Advanced computational approaches for EEG-Based decoding of neurodegenerative diseases

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING
Ph.D. Program

SSD: IBIO-01/A - BIOENGINEERING

Final Dissertation

Advanced Computational Approaches for EEG-Based Decoding of Neurodegenerative Diseases

---

by

Elena Sibilano

Supervisor:
Prof. Vitoantonio Bevilacqua, Ph.D.
Co-supervisors:
Prof. Antonio Brunetti, Ph.D.
Prof. Alberto Mazzoni, Ph.D.

Coordinator of Ph.D. Program:
Prof. Mario Carpentieri, Ph.D.

Course n°37, 01/11/2021 - 31/10/2024

Politecnico di Bari

LIBERATORIA PER L'ARCHIVIAZIONE DELLA TESI DI DOTTORATO

Al Magnifico Rettore
del Politecnico di Bari

La sottoscritta **Elena Sibilano** nata a **Terlizzi** il **04/04/1997**

residente a **Terlizzi (BA)** in **viale Roma 128** e-mail **elena.sibilano@poliba.it**

iscritta al 3° anno di Corso di Dottorato di Ricerca in **Ingegneria Elettrica e dell'Informazione** ciclo **XXXVII**

ed essendo stata ammessa a sostenere l'esame finale con la prevista discussione della tesi dal titolo:

Advanced Computational Approaches for EEG-Based Decoding of Neurodegenerative Diseases

DICHIARA

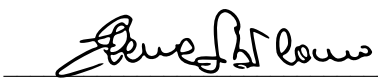
1) di essere consapevole che, ai sensi del D.P.R. n. 445 del 28.12.2000, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono puniti ai sensi del codice penale e delle Leggi speciali in materia, e che nel caso ricorressero dette ipotesi, decade fin dall'inizio e senza necessità di nessuna formalità dai benefici conseguenti al provvedimento emanato sulla base di tali dichiarazioni;
2) di essere iscritta al Corso di Dottorato di ricerca in Ingegneria Elettrica e dell'Informazione ciclo XXXVII, corso attivato ai sensi del "*Regolamento dei Corsi di Dottorato di ricerca del Politecnico di Bari"*, emanato con D.R. n.286 del 01.07.2013;
3) di essere pienamente a conoscenza delle disposizioni contenute nel predetto Regolamento in merito alla procedura di deposito, pubblicazione e autoarchiviazione della tesi di dottorato nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica;
4) di essere consapevole che attraverso l'autoarchiviazione delle tesi nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica del Politecnico di Bari (IRIS-POLIBA), l'Ateneo archivierà e renderà consultabile in rete (nel rispetto della Policy di Ateneo di cui al D.R. 642 del 13.11.2015) il testo completo della tesi di dottorato, fatta salva la possibilità di sottoscrizione di apposite licenze per le relative condizioni di utilizzo (di cui al sito http://www.creativecommons.it/Licenze), e fatte salve, altresì, le eventuali esigenze di "embargo", legate a strette considerazioni sulla tutelabilità e sfruttamento industriale/commerciale dei contenuti della tesi, da rappresentarsi mediante compilazione e sottoscrizione del modulo in calce (Richiesta di embargo);
5) che la tesi da depositare in IRIS-POLIBA, in formato digitale (PDF/A) sarà del tutto identica a quelle **consegnate**/inviate/da inviarsi ai componenti della commissione per l'esame finale e a qualsiasi altra copia depositata presso gli Uffici del Politecnico di Bari in forma cartacea o digitale, ovvero a quella da discutere in sede di esame finale, a quella da depositare, a cura dell'Ateneo, presso le Biblioteche Nazionali Centrali di Roma e Firenze e presso tutti gli Uffici competenti per legge al momento del deposito stesso, e che di conseguenza va esclusa qualsiasi responsabilità del Politecnico di Bari per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi;
6) che il contenuto e l'organizzazione della tesi è opera originale realizzata dalla sottoscritta e non compromette in alcun modo i diritti di terzi, ivi compresi quelli relativi alla sicurezza dei dati personali; che pertanto il Politecnico di Bari ed i suoi funzionari sono in ogni caso esenti da responsabilità di qualsivoglia natura: civile, amministrativa e penale e saranno dalla sottoscritta tenuti indenni da qualsiasi richiesta o rivendicazione da parte di terzi;
7) che il contenuto della tesi non infrange in alcun modo il diritto d'Autore né gli obblighi connessi alla salvaguardia di diritti morali od economici di altri autori o di altri aventi diritto, sia per testi, immagini, foto, tabelle, o altre parti di cui la tesi è composta.

Luogo e data **Bari, 10/12/2024** Firma 

La sottoscritta, con l'autoarchiviazione della propria tesi di dottorato nell'Archivio Istituzionale ad accesso aperto del Politecnico di Bari (POLIBA-IRIS), pur mantenendo su di essa tutti i diritti d'autore, morali ed economici, ai sensi della normativa vigente (Legge 633/1941 e ss.mm.ii.),

CONCEDE

- al Politecnico di Bari il permesso di trasferire l'opera su qualsiasi supporto e di convertirla in qualsiasi formato al fine di una corretta conservazione nel tempo. Il Politecnico di Bari garantisce che non verrà effettuata alcuna modifica al contenuto e alla struttura dell'opera.
- al Politecnico di Bari la possibilità di riprodurre l'opera in più di una copia per fini di sicurezza, back-up e conservazione.

Luogo e data **Bari, 10/12/2024** Firma 

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING
Ph.D. Program

SSD: IBIO-01/A - BIOENGINEERING

Final Dissertation

Advanced Computational Approaches for EEG-Based Decoding of Neurodegenerative Diseases

---

by

Elena Sibilano

Referees:
Prof. Daniele Marinazzo, Ph.D.
Prof. Sara Invitto, Ph.D.

Supervisor:
Prof. Vitoantonio Bevilacqua, Ph.D.

Co-supervisors:
Prof. Antonio Brunetti, Ph.D.
Prof. Alberto Mazzoni, Ph.D.

Coordinator of Ph.D. Program:
Prof. Mario Carpentieri, Ph.D.

Course n°37, 01/11/2021 - 31/10/2024

Abstract

The aim of this Ph.D. thesis is to illustrate the research works conducted to design and develop advanced computational frameworks for analyzing electroencephalographic (EEG) signals to improve the early diagnosis of neurodegenerative diseases (NDs). Dementia is one of the leading causes of disability and death worldwide, and the detection of its initial phases remains a critical challenge both for prognostic and therapeutic purposes. The modern conceptualization of NDs, and particularly of Alzheimer's disease, assumes cognitive decline to develop as a continuum, along which populations with still sufficient functional compensation could be targeted for early clinical trials. In this context, EEG signals can provide non-invasive and cost-effective biomarkers, holding potential for capturing neural dysfunctions associated with neurodegeneration. Nonetheless, the inherent complexity and variability of EEG result in significant challenges for accurate interpretation and analysis.

This thesis addresses how Deep Learning (DL) models, particularly Transformers, and interpretability techniques can be leveraged for robust classification of EEG data, offering insights into subtle cognitive changes in preclinical and prodromal stages and overcoming the need for domain-specific expertise to extract consistent and reliable features. Furthermore, other approaches advancing the integration of computational neuroscience with Machine Learning (ML), including biophysical modeling of neural modulation in response to specific stimuli, are explored.

In particular, the first part of the work presents a novel signal-based Deep Learning framework for distinguishing between subjective cognitive decline (SCD) and mild cognitive impairment (MCI) using resting-state EEG. The methods aim to capture prodromal signs of Alzheimer's disease through a state-of-the-art Transformer model based on the mechanism of self-attention. To enhance clinical trustworthiness and translatability, the previously described method is then integrated with interpretability tools. Specifically, the role of self-attention within Transformer models is systematically explored to explain decision-making processes, providing greater transparency into the models' focus on the input signals for differentiating SCD from MCI and proving that this information could be used to guide the identification of biomarkers of cognitive impairment in resting-state EEG.

The second part of the research work presents computational methods for analyzing evoked responses, namely event-related potentials (ERP) and event-related (de)synchronization (ERD/ERS), in neurodegeneration, exploring motor resonance in early Parkinson's disease, dynamic causal modeling for ERP classification, and the effects of sensory stimuli on electrophysiological responses in a Human-Robot Interaction scenario.

Table of contents

List of figures

List of tables

List of acronyms

Acronym	Definition
Aβ	Amyloid Beta
AD	Alzheimer's Disease
AUC	Area Under Curve
BCI	Brain-Computer Interface
CNN	Convolutional Neural Network
CSF	Cerebrospinal Fluid
CWT	Continuous Wavelet Transform
DCM	Dynamic Causal Modeling
DL	Deep Learning
DNN	Deep Neural Network
EEG	Electroencephalography
ERD	Event-Related Desynchronization
ERP	Event-Related Potential
ERS	Event-Related Synchronization
fNIRS	functional Near-Infrared Spectroscopy
HC	Healthy Control
HRI	Human-Robot Interaction

ICA	Independent Component Analysis
LOSOCV	Leave-One-Subject-Out Cross-Validation
MCI	Mild Cognitive Impairment
MEG	Magnetoencephalography
MHA	Multi-Head Attention
ML	Machine Learning
MMSE	Mini-Mental State Examination
MNS	Mirror Neuron System
MR	Motor Resonance
MRI	Magnetic Resonance Imaging
ND	Neurodegenerative Disease
NLP	Natural Language Processing
PD	Parkinson's Disease
PET	Positron Emission Tomography
RNN	Recurrent Neural Network
ROC	Receiving Operating Curve
rsEEG	resting-state Electroencephalography
SCD	Subjective Cognitive Decline
sEEG	stereotactic Electroencephalography
SNR	Signal-to-Noise Ratio
STFT	Short-Time Fourier Transform
ViT	Vision Transformer
XAI	Explainable Artificial Intelligence

Chapter 1

Introduction

The opening Chapter of this manuscript serves two purposes: to provide an overview of the motivation behind this work and to introduce the scientific challenges associated with developing intelligent systems for detecting and characterizing neurodegenerative disorders. Following this, a thorough description of the objectives and contributions to the field will be presented. Finally, a structured outline of the manuscript will guide readers through the rest of the thesis.

1.1 Motivation and Scientific Challenges

1.1.1 On the importance of non-invasive biomarkers for early diagnosis of neurodegenerative diseases

Neurodegenerative diseases (NDs) are a composite group of central nervous system disorders characterized by a chronic and selective process of loss of function and structure affecting neurons. These pathologies exhibit extreme intra- and inter-subject variability, with clinical manifestations depending on the type of neuronal systems involved during the course of the disease. NDs can lead to movement issues, known as ataxias, or impairments in mental functioning, referred to as dementias [1]. Although certain physical or cognitive symptoms linked to these disorders can be alleviated through therapy, there is currently no definitive cure for prevalent neurodegenerative conditions like Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD) and Amyotrophic Lateral Sclerosis (ALS) [2].

Ageing is one of the main risk factors for most neurodegenerative diseases [3]. Given that about 22% of the world's population is estimated to be over 60 years old by 2050 [4], early diagnosis of NDs, preventive treatment to delay their onset or improve their prognosis

are open research objectives, both from a purely clinical perspective and in fields related to or supporting clinical practice [5]. The main obstacle to reaching these objectives is the fact that in many neurodegenerative diseases, symptoms only manifest when significant neuronal loss has already occurred, but the course of the disease is now known to begin several years in advance [6, 7]. Although this knowledge has deeply shifted the modern conceptualization of these pathologies, their diagnosis is still subject to expert interpretation and requires too much time and incurs into high costs.

A traditional approach to assess the progression of neurodegenerative diseases is through neuropsychological evaluations [8]. These assessments consist of a series of tests and questionnaires designed to evaluate cognitive functions such as memory, language, attention, and executive function. While providing valuable insights into a patient's cognitive abilities, they present several limitations, including inconsistencies introduced by different examiners' expertise and lack of sensitivity for differential diagnosis [9]. As a consequence, misdiagnosis can lead to suboptimal treatment and costly investigations.

In the last century, intensive research on chemical and neuroimaging biomarkers has led to the review of diagnostic criteria for NDs, introducing objective and quantifiable indicators on the pathological changes occurring in the brain [10, 11]. These biomarkers encompass a range of physiological indicators, which can provide a reliable reflection of neurodegeneration [12].

Among the available brain imaging techniques, Electroencephalography (EEG) provides a reliable, noninvasive and cost-effective tool to investigate altered patterns in brain activity in pathological conditions [13, 14]. Recent literature in this field has expanded exponentially. However, the intrinsic complexity of the EEG signal, along with the high dimensionality, low signal-to-noise ratio and large signal variability strongly impacts the time and effort needed for its interpretation. These challenges may be better addressed by feature-based approaches that exploit more specific assumptions about the signal.

1.1.2 Unveiling the impact of Artificial Intelligence on neurodegenerative disease management

The development of intelligent systems based on Machine Learning (ML) approaches can support data management and analysis as tools for understanding the pathological mechanisms underlying neurodegenerative disorders, stratifying subgroups of patients within a single pathology, and developing new optimal and specific therapies for each patient, geared towards the frontiers of precision medicine [15].

These algorithms also forecast disease progression and prognosis, offering valuable insights into disease trajectories and supporting clinical decision-making. ML-driven image and signal processing techniques detect subtle structural and functional changes in the brain, aiding in the identification of disease-specific biomarkers [16].

Among modern ML models, Deep Learning (DL) networks are particularly suited for EEG data processing, since they can encompass large amounts of multidimensional data, and different architectures can decode directly from the time, space and frequency domains of the EEG signals. Furthermore, given the sequential nature of EEG data, DL models like recurrent neural networks (RNNs) and Transformers can capture temporal dependencies and patterns within the signal. This ability to model complex, time-dependent relationships makes DL models highly effective for tasks such as classification, feature extraction, and anomaly detection in EEG data, ultimately supporting improved accuracy in detecting cognitive and neurological states.

However, it is now known that the unexpected advancement and increasingly frequent use of DL techniques in managing big multivariate data have exposed an intrinsic problem in the decision-making process of neural network models. Although the implicit feature extraction capabilities provide solutions in contexts where traditional methods have shown limitations, the problem of deep models' *explainability* remains open. The black box nature of Deep Learning models, and thus the lack of interpretability and transparency of the implemented logic, do not allow for their complete permeability, especially in clinical settings [17]. In this light, the development of methods for visualizing, explaining, and interpreting DL models has recently attracted much attention. The emerging field of Explainable Artificial Intelligence (XAI) holds great promise for unlocking new insights into the the intricate biological processes underlying NDs and potentially bridging the gap between advanced DL models and their real-world clinical applications. In particular, when dealing with the EEG signal, there is a great need for interpretable models that can elucidate which temporal and frequency-based features contribute most to the model's predictions [18]. This interpretability is especially crucial for understanding the rationale behind a model's outputs and equip clinicians with transparent tools to both identify key biomarkers and offer insights into their relevance, ultimately advancing the personalized diagnostics and treatment for NDs.

Another important consideration in the field of neurodegenerative disease diagnosis is the limited knowledge about the generalization of existing methods. Variations in EEG signal quality, patient demographics, disease heterogeneity, and data collection protocols introduce

inconsistencies that limit the standardization diagnostic tools. Thus, developing more robust models that can adapt to a range of conditions and patient profiles is an open challenge [19].

1.2 Contribution

Given the limitations and research needs stated above, the main contributions of this Ph.D. thesis are the design, development and evaluation of the first end-to-end Deep Learning framework tailored for resting-state EEG data analysis, for discriminating between early phases of cognitive decay in the Alzheimer's spectrum, and specifically Subjective Cognitive Decline (SCD) from Mild Cognitive Impairment (MCI). The same framework is then employed to perform a multiclass classification among healthy controls (HC), SCD and MCI. The method employs a multi-head attention-based Transformer model to address the inherent challenges of EEG's sequential nature and identify the most discriminative frequency bands in the signal. It achieves state-of-the-art results, confirming that changes in relative power in the lower frequencies are indicative of diffuse slowing of brain oscillations, which is a hallmark feature in the progression of Alzheimer's.

Furthermore, a systematic interpretability workflow is integrated to enhance model's transparency, while also guiding the identification of EEG biomarkers for neurodegeneration. By leveraging the attention mechanism of the Transformer, a novel perspective of XAI techniques applied to EEG signals is proposed, which aims to find physiological correlations of the model's outcomes with pathological EEG signatures. In addition, the different role of multi-head attention and self-attention in the explainability process is investigated.

Moreover, this thesis work presents computational approaches based on statistical methodologies for the analysis of evoked potentials both in the field of neurodegenerative diseases, targeting early Parkinson's disease, and other applications in the domain of cognitive neuroscience.

1.3 Thesis Outline

After providing an introduction reporting the reference scientific context in this Chapter, the thesis is structured as follows:

- **Chapter 2** provides a theoretical background on the main concepts of this thesis, elaborating on different types of NDs and highlighting recent advances in the use of DL models in EEG signal analysis and their application in the context of neurodegeneration.

Furthermore, it provides some background on attention-based DL architectures, as well as on the concept of interpretability and XAI techniques.

- **Chapter 3** reports the first original work addressing the task of classifying subjects in the early phases of cognitive impairment, namely SCD and MCI, based on resting-state EEGs, with the aim of providing an end-to-end framework that processes signals in the time domain to characterize the stages along Alzheimer's disease continuum.

- Continuing along these lines, **Chapter 4** expands the previous research work by describing a novel framework for improving interpretability in EEG-Transformers. It describes how a systematic analysis of the model's focus on input signals and the corresponding classification outcomes can produce physiologically significant explanations.

- **Chapter 5** shifts the focus from the analysis of spontaneous EEG to the implementation of statistical methods for interpreting neural responses to specific stimuli. A range of applications of these techniques, from the characterization of early Parkinson's disease to the field of Human-Robot Interaction, is explored. Moreover, an approach based on Dynamic Causal Modeling for simulating and classifying event-related potentials is proposed.

- Finally, the conclusions of the research work described in this thesis are reported in **Chapter 6**.

Chapter 2

Background

This Chapter offers a thorough background on the key clinical and technical concepts of this thesis. Firstly, an overview of neurodegenerative diseases, mainly focusing on depicting the continuum of cognitive decline leading to dementia, is provided. The importance and the latest advances in the field of biomarkers for neurodegenerative diseases are also detailed. Then, a description of the electroencephalographic signal both from a clinical and analytical perspective is presented. The Chapter concludes with a synopsis of modern Deep Learning models and their employment in designing EEG-based computer-aided diagnostic systems in the scenario of neurodegeneration, along with a definition of explainability methods for enhancing trustworthiness of these systems.

2.1 Neurodegenerative Diseases

As evidenced in the previous Chapter, neurodegenerative diseases typically display some degree of heterogeneity, such as differences in the location of disease pathology, the extent and type of neuroinflammation, or the severity of neurodegeneration [2].

Alzheimer's disease (AD) and Parkinson's disease (PD) are the two most common neurodegenerative disorders worldwide, with AD being the most prevalent and PD the second. Both involve protein misfolding and aggregates, leading to neurotoxicity and cell death. While the exact causes of these diseases remain unclear, research suggesting they are likely the result of a combination of genetic, environmental, and lifestyle factors [20], their accurate and timely diagnosis is essential for enabling the prospective screening of ageing populations, mainly because certain subgroups, potentially identified by biomarkers, may respond more favorably to specific therapies. It's even likely that different disease-modifying treatments have optimal time windows in which they are most effective during the

disease's progression [21], i.e. during the early pre-symptomatic or prodromal stages, before significant and irreversible neurodegeneration has occurred [12].

In this Ph.D. thesis, the automatic discrimination of EEG signals of individuals with Subjective Cognitive Decline (SCD) and Mild Cognitive Impairment (MCI), which represent the preclinical and prodromal stages of AD, has been explored.

2.1.1 Alzheimer's Disease

Alzheimer's disease is a neurodegenerative disorder that gradually leads to cognitive decline, behavioral disturbances, and loss of functional independence. As the most prevalent cause of dementia worldwide and one of the leading causes of death, AD's rising prevalence is closely tied to global aging populations, making it a significant public health challenge with substantial socio-economic implications [22]. Consequently, AD has become a healthcare challenge of epidemic proportions, with no effective treatment to modify the disease's progression [23]. The irreversible nature of cognitive and behavioral decline in AD results in a devastating impact. A significant portion of these costs is associated with the institutionalization of dementia patients, a step that becomes necessary in about 50% of cases after five years and up to 90% of cases after eight years [24].

Typically, individuals affected by AD first experience episodic memory impairment, followed by additional cognitive symptoms such as language difficulties, challenges with executive and visuospatial functions, and eventually, the onset of dementia [25].

The brain changes associated with these symptoms involve plaques of a toxic protein called Amyloid Beta (Aβ) and tangles of tau proteins inside neurons. As plaques and neurofibrillary tangles accumulate, they interfere with synaptic function and neuronal health. Neuronal death results in a progressive loss of brain tissue, particularly in areas associated with memory and learning, such as the hippocampus, which manifests as brain atrophy, particularly in later stages of the disease.

2.1.1.1 Diagnosis

The conceptualization of AD as a clinical-biological entity was elaborated in the 1980s and was widely accepted and applied to clinical activity for over 30 years [26]. The National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) workgroup proposed, in 1984, the first set of structured diagnostic criteria for AD [27]. The term *probable AD* was first introduced to describe an acquired, progressive amnestic dementia for which no other cause could be

identified. This clinical diagnosis was linked to the presence of β-amyloid-containing neuritic plaques and tau-containing neurofibrillary tangles, forming the basis of a clinicopathologic model. This model was subsequently adapted in a simplified form for use in population-based studies and general clinical practice, though the uncertainty implied by the term was often overlooked. Nonetheless, the definite diagnosis of AD dementia was confirmed only by post-mortem evidence of neuropathological alterations characteristic of AD.

As public awareness of cognitive decline in later life increased, a vernacular understanding of AD emerged. In this context, Alzheimer's disease came to represent all forms of dementia not attributable to another clearly identifiable cause. However, the growing availability of biomarkers for β-amyloid and tau has highlighted the gap between the clinicopathologic, vernacular, and pathobiological models of AD.

In 2007, the NINCDS-ADRDA criteria were firstly reviewed by the International Working Group (IWG) [28]. For the first time, in vivo biomarkers were proposed to support diagnosis and characterize the clinical-biological entity of AD in the predementia stage, when symptoms do not configure the clear picture of overt dementia. The *prodromal AD* stage included a wide range of described entities, such as age-related memory impairment and cognitive decline, mild neurocognitive disorder, and Mild Cognitive Impairment (MCI). In 2010, and later with the revision of 2014, the IWG proposed a new classification, defining AD as a spectrum, including patients with overt dementia and individuals with mild or no symptoms. In this revision, a clear distinction was proposed between the clinical diagnosis and disease pathology, considering that the evidence of AD pathological changes not necessarily coincides with the stage of AD dementia [29]. With the development and increased availability of biomarkers for AD, the National Institute on Aging and Alzheimer's Association (NIA-AA) proposed a theoretical model in which the cognitive decline appears later over the disease stage, preceded by pathological and molecular alterations, and manifests initially with MCI and then with AD dementia [30].

2.1.1.2 The preclinical and prodromal phases of AD

The biological definition of AD motivated the recognition of a long pre-dementia stage, preceding the clinical appearance of symptoms, representing a potential fruitful therapeutic target [31]. In this pre-dementia stage, while pathological alteration can be already detectable, irreversible neurodegenerative processes may not yet occur, offering the possibility to change the disease course effectively. This evidence led to a new biological definition of the disease, which assumes that the cognitive decline in AD occurs over a long period and develops as a

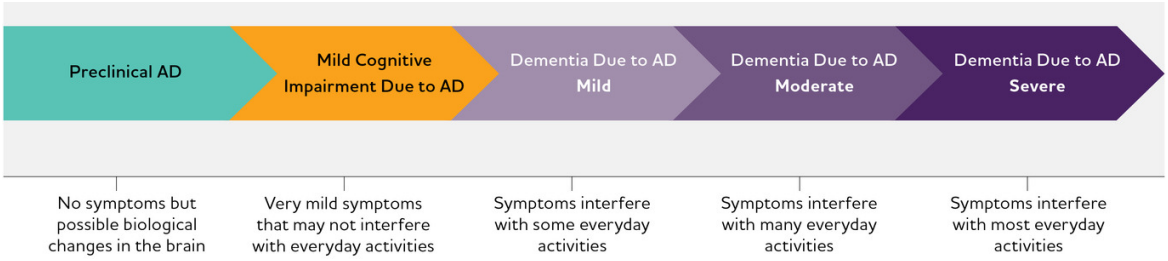


Fig. 2.1 Alzheimer's disease continuum [32]

continuum rather than as distinct, clinically-defined entities (Figure 2.1). On this continuum, three broad phases can be distinguished:

- **Preclinical AD**, which can include Subjective Cognitive Decline (SCD), where clinically unimpaired individuals with no symptoms show brain changes linked to neurodegeneration, including amyloid deposition and neurodegeneration;
- **Prodromal AD**, including MCI individuals, with a subtle cognitive decline but without impact on daily activity functioning;
- **Dementia due to AD.**

Subjective Cognitive Decline The concept of Subjective Cognitive Decline has been introduced to describe individuals who perceive a cognitive decline despite normal performance in standardized assessments [33, 34]. SCD covers all cognitive domains and is frequently self-reported as memory decline, commonly found in 25-50% of adults over 65 years old [35, 36]. Emotional factors such as anxiety and depression often accompany SCD, complicating its differential diagnosis [37, 38].

Moreover, the progression risk from SCD to dementia is heightened when subjective memory decline onset is recent (within five years) and accompanied by concern, prompting individuals to seek medical help [34]. However, only about 14% of individuals with SCD progress to dementia over long-term follow-up, with fewer than one-third developing MCI [39].

Biomarker analysis enhances etiological understanding and prognosis in SCD, particularly in detecting amyloid-β, tau, and brain atrophy [40, 41]. Abnormal biomarkers increase the likelihood of progression to AD dementia [33]. In longitudinal studies, SCD patients with positive amyloid-β markers show a higher risk of progressing to dementia, particularly when multiple biomarkers are abnormal [42].

Mild Cognitive Impairment Mild Cognitive Impairment refers to a condition characterized by mild but noticeable cognitive difficulties, which, however, do not affect the individual's ability to carry out daily life activities independently. Specifically, MCI is defined as an intermediate stage between natural cognitive decline due to aging and dementia, with patients experiencing a greater degree of cognitive impairment than expected [30]. Currently, the underlying causes of mild cognitive decline are unknown.

Historically, the amnestic MCI, which is the most frequent type, traditionally indicated the prodromal stage of typical AD dementia, while the non-amnestic MCI possibly indicated other etiologies such as frontotemporal dementia (FTD) and dementia with Lewy bodies (DLB). However, this classification is not reliable [43]. Given the heterogeneity of MCI causes, research criteria and biomarkers have been defined to support the diagnosis of MCI as a prodromal phase of Alzheimer's. According to the NIA-AA, cerebral amyloidosis and neurodegeneration are necessary to determine this type of cognitive decline [30].

2.1.1.3 Biomarkers for AD

Besides biomarkers of neuropathology, markers of neurodegenerative changes have been proposed to support AD diagnosis. Common biomarkers include amyloid-β and tau, detectable through imaging techniques like amyloid-PET and tau-PET or through cerebrospinal fluid (CSF) measurements. Amyloid-PET allows visualization of amyloid-β plaques, which typically accumulate in the neocortical association areas, particularly the medial parietal and frontal regions [44]. However, amyloid-PET positivity is also found in cognitively normal elderly individuals, complicating its use as a standalone diagnostic tool. CSF biomarkers, such as amyloid-β42 and the amyloid-β42/amyloid-β40 ratio, correlate well with amyloid-PET findings and are predictive of AD progression, particularly in patients with MCI [45]. The main limitation in using CSF measurements is the collection of samples by lumbar puncture, which is an invasive procedure not widely available.

Tau biomarkers are equally significant. Tau-PET imaging shows high accuracy in distinguishing AD patients from controls, often correlating with cognitive impairment and neurodegeneration. In particular, tau deposition varies across AD phenotypes, providing valuable insights into the clinical diversity of the disease. For example, temporoparietal tau pathology is common in typical amnestic AD, while posterior parietal and occipital tau accumulation is associated with posterior cortical atrophy [46].

Neurodegenerative biomarkers, including MRI, [18F]FDG-PET, and fluid biomarkers like neurofilaments, complement amyloid-β and tau measurements by highlighting brain atrophy and hypometabolism. Hippocampal atrophy, detectable via structural MRI, is one

of the earliest signs of AD-related neurodegeneration [47]. However, certain types of AD exhibit irregular patterns of atrophy, which may overlap with other pathologies and make diagnosis uncertain. [18F]FDG-PET is another key tool in clinical practice, revealing brain regions with hypometabolism, often correlating with tau pathology and neuronal loss.

Fluid biomarkers, including t-tau and neurofilaments, are elevated in both CSF and plasma in cases of axonal degeneration. In addition to AD, these markers are useful in other neurodegenerative diseases, such as Parkinson's Disease and ALS. Despite their promise, plasma biomarkers have yet to achieve the sensitivity and specificity needed for routine clinical use.

More recently, scientific literature has expanded with studies employing EEG as a supporting technique for a faster and simpler diagnosis of AD. Several studies proposed resting state electroencephalographic (rsEEG) rhythms as candidate biomarkers of AD [48–51]. A more comprehensive review of research in this field can be found in the work by Babiloni *et al.* [52]. Cassani *et al.* summarized EEG changes related to AD progression into four main categories: slowing, complexity reduction, synchronization decrement and neuromodulatory deficit [13]. At the MCI stage, such EEG abnormalities were found to be intermediate between healthy controls and dementia patients, and more severe compared to subjects with SCD [53]. Changes in relative and absolute power of Theta (θ) frequency band appear to be significant among AD, MCI and healthy controls at individual level [54]. Significantly higher global Delta (δ) and Theta power, lower global Alpha (α) power and a higher global peak frequency have also been found in patients with SCD that have progressed to MCI and dementia [53]. Hence, measures of EEG-recorded brain activity can represent sensitive, non-invasive markers in the prediction of clinical development of AD. This assumption holds true also when comparing EEG to other neuroimaging methods, both structural and functional [55].

2.1.2 Parkinson's disease

Parkinson's disease is the second most common neurodegenerative disorder after Alzheimer's disease, affecting an estimated 0.5-1% of individuals aged 65-69 and 1-3% of those aged 80 and older [56]. The primary pathological feature of Parkinson's disease is the degeneration of dopaminergic neurons in the substantia nigra, a region of the brain's basal ganglia. This phenomenon starts with the aggregation of α-synuclein protein into structures known as Lewy bodies. These Lewy body aggregations typically begin in the lower, or caudal, regions of the brain and progressively move toward more anterior areas, finally reaching the substantia

nigra in the midbrain, where dopaminergic neurons either die or become dysfunctional, leading to a reduction or depletion of dopamine. This decline in dopamine, a critical neurotransmitter involved in the regulation of voluntary movement, is responsible for the typical motor symptoms of PD. As the disease advances and dopamine levels continue to drop, patients progressively lose the ability to control their movements normally. While the cause and pathogenesis of selective dopamine neuron loss and α-synuclein accumulation remain unknown, increasing evidence from environmental risk factors and early-onset genetics suggests a convergence between energy metabolism and protein disposal in the development of the pathology [57]. These findings indicate that mitochondrial dysfunction and ubiquitin-proteasome system impairment may play a crucial role in the etiology of PD.

2.1.2.1 Diagnosis

PD is clinically assessed primarily through the identification of characteristic motor symptoms. A diagnosis is typically made when at least two of four motor symptoms, which include resting tremor, bradykinesia, rigidity, and postural instability, are present. However, PD is also associated with a range of non-motor comorbidities, including mental health disorders, autonomic and gastrointestinal dysfunction, and significant sleep disturbances, all of which can severely impact patients' quality of life and that of their families.

Non-dopaminergic and non-motor symptoms of PD often manifest years before motor symptoms appear, and they can dominate the clinical presentation in advanced stages, proving difficult to manage effectively [58]. This aligns with evidence suggesting that PD pathology may be at an advanced stage well before motor symptoms become clinically apparent [59]. Deficits in executive function, along with impairments in working memory and attention, are often regarded as the earliest and most prominent neuropsychological indicators of functional alterations in fronto-striatal circuits [60]. Braak *et al.* [61] proposed a staging model for PD pathology based on Lewy body distribution. In the earliest stages (stage 1), neuronal damage begins in the dorsal motor nucleus of the vagus nerve in the medulla and the olfactory bulb, representing a pre-Parkinsonian state along with stage 2, where pathological inclusions spread to the subcoeruleus-coeruleus complex and the magnocellular nucleus of the reticular formation. The condition is not classified as Parkinsonian until stage 3, with involvement of the substantia nigra, pedunculopontine nucleus, and amygdala, and then stage 4, which affects the temporal mesocortex. Late-Parkinsonian stages include stage 5, with initial neocortical involvement, and stage 6, where nearly the entire neocortex is affected.

2.1.2.2 Biomarkers for PD

Along with genetic, metabolic, and fluid biomarkers, electroencephalographic (EEG) studies have highlighted alterations in cognitive processing through both event-related potentials (ERP) and event-related desynchronization/synchronization (ERD/ERS) methods. Changes in the P3 component, such as delayed latencies or reduced amplitudes, have been observed in Parkinsonian patients [62, 63]. These changes are interpreted as indicators of cognitive slowing, particularly in stimulus classification and attention processing. ERD/ERS analysis provides additional insight by examining frequency-specific changes that reflect dynamic, transient connections between different brain regions. ERD in the Alpha band and reduced Theta-ERS at frontal regions have been linked to disruptions in basal ganglia activity and associated thalamo-cortical networks in PD, reflecting impaired auditory and visual working memory encoding and categorization processes [64, 65]. This frequency-specific approach has proven valuable in detecting subtle neurophysiological changes in PD that are not captured by traditional ERP analysis alone.

2.2 Electroencephalography

Exactly a century after its discovery by German psychiatrist Hans Berger, electroencephalography (EEG) remains the most common non-invasive technique for monitoring brain electrical activity in clinical and research settings.

The electroencephalographic signal is a measure of extracellular current flow generated by the synchronous activity of a large number of pyramidal neurons with similar spatial orientation. Cortical electrical activity exhibits oscillations characterized by different amplitudes and frequencies, referred to as rhythms. In the context of EEG rhythms, synchronization refers to the temporal dynamics of electrical activity in local cortical neuronal populations, showing collective oscillatory behavior on a macroscopic spatial scale of a few centimeters.

The amplitude of the EEG signal mainly depends on the degree of synchronization with which cortical neurons interact. Asynchronous excitation of a group of neurons generates an irregular EEG signal with low-amplitude oscillations. In contrast, synchronous excitation produces a higher-amplitude signal due to the temporal summation of individual electrical contributions. The frequency of oscillations in the EEG is linked to the pacemaker properties of thalamic neurons and feedback mechanisms occurring within the neural circuit [66].

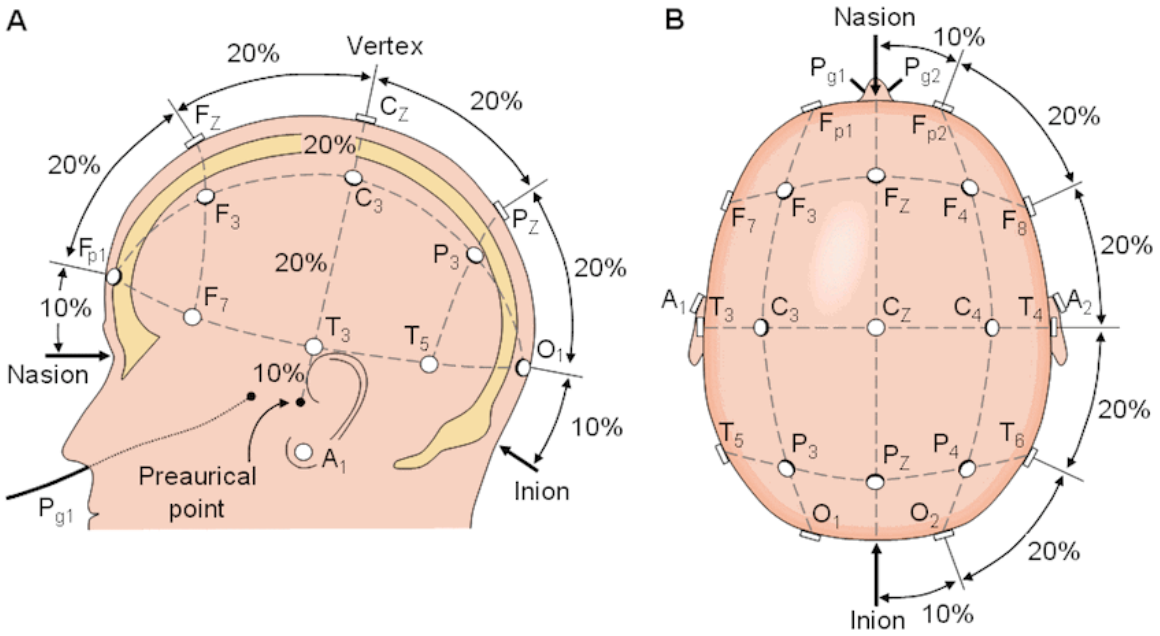


Fig. 2.2 Diagram of the international 10-20 system seen from the (A) left and (B) above the head. Each electrode is assigned a nomenclature with a letter and a number. The letters indicate the areas of the scalp: F (Frontal), C (Central), T (Temporal), P (Parietal) and O (Occipital); numbers are odd for the left side and even for the right side.

Modern EEG systems consist of scalp electrodes connected to high-impedance amplifiers and a digital data acquisition unit. Traditionally, the contact between the electrode and the skin is enhanced using electrolytic gels or abrasive pastes. Recently, dry electrodes have been used, leveraging advances in materials science and electronics to reduce preparation time. The standard 10–20 system has long been used to define electrode placement, with 21 electrodes. However, its spatial resolution is insufficient for modern brain research. To address this, high-density systems with up to 256 electrodes are now commercially available, while some specialized ultra high-density EEG systems can support over 300 electrodes.

Figure 2.2, which represents the standard 10-20 system, is provided as a reference for the next Chapters for identifying electrodes' positions and labels.

EEG traces are characterized by spontaneous voltage fluctuations associated with various mental states, levels of consciousness, or pathological disturbances. The range of the clinically relevant EEG frequency components lies between 0.1 and 100 Hz and commonly in routine clinical settings it may be more restricted (i.e., between 0.1 and 70 Hz) [67]. Oscillations have characteristic frequency bands that are clinically defined and associated with different cerebro-functional states. EEG rhythms are classified into Delta (δ), Theta (θ), Alpha (α), Beta (β), and Gamma (γ) rhythms:

- δ rhythm has oscillations at frequencies below 4 Hz. It represents the physiological rhythm observed during the third and fourth stages of human sleep and anesthesia. This rhythm can also manifest in the presence of subcortical lesions. In adults, it is typically most pronounced in the frontal regions (referred to as FIRDA - Frontal Intermittent Rhythmic Delta), while in children, it tends to be more prominent in the posterior regions (known as OIRDA - Occipital Intermittent Rhythmic Delta). The delta band is characterized by high amplitude and slow wave activity.

- θ rhythm presents oscillations in the 4–7 Hz band and is present during deep sleep states. Theta activity is typically observed in young children and may appear during drowsiness or arousal in older children and adults; however, an excess of theta activity for a given age is indicative of abnormal brain function. This rhythm can present as a focal disturbance in cases of focal subcortical lesions. Theta activity is thought to originate from the limbic system and hippocampal regions. It has been associated with anxiety, behavioral activation, and behavioral inhibition. When functioning appropriately, the theta rhythm mediates and promotes adaptive, complex behaviors such as learning and memory.

- α rhythm is characterized by oscillations in the 8–13 Hz band, with an average amplitude of 30 μV, and is recorded with closed eyes in an awake subject. Alpha is a common state for the brain and occurs whenever a person is alert (it is a marker for alertness and sleep), but not actively processing information. It emerges with closing of the eyes and with relaxation, and attenuates with eye opening or mental exertion. Alpha rhythm is more evident in the occipital cortex. The posterior basic rhythm is actually slower than 8 Hz in young children (therefore technically in the theta range). Besides the classic alpha rhythm of the visual cortex, there are rhythmic activities in the same frequency range that can be recorded from the somatosensory cortex (called the mu rhythm) and the temporal cortex (called the tau rhythm).

- β rhythm is a very fast rhythm, with oscillations between 14 and 30 Hz, a small amplitude (1–20 μV), and is associated with active cortical areas and levels of consciousness such as attention and concentration. Typically observed bilaterally in a symmetrical distribution, beta waves are most pronounced in the frontal regions. This activity is closely tied to motor behavior, often diminishing during active movements. Low-amplitude beta with a variety of frequencies is frequently linked to active, busy, or anxious thinking, as well as sustained concentration. Conversely, rhythmic beta waves exhibiting a dominant frequency pattern can be associated with various pathologies and the effects of certain medications, particularly benzodiazepines. In cases of cortical damage, beta activity may be diminished or absent. Overall, beta rhythm predominates in individuals who are alert, anxious, or have their eyes open.

- γ rhythm has oscillations with frequencies above 30 Hz and low amplitude. It is linked to active information processing in the cortex and is thought to represent binding of different populations of neurons together into a network to carry out certain cognitive or motor functions.

However, it should be noted that the EEG spectrum is not configured into discrete bands, but it is instead a continuum of overlapping frequencies. The division into bands is an arbitrary practical construct developed to simplify the interpretation of EEG data in clinical and research contexts. These bands are not entirely independent of one another; instead, they often interact and influence each other dynamically, reflecting the complex and integrative nature of neural activity. Such interactions underline the importance of interpreting EEG rhythms not in isolation but as components of a broader, interconnected spectral landscape.

2.2.1 EEG Recording

Studies have made use of EEG signals under diverse recording conditions, which can be divided into two major groups:

- **Resting-State EEG**. Spontaneous EEG activity refers to brain signals recorded in the absence of any external stimuli, capturing the brain's intrinsic, background activity. This type of recording is advantageous as does not require participants to perform tasks, making the EEG acquisition process simpler, more comfortable, and less stressful, particularly beneficial for elderly participants. Resting-state EEG includes recordings in an awake resting state, typically with eyes open or closed, as well as during sleep, allowing researchers to observe baseline neural dynamics under minimal external influence [68].

- **Event-Related EEG**. EEG signals recorded in response to specific events are known as time-locked signals. These responses are also phase-locked, resulting in event-related potentials (ERPs). When the EEG response is not phase-locked, it is referred to as induced activity, which can be assessed through event-related (de)synchronization (ERD/ERS) [69, 70] or event-related oscillations (ERO). Such event-based EEG activities are linked to sensory, perceptual, motor, and cognitive processes, providing insights into various brain functions [71]. ERD and ERS are produced by a change in the synchronization of neurons that causes a decrease (or increase) in the signal amplitude of a specific frequency band during a motor task (either executed or imagined). Indeed, ERD/ERS modulations are most prominent in the EEG measured in correspondence with the sensorimotor cortex. ERD is usually observed in low-frequency bands such as mu or beta bands, whereas ERS is the result of relaxation, which is mainly observed in the beta frequency.

In this Ph.D. thesis, different methods for modeling and analyzing EEG activity both at rest and related to specific stimuli in neuropathological and physiological conditions have been addressed.

2.2.2 EEG Analysis

The analysis of EEG signals is closely tied to the extraction of quantitative parameters, such as the dominant frequency value or the similarity between two signals recorded from symmetrical derivations, either simultaneously or at different times. Without these measurements, the evaluation of the EEG signal remains subjective and is unlikely to lead to a

logical systematization [66]. Traditionally, EEG assessment has focused on frequency and amplitude measurements using simple metrics, which, however, have significant limitations, especially when dealing with large amounts of data [72]. In such cases, it is necessary not only to reduce the volume of data to be analyzed but also to verify the relationships between internal and external factors and the phenomena identified in the signal. To address these needs, a more complex form of signal analysis is required, which may also involve elements of pattern recognition. Naturally, the method chosen for the analysis must align with the specific goal of the analysis itself. In the next sections, we'll focus on providing a short background on some well-established techniques for EEG cleaning and processing, which have been employed in the works of this thesis.

Preprocessing A fundamental challenge of EEG is that neuronal signals generated in the cortex must pass through several layers of tissue with varying electrical properties and complex geometries before reaching the scalp. This process causes the signals to become attenuated and distorted, meaning EEG is less sensitive to deep cortical activity. EEG recordings are also susceptible to interference from other bodily signals, such as eye movements, cardiac activity, and muscle contractions, as well as from environmental noise. Additionally, temporary electrode detachment can degrade signal quality and further obscure relevant EEG data.

While biological artifacts have characteristic waveform shapes and can therefore be easily identified, non-physical artifacts exhibit a wide variety of morphologies, which can distort or obscure normal EEG activity. In more severe cases, artifacts can make the recording uninterpretable. The recognition and removal of artifacts in EEG traces are the focus of numerous studies and remain an ongoing challenge [73]. One effective tool for identifying these artifacts is the Independent Component Analysis.

Independent Component Analysis If the signals were modeled as a linear composition of statistically independent sources, their activities could be isolated through the use of Independent Component Analysis (ICA), a computational method developed to separate a multivariate signal into individual additive sub-components.

Decomposing the data using ICA (or any linear decomposition method) involves a linear transformation of the data from the individual scalp channels into a spatially transformed basis.

In the original recording, each signal represents the time course of voltage differences between the projections of the source onto a channel and one or more reference channels.

After ICA decomposition, each row of the data matrix represents the time course of activity of a spatially filtered process, localized to the channel associated with the component. In ICA decomposition, the independent filters are chosen to produce signals that are maximally independent in time across each channel. The information sources can represent synchronous or partially synchronous activity within one (or possibly more) cortical patches, or activity from non-cortical sources (e.g., potentials induced by eye movements or produced by individual muscle activities, line noise, etc.).

Mathematically, the observed multivariate signal $\mathbf{X} = [x_1(t), x_2(t), \ldots, x_m(t)]^T$ can be modeled as a linear mixture of independent source signals $\mathbf{S} = [s_1(t), s_2(t), \ldots, s_n(t)]^T$ as follows:

$$\mathbf{X} = \mathbf{AS} \tag{2.1}$$

where $\mathbf{A}$ is an unknown mixing matrix, and $\mathbf{S}$ represents the latent source signals that need to be recovered. The goal of ICA is to estimate both $\mathbf{A}$ and $\mathbf{S}$ based on the observed mixed signals $\mathbf{X}$. To achieve this, ICA assumes that the components in $\mathbf{S}$ are statistically independent and that at most one of them follows a Gaussian distribution.

The task of ICA is to find an unmixing matrix $\mathbf{W}$ such that:

$$\mathbf{S} = \mathbf{WX} \tag{2.2}$$

where $\mathbf{W}$ is the inverse of the mixing matrix $\mathbf{A}$, or an approximation of it. The estimation of $\mathbf{W}$ is often performed by maximizing a measure of non-Gaussianity, such as, for example, kurtosis. One popular method for this estimation is the FastICA algorithm [74], which iteratively optimizes the independence of the extracted components. The measure of non-Gaussianity is typically computed as:

$$J(\mathbf{S}) = \sum_{i=1}^{n} [G(s_i) - E[G(s_i)]] \tag{2.3}$$

where G is a non-quadratic function, and $E[G(s_i)]$ is the expected value of the non-Gaussianity function for each component s_i.

In summary, ICA transforms the mixed EEG signals $\mathbf{X}$ into statistically independent components $\mathbf{S}$, enabling the separation of neural activity from various noise sources and artifacts. The effectiveness of this technique lies in its ability to exploit the statistical independence of the underlying sources, which is a key assumption in brain signal decomposition.

Spectral analysis Due to their inherent complexity, EEG time series can be treated as realizations of a stochastic process [66]. Their statistical properties are often analyzed using traditional signal processing methods, including probability distributions and their moments (such as mean, variance, and higher-order moments), correlation functions, and power spectra. Estimating these observable parameters typically assumes stationarity, meaning the statistical characteristics of the signal do not change over the observation period. However, it is important to note that EEG is classified as a stochastic and stationary signal only over short intervals, particularly when recorded under constant conditions.

Among the most common methods for EEG spectral analysis are the Fourier Transform (FT), autoregressive (AR) and autoregressive moving average (ARMA) models, Kalman filters, and time-frequency methods. Fourier analysis can only be applied to EEG signals if short signal windows are considered, within which stationarity can be assumed, as we said. To achieve this, a window function $w(t)$ of width $\Delta\tau$ is selected and moved incrementally by τ across the signal. For each τ, the Fourier Transform is computed within the window, then shifted, and repeated across the entire signal, a technique known as Short Time Fourier Transform (STFT).

The use of a fixed-duration time window introduces a trade-off between time and frequency resolution: narrow windows provide high temporal resolution but poor frequency resolution, while wider windows yield better frequency resolution at the expense of temporal precision. Moreover, large windows can violate the assumption of stationarity. The primary limitation of STFT is its use of a fixed-width window, resulting in constant time and frequency resolution across the signal.

The Continuous Wavelet Transform (CWT) provides an optimal compromise between time and frequency resolution, making it highly effective for non-stationary signals like EEG.

Unlike the Short Time Fourier Transform (STFT), which uses a fixed window, the CWT provides a multi-resolution analysis by adapting the time and frequency resolution according to the scale of the wavelet. The CWT of a signal $x(t)$ is defined as:

$$\mathrm{CWT}(a,b) = \int_{-\infty}^{\infty} x(t)\, \psi \left(\frac{t-b}{a} \right)^{*} dt \tag{2.4}$$

where a is the scale parameter (related to frequency), b is the translation (time shift), and $\psi(t)$ is the mother wavelet. The function ψ^* represents the complex conjugate of the wavelet, and the integral computes how well the signal matches the wavelet at different scales and times.

The resulting coefficients can be visualized in time-frequency graphs called scalograms, which provide a visual representation of the signal's energy distribution across different scales and times. Examples of scalograms are shown in Figure 2.3.

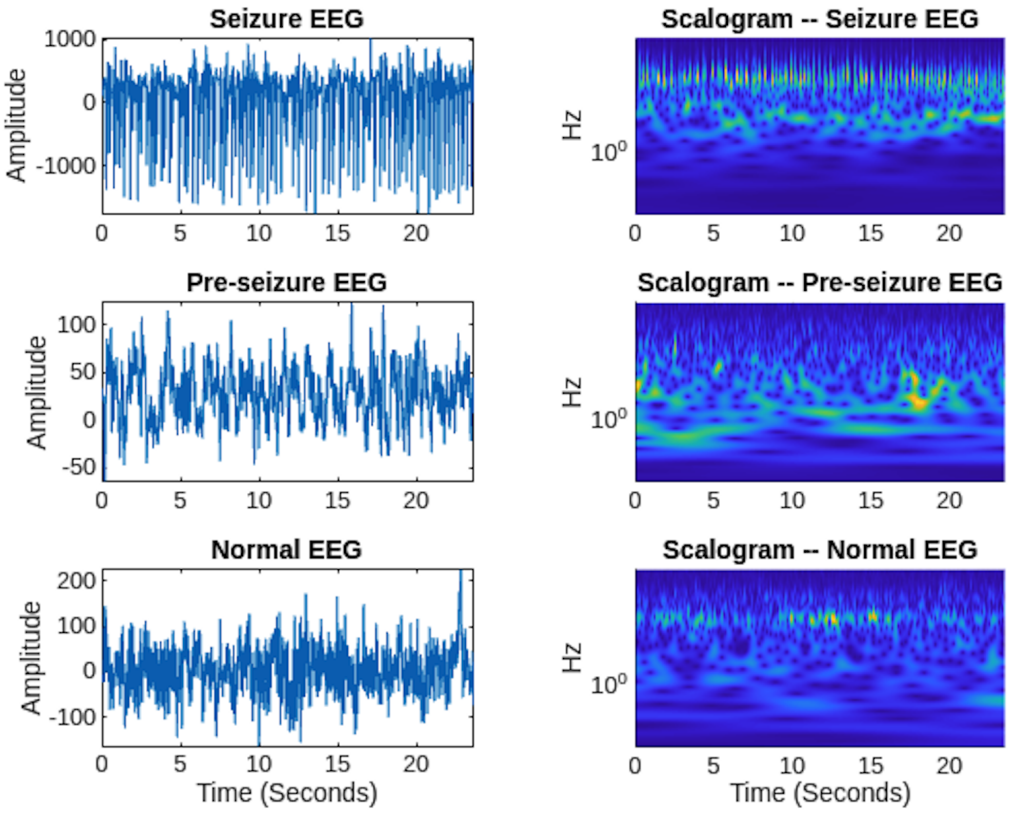


Fig. 2.3 Example of raw EEG signal segments and corresponding scalograms obtained with CWT.

In EEG analysis, CWT is particularly advantageous due to its ability to capture both low-frequency components (with wide time windows) and high-frequency components (with narrow time windows), thus adapting to the nature of the signal. Low-frequency components, such as delta waves, can be resolved with good frequency precision, while high-frequency components, like gamma waves, are resolved with better temporal precision. This multi-resolution property of CWT is highly effective for characterizing the transient and oscillatory behaviors often seen in EEG signals, especially during cognitive tasks or seizure detection.

2.2.3 Current Challenges in EEG Processing

Although many methods have become standard in the field, processing EEG signals still represents a complex task, with major challenges in signal quality due to low SNR and diverse noise sources. High-quality electrodes, optimized placement, and advanced denoising algorithms, have allowed to increase the quality of EEG and reduce the complexity of processing steps.

However, numerous challenges remain open [75]:

1. Signal complexity: the inherently high-dimensional nature of EEG data, which arises from multiple electrode placements and sampling frequencies, poses substantial challenges for subsequent analysis. The sheer volume of data complicates the extraction of meaningful features and patterns. Furthermore, EEG signals exhibit significant temporal variability, reflecting both intra- and inter-individual differences. This variability complicates the identification of consistent neural patterns across different recording sessions or experimental conditions.

2. Feature extraction, selection and dimensionality reduction: the challenge of identifying pertinent features from raw EEG data is magnified by the need for dimensionality reduction techniques that retain critical information while discarding redundant or irrelevant data.

3. Individual variability: considerable variability in EEG patterns exists across individuals, influenced by factors such as age, sex, and neurological health. This inter-subject variability complicates the development of generalized models for neural activity classification and interpretation. Also, variability in EEG signatures associated with different neurological and psychiatric conditions necessitates the use of individualized analytical approaches to enhance diagnostic accuracy.

4. Real-time processing: for applications such as brain-computer interfaces (BCIs), achieving real-time data processing and analysis is imperative. However, this necessitates the implementation of computationally efficient algorithms that can operate with minimal latency and manage manage large-scale datasets effectively through scalability.

Addressing these multifaceted challenges requires the development of advanced computational techniques, improved acquisition protocols, and a concerted effort to standardize methodologies within the field of EEG research and clinical application.

The next part of the Chapter will describe how novel computational methods can be applied to overcome these drawbacks in EEG analysis to construct reliable computer-aided EEG diagnostic systems.

2.3 Deep Learning for time-series analysis

Time series data, characterized by their inherent temporal ordering, are fundamental to a wide range of tasks involving human cognitive processing [76]. Indeed, any classification problem involving data recorded with an inherent sequence can be framed as a time-series classification task.

Building on the success of deep neural networks (DNNs) in Computer Vision, extensive research has introduced various DNN architectures for natural language processing (NLP) tasks, including machine translation [77, 78] and learning word embeddings ([79]. DNNs have also significantly advanced speech recognition, offering powerful models for acoustic processing [80]. DL methods are representation-learning techniques that involve multiple levels of abstraction, derived from nonlinear modules that transform raw data inputs into higher-level representations. As the model's depth increases, the extracted information becomes progressively more abstract. While traditional ML techniques perform well across a wide range of tasks, they are limited in their ability to process raw data without heavily relying on feature extraction processes, which convert input into a suitable representation for classifiers to recognize and discern specific patterns. On the other hand, DL techniques are designed to automatically learn hierarchical representations from raw data, eliminating the need for extensive manual feature engineering. In time series analysis, the model's depth allows it to progressively extract features that capture patterns across multiple scales, from local, short-term fluctuations to more abstract, long-term dependencies.

Figure 2.4 depicts a general scheme of time-series processing with DL models.

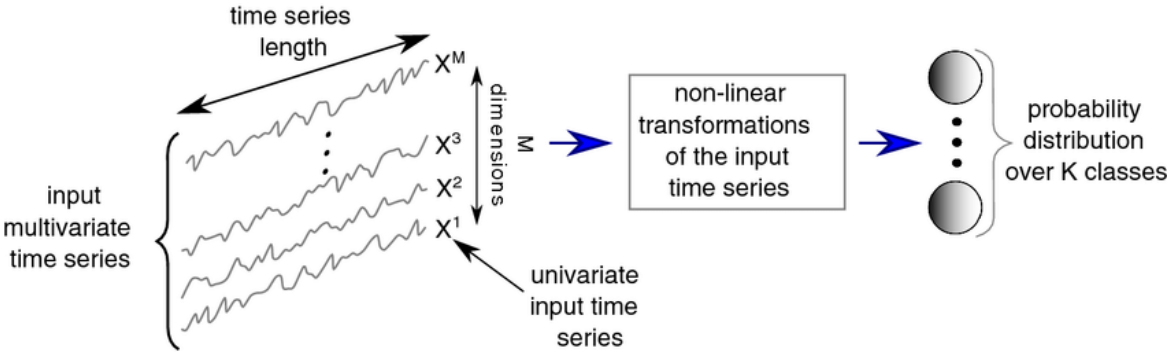


Fig. 2.4 General framework for DL-based time-series classification. From [81].

2.3.1 Evaluation metrics

In a framework exploiting DL models for time series classification, performances are evaluated using different metrics. In a binary classification task, with Positive (P) and Negative (N) classes, the results obtained from a classification system may be organised a *Confusion Matrix*, as the one reported in Table 2.1.

Table 2.1 Confusion Matrix

		Predicted Condition	
		Negative	***Positive***
True Condition	***Negative***	*TN*	*FP*
	Positive	*FN*	*TP*

Specifically, *True Positive (TP)* indicates the number of instances labelled as Positive, and correctly classified as Positive; *True Negative (TN)* refers to the number of instances labelled as Negative and correctly classified as Negative; *False Positive (FP)* refers the number of instances labelled as N but misclassified as Positive; *False Negative (FN)* indicates the number of instances labelled as P but classified as Negative.

Starting from the confusion matrix, several metrics are computed. In the following, Equations 2.5, 2.6 and 2.7 report how to compute Accuracy, Specificity, Sensitivity and F1-Score, which will be used in the subsequent Chapters.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2.5}$$

$$Specificity = \frac{TN}{N} = \frac{TN}{FP+TN} \tag{2.6}$$

$$Sensitivity\,(or\,Recall) = \frac{TP}{P} = \frac{TP}{TP+FN} \tag{2.7}$$

$$Precision = \frac{TP}{TP+FP} \tag{2.8}$$

$$F1\text{-}Score = \frac{2*TP}{2*TP+FN+FP} \tag{2.9}$$

2.3.2 Deep Learning for EEG classification

To overcome the challenges described in Section 2.2.3, new approaches are required to improve the processing of EEG towards better generalization capabilities and more flexible applications. The use of Deep Learning for EEG signal decoding and classification has increased exponentially over the years, with the development of intelligent systems for various types of clinical and non-clinical applications, such as tasks involving the recognition of emotional states [82, 83], sleep stages [84] , or motor imagery [85]. The hierarchical structure of DNNs allows features to be learned on raw or minimally preprocessed data, avoiding neural information to be lost or overlooked during feature extraction and selection pipelines [86].

The different approaches vary both in the choice of architectures and in the formulation of the input. The reviews proposed by Craick *et al.* [87] and Roy *et al.* [86] provide a detailed analysis of the published studies, offering some guidelines for design choices based on the task and the desired outcomes.

In general, the most employed architectures are Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Deep Belief Networks (DBNs). For CNNs, the input is mainly generated in the form of spectrograms or scalograms (see Section 2.2.2), whereas DBNs have shown better results when working with raw or time-domain averaged signals.

Furthermore, both studies highlight that a compelling open question in the field is whether heavy preprocessing EEG data is still necessary if DL models can effectively extract relevant features from raw, unfiltered data. Although explicit artifact removal steps may not be necessary in some cases [86], without affecting the models' performances, most studies in literature employ EEG preprocessing methods at different levels. To overcome the issues tied to the variability of data preparation, the works presented in this thesis employ a standardized EEG preprocessing pipeline, which aims to improve robustness and reproducibility of the frameworks. Further details will be provided in the dedicated sections.

More recently, attention-based models have gained success in the field of EEG data analysis [88, 89]. The next sections of the Chapter will detail how these architecture can be effectively exploited for processing EEG signals.

2.4 Attention in DL models

The year 2015 marked a pivotal shift in the evolution of Deep Learning with the rise of attention-based architectures. This shift began with the introduction of the attention mechanism in Neural Machine Translation (NMT) [90, 91] and image captioning [92]. In NMT, the goal is to learn continuous representations of sequences of variable length. At the time, Recurrent Neural Networks such as Long-Short Term Memories (LSTMs) [93] and Gated-Recurrent Units [94], were the dominant models for sequence learning. These RNNs, however, had significant limitations: their outputs depended on previous elements in the sequence, and they lacked the ability to parallelize computations, slowing down training. Additionally, their fixed-size memory struggled with long-range dependencies, creating a bottleneck in performance [95].

NMT models typically employed an encoder-decoder architecture, where both the encoder and decoder were RNNs. The encoder transformed an input sequence into a fixed-length vector, which the decoder then used to generate the output sequence one token at a time. However, this approach faced two key challenges: first, compressing the input sequence into a fixed-length vector often led to information loss [94]; second, it lacked a mechanism for aligning input and output sequences, which is a critical for tasks like translation and summarization [96]. Moreover, the decoder had no way to focus on specific, relevant input tokens when generating the output.

To address these issues, Bahdanau *et al.* [90] introduced a soft attention mechanism, enabling the model to selectively focus on relevant parts of the input when predicting each target word. This extension of the encoder-decoder architecture allowed the model to search over the input sequence, attending to the most important information for generating the target output.

In the following years, attention mechanisms rapidly expanded across neural network applications, leading to the later self-attention mechanism introduced by Vaswani *et al.* [95] in the *Transformer* architecture, which modeled interactions across the entire input sequence.

Since then, self-attention mechanisms and Transformers have become essential to sequence modeling, allowing to capture dependencies between input and output sequences, and revolutionizing the way networks process sequential data, including physiological signals.

2.4.1 Transformers and Vision Transformers

As shown in Figure 2.5, the core of a Transformer consists of an encoder and a decoder with several blocks of the same type. The encoder generates encodings of inputs, while

the decoder generates the output sequence from the encodings. Each transformer block is composed of an attention layer, a feed-forward neural network, shortcut connection and layer normalization. The attention layer is based on the concept of self-attention, which computes an attention function of the inputs to retrieve the dependencies of each element to the others.

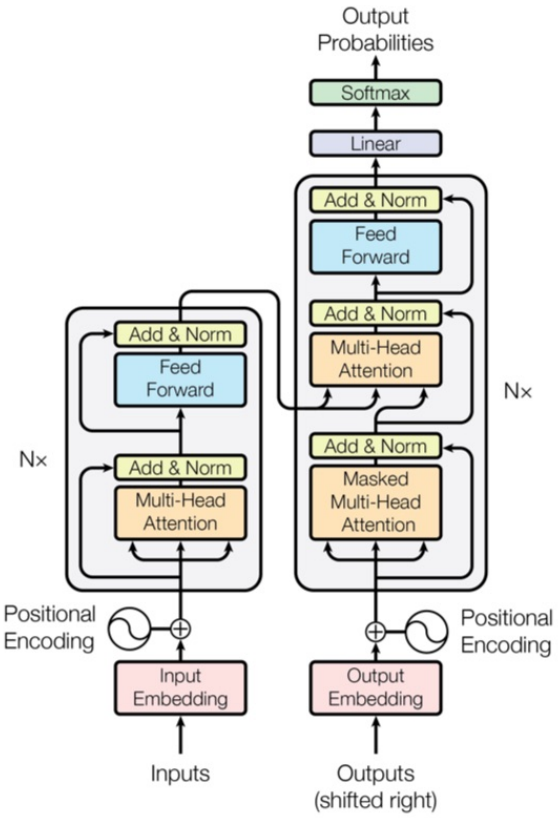


Fig. 2.5 Original Transformer architecture. Image from Vaswani *et al.* [95]

Specifically, the input vector is first transformed into three different vectors: the query vector q, the key vector k and the value vector v with dimensions $dq = dk = dv$. Vectors derived from different inputs are then merged together into three different matrices, namely Q, K and V. Subsequently, the attention function between different input vectors is calculated according to Equation 2.10.

$$Attention(Q,K,V) = softmax(\frac{Q \cdot K^T}{\sqrt{dk}}) \cdot V \tag{2.10}$$

The function computes scores between each pair of inputs, and these values impact how much attention we give to other inputs when encoding the current input. These scores are normalized for gradient stability and then translated into probabilities using the softmax

function. Finally, each value vector is multiplied by the sum of the probabilities. The subsequent layer focuses on vectors with higher probability.

The original Transformer employs layers of Multi-Head Attention (MHA), which generalise the concept of attention by computing different representation subspaces using H randomly initialized query, key and value matrices, where H is the chosen number of heads. These representations are then concatenated to feed the classification layer. This method allows the model to focus on one or more specific input positions without influencing the attention on other equally important positions at the same time.

The first fully self-attention-based architecture for Computer Vision, known as the Vision Transformer (ViT), was introduced by Dosovitskiy *et al.* [97]. In this model, an input image is divided into smaller image patches, referred to as visual tokens, which are then processed sequentially by the Transformer network. ViT directly applies the MHA mechanism to sequences of image patches for image classification tasks. Few modifications are implemented to the original architecture, even though only the transformer encoder module is kept. In such model, sequences of image patches are treated as sequences of words in NLP. 2D images are reshaped into a series of patches of dimension where C is the number of image channels, (P, P) is the resolution of each image patch, and N is the total number of resulting patches.

Instead of treating individual pixels as tokens, which would make attention computation prohibitively expensive due to its quadratic scaling with pixel count, ViT uses patches of 16×16 pixels. Each patch is flattened and linearly projected into a vector of a fixed dimension (Figure 2.6).

Since the Transformer architecture is inherently unaware of the spatial arrangement of these patches within the original image and MHA is permutation-equivariant with respect to its inputs, position embeddings are added to capture the 2D structure. ViT learns these positional relationships during training. Additionally, a learnable class embedding token, is prepended to the sequence of patches. This token is trained alongside the patches and ultimately assists in predicting the classification label via a multi-layer perceptron (MLP) head.

2.5 Explainability

In Chapter 1, we highlighted that as the use of black-box DL models has become more prevalent in high-stakes decision making, the need for greater transparency from these systems is emphasized. The risk arises when models generate decisions that are not justifiable,

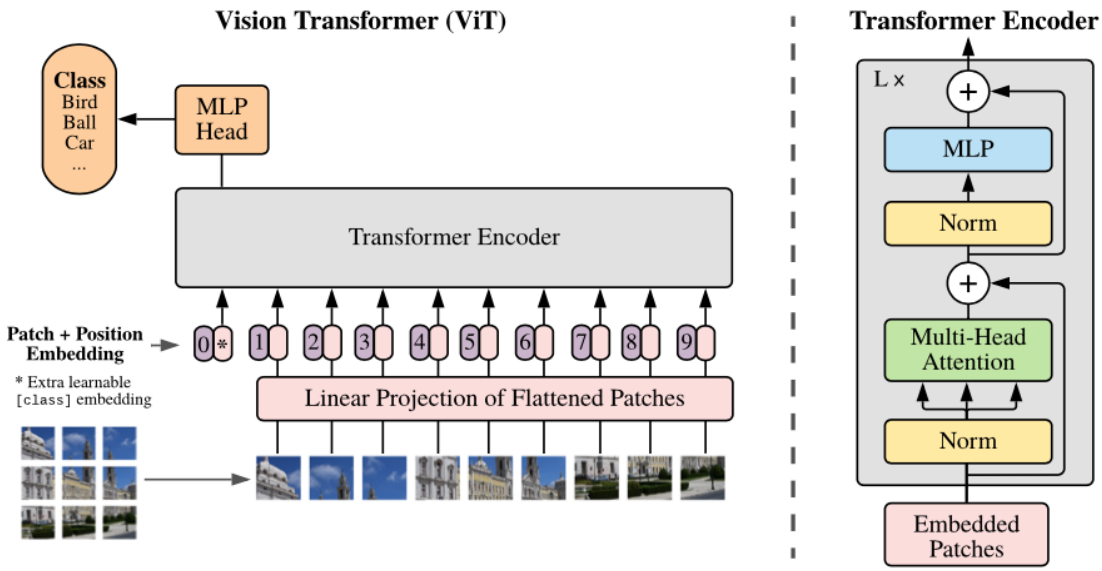


Fig. 2.6 Vision Transformer architecture. Image from Dosovitskiy *et al.* [97]

lack legitimacy, or fail to offer meaningful explanations of their behavior. In fields like precision medicine, providing detailed explanations is critical, as experts require more than just a binary prediction to support their diagnoses and make informed decisions.

The first obstacle to establishing a foundational understanding of eXplainable AI (XAI) lies in the interchangeable use of terms such as *interpretability* and *explainability* in the literature [98].

Interpretability refers to a model's ability to provide clear insights into its decision-making process for human users. An interpretable model is able to show how a decision is made for a specific input [99], by exposing the inner mechanisms through human-understandable explanations. Explainability serves as the interface between the model and the end-user, ensuring that users can understand the behavior of a system and receive clarification about why a decision is made by the model.

In the field of Computer Vision, two broad aims of work on interpretability have been recognized in the literature: transparency and post-hoc interpretation. Transparency addresses how a model functions internally, whereas post-hoc interpretations concern how the model behaves and the reasons behind its behaviour. Post-hoc methods generate explanations after the model has already been trained. Rather than modifying the original model, an external or surrogate model is used to mimic its behavior and generate explanations for users.

Post-hoc methods can be further divided into two subcategories: model-agnostic and model-specific. Model-agnostic methods are adaptable and can be applied to any model,

while model-specific methods are tailored to particular architectures. Model-agnostic post-hoc methods have gained popularity in recent research due to their flexibility and ease of integration with existing models.

Furthermore, based on how explanations can be achieved, a taxonomy of post-hoc XAI methods includes:

- Perturbation-based Methods. Models like Local Interpretable Model-agnostic Explanations (LIME) [100] approximate complex models with simpler, interpretable ones. LIME explains a black-box system by analyzing its response to small perturbations of an input. The resulting data is then used to build a local linear model that acts as a simplified proxy for the original model within the input's neighborhood. One other example is Shapley Additive Explanation Values (SHAP), which quantifies the contribution of each input features to prediction based on the Shapley values from game theory [101]. These methods, however, are computationally intensive and prone to overfitting [102].

- Backpropagation-based Methods. These techniques decompose the model predictions by first backpropagating the gradients from the predictions into input feature space and then visualizing the weights of these features in raw input. One example is GradCAM, a class-specific technique that integrates input features with the gradients of a network layer to generate explanations [103]. Due to its class-specific focus and reliable outputs, GradCAM is widely used in downstream applications like weakly-supervised semantic segmentation. However, this method relies solely on gradients from the deepest layers, leading to coarse results when these low-resolution layers are upsampled.

A third interpretability approach involves designing networks with architectures that inherently simplify the understanding of their behavior.

Indeed, while most of the explainability methods focus on giving information about how a model processes data or how it represents data internally, attention-based architectures generate explanation-producing systems by directly revealing which information flows through the network [104]. Specifically, attention can help access a model's inherent processes by showing how it assigns different weights to different inputs and parts of the input [105].

This is particularly essential for EEG-based systems, because it assesses whether the model has learned physiologically meaningful features. Foremost, interpretability allows checking whether the predictive logic of AI models conforms to specific proven physiological rules, since the predictive accuracy scores of the AI models can be deceptive.

Chapter 3

Deep Learning for the classification of SCD and MCI using rsEEG

After the introduction about the clinical context, highlighting the need for automatic identification of EEG biomarkers for diagnosing and monitoring the progression of cognitive impairment, and the description of methodologies for achieving the set goals, this Chapter reports part of the research works conducted in the field of interest of this thesis during the Ph.D. activities.

Specifically, starting from the results obtained in a precursory work based on a more traditional pipeline that involves a signal-to-image transform, the first Deep Learning framework for classifying HC, SCD and MCI subjects based on their raw resting-state EEG signals is presented.

3.1 Motivation

As described in Chapter 2, Subjective Cognitive Decline and Mild Cognitive Impairment are recognized to be part of the taxonomy of Alzheimer's disease. While MCI refers to a well-defined, intermediate stage between normal ageing and pathological status [30], many patients experience a subjective cognitive decline in memory and other cognitive domains prior to demonstrable impairment. SCD is not linked to a particular disease status itself [33]. However, it has been proved that the subjective decline, even at the stage of normal cognitive performance on mental tests, is associated with an increased risk of positive biomarkers for Alzheimer's and later conversion to dementia [106–109]. In this context, it has been established that SCD can occur at late stages of preclinical AD, before MCI is reached. This phase can be also referred to as pre-MCI or pre-prodromal AD. In particular, since

new diagnostic guidelines have been released, SCD individuals with pathological Aβ levels in cerebrospinal fluid (CSF) could be considered to be in Alzheimer's disease continuum (Section 2.1.1.2). Although the task of classifying SCD and MCI subjects from healthy controls has been addressed in several studies [110, 111], the discrimination between SCD and MCI conditions from a functional point of view is still poorly investigated in literature since anatomical and functional changes in brain between the two classes are subtler, making it a more challenging task to deal with [112]. Nevertheless, the intricacy of brain alterations in the early stages of AD makes it difficult to recognize patterns and develop accurate indicators for diagnosing and monitoring the development of AD on an individual basis [113, 114]. Furthermore, whilst advanced neuroimaging methods like PET and MRI enable to capture relevant modifications in brain processes related to AD, their use is limited in clinical settings due to cost, invasiveness and time consumption [14].

In this section of the thesis, a novel Deep Learning approach that employs a redesigned Transformer architecture for classifying rsEEG signals of HC, SCD and MCI subjects is described.

The rest of this Chapter is organised as follows: Section 3.2 summarizes the recent literature in the field and highlights the limitations of previous works. Section 3.3 provides a detailed description of the dataset employed to conduct research in this field. The last two sections, Section 3.4 and Section 3.5, present the innovative contributions to the field.

3.2 State of the art

Despite longitudinal studies have assessed the increased risk for both SCD and MCI patients to develop Alzheimer's dementia, to the best of our knowledge a limited number of works have investigated changes of distinctive biomarkers to differentiate early AD stages.

Yue *et al.* evaluated the extent of asymmetry of hippocampus and amygdala volumes from MRI scans in HC, SCD and MCI subjects [115]. They found significant differences between the latter two groups only when considering asymmetry of hippocampus, indicating that this marker could help the diagnosis of early AD stages. On the other hand, they found significant differences between HC and SCD in the volume of the right hippocampus, right amygdala and asymmetry of amygdala, and those differences were reflected in the comparison of HC and MCI. In a recent study by Li *et al.*, an approach based on ML models was exploited on features extracted from MRI data to predict the scores of cognitive tests, i.e. Mini-Mental State Examination (MMSE) or Montreal Cognitive Assessment (MoCA), of HC, SCD and MCI subjects, respectively. Results showed that imaging volumetric features of the brain

were more correlated with the scores of cognitive tests than individual features extracted from brain subregions, such as the hippocampal area [116]. Such neuroimaging-based studies, although allow to characterize SCD and MCI effectively, still require time-consuming and expensive techniques to acquire data and thus are not easily replicable.

A study by Scheijbeler *et al.* [117] used magnetoencephalography (MEG) data to compute brain network interactions in SCD and MCI patients by means of a permutation index, called inverted joint permutation entropy (*JPE_inv*), which was used to train a logistic regression model. The area under the roc curve (AUROC) value obtained with this index (0.784 for SCD-MCI classification), was higher when compared to other MEG markers. However, a limited number of 18 SCD and 18 MCI subjects was employed and thus a replication of their method on larger samples is needed.

Even fewer works have focused on the role of EEG-derived biomarkers in the classification of SCD and MCI, although a lot of work has been done in discriminating AD subjects from both MCI and HC [51, 118] also employing DL models [119–121].

Recently, quantitative electroencephalography (qEEG) was used by Engedal *et al.* to predict the conversion to dementia from a large dataset composed of 200 HC, SCD and MCI subjects for whom follow-up information was available [122]. Spectral features were extracted from the signal to calculate a Dementia Index (DI), and a statistical pattern recognition method was employed to evaluate the predictive power of the index, reaching an accuracy of 69 % in discriminating converters from non-converters. However, Engedal *et al.* predicted conversion to dementia from EEG data of subjects already diagnosed. Lazarou *et al.* [114] investigated the power of graph metrics derived from High-Density EEG (HD-EEG) to discriminate among HC, SCD, MCI and AD individuals. They expected to find differences in brain connectivity in terms of correlation matrices constructed from the EEG activity. The statistical analyses showed that SCD individuals present network values intermediate to HC and MCI, underlying a common disconnection pattern of the brain connectome in SCD but not to the same extent as in MCI. Nonetheless, in the SCD *vs* MCI comparison, classification performances of both local and global network measures, evaluated with AUROC values, were lower than 60 %. Similarly, Abazid *et al.* investigated connectivity links in the brain networks derived from rsEEG of SCD, MCI and AD patients by exploiting measures of statistical entropy and a Support Vector Machine (SVM) to discriminate the classes of patients. They demonstrated the effectiveness of the entropy measure to identify different stages of cognitive dysfunction when considering different graph parameters, reaching high accuracy levels, over 90 % [123]. However, these results depend on several stages of signal

manipulation (e.g. feature extraction, thresholding and selection) which can highly affect the classification performance.

3.3 Data description

The EEG data used in the research works presented in this Chapter, as well as the next one, have been acquired from subjects enrolled in the "PRedicting the EVolution of SubjectIvE Cognitive Decline to Alzheimer's Disease With machine learning (PREVIEW)" project, an ongoing prospective cohort study started in October 2020. [124]. The aim of the project is to investigate baseline predictors and biomarkers of Alzheimer's pathology and progression to MCI and dementia in a large cohort of patients with SCD. Specifically, patients with SCD and MCI self-referred to the Regional Reference Center for Alzheimer's Disease and Cognitive Disorders of Careggi Hospital, Florence. Age-matched healthy subjects were enrolled for cross-sectional comparison. Table 3.1 summarises some clinical and demographic information about the subjects enrolled in the study at the time the analyses were conducted. Patients were classified as SCD according to the terminology proposed by the Subjective Cognitive Decline Initiative (SCD-I) Working Group [33], which requires the subject to self-experience a persistent decline in cognitive capacity in comparison with a previously normal status and unrelated to an acute event, as well as normal age-, gender-, and education-adjusted performances on standardized cognitive tests. Patients were classified as MCI according to the NIA-AA workgroups criteria for the diagnosis of MCI [30], specifically requiring: cognitive concern reflecting a change in cognition reported by the clinician or the patient, objective evidence of impairment in one or more cognitive domains (all patients underwent an extensive neuropsychological investigation, with estimation of premorbid intelligence, and assessment of depression), preservation of independence in functional abilities and no signs of dementia. The study was approved by a local ethics committee and individual informed consent was obtained. Experimental procedures were conformed to the Declaration of Helsinki and national guidelines. Resting-state EEG data were acquired using EBNeuro's GalNt system (EBNeuro, Florence, Italy) with 64 channels digitized at a sampling rate of 512 Hz. Among the 64 electrodes, 61 electrodes covered the whole scalp to record EEG while the remaining ones recorded electrooculographic (EOG) and electrocardiographic (ECG) activity, and thus were not considered for further analysis. ERPs were also acquired during two tasks, namely a 3-choice vigilance task and a standard image recognition task, but were not employed for the purposes of this thesis. The electrodes were placed according to the $10-10$ montage system and electrode-skin impedance was set below 5 $k\Omega$. Subjects

were sat in a reclined chair for approximately 20 minutes. The acquisition protocol was structured to involve both closed and open eyes conditions.

Table 3.1 Clinical-demographic characteristics of the study population. HC: healthy controls; SCD: subjective cognitive decline; MCI: mild cognitive impairment; MMSE: mini-mental state examination; TIB: italian brief intelligence test; SD: standard deviation

Characteristics	*HC* (n = 17)	*SCD* (n = 56)	*MCI* (n = 45)
Age *(mean ± SD)*	64.29 ± 4.77	66.26 ± 8.72	74.26 ± 8.20
Females *(%)*	41.2	78.3	54.3
Age onset *(mean ± SD)*	-	55.15 ± 8.04	62.09 ± 9.97
Years of Education *(mean ± SD)*	15.50 ± 3.78	12.58 ± 3.47	10.18 ± 4.17
MMSE *(mean ± SD)*	28.92 ± 1.19	27.48 ± 2.28	27.52 ± 2.13
TIB *(mean ± SD)*	-	107.22 ± 20.48	111.00 ± 6.01

3.3.1 Data preprocessing

Raw data were preprocessed offline using Matlab R2019b (The Mathworks, Natick, MA, USA) and EEGLAB toolbox v.2021.0. In this work, a standardized pipeline, the PREP pipeline [125], was adapted and employed as a first step to clean the signal. This pipeline uses a robust re-referencing algorithm to interpolate noisy channels and leverages routines from the *cleanline* method to remove line noise components [125]. Although the biggest advantage of this approach is that it removes only deterministic line components, while preserving substantial spectral energy, it can present some drawbacks due to the assumption of signal stationarity [125]. To overcome these limitations, a 50 Hz notch filter was further applied to ensure line noise cleaning. This method can be safely applied on our data since high frequencies of the signal, which could be distorted, were not analysed [125].

A semi-automatic method employing EEGLAB's ICLabel [126] and manual choice of independent components to retain has then been applied to the signals. Lastly, epochs with excessive noise or artifacts were visually inspected and removed.

A schematic representation of the preprocessing pipeline is shown in Fig.3.1.

3.4 Preliminary results

The work entitled *A Deep Learning Framework for the Classification of Pre-prodromal and Prodromal Alzheimer's Disease Using Resting-State EEG Signals* [127] reports some of the

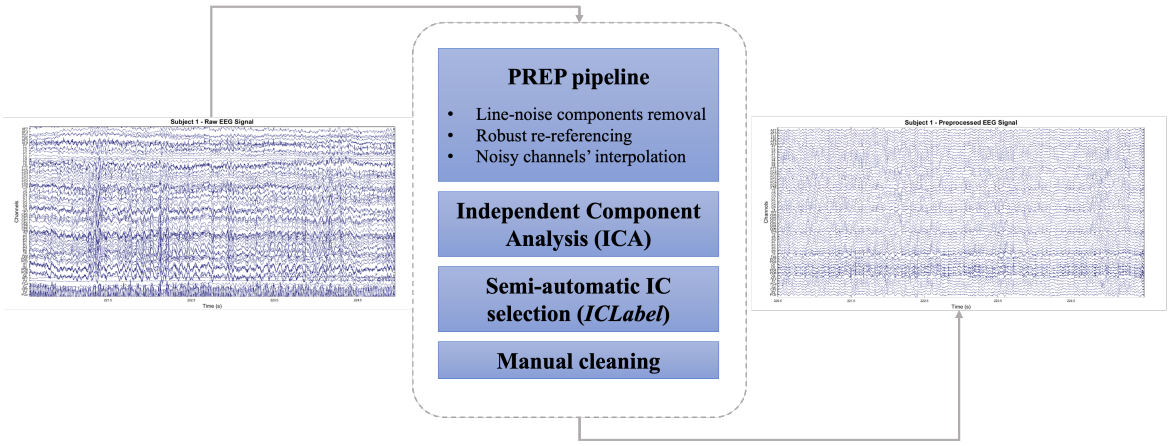


Fig. 3.1 Pipeline of the preprocessing steps applied to the EEG signals.

first results obtained by applying a DL model on rsEEG signals of SCD and MCI. It is worth noting that at the time this manuscript was produced, the collection of the PREVIEW dataset was still in process. Thus, the available EEG recordings of 35 SCD and 32 MCI subjects were employed. The key idea of this work is to use Continuous Wavelet Transform on EEG epochs to produce time-frequency representations of the input signals, and process these images using a Convolutional Recurrent Neural Network (CRNN). Several previously published studies employed DL methods based on CNNs to classify images derived from EEG signals of subjects with AD, MCI and healthy controls. Morabito *et al.* proposed a data-driven CNN based on time–frequency representations of the EEG signal to classify AD, MCI, and HC, reaching an accuracy of 82% in the three-ways classification and up to 85% when considering a binary classification between MCI-HC, MCI-AD, and AD-HC classes [128]. Similarly, a recent work by Ieracitano *et al.* exploited EEG power spectral density to construct grayscale images used as input to a customized CNN, to address the same classification task. They reported an accuracy of 83.3% and compared the results with other conventional machine learning methods (e.g., Linear Discriminant Analysis and Support Vector Machine), showing how the DL framework outperforms state-of-the-art algorithms [129]. Other studies, such as the one conducted by Kim *et al.*, employed Deep Neural Networks to address binary classifications between AD, MCI, and HC with feature- based inputs relating to the Relative Power (RP) of different frequency bands within EEG signals [130]. The maximum accuracy reached by the model was 75%. A work by Huggins et *al.* [131] proved a similar approach based on time-frequency signal representations to be effective in classifying rsEEG epochs of AD, MCI and controls. However, none of them include SCD groups.

Hence, after applying the preprocessing pipeline described in the previous section 3.3.1, 21 channels were selected to reduce the complexity of the input signals. Non-overlapping epochs of 5 s were extracted from each recording, and, for each epoch, a Continuous Wavelet Transform was applied on each EEG channel. The output coefficients of the function were plotted as scalograms, i.e., time–frequency graphs in which the energy of CWT coefficients is represented by different colors. A logarithmic scale and a colorimetric map with 256 colors were then employed to visualize and save the scalograms. The resulting 21 images for each epoch, corresponding to the EEG channels, were then tiled following the 10–10 electrode placement system, in order to retain spatial information of the channels' position on the scalp. The Deep Learning architecture incorporated a pre-trained ResNet-18 model [132], with its final layer connected to a LSTM consisting of one hidden layer with 8 units. The use of a previously trained network as backbone was needed to reduce the risk of overfitting due to the relatively small sample size. This approach enabled the model to treat each EEG recording as a series of sequential frames (i.e., scalograms) that captured the neural activity of each subject (Fig. 3.2). Each sequence was then labeled as either SCD or MCI, with the same label assigned to the corresponding subject. The dataset was split subject-wise randomly into training (70%) and test (30%) sets.

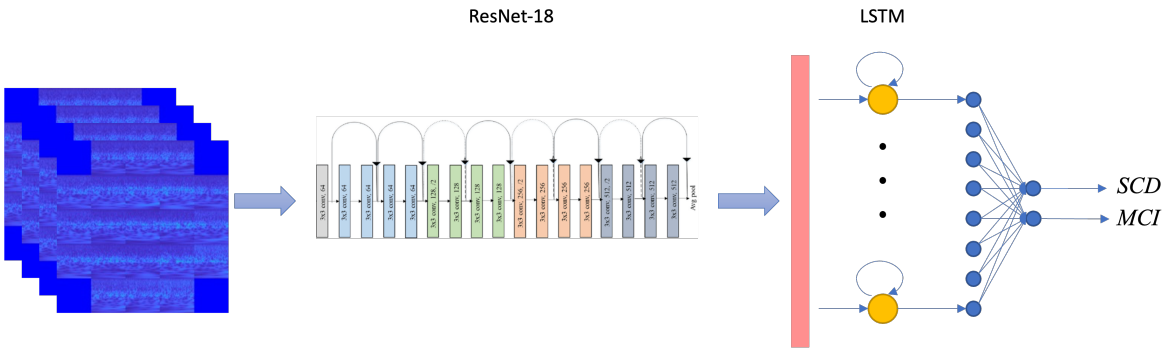


Fig. 3.2 Overall architecture of the proposed model. The time series composed of all the scalograms of a given subject in the dataset is fed to a ResNet-18 model followed by a LSTM layer composed of 8 units. Then, a fully-connected layer classifies each time series either as SCD or MCI

The classifier reached an Accuracy of 75.0%, a Sensitivity of 66.7%, and a Specificity of 81.8% on the test set for the classification of SCD and MCI.

These preliminary results were promising, mostly considering that the task of classifying SCD and MCI patients, as stated before, is very challenging [114] and that most state-of-the-art results obtained with DL approaches concerned the classification of AD, MCI, and HC subjects. Nonetheless, this work presented several limitations. Firstly, the small sample size

limited the the generalization capabilities of the framework. Furthermore, the robustness and reproducibility of the methods needed to be improved. Most importantly, the lack of a healthy control group restricted the assessment of the classifier's ability to decode neural activity when some sort of degeneration was present. The subsequent work was conducted in order to address these limitations and provide a more robust framework for this specific classification task.

3.5 Attention-based approach

The exploratory work presented in the previous section 3.4 paved the way for the application of Deep Learning to discriminate among different levels of cognitive impairment based on resting-state EEG signals. In the study *An attention-based deep learning approach for the classification of subjective cognitive decline and mild cognitive impairment using resting-state EEG* [133], an end-to-end model mainly employed in NLP, the Transformer 2.4.1, and the self-attention mechanism, were exploited to classify resting-state EEG signals of 17 HC, 56 SCD and 45 MCI subjects by focusing on the global patterns of the brain oscillatory activity. For the aim of this work, we extracted and employed only the eyes-closed (EC) epochs of the original signal for all the subjects (*mean length* = 15.03 ± 1.41 min), which represent the largest part of the protocol.

Clean data were processed and a cluster of 19 channels, namely Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2, was selected. Since these channels evenly cover the scalp area, this EEG pattern is the most employed in the literature for similar studies [134] and has been proven to ensure sufficient quality along with possible comparison with previous rsEEG findings of other projects [50]. Subsequently, the signals were bandpass filtered between 0.1 Hz and 45 Hz.

Four main frequency bands, namely Delta (δ) [0.1 - 4] Hz, Theta (θ) [4 - 8] Hz, Alpha (α) [8 - 13] Hz and Beta (β) [13 - 30] Hz were extracted from each EEG signal using designed bandpass filters, and each related dataset was created. Furthermore, in order to assess which frequency band was the most distinctive in the classification of HC, SCD and MCI, we also filtered the signals in the entire range [0.1 - 30] Hz, and an additional dataset (All-band) was generated. Gamma (γ) band [30 - 70] Hz was excluded from the analysis since the EEG signal in this band can be significantly contaminated with muscle artifacts [135]. To design filters, we used the *pop_eegfiltnew* function from EEGLAB, which has a heuristic for automatically determining the filter length and order. This function employs a zero-phase Hamming windowed sinc finite impulse response (FIR) filter [136].

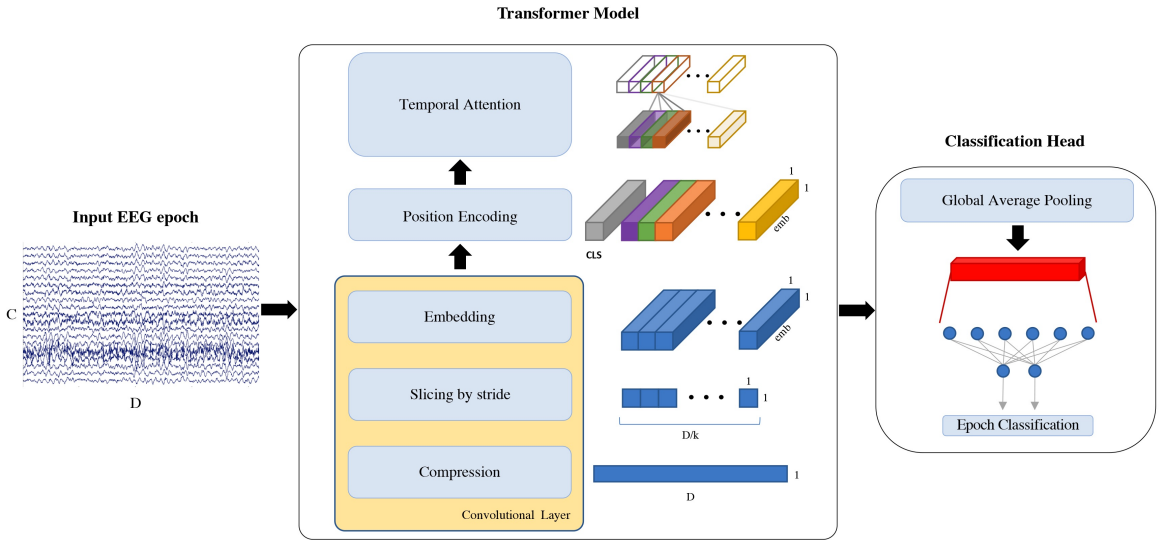


Fig. 3.3 EEG epoch classification pipeline. Each EEG segment of $C = 19$ channels and $D = 5120$ datapoints is used as input to our model, which uses a convolutional layer to compress the signal, extract slices and embed the information. $k = 31$ is the size of the kernel, $emb = 6$ is the embeddings' dimension and CLS is the classification token prepended to the input. Attention mechanism is then applied on the temporal domain and, after global average pooling, a linear layer is used to classify the input EEG epoch.

3.5.1 Proposed model

As described in the previous Chapter, the original Transformer employs layers of Multi-Head Attention (MHA), which generalise the concept of attention by computing different representation subspaces using H randomly initialized query, key and value matrices, where H is the chosen number of heads. These representations are then concatenated to feed the classification layer. This method allows the model to focus on one or more specific input positions without influencing the attention on other equally important positions at the same time.

Following the work by Song *et al.* [137], we implemented a pipeline to classify EEG epochs, as shown in Figure 3.3, by designing and training a modified version of their model on eyes-closed rsEEG signals of SCD and MCI subjects. The same pipeline was followed for the classification of HC, SCD and MCI. For this second task, the last fully connected layer was composed of three output units.

The major difference between the two architectures concerns the way attention is applied to the signals. The proposed model dismisses the spatial attention module, which is used to weight the information encoded by each EEG channel, and prioritizes the temporal domain of the signal. This difference is due to the fact that the objective of this work is to classify

resting-state signals, instead of Motor Imagery (MI) signals as in Song *et al.* [137]. In fact, while different motor imagery processes activate different areas of the cerebral cortex, and thus spatial channel information was revealed to be of fundamental importance when engaging in a MI classification task [138, 139], resting-states reflect the spontaneous brain activity, thus there is not an established spatial correlation also when investigating cognitive decline associated with Alzheimer's disease [140].

Consequently, our model aims to exploit multi-head attention to understand if temporal dependencies of the EEG sequences can highlight discriminative patterns among HC, SCD and MCI subjects. The MHA layer is included in an encoder block, which combines it with a feed-forward module, a normalization layer and dropout. The encoder block is replicated a number of times specified by the $depth$ parameter, which was set to 2, whereas the number of heads was set to 3. It is worth noting that this configuration is of low complexity and reduced computational cost since it requires fewer parameters than traditional CNNs and RNNs. A graphical representation of the implemented Transformer model is shown in Figure 3.4 with reference to SCD *vs* MCI classification.

Similarly to the original Transformer architecture, the proposed model also needs some information on the position of inputs in the time series. This is achieved by Song *et al.* by using a convolutional layer on the time dimension before compression, rather than positional encodings as in the original model [137]. Instead, we use a convolutional layer to embed channels' information, compressing it to a single channel representation, and to extract slices from EEG sequences as shown in Figure 3.3. Then, we encode the positions of all slices in the sequence, and the vector of positions is linearly added to the input. Furthermore, we prepend an extra-learnable classification token to each input sequence, which is used to predict the final class after being updated by attention, as in the ViT [97]. Compared to the original model, this position encoding method requires fewer parameters and avoids the use of an additional convolutional layer, which increases the complexity of the model. After the global average pooling, a classification head composed of a fully-connected layer, after layer normalization, is then used to classify the new representation of the input.

After preprocessing, Leave-One-Subject-Out Cross-Validation was used on the datasets (All-band, Delta, Theta, Alpha, Beta), where all subjects except one were used for training and the remaining for testing. This cross-validation strategy is the most used across studies that employ rsEEG for AD diagnosis and progression analysis [13]. EEG signals were split into 10-second epochs, and random sampling was applied to balance the classes. Z-score normalization was performed for each subject's EEG data, which was revealed to be an

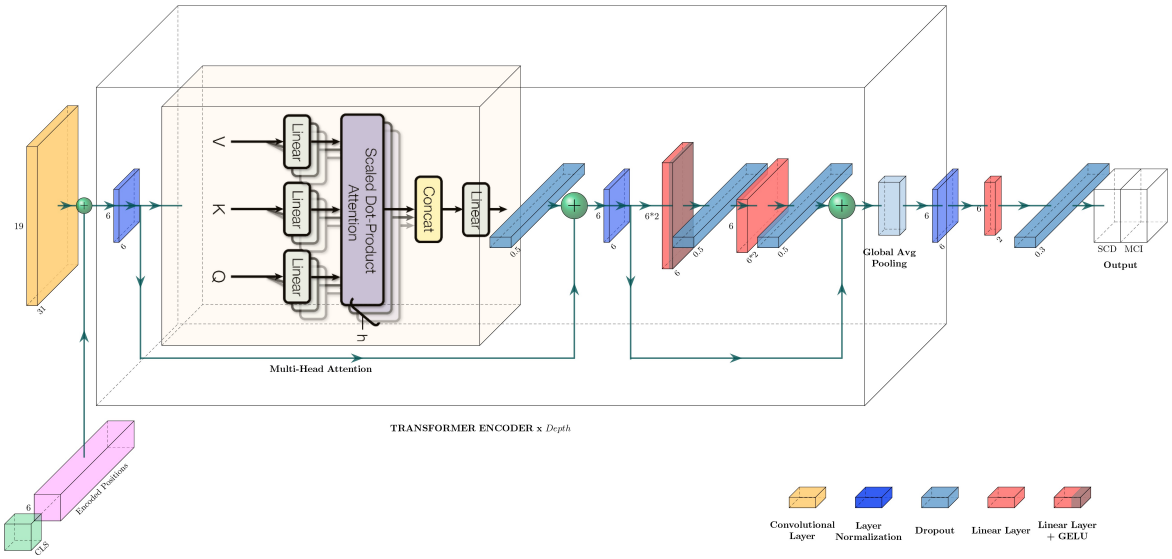


Fig. 3.4 Proposed Transformer architecture. *CLS* is the classification token, $h = 3$ is the number of heads used by Multi-Head Attention and $Depth = 2$ indicates the number of times the transformer encoder block is repeated. A legend for uncaptioned blocks is provided on the bottom right corner.

optimal normalization technique for giving models the ability to make classification across an inter-subject population [141].

The model was trained using the Adam optimizer, with a batch size of 8 and 250 iterations. Cross-Entropy was the loss function, and an early-stop mechanism prevented overfitting. Finally, classification was done at the epoch level, followed by a hard voting mechanism to predict each subject's label.

3.5.2 Results

The classification results are reported in Table 3.3 for all the datasets. The best performances have been reached by the Transformer model on Delta and Theta bands. Specifically, an Accuracy of 67.4% and a F1-Score of 67.3% were obtained for Delta, whereas a value of 65.0% was obtained for both metrics on Theta. AUC scores on Delta as well as on Theta were higher than 0.8, revealing that both the classifiers have an overall excellent diagnostic accuracy in discriminating SCD and MCI [142].

Subsequently, we evaluated the capabilities of the models on the classification of patients, which is the main objective of this study. We report the classification performances for all the datasets in terms of Accuracy (Equation 2.5), Sensitivity (Equation 2.7), Specificity (Equation 2.6) and F1-Score (Equation 2.9). The results are detailed in Table 3.4. On

Table 3.2 Confusion Matrix for SCD *vs* MCI classification

		Predicted Class	
		SCD	*MCI*
True Class	*SCD*	*TN*	*FP*
	MCI	*FN*	*TP*

Table 3.3 Per epoch classification performances. Metrics are computed on the cumulative test confusion matrix. No Information Rate (NIR) = 0.553. The 95% Confidence Interval (CI) was calculated for each set with the Clopper–Pearson method for a binomial distribution (Accuracy $>$ NIR, *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$). F1-Score is weighted for the number of samples per class.

Dataset	**Accuracy**	**CI**	**AUC**	**Sensitivity**	**Specificity**	**F1-Score**
Alpha	0.628	[0.618, 0.638]***	0.779	0.602	0.650	0.629
Beta	0.619	[0.608, 0.628]***	0.744	0.598	0.635	0.619
Delta	0.674	[0.664, 0.683]***	0.807	0.620	0.717	0.673
Theta	0.650	[0.640, 0.660]***	0.802	0.591	0.698	0.650
All-band	0.642	[0.632, 0.652]***	0.779	0.561	0.707	0.640

the Delta band, the model reached the highest value for all the computed metrics, with an Accuracy and F1-Score of 76.2%, a Sensitivity of 73.3% and a Specificity of 78.6%. It is worth noting that, when considering the epochs' classification task, both single-band Delta and Theta datasets perform better than the All-band dataset, upholding the idea that changes in particular EEG rhythms are more discriminative of SCD and MCI conditions and easier to be detected by our model. On patient-level classification, Delta outperforms all the other datasets.

Table 3.4 Per patient classification performances. Metrics are computed on the cumulative test confusion matrix. No Information Rate (NIR) = 0.554. The 95% Confidence Interval (CI) was calculated for each set with the Clopper–Pearson method for a binomial distribution (Accuracy $>$ NIR, *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$). F1-Score is weighted for the number of samples per class.

Dataset	**Accuracy**	**CI**	**Sensitivity**	**Specificity**	**F1-Score**
Alpha	0.653	[0.552, 0.745]*	0.644	0.661	0.654
Beta	0.624	[0.552, 0.718]	0.600	0.643	0.624
Delta	0.762	[0.667, 0.841]***	0.733	0.786	0.762
Theta	0.673	[0.573, 0.763]**	0.600	0.732	0.672
All-band	0.673	[0.573, 0.763]**	0.578	0.750	0.671

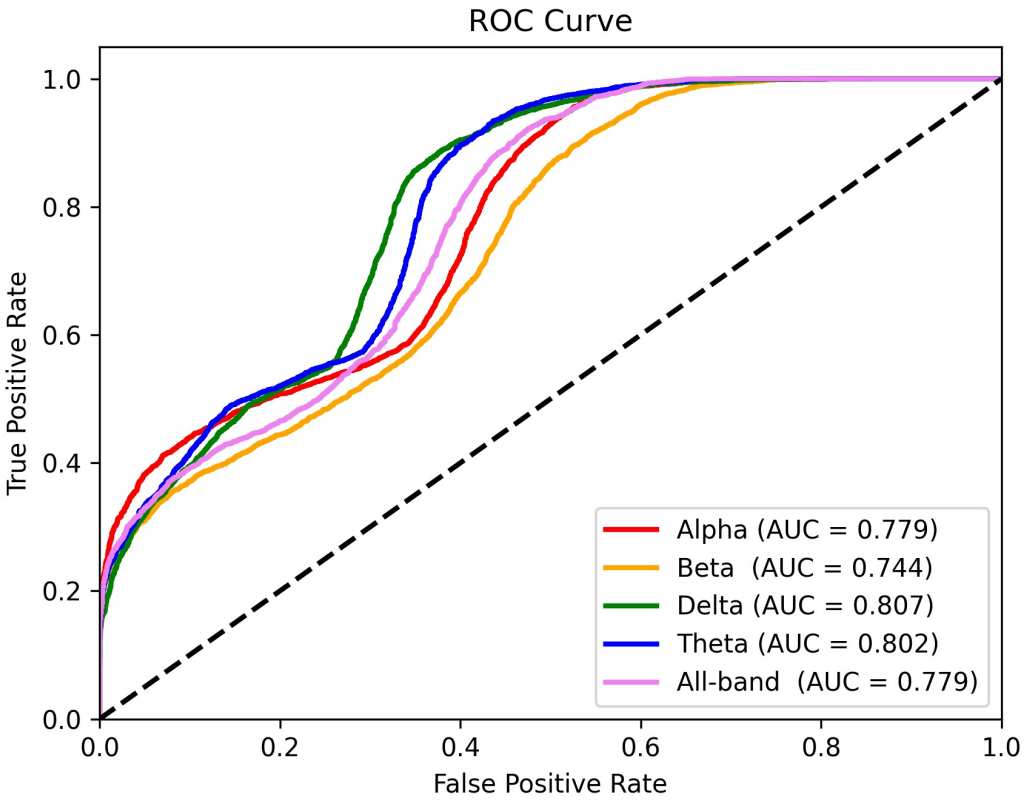


Fig. 3.5 ROC curves for SCD *vs* MCI classification on the cumulative test set.

Lastly, in order to demonstrate the efficacy of the entire workflow, we selected the best-performing frequency bands (i.e. Delta and Theta) and constructed two supplementary high-density EEG datasets, following the same pipeline, but skipping the channel selection step. Specifically, the new EEG segments used as input to the Transformer model had dimensions $C = 61$ channels and $D = 5120$ datapoints. We used the same LOSOCV approach and computed all the metrics in order to compare the results with the previous datasets. On the high-density EEG Delta dataset, we obtained an Accuracy of 62.8% and F1-Score of 62.7% on epochs' classification, while 67.3% and 67.4% were obtained for Accuracy and F1-Score on patients' classification. On the high-density EEG Theta dataset, we obtained 59.8% and 59.7% for Accuracy and F1-Score on epochs' classification, respectively. On patients' classification, Accuracy reached a value of 61.5%, whereas we obtained 61.2% for F1-Score. Although, even in this case, the Delta band shows the best results, all the metrics are lower when compared to the 19-channel datasets, meaning the information added by

using more EEG channels is not useful for our model to perform the classification of SCD and MCI subjects.

3.5.3 HC *vs* SCD *vs* MCI classification

Then, we assessed the performances of our model on the classification of HC, SCD and MCI subjects. As for the SCD *vs* MCI classification, we reported the results for both epochs and patients. In particular, Table 3.5 reports the performances on epochs in terms of Accuracy (Equation 2.5), F1-Score (Equation 2.9) and AUC, and Figure 3.6 shows the corresponding ROC curves. Specifically, the micro-average ROC curve is reported aggregating, for each dataset, the contribution of all classes.

The best performances have been reached by the Transformer model on Alpha and Theta bands. Specifically, an Accuracy of 48.8% and a F1-Score of 49.4% were obtained for Alpha, whereas values of 48.6% and 49.8% were obtained on the Theta band for the same metrics, respectively. AUC scores on both bands were higher than 0.7, revealing that the classifiers have an overall acceptable diagnostic accuracy in discriminating HC, SCD and MCI [142].

Also in this case, we evaluated the capabilities of the models to classify individual subjects. Table 3.6 reports the performances in terms of Accuracy (Equation 2.5) and F1-Score (Equation 2.9), showing that the Theta band has the highest discriminatory power with an Accuracy of 54.2% and a F1-Score of 54.9%.

Table 3.5 Per epoch HC *vs* SCD *vs* MCI classification performances. Metrics are computed on the cumulative test confusion matrix. No Information Rate (NIR) = 0.473. The 95% Confidence Interval (CI) was calculated for each set with the Clopper–Pearson method for a binomial distribution (Accuracy $>$ NIR, *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$). F1-Score is weighted for the number of samples per class.

Dataset	Accuracy	CI	AUC	F1-Score
Alpha	0.488	[0.479, 0.498]***	0.750	0.494
Beta	0.446	[0.436, 0.455]	0.693	0.448
Delta	0.449	[0.440, 0.458]	0.662	0.470
Theta	0.486	[0.476, 0.495]***	0.745	0.498
All-band	0.443	[0.434, 0.452]	0.681	0.455

As in the SCD *vs* MCI classification task, we obtained two 61-channel datasets corresponding to the bands with the best performances, namely Theta and Alpha, and compared the results with the corresponding 19-channel datasets. For Alpha, we obtained an Accuracy of 49.6% and a F1-Score of 50.1% on epochs' classification and 55.1% and 55.3% on

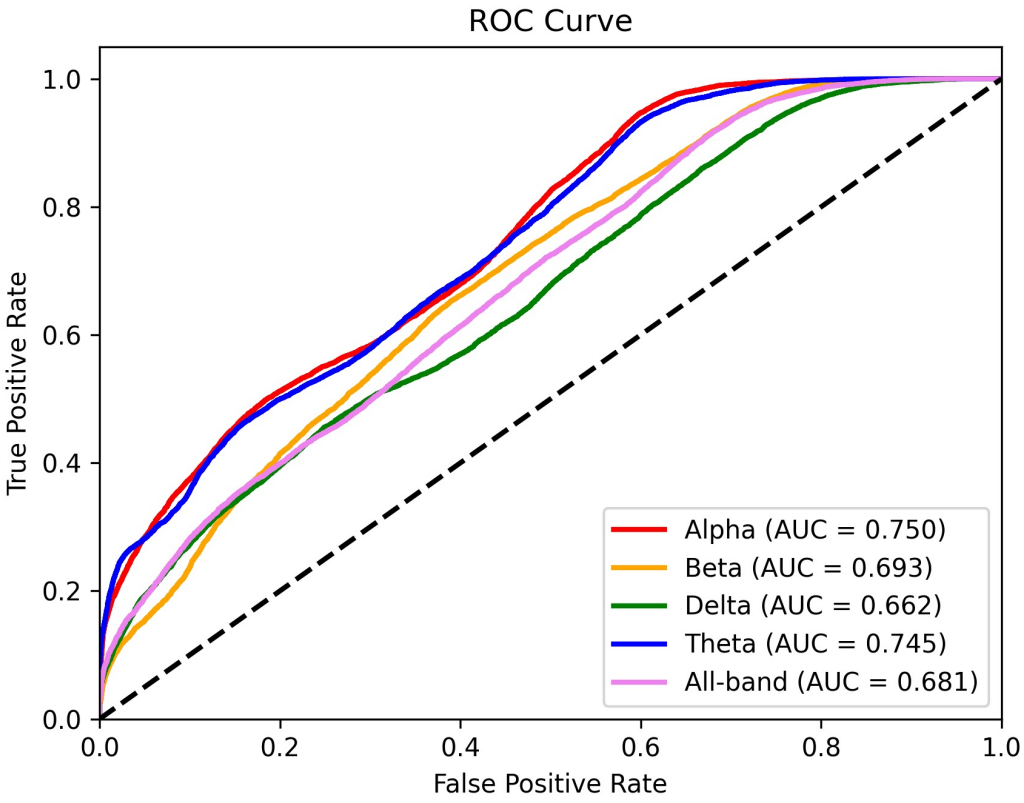


Fig. 3.6 ROC curves for HC *vs* SCD *vs* MCI classification on the cumulative test set.

patients' classification, respectively. For Theta we obtained an Accuracy of 45.9% and a F1-Score of 47.1% on epochs' classification and 48.3% and 49.8% on patients' classification. These results show that, even in the HC *vs* SCD *vs* MCI classification task, using an higher number of EEG channels does not have a significant impact on the performances of our model. In fact, although there was a small increase in the results on Alpha, on Theta, which is the best-performing band on subject-wise classification, our Transformer continues to give the highest results considering the dataset with 19 channels.

3.5.4 Performance Comparison with CNN-based models

In order to compare our model with state-of-the-art EEG classification models, we conducted experiments with some recent CNN-based models, namely DeepConvNet [143], EEGNet [144] and EEG-TCNet [145]. These architectures were mainly developed for MI-based EEG signals decoding, as well as for the classification and interpretation of EEG-based BCIs. Park *et al.* employed them in the field of the identification of preclinical AD from

Table 3.6 Per patient HC *vs* SCD *vs* MCI classification performances. Metrics are computed on the cumulative test confusion matrix. No Information Rate (NIR) = 0.475. The 95% Confidence Interval (CI) was calculated for each set with the Clopper–Pearson method for a binomial distribution (Accuracy > NIR, *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$). F1-Score is weighted for the number of samples per class.

Dataset	Accuracy	CI	F1-Score
Alpha	0.500	[0.407, 0.593]	0.503
Beta	0.500	[0.407, 0.593]	0.503
Delta	0.500	[0.407, 0.593]	0.527
Theta	0.542	[0.448, 0.634]	0.549
All-band	0.517	[0.423, 0.610]	0.532

EEG to overcome the limitation of high inter-subject variability, which affects the possibility of extracting robust handcrafted features [146]. However, to our knowledge, they have never been used for the specific classification task of discriminating SCD from MCI. It is worth noting that these models are characterized by a higher number of parameters than our Transformer. Indeed, while the Transformer contains a total of 5.2 k parameters, DeepConvNet, EEGNet and TCNet have 298.6 k, 9.8 k and 14.1 k parameters, respectively.

The models' parameters were adjusted to take EEG epochs of dimension $C \times D$ in input as our Transformer model. The comparison was performed on the best-performing datasets for both classification tasks, i.e. Delta and Theta for SCD *vs* MCI and Alpha and Theta for HC *vs* SCD *vs* MCI. Results are reported in Table 3.7 and Table 3.8, respectively. For SCD *vs* MCI classification, all the models reached comparable performances in terms of accuracy, which was always significantly higher than no-information rate for epochs classification. For the Delta band, the classification accuracy of patients was 76.2% for the Transformer, while the same metric has values of 73.3% for EEGNet, 65.3% for DeepConvNet and 70.3% for EEG-TCNet. In all the cases, except for DeepConvNet, the accuracy was always significantly higher than no-information rate ($p \leq 0.001$ for Transformer and EEGNet, $p \leq 0.01$ for EEG-TCNet).

Concerning the HC *vs* SCD *vs* MCI classification, the epochs' classification accuracy was significantly higher than no-information rate for the Transformer, for both Alpha and Theta bands, and EEG-TCNet, for the Theta band only ($p \leq 0.001$ for all the cases). However, patients' classification accuracy was not significantly higher than the no-information rate, except for the Transformer which reached a near significance ($p = 0.08$), with a value of 54.2% against 48.3% for EEGNet, 49.2% for DeepConvNet and 50.0% for EEG-TCNet.

Table 3.7 SCD *vs* MCI classification performance comparison in terms of overall accuracy on the cumulative test set of the DL models. No Information Rate (NIR) for epochs classification = 0.553; NIR for patients classification = 0.554. The 95% Confidence Interval (CI) was calculated for each set with the Clopper–Pearson method for a binomial distribution (Accuracy > NIR, *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$).

Model	**Dataset**	**Epochs**		**Patients**	
		Accuracy	*CI*	*Accuracy*	*CI*
Transformer	Delta	0.674	[0.664, 0.683]***	0.762	[0.667, 0.841]***
	Theta	0.650	[0.640, 0.660]***	0.673	[0.573, 0.763]**
EEGNet	Delta	0.726	[0.716, 0.735]***	0.733	[0.635, 0.816]***
	Theta	0.669	[0.659, 0.678]***	0.683	[0.583, 0.772]**
DeepConvNet	Delta	0.590	[0.580, 0.600]***	0.653	[0.552, 0.745]*
	Theta	0.589	[0.579, 0.599]***	0.594	[0.492, 0.691]
EEG-TCNet	Delta	0.673	[0.664, 0.683]***	0.703	[0.604, 0.790]**
	Theta	0.693	[0.683, 0.702]***	0.683	[0.583, 0.772]**

Table 3.8 HC *vs* SCD *vs* MCI classification performance comparison in terms of overall accuracy on the cumulative test set of the DL models. No Information Rate (NIR) for epochs classification = 0.473; NIR for patients classification = 0.475. The 95% Confidence Interval (CI) was calculated for each set with the Clopper–Pearson method for a binomial distribution (Accuracy > NIR, *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$).

Model	**Dataset**	**Epochs**		**Patients**	
		Accuracy	*CI*	*Accuracy*	*CI*
Transformer	Alpha	0.488	[0.479, 0.498]***	0.500	[0.407, 0.593]
	Theta	0.486	[0.476, 0.495]***	0.542	[0.448, 0.634]
EEGNet	Alpha	0.472	[0.462, 0.481]	0.483	[0.390, 0.577]
	Theta	0.451	[0.442, 0.461]	0.424	[0.333, 0.518]
DeepConvNet	Alpha	0.479	[0.469, 0.488]	0.492	[0.398, 0.585]
	Theta	0.476	[0.467, 0.486]	0.500	[0.407, 0.593]
EEG-TCNet	Alpha	0.467	[0.458, 0.477]	0.500	[0.407, 0.593]
	Theta	0.495	[0.486, 0.505]***	0.508	[0.415, 0.602]

3.5.5 General remarks

Previous studies that have addressed the task of discriminating SCD and MCI patients in the AD continuum, with statistical or traditional ML approaches, have highlighted that this problem is much more challenging than other classification tasks in the same field. This evidence can be deducted both from works that employ MRI data [112, 115] and EEG data [114]. It is also supported by other works in literature [51, 118–121], some of which show in general better classification performances than those obtained in this work but considering different classes of subjects, e.g. HC *vs* SCD, HC *vs* MCI or MCI *vs* AD.

For the SCD *vs* MCI classification task, by comparing the results on all the test sets gathered from a LOSOCV approach, we found that Delta and Theta bands had the best performances with AUC values of 0.807 and 0.802, respectively. Furthermore, the other classification metrics, i.e. Accuracy, Sensitivity, Specificity and F1-Score, were the highest on Delta, reporting a value of 67.4% for Accuracy and 67.3% for F1-Score on epoch-wise classification and a value of 72.6% for both Accuracy and F1-Score on patient-wise classification. On the same band, the model reached good Sensitivity and Specificity values, respectively of 73.3 % and 78.6%, showing it is capable of discriminating SCD and MCI subjects when they have that specific condition. For both Delta and Theta bands, the classification accuracy was significantly higher than the no-information rate ($p \leq 0.001$ and $p \leq 0.01$ for epoch-wise and patient-wise classifications, respectively), assessing that the classifier model performed better than one could do by always predicting the most common class. Indeed, changes in relative power in the lower frequencies (δ and θ) indicate a diffuse slowing of brain oscillations, which is a hallmark feature in the progression of AD [13]. In this context, EEG spectral analysis revealed that higher Delta and Theta powers are associated with clinical progression of SCD patients towards MCI and dementia, mainly when considering eyes-closed resting-state activity, as it has been done in this study [53].

Our results uphold this evidence, showing that changes in Delta and Theta are particularly useful in characterizing the brain activity of subjects affected by SCD or MCI, both when compared to other common EEG rhythms and to the All-band dataset, which includes the signals filtered in the range [0.1-30] Hz. Furthermore, the multi-head attention mechanism well captures temporal dependencies of rsEEG, highlighting their importance in the discrimination between SCD and MCI. This is supported also by a recent work by Wei *et al.*, who employed the attention mechanism to classify MCI and HC using EEG signals recorded during cognitive tasks [147]. In fact, this approach allowed to improve the performances of a traditional CNN by almost 10%, suggesting that the use of this technique should be further investigated.

On the other hand, we found that adding more spatial details by using all available 61 EEG channels, instead of a cluster of 19 channels, not only did not improve the performances of the model, but all the metrics reported lower values for both epochs' and patients' classification performed on Delta and Theta datasets. Hence, we showed that more spatial information increases the complexity and redundancy of the signal pattern produced by the selected 19 channels, which already holds enough information for the model to distinguish between the two classes.

In order to further assess the performance of our model, we added a control group of 17 healthy subjects and conducted a multiclass classification to discriminate HC, SCD and MCI simultaneously. We found that, in this case, the best-performing frequency bands were Alpha and Theta both on epochs' and patients' classification tasks. Specifically, on Alpha the AUC was 0.750, and slightly lower for Theta as shown by the ROC curves in Figure 3.6. However, Theta reported the best performances in terms of Accuracy and F1-Score when classifying subjects. In addition, Alpha and Theta bands were the only ones that reached classification accuracies significantly higher than no-information rate ($p \leq 0.001$ for both bands). These results are in line with evidence reported in literature that both SCD and MCI subjects are characterized by lower amplitude of posterior alpha rhythms in rsEEG in relation to cognitive functions when compared to controls [148] and that this feature, along with higher amplitude of δ - θ rhythms, is related to worsening of impairment over time [50].

Also in relation to this task, single-band datasets performed better than the All-band dataset, showing that specific EEG rhythms can be strong prognostic biomarkers for cognitive impairment in the context of AD. Furthermore, we conducted experiments using high-density EEG Alpha and Theta datasets and showed that, as in the previous case, increasing the number of channels does not significantly improve the capabilities of our model in discriminating among the three classes of subjects, since marginally higher results were obtained on Alpha but not on Theta.

In terms of classification performances, we compared the Transformer with three DL models based on CNNs for both binary and multiclass classification tasks. The results reported in Tables 3.7 and 3.8 show that all the models achieve overall good performances. In particular, focusing on the patients' binary classification, all the classifiers, except DeepConvNet on the Theta band, reach good accuracy levels ($> 70\%$), significantly higher than the no-information rate ($p \leq 0.001$ for Transformer and EEGNet; $p \leq 0.01$ for EEG-TCNet). The classification accuracy of patients in the multiclass approach, instead, was not significantly higher than the no-information rate in any case. Nevertheless, accuracy higher than 50% was achieved only by the Transformer and EEG-TCNet on the Theta band; in these cases, the performance on

epochs' classification was significantly higher than no-information rate ($p \leq 0.001$), meaning that both models uncovered a pattern underlying EEG data which allows the discrimination of HC, SCD and MCI subjects.

We also performed statistical analysis in order to assess the significance of our results on the cumulative test set. One-Way ANOVA was carried out for each group of data (i.e., SCD *vs* MCI on Delta and Theta bands, and HC *vs* SCD *vs* MCI on Alpha and Theta bands), considering the model as factor. The analysis did not reach the statistical significance ($p < 0.05$) in all the cases, except for the SCD *vs* MCI classification on Delta band ($p = 0.023$).

This result should be interpreted considering that we conducted the study implementing a LOSOCV approach which, in any case, allows an estimation of the generalization capabilities of the implemented models on the data of unseen subjects [149].

Despite the performances of the Transformer for the specific classification tasks are not outperforming when compared to the results obtained by CNN-based models, the use of this model still brings advantages that are worth considering. In fact, as already reported in the previous section, the Transformer model is less complex, with 5.2 k of trainable parameters, when compared to DeepConvNet, EEGNet and EEG-TCNet, which have 298.6 k, 9.8 k and 14.1 k parameters, respectively. As evidenced by a recent survey by Hu *et al.*, reducing the complexity of DL models while guaranteeing, at the same time, a sufficient level of expressive capacity by the model itself for a given task, is an open research problem [150]. In this perspective, the Transformer model already demonstrated classification capabilities comparable with more complex models.

As a final remark, in the next Chapter we'll demonstrate that the attention mechanism implemented by the Transformer, which is perfectly suited for the classification of temporal signals, allows the exploration of its interpretability capabilities by analysing temporal dependencies in the EEG signals exploiting the attention weights [151, 152].

Chapter 4

Interpretability methods for EEG-based Transformers

In the previous Chapter we demonstrated that Transformers and the self-attention mechanism can be successfully applied for the classification of EEG signals in a complex task as the discrimination of SCD from MCI patients. However, the clinical translatability of the framework remains somehow limited, due to the intrinsic complexity and black-box behaviour of the model.

In this section of the Ph.D. thesis, an extension of the previously described workflow is provided, introducing a method for visualizing and interpreting the outcome of the model as well as giving insights about its decision-making processes. Furthermore, the research work lays the foundation to the possibility of using this information to guide the identification of biomarkers of cognitive impairment in resting-state EEGs.

4.1 Motivations

Explainability and visualization methods of deep models, such as GradCAM or LIME, have already been employed in tasks for the classification of biological signals [153, 154]; however, a trustworthy understanding of DL algorithms supporting decisions in healthcare is essential and still needed [155, 156]. As seen in previous Chapters, concept of *interpretability* could represent a valid approach to deal with this problem. Transformers [95] and Vision Transformers [157] have introduced a new approach to the interpretability of deep networks in the fields of Natural Language Processing and Computer Vision through the mechanism of self-attention (see Section 2.5).

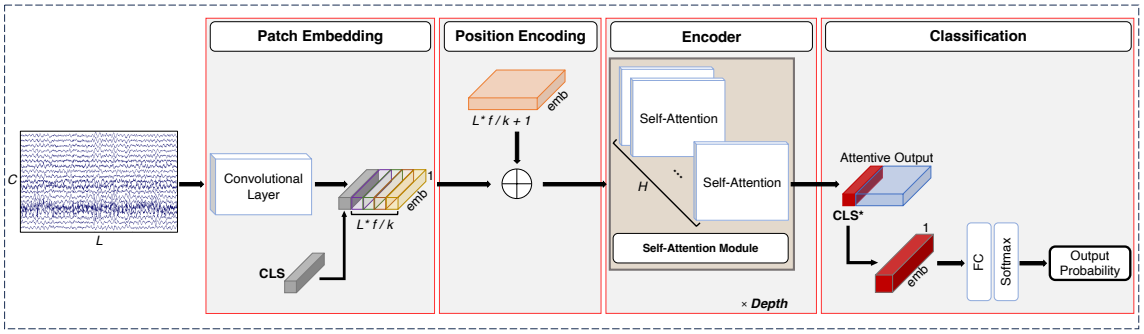


Fig. 4.1 Representation of the modules composing the proposed Transformer. C is the number of EEG channels, L is the length of the input epoch (in s), f is the EEG sampling rate (in Hz), k is the kernel size, emb is the embedding dimension and H is the number of attention heads. The classification token is denoted as CLS; the classification token updated after the Attention module is denoted as CLS*.

Various interpretability methods have been proposed for models based on Transformers [158]. Nonetheless, one effective approach is to leverage raw attention scores to visualize the portions of the input on which the model focused the most during the decision process [159, 160], particularly when working with time series [161].

On these premises, the research paper entitled *Understanding the role of self-attention in a Transformer model for the discrimination of SCD from MCI using resting-state EEG* [162] and reported in this section of the thesis aims to develop an interpretable framework for the model presented in the previous Chapter 3 in order to provide explanations for its decisions and support the identification of alterations in the brain activity of SCD and MCI patients by detecting patterns of interest in the input signals. In addition, this work aims to provide a further analysis of said method by tuning parameters and performing ablation studies on different modules of the Transformer in order to highlight the role of the self-attention component in the classification process.

4.2 Materials and methods

The dataset and the preprocessing steps used in this work are extensively described in Section 3.3. For this specific study, the binary classification task SCD *vs* MCI is considered. Specifically, epochs of rsEEG signals of 56 SCD and 45 MCI subjects were used as input to the Transformer model, whose modules are outlined in Figure 4.1 for clarity purposes.

As previously described, the model is composed of three main modules, namely patch embedding, positional encoding and the self-attention module, which is included in an encoder block. Lastly, it comprises a classification module constituted by a fully-connected layer with the *softmax* activation function.

To investigate how the traditional self-attention and Multi-Head Attention (MHA) strategies could affect the classification performances, the number of heads (H) per encoder was varied.

To avoid overevaluation of model performance, a test set was generated using 20 % of total subjects with a stratified random sampling approach. A stratified 5-fold cross-validation was employed on the remaining subjects, i.e. 43 SCD and 37 MCI, to train and validate the classification model. Using this technique, the data is divided subject-wisely into five equally-sized subsets, and the model is iteratively trained on four of these subsets and validated on the remaining one. Each subset is used as validation set exactly once.

Models were trained using Adam optimizer ($lr = 10e-4$, $\beta 1 = 0.9$, $\beta 2 = 0.999$, $eps = 1e-08$), which is the most employed method when training Transformer-based architectures [95] since it has faster convergence than non-adaptive algorithms such as SGD [163]. The value of lr was chosen by reducing it by a factor of 10 until finding an optimum in the validation set accuracy, starting from $10e-2$. Cross-Entropy was used as loss function. Batch size was set to 16 and the number of training iterations was equal to 250. In the proposed model, *emb* is set to 32 and $Depth$ is set to 2, resulting in 56194 trainable parameters.

4.3 Interpretability workflow via Self-attention

To understand the behavior of the model for the investigated classification task, it is important to know which parts of the input the model pays more attention to. To this end, we extracted weights from each attention layer of the trained models in order to identify the signal patch that contributed the most to the classification of each EEG epoch. As shown above, the classification of an EEG epoch is made upon the updated representation of the CLS token, i.e. CLS* (see Fig. 4.1). Thus, for each attention matrix, we considered the first row of values that correspond to the scaled dot-product attention of the CLS* token on the representations corresponding to the non-overlapping patches of the raw signal. This gives attention weights for each patch of the input epoch, helping evaluate their impact on the prediction.

It is worth noting that our Transformer, in its configuration with $H > 1$, uses a multi-depth and multi-head attention mechanism, which can produce different attention patterns that

can be challenging to visualize [105]. We averaged the attention scores across attention heads in order to retain all information produced by the attention module. On the other hand, we extracted different results for the first and the second encoder blocks to evaluate the contribution of each attention layer separately.

For all subjects in the dataset, we identified n patches, corresponding to n epochs of the raw signals with the highest attention weights. This means that for each epoch of length L, a patch of signal with dimension k datapoints was obtained, where k is the dimension of the kernel in the convolutional layer employed for patch embedding. In order to uphold the assumption that the highest attention weights are representative of significant changes in the EEG activity between SCD and MCI groups, we collected and concatenated 1-second long windows of the signal centered on the previously identified patches, obtaining a new set of signals for each class. Epochs belonging to the same class were then merged in a single time series. To validate the significance of the results through a comparison with a reference, we also segmented the complete signals with windows of 1 second and, once again, concatenated epochs of the same class to obtain one SCD and one MCI time series. Statistical analysis was performed on EEG data using Matlab's Letswave 7 tool. We applied the multi-sensor non-parametric cluster-based permutation Student's t-test for unpaired data [164] to compare the signals' epochs of the two groups, both for the attention-based set and the reference, which allowed us to handle the multiple comparisons problems. The calculation of the cluster-based statistics consists in grouping together neighboring t-values obtained for (space, time)-samples into clusters and summing the statistical values within each cluster. For inclusion in a cluster, only statistical values higher than the cluster-forming threshold, which was set to 0.05, are considered. Then, the significance probability is calculated with a Monte Carlo approximation based on the number of permutations. As a rule of thumb proposed in previous works, this number should be no less than 1000. Thus, to perform feasible computations, we set it to 2000.

Finally, to gain a physiological interpretation of the results, we performed time-frequency analysis by applying Continuous Wavelet Transform (CWT) to the EEG epochs and averaging the results across each group. Complex Morlet wavelet with bandwidth of 1 Hz and central frequency of 1.5 Hz was used as mother wavelet.

4.4 Results and Discussion

In this section, we extensively illustrate the results of interpretability analysis for visualizing the focus of the model on specific EEG patterns. Then, we provide results of parameter

tuning tests for choosing the best model configuration and demonstrate the efficacy of the attention module through ablation studies.

4.4.1 Interpretability analysis

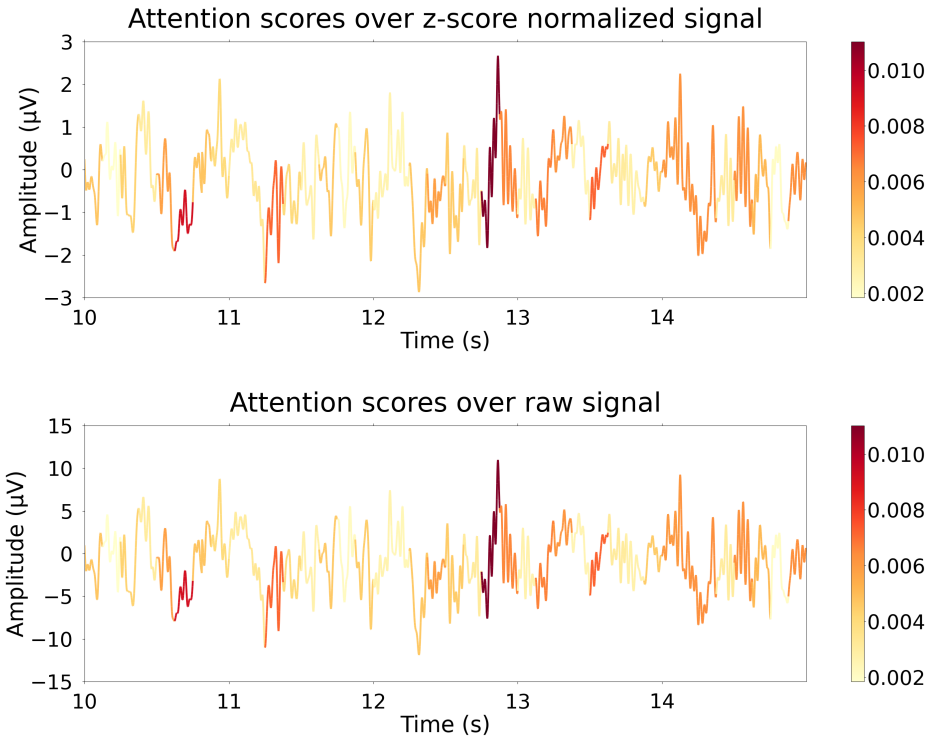
Following the approach proposed in section 4.3, the interpretability analysis was performed on the model which obtained the highest values of mean accuracy and AUC on the test set, i.e. the Transformer configuration with $L = 30$ s, $k = 64$ and $H = 8$. This configuration achieves mean accuracy of 65.4 % (95 % CI [0.637 - 0.671], p-value [Accuracy $>$ No Information Rate] = 0.00026) on epochs' classification, and 65.7 % of accuracy for subjects' classification through hard voting.

As a first attempt to visualize the attention focus, we present heat maps of attention scores on the raw EEG signals for both SCD and MCI classes. Figure 4.2 shows two examples of 5-s-long windows extracted from the corresponding 30-s epochs of one correctly classified SCD (Fig. 4.2a) and one correctly classified MCI (Fig. 4.2b) subject of the test set. For clarity purposes, the normalized and non-normalized signals of one channel, namely T3, have been plotted for both samples. Attention scores are plotted over patches of k datapoints, with dark red indicating areas with higher focus, and light yellow indicating areas with lower focus.

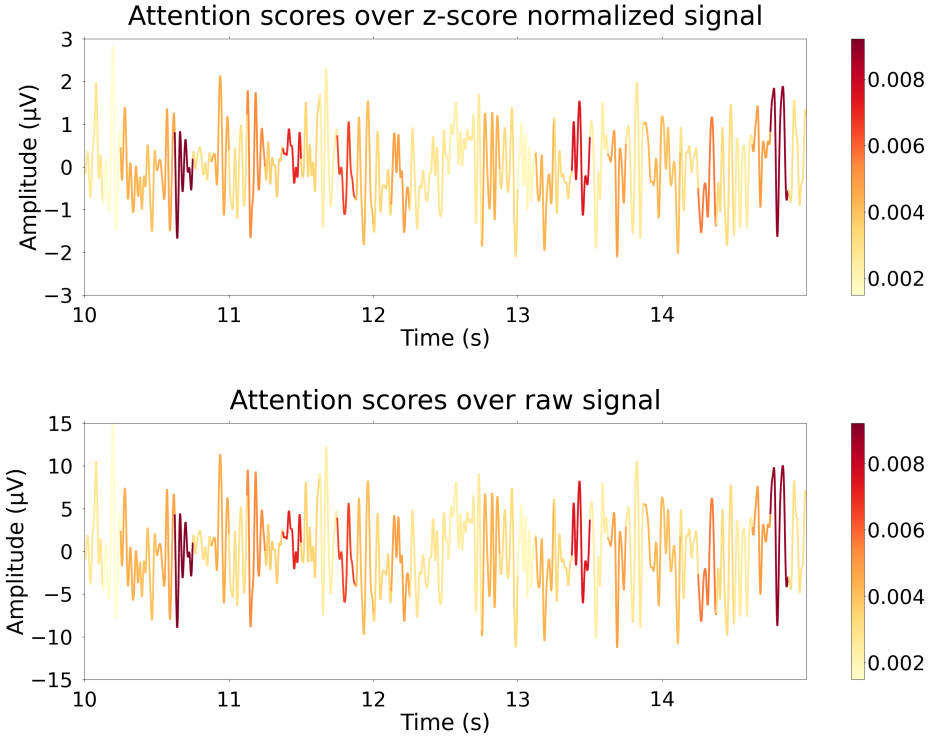
To quantitatively evaluate the contribution of the attention scores on the final classification outcome, we show results of the nonparametric cluster-based permutation Student's t-test and the corresponding time-frequency analysis with the aim of highlighting differences between the two groups. We considered channels with clustered p-value < 0.01 to be significant.

When comparing epochs of 1 s centered on patches with the highest attention, the most significant differences between the SCD and MCI signals are, indeed, located in the time interval that corresponds to those patches, i.e. from 437 ms to 562 ms since the epoch start. For instance, when considering the results of the first Transformer attention block (Fig. 4.3a) it can be noted that most statistically significant inter-group differences can be found in the central part of the time window, as shown by the corresponding scalp topographies representing clustered p-values. The most significant changes occur on clusters including the following channels: Fp1, Fp2, F3, F7, Fz, F4, F8, C3, Cz, C4, P3, P4, Pz, T5, T6, O1 and O2.

This evidence is strengthened by the results obtained on the second attention block (Fig. 4.3b). In this case, almost all statistically significant differences correspond to the highest attention scores which are located in the middle of the considered time window.



(a) SCD



(b) MCI

Fig. 4.2 Sample plots of two 5-s long EEG epochs with relative attention scores for one SCD (a) and one MCI (b) subject of the test set. Both normalized and non-normalized signals are shown.

Scalp topographies of clustered p-values show that the significant clusters include the Fp1, Fp2, F7, F3, F4, F8, C4, Cz, T3, P3, Pz and P4 channels.

The clusters found in both cases indicate brain regions that are congruent with scientific evidence from cross-sectional and longitudinal studies on the cognitive spectrum of AD. As reported in [53], the left posterior parietal and left and right temporo-occipital regions (which are represented by P3, P4, T6 and O2 electrodes) were consistently described as the most discriminative brain areas between controls, MCI and AD, while the left posterior temporal region and fronto-central midline (corresponding to T5, Fz, Cz and Pz channels) as important in the prediction of clinical progression in patients with SCD.

On the other hand, statistical analysis performed on the reference dataset, i.e. considering all epochs of 1 s extracted from the input signal, regardless of weights attributed by attention, found no significant channels at any time instant ($p > 0.01$). This result confirms that, although mean classification accuracy on the test set is not optimal, the Transformer is able to capture global temporal dependencies of the signal that allow the classification of each epoch with good discrimination capability.

However, differently from other studies that applied an interpretability approach based on attention scores to EEG signals in the context of sleep stage classification [151, 165] or motor imagery paradigms [166, 167], these features are not easily detectable and do not provide enough explanations in the time domain.

Hence, on the basis of the findings derived from the statistical analysis, we report scalp topographies of the average power CWT for SCD and MCI subject groups based on the results of the first Transformer block ($Depth = 1$). In particular, Fig. 4.5 shows CWT maps averaged across the whole 1-s interval (first and third row) and the interval of interest (second and fourth row) for delta (Fig.4.5a) and alpha (Fig.4.5b) frequency bands, respectively. Of notice, differences between the groups are once again more evident when considering the time interval corresponding to the highest attention scores, rather than the entire time window. The maps confirm that subjects belonging to the MCI group show a lower power in high frequencies and higher power in low frequencies in accordance with state-of-the-art results in the context of AD characterization from rsEEG [13, 53, 168]. In addition, these explanations keep with expectations of our previous work [133].

Additionally, we compared these maps with the ones obtained on the reference dataset, and found that in the latter the differences between the groups do not correspond to specific time intervals, in accordance with the results of the aforementioned statistical analysis.

To further understand the role of the multi-head attention mechanism, we repeated the analyses for the baseline model with single-head self-attention, which achieves a mean

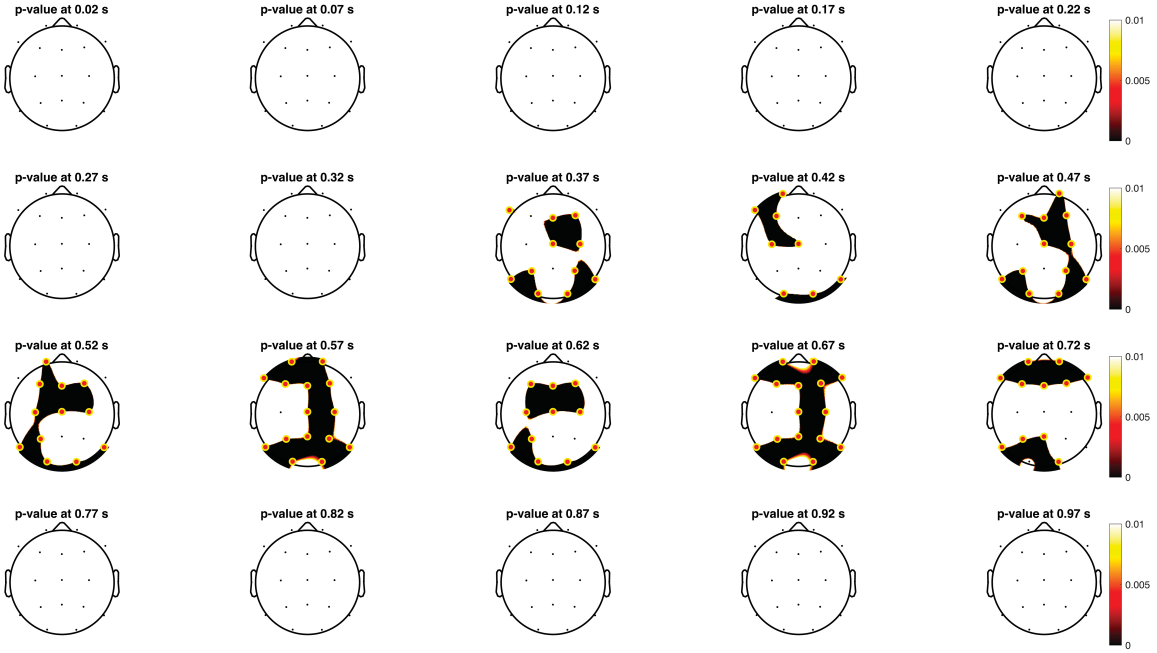
accuracy of 59.5 % (95 % CI [0.577 - 0.612], p-value [Accuracy $>$ No Information Rate] > 0.05) on epochs' classification and 61.9 % on patients' classification. As expected, and as found by [166], the attention activation of a single head is similar to the one obtained by averaging multiple heads, with significantly different patches (p-value < 0.01) between SCD and MCI groups corresponding to the highest attention scores, but resulting in more sparse and less consistent channel clusters, particularly when considering the results obtained on the first Transformer block. For $Depth = 1$, significant clusters include Fp2, F4, F7, F8, Cz, C3, C4, P3, Pz, T3, T4, T5, T6, and O2 (Fig. 4.4a). For $Depth = 2$, significant channels are Fp1, Fp2, F3, F7, F8, Fz, Cz, C3, C4, Pz, P4, T3, T4, T5, T6, O1 and O2 (Fig. 4.4b). This is explained considering that the baseline model, for the same model depth, has lower performances which do not reach the statistical significance in terms of accuracy; such a result is in line with [169], who report that single-head attention necessitates deeper models to prove more effective than MHA, but increasing the model complexity. Thus, the attention focus is less indicative of discriminative EEG features. Consequently, the spectral analysis obtained with CWT shows similar outcomes, with changes in activation between groups mostly gathered in the central part of the window, but being less enhanced, especially in the lower frequencies, for both *Depths*.

4.4.2 Hyperparamter tuning

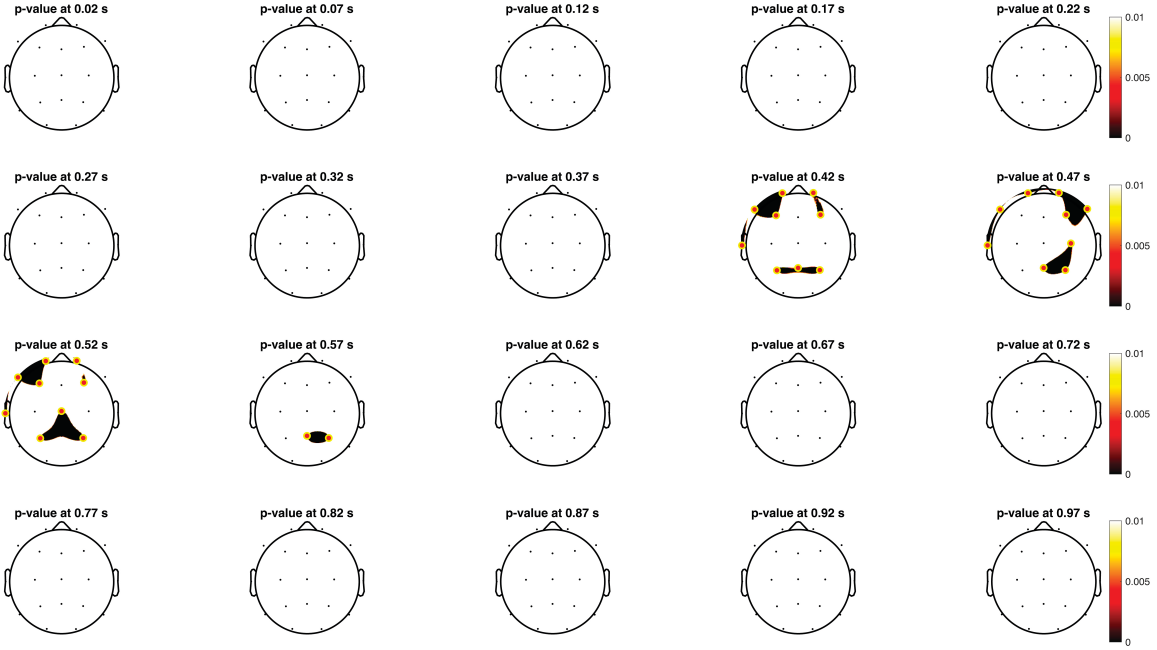
We conducted experiments to identify the best model's parameters to achieve optimal classification performances. We varied two parameters that influence the construction of the input, namely the time duration of input EEG epochs and the design of the convolutional kernel, and also investigated the influence of the number H of attention heads in the attention layer, known to impact feature learning.

In particular, three different lengths of input epochs (10, 30 and 60 seconds) and five different kernel sizes (16, 32, 64, 128 and 512) were tested and compared to identify the combination with the highest classification performance. Table 4.1 reports mean results for all the considered metrics on epochs' and patients' classification. The highest levels of mean accuracy are reached with a kernel size of 64, with values of 65.4 % and 63.0 % for epochs of 30 s and 60 s respectively, and a kernel size of 32 on epochs of 10 s with a value of 63.4 %.

By contrast, the lowest results are yielded when using kernel sizes of 512 (52.4 % on epochs of 60 s) and 16 (54.3 % on epochs of 30 s). Although the differences are not significant ($p > 0.05$), in accordance with Song *et al.* [166], we found that large kernel sizes tend to flatten temporal features and reduce the learning of global dependencies, while small

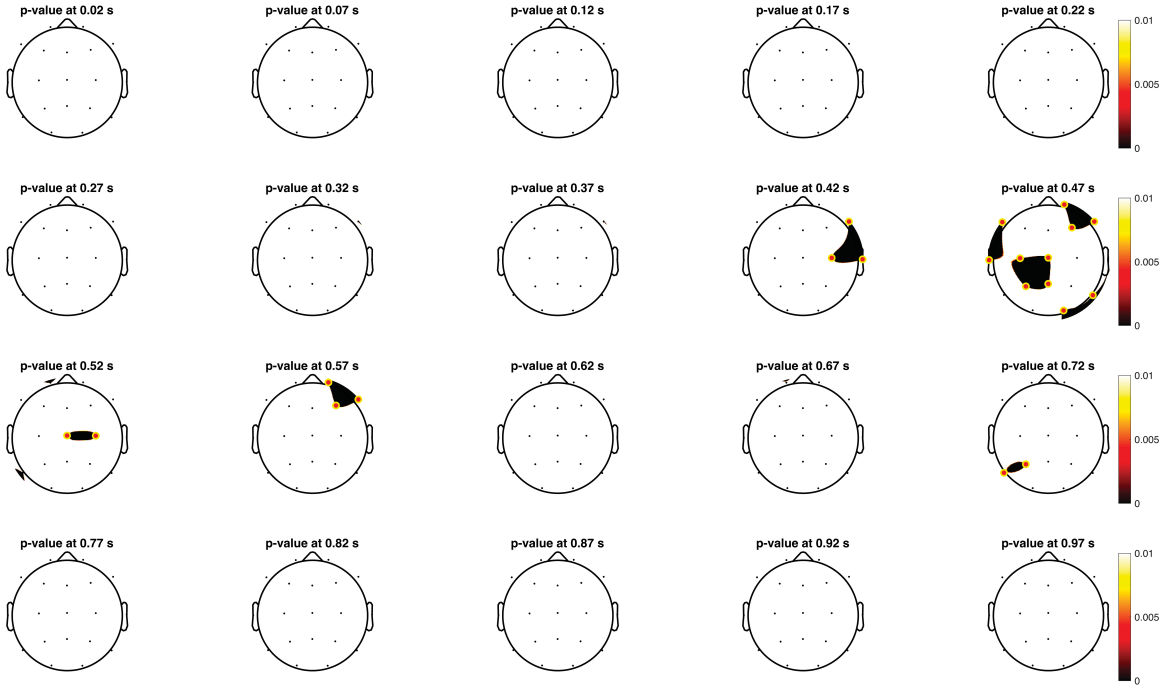


(a)

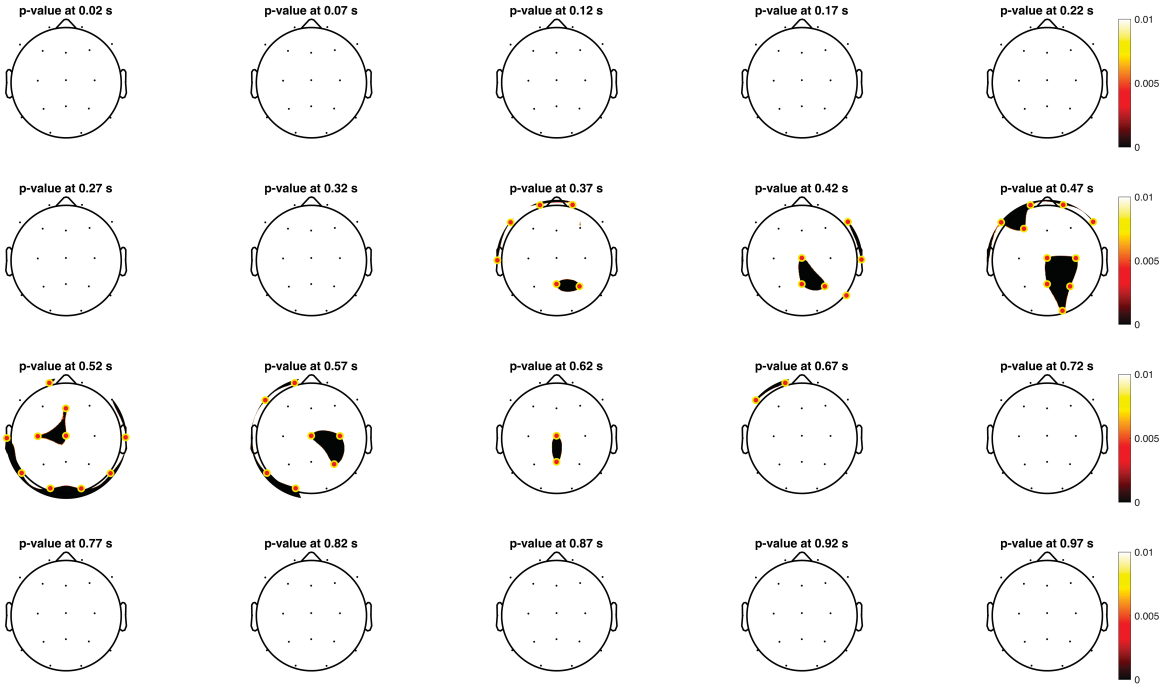


(b)

Fig. 4.3 The results of the cluster permutation Student's t-test for multi-head attention model, $Depth = 1$ (a) and $Depth = 2$ (b). Clustered p-values over time are plotted on scalp maps at 50 ms intervals. Significant channels are yellow-circled and highlighted.

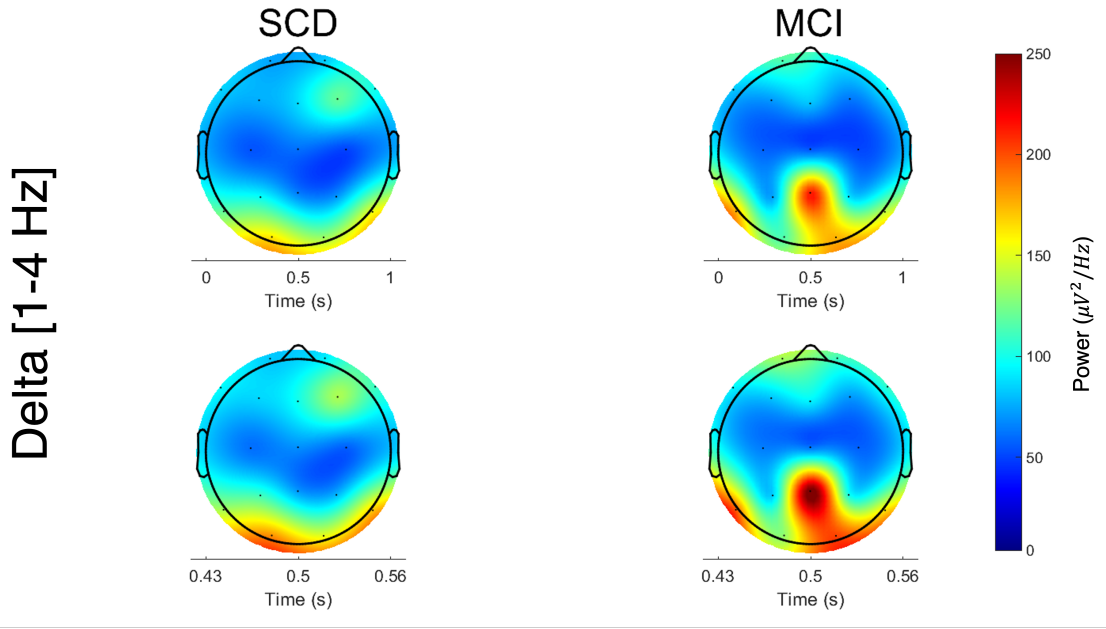


(a)

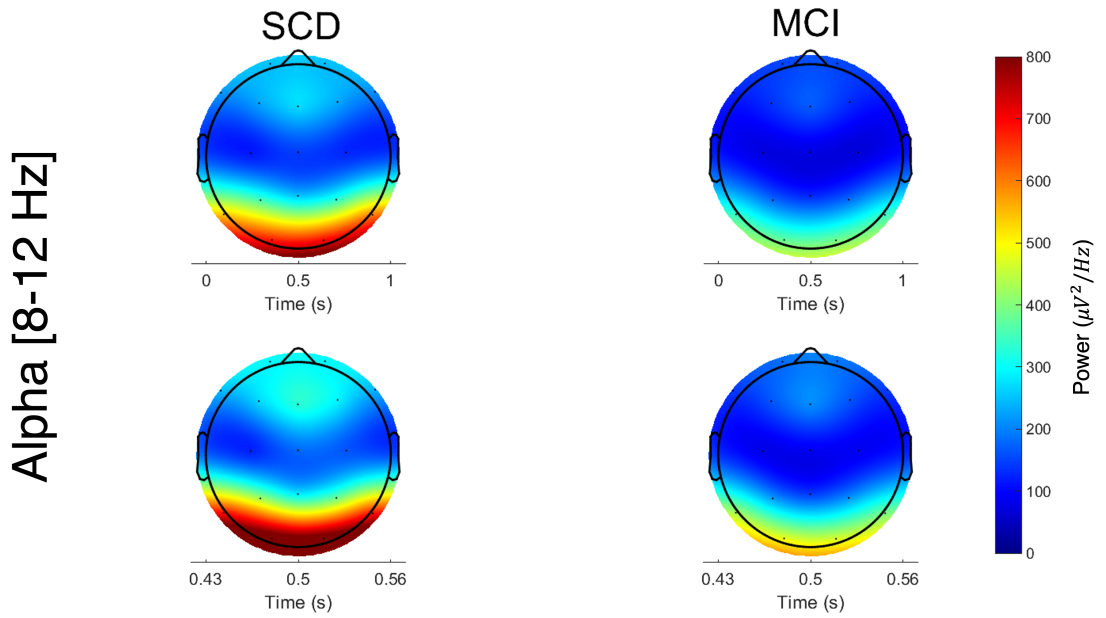


(b)

Fig. 4.4 The results of the cluster permutation Student's t-test for single-head self-attention model, $Depth = 1$ (a) and $Depth = 2$ (b). Clustered p-values over time are plotted on scalp maps at 50 ms intervals. Significant channels are yellow-circled and highlighted.



(a)



(b)

Fig. 4.5 Scalp topographies of Average Continuous Wavelet Transform of EEG signals segmented based on attention scores of the first Transformer block for SCD and MCI groups. (a) Average CWT in delta band (1-4 Hz) across the whole second interval (first row) and the interval of interest (second row). (b) Average CWT in alpha band (8-12 Hz) across the whole second interval (first row) and the interval of interest (second row).

Table 4.1 Classification results on the epochs' test set for different input configurations, expressed as mean ± standard deviation.

Epoch length	Kernel	Accuracy	Sensitivity	Specificity	Precision	F1	AUC
10	16	0.61 ± 0.01	0.59 ± 0.12	0.62 ± 0.09	0.48 ± 0.01	0.52 ± 0.04	0.60 ± 0.02
	32	0.63 ± 0.03	0.63 ± 0.13	0.64 ± 0.03	0.51 ± 0.03	0.56 ± 0.07	0.63 ± 0.05
	64	0.62 ± 0.06	0.54 ± 0.06	0.66 ± 0.07	0.50 ± 0.08	0.52 ± 0.06	0.60 ± 0.06
	128	0.47 ± 0.08	0.54 ± 0.15	0.42 ± 0.19	0.36 ± 0.05	0.43 ± 0.07	0.52 ± 0.06
	512	0.57 ± 0.03	0.53 ± 0.04	0.59 ± 0.02	0.44 ± 0.03	0.48 ± 0.03	0.59 ± 0.04
30	16	0.56 ± 0.06	0.57 ± 0.08	0.55 ± 0.05	0.44 ± 0.05	0.49 ± 0.06	0.56 ± 0.06
	32	0.62 ± 0.05	0.61 ± 0.11	0.62 ± 0.09	0.50 ± 0.06	0.54 ± 0.06	0.61 ± 0.05
	64	$\mathbf{0.65 \pm 0.05}$	0.58 ± 0.11	0.70 ± 0.06	0.54 ± 0.06	0.56 ± 0.07	$\mathbf{0.64 \pm 0.06}$
	128	0.62 ± 0.12	0.46 ± 0.15	$\mathbf{0.72 \pm 0.26}$	$\mathbf{0.57 \pm 0.14}$	0.48 ± 0.08	0.62 ± 0.12
	512	0.55 ± 0.03	0.57 ± 0.11	0.54 ± 0.09	0.43 ± 0.03	0.48 ± 0.05	0.58 ± 0.05
60	16	0.60 ± 0.08	0.64 ± 0.16	0.58 ± 0.12	0.48 ± 0.09	0.55 ± 0.11	0.61 ± 0.09
	32	0.59 ± 0.12	0.56 ± 0.14	0.62 ± 0.19	0.49 ± 0.12	0.51 ± 0.10	0.59 ± 0.11
	64	0.63 ± 0.09	$\mathbf{0.64 \pm 0.12}$	0.62 ± 0.09	0.51 ± 0.10	$\mathbf{0.57 \pm 0.11}$	0.63 ± 0.10
	128	0.59 ± 0.08	0.57 ± 0.20	0.60 ± 0.13	0.46 ± 0.07	0.50 ± 0.13	0.60 ± 0.11
	512	0.53 ± 0.05	0.49 ± 0.13	0.55 ± 0.10	0.40 ± 0.05	0.43 ± 0.07	0.52 ± 0.06

kernels produce tokens that do not contain enough information for the model to perceive local changes in the signal. On the other hand, the length of the input EEG signal seems to impact the performances of our model to a lesser extent. However, as a general remark, using very long epoch lengths (i.e. 60 s) results in a smaller dataset size which increases the risk of lowering the performance of the classification model.

We also compared the impact of choosing different numbers of heads for the attention layer, performing experiments by varying H in [1, 2, 4, 8, 16, 32]. Since each head projects the input onto a subspace of dimension $dim = \frac{emb}{H}$ to compute the context [95], the values of H were chosen based on the embedding dimension.

The results reported in Fig. 4.6 show that the effects on the performance of the model follow no evident trend ($p > 0.05$), but the highest accuracy of 65.4 % is obtained with $H = 8$, compared to 59.5 % with $H = 1$, 62.8 % with $H = 2$, 57.8 % with $H = 4$, 56.9 % with $H = 16$ and 59.3 % with $H = 32$. Also, as shown by the error bars in the same figure, setting the number of heads to 8 allowed to obtain the smallest 95 % confidence interval. Conversely, the highest confidence intervals were derived from configurations with 1 and 32 heads. This suggests that while employing a greater number of heads enables the model to identify more meaningful features, a progressive increase in the number of heads results in shorter feature lengths within each head. This, in turn, contributes to a marginal decrease in performance. This result confirms previous evidence from another study [170].

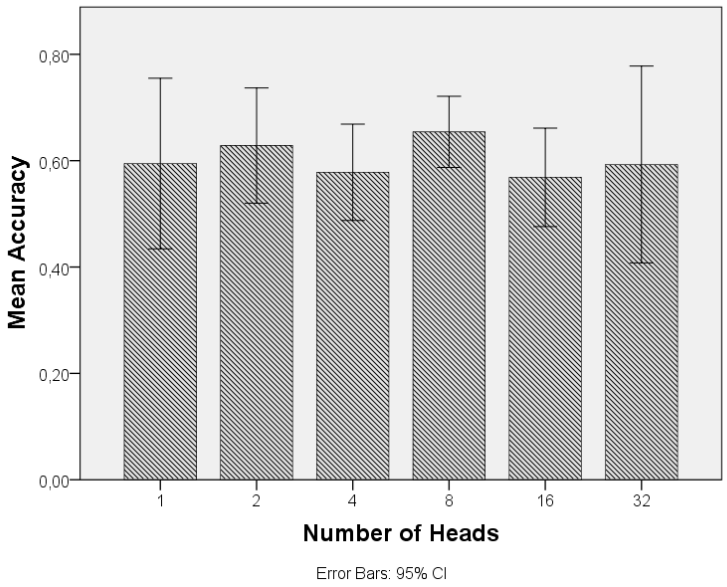


Fig. 4.6 The impact of different numbers of attention heads on the mean accuracy over folds for epoch-wise classification on the test set.

4.4.3 Ablation Study

In this section, we systematically analyze the importance of two key components of our model, namely the attention-based Transformer encoder module and the positional encoding module. An ablation study was conducted by firstly removing the Transformer encoder, i.e. the classification was performed on the input signal after convolution without applying any attention strategy. Then, we reintroduced the Transformer encoder module and dismissed the positional encoding, so that the model had no information about the position of each patch in the input sequence when performing classification. Lastly, we removed both the Transformer and the positional encoding blocks. In the study, we included results for both MHA and single-head self-attention models.

As depicted in Figure 4.7, and as already shown in Fig. 4.6, for the same input configuration, the model employing multiple heads has overall better performances than the model employing the traditional self-attention layers, which does not reach statistical significance in classification accuracy on the test set and shows high variability over the folds.

Nevertheless, the effectiveness of using an attention mechanism is confirmed by the results obtained when the Transformer block is removed, in which the mean accuracy on the test set drops significantly in the epochs' classification, decreasing by 16 % ($p = 0.004$) for the MHA configuration and by 10.5 % ($p > 0.05$) for the single-head self-attention

configuration including it. Also, in patients' classification it reduces significantly by 19 % ($p = 0.009$) in the first case and by 15.2 % ($p > 0.05$) in the latter.

The removal of the positional encoding has a different impact on the two models. For the model employing MHA, the mean accuracy over the folds decreases by 1.5% ($p > 0.05$). Although the difference is not significant, these results suggest that this model makes use of positional encoding in an informative way, but is still able to compensate for it with the attention module. Additionally, this consideration is supported by the results of the last ablation test, in which both the Transformer and the positional encoding modules are removed. In this case, the mean performances of the model are slightly higher than the case in which only the Transformer is removed, by 1.5 % epoch-wise ($p = 0.02$), proving that the positional encoding module is useful when combined with multi-head attention, but has a negative impact on the results when added to a convolutional-based model. In fact, positional information could be inherently learned by a convolutional layer with a sufficient receptive field size [171] and thus the information provided by the positional encoding in this case could produce redundancy. On the other hand, the ablation of the positional encoding module in the single-head self-attention model also seems to impact positively on the classification performances, by increasing accuracy of 2.2 %, but not significantly ($p > 0.05$), which further proves that the attention module is capable of learning positional information by itself [172]. However, this result needs further understanding [173].

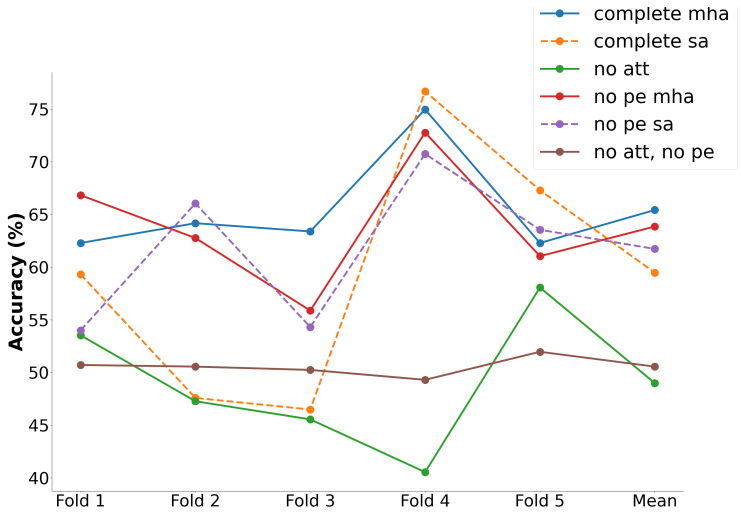


Fig. 4.7 The results of ablation study for epoch-wise classification on the test set. Accuracy values are plotted for single folds and as mean values over folds. In the legend, *att* is the attention module, *pe* is the positional encoding, *mha* is the multi-head attention and *sa* is the traditional self-attention with one head.

4.4.4 General remarks

The complexity of EEG signals poses a challenge in the identification of biomarkers that can accurately discriminate between SCD and MCI conditions. This study demonstrated that MHA can be used in an end-to-end Transformer model to automatically locate time windows of the resting-state EEG that may account for significant changes in the brain activity. The interpretability analysis showed a higher global efficacy of MHA compared to traditional self-attention approaches. Indeed, although it was previously found that the MHA-based Transformer did not outperform other investigated DL methods for the specific task, it allowed to highlight significant differences between the groups which could not be explained otherwise. In addition, the ablation study confirmed the effectiveness of introducing Transformer blocks in the proposed model, in particular when coupled with the encoding of the positions of patches in the input.

This finding suggests that this framework could serve not only to enhance the interpretability of a black-box model which achieves state-of-the-art classification performances, thus addressing the problem of the trade-off between accuracy and trustworthiness [174], but also as a guide for experts to facilitate the extraction of rsEEG markers of cognitive decay. A recent work employed the attention mechanism to design an EEG channel interpolation algorithm [175]. Similarly, this method could be exploited also in different applications to select relevant domain-specific information by taking into account short and long temporal dependencies of the signal.

Chapter 5

Computational methods for the analysis of evoked responses

So far, this thesis has dealt with the development of intelligent explainable systems for decoding and classifying spontaneous EEG signals recorded in resting-state conditions. In this last Chapter, the focus will shift on computational methods for analysing EEG responses related to internal or external events.

In particular, the first part of the Chapter describes the use of an experimental paradigm to verify the preservation of the mechanism of motor resonance in early PD patients using behavioral, hemodynamic and electrophysiological data and how this can affect rehabilitation strategies. In this work, changes in event-related desynchronizations (ERDs) of alpha rhythm are analysed as a biomarker for assessing the presence of sensorimotor network involvement during specific tasks. This study was conducted in collaboration with the Neurophysiopathology Unit at Polyclinic General Hospital of Bari, in Bari, Italy.

The second topic of the Chapter focuses on event-related potentials and proposes a different level of analysis by employing biophysical models based on Dynamic Causal Modeling (DCM) and statistical inference, coupled with ML methods. The research activities described in this section were carried out during the months spent as visiting Ph.D. student in the Department of Data Analysis at the Faculty of Psychology and Educational Sciences of Ghent University, in Ghent, Belgium.

Lastly, the Chapter reports a novel research in the field of Human-Robot Interaction (HRI) investigating how cross-modal stimulation linked to gender aspects (i.e., human pheromones and voice gender) and proxemic space variations influence behavioral and electrophysiological responses. These research activities were conducted in collaboration with the Laboratory

of Cognitive and Psychophysiological Olfactory Processes of the University of Salento, in Lecce, Italy.

5.1 Motor Resonance in Parkinson's disease

5.1.1 Motivations

The role of the motor cortex has long been known in cognitive processes such as, for example, motor planning, motor imagination, perception of action, and motor learning. The observation of action seems to involve the generation of the internal representation of that same action in the observer, a process named Motor Resonance (MR) [176, 177] and mediated by the Mirror Neuron System (MNS). Importantly, action observation determines the activation of different networks located in the visual, motor, and perceptive areas [178].

Understanding the neurophysiological mechanisms of action observation effects on the brain of neurological patients has been a hot topic in the last few years. Specifically, the progressive aging of the population poses new challenges to rehabilitation medicine, in particular for those neurodegenerative and disabling diseases such as Parkinson's disease [179]. In PD, motor resonance is often impaired due to the degeneration of dopaminergic neurons in the basal ganglia (see Section 2.1.2), which disrupts motor control and coordination. Studies suggest that individuals with PD may show reduced MR, as evidenced by altered brain activity in regions of the MNS typically associated with action observation and motor simulation, such as the premotor cortex and inferior parietal lobule [180, 181]. This impairment can contribute to difficulties in motor learning and social interactions, as patients may struggle to imitate or understand observed movements. Motor cognition appears to represent a promising field of study for the design of rehabilitation interventions for patients with PD, including those based on action observation [179].

The main objective of this study was to verify whether an experimental paradigm of action observation in a laboratory context could elicit cortical motor activation in PD patients. This specific paradigm, which involves the observation of grasping actions towards graspable or ungraspable objects through videos, had been previously employed for providing indirect evidence for the presence of MR, but never on PD patients [182–185]. In particular, the aim was to investigate how movement congruence could affect mirror mechanisms in PD and healthy controls, and whether a comparable pattern between the two groups could indicate a preservation of normal motor resonance.

5.1.2 Materials and Methods

21 PD patients and 22 sex- and age-matched controls were enrolled in the study. Inclusion criteria for PD were: diagnosis of idiopathic Parkinson's disease, Hoehn-Yahr stage < II, age between 40–80 years, MMSE > 24, absence of significant visual deficits. All patients were stable without motor/non-motor fluctuations and dyskinesias. The experimental protocol was designed as follows: in a first session, participants observed videos of grasping actions directed towards a graspable or an ungraspable object and were instructed to respond the instant the agent touched the object (Time-to-contact detection session). In a different experimental session, instead, participants were instructed to watch and pay attention to the videos (Observation-only session).

During each experimental session, the participants' cerebral hemodynamic activity was recorded using a functional Near-Infrared Spectroscopy (fNIRS) with 20 channels located on the motor and premotor brain areas. Furthermore, an EEG analysis, focused on event-related desynchronization of alpha rhythm (alpha-mu rhythm), was considered to verify the presence of a sensorimotor network involvement.

For the purposes of this Ph.D. thesis, only results relative to the EEG data analysis are reported in the next sections. The complete results were published in the work *"Effects of movement congruence on motor resonance in early Parkinson's disease"* [186].

5.1.2.1 Experimental procedure

The stimuli used in the study are described in Craighero *et al.* [176], consisting of two 2640 ms videos depicting an agent seated at a desk reaching and grasping an object, recorded from a third-person perspective. In the flat object video, the object was a parallelepiped (7 cm × 3 cm × 3 cm) oriented with its longer axis facing the agent, who naturally grasped it with fingers parallel to the frontal plane without lifting it. The sharp-tip object video replaced the parallelepiped with a polyhedron of the same dimensions using video editing software, ensuring that the kinematics of the movement remained unchanged, and the agent's fingers touched the sharp tips. Both videos featured the same moment of contact between the agent's index finger and the object (1880 ms, Frame 47). Additionally, catch-trial videos were created by stopping the videos before the agent's hand touched the objects (1520 ms, Frame 38) and extending the duration to match the experimental trials (2640 ms) by repeating the final frame. These catch trials were included to maintain participant attention but were excluded from the analysis.

The experiment was conducted using a multimodal fNIRS-EEG co-registration system. Participants, seated in front of a display and a keyboard, were first asked to grasp both objects shown in the videos using the same grip demonstrated by the agent. This task aimed to illustrate the difficulty of grasping the sharp-tip object due to its weight and sharp edges, compared to the flat object, despite both objects having the same weight. Two experimental sessions were conducted: a Time-to-contact detection session and an Observation-only session. Each session consisted of 42 randomized trials, including 30 experimental trials (15 flat object videos and 15 sharp-tip object videos) and 12 catch trials (6 of each type).

At the start of each session, participants fixated on a cross for 120 s, recording 20 s of baseline and 100 s of resting state. The type of session was announced before the resting-state period, and participants were warned again 5 s before the video began. The resting-state data were used to assess preparatory brain activity for action observation in both fNIRS and EEG modalities. Between videos, a 15 s black screen was shown, and participants were given a 5-minute break at the end of each session.

5.1.2.2 EEG recording and analysis

The EEG signal was acquired using the Micromed Brain Quick equipment at a sampling rate of 256 Hz using 61 electrodes positioned according to the 10-10 international system. To acquire also the electrooculogram (EOG), two electrodes were placed on the right and left eyebrows, respectively. The reference electrode was positioned on the nasion, and the ground electrode on Fpz. A 0.1–70 Hz band-pass filter with a 50 Hz digital filter was applied during the EEG recording. The EEG was recorded during the entire experimental procedure. The EEG data preprocessing was performed with EEGLAB. The researchers used a semi-automatic method based on visual detection and channel statistics to locate and remove the faulty recording channels. All channels with distributions far from the Gaussian one were excluded from the analyses. Ocular artifacts recorded by the EOG channels were removed by means of the ICA algorithm included in the EEGLAB tool. Next, all the EEG files were processed using Letswave 7 tool. EEG has been re-referenced to 0 value and pre-filtered with a band-pass filter in the range [1 -30] Hz. To evaluate the not phase-locked synchronization/desynchronization of alpha mu and beta mu, the researchers used a time-frequency analysis based on Continuous Wavelet Transform (CWT), with a baseline correction computed on the 20 seconds preceding the resting-state. The absolute power of the alpha (7-12 Hz) and beta (13-30 Hz) bands were considered for the single experimental conditions. In order to detect the preparation to action observation, epochs lasting 5 seconds that preceded both the start of the observation-only session and the time-to-contact detection

session were considered. To evaluate the alpha mu changes, with respect to the baseline, related to the vision of the flat and sharp tip objects, we computed the CWT in a time window from 2 seconds preceding the object grasping to 1 second following it, so the EEG was recorded simultaneously to the movement of the arm in the video.

For topographical analysis and generation of Statistical Probability Maps, we used Matlab Letswave 7 tool, applying the Student's t-test for paired data to compare the absolute power of alpha mu in single groups. The two-way ANOVA with conditions and groups as factors was also applied, to establish differences of alpha mu behaviour in resting-state preceding observation and time-to-contact detection sessions and during these sessions between flat vs sharp tip object grasping conditions. The cluster significant threshold was set to 0.05 and the number of permutations was set to 2000. For representation purposes, the Statistical Probability Maps show the significant results obtained after permutations in the range 0.001-0.01

5.1.3 Results

Alpha mu: Comparison between time preceding time-to-contact detection session and observation-only session. For the resting-state, we considered the 5 seconds preceding the time-to-contact detection session and the observation-only session.

In controls, we observed that in the lower frequencies range, the alpha rhythm was more desynchronized in the time preceding the observation-only session (Figure 5.1). However, this did not reach the statistical significance.

In PD patients, the alpha mu desynchronized in the time preceding the time-to-contact detection session (Figure 5.1). The t-test for paired data showed a significant desynchronization in the 13-18 Hz range over the central regions.

In the comparison between groups, alpha mu desynchronization was more evident in PD patients over the left fronto-central regions in the seconds preceding the time-to-contact detection session, as indicated by the results of the ANOVA test considering the session and groups as factors (Figure 5.1).

Time-to-contact detection session: comparison between flat object trials vs sharp-tip object trials. In controls, desynchronization of the mu rhythm in the alpha range was present in the second preceding the grasp of the objects. Two seconds before the flat object grasp, the alpha rhythm desynchronization prevailed in the 8-12 Hz range, on the left fronto-central electrodes, compared to sharp-tip object (Figure 5.2).

In PD patients, the desynchronization of alpha mu was also present in the second preceding and following both objects grasping. We also observed desynchronization in the range 11-13 Hz in the same time preceding the hand grasping the flat object.

The comparison between flat object trials vs sharp-tip object trials between groups was not significant (Figure 5.2).

Observation-only session: comparison between flat object trials vs sharp-tip object trials. In controls, the mu rhythm, especially in alpha range, showed a tendency to a desynchronization in the time preceding the grasp of both objects. In the time following the incongruent movement, the alpha rhythm appeared more desynchronized, though no significant change was detected with the t-test. (Figure 5.3).

Patients showed a similar mu rhythm desynchronization, especially in the alpha range, in approaching the objects grasping, though the two objects did not induce different mu rhythm behaviour.

The comparison across groups and conditions was not significant.

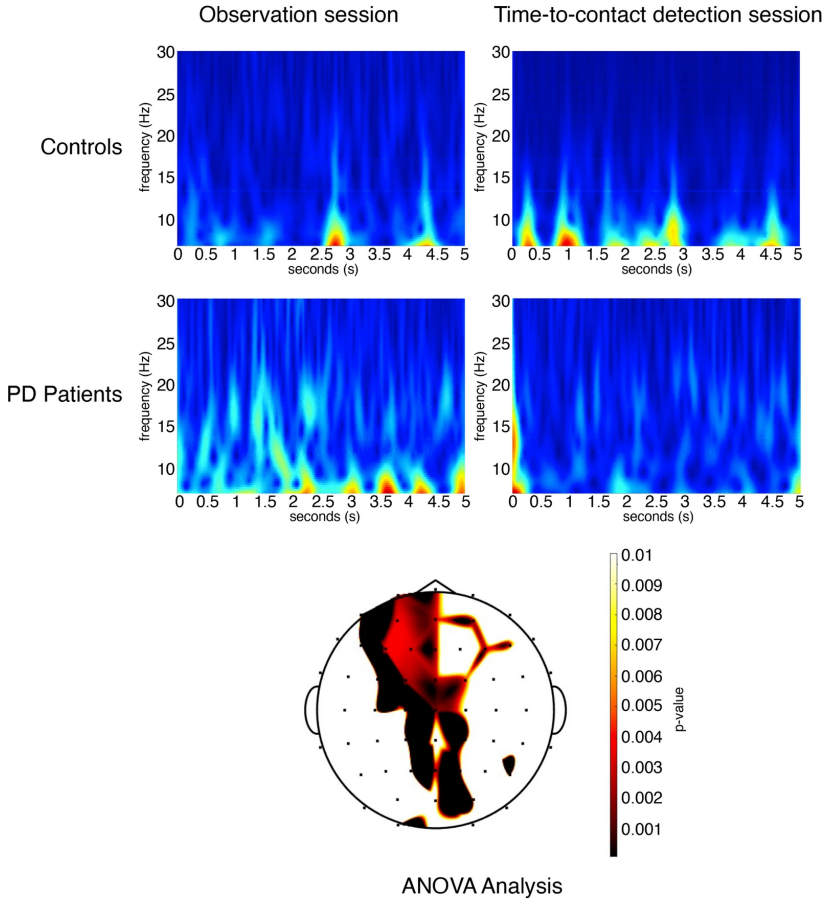


Fig. 5.1 Resting-state preceding the observation and Time-to-contact detection session. (Up) The Grand Average of Continuous Wavelet Transform of alpha-mu recorded on the C3 derivation in the 5 s of resting state preceding the observation session and Time-to-contact detection session is reported for controls and PD patients groups. In PD patients, desynchronization of EEG rhythm is evident in the 8–13 Hz range in the time preceding the Time-to-contact detection session, in controls desynchronization prevailed in the low alpha range before the observation-only session. (Bottom) The statistical map reports the p-values obtained with ANOVA analysis for the interaction group x session. It shows that alpha desynchronization was more evident in PD patients on the fronto-central electrodes for the effect of the session.

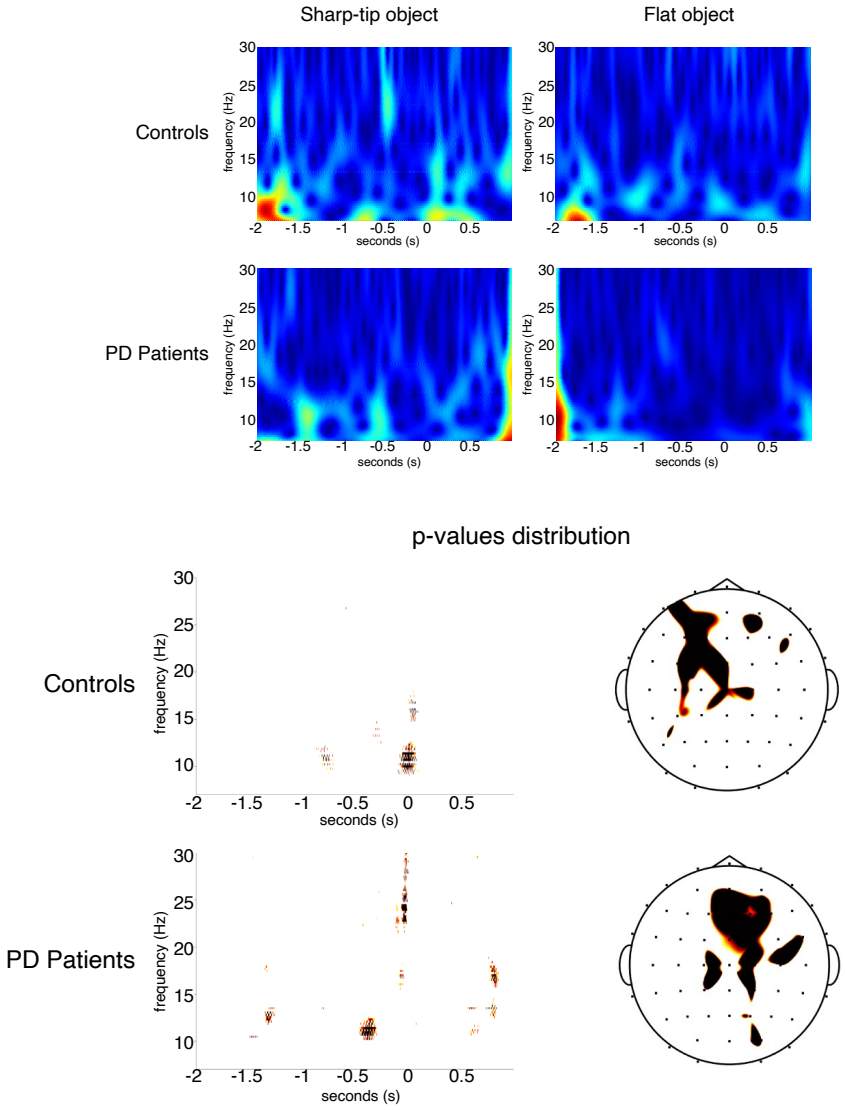


Fig. 5.2 Time-to-contact detection session: comparison between flat vs sharp-tip object. (Up) The Grand Average of time–frequency analysis of alpha-mu recorded on the C3 derivation in the 2 s preceding and 1 s following the flat and sharp-tip object grasping are reported for controls and PD patients. (Bottom) For each group, the p-values obtained with paired t-test between flat vs sharp tip object are reported on the C3 channel, and on the statistical map. Before the flat object trials, we observed that alpha-mu desynchronization prevailed in the 8–9.5 Hz range in the 2 s time in controls, and in the 1 s time in the 11–13 Hz range in PD patients.

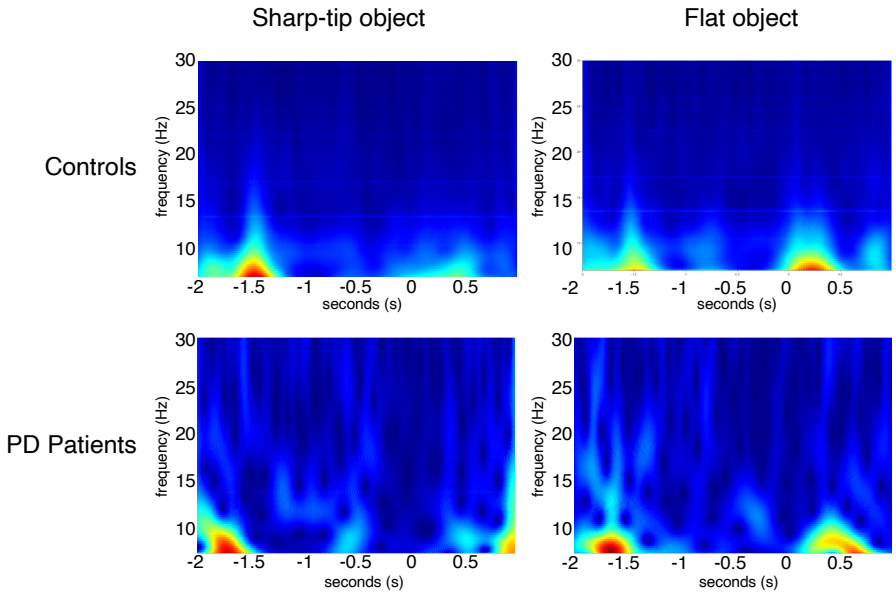


Fig. 5.3 Observation-only session: comparison between flat vs sharp-tip object. The Grand Average of time–frequency analysis of alpha-mu recorded on the C3 derivation in the 3 s preceding and 1 s following the flat and sharp tip object grasping are reported in controls and PD patients.

5.1.4 Discussion

Recent works reported that action observation therapy (AOT) has shown evidence of efficacy as a rehabilitation strategy in PD patients [187, 188]. Such an approach for therapy, in fact, was revealed to be effective in both single-session experiments and long-term therapeutic programs; in addition, a recent review work evidenced how this kind of approach was easier to apply respect to others, such as those based on Motor Imagery [180]. However, mirroring circuits activation seems weaker in PD patients during the observation of others' gait, as compared to controls.

Here, the first evidence of active motor resonance mechanisms in early PD was provided, with active response facilitation obtained with the observation of movement with explicit semantic clues. Such phenomenon could compensate for a possible initial failure in motor programming, as shown by the good performance PD patients demonstrated in motor reaction after the more suitable movement observation.

Overall, the behavioral, metabolic, and EEG data suggested that MR mechanisms are preserved in early-stage PD patients, since no significant differences were found between

groups, and similar cortical activation during the observation of congruent movements was observed [186].

Focusing on EEG results, in the resting state, PD patients exhibited greater alpha mu desynchronization compared to controls. Both PD patients and controls experienced alpha mu desynchronization when observing flat object grasping before responding, while this effect was less pronounced when observing sharp-tip objects. Among controls, observing the sharp-tip grasping triggered desynchronization of the alpha mu rhythm, likely indicating cortical activation in response to the anticipation of an incorrect action by the agent.

Nonetheless, some differences between PD patients and healthy individuals were evident. In resting-state, alpha mu desynchronization prevailed in PD patients on the left fronto-central regions, in the resting state before the task requiring a behavioural motor reaction. In PD patients, the preparation for active movement could request additional resources with respect to controls, so we could not exclude that compensatory phenomena of cortical activation may support motor reaction.

For Time-to-contact detection session, the time-frequency analysis showed a desynchronization of alpha rhythm in the 1 s time preceding and following the movement observation, which was similar for the 2 objects and for the 2 groups. This is in line with previous studies, showing that changes in EEG mu activity provide a valid means for the study of human neural mirroring. Similarly to fNIRS results, in both PD and control groups, we observed a prevalent alpha mu desynchronization in the time preceding the vision of the flat object grasping. The desynchronization was represented on the parieto-occipital and central electrodes in controls and left prefrontal and posterior central electrodes in patients. No significant differences were detectable between groups with regard to the spatial distribution of the desynchronization induced by the congruent movement. Contamination with occipital alpha suppression is possible during a visual task, and the lack of topographic specificity of mu desynchronization may be a result of more general attention processes [189]. The contribution of neuroimaging methods, such as fMRI and fNIRS could further clarify the role of cortical regions involved in mirroring phenomena, as in the present study, in which an increase of oxyhemoglobin levels was detected in the motor network. The lack of statistical differences in alpha mu desynchronization modality between patients and controls is a confirmation of the substantial integrity of motor resonance mechanisms in early PD.

For Observation-only session, the alpha mu showed desynchronization in the second preceding and following the object grasping, in a more evident way in controls for the ungraspable object. This type of EEG phenomenon could be explained with a sort of mirroring activation due to other potential motor failures and was strictly time-related to the

vision of the hand grasping. The alpha mu desynchronization is associated with execution more than observation [189], and this could explain the contradictory results obtained with the 2 brain functional analysis methods. The phenomenon we observed in controls with the fNIRS method, consistent with a cortical activation induced by the more suitable movement, was computed in the global time of the task and not evident in the time-frequency EEG analysis which, on the other hand, displayed a time related cortical reaction to uncorrected movement, which was not detectable with the fNIRS method. In any case, PD patients seemed less reactive at the cortical level during the simple observation of movement, supporting the hypothesis that mirror phenomena could have a function of motor facilitation in patients with initial dysfunction of movement programming.

Based on the present results, we could suppose that modifying the content of action observation, in order to stimulate motor resonance with the use of congruent movement, could improve the efficacy of such rehabilitation strategies. Indeed, a recent work made evident the modulation induced by motor resonance in healthy subjects, linking such excitability to the efficacy of the AOT itself [190].

5.2 Statistical inference and dynamic-causal modeling

In neuroimaging, accurately characterizing the directed interactions between brain regions is critical for understanding the underlying neural dynamics responsible for cognitive processes and pathological states. Most methods for assessing connectivity in neuroimaging studies (e.g. MEG/EEG and fMRI) rely on *functional* connectivity measures, such as phase synchronization, temporal correlations, or coherence between the activity of two regions, either at the scalp or source level. Functional connectivity captures statistical dependencies between time series and is advantageous because it does not require prior assumptions about the interactions or their causal nature. However, in some cases, the primary interest lies in understanding the causal architecture of these interactions [191]. Unlike functional connectivity, Dynamic Causal Modeling (DCM) focuses on *effective* connectivity, which specifically refers to the directed influence one neuronal system exerts over another .

DCM is a computational framework that models interactions among cortical regions, allowing to make inferences about the system's parameters and investigate how these parameters are influenced by experimental factors. This inference relies on an underlying generative model, which is informed by prior knowledge of neural dynamics and is typically expressed as a set of differential equations. These equations describe the flow of neural activity between regions and how the activity changes over time [192]

This approach goes beyond purely statistical associations and aims to model the underlying neural dynamics that generate an observed evoked response, such as ERP signals. It focuses on how brain regions interact to produce these signals, using a biophysically realistic model of neural activity and connectivity. In DCM, the brain is modeled as a deterministic, nonlinear dynamic system that responds to external inputs and generates observable outputs [193].

Since ERPs provide rich temporal data, the state equations in DCM for ERPs are more detailed compared, for example, to those used for fMRI. Bayesian inference is often employed to estimate the model parameters, which include the connectivity strengths between brain regions, the time constants of neural processes, and the amplitude of neuronal responses. Bayesian inference combines prior knowledge with observed data (i.e., ERP recordings) to estimate the parameters of the DCM. This is done by specifying a prior distribution for the model parameters, which encodes initial beliefs about the causal interactions in the brain. As the model is fitted to the observed data, the prior is updated through model inversion, resulting in a posterior distribution that represents the estimated parameters with their associated uncertainty.

In mathematical terms, Bayes' theorem can be written as:

$$P(\theta|\text{data}) = \frac{P(\text{data}|\theta)P(\theta)}{P(\text{data})} \tag{5.1}$$

where:

- $P(\theta|\text{data})$ is the posterior distribution (our updated belief about the parameters θ after seeing the data),
- $P(\text{data}|\theta)$ is the likelihood, which tells us how likely the observed data is, given certain parameter values,
- $P(\theta)$ is the prior, our belief about the parameters before seeing the data,
- $P(\text{data})$ is a normalizing constant that ensures the posterior is a valid probability distribution.

Bayesian methods are particularly suited for DCM because they can accommodate uncertainty in both the model and the data, allowing for robust estimates of effective connectivity. The posterior distributions provide not just point estimates but a range of plausible values for each parameter, reflecting the variability in the data and any prior knowledge.

5.2.1 DCM-informed classification of ERPs

The idea behind this research topic is to develop a highly generalizable and modular framework that incorporates extracting ERPs from real-data of heterogeneous cohorts of subjects, fitting DCM models to the data and obtaining the models' parameters. These parameters can be viewed as biomarkers that yield a low dimensional, interpretable feature space that allows the description of differences between subjects at individual level. Thus, this work aims to demonstrate that a combination of biophyisical models and Machine Learning may outperform traditional approaches based on raw brain data.

As a first attempt to develop the aforementioned framework, data from the Human Intracranial Database [194] available on EBRAINS Knowledge Graph was used. This is a dataset of stereotactic electroencephalography (sEEG) recordings from 100 epileptic patients, collected while patients performed up to eight behavioral tasks designed to activate large-scale neural networks involved in various cognitive functions such as language, memory, visual attention, and motor behavior. The participants in the HID were patients undergoing surgical

evaluation for drug-resistant partial epilepsy. These patients were recruited because non-invasive methods failed to identify the epileptic focus, necessitating the use of intracranial EEG (iEEG) recordings. iEEG recordings were conducted using a video-iEEG monitoring system allowing simultaneous data recording from 128 depth-EEG electrode sites, sampled at 512 Hz. One of the contact sites in the white matter was chosen as a reference. In addition, all signals were re-referenced to their nearest neighbor on the same electrode before analysis.

The continuous iEEG signals were initially filtered using a band-pass filter across multiple successive 10 Hz-wide frequency ranges (e.g., 10 bands from 50–60 Hz to 140–150 Hz). For each band, the signal envelope was computed using the Hilbert transform. The resulting envelope had a time resolution of 15.625 ms. To normalize the data, the envelope signal for each band was divided by its mean over the entire recording session and then multiplied by 100, producing instantaneous envelope values as a percentage of the mean. Finally, the envelope signals from the consecutive frequency bands (spanning 10 Hz intervals between 50 and 150 Hz) were averaged to create a single time-series for the entire session. The Visual Search Task (MCSE) was selected among different tasks to extract ERPs. This task was designed to test participants' ability to find a target (a gray "T") among distractors (tilted "L" shapes) displayed on a screen. Participants were required to indicate, as quickly as possible, whether the target was located in the upper or lower part of the array by pressing one of two buttons (right index or right middle finger). The task was divided into two conditions, easy and hard. In the easy condition, distractors were black while the target was gray; in the hard condition, both the distractors and the target were gray. The task consisted of 8 blocks, each block containing 12 trials, with 6 easy and 6 hard trials presented in a pseudo-random order.

The binary classification problem was structured to discriminate trial difficulty based on the event-triggered high-gamma signal, as shown in Figure 5.4.

Of course, while this task does not provide a clinical relevance for this study, it is simple enough to assess the proposed method.

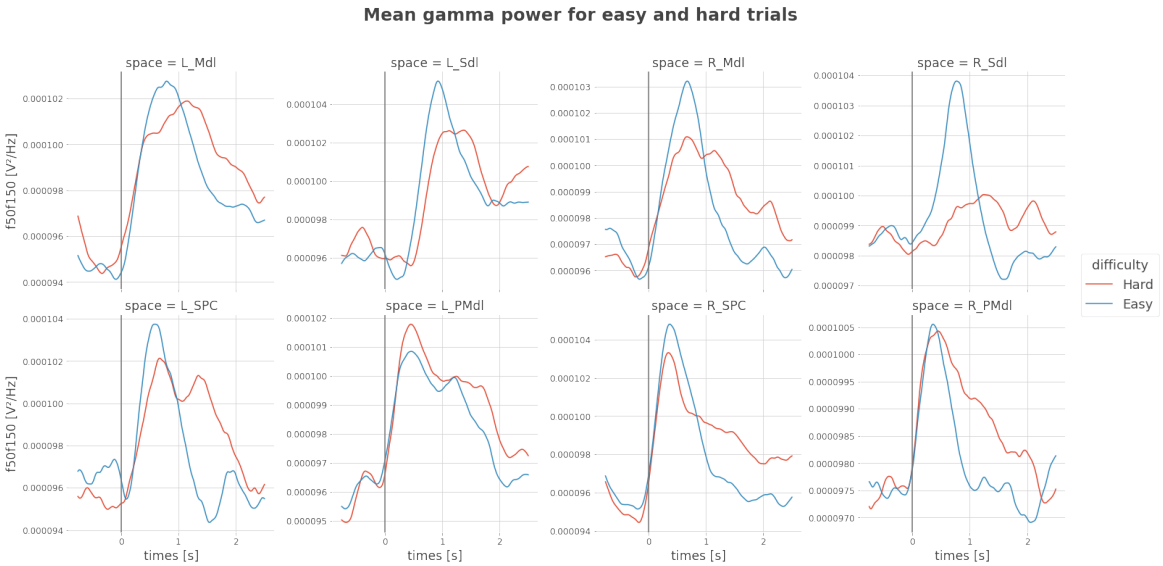


Fig. 5.4 Examples of high-gamma ERPs for easy and hard trials.

The ERP model was based on the DCM approach developed by David *et al.* [195] which incorporates the connectivity principles outlined by Felleman and Van Essen [196] to construct networks of interacting neural sources. Each source is modeled using a neural mass framework, building upon the model by Jansen and Rit [197]. This model represents the activity within a cortical area through three distinct neuronal subpopulations, corresponding to granular and agranular cortical layers. In this model, excitatory pyramidal cells, which act as the primary output neurons, receive input from both excitatory and inhibitory interneurons via intrinsic connections confined to the cortical structure. Excitatory interneurons are represented as spiny stellate cells, which are predominantly located in layer four and receive forward inputs. The excitatory pyramidal cells and inhibitory interneurons, located in agranular layers, process backward and lateral inputs within the network.

DCM parameter estimates were obtained by fitting ERPs during the easy and hard tasks. We fitted ERP recordings from different participants, thus obtaining DCM parameter estimates. Figure 5.5 shows the results obtained for different brain regions and different task difficulties.

These parameters were then used as input features to train and test a simple logistic regression model. The same model was then trained using the raw signal samples as features, and also concatenating raw signals with the parameters of the DCM fit.

Results obtained over 5-fold cross-validation demonstrated that adding the parameters to the original samples actually improved the mean classification accuracy (LR: 0.82 ± 0.15 with raw signal samples, 0.88 ± 0.14 with aggregated signals and parameters). This could

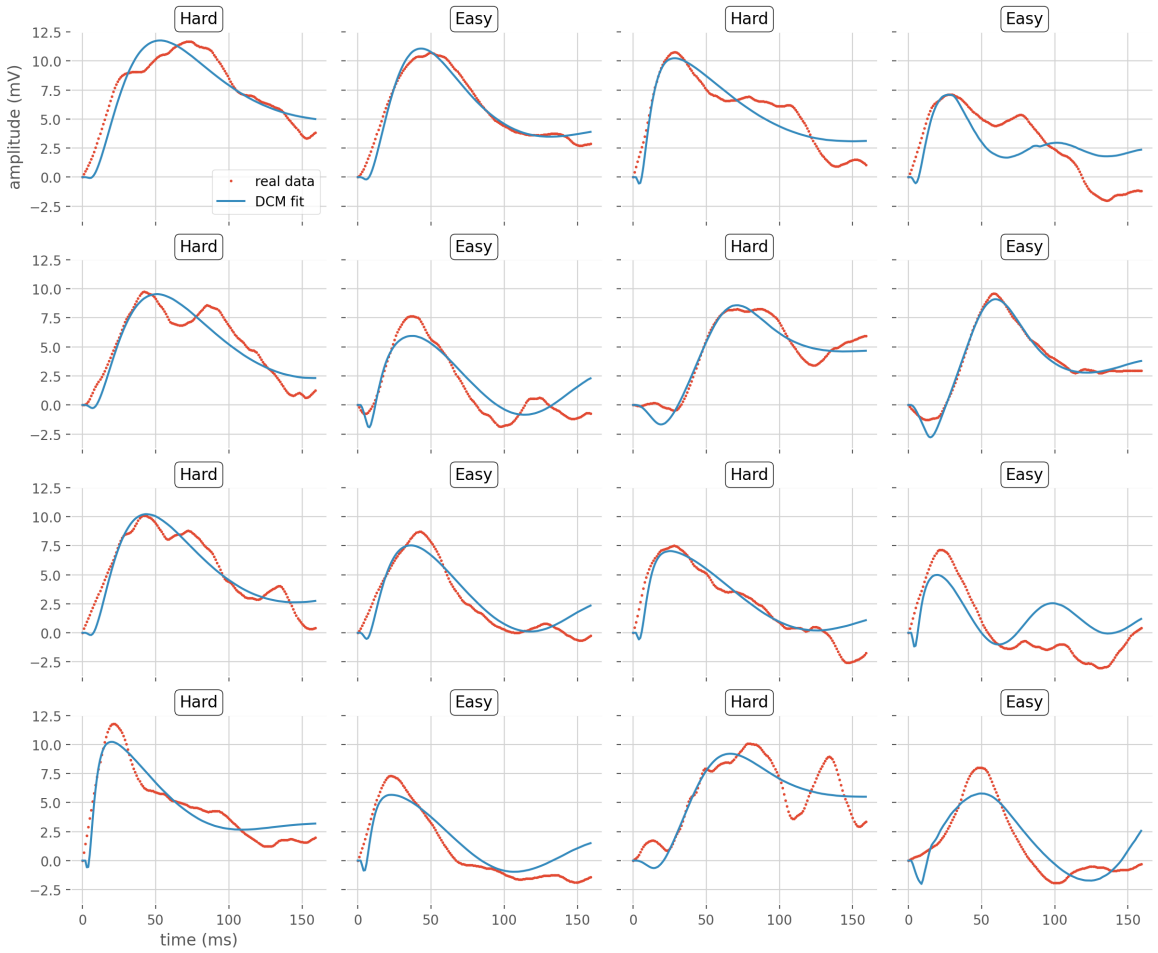


Fig. 5.5 Original and simulated ERPs obtained by fitting the DCM model. Signals are averaged over subjects and trials.

indicate that the information carried by the parameters' estimates of the DCM model is used by the classifier to discriminate between the two classes. Although these results should be confirmed and further assessed, another recent study proved a similar approach to be effective in classifying depression against controls [198], indicating that byophisical models of EEG activity could provide biomarkers of pathological states.

5.3 Effects of cross-modal stimulation in Human-Robot Interaction

5.3.1 Motivations

Robotics is increasingly vital in cognitive neuroscience, providing innovative tools and methodologies to investigate the complexities of human brain function and behavior [199]. The integration of humanoid robots and Human-Robot Interaction (HRI) paradigms facilitates a novel approach to understanding the neural and psychological processes that underpin social interaction, empathy, and the perception of agency [200, 201]. This perspective is valuable for elucidating how the brain responds to social signals, body language, and facial expressions in robotic agents, thereby enhancing our understanding of the neural foundations of social cognition [202]. One promising application of humanoid robots lies in the investigation of putative human pheromones (PP). Research over the past few decades has yielded contentious results regarding the existence and function of pheromonal signaling in humans, particularly concerning the vomeronasal system (VNS) [203, 204]. While anatomical and genetic studies suggest that the VNS is diminished or absent in humans [205, 206], evidence from other mammals indicates that pheromone-based communication remains vital [207, 208]. Notably, specific steroidal compounds, such as androstadienone and estratetraenol, have been identified as potential human pheromones, yet solid empirical support for their effects remains elusive. Methodological challenges, including difficulties in replicating results and limited sample sizes, complicate the field. Additionally, research suggests that these chemosensory mechanisms operate below conscious awareness, necessitating careful experimental design [209, 210]. Employing humanoid robots in these studies may enhance experimental control and replicability, allowing for a more nuanced exploration of how PP influences neural and behavioral responses during HRI. This innovative approach could illuminate the interplay between olfactory cues, proxemic behavior, and social communication, ultimately contributing to a deeper understanding of human social dynamics.

Recent research showed that PP affects brain activity and the sense of co-presence in a gender-dependent fashion during interaction with an embodied medium [211]. Building on these findings and the literature presented above, the aim of this research work was to investigate, during an experimental setting of HRI, the influence of cross-modal stimulation linked to gender aspects (i.e., PP and the gender of the voice) and proxemic space variations on behavioral and electrophysiological responses.

5.3.2 Experimental setup

The experiment was conducted in the Laboratory of Cognitive and Psychophysiological Olfactory Processes – INSPIRE Lab – of the University of Salento, Lecce, Italy. Fifty healthy students (women n=25; mean age=22.6 y.o., standard deviation SD=4) took part in the experiment. General information about the participants was collected through a short questionnaire. The humanoid robot NAO, created by Aldebaran Robotics, was used as a robotic interface and as a physical actuator to administer the embodied social cues stimuli, while opportunely playing a recording of a reduced Italian adaptation of the story for children "Freddie the Leaf" by Leo F. Buscaglia, lasting approximately 5 minutes. This story was previously recorded with a male (M) and a female (F) voice. During the playback of the audio, every 10 seconds, the robot would perform one out of four movements to reduce and increase the proxemic space of the participants: walk forward (S1), arm forward (S2), arm backward (S3), and walk backward (S4). The electroencephalographical signal was recorded from the scalp of the participants using a 64 active electrode cap (ActiCHamp, Brain Products, Munich, Germany), according to the international 10–10 system, with a sampling frequency of 1,000 Hz. Human putative pheromones (PP) 1,3,5(10),16-estratetraen-3-ol (Steraloids, Inc., Newport, R.I.; CAS number: 1150-90-9; Estr, E) and 4,16-Androstadien-3-one (Steraloids, Inc., Newport, R.I.; CAS number: 4075-07-4; Andr, E) were used. Vaseline oil alone (Neut, N) was used as a control substance. Three pheromonal conditions (N, E, A) were combined with two genders (F, M) of the narrative voice, resulting in a total of six experimental conditions per subject: NF, NM, EF, EM, AF, and AM. They were presented in a balanced and pseudorandomized fashion across subjects so that none of the participants underwent the same order of conditions. Before the beginning and out of sight of the subject, one experimenter (F or M depending on the condition) replaced the vial cork with a drilled one and placed the vial in a necklace with a special accommodation. Then, the collar was positioned around Nao's neck so that the pheromone could volatilize during that condition. Nao was placed 70 cm from the subject; the latter was asked to observe the robot during the listening. Each condition lasted 5 minutes during which the robot performed a movement of reduction or increase of the proxemic space every 10 seconds. Between one storytelling session and another Nao was taken out of the room so that its necklace could be replaced. In the meanwhile, the door and the window were kept open, and the subject was asked to fill out the questionnaire.

Statistical analysis was performed on EEG data using Matlab's Letswave 7 tool. We conducted a main analysis by adopting a two-way point-by-point analysis of variance (ANOVA). The factorial design included the group (men and women) as a between-subject factor,

whereas the within-subject factors were social odor condition (N, E, and A), voice (M and F), proximity space (Forward and Backward), and body (LEGS and ARMS). To estimate the significance of the amplitude responses across time and electrodes (post-stimulus), post-hoc comparisons were performed with the non-parametric cluster-based permutation Student's t-test for paired data, as in previous studies of this thesis. Multi-sensor analysis was performed in order to consider both temporal and spatial adjacency of the samples.

The spatio-temporal evolution of the significant effects is represented through topographical maps of the clusters averaged in bins of 50 ms. For all the analyses, the channels for which the t-value exceeded the statistical threshold ($p < 0.05$) were considered significant.

5.3.3 Results and discussion

The electrophysiological findings indicate that spatial variations, specifically in forward versus backward movements, significantly influence brain activation patterns, with variations also based on the moving body part (LEGS or ARMS) and narrating voice (M or F). For movement direction, forward movements generated increased activation in temporal, parietal, centro-parietal, and occipital regions across both conditions (AF and NF) (see Figure 5.6). Conversely, backward movements led to stronger frontal activation, though with latency differences between the AF and NF conditions, showing a quicker response in the AF condition.

Additionally, the results reveal distinct temporal windows for these effects, notably from 50-600 ms in AF (with a focus on 350-550 ms) and two clusters in NF (0-500 ms and 500-1000 ms). These observations suggest that different spatial movements (forward or backward) elicit unique brain activation patterns, impacting specific brain regions in ways that vary based on the narrating voice and timing.

The results further highlight significant interactions between the robot's moving body part (ARMS vs. LEGS) and experimental conditions, particularly regarding proximity space, social odors, and narrating voice.

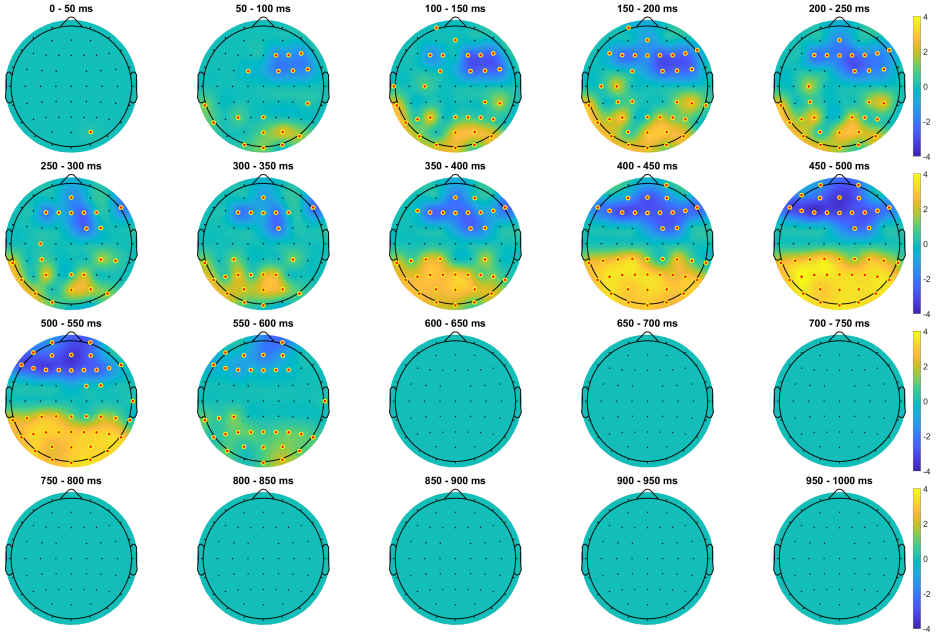
Body Part Motion (ARMS vs LEGS): The movement of the ARMS had a pronounced effect in several conditions (AF, AM, EM) (Figure 5.7). A consistent activation pattern emerged over time, beginning in the frontotemporal areas and expanding to central, parietal, temporal, and occipital regions between 350-750 ms, especially when interacting with social odor A. In contrast, the LEGS movement showed significance only in the AM condition, activating central and posterior regions with an initial response between 50-450 ms. At later time points, the arms movement dominated once more, particularly in posterior brain areas.

Interaction with Social Odors (Andr, Estr, Neuter): The Andr pheromone significantly influenced brain activity when paired with ARMS movement, marked by activation in both frontotemporal and posterior regions during specific time windows. - Estr pheromone also produced significant effects, though primarily in the central, parietal, and temporal regions, suggesting a different neural response compared to Andr. Comparisons with the Neuter odor revealed distinct patterns, especially in posterior regions during backward movements, indicating unique neural signatures for both Andr and Estr versus Neuter.

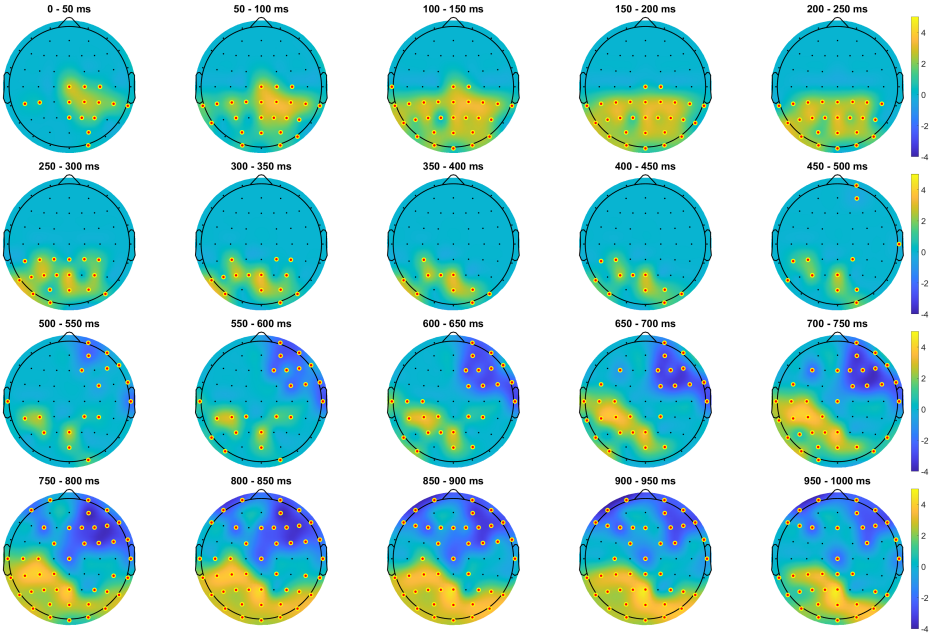
Effect of Proximity Space and Direction (Forward vs Backward): Brain activity differed noticeably between forward (S1, S2) and backward (S3, S4) robot movements (see Figure 5.6). Forward movements primarily influenced central regions, while backward movements strongly impacted posterior regions, with pronounced differences when comparing social odors (A, E, N). Significant differences were noted in backward movements across AF, AM, EM, EF, NM, and NF conditions (see Figure 5.8). The right front-lateral area was more activated in the NF condition, especially during the latter half of the post-stimulus period.

Gender Voice (Male vs Female Narrating Voice): The narrating voice's gender only significantly affected the Neuter condition (see Figure 5.8), with male voices inducing stronger activation in left-posterior regions and female voices increasing activation in front-lateral regions. Notably, no significant differences were observed for the narrating voice in the Andr or Estr conditions, suggesting that the Neuter odor enhances the voice effect.

In summary, these findings underscore the complexity of neural responses shaped by the combined influences of body part movement, proximity space, social odors, and narrating voice. Social odors, in particular, appear to modulate brain activity, especially during backward movements (see Figure 5.9), with some conditions showing enhanced activation in both frontal and posterior regions.

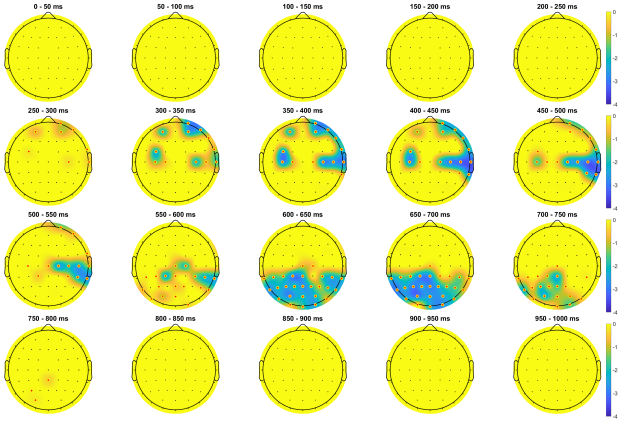


(a)

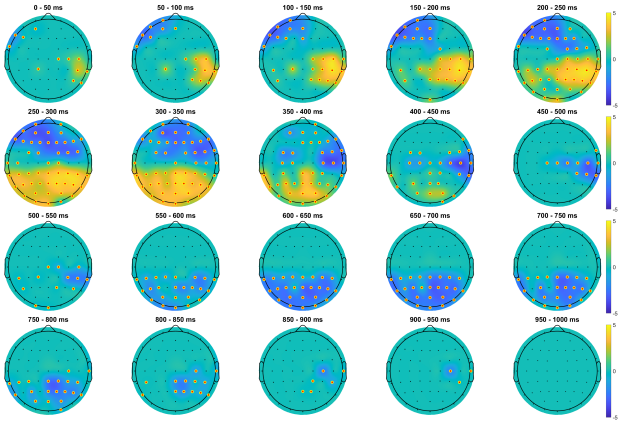


(b)

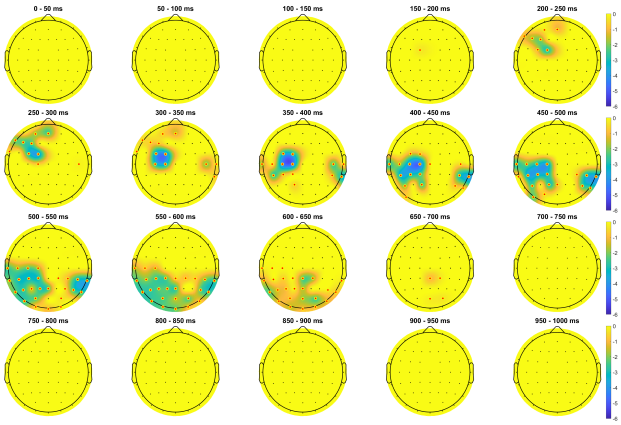
Fig. 5.6 Clustered t-value for AF (a) and NF (b) in Forward vs Backward. The color map represents the value of the t-statistic. Significant channels ($p < 0.05$) are circled and highlighted in red.



(a)

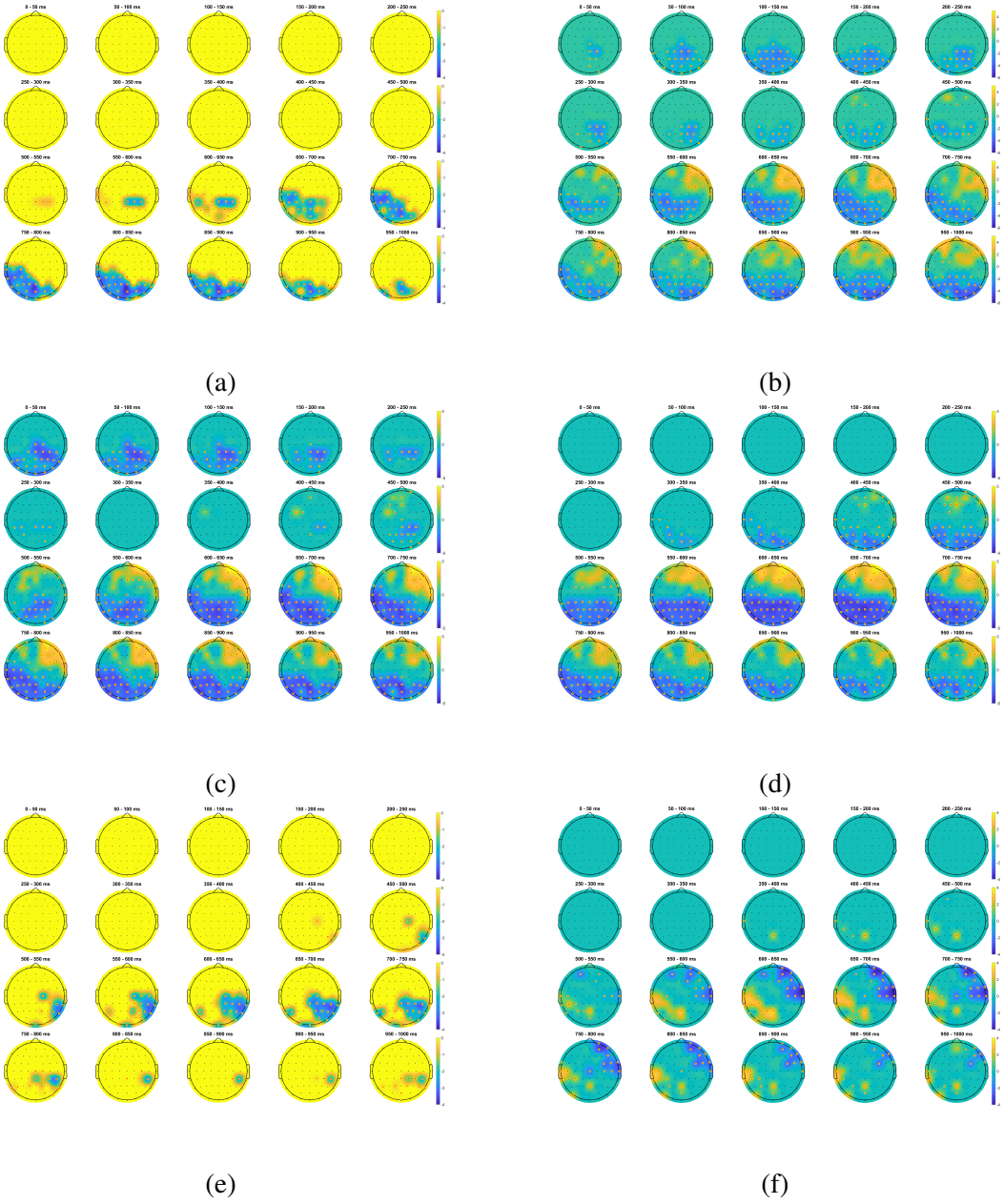


(b)



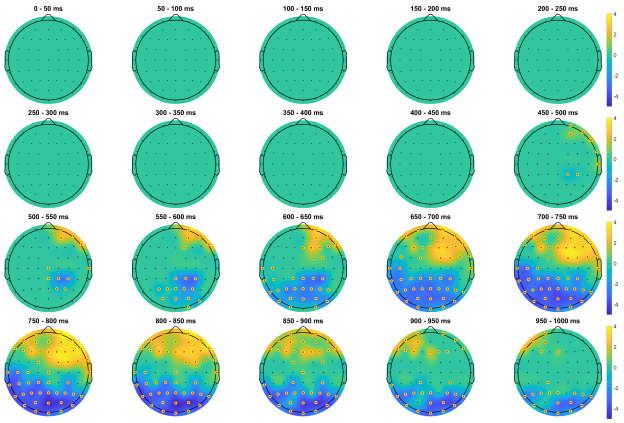
(c)

Fig. 5.7 Clustered t-value for AF (a), AM (b) and EM (c) in ARMS vs LEGS. The color map represents the value of the t-statistic. Significant channels ($p < 0.05$) are circled and highlighted in red.

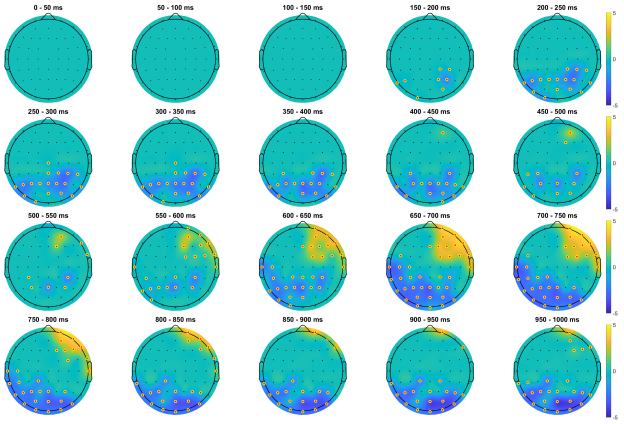


(a) (b) (c) (d) (e) (f)

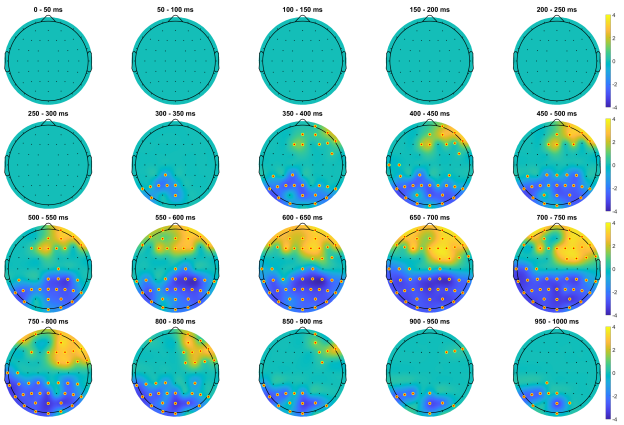
Fig. 5.8 Clustered t-value for Backward in AF vs NF (a), AM vs NF (b), EF vs NF (c), EM vs NF (d), EM vs NM (e), and NM vs NF (f). The color map represents the value of the t-statistic. Significant channels ($p < 0.05$) are circled and highlighted in red.



(a)



(b)



(c)

Fig. 5.9 Clustered t-value for Men Backward in AF vs NF (a), AM vs NF (b), EM vs NF (c). The color map represents the value of the t-statistic. Significant channels ($p < 0.05$) are circled and highlighted in red.

Chapter 6

Conclusion

The overall purpose of this Ph.D. dissertation was to conceptualize, develop and evaluate novel computational methods for processing electrophysiological signals in order to support early clinical diagnosis and progression monitoring of neurodegenerative diseases.

In particular, the work aimed to advance the field of neurodegenerative disease diagnostics by focusing on the role of Deep Learning and EEG biomarkers, specifically for the classification and differentiation of subjective cognitive decline and mild cognitive impairment.

After the introduction highlighting the need for the automatic identification of non-invasive and cost-effective tools for decoding neurodegeneration, a review of the current scientific and clinical challenges associated with the intrinsic complexity of the continuum of Alzheimer's and other neurological diseases has been detailed in Chapter 2. In this context, the potential of automatic extraction of EEG features through accurate and reliable DL models has been emphasized. This groundwork underscored the significance of resting-state EEG as a biomarker, which motivated the development of intelligent systems for EEG signal classification and analysis, and the investigation of principles behind the developed systems, by exploiting explainability techniques.

Indeed, Chapter 3 presented the first DL framework that employs the attention mechanism implemented by the Transformer model to classify patients affected by early-stage conditions of AD at individual level, using resting-state EEG signal. The results obtained by training and testing the model on EEG data corresponding to different frequency bands confirmed previous findings, which revealed a correlation between clinical progression of the disease and signal alterations in specific frequency bands, e.g. power spectrum shifts from high-frequency components (α and β) towards low-frequency components (δ and θ). The robustness of the classification model has been further confirmed by the performance obtained in the

discrimination of healthy controls, SCD and MCI. Results have been compared with other works found in the state-of-the-art analysis.

Building on those premises, in Chapter 4 a complete end-to-end framework which leverages attention scores to gain insights of the developed Transformer model was described. The focus of the Transformer, which corresponds to the highest attention on specific signal patches, is representative of hallmark EEG patterns that could allow to discriminate SCD from MCI. This could be employed as a guide for experts to facilitate the extraction of rsEEG markers of cognitive decay. Early identifying the prodromal stages of AD has become fundamental, since risky subjects might represent a target population for disease-modifying therapies. This work took an essential step toward the integration of AI tools in personalized medicine for neurodegeneration, ultimately advancing the field toward improved patient outcomes and quality of life.

Lastly, in Chapter 5, computational methods for analyzing evoked response, i.e. event-related potentials and event-related (de)synchronization, were presented as they relate to neurodegenerative conditions, with a focus on early-stage Parkinson's disease. This Chapter investigated how ERD/ERS components, associated with motor and sensory processing, could serve as markers for understanding the progression of impairment in neurodegeneration. By examining the effects of movement congruence on motor resonance, the study demonstrated the substantial preservation of motor resonance mechanisms in early PD patients and the possibility that the action observation finalized to a consequent movement can activate cortical networks in patients with no advanced motor limitations, allowing early rehabilitation interventions with specific observation paradigms.

Furthermore, a framework based on dynamic-causal modeling on ERP data to capture underlying neural dynamics was presented. This approach also supported DCM-informed classification, which demonstrated the utility of combining statistical inference with Machine Learning for ERP-based diagnostics, paving the way towards a personalized pathological modeling of neurodegenerative processes.

A comprehensive method for evaluating the impact of social and environmental factors, such as gender voices and proxemic variations, on ERP responses, revealing how these external influences could modulate neural processing in physiological conditions was also introduced. Further research could examine how factors like stress, social interaction, and environmental changes influence neural responses in neurodegenerative patients. This could provide a more holistic understanding of how external conditions impact disease progression, which may be especially relevant for creating supportive environments in clinical or home settings.

My Publications

1. **Sibilano E.**, Brunetti A., Buongiorno D., Lassi M., Grippo A., Bessi V., Micera S., Mazzoni A., Bevilacqua V. (2023). An attention-based deep learning approach for the classification of subjective cognitive decline and mild cognitive impairment using resting-state EEG. *Journal of neural engineering*, 20(1), doi:10.1088/1741-2552/acb96e.

2. **Sibilano E.**, Buongiorno D., Lassi M., Grippo A., Bessi V., Sorbi S., Mazzoni A., Bevilacqua V., Brunetti A. Understanding the role of self-attention in a Transformer model for the discrimination of SCD from MCI using resting-state EEG. *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 6, pp. 3422-3433, June 2024, doi: 10.1109/JBHI.2024.3390606.

3. Suglia V., Brunetti A., Pasquini G., Caputo M., Marvulli T.M., **Sibilano E.**, Della Bella S., Carrozza P., Beni C., Naso D., Monaco V., Cristella G., Bevilacqua V., Buongiorno D. A Serious Game for the Assessment of Visuomotor Adaptation Capabilities during Locomotion Tasks Employing an Embodied Avatar in Virtual Reality. *Sensors*. 2023;23(11):5017. doi:10.3390/s23115017

4. Gentile E., Brunetti A., Ricci K., Vecchio E., Santoro C., **Sibilano E.**, Bevilacqua V., Iliceto G., Craighero L., de Tommaso M., Effects of movement congruence on motor resonance in early Parkinson's disease. *Scientific Reports* 13, 14887 (2023). doi:10.1038/s41598-023-42112-2

5. **Sibilano E.**, Lassi M., Mazzoni A., Bevilacqua V., Brunetti A. (2023). A Deep Learning Framework for the Classification of Pre-prodromal and Prodromal Alzheimer's Disease Using Resting-State EEG Signals. In: Esposito, A., Faundez-Zanuy, M., Morabito, F.C., Pasero, E. (eds) *Applications of Artificial Intelligence and Neural Systems to Data Science. Smart Innovation, Systems and Technologies*, vol 360. Springer, Singapore. doi:10.1007/978-981-99-3592-5_9

6. **Sibilano E.**, Algieri A., Bevilacqua V., Buongiorno D., Brunetti A. (2023). Major Depressive Disorder Classification with 3D CNNs and Grad-CAM Visualization on structural Magnetic Resonance Images. In Smart Innovation, Systems and Technologies Springer, Singapore (to appear)

7. **Sibilano E.**, Suglia V., Brunetti A., Buongiorno D., Caporusso N., Guger C., Bevilacqua V. Brain-Computer Interfaces (2023). In *Psychophysiology Methods* (pp. 203-240). New York, NY: Springer US.

References

[1] Rakesh Kumar Sahoo, Tanisha Gupta, Vinay Kumar, Sarita Rani, Umesh Gupta, et al. Aetiology and pathophysiology of neurodegenerative disorders. In *Nanomedical Drug Delivery for Neurodegenerative Diseases*, pages 1–16. Elsevier, 2022.

[2] Richard NL Lamptey, Bivek Chaulagain, Riddhi Trivedi, Avinash Gothwal, Buddhadev Layek, and Jagdish Singh. A review of the common neurodegenerative disorders: current therapeutic approaches and the potential role of nanotherapeutics. *International journal of molecular sciences*, 23(3):1851, 2022.

[3] Yujun Hou, Xiuli Dan, Mansi Babbar, Yong Wei, Steen G Hasselbalch, Deborah L Croteau, and Vilhelm A Bohr. Ageing as a risk factor for neurodegenerative disease. *Nature Reviews Neurology*, 15(10):565–581, 2019.

[4] Eleni Kanasi, Srinivas Ayilavarapu, and Judith Jones. The aging population: demographics and the biology of aging. *Periodontology 2000*, 72(1):13–18, 2016.

[5] Junfang Xu, Yuqian Zhang, Chengxuan Qiu, and Feng Cheng. Global and regional economic costs of dementia: a systematic review. *The Lancet*, 390:S47, 2017.

[6] Reisa A Sperling, Paul S Aisen, Laurel A Beckett, David A Bennett, Suzanne Craft, Anne M Fagan, Takeshi Iwatsubo, Clifford R Jack Jr, Jeffrey Kaye, Thomas J Montine, et al. Toward defining the preclinical stages of alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):280–292, 2011.

[7] Carles Gaig and Eduardo Tolosa. When does parkinson's disease begin? *Movement Disorders*, 24(S2):S656–S664, 2009.

[8] Sandra Weintraub. Neuropsychological assessment in dementia diagnosis. *CONTINUUM: Lifelong Learning in Neurology*, 28(3):781–799, 2022.

[9] AD Hutchinson and Jane L Mathias. Neuropsychological deficits in frontotemporal dementia and alzheimer's disease: a meta-analytic review. *Journal of Neurology, Neurosurgery & Psychiatry*, 78(9):917–928, 2007.

[10] Clifford R Jack Jr, David A Bennett, Kaj Blennow, Maria C Carrillo, Billy Dunn, Samantha Budd Haeberlein, David M Holtzman, William Jagust, Frank Jessen, Jason Karlawish, et al. Nia-aa research framework: toward a biological definition of alzheimer's disease. *Alzheimer's & dementia*, 14(4):535–562, 2018.

[11] Ronald B Postuma and Daniela Berg. The new diagnostic criteria for parkinson's disease. *International Review of Neurobiology*, 132:55–78, 2017.

[12] Oskar Hansson. Biomarkers for neurodegenerative diseases. *Nature medicine*, 27(6): 954–963, 2021.

[13] Raymundo Cassani, Mar Estarellas, Rodrigo San-Martin, Francisco J Fraga, and Tiago H Falk. Systematic review on resting-state eeg for alzheimer's disease diagnosis and progression assessment. *Disease markers*, 2018, 2018.

[14] Paolo Maria Rossini, Riccardo Di Iorio, Francesco Vecchio, Maria Anfossi, Claudio Babiloni, Marco Bozzali, Amalia Cecilia Bruni, Stefano F Cappa, Julien Escudero, Francisco Jose Fraga, et al. Early diagnosis of alzheimer's disease: the role of biomarkers including advanced eeg signal analysis. report from the ifcn-sponsored panel of experts. *Clinical Neurophysiology*, 131(6):1287–1310, 2020.

[15] Monika A Myszczynska, Poojitha N Ojamies, Alix MB Lacoste, Daniel Neil, Amir Saffari, Richard Mead, Guillaume M Hautbergue, Joanna D Holbrook, and Laura Ferraiuolo. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nature reviews neurology*, 16(8):440–456, 2020.

[16] Loveleen Gaur. *AI and Neuro-Degenerative Diseases: Insights and Solutions*. Springer Nature, 2024.

[17] Feng-Lei Fan, Jinjun Xiong, Mengzhou Li, and Ge Wang. On interpretability of artificial neural networks: A survey. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(6):741–760, 2021.

[18] Jeffrey L Krichmar, James Leland Olds, Juan V Sanchez-Andres, and Huajin Tang. Explainable artificial intelligence and neuroscience: cross-disciplinary perspectives, 2021.

[19] Shuroug A Alowais, Sahar S Alghamdi, Nada Alsuhebany, Tariq Alqahtani, Abdulrahman I Alshaya, Sumaya N Almohareb, Atheer Aldairem, Mohammed Alrashed, Khalid Bin Saleh, Hisham A Badreldin, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC medical education*, 23(1):689, 2023.

[20] Amy R Dunn, Kristen MS O'Connell, and Catherine C Kaczorowski. Gene-by-environment interactions in alzheimer's disease and parkinson's disease. *Neuroscience & Biobehavioral Reviews*, 103:73–80, 2019.

[21] Todd E Golde. Disease-modifying therapies for alzheimer's disease: more questions than answers. *Neurotherapeutics*, 19(1):209–227, 2023.

[22] Jamal S Rana, Sadiya S Khan, Donald M Lloyd-Jones, and Stephen Sidney. Changes in mortality in top 10 causes of death from 2011 to 2018. *Journal of general internal medicine*, 36:2517–2518, 2021.

[23] Marta Crous-Bou, Carolina Minguillón, Nina Gramunt, and José Luis Molinuevo. Alzheimer's disease prevention: from risk factors to early intervention. *Alzheimer's research & therapy*, 9:1–9, 2017.

[24] Melanie Luppa, Tobias Luck, Siegfried Weyerer, Hans-Helmut König, Elmar Brähler, and Steffi G Riedel-Heller. Prediction of institutionalization in the elderly. a systematic review. *Age and ageing*, 39(1):31–38, 2010.

[25] Rawan Tarawneh and David M Holtzman. The clinical problem of symptomatic alzheimer disease and mild cognitive impairment. *Cold Spring Harbor perspectives in medicine*, 2(5):a006148, 2012.

[26] Ronald C Petersen. How early can we diagnose alzheimer disease (and is it sufficient)? the 2017 wartenberg lecture. *Neurology*, 91(9):395–402, 2018.

[27] Guy McKhann, David Drachman, Marshall Folstein, Robert Katzman, Donald Price, and Emanuel M Stadlan. Clinical diagnosis of alzheimer's disease: Report of the nincds-adrda work group* under the auspices of department of health and human services task force on alzheimer's disease. *Neurology*, 34(7):939–939, 1984.

[28] Bruno Dubois, Howard H Feldman, Claudia Jacova, Steven T DeKosky, Pascale Barberger-Gateau, Jeffrey Cummings, André Delacourte, Douglas Galasko, Serge Gauthier, Gregory Jicha, et al. Research criteria for the diagnosis of alzheimer's disease: revising the nincds–adrda criteria. *The Lancet Neurology*, 6(8):734–746, 2007.

[29] Bruno Dubois, Howard H Feldman, Claudia Jacova, Jeffrey L Cummings, Steven T DeKosky, Pascale Barberger-Gateau, André Delacourte, Douglas Galasko, Serge Gauthier, Gregory A Jicha, et al. Advancing research diagnostic criteria for alzheimer's disease: the iwg-2 criteria. *The Lancet Neurology*, 13(6):614–629, 2014.

[30] Marilyn S Albert, Steven T DeKosky, Dennis Dickson, Bruno Dubois, Howard H Feldman, Nick C Fox, Anthony Gamst, David M Holtzman, William J Jagust, Ronald C Petersen, et al. The diagnosis of mild cognitive impairment due to alzheimer's disease: recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):270–279, 2011.

[31] Bruno Dubois, Harald Hampel, Howard H Feldman, Philip Scheltens, Paul Aisen, Sandrine Andrieu, Hovagim Bakardjian, Habib Benali, Lars Bertram, Kaj Blennow, et al. Preclinical alzheimer's disease: definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, 12(3):292–323, 2016.

[32] Alzheimer's disease facts and figures. *Alzheimer's dementia: the journal of the Alzheimer's Association*, 19(4):1598–1695, 2023.

[33] Frank Jessen, Rebecca E Amariglio, Martin Van Boxtel, Monique Breteler, Mathieu Ceccaldi, Gaël Chételat, Bruno Dubois, Carole Dufouil, Kathryn A Ellis, Wiesje M Van Der Flier, et al. A conceptual framework for research on subjective cognitive decline in preclinical alzheimer's disease. *Alzheimer's & dementia*, 10(6):844–852, 2014.

[34] Frank Jessen, Rebecca E Amariglio, Rachel F Buckley, Wiesje M van der Flier, Ying Han, José Luis Molinuevo, Laura Rabin, Dorene M Rentz, Octavio Rodriguez-Gomez, Andrew J Saykin, et al. The characterisation of subjective cognitive decline. *The Lancet Neurology*, 19(3):271–278, 2020.

[35] Ran An, Yajing Gao, Xiuxiu Huang, Yi Yang, Chengfengyi Yang, and Qiaoqin Wan. Predictors of progression from subjective cognitive decline to objective cognitive impairment: a systematic review and meta-analysis of longitudinal studies. *International Journal of Nursing Studies*, 149:104629, 2024.

[36] Leonardo Zullo, Christopher Clark, Mehdi Gholam, Enrique Castelao, Armin von Gunten, Martin Preisig, and Julius Popp. Factors associated with subjective cognitive decline in dementia-free older adults—a population-based study. *International Journal of Geriatric Psychiatry*, 36(8):1188–1196, 2021.

[37] G Pusswald, D Moser, M Pflüger, A Gleiss, E Auff, E Stögmann, P Dal-Bianco, and J Lehrner. The impact of depressive symptoms on health-related quality of life in patients with subjective cognitive decline, mild cognitive impairment, and alzheimer's disease. *International Psychogeriatrics*, 28(12):2045–2054, 2016.

[38] Yi-Chia Wei, Yi-Chia Kung, Chemin Lin, Chun-Hung Yeh, Pin-Yuan Chen, Wen-Yi Huang, Yu-Chiau Shyu, Ching-Po Lin, and Chih-Ken Chen. Differential neuropsychiatric associations of plasma biomarkers in older adults with major depression and subjective cognitive decline. *Translational Psychiatry*, 14(1):333, 2024.

[39] Alex J Mitchell, Helen Beaumont, David Ferguson, Motahare Yadegarfar, and Brendon Stubbs. Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: meta-analysis. *Acta Psychiatrica Scandinavica*, 130 (6):439–451, 2014.

[40] A. C. van Harten, L. Si, and et al. Prevalence of subjective cognitive decline in older adults. *Journal of Alzheimer's Disease*, 63(2):673–678, 2018.

[41] Audrey Perrotin, Robin de Flores, Franck Lamberton, Geraldine Poisnel, Renaud La Joie, Vincent de la Sayette, Florence Mezenge, Clemence Tomadesso, Brigitte Landeau, Beatrice Desgranges, et al. Hippocampal subfield volumetry and 3d surface mapping in subjective cognitive decline. *Journal of Alzheimer's Disease*, 48(s1): S141–S150, 2015.

[42] Steffen Wolfsgruber, Alexandra Polcher, Alexander Koppara, Luca Kleineidam, Lutz Frölich, Oliver Peters, Michael Hüll, Eckart Rüther, Jens Wiltfang, Wolfgang Maier, et al. Cerebrospinal fluid biomarkers and clinical progression in patients with subjective cognitive decline and mild cognitive impairment. *Journal of Alzheimer's Disease*, 58(3):939–950, 2017.

[43] Ronald C Petersen. Mild cognitive impairment. *CONTINUUM: lifelong Learning in Neurology*, 22(2):404–418, 2016.

[44] Osama Sabri, Marwan N Sabbagh, John Seibyl, Henryk Barthel, Hiroyasu Akatsu, Yasuomi Ouchi, Kohei Senda, Shigeo Murayama, Kenji Ishii, Masaki Takao, et al. Florbetaben pet imaging to detect amyloid beta plaques in alzheimer's disease: phase 3 study. *Alzheimer's dementia*, 11(8):964–974, 2015.

[45] Niklas Mattsson, Henrik Zetterberg, Oskar Hansson, Niels Andreasen, Lucilla Parnetti, Michael Jonsson, Sanna-Kaisa Herukka, Wiesje M Van der Flier, Marinus A Blankenstein, Michael Ewers, et al. Csf biomarkers and incipient alzheimer disease in patients with mild cognitive impairment. *Jama*, 302(4):385–393, 2009.

[46] Rik Ossenkoppele, Daniel R Schonhaut, Michael Schöll, Samuel N Lockhart, Nagehan Ayakta, Suzanne L Baker, James P O'Neil, Mustafa Janabi, Andreas Lazaris, Averill Cantwell, et al. Tau pet patterns mirror clinical and neuroanatomical variability in alzheimer's disease. *Brain*, 139(5):1551–1567, 2016.

[47] Avinash Vijayakumar and Abhishek Vijayakumar. Comparison of hippocampal volume in dementia subtypes. *International Scholarly Research Notices*, 2013(1):174524, 2013.

[48] Yue Ding, Yinxue Chu, Meng Liu, Zhenhua Ling, Shijin Wang, Xin Li, and Yunxia Li. Fully automated discrimination of alzheimer's disease using resting-state electroencephalography signals. *Quantitative Imaging in Medicine and Surgery*, 12(2):1063, 2022.

[49] Majid Torabinikjeh, Vahid Asayesh, Mahdi Dehghani, Aliakbar Kouchakzadeh, Hanie Marhamati, and Shahriar Gharibzadeh. Correlations of frontal resting-state eeg markers with mmse scores in patients with alzheimer's disease. *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, 58(1):1–7, 2022.

[50] Claudio Babiloni, Susanna Lopez, Claudio Del Percio, Giuseppe Noce, Maria Teresa Pascarelli, Roberta Lizio, Stefan J Teipel, Gabriel González-Escamilla, Hovagim Bakardjian, Nathalie George, et al. Resting-state posterior alpha rhythms are abnormal in subjective memory complaint seniors with preclinical alzheimer's neuropathology and high education level: the insight-pread study. *Neurobiology of Aging*, 90:43–59, 2020.

[51] Antonio I Triggiani, Vitoantonio Bevilacqua, Antonio Brunetti, Roberta Lizio, Giacomo Tattoli, Fabio Cassano, Andrea Soricelli, Raffaele Ferri, Flavio Nobili, Loreto

Gesualdo, et al. Classification of healthy subjects and alzheimer's disease patients with dementia from cortical sources of resting state eeg rhythms: a study using artificial neural networks. *Frontiers in neuroscience*, 10:604, 2017.

[52] Claudio Babiloni, Xianghong Arakaki, Hamed Azami, Karim Bennys, Katarzyna Blinowska, Laura Bonanni, Ana Bujan, Maria C Carrillo, Andrzej Cichocki, Jaisalmer de Frutos-Lucas, et al. Measures of resting state eeg rhythms for clinical trials in alzheimer's disease: Recommendations of an expert panel. *Alzheimer's & Dementia*, 17(9):1528–1553, 2021.

[53] Alida A Gouw, Astrid M Alsema, Betty M Tijms, Andreas Borta, Philip Scheltens, Cornelis J Stam, and Wiesje M van der Flier. Eeg spectral analysis as a putative early prognostic biomarker in nondemented, amyloid positive subjects. *Neurobiology of Aging*, 57:133–142, 2017.

[54] Christian Sandøe Musaeus, Knut Engedal, Peter Høgh, Vesna Jelic, Morten Mørup, Mala Naik, Anne-Rita Oeksengaard, Jon Snaedal, Lars-Olof Wahlund, Gunhild Waldemar, et al. Eeg theta power is an early marker of cognitive decline in dementia due to alzheimer's disease. *Journal of Alzheimer's Disease*, 64(4):1359–1371, 2018.

[55] Ashleigh F Parker, Lisa Ohlhauser, Vanessa Scarapicchia, Colette M Smart, Cassandra Szoeke, and Jodie R Gawryluk. A systematic review of neuroimaging studies comparing individuals with subjective cognitive decline to healthy controls. *Journal of Alzheimer's Disease*, (Preprint):1–23, 2022.

[56] Antonina Kouli, Kelli M Torsney, and Wei-Li Kuan. Parkinson's disease: etiology, neuropathology, and pathogenesis. *Exon Publications*, pages 3–26, 2018.

[57] Adina N MacMahon Copas, Sarah F McComish, Jean M Fletcher, and Maeve A Caldwell. The pathogenesis of parkinson's disease: a complex interplay between astrocytes, microglia, and t lymphocytes? *Frontiers in neurology*, 12:666737, 2021.

[58] K Ray Chaudhuri, Daniel G Healy, and Anthony HV Schapira. Non-motor symptoms of parkinson's disease: diagnosis and management. *The Lancet Neurology*, 5(3): 235–245, 2006.

[59] Lorraine V Kalia and Anthony E Lang. Parkinson's disease. *The Lancet*, 386(9996): 896–912, 2015.

[60] Simon JG Lewis, Roshan Cools, Trevor W Robbins, Anja Dove, Roger A Barker, and Adrian M Owen. Using executive heterogeneity to explore the nature of working memory deficits in parkinson's disease. *Neuropsychologia*, 41(6):645–654, 2003.

[61] Heiko Braak, Kelly Del Tredici, Udo Rüb, Rob AI De Vos, Ernst NH Jansen Steur, and Eva Braak. Staging of brain pathology related to sporadic parkinson's disease. *Neurobiology of aging*, 24(2):197–211, 2003.

[62] EV Evarts, H Teräväinen, and DB Calne. Reaction time in parkinson's disease. *Brain: a journal of neurology*, 104(Pt 1):167–186, 1981.

[63] Andrea Antal, Szabolcs Kéri, György Dibó, György Benedek, Zoltán Janka, László Vécsei, and Ivan Bodis-Wollner. Electrophysiological correlates of visual categorization: evidence for cognitive dysfunctions in early parkinson's disease. *Cognitive brain research*, 13(2):153–158, 2002.

[64] Christina Schmiedt, Anette Meistrowitz, Günter Schwendemann, Manfred Herrmann, and Canan Basar-Eroglu. Theta and alpha oscillations reflect differences in memory strategy and visual discrimination performance in patients with parkinson's disease. *Neuroscience letters*, 388(3):138–143, 2005.

[65] Juliana Dushanova, Dolja Philipova, and Gloria Nikolova. Event-related desynchronization/synchronization during discrimination task conditions in patients with parkinson's disease. *Cellular and molecular neurobiology*, 29:971–980, 2009.

[66] Ernst Niedermeyer. *Niedermeyer's electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2011.

[67] Fernando Lopes Da Silva. Eeg: origin and measurement. In *EEG-fMRI: physiological basis, technique, and applications*, pages 23–48. Springer, 2023.

[68] Fabrizio Vecchio, Claudio Babiloni, Roberta Lizio, Fabrizio De Vico Fallani, Katarzyna Blinowska, Giulio Verrienti, Giovanni Frisoni, and Paolo M Rossini. Resting state cortical eeg rhythms in alzheimer's disease: toward eeg markers for clinical applications: a review. *Supplements to Clinical neurophysiology*, 62:223–236, 2013.

[69] MX Cohen. *Analyzing neural time series data: theory and practice*. MIT press, 2014.

[70] Francisco J Fraga, Leonardo A Ferreira, Tiago H Falk, Erin Johns, and Natalie D Phillips. Event-related synchronisation responses to n-back memory tasks discriminate between healthy ageing, mild cognitive impairment, and mild alzheimer's disease. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 964–968. IEEE, 2017.

[71] Leif Sörnmo and Pablo Laguna. *Bioelectrical signal processing in cardiac and neurological applications*. Academic press, 2005.

[72] F Lopes Da Silva. Eeg analysis: theory and practice. *Electroencephalography: basic principles, clinical applications and related fields*, pages 1125–1159, 1999.

[73] Shayan Motamedi-Fakhr, Mohamed Moshrefi-Torbati, Martyn Hill, Catherine M Hill, and Paul R White. Signal processing techniques applied to human sleep eeg signals—a review. *Biomedical Signal Processing and Control*, 10:21–33, 2014.

[74] Erkki Oja, Juha Karhunen, Harri Valpola, Jaakko Särelä, Mika Inki, Antti Honkela, Alexander Ilin, Karthikesh Raju, Tapani Ristaniemi, and Ella Bingham. Independent component analysis and blind source separation. *Helsinki Univ. Technol., Espoo, Finland, Tech. Rep*, 2003.

[75] Nisreen S Amer and Samir Brahim Belhaouari. Eeg signal processing for medical diagnosis, healthcare, and monitoring: A comprehensive review. *IEEE Access*, 2023.

[76] Martin Längkvist, Lars Karlsson, and Amy Loutfi. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern recognition letters*, 42: 11–24, 2014.

[77] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *International Conference on Learning Representations*, 2015.

[78] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*, pages 3104–3112, 2014.

[79] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *Proceedings of Workshop at ICLR*, 2013.

[80] Geoffrey Hinton, Li Deng, Dong Yu, George E. Dahl, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[81] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33(4):917–963, 2019.

[82] R. Qiao, C. Qing, T. Zhang, X. Xing, and X. Xu. A novel deep-learning based framework for multi-subject emotion recognition. In *2017 4th International Conference on Information, Cybernetics and Computational Social Systems (ICCSS)*, pages 181–185, 2017. doi: 10.1109/ICCSS.2017.8091408.

[83] H. Xu and K. N. Plataniotis. Affective states classification using eeg and semi-supervised deep learning approaches. In *2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6, 2016. doi: 10.1109/MMSP.2016.7813351.

[84] A. Vilamala, K. H. Madsen, and L. K. Hansen. Deep convolutional neural networks for interpretable analysis of eeg sleep stage scoring. *CoRR*, abs/1710.00633, 2017. URL http://arxiv.org/abs/1710.00633.

[85] Y. R. Tabar and U. Halici. A novel deep learning approach for classification of eeg motor imagery signals. *Journal of Neural Engineering*, 14(1):016003, 2017. doi: 10.1088/1741-2560/14/1/016003.

[86] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, 16(5):051001, 2019.

[87] Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. Deep learning for electroencephalogram (eeg) classification tasks: a review. *Journal of neural engineering*, 16(3):031001, 2019.

[88] Yongling Xu, Yang Du, Ling Li, Honghao Lai, Jing Zou, Tianying Zhou, Lushan Xiao, Li Liu, and Pengcheng Ma. Amdet: Attention based multiple dimensions eeg transformer for emotion recognition. *IEEE Transactions on Affective Computing*, 2023.

[89] Chao Jiang, Yingying Dai, Yunheng Ding, Xi Chen, Yingjie Li, and Yingying Tang. Tsann-tg: Temporal–spatial attention neural networks with task-specific graph for eeg emotion recognition. *Brain Sciences*, 14(5):516, 2024.

[90] Dzmitry Bahdanau. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[91] Minh-Thang Luong. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.

[92] Kelvin Xu. Show, attend and tell: Neural image caption generation with visual attention. *arXiv preprint arXiv:1502.03044*, 2015.

[93] S Hochreiter. Long short-term memory. *Neural Computation MIT-Press*, 1997.

[94] Kyunghyun Cho. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

[95] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[96] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine*, 13(3):55–75, 2018.

[97] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[98] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina,

Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58: 82–115, 2020.

[99] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019.

[100] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[101] M Scott, Lee Su-In, et al. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30:4765–4774, 2017.

[102] Xinliang Zhou, Chenyu Liu, Liming Zhai, Ziyu Jia, Cuntai Guan, and Yang Liu. Interpretable and robust ai in eeg systems: A survey. *arXiv preprint arXiv:2304.10755*, 2023.

[103] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

[104] Leilani H Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pages 80–89. IEEE, 2018.

[105] Jesse Vig. A multiscale visualization of attention in the transformer model. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 37–42, 2019.

[106] Carole Dufouil, Rebecca Fuhrer, and Annick Alpérovitch. Subjective cognitive complaints and cognitive decline: consequence or predictor? the epidemiology of vascular aging study. *Journal of the American Geriatrics Society*, 53(4):616–621, 2005.

[107] Lorena Rami, Juan Fortea, Beatriz Bosch, Cristina Solé-Padullés, Albert Lladó, Alex Iranzo, Raquel Sánchez-Valle, and Jose Luis Molinuevo. Cerebrospinal fluid biomarkers and memory present distinct associations along the continuum from healthy subjects to ad patients. *Journal of Alzheimer's Disease*, 23(2):319–326, 2011.

[108] Rebecca E Amariglio, J Alex Becker, Jeremy Carmasin, Lauren P Wadsworth, Natacha Lorius, Caroline Sullivan, Jacqueline E Maye, Christopher Gidicsin, Lesley C Pepin, Reisa A Sperling, et al. Subjective cognitive complaints and amyloid burden in cognitively normal older individuals. *Neuropsychologia*, 50(12):2880–2886, 2012.

[109] Yu Sun, Fu-Chi Yang, Ching-Po Lin, and Ying Han. Biochemical and neuroimaging studies in subjective cognitive decline: progress and perspectives. *CNS Neuroscience & Therapeutics*, 21(10):768–775, 2015.

[110] Haifeng Chen, Weikai Li, Xiaoning Sheng, Qing Ye, Hui Zhao, Yun Xu, and Feng Bai. Machine learning based on the multimodal connectome can predict the preclinical stage of alzheimer's disease: a preliminary study. *European Radiology*, 32(1):448–459, 2022.

[111] Tiantian Liu, Yonghao Wang, Tianyi Yan, Yunlei Liu, Rong Xu, Jiancheng Li, and Yunyan Xie. Preclinical stages of alzheimer's disease classification by a rs-fmri study. In *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–6. IEEE, 2018.

[112] Weijie Huang, Xuanyu Li, Xin Li, Guixia Kang, Ying Han, and Ni Shu. Combined support vector machine classifier and brain structural network features for the individual classification of amnestic mild cognitive impairment and subjective cognitive decline patients. *Frontiers in aging neuroscience*, 13, 2021.

[113] Saima Rathore, Mohamad Habes, Muhammad Aksam Iftikhar, Amanda Shacklett, and Christos Davatzikos. A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages. *NeuroImage*, 155:530–548, 2017.

[114] Ioulietta Lazarou, Kostas Georgiadis, Spiros Nikolopoulos, Vangelis P Oikonomou, Anthoula Tsolaki, Ioannis Kompatsiaris, Magda Tsolaki, and Dimitris Kugiumtzis. A novel connectome-based electrophysiological study of subjective cognitive decline related to alzheimer's disease by using resting-state high-density eeg egi ges 300. *Brain Sciences*, 10(6):392, 2020.

[115] Ling Yue, Tao Wang, Jingyi Wang, Guanjun Li, Jinghua Wang, Xia Li, Wei Li, Mingxing Hu, and Shifu Xiao. Asymmetry of hippocampus and amygdala defect in subjective cognitive decline among the community dwelling chinese. *Frontiers in psychiatry*, 9:226, 2018.

[116] Aojie Li, Ling Yue, Shifu Xiao, and Manhua Liu. Cognitive function assessment and prediction for subjective cognitive decline and mild cognitive impairment. *Brain imaging and behavior*, 16(2):645–658, 2022.

[117] Elliz P Scheijbeler, Anne M van Nifterick, Cornelis J Stam, Arjan Hillebrand, Alida A Gouw, and Willem de Haan. Network-level permutation entropy of resting-state meg recordings: A novel biomarker for early-stage alzheimer's disease? *Network Neuroscience*, 6(2):382–400, 2022.

[118] Serafettin Gunes, Yumi Aizawa, Takuma Sugashi, Masahiro Sugimoto, and Pedro Pereira Rodrigues. Biomarkers for alzheimer's disease in the current state: A narrative review. *International Journal of Molecular Sciences*, 23(9):4962, 2022.

[119] Eduardo Perez-Valero, Miguel Ángel Lopez-Gordo, Christian Morillas Gutiérrez, Ismael Carrera-Muñoz, and Rosa M Vílchez-Carrillo. A self-driven approach for multi-class discrimination in alzheimer's disease based on wearable eeg. *Computer Methods and Programs in Biomedicine*, 220:106841, 2022.

[120] Ashik Mostafa Alvi, Siuly Siuly, Hua Wang, Kate Wang, and Frank Whittaker. A deep learning based framework for diagnosis of mild cognitive impairment. *Knowledge-Based Systems*, 248:108815, 2022.

[121] Saman Fouladi, Ali A Safaei, Nadia Mammone, Foad Ghaderi, and MJ Ebadi. Efficient deep neural networks for classification of alzheimer's disease and mild cognitive impairment from scalp eeg recordings. *Cognitive Computation*, pages 1–22, 2022.

[122] Knut Engedal, Maria Lage Barca, Peter Høgh, Birgitte Bo Andersen, Nanna Winther Dombernowsky, Mala Naik, Thorkell Eli Gudmundsson, Anne-Rita Øksengaard, Lars-Olof Wahlund, and Jon Snaedal. The power of eeg to predict conversion from mild cognitive impairment and subjective cognitive decline to dementia. *Dementia and Geriatric Cognitive Disorders*, 49(1):38–47, 2020.

[123] Majd Abazid, Nesma Houmani, Jérôme Boudy, Bernadette Dorizzi, Jean Mariani, and Kiyoka Kinugawa. A comparative study of functional connectivity measures for brain network analysis in the context of ad detection with eeg. *Entropy*, 23(11):1553, 2021.

[124] Salvatore Mazzeo, Michael Lassi, Sonia Padiglioni, Alberto Arturo Vergani, Valentina Moschini, Maenia Scarpino, Giulia Giacomucci, Rachele Burali, Carmen Morinelli, Carlo Fabbiani, et al. Predicting the evolution of subjective cognitive decline to alzheimer's disease with machine learning: the preview study protocol. *BMC Neurology*, 23(1):300, 2023. ISSN 1471-2377. doi: 10.1186/s12883-023-03347-8.

[125] Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyung-Min Su, and Kay A Robbins. The prep pipeline: standardized preprocessing for large-scale eeg analysis. *Frontiers in neuroinformatics*, 9:16, 2015.

[126] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197, 2019.

[127] Elena Sibilano, Michael Lassi, Alberto Mazzoni, Vitoantonio Bevilacqua, and Antonio Brunetti. A deep learning framework for the classification of pre-prodromal and prodromal alzheimer's disease using resting-state eeg signals. In *Applications of Artificial Intelligence and Neural Systems to Data Science*, pages 93–101. Springer, 2023.

[128] Francesco Carlo Morabito, Maurizio Campolo, Cosimo Ieracitano, Javad Mohammad Ebadi, Lilla Bonanno, Alessia Bramanti, Simona Desalvo, Nadia Mammone, and Placido Bramanti. Deep convolutional neural networks for classification of mild

cognitive impaired and alzheimer's disease patients from scalp eeg recordings. In *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, pages 1–6. IEEE, 2016.

[129] Cosimo Ieracitano, Nadia Mammone, Alessia Bramanti, Amir Hussain, and Francesco C Morabito. A convolutional neural network approach for classification of dementia stages based on 2d-spectral representation of eeg recordings. *Neurocomputing*, 323:96–107, 2019.

[130] Donghyeon Kim and Kiseon Kim. Detection of early stage alzheimer's disease using eeg relative power with deep neural network. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 352–355. IEEE, 2018.

[131] Cameron J Huggins, Javier Escudero, Mario A Parra, Brian Scally, Renato Anghinah, Amanda Vitória Lacerda De Araújo, Luis F Basile, and Daniel Abasolo. Deep learning of resting-state electroencephalogram signals for three-class classification of alzheimer's disease, mild cognitive impairment and healthy ageing. *Journal of Neural Engineering*, 18(4):046087, 2021.

[132] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[133] Elena Sibilano, Antonio Brunetti, Domenico Buongiorno, Michael Lassi, Antonello Grippo, Valentina Bessi, Silvestro Micera, Alberto Mazzoni, and Vitoantonio Bevilacqua. An attention-based deep learning approach for the classification of subjective cognitive decline and mild cognitive impairment using resting-state eeg. *Journal of Neural Engineering*, 20(1):016048, 2023.

[134] Raffaele Ferri, Claudio Babiloni, Vania Karami, Antonio Ivano Triggiani, Filippo Carducci, Giuseppe Noce, Roberta Lizio, Maria T Pascarelli, Andrea Soricelli, Francesco Amenta, et al. Stacked autoencoders as new models for an accurate alzheimer's disease classification support using resting-state eeg and mri measurements. *Clinical Neurophysiology*, 132(1):232–245, 2021.

[135] Emma M Whitham, Kenneth J Pope, Sean P Fitzgibbon, Trent Lewis, C Richard Clark, Stephen Loveless, Marita Broberg, Angus Wallace, Dylan DeLosAngeles, Peter Lillie, et al. Scalp electrical recording during paralysis: quantitative evidence that eeg frequencies above 20 hz are contaminated by emg. *Clinical neurophysiology*, 118(8): 1877–1888, 2007.

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

[137] Yonghao Song, Xueyu Jia, Lie Yang, and Longhan Xie. Transformer-based spatial-temporal feature learning for eeg decoding. *arXiv preprint arXiv:2106.11170*, 2021.

[138] Md Rahman, Mohammad Shorif Uddin, Mohiuddin Ahmad, et al. Modeling and classification of voluntary and imagery movements for brain–computer interface from fnir and eeg signals through convolutional neural network. *Health Information Science and Systems*, 7(1):1–22, 2019.

[139] MKM Rahman and Md A Mannan Joadder. A space-frequency localized approach of spatial filtering for motor imagery classification. *Health Information Science and Systems*, 8(1):1–8, 2020.

[140] Amir H Meghdadi, Marija Stevanović Karić, Marissa McConnell, Greg Rupp, Christian Richard, Joanne Hamilton, David Salat, and Chris Berka. Resting state eeg biomarkers of cognitive decline associated with alzheimer's disease and mild cognitive impairment. *PloS one*, 16(2):e0244180, 2021.

[141] Miguel Arevalillo-Herráez, Maximo Cobos, Sandra Roger, and Miguel García-Pineda. Combining inter-subject modeling with a subject-based data transformation to improve affect recognition from eeg signals. *Sensors*, 19(13):2999, 2019.

[142] Jayawant N Mandrekar. Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, 5(9):1315–1316, 2010.

[143] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420, 2017.

[144] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

[145] Thorir Mar Ingolfsson, Michael Hersche, Xiaying Wang, Nobuaki Kobayashi, Lukas Cavigelli, and Luca Benini. Eeg-tcnet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2958–2965. IEEE, 2020.

[146] Jinhee Park, Sehyeon Jang, Jeonghwan Gwak, Byeong C Kim, Jang Jae Lee, Kyu Yeong Choi, Kun Ho Lee, Sung Chan Jun, Gil-Jin Jang, and Sangtae Ahn. Individualized diagnosis of preclinical alzheimer's disease using deep neural networks. *Expert Systems with Applications*, 210:118511, 2022.

[147] Jianing Wei, Wendong Xiao, Sen Zhang, and Pengyun Wang. Mild cognitive impairment classification convolutional neural network with attention mechanism. In *2020 IEEE 16th International Conference on Control & Automation (ICCA)*, pages 1074–1078. IEEE, 2020.

[148] Claudio Babiloni, Pieter Jelle Visser, Giovanni Frisoni, Peter Paul De Deyn, Lorena Bresciani, Vesna Jelic, Guy Nagels, Guido Rodriguez, Paolo M Rossini, Fabrizio Vecchio, et al. Cortical sources of resting eeg rhythms in mild cognitive impairment and subjective memory complaint. *Neurobiology of Aging*, 31(10):1787–1798, 2010.

[149] Davoud Gholamiangonabadi, Nikita Kiselov, and Katarina Grolinger. Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection. *IEEE Access*, 8:133982–133994, 2020.

[150] Xia Hu, Lingyang Chu, Jian Pei, Weiqing Liu, and Jiang Bian. Model complexity of deep learning: A survey. *Knowledge and Information Systems*, 63(10):2585–2619, 2021.

[151] Jathurshan Pradeepkumar, Mithunjha Anandakumar, Vinith Kugathasan, Dhinesh Suntharalingham, Simon L Kappel, Anjula C De Silva, and Chamira US Edussooriya. Towards interpretable sleep stage classification using cross-modal transformers. *arXiv preprint arXiv:2208.06991*, 2022.

[152] Jin Xie, Jie Zhang, Jiayao Sun, Zheng Ma, Liuni Qin, Guanglin Li, Huihui Zhou, and Yang Zhan. A transformer-based approach combining deep learning network and spatial-temporal information for raw eeg classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:2126–2136, 2022.

[153] V Jahmunah, Eddie Yin Kwee Ng, Ru-San Tan, Shu Lih Oh, and U Rajendra Acharya. Explainable detection of myocardial infarction using deep learning models with grad-cam technique on ecg signals. *Computers in Biology and Medicine*, 146:105550, 2022.

[154] Yurong Li, Hao Yang, Jixiang Li, Dongyi Chen, and Min Du. Eeg-based intention recognition with deep recurrent-convolution neural network: Performance and channel selection by grad-cam. *Neurocomputing*, 415:225–233, 2020.

[155] Alvaro Fernandez-Quilez. Deep learning in radiology: ethics of data and on the value of algorithm transparency, interpretability and explainability. *AI and Ethics*, 3(1): 257–265, 2023.

[156] Xuhong Li, Haoyi Xiong, Xingjian Li, Xuanyu Wu, Xiao Zhang, Ji Liu, Jiang Bian, and Dejing Dou. Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. *Knowledge and Information Systems*, 64(12):3197–3234, 2022.

[157] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

[158] Hila Chefer, Shir Gur, and Lior Wolf. Transformer interpretability beyond attention visualization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 782–791, 2021.

[159] Jesse Vig and Yonatan Belinkov. Analyzing the structure of attention in a transformer language model. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 63–76, 2019.

[160] Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D Manning. What does bert look at? an analysis of bert's attention. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 276–286, 2019.

[161] Leonid Schwenke and Martin Atzmueller. Show me what you're looking for: visualizing abstracted transformer attention for enhancing their local interpretability on time series data. In *The International FLAIRS Conference Proceedings*, volume 34, 2021.

[162] Elena Sibilano, Domenico Buongiorno, Michael Lassi, Antonello Grippo, Valentina Bessi, Sandro Sorbi, Alberto Mazzoni, Vitoantonio Bevilacqua, and Antonio Brunetti. Understanding the role of self-attention in a transformer model for the discrimination of scd from mci using resting-state eeg. *IEEE Journal of Biomedical and Health Informatics*, 2024.

[163] Yan Pan and Yuanzhi Li. Toward understanding why adam converges faster than sgd for transformers. *arXiv preprint arXiv:2306.00204*, 2023.

[164] Eric Maris and Robert Oostenveld. Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods*, 164(1):177–190, 2007.

[165] Huy Phan, Kaare Mikkelsen, Oliver Y Chén, Philipp Koch, Alfred Mertins, and Maarten De Vos. Sleeptransformer: Automatic sleep staging with interpretability and uncertainty quantification. *IEEE Transactions on Biomedical Engineering*, 69(8): 2456–2467, 2022.

[166] Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022.

[167] Chen-Chen Fan, Hongjun Yang, Zeng-Guang Hou, Zhen-Liang Ni, Sheng Chen, and Zhijie Fang. Bilinear neural network with 3-d attention for brain decoding of motor imagery movements from the human eeg. *Cognitive Neurodynamics*, 15:181–189, 2021.

[168] Michael Lassi, Carlo Fabbiani, Salvatore Mazzeo, Rachele Burali, Alberto Arturo Vergani, Giulia Giacomucci, Valentina Moschini, Carmen Morinelli, Filippo Emiliani, Maenia Scarpino, et al. Degradation of eeg microstates patterns in subjective cognitive decline and mild cognitive impairment: Early biomarkers along the alzheimer's disease continuum? *NeuroImage: Clinical*, 38:103407, 2023.

[169] Liyuan Liu, Jialu Liu, and Jiawei Han. Multi-head or single-head? an empirical comparison for transformer training. *arXiv preprint arXiv:2106.09650*, 2021.

[170] Emadeldeen Eldele, Zhenghua Chen, Chengyu Liu, Min Wu, Chee-Keong Kwoh, Xiaoli Li, and Cuntai Guan. An attention-based deep learning approach for sleep stage classification with single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:809–818, 2021.

[171] Jinhwan Park and Wonyong Sung. Effect of adding positional information on convolutional neural networks for end-to-end speech recognition. In *INTERSPEECH*, pages 46–50, 2020.

[172] Adi Haviv, Ori Ram, Ofir Press, Peter Izsak, and Omer Levy. Transformer language models without positional encodings still learn positional information. *arXiv preprint arXiv:2203.16634*, 2022.

[173] Wei Wei, Zanbo Wang, Xianling Mao, Guangyou Zhou, Pan Zhou, and Sheng Jiang. Position-aware self-attention based neural sequence labeling. *Pattern Recognition*, 110:107636, 2021.

[174] Vimbi Viswan, Noushath Shaffi, Mufti Mahmud, Karthikeyan Subramanian, and Faizal Hajamohideen. Explainable artificial intelligence in alzheimer's disease classification: A systematic review. *Cognitive Computation*, 16(1):1–44, 2024.

[175] Renjie Liu and Zaijun Wang. Assigning channel weights using an attention mechanism: an eeg interpolation algorithm. *Frontiers in Neuroscience*, 17:1251677, 2023.

[176] Laila Craighero and Stefano Mele. Equal kinematics and visual context but different purposes: Observer's moral rules modulate motor resonance. *Cortex*, 104:1–11, 2018.

[177] Giacomo Rizzolatti and Laila Craighero. The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192, 2004. doi: 10.1146/annurev.neuro.27.070203.144230.

[178] Giulia Bommarito, Martina Putzolu, Laura Avanzino, Carola Cosentino, Alessandro Botta, Roberta Marchese, Matilde Inglese, and Elisa Pelosin. Functional correlates of action observation of gait in patients with parkinson's disease. *Neural Plasticity*, 2020 (1):8869201, 2020.

[179] Ioannis Giannakopoulos, Paraskevi Karanika, Charalambos Papaxanthis, and Panagiotis Tsaklis. The effects of action observation therapy as a rehabilitation tool in parkinson's disease patients: A systematic review. *International Journal of Environmental Research and Public Health*, 19(6):3311, 2022.

[180] Daniele Caligiore, Rick CW Helmich, Mark Hallett, Ahmed A Moustafa, Lars Timmermann, Ivan Toni, and Gianluca Baldassarre. Parkinson's disease as a system-level disorder. *NPJ Parkinson's Disease*, 3(1):1–9, 2017.

[181] Leonardo Marinelli, Angelo Quartarone, Mark Hallett, John Rothwell, and Mario Manto. Action observation treatment for rehabilitation in parkinson's disease: a pilot study. *Frontiers in Neurology*, 10:1089, 2019.

[182] Laila Craighero et al. Temporal prediction of touch instant during observation of human and robot grasping. *Brain Research Bulletin*, 75:770–774, 2008.

[183] Elena Gentile et al. Movement observation activates motor cortex in fibromyalgia patients: A fnirs study. *Scientific Reports*, 12(1):1–14, 2022.

[184] Laila Craighero and Valentina Zorzi. Hand–foot motor priming in the presence of temporary inability to use hands. *Visual cognition*, 20(1):77–93, 2012.

[185] Laila Craighero, Sonia Mele, and Valentina Zorzi. An object-identity probability cueing paradigm during grasping observation: The facilitating effect is present only when the observed kinematics is suitable for the cued object. *Frontiers in Psychology*, 6:1479, 2015.

[186] Eleonora Gentile, Antonio Brunetti, Katia Ricci, Eleonora Vecchio, Carlo Santoro, Elena Sibilano, Vitoantonio Bevilacqua, Giovanni Iliceto, Laila Craighero, and Marina de Tommaso. Effects of movement congruence on motor resonance in early parkinson's disease. *Scientific Reports*, 13(1):14887, 2023.

[187] Silvia Lahuerta-Martín, Rocío Llamas-Ramos, and Inés Llamas-Ramos. Effectiveness of therapies based on mirror neuron system to treat gait in patients with parkinson's disease—a systematic review. *Journal of Clinical Medicine*, 11(14):4236, 2022.

[188] Giacomo Rizzolatti, Maddalena Fabbri-Destro, Arturo Nuara, Roberto Gatti, and Pietro Avanzini. The role of mirror mechanism in the recovery, maintenance, and acquisition of motor abilities. *Neuroscience & Biobehavioral Reviews*, 127:404–423, 2021.

[189] Nathan A Fox, Marian J Bakermans-Kranenburg, Kathryn H Yoo, Lindsay C Bowman, Erin N Cannon, Ross E Vanderwert, Pier F Ferrari, and Marinus H Van IJzendoorn. Assessing human mirror activity with eeg mu rhythm: A meta-analysis. *Psychological bulletin*, 142(3):291, 2016.

[190] Arturo Nuara, Maria Chiara Bazzini, Pasquale Cardellicchio, Emilia Scalona, Doriana De Marco, Giacomo Rizzolatti, Maddalena Fabbri-Destro, and Pietro Avanzini. The value of corticospinal excitability and intracortical inhibition in predicting motor skill improvement driven by action observation. *Neuroimage*, 266:119825, 2023.

[191] Jun Cao, Yifan Zhao, Xiaocai Shan, Hua-liang Wei, Yuzhu Guo, Liangyu Chen, John Ahmet Erkoyuncu, and Ptolemaios Georgios Sarrigiannis. Brain functional and effective connectivity based on electroencephalography recordings: A review. *Human brain mapping*, 43(2):860–879, 2022.

[192] Andre C Marreiros, Klaas Enno Stephan, and Karl J Friston. Dynamic causal modeling. *Scholarpedia*, 5(7):9568, 2010.

[193] Karl J Friston, Lee Harrison, and Will Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, 2003.

[194] Jean-Philippe Lachaux, Sylvain Rheims, Benoit Chatard, Maryne Dupin, and Olivier Bertrand. Human intracranial database (release-5), 2023. URL https://search.kg.ebrains.eu/instances/bb13d2d1-4609-4790-a20b-678836ad486f.

[195] Olivier David, Stefan J Kiebel, Lee M Harrison, Jérémie Mattout, James M Kilner, and Karl J Friston. Dynamic causal modeling of evoked responses in eeg and meg. *NeuroImage*, 30(4):1255–1272, 2006.

[196] Daniel J Felleman and David C Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1(1):1–47, 1991.

[197] Ben H Jansen and Vincent G Rit. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological cybernetics*, 73(4):357–366, 1995.

[198] DA Pinotsis, S Fitzgerald, C See, A Sementsova, and AS Widge. Toward biophysical markers of depression vulnerability. *Frontiers in Psychiatry*, 13:938694, 2022.

[199] Dario Floreano, Auke Jan Ijspeert, and Stefan Schaal. Robotics and neuroscience. *Current Biology*, 24(18):R910–R920, 2014.

[200] Jenna H Chin, Kerstin S Haring, and Pilyoung Kim. Understanding the neural mechanisms of empathy toward robots to shape future applications. *Frontiers in Neurorobotics*, 17:1145989, 2023.

[201] Laura Miraglia, Cinzia Di Dio, Federico Manzi, Takayuki Kanda, Angelo Cangelosi, Shoji Itakura, Hiroshi Ishiguro, Davide Massaro, Peter Fonagy, and Antonella Marchetti. Shared knowledge in human-robot interaction (hri). *International Journal of Social Robotics*, 16(1):59–75, 2024.

[202] Anna Henschel, Ruud Hortensius, and Emily S Cross. Social cognition in the age of human–robot interaction. *Trends in Neurosciences*, 43(6):373–384, 2020.

[203] Michael Meredith. Human vomeronasal organ function: a critical review of best and worst cases. *Chemical senses*, 26(4):433–445, 2001.

[204] Timothy D Smith, Jeffrey T Laitman, and Kunwar P Bhatnagar. The shrinking anthropoid nose, the human vomeronasal organ, and the language of anatomical reduction. *The Anatomical Record*, 297(11):2196–2204, 2014.

[205] Ivan Rodriguez and Peter Mombaerts. Novel human vomeronasal receptor-like genes reveal species-specific families. *Current biology*, 12(12):R409–R411, 2002.

[206] Yoav Gilad, Carlos D Bustamante, Doron Lancet, and Svante Pääbo. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *The American Journal of Human Genetics*, 73(3):489–501, 2003.

[207] Robert E Johnston. Pheromones, the vomeronasal system, and communication: from hormonal responses to individual recognition. *Annals of the New York Academy of Sciences*, 855(1):333–348, 1998.

[208] Michael J Baum and James A Cherry. Processing by the main olfactory system of chemosignals that facilitate mammalian reproduction. *Hormones and behavior*, 68: 53–64, 2015.

[209] Bettina M Pause. Processing of body odor signals by the human brain. *Chemosensory perception*, 5:55–63, 2012.

[210] Johan N Lundström, Julie A Boyle, Robert J Zatorre, and Marilyn Jones-Gotman. Functional neuronal processing of body odors differs from that of similar common odors. *Cerebral Cortex*, 18(6):1466–1474, 2008.

[211] Sara Invitto, Soheil Keshmiri, Andrea Mazzatenta, Alberto Grasso, Daniele Romano, Fabio Bona, Masahiro Shiomi, Hidenobu Sumioka, and Hiroshi Ishiguro. Perception of social odor and gender-related differences investigated through the use of transfer entropy and embodied medium. *Frontiers in Systems Neuroscience*, 15:650528, 2021.