources, acquisition settings or patient populations.

Moreover, multiclass EEG-based neurological disease classification remains less explored than binary classification. In particular, few studies have investigated unified frameworks capable of integrating heterogeneous EEG datasets and simultaneously discriminating among multiple neurological conditions.

These gaps motivate the present work, which investigates EEG-based neurological disease classification through a unified framework involving multiple publicly available datasets acquired under heterogeneous experimental conditions. Both binary and multiclass scenarios are considered, together with hierarchical classification strategies, feature selection, hyperparameter optimization and data augmentation techniques, with the aim of evaluating their effectiveness in increasingly realistic diagnostic settings.

2.5 Thesis Contributions

Based on the research gaps identified in the previous section, the main contributions of this work can be summarized as follows:

- Investigation of EEG-based neurological disease classification, focusing on the discrimination between healthy subjects and patients affected by Alzheimer's Disease, Epilepsy and Amyotrophic Lateral Sclerosis through machine learning techniques.
- Development and evaluation of binary classification scenarios, aimed at separately distinguishing healthy subjects from each neurological condition using Support Vector Machine models.
- Exploration of multiple multiclass classification strategies, including a standard multiclass SVM approach and hierarchical classification frameworks designed to address the complexity of distinguishing multiple neurological disorders.
- Implementation of hierarchical classification architectures, including a two-level approach separating healthy and pathological subjects before multiclass discrimination among diseased individuals
- Assessment of data augmentation methodologies, through the integration of SMOTE and CTGAN to mitigate data imbalance and improve model robustness.
- Comparative analysis of classification paradigms, aimed at evaluating their effectiveness, strengths and limitations in EEG-based neurological disease discrimination.

In general, this thesis seeks to contribute to understanding how machine learning techniques applied to EEG signals may support the characterization and classification of neurological disorders within increasingly complex diagnostic scenarios.

Chapter 3

Materials and Methods

3.1 Overview of the Proposed Framework

The objective of this study is to investigate the feasibility of EEG-based machine learning techniques for the automated discrimination of neurological disorders, with particular focus on Alzheimer's disease, Epilepsy and Amyotrophic Lateral Sclerosis.

To achieve this objective, a unified computational framework was developed and applied to multiple publicly available EEG datasets acquired under different experimental conditions and distributed in different file formats. Despite these differences, all recordings were processed through a common analysis pipeline designed to have consistency across datasets and facilitate comparative evaluation.

The proposed framework consists of four main stages: data acquisition, signal preprocessing, feature extraction and classification. First, EEG recordings from the selected datasets were imported and standardized. Subsequently, preprocessing operations including re-referencing, filtering, resampling, epoch segmentation and artifact rejection were applied to improve signal quality and ensure uniformity among recordings.

After preprocessing, a feature extraction procedure based on spectral and statistical descriptors was employed to transform raw EEG signals into compact numerical representations suitable for machine learning analysis. Features were extracted from the canonical EEG frequency bands and included both statistical measures and spectral ratios commonly associated with neurological alterations.

The resulting feature tables were then used to train and evaluate several classification models. Initially, binary classification experiments were conducted to discriminate healthy subjects from each neurological condition individually. Subsequently, different multiclass classification strategies were investigated, including

direct multiclass classification, hierarchical classification frameworks and disease-grouping approaches. In addition, the impact of data augmentation techniques, namely SMOTE and CTGAN, was evaluated to address potential class imbalance issues.

Figure 3.1 shows the workflow of the proposed methodology.

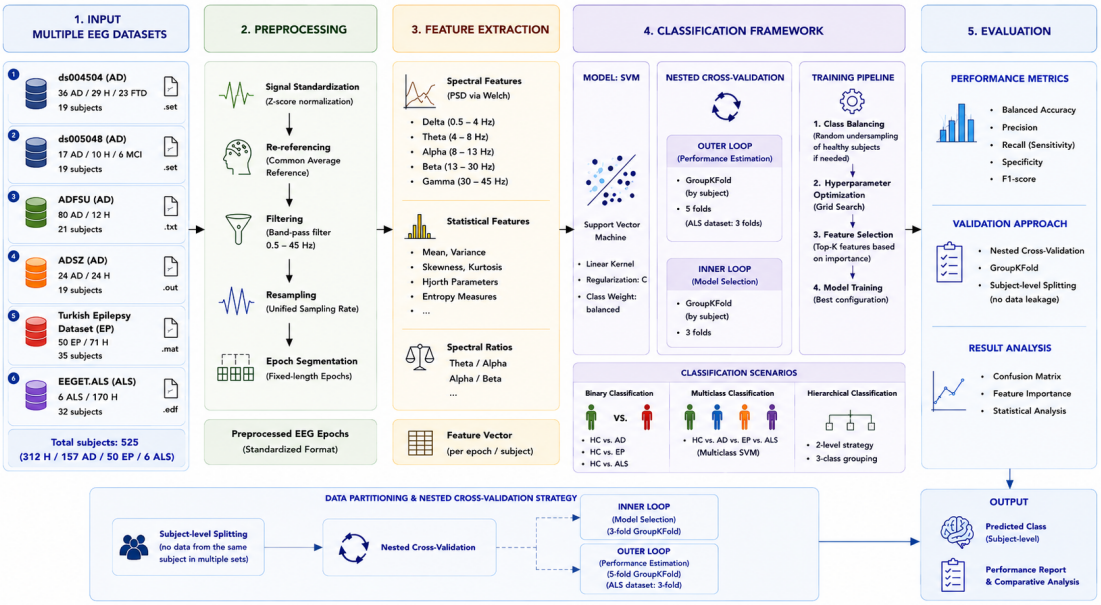

Figure 3.1. Workflow of the proposed methodology

3.2 EEG Datasets

The experimental evaluation conducted in this work is based on multiple publicly available EEG datasets covering three neurological conditions: Alzheimer's disease, Epilepsy and Amyotrophic Lateral Sclerosis. The adoption of multiple datasets was motivated by the need to evaluate the robustness and generalizability of the proposed framework across different patient populations, acquisition protocols and recording environments.

The selected datasets were obtained from publicly accessible repositories and distributed in different file formats, including EEGLAB (.set), European Data Format (.edf), MATLAB (.mat), text-based (.txt) and output (.out) files. Despite these differences, all recordings were processed through the same preprocessing and feature extraction pipeline, ensuring methodological consistency throughout the study.

An additional advantage of this approach is the inclusion of EEG recordings acquired under different experimental conditions and using different acquisition systems. Consequently, the resulting classification models are evaluated on data characterized by more variability than studies based on a single dataset, providing a more realistic assessment of model robustness.

Table 3.1 summarizes the datasets used in this work.

Dataset	**Disease**	**Subjects**	**Channels**	**Format**
ds004504	AD	36 AD / 29 HC / 23 FTD	19	.set
ds005048	AD	17 AD / 10 HC / 6 MCI	19	.set
ADFSU	AD	80 AD / 12 HC	21	.txt
ADSZ	AD	24 AD / 24 HC	19	.out
Turkish Epilepsy Dataset	EP	50 EP / 71 HC	35	.mat
EEGET-ALS	ALS	6 ALS / 170 HC	32	.edf

Table 3.1. Datasets used in this work. Abbreviations: AD: Alzheimer's disease; EP: Epilepsy; ALS: Amyotrophic Lateral Sclerosis; FTD: Frontotemporal Dementia; MCI: Mild Cognitive Impairment; HC: Healthy Controls.

3.2.1 Alzheimer's Disease Datasets

To investigate EEG-based Alzheimer's disease classification, four publicly available datasets were considered. The use of multiple datasets was intended to increase the variability of the analyzed population, reduce the dependence of the classification models on a specific acquisition protocol or recording environment and increase the number of subjects.

ds004504

The ds004504 dataset was obtained from the OpenNeuro repository [29] and was originally introduced by Miltiadous et al. The dataset contains resting-state EEG recordings collected from subjects diagnosed with Alzheimer's disease, Frontotemporal Dementia (FTD) and Healthy Controls (HC).

Recordings were acquired using the international 10–20 electrode placement system and include 19 EEG channels. For the purposes of this work, only the Alzheimer's disease and healthy control groups were considered, while subjects diagnosed with frontotemporal dementia were excluded from the classification experiments.

ds005048

The ds005048 was collected by Lahijanian et al. and is publicly available through the OpenNeuro repository [30]. This database includes EEG recordings acquired during an auditory stimulation paradigm designed to investigate cognitive responses associated with Alzheimer's disease.

The dataset comprises recordings from Alzheimer's disease patients, Healthy Controls and subjects diagnosed with Mild Cognitive Impairment (MCI). In this work, only AD patients and HC subjects were considered for classification purposes. All EEG data were recorded using 19 monopolar channels based on the standard 10/20 system.

The inclusion of this dataset introduces additional variability into the experimental framework by combining resting-state and task-related EEG recordings, enabling the assessment of model robustness under different acquisition conditions.

ADFSU

The ADFSU dataset was introduced by Vicchietti et al. [31] and is publicly available through the Open Science Framework repository. The dataset contains EEG recordings acquired from 92 subjects, including 80 patients diagnosed with AD and 12 HC.

EEG recordings were collected according to the international 10–20 electrode placement system from 21 channels. Since both eyes-open and eyes-closed recordings are available, only the eyes-closed condition was considered in this work. This choice was motivated by the fact that resting-state eyes-closed EEG is one of the most commonly adopted paradigms in Alzheimer's disease research and facilitates comparison with other public datasets.

ADSZ

The ADSZ dataset was originally introduced by Pineda et al. and later made publicly available by Alves et al. through the Figshare repository [32]. The dataset contains EEG recordings collected from 48 subjects, equally divided between Alzheimer's disease patients and Healthy Controls.

The dataset was originally created to support the investigation of electrophysiological alterations associated with both Alzheimer's disease and schizophrenia. However, only the Alzheimer's disease and healthy control groups were considered in the present study.

3.2.2 Epilepsy Dataset

The epilepsy classification experiments were conducted using the Turkish Epilepsy Dataset introduced by Tasci et al. [33]. The dataset contains EEG recordings collected from 121 subjects, including 71 Epilepsy patients and 50 Healthy Controls. Recordings were acquired using 35 EEG channels and represent one of the largest publicly available datasets specifically designed for EEG-based epilepsy detection. Unlike the other datasets considered in this study, the Turkish Epilepsy Dataset was originally provided as MATLAB files containing recordings from multiple sensor modalities. Therefore, an additional conversion step was required before EEG preprocessing. Only the 19 channels corresponding to the standard EEG montage were retained, while the remaining signals were excluded from the analysis. This choice was made to ensure compatibility with the other datasets adopted in this work, which predominantly followed the international 10–20 electrode placement system. The resulting EEG recordings were subsequently processed using the same preprocessing pipeline applied to all datasets.

Compared to commonly used epilepsy datasets, the Turkish Epilepsy Dataset provides a larger and more heterogeneous population, allowing a more realistic evaluation of machine learning models and reducing the risk of overly optimistic performance estimates. For these reasons, it was selected as the benchmark dataset for epilepsy classification in this study.

3.2.3 ALS Dataset

The EEGET-ALS dataset [28] was adopted for the investigation of amyotrophic lateral sclerosis classification. The dataset was recently released as a publicly available resource specifically designed to support EEG-based studies on ALS patients.

The dataset contains EEG recordings acquired from 6 ALS patients and 170 HC subjects using a 32-channel acquisition system. EEG signals were recorded according to a standardized experimental protocol and distributed in European Data Format (EDF), facilitating interoperability across different EEG processing platforms.

Although the number of ALS patients is relatively limited, the EEGET-ALS dataset represents one of the few publicly available EEG resources specifically dedicated to ALS research.

3.3 EEG Preprocessing

EEG recordings obtained from the selected datasets were acquired using different hardware configurations, experimental protocols, sampling frequencies and file formats. Consequently, a standardized preprocessing pipeline was applied to all recordings in order to ensure consistency between datasets and improve the quality of the extracted features.

The preprocessing procedure was implemented using the Python-based MNE framework and consisted of signal standardization, re-referencing, band-pass filtering, resampling, epoch segmentation and artifact rejection. The same preprocessing steps were applied to all datasets regardless of the original file format, enabling the construction of a unified analysis framework.

The preprocessing parameters adopted in this work (band-pass filter between 0.5 and 45 Hz, resampling frequency of 250 Hz, epoch duration of 4 s, and artifact rejection threshold of 250 µV) were selected based on a combination of literature evidence and preliminary experimental optimization. Multiple configurations were evaluated through exploratory analyses and hyperparameter tuning procedures, and the final values corresponded to the configuration that produced the most stable and accurate classification performance.

3.3.1 Signal Standardization

The EEG datasets adopted in this work were distributed in heterogeneous file formats, including EEGLAB (.set), European Data Format (.edf), MATLAB (.mat), text-based (.txt), and output (.out) files.

Regardless of their original format, all recordings were imported into a common in-memory representation using MNE-Python. Signal amplitudes were converted to microvolts (μV) whenever necessary while channel labels were harmonized according to the International 10–20 nomenclature. This step ensured consistent numerical ranges across recordings and facilitated the application of a unified preprocessing and feature extraction pipeline. Although the number of available electrodes varied across datasets, equivalent scalp locations were assigned consistent channel names. Following filtering, all recordings were resampled to a common sampling frequency of 250 Hz.

After preprocessing and feature extraction, the resulting subject-level feature vectors were stored in comma-separated value (.csv) files. Each row corresponded to one subject and contained the extracted EEG descriptors together with the subject identifier and the corresponding class label.

3.3.2 Re-referencing

After signal standardization, EEG recordings were re-referenced using the common average reference technique. In this approach, the average signal across all available electrodes is computed and subtracted from each channel.

Common average referencing is widely adopted in EEG analysis because it reduces the influence of the original reference electrode and attenuates spatially common noise components. Moreover, it provides a more balanced representation of brain activity and improves the comparability of recordings acquired using different hardware configurations.

The average reference was applied uniformly to all EEG recordings prior to subsequent preprocessing operations.

3.3.3 Filtering

To improve the quality of the EEG signals and remove unwanted noise, a zero-phase finite impulse response (FIR) band-pass filter was applied between 0.5 Hz and 45 Hz.

The filter length was automatically determined by MNE-Python according to the original sampling frequency and the transition bandwidth. Since filtering was performed before resampling and the datasets had different original sampling frequencies, the resulting FIR order varied across datasets. Specifically, filter orders of 844, 1650, 1690, and 3300 were obtained for sampling frequencies of 128 Hz, 250 Hz, 256 Hz and 500 Hz, respectively.

The lower limit of 0.5 Hz was chosen to eliminate slow signal drifts and baseline shifts that often occur during EEG recording. The upper limit of 45 Hz was selected to reduce high-frequency noise and muscle-related artifacts, while still preserving the brain rhythms most relevant for neurological analysis.

This frequency range includes all major EEG bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and low-gamma (30–45 Hz). These bands are widely used as key features for characterizing neurological disorders such as Alzheimer's disease.

3.3.4 Resampling

Since the adopted datasets were originally acquired at different sampling frequencies, all recordings were resampled to a common frequency of 250 Hz.

This step was performed to ensure consistency across datasets and simplify subsequent signal processing operations. The selected sampling frequency preserves the spectral content of interest within the analyzed frequency range while reducing computational complexity and memory requirements.

By resampling all recordings to a common frequency, feature extraction procedures could be applied uniformly across datasets without introducing dataset-specific processing configurations.

3.3.5 Epoch Segmentation

After preprocessing, continuous EEG recordings were segmented into fixed-length temporal windows (epochs) to facilitate feature extraction and machine learning analysis.

Epoching is a widely adopted procedure in EEG processing, as it transforms long continuous recordings into shorter and more homogeneous segments while preserving the temporal dynamics of brain activity and increases the number of available observations.

In the proposed framework, non-overlapping epochs of 4 seconds were generated from each recording. Given the standardized sampling frequency of 250 Hz, each epoch contained 1000 samples per channel.

Non-overlapping segmentation was adopted to avoid generating highly correlated samples from adjacent windows, reducing redundancy while preserving subject-level independence throughout the subsequent classification pipeline.

The choice of the epoch length was supported by preliminary experiments performed using different epoch durations (e.g., 3 s, 4 s, and 6 s). Across the considered neurological disorders, 4-s epochs consistently provided the highest Balanced Accuracy while preserving a sufficient number of training samples. For example, in the first version of the framework, on the ds004504 Alzheimer's disease dataset, the Balanced Accuracy increased from 75.1% using 3-s epochs to 84.6% using 4-s epochs. Similarly, for the ALS dataset, 4-s epochs achieved a Balanced Accuracy of 85.2%, compared with 78.6% obtained using 6-s epochs. Since this epoch duration also provided the best compromise across the heterogeneous datasets included in this work, 4-s non-overlapping epochs were adopted throughout all experiments.

Recordings whose duration was shorter than the selected epoch length were excluded from further analysis.

3.3.6 Artifact Rejection

Following epoch segmentation, an artifact rejection procedure was applied to identify and remove EEG segments potentially contaminated by non-neural activity. EEG recordings are often affected by various sources of interference, including eye blinks, eye movements, muscle contractions, electrode displacement, and environmental noise. These artifacts can alter signal characteristics and negatively impact the reliability of subsequent feature extraction and classification stages.

To mitigate this issue, an automatic artifact rejection strategy was adopted. Specifically, epochs were discarded whenever the absolute EEG amplitude exceeded a predefined threshold of 250 µV in any channel. This procedure was implemented using the rejection mechanism provided by the MNE framework.

The selection of the rejection threshold was guided by both literature evidence and preliminary experimental analyses. The final value of 250 µV was selected as it provided the best compromise between artifact removal and classification performance across the heterogeneous datasets considered in this work.

Only epochs satisfying the quality criteria were maintained for subsequent feature extraction and machine learning analysis.

3.4 Feature Extraction

Following preprocessing, each EEG recording was transformed into a numerical table for machine learning analysis. Since raw EEG signals consist of high-dimensional temporal sequences, direct classification is often challenging and computationally expensive. Therefore, a feature extraction procedure was adopted to summarize the most relevant spectral characteristics of the recordings.

The proposed framework is based on power spectral analysis combined with statistical descriptors computed over the canonical EEG frequency bands. In addition, several spectral ratios commonly associated with neurological alterations were extracted. The resulting feature vectors were subsequently used as input for the classification models described in the following sections.

For each preprocessed epoch, spectral information was extracted through Power Spectral Density estimation. PSD analysis represents one of the most widely adopted approaches in EEG research because it provides a compact representation of the distribution of signal power across different frequency components [34].

In this work, PSD was estimated independently for each EEG channel using Welch's method as implemented in MNE-Python. Each 4-s epoch contained 1000 samples at the standardized sampling frequency of 250 Hz and was analyzed as a single 1000-sample segment. Before spectral estimation, the temporal mean of each segment was removed and PSD values between 0.5 Hz and 45 Hz were retained. Compared to a standard Fourier transform, Welch's approach reduces variance in the spectral estimate and provides more stable power measurements, making it particularly suitable for biomedical signals characterized by high variability [35].

The PSD was computed within the frequency range between 0.5 Hz and 45 Hz, corresponding to the bandwidth preserved during preprocessing. Spectral power values were calculated independently for each epoch and each EEG channel, generating a frequency-domain representation of the signal that served as the basis for subsequent feature extraction.

After PSD estimation, spectral power values were grouped according to the canonical EEG frequency bands (Table 1.1).

For each channel and each epoch, the average spectral power within every frequency band was computed.

For each EEG frequency band and each recording channel, a set of statistical descriptors was computed from the corresponding band-power values across all available epochs.

The extracted statistics were:

- Mean, representing the average spectral power;
- Variance, measuring the variability of spectral power across epochs;
- Skewness, quantifying the asymmetry of the power distribution;

- Kurtosis, describing the concentration of observations around the mean and the presence of extreme values.

Consequently, each EEG channel was represented by four statistical descriptors for every frequency band. Considering the five analyzed EEG bands (delta, theta, alpha, beta, and gamma), a total of twenty statistical features were extracted per channel.

These descriptors were selected because they capture complementary aspects of EEG dynamics, including average activity levels, temporal variability, and higher-order distributional characteristics that may differ between healthy subjects and neurological patients.

In addition to statistical descriptors, several spectral power ratios were extracted because they provide a representation of the relative balance between slow and fast brain oscillations.

The following ratios were computed for each EEG channel:

- Theta/Alpha
- Delta/Alpha
- Delta/Beta
- Theta/Beta
- Slow/Fast

where the Slow/Fast ratio was defined as: $\frac{\text{Delta+Theta}}{\text{Alpha+Beta}}$.

These ratios were included because numerous studies have reported alterations in the balance between slow-frequency and fast-frequency EEG activity in neurological disorders.

After feature extraction, all descriptors were organized into a structured tabular dataset suitable for machine learning analysis. Each row of the resulting feature matrix corresponded to a single subject, while columns represented the extracted EEG features.

In addition to the feature values, two metadata fields were included:

- subject_id, uniquely identifying each participant;
- label, indicating the class associated with the subject.

For binary classification experiments, labels were encoded as:

- 0 = pathological subject;
- 1 = healthy control.

The remaining columns contained the extracted EEG descriptors. Specifically, for each EEG channel, five frequency bands were analyzed and four statistical metrics were computed, resulting in:

$$N_{\text{channels}} \times 5 \times 4$$

statistical features.

Furthermore, five spectral ratios were extracted for each channel, yielding:

$$N_{\text{channels}} \times 5$$

additional features.

Consequently, for example, for recordings containing 19 EEG channels, the total number of extracted features was:

$$(19 \times 5 \times 4) + (19 \times 5) = 475$$

features per subject.

The resulting feature matrix, in which each row corresponded to a single subject, was used as input for all classification models investigated in this study. Each participant was therefore represented by a high-dimensional feature vector obtained by aggregating spectral information across all valid EEG epochs.

Since the number of extracted features was comparable to the number of available subjects, this representation could increase the risk of overfitting and the effects associated with the curse of dimensionality. To mitigate these issues, the feature-selection procedures described in Section 3.5.5 were applied within the classification framework, reducing the effective dimensionality of the input space and retaining only the most informative EEG descriptors.

3.5 Classification Framework

After feature extraction, the resulting feature matrix was used to train and evaluate machine learning models for neurological disease classification. The proposed classification framework was designed to maximize model robustness while minimizing the risk of overfitting, particularly given the limited number of available subjects and the relatively high dimensionality of the extracted feature space.

The classification procedure consisted of different stages, including class balancing, hyperparameter optimization, feature selection and performance evaluation. To ensure an unbiased assessment of model performance, all optimization procedures were embedded within a nested cross-validation framework.

3.5.1 Support Vector Machine Classifier

Support Vector Machines were selected as the primary classification model throughout this study. SVMs are widely used in biomedical machine learning applications due to their effectiveness in high-dimensional feature spaces and their ability to achieve good generalization performance even when the number of features exceeds the number of available samples.

Given a set of training samples, an SVM aims to identify the optimal decision boundary that maximizes the separation margin between classes. Through the use of kernel functions, SVMs can also model non-linear decision boundaries when the data are not linearly separable.

The choice of SVM was motivated by several factors. First, EEG-based classification problems often involve a relatively small number of subjects combined with a large number of extracted features. Second, numerous studies reviewed in Chapter 2 reported competitive performance using SVM-based approaches for neurological disease classification.

3.5.2 Nested Cross-Validation Strategy

To obtain an unbiased estimate of model performance and prevent information leakage during model selection, all classification experiments were conducted within a nested cross-validation framework.

The adopted strategy consisted of two validation levels. The outer loop was used to estimate the generalization performance of the classifier, whereas the inner loop was employed for model selection and hyperparameter optimization.

Since multiple EEG epochs originated from the same subject, preserving subject independence between training and test sets was considered essential. Therefore, GroupKFold cross-validation was adopted, using the subject identifier as grouping variable. This approach guarantees that all epochs belonging to a given subject

are assigned exclusively to either the training set or the test set within a fold, preventing data leakage across partitions.

For the majority of the experiments, the outer loop consisted of 5 folds, while the inner loop consisted of 3 folds. However, for the ALS dataset, which contains only six ALS patients, a 3-fold outer cross-validation scheme was adopted to ensure that pathological subjects were represented in each fold and to avoid excessively small test partitions.

Performance evaluation was primarily based on balanced accuracy, which provides a more reliable assessment than conventional accuracy in the presence of class imbalance.

3.5.3 Class Balancing Strategy

Some of the considered datasets exhibited substantial class imbalance between healthy controls and pathological subjects. Such imbalance may bias the learning process towards the majority class and lead to overly optimistic performance estimates.

To mitigate this issue, an optional class balancing procedure was introduced for binary classification experiments characterized by pronounced class asymmetry. The balancing strategy consisted of randomly selecting a subset of healthy subjects in order to obtain a more balanced class distribution while preserving subject independence.

The impact of class balancing was evaluated experimentally by comparing classification performance with and without balancing. This analysis allowed the assessment of whether reducing class imbalance could improve model generalization and classification fairness.

3.5.4 Feature Importance Analysis

To better understand which EEG features contributed most to the classification process, a feature importance analysis was performed using the coefficients of the linear SVM models.

After each nested cross-validation experiment, the coefficients of the best linear SVM model obtained in each outer fold were extracted. Since all features were standardized within the classification pipeline using StandardScaler before model training, the magnitudes of the SVM coefficients can be directly compared across features. Under this condition, features associated with larger absolute coefficient values have a greater influence on the decision boundary. Therefore, the absolute values of the coefficients were adopted as a measure of feature importance.

For each feature, the absolute coefficient values were averaged to obtain a global importance score. Features were subsequently ranked according to their global

importance scores, allowing the identification of the most discriminative EEG descriptors for each classification task.

This ranking served as the basis for the feature selection procedure described in the following section.

3.5.5 Feature Selection

Following feature importance analysis, a feature selection procedure was conducted to investigate whether classification performance could be improved using a reduced subset of highly discriminative features.

Features were first ranked according to their global importance scores computed in the previous section. Starting from this ordered list, several candidate subsets containing the Top-K ranked features were generated, where K ranged from small feature subsets to the complete feature space.

For each candidate subset, the entire nested cross-validation procedure was repeated and classification performance was assessed in terms of Balanced Accuracy. Rather than selecting an arbitrary number of features, different values of K were systematically investigated. The number of evaluated features depended on the specific classification task and on the total number of available descriptors, progressively increasing from small subsets to the complete feature set. The subset producing the highest Balanced Accuracy was selected as the optimal feature configuration for the considered classification task.

This strategy allowed the evaluation of the trade-off between model complexity and predictive performance, while also improving the interpretability of the resulting classifiers.

3.5.6 Hyperparameter Optimization

To further improve classifier performance, automated hyperparameter optimization was performed using the Optuna framework.

For each outer cross-validation fold, Optuna optimized the Support Vector Machine (SVM) hyperparameters within the inner GroupKFold validation loop. The optimization objective was the mean Balanced Accuracy computed across the inner validation folds.

The search space explored by Optuna included the following hyperparameters:

- the regularization parameter C, sampled on a logarithmic scale in the range $[10^{-3},10^{3}]$;
- the kernel type, selected between linear and rbf;

- the γ parameter for the RBF kernel, sampled on a logarithmic scale in the range $[10^{-4},10^{1}]$. When the linear kernel was selected, the default value scale was adopted.

For each outer fold, a total of 50 optimization trials were performed. Furthermore, Optuna's MedianPruner was employed to terminate unpromising trials early after an initial warm-up period, reducing computational cost while maintaining an efficient exploration of the search space.

Once the optimal hyperparameter configuration had been identified within the inner cross-validation loop, the corresponding SVM model was retrained using the complete outer training set and subsequently evaluated on the corresponding outer test set. This procedure ensured that the test data remained completely unseen during both model selection and hyperparameter optimization.

Hyperparameter optimization was conducted twice: first using the complete feature set and subsequently using the optimal feature subset identified through the feature selection procedure. This enabled the assessment of the combined impact of feature selection and hyperparameter optimization on classification performance.

The best hyperparameter configurations identified by Optuna for the different classification experiments are reported in Appendix A.

3.5.7 Evaluation Metrics

To assess the performance of the proposed classification framework, multiple evaluation metrics were considered throughout the experimental analysis. Since several of the adopted datasets exhibit class imbalance, relying exclusively on accuracy could provide misleading estimates of classifier performance. Consequently, Balanced Accuracy was adopted as the primary evaluation metric, while additional measures were used to provide complementary information regarding the classifier behavior.

Accuracy

Accuracy is the proportion of correct predictions (both positive and negative) to total predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, TN, FP and FN denote true positives, true negatives, false positives and false negatives, respectively. Although accuracy is one of the most commonly reported performance measures, it may produce optimistic results when class distributions are highly imbalanced.

Balanced Accuracy

Balanced Accuracy was selected as the primary evaluation metric throughout this work. Unlike conventional accuracy, Balanced Accuracy assigns equal importance to each class by averaging the sensitivity and specificity:

$$\text{Balanced Accuracy} = \frac{Sensitivity + Specificity}{2}$$

This metric involves unequal class distributions, as it avoids bias in favor of the majority class. Moreover, Balanced Accuracy was also adopted as the optimization objective during hyperparameter tuning with Optuna.

Sensitivity (Recall)

Sensitivity, also referred to as Recall or True Positive Rate, measures the proportion of correctly identified positive samples and is defined as

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

High sensitivity indicates that the classifier successfully identifies pathological subjects while minimizing false negatives.

Specificity

Specificity measures the proportion of correctly identified negative samples and is defined as

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

A high specificity value indicates that healthy subjects are correctly recognized, reducing the occurrence of false positive diagnoses.

Precision

Precision evaluates the reliability of positive predictions and is defined as

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

F1-score

The F1-score combines Precision and Recall into a single metric:

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

The F1-score is particularly useful when both false positives and false negatives are considered equally important.

ROC-AUC

The Receiver Operating Characteristic Area Under the Curve (ROC-AUC) is a threshold-independent evaluation metric used to assess the discriminative capability of a classifier. The ROC curve represents the relationship between the True Positive Rate, and the False Positive Rate at different classification thresholds.

The False Positive Rate is defined as:

$$\text{FPR} = \frac{FP}{FP + TN}$$

The ROC-AUC value summarizes the overall ability of the model to distinguish between positive and negative classes. A value close to 1 indicates excellent discriminative performance, whereas a value close to 0.5 corresponds to a classifier with performance comparable to random guessing.

Chapter 4

Experimental Results

4.1 Binary Classification

4.1.1 Alzheimer's Disease Classification

The first set of experiments focused on the binary discrimination between HC (74) and AD (157). The objective is to evaluate the effectiveness of the proposed classification pipeline and to quantify the contribution of feature selection and hyperparameter optimization to the overall classification performance.

As a baseline, the complete feature set (475 features) was used to train a linear Support Vector Machine within the Nested GroupKFold cross-validation framework described in Chapter 3. The baseline classifier achieved an Accuracy of 84.87%, a Balanced Accuracy of 81.13% and a ROC-AUC of 89.21%, demonstrating that the extracted EEG spectral descriptors already provided substantial discriminative information for Alzheimer's disease recognition.

Following the baseline experiment, a feature importance analysis was performed using the coefficients of the linear SVM. Features were ranked according to their average contribution across the outer cross-validation folds, and several Top-K feature subsets were subsequently evaluated. Rather than selecting an arbitrary number of descriptors, the complete nested cross-validation procedure was repeated for each candidate subset in order to identify the optimal feature configuration.

Hyperparameter optimization was then performed using the Optuna framework. Two optimization strategies were investigated: the first considered the complete feature set, while the second combined Optuna with the optimal subset of selected features. The performance obtained by the different configurations is summarized in Table 4.1.

Method	**Accuracy**	**Balanced Accuracy**
Baseline SVM	84.87 ± 2.96	81.13 ± 4.29
Top-50 Features	91.79 ± 2.07	91.69 ± 2.65
Optuna All Features (475)	90.06 ± 5.37	89.50 ± 6.15
Optuna + Top-50 Features	**92.22 ± 3.21**	**92.02 ± 3.19**

Table 4.1. Performance obtained for Alzheimer's disease binary classification using different stages of the proposed classification pipeline.

The results clearly show the progressive improvement obtained throughout the proposed pipeline. Feature selection alone substantially improved the baseline model, increasing the Balanced Accuracy from 81.13% to 91.69%. This result indicates that a relatively small subset of EEG descriptors contains most of the discriminative information required to distinguish Alzheimer's disease patients from healthy controls.

Hyperparameter optimization further enhanced the classifier performance. When Optuna was applied to the complete feature set, the Balanced Accuracy increased to 89.50%, confirming the effectiveness of automated hyperparameter tuning. However, the best overall performance was achieved by combining Optuna with the selected Top-50 features, reaching an Accuracy of 92.22% and a Balanced Accuracy of 92.02%.

To further investigate the influence of feature dimensionality, different Top-K feature subsets were evaluated following the ranking procedure described in Chapter 3. For each Top-K subset, the complete nested cross-validation procedure was repeated. Figure 4.1 reports the Balanced Accuracy obtained for each evaluated subset.

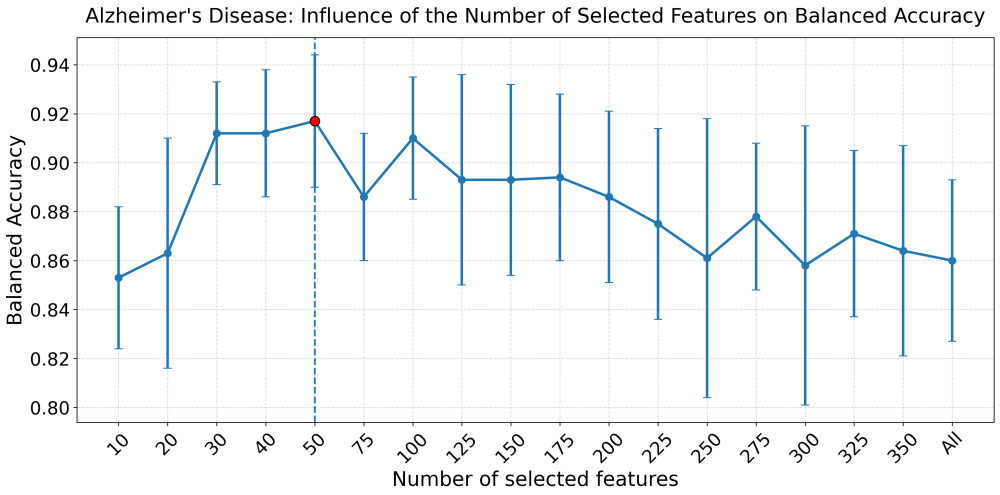

Figure 4.1. Balanced Accuracy obtained for different Top-K feature subsets during the feature selection procedure for Alzheimer's disease classification. Error bars represent the standard deviation across the outer folds of the nested cross-validation.

As shown in Figure 4.1, the classification performance does not monotonically increase with the number of selected features. Instead, the Balanced Accuracy reaches its maximum when approximately 50 features are retained. Beyond this point, progressively adding lower-ranked descriptors leads to a gradual decrease in performance, suggesting that additional features mainly introduce redundancy and noise rather than useful discriminative information.

Overall, these results demonstrate that feature selection had the greatest impact on classification performance, simultaneously reducing the feature space by almost 90% (from 475 to 50 features) while improving the Balanced Accuracy by more than 10 percentage points. The subsequent application of automated hyperparameter optimization further refined the classifier, leading to the best overall performance achieved for Alzheimer's disease classification.

4.1.2 Epilepsy Classification

The second set of experiments focused on the binary discrimination between EP (50) and HC (71). As for the Alzheimer's disease experiments, the objective was to evaluate the effectiveness of the proposed classification framework and to quantify the contribution of feature selection and hyperparameter optimization.

As a baseline, the complete feature set was used to train a linear Support Vector Machine within the Nested GroupKFold cross-validation framework described in Chapter 3. The baseline classifier achieved an Accuracy of 80.93%, a Balanced Accuracy of 80.48% and a ROC-AUC of 84.25%, confirming that the extracted EEG features were able to capture relevant electrophysiological differences between epileptic patients and healthy controls.

Following the baseline experiment, feature importance analysis was performed using the coefficients of the linear SVM classifier. Features were ranked according to their average contribution across the outer cross-validation folds, and several Top-K feature subsets were subsequently evaluated. Hyperparameter optimization was then performed using Optuna, both on the complete feature set and on the optimal subset of selected features. The performance obtained by the different configurations is summarized in Table 4.2.

Method	**Accuracy**	**Balanced Accuracy**
Baseline SVM	80.93 ± 5.77	80.48 ± 5.37
Top-50 Features	86.77 ± 4.88	86.62 ± 5.26
Optuna All Features (475)	82.60 ± 7.24	81.90 ± 6.92
Optuna + Top-50 Features	**87.63 ± 5.17**	**87.05 ± 6.32**

Table 4.2. Performance obtained for Epilepsy binary classification using different stages of the proposed classification pipeline.

The obtained results confirm the effectiveness of the proposed methodology. Feature selection substantially improved the baseline classifier, increasing the Balanced Accuracy from 80.48% to 86.62%. This finding indicates that, similarly to the Alzheimer's disease experiments, only a relatively small subset of EEG descriptors is required to effectively discriminate epileptic patients from healthy subjects.

Hyperparameter optimization performed on the complete feature space produced a moderate improvement over the baseline model, reaching a Balanced Accuracy of 81.90%. However, the best overall performance was obtained when Optuna was combined with the selected Top-50 features, achieving an Accuracy of 87.63% and a Balanced Accuracy of 87.05%.

To further investigate the influence of feature dimensionality, different Top-K feature subsets were evaluated following the ranking procedure described in Chapter 3. Figure 4.2 reports the Balanced Accuracy obtained for each evaluated subset.

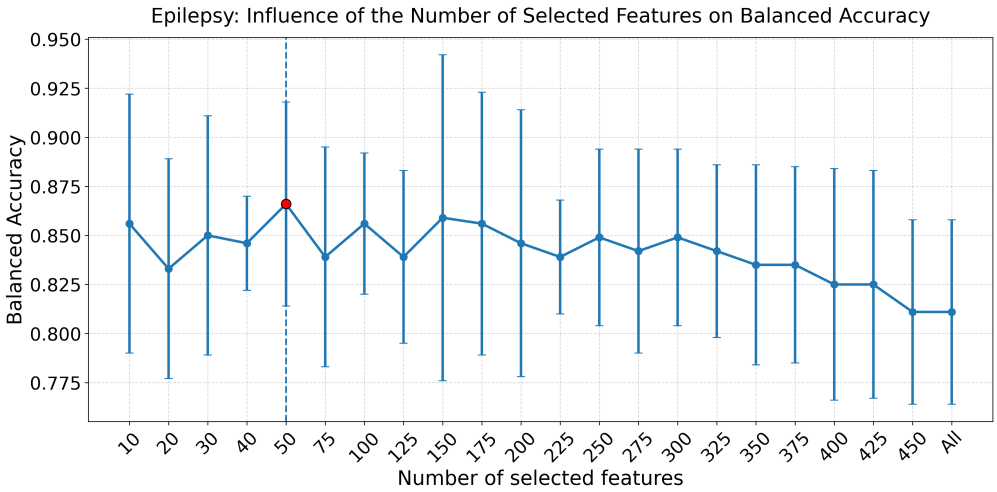

Figure 4.2. Balanced Accuracy obtained for different Top-K feature subsets during the feature selection procedure for epilepsy classification. Error bars represent the standard deviation across the outer folds of the nested cross-validation.

As shown in Figure 4.2, the classifier reaches its highest Balanced Accuracy when the Top-50 features are retained. Increasing the number of selected descriptors beyond this point does not improve classification performance and generally results in a slight performance degradation, suggesting that additional features mainly contribute redundant information.

Overall, the epilepsy experiments confirm the observations obtained for Alzheimer's disease. Feature selection represents the most influential step of the proposed pipeline, while automated hyperparameter optimization provides an additional improvement when applied to the reduced feature space. These results demonstrate that the proposed framework is capable of identifying compact and highly discriminative EEG representations across different neurological disorders.

4.1.3 Amyotrophic Lateral Sclerosis Classification

The third set of experiments focused on the binary discrimination between HC (167) and patients affected by ALS (6). Unlike the Alzheimer's disease and epilepsy datasets, the EEGET-ALS dataset contains multiple recording sessions acquired from the same ALS patient. Specifically, each patient performed nine experimental scenarios across multiple acquisition times. To maximize the amount of available data, EEG segments corresponding to the same scenario collected at different acquisition times were combined, resulting in multiple EEG segments for each ALS subject.

After preprocessing and quality control, the final dataset consisted of 167 subjects and 1490 EEG segments, including 54 pathological segments for 6 patients and 1436 healthy segments for 167 patients. This substantial class imbalance motivated the adoption of Balanced Accuracy as the primary performance metric.

Consequently, the proposed framework was evaluated at both the segment level and the subject level. While the segment-level evaluation assesses the classification of individual EEG segments, the subject-level evaluation aggregates the predictions belonging to the same patient, providing a more clinically meaningful assessment of the classifier performance.

As a baseline, the complete feature set was used to train a linear Support Vector Machine within the Nested GroupKFold cross-validation framework described in Chapter 3. Feature importance analysis was subsequently performed using the coefficients of the linear SVM, allowing the extracted EEG descriptors to be ranked according to their average contribution across the outer cross-validation folds. As in the previous experiments, several Top-K feature subsets were evaluated and hyperparameter optimization was performed using the Optuna framework. The performance obtained by the different configurations is summarized in Table 4.3 for the segment-level analysis and in Table 4.4 for the subject-level analysis.

Method	**Accuracy**	**Balanced Accuracy**
Baseline SVM	**98.66 ± 0.38**	81.87 ± 2.95
Top-10 Features	98.12 ± 0.76	**95.69 ± 4.05**
Optuna All Features (800)	96.18 ± 2.42	85.17 ± 8.73
Optuna + Top-20 Features	97.45 ± 1.97	94.16 ± 4.48

Table 4.3. Segment-level performance obtained for ALS binary classification using different stages of the proposed classification pipeline.

Method	**Accuracy**	**Balanced Accuracy**
Baseline SVM	98.84 ± 8.20	83.33 ± 10.39
Top-20 Features	**99.42 ± 1.18**	**91.67 ± 7.90**
Optuna All Features (800)	97.11 ± 0.82	82.44 ± 7.56
Optuna + Top-20 Features	98.84 ± 8.20	91.37 ± 7.66

Table 4.4. Subject-level performance obtained for ALS binary classification using different stages of the proposed classification pipeline.

At first glance, all investigated models achieved extremely high classification accuracy, with values consistently above 96%. However, because of the pronounced class imbalance, Accuracy alone provides a potentially misleading estimate of classifier performance. For this reason, the discussion primarily focuses on Balanced Accuracy, which equally weights the classification performance of both classes.

Considering the segment-level evaluation (Table 4.3), feature selection substantially improved the baseline classifier, increasing the Balanced Accuracy from 81.87% to 95.69%. Interestingly, the optimal feature subset consisted of only 10 descriptors, considerably fewer than those required for the Alzheimer's disease and epilepsy classification tasks. This finding suggests that a relatively compact subset of EEG features is sufficient to effectively discriminate ALS patients from healthy controls. Considering the limited number of available ALS subjects, reducing the dimensionality of the feature space also mitigates the risk of overfitting, leading to a more robust and generalizable classification model.

Hyperparameter optimization performed on the complete feature space increased the segment-level Balanced Accuracy to 85.17%, representing an improvement over the baseline model but remaining inferior to the results obtained using the selected Top-20 features. Similarly to the previous experiments, combining feature selection with Optuna produced a comparable Balanced Accuracy (94.16%), confirming that feature selection represented the main contributor to the observed performance improvement.

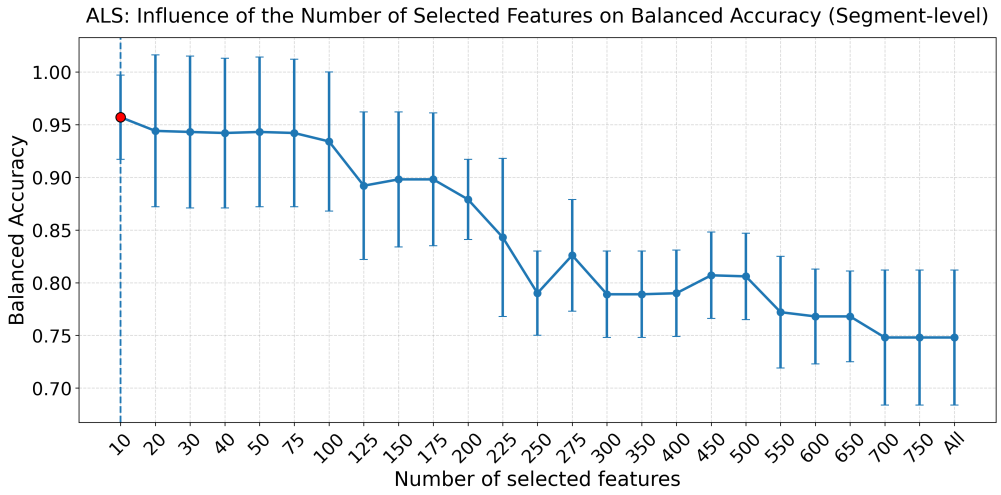

Figure 4.3. Segment-level Balanced Accuracy obtained for different Top-K feature subsets during the feature selection procedure for ALS classification. Error bars represent the standard deviation across the outer folds of the nested cross-validation.

The influence of feature dimensionality at the segment level was further investigated by evaluating multiple Top-K feature subsets following the ranking procedure described in Chapter 3. The corresponding results are reported in Figure 4.3.

As shown in Figure 4.3, the highest Balanced Accuracy was achieved using the Top-10 features. Increasing the number of selected descriptors beyond this point did not improve the classifier performance and generally resulted in a slight decrease in Balanced Accuracy. This behaviour suggests that the most discriminative information is concentrated within a relatively small subset of EEG features. Furthermore, considering the limited number of independent ALS subjects available in the dataset, increasing the feature dimensionality without a proportional increase in the number of training samples may aggravate the curse of dimensionality, making the classification problem more prone to overfitting and reducing the generalization capability of the classifier. Therefore, selecting a compact subset of highly informative features not only improves interpretability but also contributes to better generalization performance.

The subject-level evaluation (Table 4.4) confirmed the same overall trend. Aggregating the predictions belonging to the same patient increased the robustness of the classification, leading to a Balanced Accuracy of 91.67% when the Top-20 feature subset was employed. Figure 4.4 illustrates the subject-level Balanced Accuracy obtained for the different Top-K feature configurations.

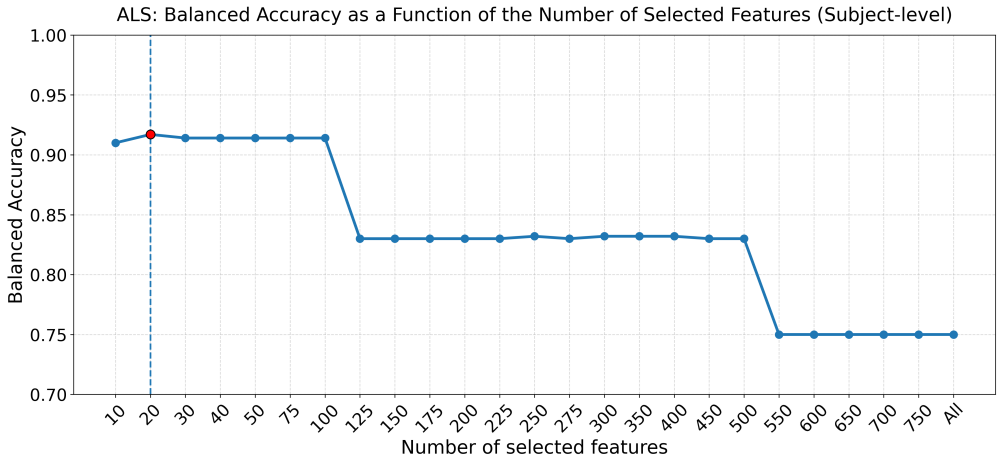

Figure 4.4. Subject-level Balanced Accuracy obtained for different Top-K feature subsets during the feature selection procedure for ALS classification.

Unlike the segment-level evaluation, the subject-level performance remains relatively stable once the optimal subset of features has been selected. This behaviour is expected since aggregating multiple EEG segments belonging to the same patient reduces the variability associated with individual recordings and provides a more robust representation of each subject.

Overall, the ALS experiments further confirm the effectiveness of the proposed classification framework. Consistently with the Alzheimer's disease and epilepsy experiments, feature selection represented the most influential step of the pipeline, substantially improving the Balanced Accuracy while simultaneously reducing the dimensionality of the feature space. The results suggest that limiting the number of input features mitigates the negative effects associated with high-dimensional feature spaces, particularly when only a limited number of independent subjects is available. Hyperparameter optimization provided a further refinement of the classifier, although its contribution was smaller than that achieved through feature selection alone.

4.2 Multiclass Classification

After validating the proposed framework on the three binary classification tasks, the experiments were extended to a more challenging multiclass scenario involving Healthy subjects together with patients affected by Alzheimer's disease, Epilepsy and Amyotrophic Lateral Sclerosis. Unlike the previous experiments, the classifier was required to simultaneously discriminate among four different classes, substantially increasing the complexity of the classification task.

The final multiclass dataset consisted of 525 subjects and 1842 EEG segments, including 312 HC, 157 AD, 50 EP and 6 ALS patients. As expected, this distribution resulted in a considerable class imbalance, particularly due to the limited number of ALS subjects.

The multiclass classifier was evaluated using the same nested cross-validation protocol adopted throughout the previous experiments. Specifically, a three-fold outer GroupKFold was employed to estimate the generalization performance, while a three-fold inner GroupKFold was used for model selection and hyperparameter optimization. Subject identifiers were used as grouping variables to ensure that EEG segments belonging to the same subject were never simultaneously assigned to both training and test sets, thus preventing information leakage.

Similarly to the binary experiments, a baseline linear SVM was first trained using the complete feature set. Feature importance analysis was then performed to rank the extracted EEG descriptors according to their discriminative contribution. Several Top-K feature subsets were subsequently evaluated, followed by hyperparameter optimization using the Optuna framework.

Since the final objective of the proposed system is the correct classification of individual patients rather than individual EEG segments, the following discussion focuses on the subject-level evaluation. The corresponding segment-level results are reported in Appendix B and show the same overall trends.

Table 4.5 summarizes the subject-level classification performance obtained by the proposed framework.

Method	**Accuracy**	**Balanced Accuracy**	**Macro F1**
Baseline SVM	**75.62 ± 1.42**	79.07 ± 7.75	0.655 ± 0.035
Top-125 Features	75.61 ± 1.45	**85.33 ± 5.68**	**0.664 ± 0.052**
Optuna All Features (375)	73.90 ± 1.58	75.79 ± 4.56	0.637 ± 0.028
Optuna + Top-125 Features	73.33 ± 1.68	84.22 ± 4.24	0.645 ± 0.025

Table 4.5. Subject-level performance obtained for the multiclass classification problem.

The obtained results demonstrate that the proposed framework successfully generalizes from binary to multiclass EEG classification.

The baseline classifier achieved a Balanced Accuracy of 79.07%, already indicating a good capability to discriminate among the four neurological conditions. Nevertheless, feature selection substantially improved the classification performance, increasing the Balanced Accuracy to 85.33% while reducing the feature space from 375 to only 125 descriptors. This confirms the observations made throughout the binary classification experiments, where feature selection consistently represented the most influential stage of the proposed pipeline.

Combining feature selection with Optuna yielded a Balanced Accuracy of 84.22%, confirming that the largest performance gains originated from the reduction of the feature space rather than from hyperparameter optimization.

Interestingly, the multiclass problem required a considerably larger feature subset than the binary classification tasks. While Alzheimer's disease and epilepsy achieved their best performance using approximately 50 features and ALS required only 20 descriptors, the multiclass classifier reached its optimum with 125 selected features. This behaviour is expected, since simultaneously discriminating four neurological conditions requires a richer representation capable of capturing disease-specific EEG patterns across multiple classes.

To further investigate the behaviour of the proposed classifier, the subject-level confusion matrix obtained using the baseline linear SVM was analyzed (Figure 4.5).

The baseline classifier correctly recognized the majority of healthy subjects, Alzheimer's disease and epilepsy patients. Epilepsy represented the easiest class to identify, whereas ALS remained the most challenging due to the extremely limited number of available subjects. Most classification errors involved ALS subjects being incorrectly assigned to the healthy class, reflecting the severe class imbalance characterizing the dataset.

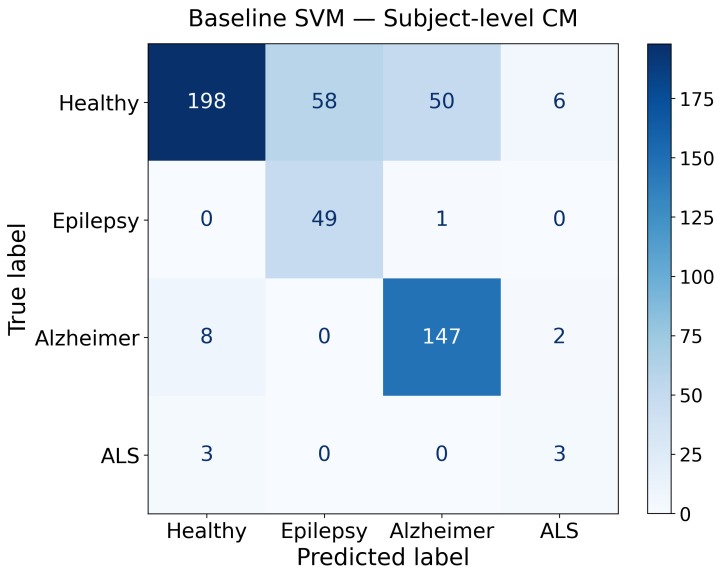

Figure 4.5. Multiclass: Baseline SVM Confusion Matrix.

The subject-level confusion matrix obtained using the final optimized model (Optuna + Top-125 features) is reported in Figure 4.6.

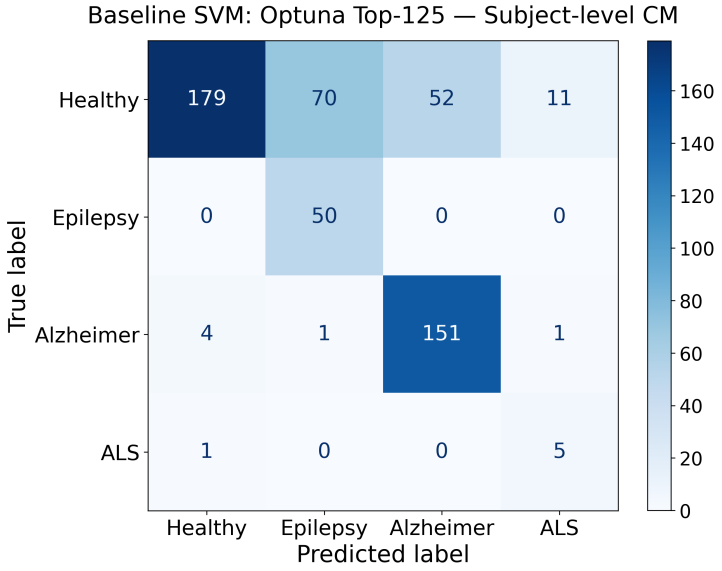

Figure 4.6. Multiclass: Baseline SVM Confusion Matrix using Optuna + Top-125.

Compared with the baseline classifier, the optimized framework substantially improves the recognition of the minority classes, particularly ALS, while maintaining excellent performance for epilepsy and slightly improving the classification of Alzheimer's disease. Although a small reduction is observed in the number of correctly classified healthy subjects, this behaviour represents a desirable trade-off. From a clinical perspective, falsely identifying a healthy subject as pathological is generally less critical than failing to detect a neurological disorder, since additional clinical examinations can subsequently exclude the disease. Consequently, the increase in Balanced Accuracy mainly reflects a more reliable recognition of pathological subjects rather than an improvement limited to the already dominant healthy class.

Overall, the multiclass experiments demonstrate that the proposed framework successfully extends the binary classification pipeline to a substantially more challenging four-class problem. Despite the increased complexity and severe class imbalance, the proposed methodology achieved competitive subject-level performance, with feature selection consistently representing the main contributor to the observed improvement. The obtained results further confirm the effectiveness of the proposed EEG feature representation for the automated discrimination of multiple neurological disorders.

4.3 Two-Stage Classification Analysis

Although the standard multiclass classifier demonstrated that the proposed framework is capable of simultaneously discriminating between healthy subjects and three neurological disorders, this formulation does not explicitly reflect the diagnostic workflow typically adopted in clinical practice. Physicians generally perform the diagnostic process in two consecutive stages. First, they determine whether the patient presents signs of neurological impairment. Subsequently, only pathological subjects undergo a differential diagnosis aimed at identifying the specific neurological disorder.

Motivated by this observation, an additional set of experiments was conducted to independently investigate these two classification stages. Rather than directly solving the four-class problem, two independent classifiers were developed. The first classifier distinguishes healthy subjects from pathological subjects, whereas the second classifier considers only pathological EEG recordings and discriminates among Alzheimer's disease, epilepsy and Amyotrophic Lateral Sclerosis.

It is important to emphasize that this experiment does not represent a complete hierarchical classifier. Instead, it aims to evaluate the capability of the proposed feature extraction and classification framework to solve the two classification tasks independently. A fully integrated leakage-free hierarchical architecture, in which the predictions of the first classifier determine the input of the second stage, is presented in the following section.

4.3.1 Healthy vs Disease Classification

The first stage addresses the binary discrimination between healthy subjects and patients affected by neurological disorders. All pathological subjects, including Alzheimer's disease, Epilepsy and ALS, were grouped into a single disease class, resulting in a binary screening problem aimed at identifying the presence of neurological impairment.

The same experimental pipeline described in Chapter 3 was adopted. A baseline linear Support Vector Machine was initially trained using the complete feature set. Feature importance analysis was then performed, followed by Top-K feature selection and Optuna-based hyperparameter optimization.

Table 4.6 summarizes the subject-level performance obtained for the Healthy vs Disease classification task.

Method	**Accuracy**	**Balanced Accuracy**	**Macro F1**
Baseline SVM	74.10 ± 3.18	77.16 ± 2.35	0.740 ± 0.032
Top-100 Features	**75.81 ± 1.00**	**79.08 ± 1.17**	**0.757 ± 0.010**
Optuna All Features (375)	75.24 ± 3.57	78.65 ± 2.89	0.752 ± 0.036
Optuna + Top-100 Features	75.43 ± 2.14	78.81 ± 1.50	0.754 ± 0.022

Table 4.6. Subject-level performance obtained for the Healthy vs Disease classification task.

The obtained results demonstrate that the proposed EEG representation reliably separates healthy subjects from patients affected by neurological disorders.

The baseline classifier achieved a Balanced Accuracy of 77.16%, already providing a satisfactory screening capability. Selecting the 100 most discriminative features further improved the Balanced Accuracy to 79.08%, confirming once again the positive contribution of feature selection. Hyperparameter optimization produced only marginal additional improvements, with the final optimized configuration reaching a Balanced Accuracy of 78.81%.

Figures 4.7 and 4.8 report the subject-level confusion matrices obtained before and after feature selection.

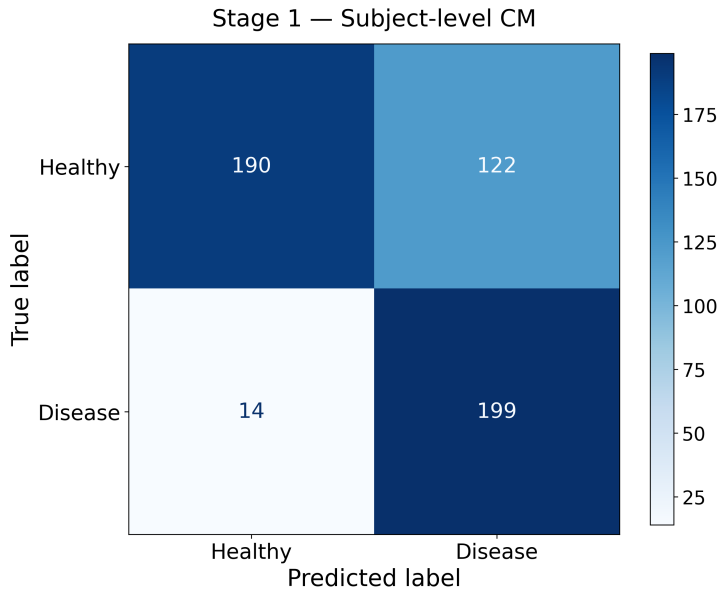

Figure 4.7. Healthy vs Disease: Baseline Confusion Matrix.

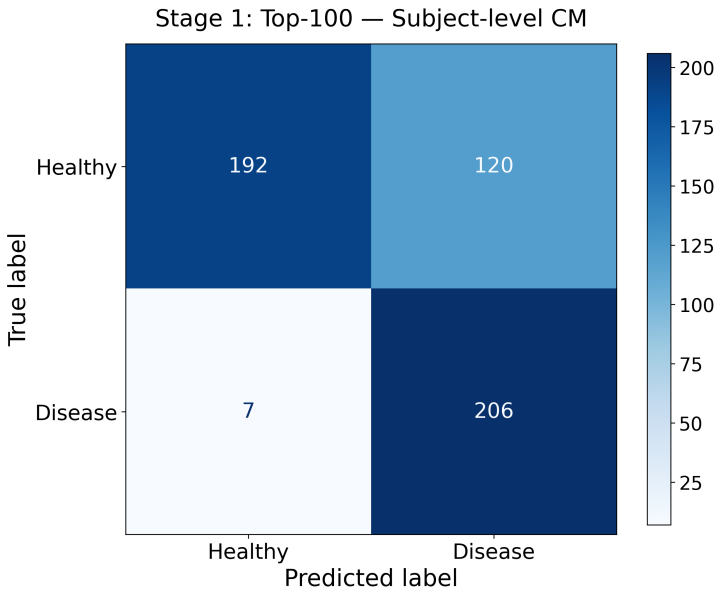

Figure 4.8. Healthy vs Disease: Confusion Matrix obtained using the Top-100 Feature subset.

Compared with the baseline classifier, the optimized model exhibits fewer false positive and false negative predictions, leading to a more balanced discrimination between healthy and pathological subjects. From a clinical perspective, this behaviour is particularly desirable since the screening stage aims to minimize the number of undetected neurological disorders while maintaining a satisfactory recognition of healthy individuals.

Although the subject-level evaluation yielded lower performance than the segment-level analysis, the same overall trend was observed. Feature selection consistently improved the classification performance, while hyperparameter optimization provided only marginal additional benefits.

4.3.2 Disease-Specific Classification

The second stage investigates a different classification problem. Instead of distinguishing healthy subjects from patients, only pathological EEG recordings are considered. Consequently, the classifier performs a differential diagnosis among Alzheimer's disease, Epilepsy and Amyotrophic Lateral Sclerosis.

Since healthy subjects are excluded from this experiment, the classification task becomes intrinsically less complex than the original four-class problem. Nevertheless, accurately distinguishing among neurological disorders remains clinically relevant, as different diseases may exhibit partially overlapping EEG alterations.

The same experimental pipeline was adopted. A baseline Support Vector Machine was first trained using the complete feature space, followed by feature ranking, Top-K feature selection and Optuna-based hyperparameter optimization.

Since this experiment aims at evaluating the capability of the proposed framework to correctly identify the neurological disorder affecting each patient, the following discussion focuses on the subject-level evaluation. Table 4.7 shows the subject-level results.

Method	**Accuracy**	**Balanced Accuracy**	**Macro F1**
Baseline SVM	95.77 ± 2.30	97.61 ± 1.65	0.866 ± 0.047
Top-50 Features	**98.59 ± 1.15**	**99.36 ± 0.52**	**0.957 ± 0.053**
Optuna All Features (375)	98.59 ± 0.00	98.91 ± 0.68	0.946 ± 0.026
Optuna + Top-50 Features	98.59 ± 0.66	99.36 ± 0.30	0.930 ± 0.033

Table 4.7. Subject-level performance obtained for the disease-specific classification task (Alzheimer's disease, Epilepsy and ALS).

The obtained results demonstrate that, once pathological subjects have been identified, the proposed framework is capable of distinguishing among Alzheimer's disease, Epilepsy and ALS with remarkably high performance.

The baseline classifier already achieved a Balanced Accuracy of 97.61%, which increased to 99.36% after selecting the 50 most informative features. Hyperparameter optimization alone produced only marginal improvements, whereas combining feature selection with Optuna maintained an almost identical performance.

These findings indicate that the extracted EEG descriptors contain highly discriminative disease-specific information. Moreover, only a relatively small subset of features is required to accurately separate the three neurological disorders, further confirming the effectiveness of the proposed feature selection strategy.

The corresponding segment-level performances are reported in Appendix B and exhibit the same overall behaviour.

Figures 4.9 and 4.10 report the subject-level confusion matrices obtained before and after feature selection.

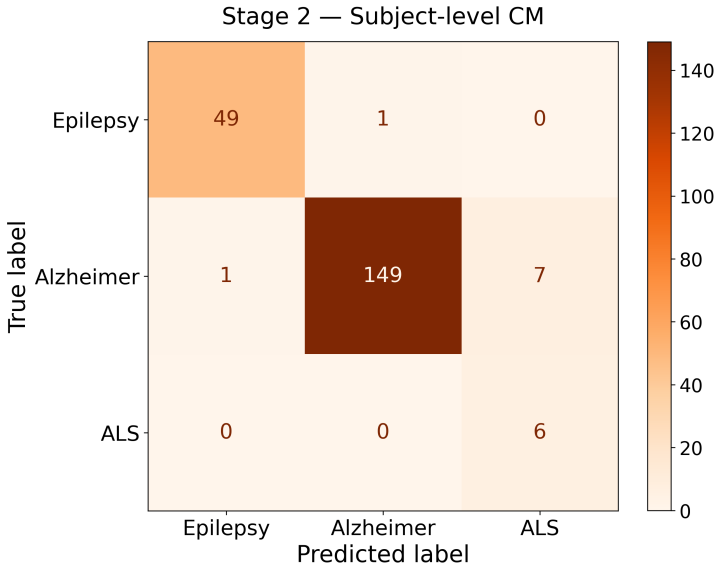

Figure 4.9. Disease-specific Classification: Baseline Confusion Matrix.

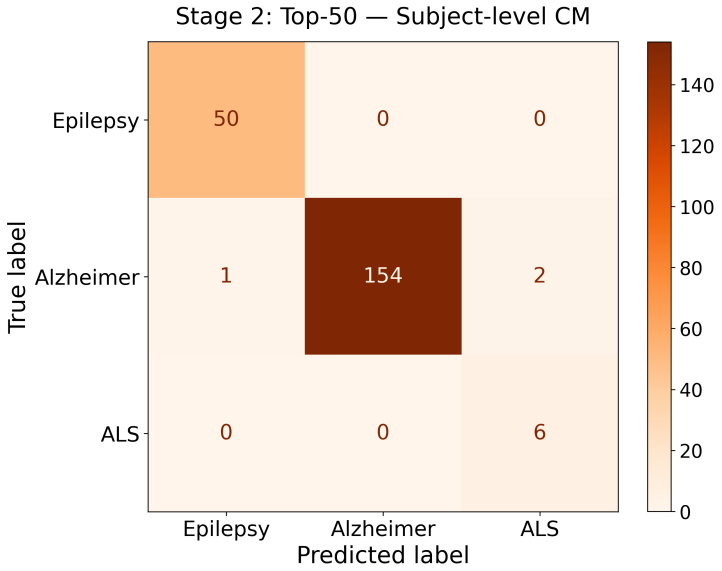

Figure 4.10. Disease-specific Classification: Baseline Confusion Matrix using the Top-50 Feature subset.

Compared with the baseline classifier, the optimized model achieves an almost perfect recognition of the three neurological disorders, with only a very limited number of misclassified subjects. The improvements mainly concern Alzheimer's disease and ALS, while epilepsy is correctly identified in nearly all cases.

The two-stage analysis provides useful insights into the intrinsic difficulty of the proposed diagnostic workflow. The first-stage classifier demonstrated that healthy EEG recordings could be reliably distinguished from pathological recordings, achieving a balanced accuracy of 79%. Among pathological subjects, the second-stage classifier achieved almost perfect discrimination among Alzheimer's disease, epilepsy, and ALS, with a Balanced Accuracy close to 99%.

Although these results demonstrate the strong discriminative capability of the proposed EEG feature extraction framework, the second-stage classifier was evaluated under an idealized assumption, namely that the separation between healthy and pathological subjects was already known. This experimental setting was intentionally adopted to assess the maximum achievable performance of the disease-specific classifier independently of the screening stage.

In a real clinical scenario, however, the disease-specific classifier would receive as input only the subjects predicted as pathological by the screening classifier. Consequently, classification errors produced during the first stage would propagate to the second stage and could affect the overall diagnostic performance.

To evaluate the proposed framework under more realistic conditions, the following section introduces a complete leakage-free hierarchical cascade, in which only the subjects predicted as pathological by the first-stage classifier are forwarded to the disease-specific classifier. This configuration provides a more realistic representation of the intended clinical workflow while eliminating information leakage between the two classification stages.

4.4 Leakage-Free Hierarchical Cascade

Although the two-stage analysis demonstrated that the proposed EEG descriptors can accurately distinguish neurological disorders once pathological subjects have been identified, the second-stage classifier was evaluated using the ground-truth disease labels. Such an experimental setting does not reflect real clinical practice, where the disease classifier receives only the subjects identified as pathological during the screening stage.

To overcome this limitation, a complete leakage-free hierarchical cascade was developed. In this framework, the first-stage classifier distinguishes healthy subjects from pathological subjects. Only those subjects predicted as pathological are subsequently forwarded to the second-stage classifier, which identifies the specific neurological disorder. Consequently, classification errors generated during the screening stage propagate to the second stage, providing a more realistic evaluation of the complete diagnostic workflow.

The same Nested GroupKFold strategy described in Chapter 3 was adopted throughout the entire cascade. Subject grouping was preserved during both stages of the pipeline, ensuring that no information leakage occurred between the training and testing sets.

Unlike the previous two-stage analysis, the second classifier is trained exclusively on the training partition of each outer fold, thus avoiding information leakage between the two stages. Consequently, this experimental setup provides a more realistic evaluation of the proposed diagnostic framework.

Table 4.8 summarizes the subject-level performance obtained by the complete hierarchical cascade.

Method	**Accuracy**	**Balanced Accuracy**	**Macro F1**	**ROC-AUC**
Baseline Cascade	73.15 ± 2.74	82.64 ± 7.12	0.631 ± 0.047	0.870 ± 0.038
Top-150 Features	74.11 ± 2.23	83.77 ± 6.60	0.632 ± 0.030	0.878 ± 0.033
Optuna All Features (375)	73.34 ± 2.59	82.80 ± 6.89	0.635 ± 0.046	0.871 ± 0.037
Optuna + Top-150 Features	$\mathbf{73.72 \pm 1.80}$	$\mathbf{83.83 \pm 5.90}$	$\mathbf{0.627 \pm 0.004}$	$\mathbf{0.878 \pm 0.029}$

Table 4.8. Subject-level performance obtained for the leakage-free hierarchical cascade.

The obtained results differ from those observed during the previous two-stage analysis. Since the second-stage classifier receives only the subjects predicted as pathological by the first-stage classifier, classification errors generated during the

screening stage inevitably propagate throughout the cascade. Consequently, the overall classification problem becomes considerably more challenging.

Despite this increased complexity, the proposed hierarchical framework achieved a subject-level Balanced Accuracy of 82.64% using the complete feature space. Selecting the 150 most discriminative features improved classification performance, increasing the Balanced Accuracy to 83.83%.

To further investigate the behaviour of the proposed cascade, classification errors were analyzed separately for the two stages of the diagnostic pipeline.

Out of 1842 evaluated EEG segments, 168 errors (9.1%) originated from the first screening stage. More specifically, the binary classifier incorrectly classified 22 pathological segments as healthy (false negatives) and 146 healthy segments as pathological (false positives). Once the pathological subjects had been identified, the disease-specific classifier generated only 9 additional classification errors, corresponding to 3.8% of the correctly detected pathological segments.

These findings indicate that the overall performance of the hierarchical cascade is primarily limited by the first screening stage rather than by the disease-specific classifier. Once neurological impairment has been correctly identified, the second-stage classifier is capable of distinguishing among Alzheimer's disease, epilepsy and ALS with very high reliability.

The subject-level confusion matrix of the complete hierarchical cascade is reported in Figures 4.11 and 4.12.

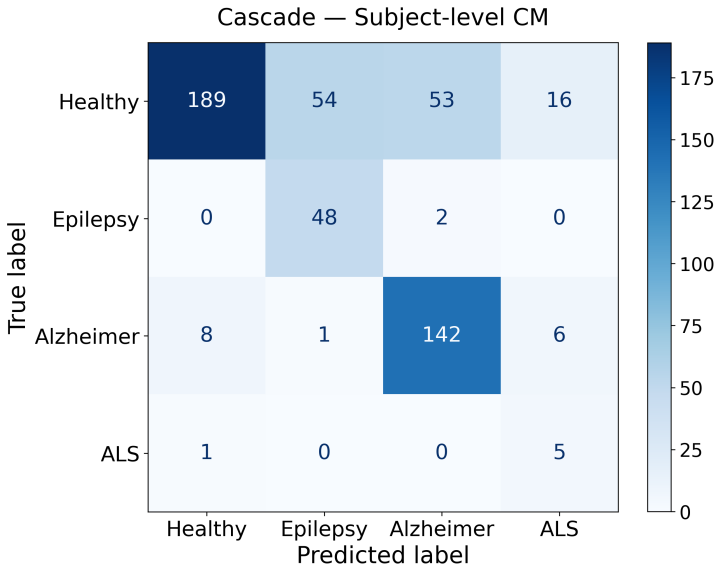

Figure 4.11. Cascade: Baseline Confusion Matrix.

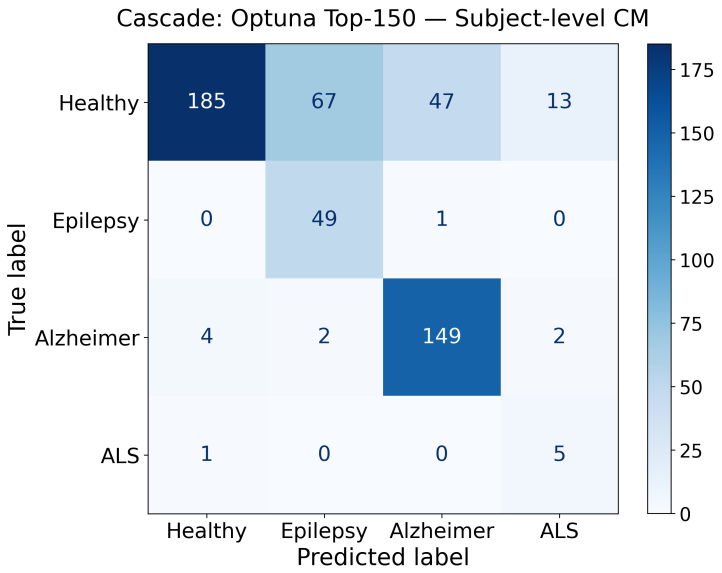

Figure 4.12. Cascade: Baseline Confusion Matrix using Optuna + Top-150.

Most healthy subjects and epilepsy patients were correctly classified, while the majority of the remaining errors originated from the first-stage screening classifier. In particular, pathological subjects incorrectly classified as healthy during the screening stage could not be recovered by the disease-specific classifier, since they never reached the second stage of the cascade.

These findings further indicate that improving the screening stage is likely to provide the largest overall benefit to the complete diagnostic workflow, since even small improvements in the first-stage sensitivity would directly translate into better end-to-end classification performance.

4.4.1 Class-wise Performance Analysis

To better understand the behaviour of the proposed hierarchical cascade, additional performance metrics were analyzed for each neurological class using the final optimized model (Optuna + Top-150 features). Table 4.9 reports the class-wise subject-level performance obtained by the final optimized hierarchical cascade. These metrics provide a more detailed understanding of the behaviour of the proposed framework than the overall performance measures reported in Table 4.8.

Class	**Recall**	**Specificity**	**Precision**	**F_1-score**	**ROC-AUC**
Healthy	0.590	0.977	0.974	0.735	0.783
Epilepsy	0.980	0.857	0.419	0.587	0.918
Alzheimer's	0.949	0.867	0.753	0.839	0.908
ALS	0.833	0.969	0.238	0.370	0.901

Table 4.9. Class-wise subject-level performance obtained by the final optimized leakage-free hierarchical cascade (Optuna + Top-150 features). All metrics are reported on a scale from 0 to 1.

Alzheimer's disease achieved the highest overall classification performance among the pathological classes, with a sensitivity of 94.90%, a precision of 75.25% and an F_1-score of 83.94%. These results indicate that the proposed framework is able to reliably recognize Alzheimer's patients while maintaining a good balance between false positives and false negatives.

Epilepsy also exhibited an excellent sensitivity (98.00%), demonstrating that almost all epileptic subjects were correctly identified. However, the corresponding precision was considerably lower (41.88%), indicating that a substantial number of subjects predicted as epileptic actually belonged to other classes. This behaviour reflects the conservative nature of the screening stage, which tends to favour the detection of pathological subjects at the expense of increasing the number of false positive classifications.

Despite the extremely limited number of available ALS subjects, the proposed framework correctly identified 83.33% of them. Although the precision remained relatively low (23.81%), the specificity reached 96.92%, confirming that ALS predictions were generated only in a limited number of healthy subjects. Considering that only six ALS patients were available, these results should be interpreted with caution but nevertheless demonstrate the capability of the proposed framework to capture disease-specific EEG alterations even under severe class imbalance.

Finally, healthy subjects exhibited the lowest sensitivity (58.97%) but an extremely high specificity (97.65%) and precision (97.35%). This behaviour indicates that most classification errors originated from healthy subjects being classified as

pathological rather than from pathological subjects being classified as healthy.

From a clinical perspective, this behaviour represents a desirable trade-off. The hierarchical cascade intentionally favors sensitivity toward neurological disorders, accepting a higher number of false positive diagnoses among healthy subjects. Consequently, several healthy individuals are referred to the second-stage classifier and may undergo additional diagnostic examinations. Although this slightly reduces the overall classification accuracy, it substantially decreases the risk of overlooking neurological diseases, which would represent a considerably more critical error in clinical practice.

Conversely, only a limited number of pathological subjects are incorrectly classified as healthy during the screening stage. Once a subject reaches the disease-specific classifier, the probability of correctly identifying the underlying neurological disorder becomes very high, as demonstrated by the excellent sensitivities obtained for Alzheimer's disease, Epilepsy and ALS.

Overall, these findings indicate that the proposed leakage-free hierarchical cascade successfully prioritizes the identification of neurological disorders over the correct recognition of healthy subjects. This design choice is consistent with real clinical screening applications, where generating additional follow-up examinations for healthy individuals is generally preferable to failing to detect patients affected by neurological diseases.

4.5 Data Augmentation using SMOTE and CT-GAN

The experiments presented in the previous sections highlighted that one of the main limitations of the proposed hierarchical classification framework is the severe class imbalance characterizing the available EEG datasets, particularly due to the limited number of ALS subjects. To investigate whether balancing the training data could improve the classification performance, two widely adopted data augmentation techniques were evaluated: Synthetic Minority Over-sampling Technique (SMOTE) and Conditional Tabular Generative Adversarial Networks (CTGAN).

SMOTE generates synthetic feature vectors through linear interpolation between neighboring minority-class samples, whereas CTGAN employs a generative adversarial network specifically designed for tabular data to learn the underlying feature distribution and generate more realistic synthetic observations.

To prevent information leakage, synthetic samples were generated exclusively from the training partition of each outer fold of the Nested GroupKFold procedure. Consequently, validation and test sets always contained only real EEG recordings, ensuring a fair comparison with the previously presented experiments.

Method	**Accuracy**	**Balanced Accuracy**	**Macro F1**
SMOTE	84.57 ± 3.12	75.97 ± 15.08	0.730 ± 0.143
CTGAN	76.97 ± 2.97	82.04 ± 11.00	0.652 ± 0.095

Table 4.10. Subject-level performance obtained using SMOTE and CT-GAN data augmentation.

The subject-level results reported in Table 4.10 indicate that SMOTE did not improve the overall performance of the proposed hierarchical framework. Although a satisfactory overall Accuracy of 84.57% was achieved, the corresponding Balanced Accuracy decreased compared with the baseline hierarchical cascade presented in the previous section.

The subject-level confusion matrix obtained using SMOTE is reported in Figure 4.13. While the number of correctly classified healthy EEG segments increased considerably, the recognition of minority classes, particularly ALS and epilepsy, deteriorated. Consequently, the increase in overall Accuracy mainly reflects the highly imbalanced class distribution rather than a genuine improvement in disease recognition.

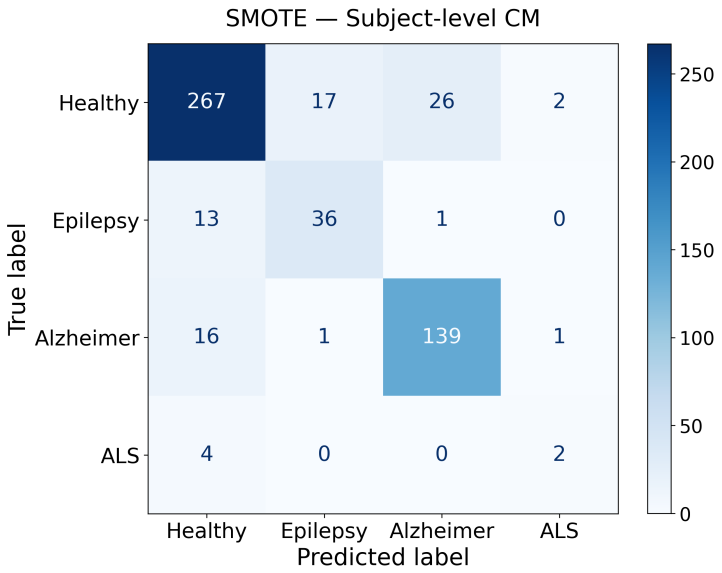

Figure 4.13. SMOTE: Confusion matrix.

These findings suggest that the linear interpolation strategy adopted by SMOTE is not well suited for the handcrafted EEG descriptors employed in this work, as the generated synthetic samples do not accurately preserve the complex distribution of pathological EEG features.

Compared with SMOTE, CTGAN achieved a considerably higher Balanced Accuracy, indicating a more balanced recognition of the different neurological classes despite a lower overall Accuracy.

Inspection of the CTGAN confusion matrix demonstrates a substantially better ability to preserve discrimination among minority neurological disorders. Compared with SMOTE, CTGAN generated fewer misclassifications involving pathological classes, indicating that synthetic samples better approximated the original feature distribution.

The CTGAN subject-level confusion matrix is reported in Figure 4.14.

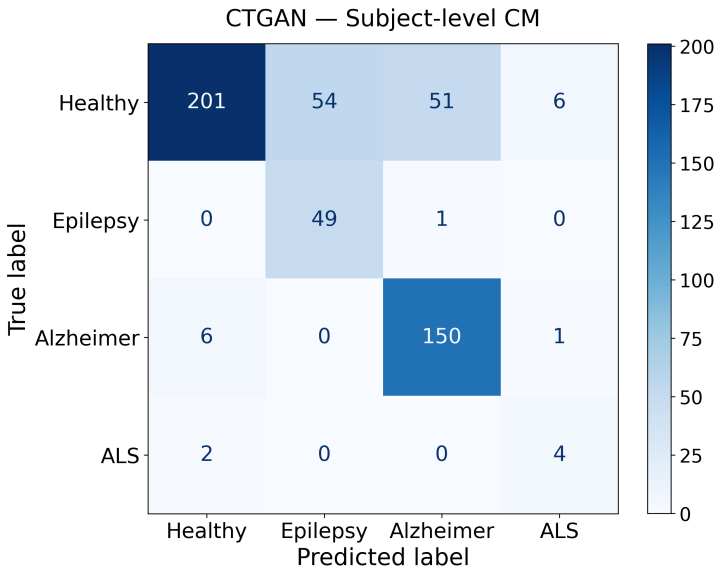

Figure 4.14. CTGAN: Confusion Matrix.

Overall, the data augmentation experiments indicate that neither SMOTE nor CTGAN was able to outperform the classifier trained exclusively on real EEG recordings. Although CTGAN produced more balanced predictions than SMOTE, the performance remained inferior to that achieved by the proposed leakage-free hierarchical cascade without artificial data augmentation.

These findings suggest that the handcrafted EEG descriptors extracted in this work already provide a sufficiently informative representation of the underlying neurological conditions. Consequently, generating additional synthetic samples does not substantially improve the generalization capability of the classifier. Moreover, the results indicate that preserving the original feature distribution is more beneficial than artificially increasing the number of training samples when dealing with heterogeneous EEG datasets collected under different acquisition protocols.

The obtained results also highlight that data augmentation should not be considered a universal solution for class imbalance. While synthetic data generation has proved effective in several biomedical applications, its effectiveness strongly depends on the characteristics of the extracted feature space. In the present study, the handcrafted statistical EEG descriptors appear to capture disease-specific information that is difficult to reproduce through synthetic feature generation.

4.6 Three-class Classification

The experiments presented so far highlighted that the ALS class represents the main limitation of the proposed multiclass framework due to the extremely limited number of available subjects. Although the leakage-free hierarchical cascade substantially improved the overall diagnostic process, directly discriminating Alzheimer's disease, epilepsy and ALS remains particularly challenging because ALS is represented by only six subjects.

To mitigate this limitation, an additional clinically-guided hierarchical classification strategy was investigated: Alzheimer's disease and ALS were initially merged into a common neurodegenerative class, whereas epilepsy was maintained as an independent neurological disorder.

The resulting framework consists of two consecutive classification stages. The first classifier distinguishes among healthy subjects, epilepsy and neurodegenerative diseases, while the second classifier further discriminates Alzheimer's disease from ALS only within the neurodegenerative group.

The same experimental protocol adopted throughout the previous experiments was employed. Feature selection and Optuna-based hyperparameter optimization were independently applied to both classification stages.

Method	**Balanced Accuracy (%)**	**Macro F1**
Baseline	83.67 ± 4.20	$\mathbf{0.730 \pm 0.054}$
Top-125 Features	$\mathbf{84.89 \pm 1.30}$	0.713 ± 0.038
Optuna	84.23 ± 3.76	0.707 ± 0.049
Optuna + Top-125 Features	84.33 ± 1.29	0.715 ± 0.038

Table 4.11. Subject-level performance obtained for the first stage of the three–class hierarchical framework (Healthy vs Epilepsy vs Neurodegenerative Diseases).

Method	**Balanced Accuracy (%)**	**Macro F1**
Baseline	98.08 ± 1.57	0.851 ± 0.112
Top-40 Features	99.68 ± 0.90	0.965 ± 0.049
Optuna	97.45 ± 1.20	0.865 ± 0.098
Optuna + Top-40 Features	$\mathbf{99.68 \pm 0.45}$	$\mathbf{0.965 \pm 0.049}$

Table 4.12. Subject-level performance obtained for the second stage of the three-class hierarchical framework (Alzheimer's disease vs ALS).

The subject-level results demonstrate that incorporating prior clinical knowledge into the hierarchical organization of neurological disorders effectively simplifies the classification problem.

During the first stage, the proposed framework successfully separates healthy subjects, epilepsy and neurodegenerative diseases, achieving a Balanced Accuracy of 84.89% after feature selection. Similarly to the previous experiments, feature selection provided the largest performance improvement, whereas hyperparameter optimization produced only marginal additional benefits.

Once neurodegenerative subjects have been identified, the second classification stage achieves an almost perfect discrimination between Alzheimer's disease and ALS, reaching a subject-level Balanced Accuracy of 99.68%. This result indicates that separating the broad diagnostic problem into two consecutive and clinically meaningful decisions substantially reduces the complexity associated with directly discriminating all neurological disorders simultaneously.

Compared with the leakage-free hierarchical cascade presented in the previous section, the proposed strategy benefits from the fact that Alzheimer's disease and ALS share common neurodegenerative characteristics. By first grouping these disorders into a single class, the first classifier focuses on recognizing broader pathological patterns, leaving the more challenging discrimination between Alzheimer's disease and ALS to a dedicated classifier.

Figures 4.15 and 4.16 illustrate the confusion matrices obtained for the first stage of the proposed hierarchical framework, where healthy controls, epilepsy and neurodegenerative diseases are discriminated. After feature selection, the confusion matrix shows a reduction in misclassified subjects, confirming that selecting the most informative EEG features improves the separation among the three groups.

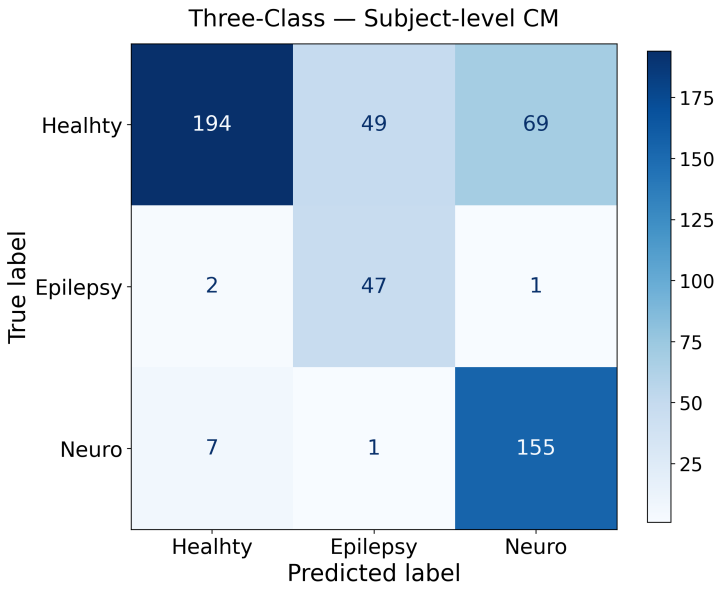

Figure 4.15. Three-Class: Baseline Confusion Matrix.

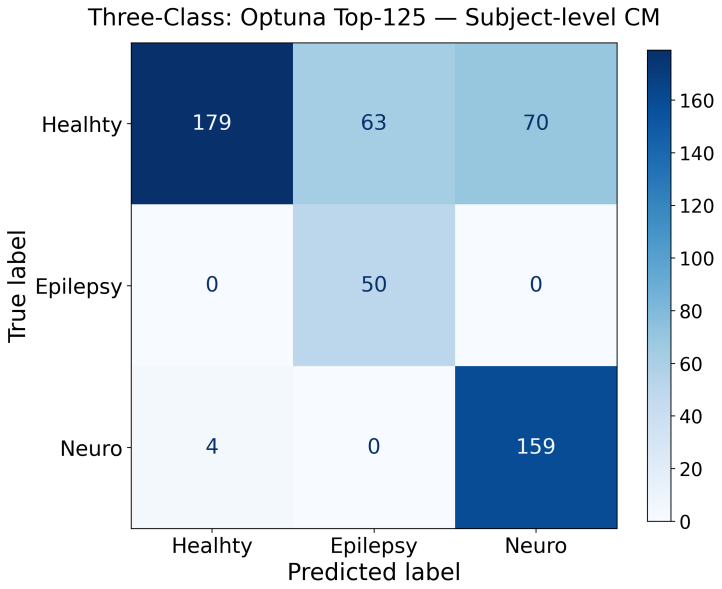

Figure 4.16. Three-Class: Baseline Confusion Matrix using Optuna + Top-125.

Figures 4.17 and 4.18 report the confusion matrices for the second stage, which focuses exclusively on the discrimination between Alzheimer's disease and ALS. In

agreement with the quantitative results reported in Table 4.12, only a very limited number of misclassifications is observed, with the optimized model achieving an almost perfect separation between the two neurodegenerative conditions. This finding suggests that, once broader pathological categories have been identified, the EEG features extracted by the proposed framework are sufficiently discriminative to reliably distinguish Alzheimer's disease from ALS despite the severe class imbalance.

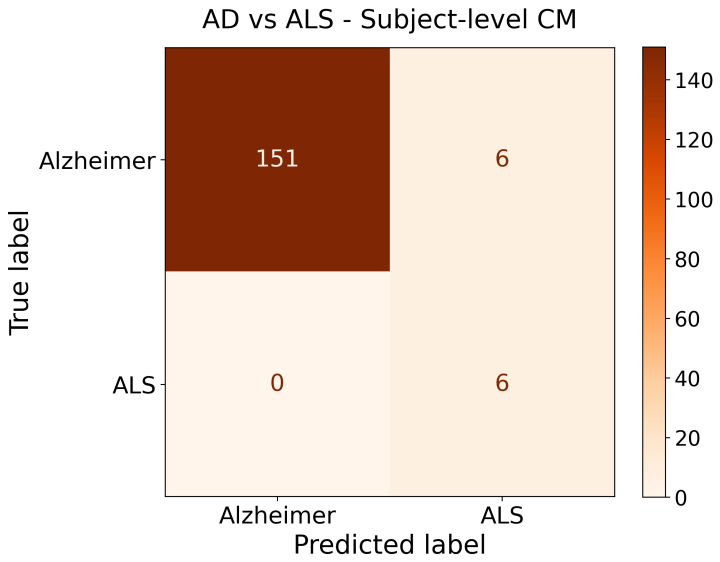

Figure 4.17. AD vs ALS: Baseline Confusion Matrix.

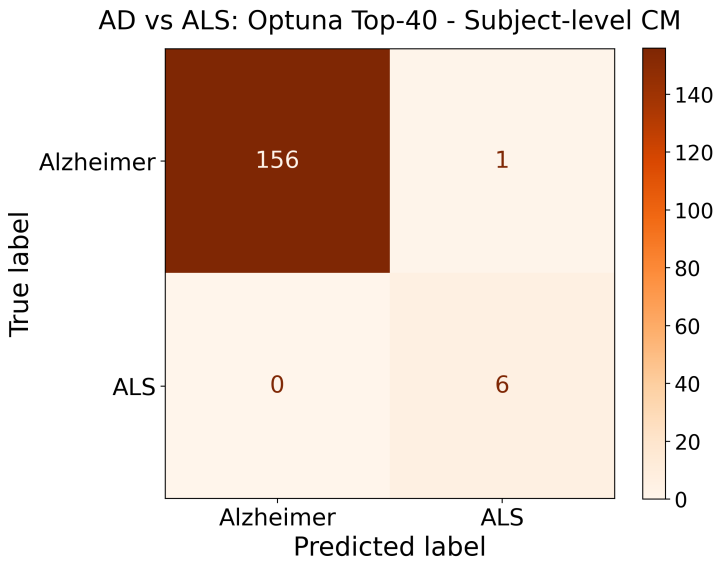

Figure 4.18. AD vs ALS: Baseline Confusion Matrix using Optuna + Top-40.

Overall, these experiments demonstrate that incorporating clinical knowledge into the hierarchical organization of neurological disorders represents an effective strategy for addressing severe class imbalance. Rather than directly solving a highly complex multiclass problem, decomposing the diagnostic task into clinically meaningful stages substantially improves the discrimination of rare neurological conditions while maintaining excellent overall classification performance. These findings further support the potential of hierarchical machine learning approaches for EEG-based neurological disease classification.

4.7 Comparison of the Proposed Strategies

This section summarizes the performance obtained by the different classification strategies investigated throughout this work. Rather than focusing on individual experiments, the objective is to provide a global comparison of the proposed approaches and highlight their respective strengths and limitations.

Since the considered datasets exhibit different degrees of class imbalance, Balanced Accuracy was adopted as the primary evaluation metric. This metric provides a more reliable estimate of the classifier performance than conventional Accuracy, particularly for the ALS dataset and the multiclass classification problems, where the number of healthy subjects considerably exceeds the number of pathological subjects.

Classification Task	**Best Strategy**	**Balanced Accuracy**
Binary Classification		
Alzheimer's Disease	Top-50 Features + Optuna	**92.02%**
Epilepsy	Top-50 Features + Optuna	**87.05%**
Amyotrophic Lateral Sclerosis	Top-20 Features	**91.67%**
Direct Multiclass Classification		
Standard Multiclass	Top-125 Features	**85.33%**
Hierarchical Classification		
Healthy vs Disease	Top-100 Features	**79.08%**
Disease-Specific Classification	Top-50 Features	**99.36%**
Leakage-Free Hierarchical Cascade	Top-150 Features	**83.83%**
Three-Class Strategy	Top-125 Features	**84.89%**
AD vs ALS	Top-40 Features + Optuna	**99.68%**
Data Augmentation		
SMOTE	SVM	**75.97%**
CTGAN	SVM	**82.04%**

Table 4.13. Overall comparison of the best subject-level balanced accuracy obtained for the different classification strategies investigated in this work.

Table 4.13 summarizes the highest Balanced Accuracy obtained for each classification task investigated throughout this work.

Several common trends can be identified. First, feature selection consistently represented the most influential stage of the proposed classification pipeline. In nearly all experiments, selecting a reduced subset of highly discriminative EEG descriptors substantially improved the classifier performance while simultaneously reducing the dimensionality of the feature space. This observation indicates that the extracted statistical EEG features contain a significant amount of redundant information and that a relatively small subset of descriptors is sufficient to accurately characterize the considered neurological disorders.

Hyperparameter optimization using Optuna generally provided additional improvements, although its contribution was less pronounced than that of feature selection. In most cases, the combination of Top-K feature selection and Optuna produced the highest overall Balanced Accuracy, confirming the complementary nature of these two optimization stages.

The comparison among the different classification frameworks also highlights the progressive evolution of the proposed methodology. While the standard multiclass classifier demonstrated the feasibility of simultaneously discriminating multiple neurological disorders, the subsequent hierarchical approaches provided a more realistic representation of the clinical diagnostic workflow.

The leakage-free hierarchical cascade confirmed that excellent classification performance can be maintained even when disease-specific classification is performed exclusively on subjects predicted as pathological by an independent screening classifier. Furthermore, the three-class strategy demonstrated that incorporating prior medical knowledge into the organization of neurological disorders can simplify the classification problem, particularly for underrepresented diseases such as ALS.

Finally, the experiments involving SMOTE and CTGAN showed that synthetic data augmentation did not improve the classification performance obtained using the original EEG recordings. Although CTGAN produced more balanced results than SMOTE, neither augmentation strategy outperformed the classifiers trained exclusively on real data, suggesting that the proposed statistical EEG descriptors already provide a highly informative representation of the considered neurological disorders.

The experimental results presented throughout this chapter demonstrate that the proposed EEG-based framework is capable of accurately discriminating multiple neurological disorders using handcrafted statistical spectral features extracted from heterogeneous public datasets. The progressive refinement of the classification pipeline, including feature selection, hyperparameter optimization and clinically-guided hierarchical organization, consistently improved the robustness and interpretability of the proposed models.

4.8 Comparison with State-of-the-Art Studies

The experimental results presented in the previous sections demonstrate the effectiveness of the proposed EEG-based framework for the classification of neurological disorders under binary, multiclass and hierarchical classification scenarios. To better contextualize these findings, the proposed methodology is compared with representative studies selected from the literature.

Unlike the Related Work presented in Chapter 2, which provides a general overview of previous EEG-based machine learning approaches, this section also shows the results of the previous EEG-based neurological disease classification studies.

The comparison reported in Table 4.14 highlights several methodological differences between the proposed framework and previous EEG-based neurological disease classification studies.

A first important difference concerns the scope of the investigated classification problems. Most existing studies focus on a single neurological disorder and formulate the problem as a binary classification task between healthy controls and patients. In contrast, the proposed framework investigates three distinct neurological disorders, namely Alzheimer's disease, epilepsy and Amyotrophic Lateral Sclerosis, within a unified computational pipeline. Besides conventional binary classification, multiclass, hierarchical and leakage-free cascade classification strategies are systematically evaluated.

Another relevant aspect is the adopted validation methodology. Several previous studies employ conventional cross-validation strategies or leave-one-subject-out validation. Although these approaches provide reliable performance estimates, they generally investigate only a single classification scenario. In the proposed work, all experiments were evaluated using a subject-independent Nested GroupKFold cross-validation protocol, ensuring that EEG segments belonging to the same subject never appear simultaneously in the training and testing sets. Furthermore, feature selection and hyperparameter optimization were performed exclusively within the training folds, preventing information leakage during model development.

Study	**Subjects**	**Validation**	**Result**
Akbar et al. (2026) [22]	36 AD/29 HC/ 23 FTD	Nested CV	ACC: 87.8% F1-Macro: 0.85 AUC: 0.88
Perez-Valero et al. (2023) [23]	8 AD/8 HC	LOSO CV	ACC: 87%
Chu et al. (2023) [24]	112 AD/93 HC	Cross-validation	ACC: 81% SEN: 82% SPE: 80%
Tran et al. (2022) [25]	5 EP/5 HC	Cross-validation	ACC: 98.4% PRE: 96% REC: 96%
Lih et al. (2023) [26]	50 EP/71 HC	10-fold CV	ACC: 85% SEN: 82% SPE: 87%
Liu et al. (2024) [27]	24 EP	Cross-validation	ACC: 98.13% SEN: 98.03% SPE: 98.23%
Ngo et al. (2024) [28]	6 ALS/170 HC	Cross-validation	ACC: 89.4%
Proposed study	157 AD/74 HC	Nested GroupKFold CV	ACC: 92.22% BAC: 92.02%
Proposed study	50 EP/71 HC	Nested GroupKFold CV	ACC: 87.63% BAC: 87.05%
Proposed study	6 ALS/167 HC	Nested GroupKFold CV	ACC: 99.42% BAC: 91.67%
Proposed study	157 AD/50 EP/ 6 ALS/ 312 HC	Nested GroupKFold CV	ACC: 73.72% BAC: 83.83% F1-Macro: 0.627 AUC: 0.878

Table 4.14. Comparison between representative state-of-the-art EEG-based neurological disease classification studies and the proposed framework. Reported results are not directly comparable because of differences in datasets, datasets sizes, EEG acquisition protocols, validation strategies, task complexity and evaluation metrics.

The proposed framework also differs from previous studies in terms of dataset heterogeneity. Most published works are based on a single EEG dataset acquired under homogeneous experimental conditions. Conversely, this study integrates multiple publicly available datasets characterized by different acquisition protocols, recording devices, signal formats and numbers of EEG channels. Although this considerably increases the complexity of the classification task, it also provides a more realistic assessment of model robustness and generalization capability.

From the experimental perspective, the obtained classification performances are competitive with those reported in the literature. Nevertheless, direct numerical comparisons should be interpreted with caution because different studies employ different datasets, preprocessing pipelines, feature extraction procedures and evaluation protocols. In particular, several published works report very high classification accuracies using relatively small datasets or homogeneous acquisition settings, conditions that may facilitate classification but reduce generalizability.

Overall, the proposed framework achieves competitive performance while adopting a rigorous validation protocol and addressing substantially more challenging classification problems than those commonly investigated in previous EEG-based neurological disease classification studies.

Chapter 5

Discussion

5.1 Main Findings

The experimental results presented in Chapter 4 demonstrate that the proposed EEG-based framework is capable of accurately discriminating multiple neurological disorders using handcrafted statistical spectral features extracted from heterogeneous public datasets. Despite relying on a relatively simple machine learning architecture based on Support Vector Machines, the proposed methodology consistently achieved competitive classification performance across all investigated tasks.

One of the most significant findings emerging from this work concerns the effectiveness of feature selection. Although several hundred statistical descriptors were initially extracted from each EEG recording, only a relatively small subset of highly discriminative features was required to maximize classification performance. This behaviour was consistently observed across all binary, multiclass and hierarchical classification experiments. In most cases, reducing the dimensionality of the feature space not only simplified the resulting models but also improved their generalization capability, suggesting that a substantial proportion of the extracted features contained redundant or weakly informative information.

Another important observation concerns the relative contribution of hyperparameter optimization. While Optuna generally produced additional improvements over the baseline classifiers, its impact was consistently smaller than that obtained through feature selection. These findings indicate that, for the proposed framework, the quality of the EEG representation plays a more important role than the fine optimization of the classifier parameters. Consequently, the identification of informative EEG descriptors should be considered the primary factor determining classification performance.

The experiments also highlighted the importance of selecting appropriate evaluation metrics. Throughout the study, Balanced Accuracy proved to be considerably more informative than conventional Accuracy due to the severe class imbalance characterizing several of the investigated datasets, particularly those including ALS patients. In multiple experiments, high Accuracy values were accompanied by substantially lower Balanced Accuracy, indicating that conventional Accuracy alone may provide an overly optimistic estimation of classifier performance. This observation further justifies the choice of Balanced Accuracy as the primary evaluation metric adopted throughout this work.

The progressive evolution of the proposed classification framework provides further useful insights into the complexity of automatic diagnosis of neurological diseases. The standard multiclass classifier demonstrated that simultaneous discrimination between different neurological diseases is possible using manually defined EEG features. However, subsequent experiments showed that breaking the diagnostic process into simpler hierarchical steps substantially improves both the interpretability and robustness of the overall framework.

In particular, the leakage-free hierarchical cascade demonstrated that reliable disease-specific classification is possible even when the second-level classifier receives only subjects identified as pathological during the initial screening phase. Furthermore, the three-class strategy demonstrated that incorporating prior medical knowledge into the organization of neurological diseases is an effective approach to mitigate severe class imbalance while maintaining excellent diagnostic performance.

Overall, the experimental analysis demonstrates that accurate EEG-based neurological disease classification does not necessarily require highly complex deep learning architectures. Instead, carefully designed statistical feature extraction, rigorous validation procedures, appropriate feature selection and clinically meaningful hierarchical classification strategies can together provide highly competitive diagnostic performance while maintaining excellent model interpretability.

5.2 Clinical Implications

The results obtained in this work have several implications for the development of computer-aided diagnostic systems based on EEG recordings.

First, the proposed framework confirms the potential of EEG as a non-invasive and cost-effective tool for supporting the diagnosis of neurological disorders. These characteristics make EEG particularly attractive as a first-line screening modality, especially in healthcare environments with limited access to advanced neuroimaging technologies.

A second important aspect concerns the interpretability of the proposed methodology. Unlike many recent deep learning approaches, whose decision-making process is often difficult to explain, the framework developed in this work is based on handcrafted statistical spectral descriptors that possess a clear neurophysiological interpretation. Furthermore, the use of a linear Support Vector Machine allows the contribution of each feature to be quantified through the analysis of the model coefficients, making it possible to identify the EEG characteristics that most strongly influence the classification process. Such interpretability is particularly desirable in clinical applications, where understanding the rationale behind an automated decision can increase clinicians' confidence in the system.

Another relevant observation concerns the hierarchical organization of the proposed classification framework. Clinical diagnosis rarely consists of a single decision; instead, physicians progressively reduce the range of possible diagnoses by combining clinical evidence collected at different stages of the diagnostic process. The hierarchical strategies investigated in this work reproduce this reasoning. The initial classification stages identify whether a subject is healthy or affected by a neurological disorder, while the subsequent classifiers progressively differentiate among specific diseases. This organization provides a workflow that is more consistent with routine clinical practice.

Finally, the experimental results demonstrate that high diagnostic performance can be achieved without relying on highly complex machine learning architectures. The combination of carefully designed EEG preprocessing, statistical feature extraction, rigorous validation procedures and hierarchical classification proved sufficient to accurately discriminate multiple neurological disorders. This finding suggests that relatively simple and interpretable machine learning models remain highly competitive for EEG-based diagnostic applications, particularly when limited amounts of training data are available.

5.3 Limitations

Despite the promising results obtained throughout the experimental evaluation, several limitations should be acknowledged.

The first limitation concerns the size and distribution of the available datasets. Although the proposed framework was evaluated on multiple publicly available EEG databases comprising more than five hundred subjects, the individual datasets remain relatively small when considered separately. This limitation is particularly evident for Amyotrophic Lateral Sclerosis, where only six patients were available. Such an imbalance inevitably affects the reliability of the statistical estimates and represents one of the main challenges addressed throughout this work. While the proposed hierarchical strategies partially mitigated this issue, larger and more balanced datasets would be necessary to further validate the generalization of the proposed framework.

A second limitation is related to the heterogeneity of the considered datasets. The EEG recordings were collected by different research groups using different acquisition protocols, recording devices, electrode configurations and experimental conditions. Although a common preprocessing pipeline was adopted to minimize these differences, residual variability among datasets cannot be completely eliminated. Consequently, part of the classification performance may still be influenced by dataset-specific characteristics rather than exclusively by disease-related EEG alterations.

Another important limitation concerns the adopted feature representation. The proposed methodology relies entirely on handcrafted statistical spectral descriptors extracted from predefined EEG frequency bands. Although these features demonstrated excellent discriminative capacity and offer a high degree of interpretability, they may not fully capture the complex temporal and spatial dynamics of EEG activity.

Finally, the proposed framework was validated exclusively using publicly available EEG recordings acquired under controlled experimental conditions. Before being considered for routine clinical use, the methodology should be evaluated on larger prospective cohorts collected in real clinical environments, where EEG recordings are typically characterized by greater variability, different patient populations and a wider range of recording conditions.

Chapter 6

Conclusions and Future Research

This study presented a complete machine learning framework for the automatic classification of neurological disorders using EEG recordings. The proposed methodology integrates standardized EEG preprocessing, statistical spectral feature extraction, feature selection, hyperparameter optimization and hierarchical classification strategies into an interpretable diagnostic pipeline.

The experimental evaluation was conducted on multiple publicly available EEG datasets including healthy subjects together with patients affected by Alzheimer's disease, epilepsy and Amyotrophic Lateral Sclerosis. Unlike many previous studies focusing on a single neurological disorder, this work investigated both binary and multiclass classification scenarios, progressively extending the proposed framework towards increasingly realistic diagnostic settings.

The obtained results demonstrate that handcrafted statistical EEG features provide a reliable representation for neurological disease classification. Across all investigated tasks, feature selection consistently represented the most influential stage of the classification pipeline, significantly improving the Balanced Accuracy while simultaneously reducing the dimensionality of the feature space. Hyperparameter optimization further refined the classifiers, although its contribution was generally smaller than that achieved through feature selection.

The experimental analysis also highlighted the importance of adopting appropriate validation strategies. The use of Nested GroupKFold cross-validation ensured subject-independent evaluation and prevented information leakage between training and testing partitions, providing a rigorous assessment of the proposed models. Furthermore, Balanced Accuracy proved to be the most informative performance metric for the considered datasets, where substantial class imbalance

could otherwise produce overly optimistic estimates when relying solely on conventional Accuracy.

Several classification strategies were investigated throughout this work. Besides conventional binary and multiclass classification, hierarchical approaches were proposed to progressively simplify the diagnostic problem. The leakage-free hierarchical cascade demonstrated that disease-specific classification can be successfully performed following an initial pathological screening stage while preserving high diagnostic performance. Moreover, the clinically-guided hierarchical framework, inspired by the natural organization of neurological disorders, further improved the classification process by grouping related diseases before performing disease-specific discrimination. This strategy proved particularly effective for underrepresented conditions such as Amyotrophic Lateral Sclerosis and represents the main methodological contribution of this work.

Additional experiments involving synthetic data augmentation through SMOTE and CTGAN showed that generating artificial feature vectors did not improve the performance obtained using real EEG recordings. These findings suggest that, for the handcrafted statistical descriptors considered in this work, careful feature engineering and rigorous model design play a more important role than synthetic data generation.

Overall, the results obtained throughout this study demonstrate that accurate EEG-based neurological disease classification does not necessarily require highly complex deep learning architectures. Instead, a carefully designed pipeline combining robust preprocessing, interpretable statistical descriptors, rigorous validation procedures and clinically meaningful hierarchical classification can achieve excellent diagnostic performance while maintaining transparency and interpretability. These characteristics make the proposed framework a promising decision-support tool for future EEG-based clinical applications.

6.1 Future Research

Although the proposed framework achieved promising results across multiple neurological disorders, several directions can be explored to further extend the present work.

A first research direction concerns the inclusion of additional neurological and neurodegenerative diseases. The current study focused on Alzheimer's disease, epilepsy and Amyotrophic Lateral Sclerosis; however, the hierarchical classification framework is sufficiently general to accommodate a broader range of neurological conditions. Extending the proposed methodology to larger clinical cohorts involving additional pathologies would allow the development of a more comprehensive EEG-based diagnostic support system.

Another important direction involves the validation of the proposed framework on larger and more balanced datasets. In particular, the limited number of ALS patients represents one of the main constraints of the present study. Future investigations based on larger clinical databases would provide a more reliable assessment of the robustness and generalization capability of the proposed models.

A particularly promising extension concerns the integration of multimodal physiological information. Preliminary investigations conducted during this work explored the combination of EEG and eye-tracking signals for the analysis of ALS patients. Although this multimodal evaluation was intentionally kept outside the main scope of the work due to the limited availability of eye-tracking recordings, the preliminary results indicated an improvement of approximately 9% in Balanced Accuracy compared with the EEG-only baseline. Interestingly, eye-tracking alone was not sufficient to accurately discriminate the considered neurological conditions. However, when combined with EEG, it consistently improved the classification performance, suggesting that behavioural information complements rather than replaces electrophysiological measurements.

Finally, the proposed framework is not limited to neurological disease diagnosis. The combination of standardized EEG preprocessing, statistical feature extraction, feature selection and hierarchical machine learning can be naturally extended to several EEG-based pattern recognition problems, including cognitive state assessment, mental workload estimation, emotion recognition and human action recognition. These applications demonstrate that the methodology developed in this work may represent a general and flexible framework for EEG signal analysis beyond the clinical domain.

In conclusion, the framework proposed in this study provides a solid foundation for future developments in EEG-based artificial intelligence systems. By combining interpretable machine learning techniques with clinically meaningful classification strategies, it contributes to the development of reliable, transparent and scalable

decision-support tools for both neurological diagnosis and broader EEG-based applications.

Bibliography

[1] World Health Organization. Global status report on neurology. Technical report, World Health Organization, Geneva, Switzerland, October 2025. Accessed: 2026-07-12.

[2] N. Bogdanovic, O. Hansson, H. Zetterberg, H. Basun, M. Ingelsson, L. Lannfelt, and K. Blennow. Alzheimer's disease – the most common cause of dementia. *Lakartidningen*, 117:FZHM, March 2020. Originally published in Swedish: "Alzheimers sjukdom – diagnostik och behandling i dag och i framtiden".

[3] H. Anwar, Q. U. Khan, N. Nadeem, I. Pervaiz, M. Ali, and F. F. Cheema. Epileptic seizures. *Discoveries (Craiova)*, 8(2):e110, 2020.

[4] J. Ravits, D. Ferrey, B. Gundogdu, W. Qayoumi, and C. Zale. Amyotrophic lateral sclerosis: A review. *JAMA*, 335(22):1970–1982, June 2026.

[5] JiaoJiao Guo and Christos Panagiotis Lisgaras. The use of electroencephalography in neurodegenerative disease and its utility in dementia. *npj Dementia*, 2(1):36, 2026.

[6] Gregory A. Light, Leanne E. Williams, F. Minow, J. Sprock, Anthony J. Rissling, R. Sharp, Neal R. Swerdlow, and David L. Braff. Electroencephalography (eeg) and event-related potentials (erps) with human participants. *Current Protocols in Neuroscience*, 52(1):Unit 6.25, 2010.

[7] Ahmad Chaddad, Yifan Wu, R. Kateb, and Ahmed Bouridane. Electroencephalography signal processing: A comprehensive review and analysis of methods and techniques. *Sensors*, 23(14):6434, 2023.

[8] Rui Lyu. Deep learning approaches for eeg-based healthcare applications: A comprehensive review. *Frontiers in Human Neuroscience*, 19:1689073, 2026.

[9] Wanzeng Kong and Xuanyu Jin. *Basics of EEG Signals*, pages 15–31. Springer Nature Singapore, Singapore, 2025.

[10] Nicole Greig. What are brain waves? 5 types of brain waves and their functions, 2024.

[11] J. Fernandez, B. Innocenti, and B. López. EEG classification for neurological disorders using frequency band deciles. *Scientific Reports*, 15(1):45142, 2025.

[12] Chaitanya S. Nayak and A. C. Anilkumar. *Normal EEG Waveforms*. StatPearls Publishing, Treasure Island, FL, 2026. Updated August 3, 2025.

[13] C. Simfukwe, S. S. A. An, and Y. C. Youn. Time-frequency domain analysis of quantitative electroencephalography as a biomarker for dementia. *Diagnostics*, 15(12):1509, 2025.

[14] Prasanna D. Emmady, Robert M. D. Asuncion, and A. C. Anilkumar. *EEG Abnormal Waveforms*. StatPearls Publishing, Treasure Island, FL, 2026. Updated April 6, 2025.

[15] J. Chmiel and M. Stępień-Słodkowska. Resting-state EEG oscillations in amyotrophic lateral sclerosis (ALS): Toward mechanistic insights and clinical markers. *Journal of Clinical Medicine*, 14(2):545, 2025.

[16] Hakan Uyanik, Abdulkadir Sengur, Massimo Salvi, Ru-San Tan, Jen Hong Tan, and U. Rajendra Acharya. Automated detection of neurological and mental health disorders using eeg signals and artificial intelligence: A systematic review. *WIREs Data Mining and Knowledge Discovery*, 15(1):e70002, 2025. e70002 DMKD-00820.R1.

[17] Stefania Coelli, Alessandra Calcagno, Chiara Maria Cassani, Federico Temporiti, Pierluigi Reali, Roberto Gatti, Manuela Galli, and Anna Maria Bianchi. Selecting methods for a modular eeg pre-processing pipeline: An objective comparison. *Biomedical Signal Processing and Control*, 90:105830, 2024.

[18] A. K. Singh and S. Krishnan. Trends in EEG signal feature extraction applications. *Frontiers in Artificial Intelligence*, 5:1072801, 2023.

[19] Rosita Guido, Stefania Ferrisi, Danilo Lofaro, and Domenico Conforti. An overview on the advancements of support vector machine models in healthcare applications: A review. *Information*, 15(4), 2024.

[20] Surani Matharaarachchi, Mike Domaratzki, and Saman Muthukumarana. Enhancing smote for imbalanced data with abnormal minority instances. *Machine Learning with Applications*, 18:100597, 2024.

[21] Pankaj Yadav, Gulshan Sihag, and Vivek Vijay. Rebalancing the scales: A systematic mapping study of generative adversarial networks (GANs) in addressing data imbalance. *arXiv preprint arXiv:2502.16535*, 2025. Version 1, submitted February 23, 2025.

[22] F. Akbar, Y. Alkhrijah, S. M. Usman, S. Khalid, I. Ihsan, and M. A. Alawad. A deep-svm hybrid framework with enhanced eeg feature engineering and shap-based explainability for alzheimer's classification. *Scientific Reports*, 2026. Online ahead of print.

[23] Eduardo Perez-Valero, Christian Morillas, Miguel A. Lopez-Gordo, and Jesus Minguillon. Supporting the detection of early alzheimer's disease with a four-channel eeg analysis. *International Journal of Neural Systems*, 33(04):2350021, 2023. PMID: 36803195.

[24] Kwo-Ta Chu, Weng-Chi Lei, Ming-Hsiu Wu, Jong-Ling Fuh, Shuu-Jiun Wang, Isobel T. French, Wen-Sheng Chang, Chi-Fu Chang, Norden E. Huang, Wei-Kuang Liang, and Chi-Hung Juan. A holo-spectral eeg analysis provides an early detection of cognitive decline and predicts the progression to alzheimer's disease. *Frontiers in Aging Neuroscience*, Volume 15 - 2023, 2023.

[25] Ly V. Tran, Hieu M. Tran, Tuan M. Le, Tri T. M. Huynh, Hung T. Tran, and Son V. T. Dao. Application of machine learning in epileptic seizure detection. *Diagnostics*, 12(11), 2022.

[26] Oh Shu Lih, V. Jahmunah, Elizabeth Emma Palmer, Prabal D. Barua, Sengul Dogan, Turker Tuncer, Salvador García, Filippo Molinari, and U Rajendra Acharya. Epilepsynet: Novel automated detection of epilepsy using transformer model with eeg signals from 121 patient population. *Computers in Biology and Medicine*, 164:107312, 2023.

[27] Xin Liu, Chunyang Li, Xicheng Lou, Haohuan Kong, Xinwei Li, Zhangyong Li, and Lisha Zhong. Epileptic seizure prediction based on eeg using pseudo-three-dimensional cnn. *Frontiers in Neuroinformatics*, Volume 18 - 2024, 2024.

[28] Thi Duyen Ngo, Hai Dang Kieu, Minh Hoa Nguyen, Hoang-Anh Nguyen The, Van Mao Can, Ba Hung Nguyen, and Thanh Ha Le. An eeg & eye-tracking dataset of als patients & healthy people during eye-tracking-based spelling system usage. *Scientific Data*, 11(1):664, 2024.

[29] Andreas Miltiadous, Katerina D. Tzimourta, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Dimitrios G. Tsalikakis, Pantelis Angelidis, Markos G. Tsipouras, Evripidis Glavas, Nikolaos Giannakeas, and Alexandros T. Tzallas. "a dataset of eeg recordings from: Alzheimer's disease, frontotemporal dementia and healthy subjects", 2024.

[30] Mojtaba Lahijanian, Hamid Aghajan, and Zahra Vahabi. "40hz auditory entrainment", 2025.

[31] Mário L. Vicchietti, Fernando M. Ramos, Luiz E. Betting, and Andriana S. L. O. Campanharo. Computational methods of eeg signals analysis for alzheimer's disease classification, 2024. Project repository.

[32] Aruane Pineda and Caroline Alves. Alzheimer's disease and schizophrenia, 2022. Dataset.

[33] Irem Tasci, Burak Tasci, Prabal D. Barua, Sengul Dogan, Turker Tuncer, Elizabeth Emma Palmer, Hamido Fujita, and U. Rajendra Acharya. Epilepsy detection in 121 patient populations using hypercube pattern from eeg signals. *Information Fusion*, 96:252–268, 2023.

[34] Sayantan Acharya, Abbas Khosravi, Douglas Creighton, Roohallah Alizadehsani, and U Rajendra Acharya. Neurostressology: A systematic review of eeg-based automated mental stress perspectives. *Information Fusion*, 124:103368, 2025.

[35] Pranay Rahi and Dr. Rajesh Mehra. Analysis of power spectrum estimation using welch method for various window techniques. *International Journal of Emerging Technologies and Engineering*, pages 106–109, 08 2014.

Appendix A

Additional Experimental Details

A.1 Optuna Hyperparameter Optimization

This appendix reports the optimal hyperparameter configurations identified by Optuna during the nested cross-validation experiments.

For each outer cross-validation fold, Optuna optimized the Support Vector Machine (SVM) hyperparameters by maximizing the mean Balanced Accuracy obtained across the inner `GroupKFold` validation folds.

The search space adopted throughout all experiments was defined as follows:

- Regularization parameter: $C \in [10^{-3}, 10^{3}]$;
- Kernel: linear or rbf;
- RBF kernel parameter: $\gamma \in [10^{-4}, 10^{1}]$. When the linear kernel was selected, the default value scale was adopted.

For each outer fold, 50 optimization trials were performed using Optuna's MedianPruner strategy.

A.2 Binary Classification

A.2.1 Alzheimer's Disease

Feature Set	Fold	Kernel	C	γ
All	1	RBF	2.966	3.78×10^{-4}
All	2	RBF	58.647	5.62×10^{-4}
All	3	RBF	169.798	1.52×10^{-3}
All	4	Linear	2.02×10^{-3}	–
All	5	RBF	27.526	6.50×10^{-4}
Top-50	1	Linear	0.461	–
Top-50	2	Linear	0.0127	–
Top-50	3	Linear	0.0200	–
Top-50	4	Linear	0.0645	–
Top-50	5	RBF	28.295	4.12×10^{-4}

Table A.1. Best hyperparameter configurations identified by Optuna for the Alzheimer's disease classification experiments.

A.2.2 Epilepsy

Feature Set	Fold	Kernel	C	γ
All	1	RBF	30.222	1.01×10^{-3}
All	2	RBF	1.476	2.80×10^{-3}
All	3	RBF	10.893	2.97×10^{-4}
All	4	RBF	4.164	6.72×10^{-4}
All	5	Linear	3.44×10^{-3}	–
Top-50	1	Linear	0.136	–
Top-50	2	Linear	0.0410	–
Top-50	3	Linear	0.0401	–
Top-50	4	Linear	0.0221	–
Top-50	5	Linear	0.0321	–

Table A.2. Best hyperparameter configurations identified by Optuna for the epilepsy classification experiments.

A.2.3 Amyotrophic Lateral Sclerosis

Feature Set	Fold	Kernel	C	γ
All	1	RBF	0.0272	8.54×10^{-4}
All	2	–	–	–
All	3	RBF	0.0737	1.71×10^{-4}
Top-20	1	RBF	0.253	8.77×10^{-2}
Top-20	2	–	–	–
Top-20	3	Linear	5.01×10^{-3}	–

Table A.3. Best hyperparameter configurations identified by Optuna for the ALS classification experiments.

A.3 Multiclass Classification

A.3.1 Standard Multiclass SVM

Feature Set	Fold	Kernel	C	γ
Top-125	1	RBF	0.1697	8.90×10^{-4}
Top-125	2	Linear	0.0010	–
Top-125	3	RBF	0.1174	1.11×10^{-3}

Table A.4. Best hyperparameter configurations identified by Optuna for the standard multiclass SVM experiments.

A.4 Hierarchical Classification Strategies

The same Optuna optimization procedure described above was also adopted for the hierarchical and three-class classification strategies. The hyperparameter search was performed using the same search space and optimization protocol adopted for the binary and standard multiclass classifiers. Since these experiments followed an identical optimization procedure, only the resulting classification performances are discussed in Chapter 4.

Appendix B

Additional Segment-Level Results

The main body of this thesis focuses on the subject-level evaluation, which better reflects the intended clinical application of the proposed EEG-based classification framework. Nevertheless, for completeness and reproducibility, this appendix reports the corresponding segment-level performance obtained for all the investigated classification strategies. Although the absolute performance values differ from those observed at the subject level, the same overall trends are consistently observed across all experiments.

B.1 Standard Multiclass Classification

Method	**Accuracy**	**Balanced Accuracy**	**F1-Macro**
Baseline SVM	**89.96 ± 0.73**	84.44	0.708
Top-125 Features	88.87 ± 1.15	91.08 ± 3.37	**0.713**
Optuna All Features (375)	89.36 ± 0.20	86.97	0.702
Optuna + Top-125 Features	88.38 ± 1.68	**92.44 ± 4.24**	0.711

Table B.1. Segment-level performance obtained for the standard multiclass classification task.

B.2 Healthy vs Disease Classification

Method	Accuracy	Balanced Accuracy	F1-Macro
Baseline SVM	90.23 ± 1.22	90.59 ± 3.57	0.831 ± 0.022
Top-100 Features	90.99 ± 0.80	92.80 ± 2.09	0.846 ± 0.013
Optuna All Features (375)	–	91.91 ± 2.61	–
Optuna + Top-100 Features	–	**92.84 ± 2.11**	–

Table B.2. Segment-level performance obtained for the Healthy vs Disease classification task.

B.3 Disease-Specific Classification

Method	Accuracy	Balanced Accuracy	F1-Macro
Baseline SVM	96.53 ± 1.90	97.61 ± 1.65	0.963 ± 0.021
Top-50 Features	98.85 ± 0.94	99.36 ± 0.52	0.988 ± 0.010
Optuna All Features (375)	–	98.26 ± 0.79	–
Optuna + Top-50 Features	–	**99.37 ± 0.51**	–

Table B.3. Segment-level performance obtained for the disease-specific classification task (Alzheimer's disease, Epilepsy and ALS).

B.4 Leakage-Free Hierarchical Cascade

Method	Accuracy	Balanced Accuracy	F1-Macro
Baseline Cascade	90.39 ± 0.61	87.77 ± 5.70	0.730 ± 0.038
Top-150 Features	**90.66 ± 0.47**	**91.46 ± 6.26**	**0.743 ± 0.034**
Optuna All Features (375)	90.39 ± 0.48	87.91	0.730
Optuna + Top-150 Features	89.47 ± 0.89	91.10 ± 4.98	0.720 ± 0.026

Table B.4. Segment-level performance obtained for the leakage-free hierarchical cascade.

B.5 Data Augmentation using SMOTE and CT-GAN

Method	Accuracy	Balanced Accuracy	Macro F1
SMOTE	93.38 ± 2.97	78.82 ± 14.27	0.722 ± 0.130
CTGAN	90.44 ± 1.47	88.48 ± 10.38	0.715 ± 0.091

Table B.5. Segment-level performance obtained using SMOTE and CT-GAN data augmentation.

B.6 Clinically-Guided Three-Class Classification

B.6.1 Stage 1: Healthy vs Epilepsy vs Neurodegenerative Diseases

Method	Segment Balanced Accuracy
Baseline	92.05 ± 5.46
Top-125 Features	**94.73**
Optuna All Features (375)	93.29
Optuna + Top-125 Features	94.70

Table B.6. Segment-level performance obtained for the first stage of the clinically-guided hierarchical framework.

B.6.2 Stage 2: Alzheimer's Disease vs ALS

Method	Segment Balanced Accuracy
Baseline	95.30
Top-40 Features	**99.68**
Optuna All Features (375)	93.51
Optuna + Top-40 Features	97.72

Table B.7. Segment-level performance obtained for the second stage of the clinically-guided hierarchical framework.