Explainable Deep Learning-Based EEG Analysis for Biomarker Discovery and Its Application on Depersonalisation/derealisation Disorder

By
Abbas Salami

A thesis submitted for the degree of PhD

School of Computer Science and Electronic Engineering,
University of Essex

April 2023

© Copyright by Abbas Salami, 2023
All Rights Reserved

© Copyright by Abbas Salami, 2023
All Rights Reserved

Declaration

I declare that this thesis has been composed solely by myself and has not been submitted, in whole or part, in any previous application for a degree. Except where states otherwise by reference or acknowledgement, the work presented is entirely my own. However, some sentences and illustrations in this thesis have been quoted verbatim from the following articles:

- Salami, A., Andreu-Perez, J., & Gillmeister, H. (2020). Symptoms of depersonalisation/derealisation disorder as measured by brain electrical activity: A systematic review. *Neuroscience & Biobehavioral Reviews, 118, 524-537*.
- Salami, A., Andreu-Perez, J., & Gillmeister, H. (2020, December). Towards decoding of depersonalisation disorder using EEG: A time series analysis using CDTW. In 2020 *IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 548-553)*.
- Salami, A., Andreu-Perez, J., & Gillmeister, H. (2022). EEG-ITNet: An Explainable Inception Temporal Convolutional Network for Motor Imagery Classification. *IEEE Access, 10, 36672-36685*.
- Salami, A., Andreu-Perez, J., & Gillmeister, H.. Finding Neural Correlates of Depersonalisation Disorder via Explainable CNN-based Analysis Guided by Clinical Assessment Scores. *Artificial Intelligence in Medicine (Under review)*

Abstract

Mental health disorders are typically diagnosed based on subjective reports (e.g., through questionnaires) followed by clinical interviews to evaluate self-reported symptoms. Therefore, considering the interconnected nature of psychiatric disorders, their accurate diagnosis is a real challenge without indicators of underlying physiological dysfunction. Depersonalisation/derealisation disorder (DPD) is an example of dissociative disorder characterised mainly by persistent disembodiment, detachment from the surroundings, and feeling of emotional numbness. Its underlying neural correlates have been investigated to understand and help with a more accurate and in-time diagnosis of the disorder. However, in terms of EEG studies, which hold great importance due to their convenient and inexpensive nature, the literature has often been based on hypotheses proposed by experts in the field, meaning it requires prior knowledge of the disorder. In addition, participants labelling in research experiments are often derived from the outcome of the Cambridge Depersonalisation Scale (CDS), a subjective assessment to quantify the level of depersonalisation/derealisation. As a result, I aimed to propose a novel EEG processing pipeline based on deep neural networks to discover electrophysiological DPD biomarkers. My deep learning model requires no prior knowledge or assumption of the disorder. In addition, the structure of the proposed model targets the unreliability of CDS scores by using them as prior information only to guide the unsupervised learning task in a multi-task learning scenario. I have also presented new ways of network visualisation to investigate spectral, spatial, and temporal information derived in the learning process and have provided neuroscientific evidence supporting the reliability of my results. I have also applied the visualisation approach to a novel motor imagery BCI system called EEG-ITNet to represent the future of more robust, interpretable, and high-accuracy BCI systems. The proposed EEG analytics in this thesis could also be

applied to investigate other mental disorders currently diagnosed based on clinical assessment scores.

Table of Contents

List of Tables

List of Figures

Acknowledgement

I would like to express my gratitude for the support and guidance of my supervisors, Dr Javier Andreu-Perez and Dr Helge Gillmeister, throughout my PhD. Working with Javier was an invaluable experience, whose mentorship help me develop personally and professionally as a researcher. In addition, I sincerely appreciate the valuable time, passion, and level of expertise that Helge offered me, which made my journey delightful and helped me develop a strong and confident character. Indeed, none of this would have been possible or even made sense without the relentless support, kindness, and love I have received from my beloved wife over all these years.

1 Introduction

In this chapter, I outline my motivation to pursue this research project and the line of arguments in choosing the research method. I highlight the shortcomings of the literature and the novelties of my work to address those shortcomings. The chapter also presents my contributions to the science and the structure of the whole thesis.

1.1 Motivation

Nowadays, mental health disorders are typically diagnosed based on subjective reports (e.g., through questionnaires) followed by clinical interviews to evaluate self-reported symptoms. Therefore, diagnosing mental health disorders is a real challenge without indicators of underlying physiological dysfunction. Every patient may show a range of cognitive, behavioural, and physical symptoms, which are collated to find the best-fitted diagnosis and treatment. However, literature has indicated a deeply interconnected nature of psychiatric disorders and has argued that finding predefined boundaries is unlikely [1]. As a result, a delay in accurate diagnosis or even misdiagnosis is highly probable.

My motivation to conduct this research was formed around one of the most prevalent under-diagnosed psychiatric disorders, Depersonalisation/derealisation disorder (DPD). Although studies have shown the prevalence of DPD to be around 1–2% of the population [2-4], making it as common as some well-known mental disorders such as schizophrenia and Obsessive-compulsive Disorder (OCD), it typically takes seven to 12 years on average to be accurately diagnosed. Therefore, it is a critical endeavour for clinical practice and research to identify DPD, its risks and neuroprotective factors, and find alternative ways for its faster and more accurate diagnosis. Accordingly, I decided to propose Electroencephalogram (EEG) analytics based on deep learning algorithms to investigate the underlying neural patterns in DPD and find potential electrophysiological biomarkers to help with its accurate and timely diagnosis. It is important to note that the term "biomarker" used throughout my thesis refers

to a unique, abnormal neural pattern activated by appropriate stimuli that can explain DPD symptoms. Furthermore, I intended to propose explainable visualisation approach in deep learning algorithms to take a step toward gaining the attention and trust of clinicians and practitioners in the capability of Artificial Intelligence (AI) for medical investigation and diagnosis.

1.2 Justification of Methodology

I argued that finding potential neural biomarkers can be a huge step toward understanding DPD and supporting its accurate and timely diagnosis. Among the neuroimaging techniques for DPD analysis within the central nervous system, EEG holds great promise because of its non-invasive nature, low costs, and simple setup. Besides, it has shown promising capability in discovering the underlying neural correlates of many psychological and mental disorders and is a reliable method for medical diagnosis [5]. However, research on the neural correlates of DPD using EEG is often formed around a hypothesis from an expert in the field. Then, the reliability of the potential biomarker as a discriminative factor in DPD is validated using statistical tests. In other words, finding a potential biomarker is based on trial and error and requires an expert's prior knowledge. In addition, labels of participants in experimental research as either patient or control (or high vs low symptomology) are often assigned according to a clinical questionnaire and the most common way to do so for DPD is currently based on the Cambridge Depersonalisation Scale (CDS) [6]. However, the CDS score is not always reliable, and there is sometimes disagreement between the outcome of the CDS questionnaire and the clinicians' diagnosis [7]. Besides, finding an agreed threshold by which a participant would be considered a DPD patient is challenging.

In this research, I employed deep learning algorithms to overcome the above shortcomings in the experimental literature. From several machine learning algorithms for

EEG analysis, deep learning is a reliable technique with proven outstanding results in identifying several neurological and psychological disorders [8, 9]. In addition, end-to-end deep learning architectures are particularly noteworthy since they eliminate the need for handcrafted features and prior information regarding the dataset. Accordingly, I proposed an automated biomarker discovery protocol using a novel end-to-end multi-input multi-output deep learning structure. I showed how CDS scores could help conduct (but not directly imply) the learning process. Regressing CDS scores was not the main goal of the learning process but separating the neural data between the control and patient. This way, I put less stress on the CDS scores as a subjective measure to accurately reflect the level of depersonalisation, but I took advantage of them as prior information to guide the unsupervised part of my learning process to find more reliable discriminative neural features.

1.3 Contribution to Science

Here, I summarise the highlights and novelties of my work and its outputs in the form of scientific publications.

1.3.1 Highlights and Novelties

The highlight and novelties of this work are:

a) Proposing a novel deep learning framework for EEG signal analysis to find neural patterns associated with DPD symptoms without requiring prior knowledge of the disorder. My method uses clinical assessment scores while addressing their uncertainty in the learning process.
b) Proposing new ways of network visualisation, investigating spectral, spatial, and temporal information derived in a deep learning process.

1.3.2 Outputs

My contribution to science comprises the following four peer-reviewed publications:

1. Salami, A., Andreu-Perez, J., & Gillmeister, H. (2020). Symptoms of depersonalisation/derealisation disorder as measured by brain electrical activity: A systematic review. *Neuroscience & Biobehavioral Reviews, 118, 524-537*.

   Link to the publication

   As the first step in my research, I aimed to provide a systematic review of the symptoms of DPD as measured by brain electrical activity. I comprehensively described research targeting the neural correlates of core DPD symptoms, covering publications between 1992 and 2020 that have used electrophysiological techniques. The aim was to investigate the diagnostic potential of these relatively inexpensive and convenient neuroimaging tools. I reviewed the EEG power spectrum, components of the event-related potential (ERP), as well as vestibular and heartbeat-evoked potentials as likely electrophysiological biomarkers to study DPD symptoms. I suggested tools for prospective studies of electrophysiological DPD biomarkers, which are urgently needed to fully develop their diagnostic potential.

2. Salami, A., Andreu-Perez, J., & Gillmeister, H. (2022). EEG-ITNet: An Explainable Inception Temporal Convolutional Network for Motor Imagery Classification. *IEEE Access, 10, 36672-36685*.

   Link to the publication

   As part of my investigations to develop the final explainable biomarker discovery protocol based on deep learning, I proposed the end-to-end EEG-ITNet, an explainable inception temporal convolutional network for motor imagery classification. I specifically addressed the lack of interpretability of deep learning models, which makes them not quite favoured by the neuroscience community. I

comprehensively explained and supported the validity of network illustration from a neuroscientific perspective.

3. Salami, A., Andreu-Perez, J., & Gillmeister, H.. Finding Neural Correlates of Depersonalisation Disorder via Explainable CNN-based Analysis Guided by Clinical Assessment Scores. *Artificial Intelligence in Medicine (Under review)*

   My explainable deep learning-based EEG analysis for biomarker discovery was thoroughly explained and presented in my last publication. I argued that EEG studies on finding discriminative neural factors are often based on hypotheses proposed by experts in the field, meaning it requires prior knowledge of the disorder. In addition, participant labelling in research experiments is often derived from the outcome of the CDS, while its threshold and reliability might be challenged. As a result, I aimed to propose a novel end-to-end EEG processing pipeline based on deep neural networks for DPD biomarker discovery, which requires no prior knowledge or assumption of the disorder. In addition, the multi-task learning nature of the proposed deep model targets the uncertainty in CDS scores by using them as prior information only to guide the unsupervised learning task. A comprehensive evaluation has been done to confirm the significance of my proposed deep structure, including new ways of network visualisation to investigate spectral, spatial, and temporal information derived in the learning process. I also argued that the proposed EEG analytics could also be applied to investigate other psychological and mental disorders currently indicated on the basis of clinical assessment scores.

4. Salami, A., Andreu-Perez, J., & Gillmeister, H. (2020, December). Towards decoding of depersonalisation disorder using EEG: A time series analysis using CDTW. In 2020 *IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 548-553).*

   Link to the publication

To emphasise the importance of my research, I showed that we could discriminate between DPD patients and healthy individuals using a reliable biomarker in their neural activities. I used one of the EEG biomarkers found in my analysis as a feature and performed a classification task using time series analysis. 85% accuracy (Kappa 0.7) was obtained in my investigation using leave-one-subject-out cross-validation.

1.4 Structure of the Thesis

The structure of the rest of my PhD thesis is as follows:

- In chapter 2, I first introduce DPD and its symptoms and then provide the fundamentals needed to read and understand the rest of the thesis, including essential EEG acquisition, signal processing, and machine learning concepts.
- In chapter 3, I exhaustively discuss the structural and functional neurophysiology of DPD and provide a systematic review of research targeting the neural correlates of core DPD symptoms, focusing on studies aiming to characterise the cardinal symptoms of the disorder by measuring the electrical activity of the brain. The chapter also recalls the application of deep learning in neuroimaging signal analysis and introduces some state-of-the-art methods, stating their pros and cons.
- In chapter 4, I describe the employed pre-processing pipeline, followed by in detailed description of my deep learning architectures, including score-guided DPD biomarker discovery system and EEG-ITNet, an explainable neural network for motor imagery Brain-computer Interfaces (BCIs). The chapter also holds my visualisation techniques and how to explain and interpret the learning process. In addition, the chapter contains details of the two DPD datasets and two motor imagery datasets used in my PhD project. Finally, I emphasise the importance of finding a reliable biomarker to aid the DPD diagnostic process by performing a classification task using time series analysis.

- In chapter 5, I present the application of my explainable deep models in motor imagery BCI and also my results on the effectiveness of the proposed biomarker discovery system for handling uncertainty in clinical assessment scores and its ability to find reliable electrophysiological biomarkers. I evaluate each part of my analysis and network independently and also as a whole system. In addition, I provide the results of my initial investigations on the reliability of electrophysiological biomarkers in supporting DPD diagnosis.
- In chapter 6, I comprehensively discuss and validate the outputs of both networks designed as motor imagery BCI and DPD biomarker discovery system and provide neuroscientific evidence supporting my results.
- In chapter 7, a conclusion for my PhD research and a discussion on its limitation can be found. The chapter also includes recommendations for potential future work.

2 Background

In this chapter, I first introduce DPD and then explain the technical background needed to follow and understand the rest of the thesis. The chapter includes basic concepts in EEG signal acquisition and processing and some deep learning theories.

2.1 Depersonalisation/derealisation Disorder (DPD)

Depersonalisation/derealisation refers to a state of mind in which a person feels detached and disconnected from their bodies and own senses as well as from their surroundings [10]. In traumatic situations, or when the brain faces a high level of stress or anxiety, prefrontal inhibition of the limbic emotional response system serves to protect the organism from overwhelming sensations and emotions [11, 12]. As a result, individuals may experience temporary emotional numbing, disembodiment, out-of-body experiences, or a sense of unreality about the outside world [2]. For instance, temporary occurrences of depersonalisation have been reported by almost 50% of college students [15]. Fatigue [16], sleep deprivation [17], or travelling to unfamiliar places can also be the cause of transient depersonalisation/derealisation [18]. However, in cases where the symptoms are chronic, it is considered a type of dissociative disorder (Depersonalisation/derealisation disorder (DPD); Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) [19]).

Although the exact cause of DPD is not yet known, traumatic experiences and childhood anxiety are thought to be common triggers [3, 20]. It can also be provoked by intense stress, depression, panic attacks, and ingestion of psychoactive substances [21-23]. Patients experience persistent and frequent feelings of disembodiment or detachment from their physical self as well as emotional numbness that may extend beyond the present moment to include memories and imagination. Since derealisation is an inseparable aspect of DPD in most cases, the symptoms may also include detachment from surroundings, as if the world

around the patient is unreal, or a weakened ability to respond to emotional circumstances. However, the capacity for emotional expression and reality testing remains intact [24, 25].

It is important to differentiate transient depersonalisation/derealisation symptoms and the chronic type. While the former is common and often needs no intervention, the latter can profoundly affect the quality of life for patients and intervene in their daily activities and social relationships. The literature has reported DPD to be accompanied by anxiety, depression or schizophrenia [26], as well as difficulties in concentration and memory retrieval [27]. However, symptoms of depersonalisation and derealisation secondary to another medical or psychiatric diagnosis should be identified and distinguished from primary DPD to treat the underlying problem [28]. Therefore, identifying DPD, as well as its risks and neuroprotective factors, at early stages should thus be a critical endeavour for clinical practice and research.

DPD has a prevalence of about 1-2% of the population [2-4], which is comparable to that of schizophrenia and OCD, with an equal gender ratio and an average onset age in early adulthood [23, 29-31]. Nevertheless, DPD is one of the most prevalent under-diagnosed psychiatric disorders [28, 32]. Generally, there are no medical laboratory tests for the diagnosis of dissociative disorders, and since patients find it difficult to describe the symptoms of depersonalisation and derealisation, it currently takes an average of seven to 12 years to correctly diagnose DPD [28]. Diagnosis is hampered by a lack of awareness of DPD among medical practitioners [33], and its symptomatology overlap with medical conditions such as epilepsy and migraine [34, 35] and psychiatric conditions such as depression and Post-traumatic Stress Disorder (PTSD) [36]. Therefore, delineating the neurophysiological correlates of DPD may be of great importance for an early diagnosis of DPD as it discriminates it from its transient form and from other conditions.

2.1.1 Cambridge Depersonalization Scale (CDS)

CDS is a standard scale and the most well-known benchmark questionnaire to evaluate the frequency and duration of DPD symptoms, comprising 29 items to quantify the level of depersonalisation/derealisation in individuals. It consists of items such as "Parts of my body feel as if they didn't belong to me" or "My surroundings feel detached or unreal, as if there were a veil between me and the outside world". Through measuring the frequency and duration of symptoms over a recent period of six months on a combined 10-point scale, the level of depersonalisation/derealisation is then evaluated based on the overall score, which could be between 0 and 290. According to the ROC curve analysis and finding the best compromise between true positive and false positive rates in a sample of 77 subjects, Sierra et al. [6] proposed a cut-off point of 70 for the CDS score. However, there is sometimes disagreement between the outcome of the CDS questionnaire and clinicians' diagnosis [7]. In addition, some studies [37-39] have suggested a threshold of 50 on different, nonclinical datasets, challenging the utility of the clinical cut-off point for the purpose of cognitive neuroscience research.

2.2 EEG Signal Acquisition

In this section, I briefly define common terminologies in EEG signal acquisition. The datasets used in my research to investigate neural signatures of depersonalisation symptoms were the outcome of experiments designed to measure brain responses to specific events or stimuli. More specifically, they consist of visual-tactile stimulation, which provokes somatosensory responses. These types of datasets are designed to record Event-related Potentials (ERPs), which are the main focus of this section.

2.2.1 What Is EEG

EEG represents the electrical activities of a population of neurons in the human brain, which can be measured by electrodes placed on the surface of the scalp. EEG acquisition system often comprises scalp electrodes, a signal amplifier, an analogue-to-digital converter, and a transmission module to transfer the signals to a computer. The conversion rate (measured in Hz) is called sampling frequency which is defined as the number of samples per second taken from a continuous/analogue signal to make a discrete/digital signal.

2.2.2 EEG Artefacts

EEG signals can be affected and deteriorated by artefacts, which refer to any physiological, electrical, and environmental signals integrated with our desired neural signal during EEG recording. For example, artefacts could be interference from other physiological signals, such as heart rate, electrode detachment, or signal distortion resulting from the cortex's behaviour as a volume conductor [40].

2.2.3 Event-related Potential (ERP)

Event-related Potential (ERP) is a neural pattern generated by the brain in response to a specific event or stimulus [41]. In cases where the stimulus is tactile, it is referred to as Somatosensory Evoked Potential (SEP). Each ERP signal contains several components corresponding to several brain mechanisms. Temporal averaging over several trials can be used to extract and visualise the components of ERP. The idea is based on the assumption that each ERP signal contains an Additive White Gaussian Noise (AWGN). So averaging over many trials can reject noise from the signal (because AWGN has zero mean) and increase the Signal-to-noise Ratio (SNR). A sample ERP signal and its common components are depicted in **Figure 2.1**.

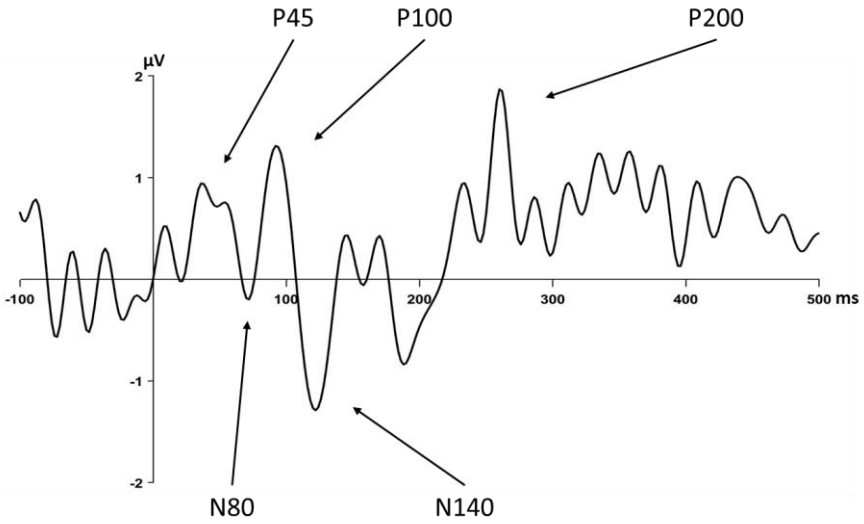

Figure 2.1 Different SEP components

2.2.3.1 Baseline Correction

In ERP signal processing, baseline correction is a critical stage that uses EEG activity over a pre-stimulus time window (also called the baseline period) to correct activity over a post-stimulus interval. The average signal amplitude over the baseline period gets subtracted from the signal following the event.

2.3 Signal Processing

Here, I introduce the main signal processing approaches I used in my PhD study to help the reader understand the following chapters better.

2.3.1 Continuous Dynamic Time Warping (CDTW)

Dynamic Time Warping (DTW) is used to find the similarity between two time series with possible delays and different speeds with respect to each other [42]. Consider two different multi-dimensional time series $A(t), t = 1, \dots, T_1$ and $B(t), t = 1, \dots, T_2$. The general idea is to find a warping function $\psi = [\psi_1(t), \psi_2(t)]^T$ between the two time series which minimises the following similarity measure:

$$D(A,B) = \sum_{t=2}^{T} d\left(\left(A\left(\psi_1(t-1)\right), B\left(\psi_2(t-1)\right)\right), \left(A\left(\psi_1(t)\right), B\left(\psi_2(t)\right)\right)\right)$$
$$= \sum_{t=2}^{T} \left\| \overrightarrow{A\left(\psi_1(t)\right)B\left(\psi_2(t)\right)} - \overrightarrow{A\left(\psi_1(t-1)\right)B\left(\psi_2(t-1)\right)} \right\|^2 \quad (2.1)$$

where $\|.\|$ is the Frobenius norm. In the warping or matching function ψ, a point $A(\psi_1(t))$ corresponds to a point $B(\psi_2(t))$ for $t = 1, \dots, T$. Minimisation of the similarity measure D can be done using dynamic programming in the so-called *a multi-stage decision process* [43] based on the following recursive formula:

$$D(t) = \min_{\psi_1(t-1)} \left\{ D(t-1) + d\left(\left(A\left(\psi_1(t-1)\right), B\left(\psi_2(t-1)\right)\right), \left(A\left(\psi_1(t)\right), B\left(\psi_2(t)\right)\right)\right)\right\} \quad (2.2)$$

Note that in general $A(t)$ and $B(t)$ can be samples of two continues time series; however, in case the two time series are discrete, $\psi_1(t)$ and $\psi_2(t)$ can only take values on $\{1, \dots, T_1\}$ and $\{1, \dots, T_2\}$ respectively. Therefore, the warping function can be derived from the similarity matrix calculated as follows:

$$D(i,j) = \| A(i) - B(j) \| + \min\{D(i-1,j), D(i,j-1), D(i-1,j-1)\} \quad (2.3)$$
$$i = 1, \dots, T_1 \; j = 1, \dots, T_2$$

Then a warping path (alignment path) is found, starting from the top right corner of the distance matrix to the bottom left corner, by selecting the cells with the minimum value in each step from the adjacent cells and keeping the slope of the path always non-negative. The similarity measure between the two time series is the sum of distances on the warping path. An example of the warping plane and the warping path is illustrated in **Figure 2.2**. In this example, the aligned signals after DTW are as follows:

$$A' = (A^1, A^1, A^1, A^1, A^2, A^2, A^2, A^3, A^4, A^5, A^6, A^7, A^7, A^8, A^9, A^{10})$$
$$B' = (B^1, B^2, B^3, B^4, B^4, B^5, B^6, B^6, B^7, B^7, B^7, B^7, B^8, B^9, B^9, B^{10})$$

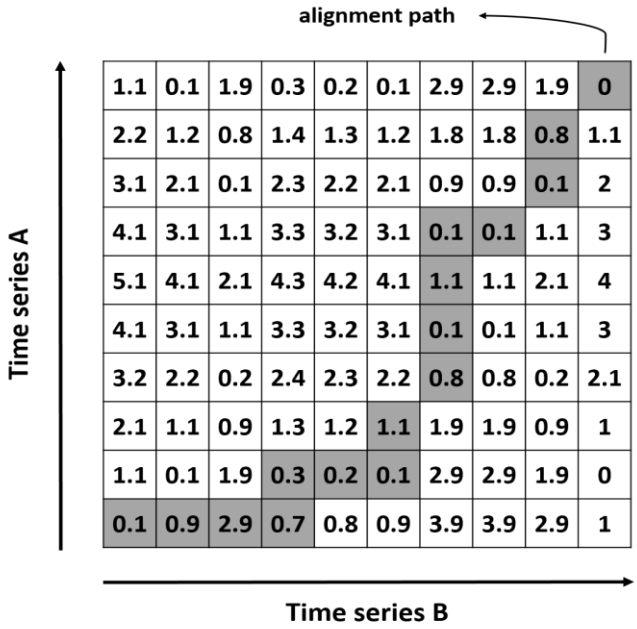

Figure 2.2 Sample illustration of the distance matrix and the alignment path in DTW

As mentioned earlier, $A(t)$ and $B(t)$ can be a subsample of two continuous time series. Continuous Dynamic Time Warping (CDTW) is a continuous version of DTW in which each point of one time series can be matched with any value between two samples of the other time series (more details can be found in [44]).

ERP signals are an example of time series, and therefore, I applied CDTW in section 4.5.2 to address the possible distortion and time shift in the ERP components. My goal was to examine the potential of using a reliable electrophysiological biomarker to distinguish between individuals with low and high levels of depersonalisation.

2.3.2 Resampling-average Method

In ERP analysis, the existing noise in the experiment is assumed to be generated from a normal distribution with zero mean. Therefore, it was mentioned that the reason behind taking the average over multiple epochs time-locked to the stimuli of interest is to cancel out the noise and random brain activities that can affect ERP components. However, in EEG studies, this approach results in a single grand-average ERP for each subject per condition,

making it impractical to train a machine learning algorithm due to insufficient data samples. To overcome this problem, the resampling-average method (an idea derived from [45]) can be used to generate multiple ERP waves for efficient training of machine learning algorithms and consequently find discriminative features among the groups of interest. The resampling-average method used in my study is portrayed in **Figure 2.3** and works as follows. For each condition in the experiment (explained in detail in chapter 4), I randomly selected a subset of *r* trials of that condition and calculated the average ERP on that subset to form a single input data for my deep learning model. To enhance the output of the resampling-average method further, the trials that dropped out for averaging were randomly selected from those with high variance in the ERP pre-stimulus (baseline) signal. This was based on the assumption that a high-quality ERP signal should have a low variance in the pre-stimulus signal. By repeating this process *n* times, we can generate *n* ERP waves. It is important to note that although the generated samples are not independent of each other, using this method can reduce the impact of outliers in the EEG dataset and generate enough data for training an effective deep neural network.

2.4 Deep Learning

Artificial Neural Network (ANN) is a subset of machine learning, which reflects the human brain's behaviour and mimics how biological neurons signal to one another. Like other machine learning algorithms, ANNs are designed to predict the outcomes without being

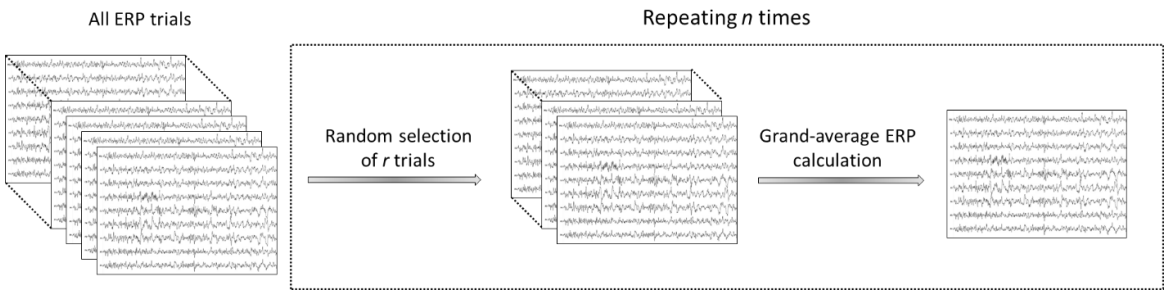

Figure 2.3 Resampling-average method to generate ERP samples.

explicitly programmed to do so by learning the hidden patterns in the available data. To explain the basic building blocks of ANN and how it functions, we can look at its elementary entity, the artificial neuron.

2.4.1 Artificial Neuron

Assume a set of *d*-dimensional input data points $X = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ where $\bar{x}_i \in \mathbb{R}^d$ and x_{ij} is the *j*-th dimension of $\bar{x}_i$. Having a weight vector $\bar{w} = [w_1, w_2, \dots, w_d]^T$ and a scalar bias term b, the output $\hat{y}_i$ of the neuron can be calculated as follows:

$$\hat{y}_i = \phi\left(\sum_{j=1}^{d} w_j x_{ij} + b\right) = \phi(\bar{w}^T \bar{x}_i + b) \tag{2.4}$$

where ϕ is the activation function. Notice that the bias term and the activation function are optional, and the model in **Figure 2.4** will represent a linear regression if no activation is applied. However, activation functions are used to empower ANNs to learn nonlinear decision boundaries and give them the flexibility to be used in different scenarios. So, for instance, the sigmoid function can be used to have a logistic regression instead of a linear one.

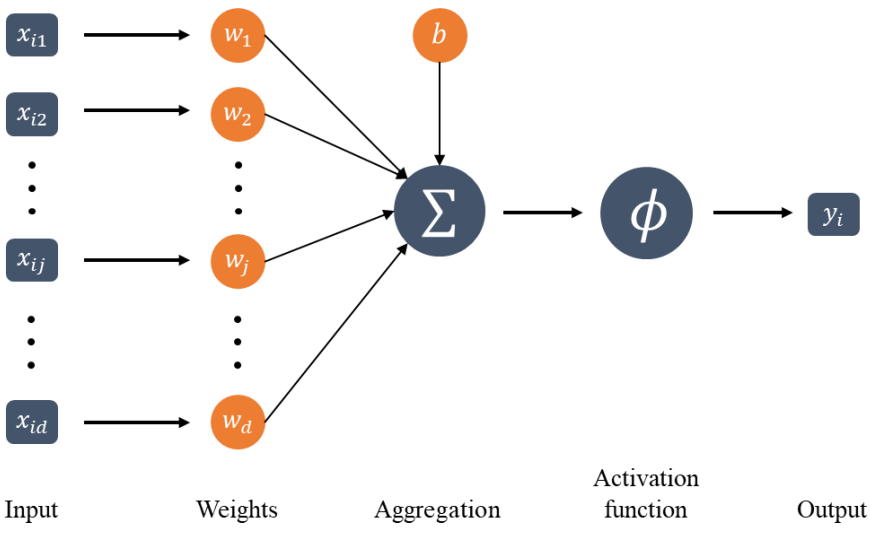

Figure 2.4 Artificial neuron

For this simple model, assume a binary classification scenario where the goal is to find the optimum values for $\overline{w}$ and b that result in an accurate prediction of the true class labels $y_i \in \{0,1\}$ over a training set $\{X, Y\}$, where $Y = \{y_1, y_2, \dots, y_n\}$ is the set of true labels. The above optimisation problem can be solved by defining an appropriate loss function. For instance, Log Loss [46], which indicates how close the prediction probability is to the corresponding true label, can be used as the loss function in this simple example. Accordingly, the cost function, which is defined as an average of a loss function over an entire training dataset, can be written as follows:

$$J(\theta) = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1 - y_i)\log\,(1 - \hat{y}_i) \tag{2.5}$$

where θ refers to the set of network parameters, including the weight vector $\overline{w}$ and the bias term b. As mentioned earlier, the goal is to solve the optimisation problem of finding the values of network parameters θ that minimise the cost function $J(\theta)$, which can be achieved using optimisation algorithms, such as the well-known iterative gradient descent [47]. Notice that the choice of loss function depends on the scenario and its possible outputs. I will later introduce the unique loss function I used in my research, which was proposed according to my goals and the nature of my dataset.

2.4.2 Convolutional Neural Network (CNN)

A multi-channel EEG signal can be seen as a 2-dimensional image, where each pixel represents the EEG sample at a single time point over a single electrode. Therefore, Convolutional Neural Networks (CNNs), a class of ANN initially developed for image recognition and tasks involving pixel data processing, can be applied to analyse EEG signals and extract information from them [48]. For that, I have taken advantage of two algorithms in the CNN literature in my research and design, including the inception module and Temporal

Convolutional Network (TCN). However, explaining the learned features and their connection with known EEG components was the challenge in my study.

The basic components of CNN are convolutional, pooling, flatten, and fully connected layers, as depicted in **Figure 2.5**. The convolution process involves a kernel inside CNN core layers moving across the receptive fields of an image. In other words, it is a process where a kernel matrix (simply a matrix of numbers) is passed over the input image and transforms the image based on the values from the filter. Using multiple kernels and concatenating them after convolution (different colours in **Figure 2.5**) along so called convolutional channel dimension, we map the input into a new (often with a higher dimension) space. During the learning process, the weight matrices are optimised to find the mapping that reveals the discriminative information more evident. A pooling layer is then applied to reduce the data dimension, which could be simply the average of adjacent pixels. The aim of the flatten layer is to reform the output of a convolutional layer into a single one-dimensional vector, converting the multi-dimensional input to a vector similar to our input data $\bar{x}_i$ explained in the artificial neuron. Note that the activation functions can be applied throughout the network, often after each matrix multiplication, to empower the network to learn nonlinear

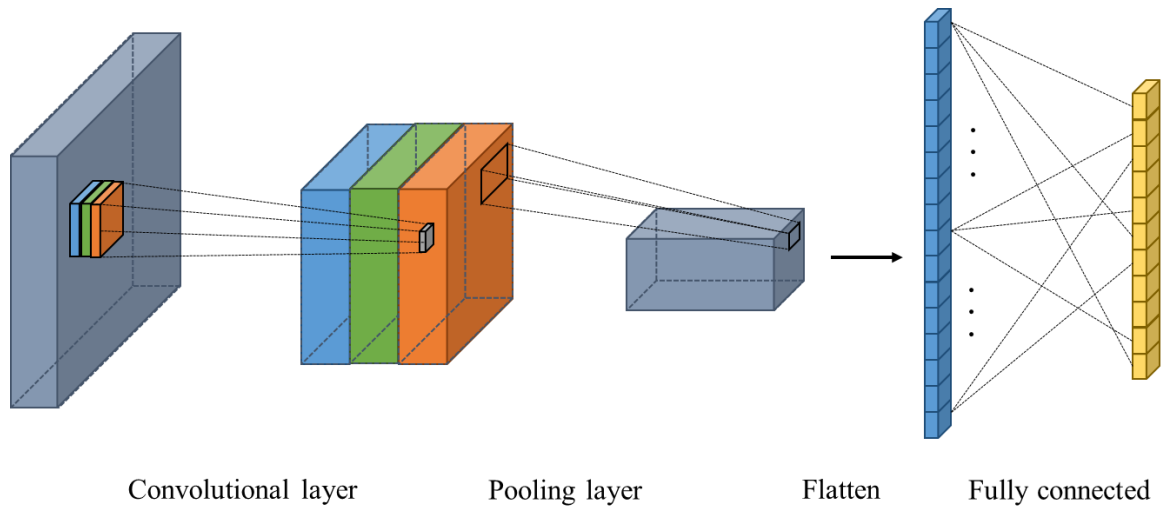

Figure 2.5 Simple CNN architechture

patterns in the data. For more information on CNN and the math behind it, please refer to [47].

2.4.2.1 Inception

The naive inception module was initially proposed in computer vision to tackle the drawbacks of using deeper, wider, and more complex networks for image classification [49]. To understand how inception works, assume a convolutional layer designed to break the multi-channel EEG signal into its informative frequency sub-bands (more details provided in chapter 4). The idea of an inception module is to use several parallel (rather than sequential) convolutional layers with various filter sizes and stack their output along the convolutional channel dimension (**Figure 2.6-b**), rather than adopting a single convolutional layer with a

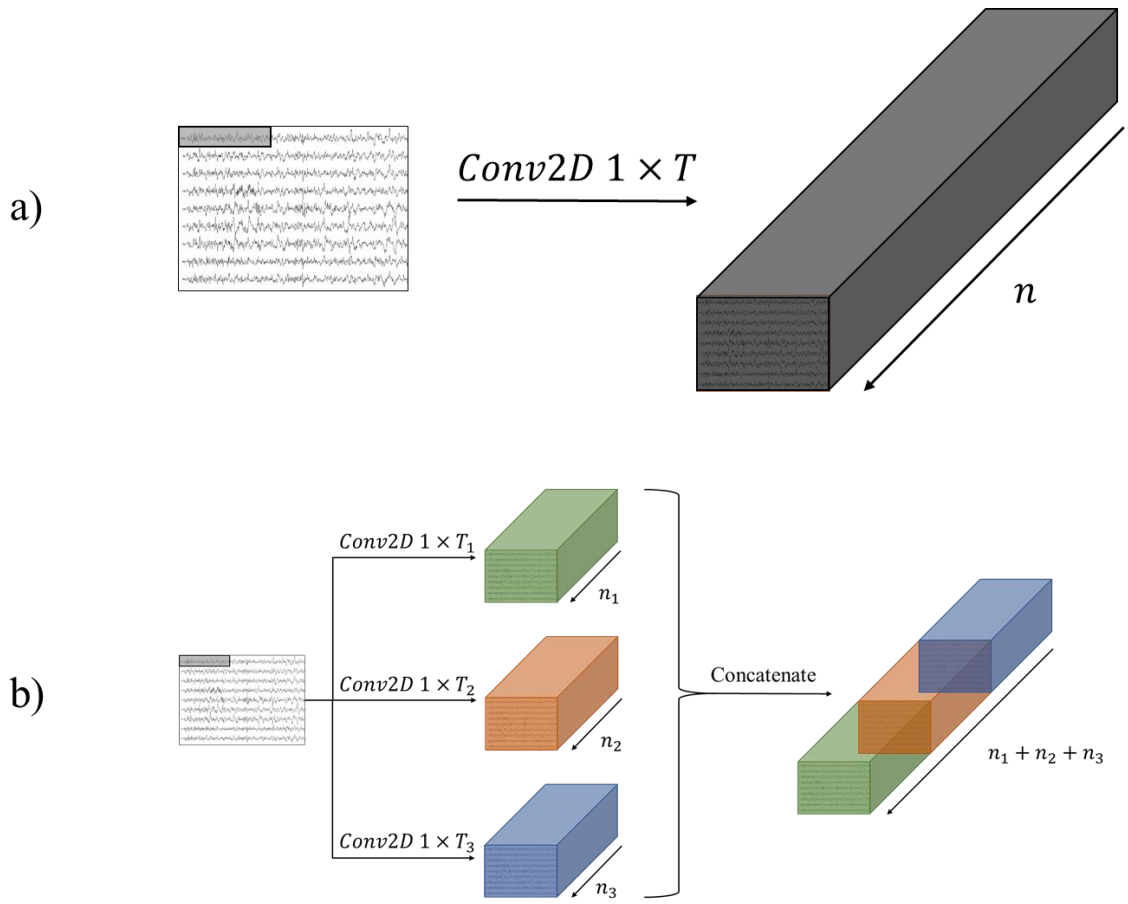

Figure 2.6 Conventional convolutional layer (a) and its alternative with inception modules (b)

fixed-length kernel (**Figure 2.6-a**). Inception modules are effective, especially in the case of small training datasets or limited computational capacity, where it is not feasible to use a deep, complex neural network with a high number of parameters. This is because the network gets either more prone to overfitting or goes without computational resources [49].

2.4.2.2 Temporal Convolutional Network (TCN)

Convolution with dilation is the foundation of TCNs initially proposed by Bai et al. [50]. Dilation refers to expanding the kernel by inserting holes between its consecutive elements. The TCN term was first coined by Lea et al. [51] and used as a new and effective class of temporal models for action segmentation and detection. Since then, TCNs have been suggested as a promising alternative to Recurrent Neural Networks (RNNs) [47] for sequence modelling by providing a unified approach to encode spatial-temporal information and capture high-level temporal information hierarchically. In the area of EEG signal analysis, TCNs have been recently applied and demonstrated promising results for temporal analysis of EEG time series with faster computation than RNNs [52, 53].

Adding dilation to a series of convolutional layers can increase the receptive field of the network. Therefore, in my design, I applied this modification of CNN consisting of a series of causal convolutional layers with varying dilation (powers of a base dilation) to take advantage of a more extended history coverage and extract rich temporal features at each time step. By causal convolutional, I mean that the output of the convolutional layer at each time step depends only on the earlier time steps of the input time series (**Figure 2.7**). A leading zero padding [47] is applied to achieve this and ensure an output sequence with equal length to the input sequence. Mathematically, if we have n convolutional layers stacked on one another with a dilation base of b and a kernel size of T, then the TCN receptive field r can be calculated as follows:

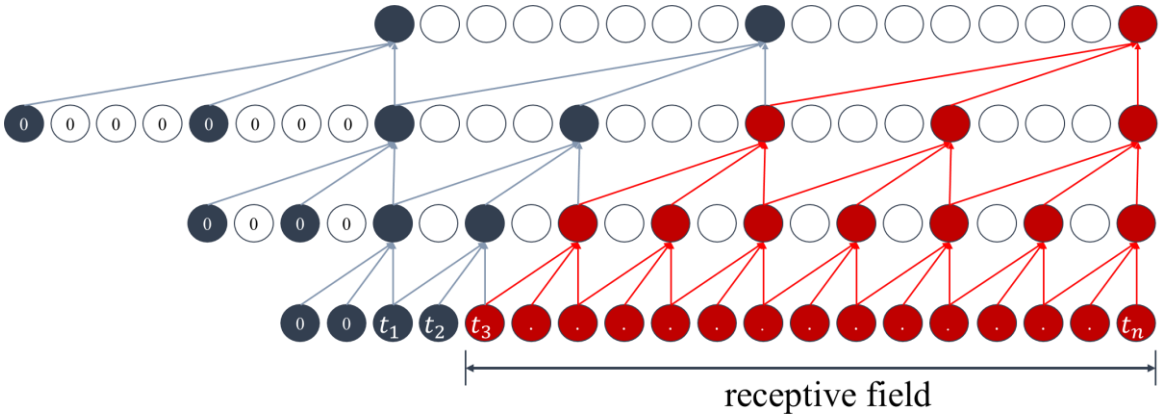

Figure 2.7 Three causal convolutional layers with leading zero padding and the corresponding receptive field. In this example, the kernel size is 3 and the dilation based is 2.

$$r = 1 + (T - 1)\frac{b^n - 1}{b - 1} \tag{2.6}$$

Notice that the kernel size T needs to be selected as a number greater than the dilation base b to avoid having holes in the receptive field. By holes, I mean samples in the input sequence that do not affect the output sequence. To further enhance the performance of TCN, its representation in the form of residual blocks was proposed [50]. To extract temporal features in my motor imagery BCI investigations, I designed a TCN as a series of residual blocks, each consisting of multiple causal convolutional layers with the same dilation. The dilation rate increases for each consecutive residual block as the powers of the dilation base b. In this case, with having n residual blocks with m convolutional layers in each, a dilation base of b, and a kernel size of T, the receptive field of the TCN will be calculated as follows:

$$r = 1 + m(T - 1)\frac{b^n - 1}{b - 1} \tag{2.7}$$

So (2.7) can be used to determine the number of residual blocks and causal convolutions in each, dilation base, and the kernel size needed to design a TCN that can account for r time steps back in time. Of course, the residual blocks also contain activation functions and dropout to account for nonlinearity and avoid overfitting the network, respectively. Batch

normalization is often applied after convolutional layers to tackle the exploding gradient problem [47].

3 Literature Review

In this chapter, I first present the outcome of my systematic review on the neural correlates of core DPD symptoms, covering publications between 1992 and 2020 that have used electrophysiological techniques. The aim was to investigate the diagnostic potential of these relatively inexpensive and convenient neuroimaging tools. I reviewed the EEG power spectrum, components of the ERP, as well as vestibular and heartbeat-evoked potentials as likely electrophysiological biomarkers to study DPD symptoms. I argue that acute anxiety- or trauma-related impairments in the integration of interoceptive and exteroceptive signals play a key role in the formation of DPD symptoms and that future research needs analysis methods that can take this integration into account. I also suggest tools for prospective studies of electrophysiological DPD biomarkers, which are urgently needed to develop their diagnostic potential fully.

In the second part of this chapter, I review the application of deep learning for EEG signal analysis, stating the pros and cons of existing architectures and their shortcomings in handling my scenario. I discuss why I need a structure that can handle clinical assessment scores as prior information rather than entirely relying on them in my multi-task learning scenario (or generally for analysing any mental disorders assessed based on clinical assessment scores). In addition, I indicate the need for a more explainable visualisation technique for the learning process, which could help to intelligibly investigate and discuss my findings from a cognitive neuroscientific point of view.

3.1 Symptoms of Depersonalisation/derealisation Disorder as Measured by Brain Electrical Activity

In this review section, I provide an overview of the structural and functional neurophysiology of DPD, with a particular focus on studies aiming to characterise the cardinal symptoms of the disorder, such as feelings of disembodiment and emotional numbing by measuring the

electrical activity of the brain. Electrophysiological neuroimaging techniques are of great interest because of their ease of application and cost-effectiveness for clinical practice. Therefore, I intend to identify and introduce electrophysiological biomarkers associated with DPD symptoms, which may potentially help with the early recognition of this under-diagnosed psychiatric condition. My review demonstrates both the urgent need to replicate promising findings on a larger scale and the potential for further electrophysiological pattern analysis to characterise DPD.

3.1.1 Introduction

Several neuroimaging studies in the literature have targeted central neural patterns and possible abnormal activities in DPD with functional Magnetic Resonance Imaging (fMRI) [10, 60], Positron Emission Tomography (PET) [61], and EEG. These studies predominantly compare the neural substrates of DPD patients with control subjects and have mainly focused on two core aspects of DPD, emotional numbing and disembodiment [62]. For instance, various fMRI studies [63-65] have investigated the neural responses of DPD patients to emotional versus neutral stimuli [66, 67]. Results illustrate that emotional numbing (the attenuation of emotional experiences as a result of inhibitory processes) in DPD is associated with reduced activity in brain areas responsible for emotional processing, particularly the insula and limbic regions, including the hypothalamus and amygdala [68]. Lemche et al. [63] showed an inverse relationship between activity in the hypothalamus and amygdala and positive and negative emotional stimuli intensity in a group of DPD patients compared with controls. fMRI studies also showed increased activation of the right ventrolateral prefrontal cortex in DPD patients exhibiting emotional numbness in response to aversive stimuli [66, 67].

Similarly, increased activation of the dorsal prefrontal cortex, which plays a role in emotional suppression [69], was found during processing both positive and negative emotional

facial expressions in DPD patients [63]. Dorsal prefrontal activation was inversely related to skin conductance levels. This suggests that prefrontal regions actively suppress the operation of emotional cortical and limbic regions. In line with this proposal, fifteen minutes of 1Hz inhibitory rTMS to the right ventrolateral prefrontal cortex was found to result in an increase of skin conductance capacity, which reflects the capacity of autonomic responses to emotional stimuli [59].

Among the neuroimaging techniques for analysing brain activity, EEG holds great promise as a diagnostic tool because of its non-invasive nature, low costs and simple setup. It provides information about the ongoing neural processes in the cerebral cortex with high temporal resolution. Therefore, I specifically intended to provide a review of studies on DPD using electrophysiological signals to detect and introduce electrophysiological biomarkers associated with DPD symptoms. My review also addresses some recent developments in the theories of self-consciousness that can potentially help to explain the unique symptomatology of DPD.

3.1.2 Review Methodology

The papers reported in this review are exclusively based on electrophysiological approaches. They comprise all the papers that have tried to explain symptoms of DPD using scalp electrophysiological signals (for a review of studies on DPD using other behavioural or neuroimaging techniques, see [2, 25]). I categorized my search based on the four major and distinct symptoms of DPD derived from factor analysis on the CDS [6, 72]. In addition, I describe studies related to ‘other symptoms’ of DPD, such as those related to perception, attention and working memory (see **Table 3.1**). I also explicitly distinguished between papers based on whether they investigated transient or chronic DPD, with transient DPD defined as episodes of depersonalisation in healthy individuals and those with a primary diagnosis of another illness. Since the essence of depersonalisation is a self-protective mechanism, its momentary symptoms can emerge in many healthy individuals during their lifespan as the

brain's response to an overwhelmingly stressful or traumatic situation in order to reduce its repercussions by creating a sense of physical and emotional numbness. Studies have reported the prevalence of transient depersonalisation in the range of 34 to 70% in the non-clinical population [28, 73]. Besides, although a comprehensive study has confirmed DPD as a distinct disorder [74], transient depersonalisation and derealisation are also common symptoms in several major psychiatric illnesses such as anxiety [29], panic attacks [75], burnout syndrome [76], and PTSD [77]. For instance, episodic depersonalisation attacks along with panic attacks have been jointly observed in several cases [78, 79], and this symptomatic pattern has been considered a distinct disorder, the "phobic anxiety-depersonalisation syndrome" [80]. The dissociative subtype of PTSD has been distinguished from its nondissociative subtype symptomatically and by distinct patterns of central nervous system activities [81]. In the dissociative subtype, patients show episodic depersonalisation symptoms, which correlate negatively with the activation of amygdala and right anterior insula [82].

Symptom	**Description**
Disembodiment feelings (desomatisation)	Lack of body ownership or loss of agency
Emotional numbing (de-affectualisation)	Attenuation in emotional responsiveness
Anomalous subjective recall (de-ideation)	Disassociation between an incident and personal feeling in memory retrieval
Alienation from surroundings (derealisation)	Detachment of the self from its surroundings
Other symptoms and processing differences	Impaired attentional functioning and processing speed or perceptual organisation

Table 3.1 Five major symptoms of DPD (adapted from [25, 33] and other associated processing differences)

3.1.2.1 Search Keywords and Information Sources

Several combinations of keywords were used in my systematic search. The search keywords consisted of one word from Set 1 and one word from Set 2 as follows (notice that, unlike Scopus, distinct results will be derived for British and American spelling of the keywords in Google Scholar):

Set 1: "Depersonalization", "Depersonalisation", "Derealization", "Derealisation"

Set 2: "Electrophysiological", "EEG", "MEG", "Biomarker", "Interoception", "Exteroception"

The search was conducted in two major electronic databases, Google Scholar and Scopus. The search was not restricted to any year range, and selected papers covered publications from 1992 to February 2020. References of all the relevant papers were also scanned to find further potential studies in the field.

3.1.2.2 Articles Overview

Papers selected (based on their title) from Google Scholar and Scopus created an initial database of 104 studies after removing duplicate papers. The papers were manually filtered, and the number was reduced to 70 based on their abstracts. Papers concerned with (neuroimaging) techniques other than electrophysiology were excluded from the review. A final assessment of the papers was conducted based on an evaluation of the full text. The final number of selected papers was 10, and these are presented as an overview in **Table 3.2**, and described in detail in the next section.

3.1.3 Electrophysiological Studies of DPD

Earlier studies sought to investigate the oscillatory signatures associated with the experience of depersonalisation. Locatelli et al. [83] examined EEG patterns in depersonalisation states. They aimed to observe the probable dysregulation in the temporolimbic regions of the brain

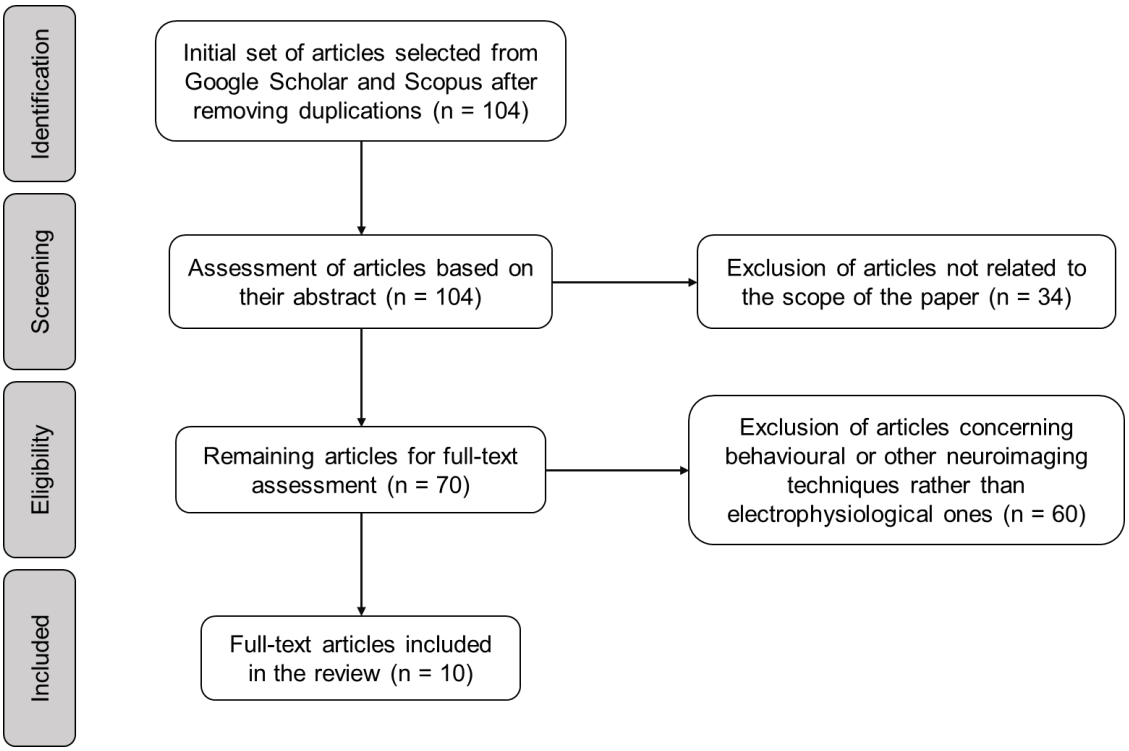

Figure 3.1 Flow diagram of the article selection process through the systematic review

in healthy subjects and panic disorder patients with and without depersonalisation/derealisation using an odour discrimination task. The power of the EEG signals in eight separate frequency bands covering 1 to 30Hz was analysed at six temporal electrodes. The results revealed bilateral abnormalities in EEG of the temporal lobe in patients with depersonalisation compared to panic disorder patients without depersonalisation or derealisation. Patients with depersonalisation/derealisation showed increased power in the delta band and a bilateral lack of responsiveness in the upper alpha band during odour stimulation. Hollander et al. [24] reported EEG power changes in depersonalisation in a different frequency band. They investigated the neurophysiological basis of depersonalisation in a 23-year-old man who had reported depersonalisation and derealisation symptoms after suffering from severe anxiety for a period. Although the long-term resting-state EEG was reported as normal, brain electrical activity mapping revealed frontal alpha overactivation and increased temporal theta activity in the left hemisphere. They also reported enhanced N200

Study (number of participants)	Task	Electrophysiological biomarker	Target symptom (transiency)	Findings
(Hollander et al., 1992) (n = 1) [24]	None specified (a case report)	EEG power spectrum (alpha, theta); ERPs (visual and auditory; cognitive N200)	None specified (chronic)	Increased frontal alpha and temporal theta activity and enhanced N200 in visual and auditory ERPs over left brain regions in a DPD patient (case study).
(Locatelli et al., 1993) (n = 37) [80]	Odour discrimination task	EEG power spectrum (delta, upper alpha)	None specified (transient)	Increased power in the delta band and a bilateral lack of responsiveness in the upper alpha band during odour discrimination in panic disorder patients with versus without depersonalisation (unable to obtain effect size from available data).
(Raimo et al., 1999) (n = 1) [81]	None specified (a case report)	EEG power spectrum (theta)	None specified (transient)	Significantly higher relative power in theta band during induced depersonalisation episodes by absolute alcohol consumption compared with asymptomatic episodes (case study).
(Hayashi et al., 2010) (n = 70) [82]	Intermittent Photic Stimulation (IPS)	EEG power spectrum (theta)	None specified (transient)	Abnormal theta activity (repeated slow waves) in 17 of 70 panic disorder patients, which was associated with three panic disorder symptoms (depersonalisation, nausea, and paresthesias) (OR = 13.92 for association with depersonalization/derealization).
(Schulz et al., 2015) (n = 47) [136]	Heartbeat perception task	HEPs	Disembodiment (chronic)	Reduced differential HEP amplitudes after 500ms between rest and heartbeat tracking in DPD patients (n= 23) compared with healthy controls (n = 24) (η_p^2 = 0.10 for the two-way interaction between group and experimental task).

(Adler et al., 2016) (n = 27) [83]	Somatosensory resonance paradigm	ERPs (somatosensory; perceptual P45, cognitive P200)	Disembodiment (transient)	Reduced somatosensory resonance for self-related information at P45 (no changes for other-related information at N80) in central-parietal electrodes in individuals with high (n = 14) versus low (n = 13) CDS scores (P45: η_p^2 = 0.46 for self-face touch in low CDS group only; N80: η_p^2 = 0.29 for other-face touch in all participants). Reduced self-other differences in resonance at cognitive P200 in fronto-central electrodes in individuals with high versus low CDS scores (η_p^2 = 0.06 versus η_p^2 = 0.41 for interaction between observed touch and observed face in each group).
(Quaedflieg et al., 2013) (n = 30) [148]	Emotion-induced blindness (EIB)	ERPs (visual; cognitive 200-300ms)	Emotional numbing (transient)	Smaller difference in ERP amplitudes in a 200–300ms time window between emotional and neutral distractors in frontal electrodes in individuals with high (n = 15) versus low (n = 15) CDS scores (*d* = 0.71 for differential activation at Fz in 200–300ms time window between the two groups).
(Papageorgiou et al., 2002) (n = 30) [166]	Working memory test	ERPs (auditory; cognitive P300)	Other (working memory) (transient)	Lower P300 amplitudes (no changes in latency) in response to single sound presenting the start of working memory tests at central posterior brain regions in healthy individuals with transient depersonalisation (n = 15) compared with a control group (n = 15) (*d* = 1.62 for group difference in P300 amplitudes).

(Wise et al., 2009) (n = 75) [180]	Auditory oddball task	ERPs (auditory; cognitive P300)	Other (attentional dysfunction) (transient)	Reduced P300 latency (no changes in amplitude) in response to striking (target) tone in panic disorder patients with transient depersonalisation (n = 25) compared with matched healthy controls (n = 50) (d = 0.62 for group difference in P300 latency).
(Schabinger et al., 2018) (n = 28) [201]	Spatial cueing paradigm	ERPs (visual; perceptual P1)	Other (attentional dysfunction) (chronic)	Reduced attentional suppression of irrelevant sensory inputs at P1 in occipito-parietal electrodes in DPD patients (n = 14) compared with anxiety- and depression-matched psychosomatic patients without DPD (n = 14) (η_p^2 = 0.66 vs. η_p^2 = 0.19 for effect of cue validity (valid vs .neutral vs. invalid) in DPD vs. control group).

Table 3.2 Studies using electrophysiological neuroimaging techniques to characterise transient and chronic depersonalisation symptoms. Note that studies are ordered according to target symptom, and then chronologically. Reported effect sizes (interpretation): d: Cohen's d effect size (small (0.2), medium (0.5), and large (0.8)). η_p^2: partial eta squared effect size (small (0.01), medium (0.06), and large (0.14)). OR: odds ratio effect size (small (1.5), medium (2.5), and large (4.3)).

components of visual and auditory ERPs (~200ms post-stimulus) in the left temporal areas. DPD was frequently found to be associated with abnormal theta activity over temporal regions. In an alcohol-induced depersonalisation state, Raimo et al. [84] observed a significantly higher relative power in the theta band compared with asymptomatic episodes. The presence of abnormal theta activities in the depersonalisation state was also confirmed in [85] and associated with a large effect size. Hayashi et al. [85] showed that depersonalisation symptoms in panic disorder patients could induce abnormalities in the EEG pattern of patients, characterised by repeated slow wave (in the range of theta) bursts.

3.1.3.1 Disembodiment Feelings (Desomatisation)

One of the cardinal symptoms of depersonalisation is disembodiment – feeling detached or estranged from one's own body parts or whole body when looking at them directly or in a mirror. DPD patients also often complain about a lack of agency - feeling as if their speech or movements are robotic and not their own. In other words, the disorder is characterized by frequent and persistent experiences of losing the physical sense of self (the feeling of oneself as the bodily subject of one's experiences). Adler et al. [37] tried to explain disembodiment in DPD patients using a somatosensory resonance paradigm, which taps into the human mirror neuron system [86, 87]. The mirror neuron system, which resides in the premotor cortex and inferior parietal cortex, as well as associated regions of sensory, motor and emotional processing [88], is active both when we perform an action (action execution) and when we observe someone else engaging in a similar action (action observation) [89] and is thought to encode the functional goal of action [90]. The mirror neuron system plays a crucial role in the adult representation of the bodily self [91-93] and its development from infancy based on contingencies between sensory and motor information [94, 95]. Adler et al. [37] investigated the mirroring mechanism using SEPs for self-related information (synchronous visual-tactile stimulation on one's own face) and other-related information (synchronous visual-tactile

stimulation on someone else's face). For people with very infrequent symptoms of depersonalisation (low CDS scores), mirroring effects for self-related information were observed at early stages of somatosensory processing (P45 component, ~45ms post-stimulus), while for other-related information, they occurred at later stages (N80 component, ~80ms post-stimulus) (both findings associated with a large effect size). For persons with high levels of depersonalisation (high CDS scores), however, the authors did not observe a self-related mirroring effect at P45. Still, they found other-related mirroring at N80 (finding associated with a large effect size). At later cognitive stages (P200 component, ~200ms post-stimulus), mirroring effects for self- and other-related information differed for individuals with low CDS scores but not for those with high CDS scores (findings associated with large and medium effect size, respectively). In accordance with [96], Adler et al. suggested the lack of early (implicit) self-related information processing as a potential biomarker to explain disembodiment feelings in DPD. A lack of self-other differentiation at later cognitive stages (P200) may also contribute to this aspect of the DPD phenomenology.

In order to explain disembodiment in DPD, it is important to understand how the sense of bodily self-attribution forms in humans. The “rubber hand illusion” [97-99] and its full-body virtual reality equivalent [100-102] have been instrumental in showing that the integration of multisensory information from our environment, specifically our bodily stimuli, forms the sense of bodily self-consciousness [99, 103-105] (for review see [106]). In the illusion, a participant's real hand receives tactile stimulation synchronously with a rubber hand, while they only see the rubber hand being touched. Synchronous (but not asynchronous) visuotactile stimuli result in changes in body ownership (the feeling that the rubber hand is part of one's own body) and self-location (the felt location of one's own hand shifts toward the rubber hand).

The multimodal integration of exteroceptive signals, such as the tactile and visual stimuli employed in the rubber hand illusion and Adler et al. 's study, is thought to occur in several temporal and parietal lobe areas [107], especially the insula [108, 109], and disruptions in this process have been proposed as a possible explanation for feelings of disembodiment in DPD [39, 110]. Indeed, participants with frequent symptoms of depersonalisation (high CDS scores) appear to be more susceptible to feelings of illusory ownership of a rubber hand than those with infrequent symptoms [38]. It is worth noting that this association between biases in multisensory information processing and anomalous bodily experiences is not limited to the clinical population. For instance, Braithwaite et al. [111] exposed the rubber hand to a realistic threat following periods of synchronous vs asynchronous visual-tactile stimulation. They found no differences in Galvanic Skin Responses (GSRs) in a group of persons predisposed to out-of-body experiences. In contrast, a control group showed elevated threat-related GSRs after synchronous compared to asynchronous visual-tactile stimulation. The GSR tracks changes in the conductivity of human skin due to sweating, reflective of the arousal related to the intensity of emotional states.

In addition to the integration of exteroceptive signals, several studies have confirmed the role of interoceptive signals in bodily self-consciousness [112, 113] (for a review, see [114]). Interoception refers to the processing of signals originating from visceral organs and representing the body's internal state [115]. The main brain regions responsible for interoception are the insula, cingulate cortex, amygdala, and somatosensory cortex [116]. Notice that the role of these brain regions has been revealed before in depersonalisation [117], and abnormal activities in those brain areas have been observed in DPD patients [68, 118-120]. Since the activity of the insula is attenuated in DPD [67, 68], attempts to explain disembodiment in DPD based on deficient interoception were made in 2014 [121, 122]. Interoceptive functioning was evaluated based on two heartbeat detection tasks (Schandry

and Whitehead tasks [123, 124]) in both studies. However, Michal et al. 's DPD patients performed similarly to healthy subjects in heartbeat detection tasks, while Sedeño et al. 's DPD patients showed impaired performance. This contradiction may be explained by the inability of interoceptive accuracy tests to reliably measure body awareness [125]. In fact, Michal et al. [121] argued that patients with DPD have specific difficulties with sustained attention to interoceptive signals as they show decreasing accuracy in the heartbeat perception task over time, while healthy observers typically increase their accuracy. This pattern may suggest deficits in sustaining attention to bodily sensations, rather than with bodily sensations per se, in patients with DPD. However, a more robust index is needed for the investigation of interoception and bodily self-consciousness in DPD.

One of the main indices of interoceptive signal processing are Heartbeat-evoked Potentials (HEPs) [126-128], which correspond to the processing of cardiac signals in the brain. HEPs are obtained by averaging brain signals time-locked with heartbeats in frontocentral regions of the brain, with the insula as their primary origin [129]. They normally can be observed from 200 to 500ms after the occurrence of the R-peak in the human electrocardiogram [127, 130]. Changes in HEPs can represent interoception and the level of attention to internal signals. They also reflect distinct attentional mechanisms for interoceptive and exteroceptive signals, such that interoception (tapping in line with one's own heartbeat) yields larger HEP amplitudes compared with exteroception (tapping in line with a recording of a simulated heartbeat) [131]. This difference was shown to be greater for people who lack interoceptive awareness, measured through the heartbeat detection task [123], and who may thus need more attentional effort to perceive their visceral signals [132]. In this regard, and to understand the role of interoceptive signals in bodily self-consciousness, Park et al. [133] investigated the association between HEPs and bodily self-consciousness. They showed that HEPs correspond to the strength of an induced illusory sense of self in a

full-body version of the rubber hand illusion. In addition, the authors confirmed the role of the insula, cingulate cortex, and amygdala in interoceptive signal processing and bodily self-consciousness, which had been reported before [134, 135].

Schulz et al. [136] studied the patterns of HEPs in people with DPD and compared them with healthy subjects through a heartbeat perception task. The objective was to find an association between feelings of disembodiment in DPD and the cortical representation of HEPs as an interoceptive signal. Healthy participants showed differential HEP amplitudes after 500ms from the onset of HEP between rest and heartbeat perception task, but such a difference was not found in DPD patients (finding associated with a relatively large effect size). A later study based on the analysis of cardiac responses to startle stimuli demonstrated that the altered pattern of visceral-afferent signals in DPD is not limited to the cortical level, as similar effects were observed in the brainstem [137]. It is worth noting that according to their earlier report [121], DPD patients' performance in the heartbeat perception task decreased over time. Patients with DPD showed higher initial performance than healthy controls but showed a decrease over time, while healthy individuals showed an increase over time, resulting in a lack of difference in performance between the two groups overall.

Since it is likely that both interoception and exteroception are involved in the formation of self-consciousness, researchers have attempted to define the role of their integration in bodily self-consciousness [138, 139]. For example, Sel et al. [140] investigated whether the integration of visual and cardiac information, as an interoceptive and exteroceptive signal, respectively, can modify self-face recognition and neural responses to heartbeats. For this purpose, a modified version of the enfacement illusion [141] as a multisensory integration method was used. The participants saw an unfamiliar face being morphed with their own face and also integrated with a pulsing shade synchronous or asynchronous with their heartbeats. The results showed changes in the HEP amplitude between 195 and 289ms after the R-peak

at centroparietal sites of the right hemisphere, with reduced HEP amplitudes during synchronous cardio-visual stimulation compared with asynchronous stimulation. The authors explained the reduced HEP amplitude as a result of a conflict created by the external representation of private information (heartbeats) by an external agent (someone else's face). Based on their argument, the brain resolves this conflict by attenuating the prominence of interoceptive sensations in the formation of self-awareness, which manifests as reduced HEP amplitudes and therefore leads to an increased perceived similarity between self and other. Similar results were later obtained by Heydrich et al. [142] for cardio-visual stimulation on a full body. Instead of HEPs, they used SEPs as a possible index for changes in self-consciousness. Self-identification with the virtual body modulated SEP component P45, which had previously been shown to reflect reduced implicit self-related information processing in DPD [37].

In addition to impairments in the integration of exteroceptive and interoceptive body-related signals, feelings of disembodiment in DPD may be exacerbated by deficits in attentional processes [143, 144], and this should be taken into account in future studies of DPD [145].

3.1.3.2 Emotional Numbing (De-affectualisation)

A second core symptom of DPD is emotional numbness [146, 147]. Self-reports from patients have asserted the lack of emotional responsivity to external stimuli in DPD [23], and fMRI studies have shown reduced activity in emotion processing regions such as the amygdala, hippocampus, temporal gyrus and anterior insula [66]. Quaedflieg et al. [148] examined the hypothesis of whether the emotion-induced blindness effect differs in individuals with high versus low levels of depersonalisation determined on the basis of scores on the CDS. Emotion-induced blindness refers to a phenomenon in which one emotionally striking stimulus draws the attention of an individual to such an extent that it reduces the processing capability for further subsequent signals [149]. Due to the lack of patients' emotional responsivity [23], an

inverse relationship between the level of depersonalisation and emotion-induced blindness was expected. The Emotional Scenes Task [149] was used to present emotional distractors 200ms before the target in order to reduce the ability for correct target detection. As expected, individuals with high levels of depersonalisation (high CDS scores) showed slightly, but not significantly, less emotion-induced blindness than those with low levels of depersonalisation (low CDS scores). The authors also examined visual ERPs during the above task for the two groups. They found a meaningful positive correlation between the magnitude of emotion-induced blindness and the ERP difference wave of emotional versus neutral distractors in a 200–300ms time window at central and frontal electrodes. Interestingly, the ERP difference wave at frontal sites was significantly smaller for high CDS compared to low CDS individuals in the 200–300ms time window (finding associated with a relatively large effect size). Therefore, the authors explained that the impact of an emotional distractor on subsequent processing is less in people with high level of depersonalisation and is found at relatively early stages (200-300ms) of information processing. They also showed that the lack of impact from emotional distractors on these ERP components was associated with the derealisation factor of the CDS rather than with levels of anxiety. In fact, depression and anxiety were related to ERP difference waves in the 600–700ms time window at frontal electrodes. Their findings confirm that DPD is a distinct psychological phenomenon from anxiety or depression [150, 151].

The attenuation of the emotional response in DPD is thought to be caused by decreased activities in emotional cortical (insula) and limbic (hypothalamus, amygdala) regions as well as increased activities of the prefrontal cortex [25, 63, 64, 68, 147]. The activation of the posterior dorsal prefrontal cortex, specifically, is thought to represent true inhibition of the intensity of an emotional stimulus [152, 153] through the functional coupling between prefrontal and limbic regions [154]. In sum, it is hypothesised that emotional numbing in DPD is due to early overactivation of these prefrontal regions, which triggers the reduction of activity

in limbic areas. The activation of this early defensive mechanism could be due to the higher sensitivity of the brain to perceive an external emotional stimulus as a threat [14].

This prefrontal-to-limbic suppressive mechanism can be further investigated using electrophysiological signals. To do so, I suggest that researchers might consider EEG and ERP markers associated with selective cortical inhibition of affective (e.g. aversive) processing, such as alpha wave activity [155], frontal alpha asymmetries [156], delta–beta coherence [157], theta/beta ratio [158], and relatively early (200-300ms) posterior negativities in visual ERPs, which are thought to be a consequence of emotional stimulus appraisal in the amygdala [159, 160], in addition to the frontal and central emotional difference waves in the same time range (200-300ms) that [148] showed to be attenuated in DPD.

3.1.3.3 Anomalous Subjective Recall (De-ideation)

Although functionally intact, memories in DPD can be subject to fragmentation, where patients have difficulty forming sequential and coherent narratives of events [58]. DPD patients might also complain of their memories being “colourless” ([25], p.4). Although patients can recall autobiographical memories, they describe them as if they did not personally experience them and as if they were an outside observer of the incident. Sierra and David [25] argued that autobiographical memory recall entails two aspects, including retrieval of the incident and retrieval of the particular feelings during that incident. Although the former aspect is intact in DPD patients, the absence of the latter [66] results in actual memories becoming colourless and like a dream. Similarly, a study of visual imagery and perception by Lambert et al. [161], compared patients with DPD and a group of healthy individuals. DPD patients performed as well as the control group in the visual perception test [162] but showed weakened ability in the imagination of visual information. Since the recall of autobiographical events and the projection of one's self into an imaginary future are similarly constructive processes [163], with largely overlapping neural underpinnings in limbic (hippocampus) and medial prefrontal,

medial parietal and temporal cortical regions [164, 165], it is conceivable that both may feel similarly self-detached in DPD.

However, in addition to the questioning of de-ideation as an independent concept in the factor structure of the CDS [146], subjective recall and imagination in DPD have not been adequately studied. In fact, no article was found in the literature on the investigation of de-ideation using electrophysiological signals. However, Papageorgiou et al. [166] observed an altered pattern of P300 during a working memory task in DPD patients. This encoding-related electrophysiological signature could therefore be a potential biomarker to investigate de-ideation in DPD [167]. Moreover, observation of N200 and frontally distributed N400 (FN400) components of ERP in DPD patients during a memory recall task might also help to discover the underlying nature of de-ideation. A study by Proverbio et al. [168] showed no distinction in N200 and FN400 components between the retrieval of an old memory and the retrieval of a more recent but emotionally salient memory, and both scenarios evoked smaller components in comparison with a recent neutral memory. Since these ERP components index familiarity of a stimulus [169, 170], the authors argued that both time and emotional valence have effects on memory recall. Enhanced N200 and FN400 components in DPD patients could therefore mark patients' subjective unfamiliarity with their memory. Finally, an additional biomarker of altered self-related memory and imagination in DPD may be occipital alpha wave activity. Resting-state occipital alpha is an index of visual cortical excitability [171], which has recently been associated with individual differences in the strength of visual imagery [172].

3.1.3.4 Alienation from Surroundings (Derealisation)

DPD refers to a chronic condition and entails, as some researchers proposed [62, 72], four distinct symptoms, one of which is derealisation. However, in the case of transient depersonalisation, derealisation is a distinct phenomenon [173], characterised by detachment from surroundings rather than from bodily self, with possible distinct neurobiological

mechanisms [174]. For instance, in a recent study by Heydrich et al. [175], the authors investigated the brain mechanism of depersonalisation- and derealisation-like symptoms in patients with epilepsy. Patients were divided into three groups those with only depersonalisation-like symptoms, those with only derealisation-like symptoms, and a control group consisting of patients with temporal lobe epilepsy who had experienced Déjà vu or experiential hallucinations. The results from multimodal neuroimaging study revealed that the majority of patients in the first group suffered from frontal lobe epilepsy while the second group mostly suffered from temporal lobe epilepsy. The epileptogenic zone in the group of patients with depersonalisation-like symptoms extended from the mediodorsal premotor cortex to the medial prefrontal cortex. Heydrich et al. 's results thus showed not only that depersonalisation and derealisation are two distinct transient phenomena, but also that they are associated with two different sources of impairments.

No study was found in the DPD literature explicitly targeting derealisation symptoms using electrophysiological methods. I propose that ERP signatures of familiarity and of attentional engagement could serve as potential biomarkers of derealisation. Since derealisation is characterised by a sense of unfamiliarity with one's surroundings, including with spaces and objects that are intimately known, N200 and FN400, which previously were highlighted as familiarity indexes [169, 170], may differ less in DPD patients than in controls during exposure to familiar versus unfamiliar scenery. Investigating the allocation of attention within such scenes may also help to delineate the underlying nature of derealisation, as unfamiliar contexts are likely to present a greater source of distraction than familiar contexts [176-178]. P300 has been shown to be associated with cortical engagement in attentional tasks [179] and may serve as a biomarker for feelings of alienation from one's surroundings. Indeed, abnormal P300 patterns have already been observed in DPD patients in other tasks [166, 180]. I further propose EEG/ERP indices of spatial cognition within egocentric and allocation reference frames as

potential measures of derealisation symptoms. For example, egocentric (vs allocentric) encoding of object locations in space has been associated with larger N1 amplitudes and longer N2 latencies at left and bilateral inferior parietal sites, respectively [181], possibly resulting from differential spatial discrimination and frame-dependent localisation processes at these stages [182-184]. Gramann and colleagues [185] found that the use of an egocentric (vs. allocentric) reference frame during spatial navigation was associated with greater alpha suppression in or near the right primary visual cortex (vs in occipito-temporal, bilateral inferior parietal, and retrosplenial cortical areas). As a result of incoherent spatial referencing between body and environment, such EEG/ERP signatures may be expected to be altered in persons experiencing derealisation symptoms.

Another way to investigate derealisation (and other DPD) symptoms may be through the probing of the vestibular system. The vestibular system includes sensory organs located in the inner ear, which send information about the head's position, spatial orientation, and motion to the brain [186], where vestibular signals are processed in distributed regions from the temporo-parietal cortex to the prefrontal cortex [187]. The vestibular system plays a crucial role, not only in the sense of balance, motor coordination, and spatial orientation but also more broadly for egocentric self-awareness [188] by providing a gravitational reference to other bodily signals [189]. When signalling pathways are disturbed in peripheral vestibular disease, the brain fails to generate a coherent spatial representation of the body with respect to the external world [190]. An incoherent spatial frame of reference may also result in feelings of detachment from the world (derealisation). Several studies in the literature indeed report a higher tendency for DPD symptoms among patients with peripheral vestibular disease than among healthy individuals. Sang et al. [191] first showed this in a sample group of 171 subjects (121 healthy, 50 patients), and later [192] reported that levels of depersonalisation were negatively correlated with patients' ability to estimate spatial orientation in an environment. Further substantiating

the potential involvement of vestibular signals in DPD symptoms, Tschan et al. [193] found that detachment from memory, derealisation and disembodiment in the general population were the three most substantial DPD symptoms associated with feelings of vertigo and dizziness, the most frequent symptoms seen in vestibular patients [191].

Future studies of DPD, specifically those interested in derealisation, may thus consider activating the vestibular system through passive full-body motion or through direct artificial stimulation of the (otolith) vestibular system [194], and probing EEG / ERP markers pertaining to cortical vestibular processing (vestibular evoked potentials, EEG power and EEG microstates). For instance, several vestibular evoked potentials can be measured over posterior, frontal and central sites within 500ms of passive motion, acoustic or galvanic stimulation [195-197]. Evoked potentials can reflect different motion parameters and have been source-localised to the cingulate sulcus visual area and the opercular-insular region [195]. These studies have also identified evoked beta- and mu-band responses in central electrodes. EEG studies have further shown that motion-induced vestibular stimulation causes bilateral temporal-parietal suppression of alpha oscillatory activity [198, 199], which was found to be attenuated in patients with the vestibular loss [198]. Cortical EEG/ERP signatures like these may thus be helpful in studying the vestibular system in DPD and egocentric self-awareness in general.

3.1.3.5 Other Symptoms of DPD

Standard neuropsychological tests have detected broad perceptual and attentional alterations in the pathophysiology of depersonalisation/derealisation [144, 200]. Already Hollander et al. [24] suggested that depersonalisation is marked by dysfunctions in the emotional modulation and integration of perceptual information, and they reported increased absolute values of visual and auditory N200 components of ERPs over the left temporal cortex in a DPD patient, in addition to abnormal theta oscillatory activity.

An electrophysiological alteration in early attentional functioning in DPD was later verified in [201]. Schabinger et al. used a spatial cueing paradigm [202] to investigate the selective attentional mechanisms [203, 204] in DPD and psychosomatic control patients who were matched for depression and anxiety. Each trial in this paradigm consisted of a target (ellipse) or non-target (circle) visual stimulus in either the left or right area of the screen, proceeded by a central spatial cue (arrow) indicating correctly, incorrectly, or neutrally the location of the upcoming stimulus. Visual ERPs in response to cued stimuli were investigated for the two groups. Schabinger et al. found diminished suppression of stimuli at to-be-ignored locations at the early sensory P1 component in DPD patients compared to control patiends (findings associated with a large effect size) while attentional effects at sensory N1 (enhancement of stimuli at to-be-attended locations) and later cognitive components (including P300) were similar across patient groups. It was proposed that the lack of early-stage suppression of irrelevant sensory inputs might be responsible for the distractibility reported by DPD patients. Schabinger et al. suggested visual ERP component P1 as a potential biomarker for deficient attentional functioning in chronic DPD.

Papageorgiou et al. examined the potential alteration in the P300 ERP component in transient depersonalisation [166]. In a working memory test, lower P300 amplitudes (but no changes in latency) at central posterior brain regions were observed in individuals with transient depersonalisation experiences compared with a control group (finding associated with a large effect size). Since high P300 amplitudes are typically evoked by conditions which demand more attention [205, 206], the findings of the above study appear to confirm attention problems in depersonalisation states. However, a later study [180] reported a related but inverse finding regarding amplitude and latency changes in P300 in the depersonalisation state. Analysis of ERP components during an oddball auditory task in patients with panic disorder revealed that patients who had experienced depersonalisation symptoms showed reduced P300

latency (but no changes in amplitude) to a striking target sound compared with healthy individuals (finding associated with a medium effect size). Reduced P300 latency indicates accelerated information processing and stimulus classification [41, 207]. No such reduction in P300 latency was observed in the comparison between panic disorder patients without depersonalisation and the control group. In addition to the unchanged attentional effects at P300 in a visual-spatial task comparing DPD with psychosomatic control patients [201], longer (rather than reduced) P300 latencies were reported in panic disorder patients compared to healthy individuals [208]. In sum, although P300 may be a valuable electrophysiological biomarker for attentional deficits in depersonalisation/derealisation, more studies with carefully designed tasks are needed to examine its precise expression in each task and for each group of patients. Additional ERP biomarkers for abnormalities in perceptual-attentional systems in DPD may be found in P1 [201] and N200 [24] components, and these also require substantial replication through new studies of transient and chronic DPD.

3.1.4 Summary

DPD can profoundly affect the quality of life of patients and interfere with their social relationships and daily activities. It usually takes several years to be correctly diagnosed [28], and the disorder's symptoms can be intolerable until then. I provided a systematic review of the studies targeting transient and chronic symptoms of depersonalisation using electrophysiological neuroimaging techniques. The aim was to describe what is presently known about the neurophysiological correlates of DPD symptoms and to make recommendations for further study to improve the diagnostic potential of this neuroimaging tool.

Before I summarise the EEG/ERP indices of DPD, it is worth noting the sparse use of electrophysiology to delineate the neurophysiological correlates of DPD symptoms. Only ten studies satisfied my criteria for inclusion in this review, two of which were single-case

studies. Yet, EEG/ERP methodologies are powerful techniques for eliciting human perception, cognition and action independently of participants' cultural background or education levels. Unlike fMRI, EEG/ERPs are direct measures of neural activity. Inexpensive, easy to implement, and well tolerated by patients, their diagnostic potential has been successfully studied in other disorders, including schizophrenia [209], which has a similar prevalence as DPD. I would therefore urge researchers interested in depersonalisation/derealisation to invest more resources into these techniques to develop their potential in supporting DPD diagnostic process beyond what is presently known.

Several studies have shown abnormal EEG activities in theta band in DPD patients, and the severity of symptoms was found to be associated with increased theta activity. Higher theta wave synchronisation has been observed for emotional compared to neutral stimuli, and this synchronisation occurs earlier (around 200ms) when emotional stimuli are processed implicitly but later (around 300ms) during the explicit recognition of facial emotions [210]. Therefore, increased theta activity in DPD patients might be associated with their greater effort or involvement in processing emotional information (specifically unpleasant emotional information [56, 147, 148]). Another study [211] showed that the theta wave represents the suppression of nontarget stimuli in the go/nogo task, which requires both selective inhibition and arousal. Therefore, theta activity might also represent the greater effort in the selective suppression of processing, which has been observed to be deficient in DPD patients [201]. Theta further plays a role in memory maintenance in that higher theta activity is associated with the need for greater working memory capacity [212, 213]. Thus, the theta band of the EEG power spectrum is likely to serve as a potential electrophysiological biomarker to study DPD symptoms related to emotion, attention/inhibition and working memory. Future studies of DPD should therefore focus on this oscillatory signature and investigate (a) the temporal dynamics of event-related synchronisation in response to emotional stimuli and (b)

oscillatory power during transient episodes of depersonalisation and in patients with chronic DPD, for example. Note that it is important for future studies to consider a relatively prolonged time window in order to analyse low-frequency components of EEG such as theta (practically, the time window needs to contain at least three cycles of the target frequency), which may affect stimulus design.

Analysis of cortical event-related and heartbeat-evoked potentials (ERPs and HEPs) can also reveal valuable information regarding the underlying nature of DPD symptoms, as well as of the sensorimotor integrative processes contributing to bodily self-consciousness in general. HEPs and some ERP components have been introduced as potential valuable indices to investigate DPD symptoms related to disembodiment. Somatosensory P45 reflects processing in the primary somatosensory cortex and is known to be involved in attributing body ownership [37, 214, 215]. Cortical HEPs are thought to primarily derive from insula activity, which is attenuated in DPD [67, 68]. A lack of P45 modulation during visual-tactile stimulation related to the self [37], and a lack of HEP modulation during focused attention to one's own heartbeat [136], may both serve as an electrophysiological biomarker for feelings of disembodiment (desomatisation) in transient and chronic DPD. Both somatosensory cortex and insula are part of the networks responsible for interoception [112], and the integration of interoceptive, exteroceptive, and interoceptive with exteroceptive sensory signals have been recently proposed as critical for generating the moment-to-moment feeling of self-consciousness. When integrative processes like these are transiently or chronically dysfunctional, feelings of disembodiment may ensue and may be further exacerbated by abnormal attentional processes [145]. My review has highlighted the potential of somatosensory P45 and of HEPs for measuring these processes in health and disease. It is worth mentioning that there might be a link between somatosensory P45 and cortical HEP in the investigation of DPD symptoms. A recent study [216] showed that the activation of the

somatosensory P45 component in response to personal visual-tactile stimulation (touch on the subject's own hand) is stronger in people with a higher level of interoceptive awareness. If interoceptive awareness modules HEPs [217], measuring both P45 and HEPs in the same study may be able to test the relationship between interoceptive and exteroceptive awareness [218]. In addition, researchers may consider the inclusion of electrophysiological markers for proprioceptive and vestibular signal processing to investigate the bodily self in health and disease. Vestibular processes are increasingly recognised as critically involved in feelings of body ownership and egocentric perception and vestibular disturbances bear a strong link with several cardinal DPD symptoms [193, 219].

Another potential set of ERP biomarkers for DPD symptoms occurs around 200-300ms [148], including N200 [24], P200 [37] and P300 [166, 180]. In this review, I have referred to enhanced temporal N200 as an aspect of alterations in the perceptual-attentional system [24], to lack of self-other differentiation in somatosensory resonance at frontocentral P200 as an aspect of disembodiment [37], to lack of frontocentral 200-300ms differences [148] as an aspect of emotional numbing, and to reduced amplitude/latency at centro-parietal P300 [166, 180] as an aspect of working memory/attentional dysfunction. I have also suggested the potential of additional markers in this time range, such as the N200 and frontal N400 related to stimulus familiarity. Partly due to inconsistencies in the precise expression of P300 abnormalities in DPD, further studies are urgently needed to systematically confirm P300 as a potential electrophysiological biomarker for investigating DPD symptomatology related to attentional and working memory dysfunction, as well as to symptoms of de-ideation and derealisation. Caution should be applied when regarding the association between P300 and DPD symptoms because P300 changes also occur with a number of other pathologies and physiological states (e.g. [220-222]) and are thus not necessarily selective for depersonalisation.

Reduced brain activities have been observed in DPD patients in sensory information processing units [68] as well as regions responsible for the processing of visceral signals [117]. Besides, the impairment is found mainly in the early stages of information processing [37, 148, 201]. This impairment in the implicit processing of multimodal interoceptive and exteroceptive signals could be a result of long-time severe stress or anxiety [13], which may damage the sensory processing units and reduce their processing capacity [14]. For instance, one dominant theory to explain emotional numbing in DPD [14] defines a threshold for the level of anxiety (or any unpleasant salient stimuli) after which the emotional processing units (including the anterior insula and amygdala) discontinue translating emotions into perceived feelings, and DPD is associated with abnormalities in this threshold or in how quickly it is reached [59]. The higher GSR baseline and the earlier peak in GSR response of DPD patients to emotional stimuli in [58] represent a faster saturation of emotional capacity [147]. Further, reduced capacity in sensory processing units may also be the cause of concentration problems in DPD, in a way that there may be a commensurate loss in the capacity to filter relevant from irrelevant signals in sensory information. A related framework within which symptoms of DPD may be explained is based on the loss of the brain's ability to make and update predictions about the internal body state and the outside world [145, 223, 224]. With reduced activity in emotional information processing units (such as insula), the brain faces a relative sparsity of information for maintaining precision in predictions about one's self and the outside world. The increasing inconsistencies may cause further suppression of emotional processing and, in turn, give rise to DPD symptoms such as feelings of disembodiment or derealisation. Conflict-monitoring-related electrophysiological markers such as the Mismatch Negativity (MMN) and Error-related Negativity (ERN) are good candidates for investigating this theory further. MMN is a pre-attentive fronto-central response around 100-200ms after omitted or perceptually deviant sensory events in a stimulus sequence and is generated by a

network of temporal-prefrontal cortical regions [225]. ERN is generated by a region in the Anterior Cingulate Cortex (ACC) [226] and emerges around 100ms after the onset of an erroneous motor response, even when the observer is unaware of making an error and even when the error cannot be corrected [159]. Recent studies have begun to describe both MMN and ERN within a predictive-coding framework as indexes of the brain's monitoring of bodily processes, where MMN may represent the failed prediction of visceral and other sensory inputs and thus the failed suppression of the prediction error [225, 227]. ERN may represent the monitoring of errors committed as a consequence of failed interoceptive and exteroceptive predictions, with imbalances in monitoring shown to be related to anxiety pathology [228-230]. Thus, MMN and ERN may be useful for indexing errors in predicting the internal and external state of the body in persons with DPD.

There is not enough evidence to confirm a clear hemispheric lateralisation of DPD symptoms' neurophysiological correlates. Nevertheless, it is likely that such hemispheric biases will eventually emerge with more targeted research. Some studies have already shown abnormalities in left-hemispheric activation of DPD patients [24, 231] and of individuals with more frequent dissociative experiences [232]. Furthermore, based on the dominant role of the right hemisphere in emotional [233] and self-related processing [91], right-hemispheric biases in the dysregulation of emotional processing and disembodiment in DPD may be expected. In support of this argument, it may also be interesting to note that the right hemisphere may be more involved in the perception and processing of negative emotions, while the left hemisphere deals predominantly with positive sensations [234], and that impairment in emotional responsivity in DPD is often exclusive to unpleasant and threatening stimuli rather than pleasant ones [54-56]. In addition, inhibitory rTMS applied to the right ventrolateral prefrontal cortex can increase arousal capacity [59] and rTMS applied to the right temporoparietal junction can reduce symptoms (as measured by CDS total score) in

DPD patients [235]. Increased activation in the angular gyrus of the right parietal lobe has also been found to be correlated with the level of depersonalisation [61]. However, the DPD literature currently faces a lack of direct investigations of hemispheric differences related to DPD symptoms; more studies on this are urgently needed.

A schematic of the notable biological signals and associated electrophysiological biomarkers introduced in this review are depicted in **Figure 3.2**. Future studies in this area should consider both indexes of interoceptive (such as HEPs) and exteroceptive signals (such as visual or somatosensory ERPs), vestibular signals, and especially their interaction (including monitoring ERP markers like MMN, ERN). For a comprehensive picture of DPD, it is useful to jointly investigate both peripheral (autonomic) and central (cerebral) bodily responses, and both early (perceptual) and later (cognitive) stages of central information processing. Additionally, there is a lack of research in the analysis of EEG power spectra and their relative ratios and coherence. As this review has uncovered that these may yield promising biomarkers for all cardinal DPD symptoms, further studies need to consider different EEG waveforms and their roles in the formation of these symptoms. For such analyses, the use of time-frequency and phase-based signal processing [236] can be very insightful since studies have suggested phase alignment as a fundamental phenomenon underlying the generation of ERPs and HEPs rather than evoked potentials [129, 237].

Further, the association between the vestibular system and the DPD symptoms can help the design of experimental paradigms in future studies [238]. In this regard, a study by Sang et al. [191] proposes caloric vestibular stimulation as an effective way to provoke depersonalisation/derealisation-like symptoms in the non-clinical population. The induced symptoms were similar to those experienced by patients with vestibular disease. However, it should be carefully considered that there is also a close link between the vestibular system dysfunction and out-of-body experiences [239-241]. For example, Lopez et al. [240] reported

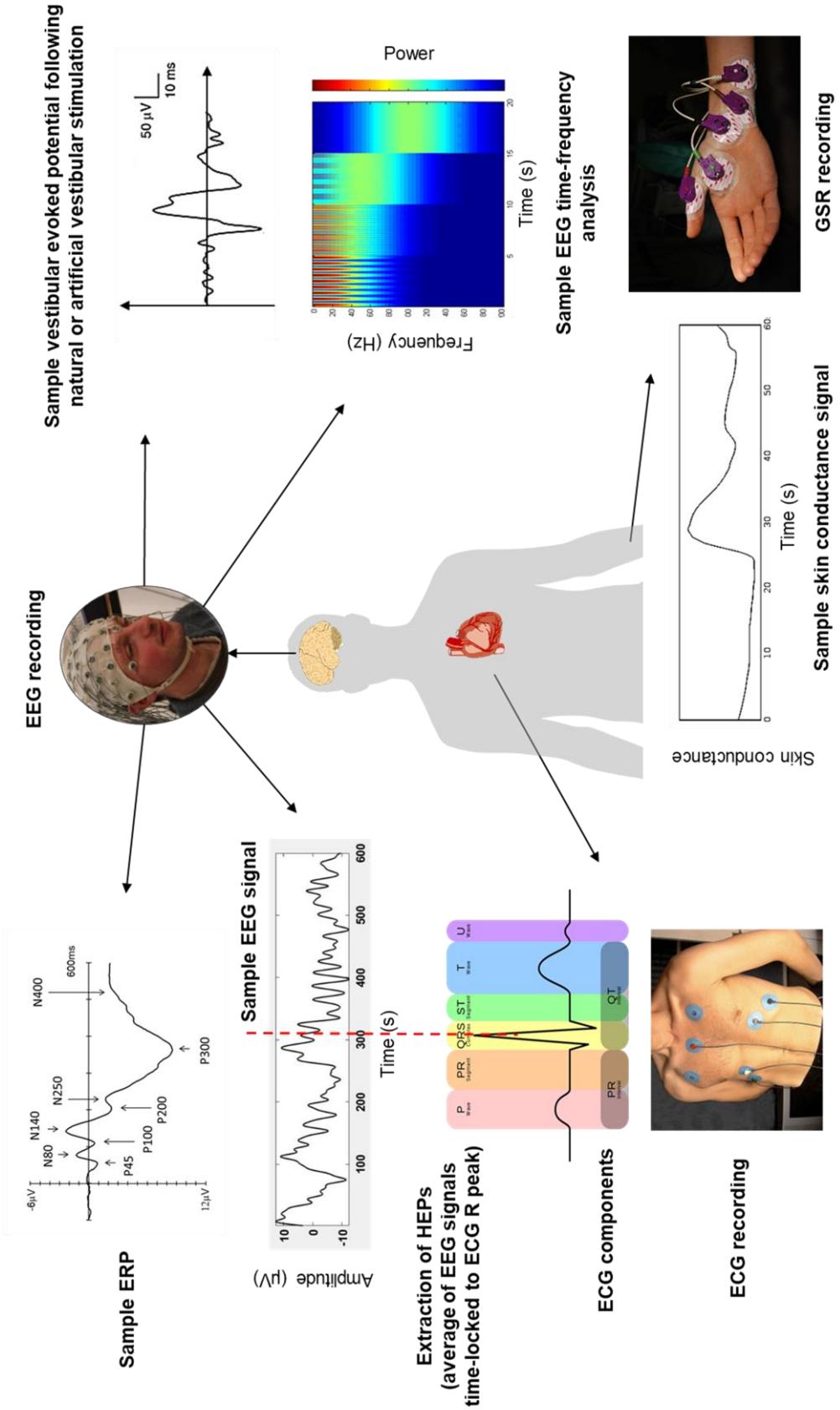

Figure 3.2 An overview of the biological signals and relevant electrophysiological biomarkers introduced in this review

a higher tendency for out-of-body experiences in patients with dizziness, and the relationship was more significant in patients with a peripheral vestibular disorder. Since out-of-body experiences are not that common in depersonalisation/derealisation [25], using methods to dysregulate the vestibular system (such as caloric vestibular stimulation) to induce depersonalisation/derealisation symptoms should be done carefully to prevent misinterpretation.

3.2 Deep Learning for EEG Signal Analysis

My research project falls into the area of BCIs. Therefore, I first provide a brief introduction to BCIs and then discuss the shortcomings of the available models and algorithms to address all aspects of my scenario.

3.2.1 Brain-computer Interfaces (BCIs)

BCIs were developed to provide direct communication between the human brain and external devices through the acquisition of neural signals and their translation and transmission as control commands to assistive devices [242]. Although initially designed to help people with cognitive or physical impairments, BCIs have a wide range of applications nowadays, ranging from medical and neuroergonomics to entertainment, intelligent environments, and even security and authentication [243]. EEG-based BCI systems are among the popular techniques due to the convenient and inexpensive nature of EEG signals. The electrical activities of a population of neurons can generate adequately high electrical fields that can be measured by electrodes placed on the surface of the scalp. However, the recorded signals have a low SNR due to several factors, such as interference from other physiological signals, electrode detachment, or signal distortion resulting from the cortex's behaviour as a volume conductor [40]. Therefore, advanced signal processing and machine learning algorithms have

been extensively applied in this area to increase the SNR and the quality of the extracted neural patterns of interest [244].

3.2.2 Existing Deep Architectures and Their Shortcomings

Deep learning has seen considerable use in the analysis of medical imaging data, including EEG signals [245]. The investigations in this area have been mainly based on supervised learning and with the purpose of performing classification tasks, especially in the field of BCI. For instance, Rezaei Tabar et al. [246] used a stacked autoencoder after a pre-trained CNN to perform binary classification in a motor imagery BCI task. They used the short-time Fourier transform of multi-channel EEG signals as an input to enrich the extracted feature vector from the CNN. Although they managed to improve the classification accuracy compared to the state-of-the-art approaches, they did not provide any insight into the learned features and the interpretation of discriminative features used to perform the classification task. Accordingly, more recent publications have focused not only on proposing novel architectures to analyse EEG data but also on extracting and interpreting features of the network from a neuroscientific perspective.

Schirrmeister et al. [247] investigated different end-to-end CNN architectures, such as shallow CNN, deep CNN, and residual CNN, for EEG analysis. They reported relatively higher performance than the well-known Filter Bank Common Spatial Pattern (FBCSP) [248] and showed deep learning as a promising tool for analysing EEG signals. They also used two novel feature visualisation techniques based on the correlation between several major frequency bands and the output of CNN units to understand the level of contribution of each feature in the CNN final decision. The limitation of their visualisation method was its dependency on known features, which makes it impracticable in cases where there is no prior information on discriminative features. In another influential research study, Lawhern et al. [249] introduced EEGNet as a compact CNN capable of analysing different EEG-based BCI

paradigms. They also showed how the filter weights of convolutional layers could be visualised to represent and portray the learned features. Although EEGNet showed promising results as a general model, it was still only a supervised learning algorithm. Hence for my DPD scenario or generally for analysing mental disorders based on clinical assessment scores, a multi-task learning structure is needed to handle clinical assessment scores as prior information rather than entirely relying on them. In addition, I intended to find a more explainable visualisation technique for the learning process, which helps to intelligibly investigate and discuss my findings from a cognitive neuroscientific point of view.

Therefore, I aimed to propose a novel end-to-end EEG processing pipeline based on deep neural networks for DPD biomarker discovery, which can employ the CDS scores as prior information in the learning process and provide explainability to discuss and validate its findings. I will argue that the proposed EEG analytics could also be applied to investigate other psychological and mental disorders currently indicated on the basis of clinical assessment scores.

4 Proposed Methods

In this chapter, I first present my EEG pre-processing pipeline to clean and prepare the raw EEG data before feeding it to end-to-end deep models. The chapter then details my proposed EEG analytics based on deep learning algorithms acting as a biomarker discovery system for DPD. In addition, the chapter includes the visualisation technique used to evaluate and explain the learning process. As a part of my project and to show the reliability of my visualisation technique, I applied the approach to a novel convolutional network called EEG-ITNet, designed for motor imagery classification. The EEG-ITNet structure has been extensively explained in this chapter. Finally, to show the importance of finding a reliable electrophysiological biomarker for supporting DPD diagnosis, I present a waveform matching algorithm to perform a classification task on an auxiliary DPD dataset.

4.1 Proposed Pre-Processing Pipeline

The raw EEG data was fed into my proposed automated preprocessing pipeline, detailed in **Figure 4.1**. The pipeline starts with downsampling to 250Hz to reduce the size of the dataset and decrease the processing time without considerable loss of information. Then, a high-pass Finite Impulse Response (FIR) filter with a 1Hz cut-off frequency was applied to the multi-channel EEG signals to remove the DC offset and low-frequency artefacts, followed by a low-pass filter with a 40Hz cut-off frequency to remove high-frequency artefacts and 50Hz

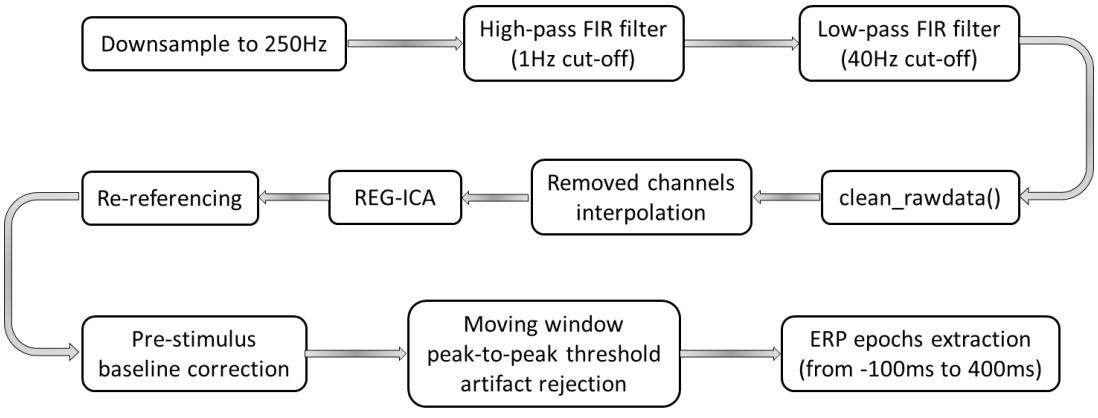

Figure 4.1 An overview of the proposed automated preprocessing pipeline

line noise. One might argue that a relatively high 1Hz cut-off frequency can be detrimental in terms of affecting ERP components. However, since somatosensory processing is mainly associated with early high-frequency ERP components [250], using a 1Hz cut-off frequency was not only unlikely to have a negative impact on my results but was also crucial to correct significant signal distortion in my dataset and save as many trials as possible. Nevertheless, in the case of no information on the nature of the experiment or high-quality recordings, a high-pass filter with a lower cut-off frequency is generally advisable. In the next stage, the clean_rawdata() plugin in EEGLAB was used to detect and remove corrupted channels automatically, including the ones with a constant pattern, excessive noise, or poor scalp-surface contact. clean_rawdata() is the offline version of Artefact Subspace Reconstruction (ASR) method proposed by Christian Kothe (details can be found in [251]). Next, those rejected EEG channels were interpolated using other nearby channels. Using data from two EOG channels, I then used the REG-ICA algorithm [252] to remove blinks and other EOG artefacts from my EEG signals. REG-ICA is a hybrid algorithm for EOG artefact rejection based on Independent Component Analysis (ICA). The method applies a regression algorithm to compare independent components with EOG channels to decontaminate them. I used Preconditioned ICA (PICARD) [253, 254] as the ICA algorithm and Least Mean Square (LMS) as the regression algorithm. After artefact rejection, I used the average of all electrodes to re-reference EEG voltages, followed by 100ms pre-tactile-stimulus baseline correction. I also applied moving window peak-to-peak threshold artefact rejection to exclude any trial that was not cleaned during the earlier steps. Finally, ERP epochs were extracted from 100ms before tactile stimulus onset to 400ms after tactile stimulus onset.

4.2 Score-guided Biomarker Discovery System for DPD

I argued that EEG studies on finding discriminative neural factors are often based on hypotheses proposed by experts in the field, meaning it requires prior knowledge of the

disorder. In addition, participant labelling in research experiments is often derived from scores on the CDS, the threshold and reliability of which might be challenged. As a result, I aimed to propose a novel end-to-end EEG processing pipeline based on deep neural networks for DPD biomarker discovery, which requires no prior knowledge or assumption of the disorder. In addition, the multi-task learning nature of the proposed deep model targets the potential unreliability of CDS scores by using them as prior information only to guide the unsupervised learning task. Thus, this section introduces my main DPD dataset, which I used to train and evaluate the score-guided biomarker discovery system. Note that I will later introduce another DPD dataset with fewer participants that I used to evaluate the reliability of identified biomarkers. The second dataset will be referred to as the auxiliary DPD dataset. After introducing the main DPD dataset, I fully demonstrate my deep learning structure, its corresponding loss function, and the learning process.

4.2.1 Main DPD Dataset

The main DPD dataset used in my study was collected before by [255]. In the original study, which was inspired by Adler et al. [37], emotional primes and a tactile mirroring task were used to examine the cardinal symptoms of DPD, including emotional numbness and disembodiment. The study was approved by the Human Research Ethics Committee of the authors' institution, and all participants gave informed written consent to take part. The dataset consists of 50 participants who initially took the self-rating CDS questionnaire to quantify their (trait) level of depersonalisation. With a threshold of 50 on CDS scores, 21 subjects were evaluated as participants with a low level of depersonalisation, and 29 subjects were considered individuals with a high level of depersonalisation, henceforth termed the control group and the DPD group, respectively. It is important to recall that the DPD diagnostic process is not solely based on CDS, and the outcome of the questionnaire only helps the clinician in their final diagnosis of DPD as a primary condition [28]. The diagnostic

process also involves several examinations, including but not limited to physical exams, lab tests, and psychiatric evaluations. Nevertheless, since there is not always access to clinically diagnosed DPD patients in experimental research, CDS is often used as a primary metric to label participants as those with low and high levels of depersonalisation.

In addition to the CDS, participants' levels of depression and anxiety were also recorded based on Patient Health Questionnaire-9 (PHQ9) and State-trait Inventory for Cognitive and Somatic Anxiety (STICSA), respectively. The aim was to control the effect of depression and anxiety, which are highly comorbid with depersonalisation [256, 257], in the analysis. Participants also completed the Operationalised Psychodynamic Diagnosis-Structure Questionnaire (OPD-SQ) self-object differentiation subscale, which has been associated with dissociation [258], and the Multidimensional Assessment of Interoceptive Awareness (MAIA), which assesses eight different dimensions of subjective interoception [259].

Each session in the experiment consisted of two types of trials; tactile stimulation following an emotional prime and tactile stimulation following a neutral prime. The emotional prime was in the form of a happy or angry face and voice, with happy and neutral primes and angry and neutral primes forming two distinct datasets. During the experiment, subjects with normal or corrected-to-normal vision were asked to sit in front of a computer screen. Each trial started with a 400ms window of emotional or neutral prime, followed by a 600ms neutral hand stimulus (subjects observed a left or right hand and a pencil against a white background). The next 200ms time window comprised either the pencil touching the participant's hand (touch condition) or the space next to the hand (no-touch condition) so that the perceived distance of the pencil travelling would be the same. Finally, each trial ended with replaying the neutral hand stimulus for 600ms. In the 200ms time window of both touch and no-touch conditions, the participants received a 200ms tactile stimulus to their same hand, resulting in a synchronous visual-tactile stimulation for the touch condition. All trial

types within each set were randomly intermixed with equal frequency. The animated schematics and timing scheme of each trial are presented in **Figure 4.2**. EEG signals were recorded during each session using a Compumedics Neuroscan SynAmps RT 64-channel amplifier and an EasyCap scalp electrode cap at a 1000-Hz sampling frequency and an online filter of 0.01-100 Hz.

The main DPD Dataset holds several types of trials as it contains two subsets (happy and angry), two conditions (touch and no-touch), and two types of visual stimulation (emotional and neutral prime). Since studies have shown abnormal autonomic nervous system responses in both the brain and the autonomic nervous system of DPD patients [260], which tend to be more evident for unpleasant and threatening emotional stimuli [54], I decided to only focus on the angry set and synchronous visual-tactile stimulation (touch condition) following emotional primes (angry faces and sounds). However, even by focusing on a single subset and a single condition after a single prime, one would still end up with two types of trials representing tactile stimulus to the left or right hand. It is essential to consider that those trials still need to be analysed separately since somatosensory processing in the brain is only initially lateralised to the hemisphere contralateral to the touch. Accordingly, one should not

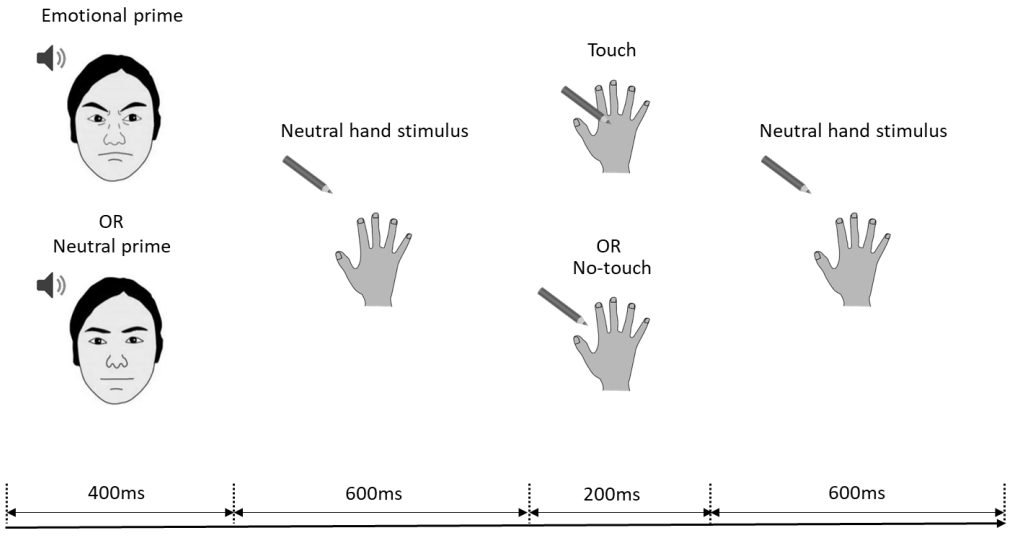

Figure 4.2 The schematics and timing of each trial in the main DPD dataset

combine those trials by remapping the electrodes in one of the conditions unless one is interested in only early ERP components.

Based on the quality and quantity of processed EEG signals for each participant, data from 7 participants were excluded from further analysis in the main DPD dataset, resulting in a total of 19 and 24 control and DPD participants, respectively. Furthermore, I disregarded almost half of the EEG channels and analysed only those presented in **Figure 4.3** to simplify my analysis. Note that the channel selection was in line with the nature of my experiment, which contained visual and tactical stimulation.

4.2.2 Network Architecture

The deep learning architecture proposed in my research as a biomarker discovery system is a multi-input multi-output deep neural network, as depicted schematically in **Figure 4.4**. The two input branches of the network consist of sequences of layers similar to the structure of the well-known EEGNet [249] to analyse the trials with tactile stimuli to the left and right hand separately. Generally, each branch starts with a 2D convolutional layer with a kernel

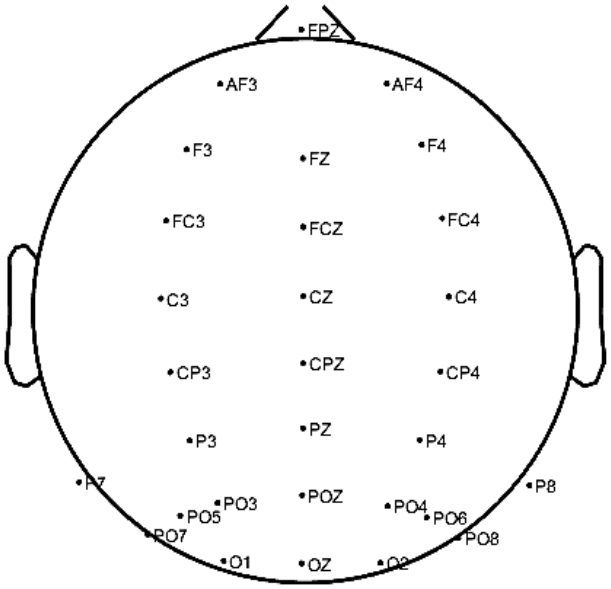

Figure 4.3 The names and placement of electrodes used in the research

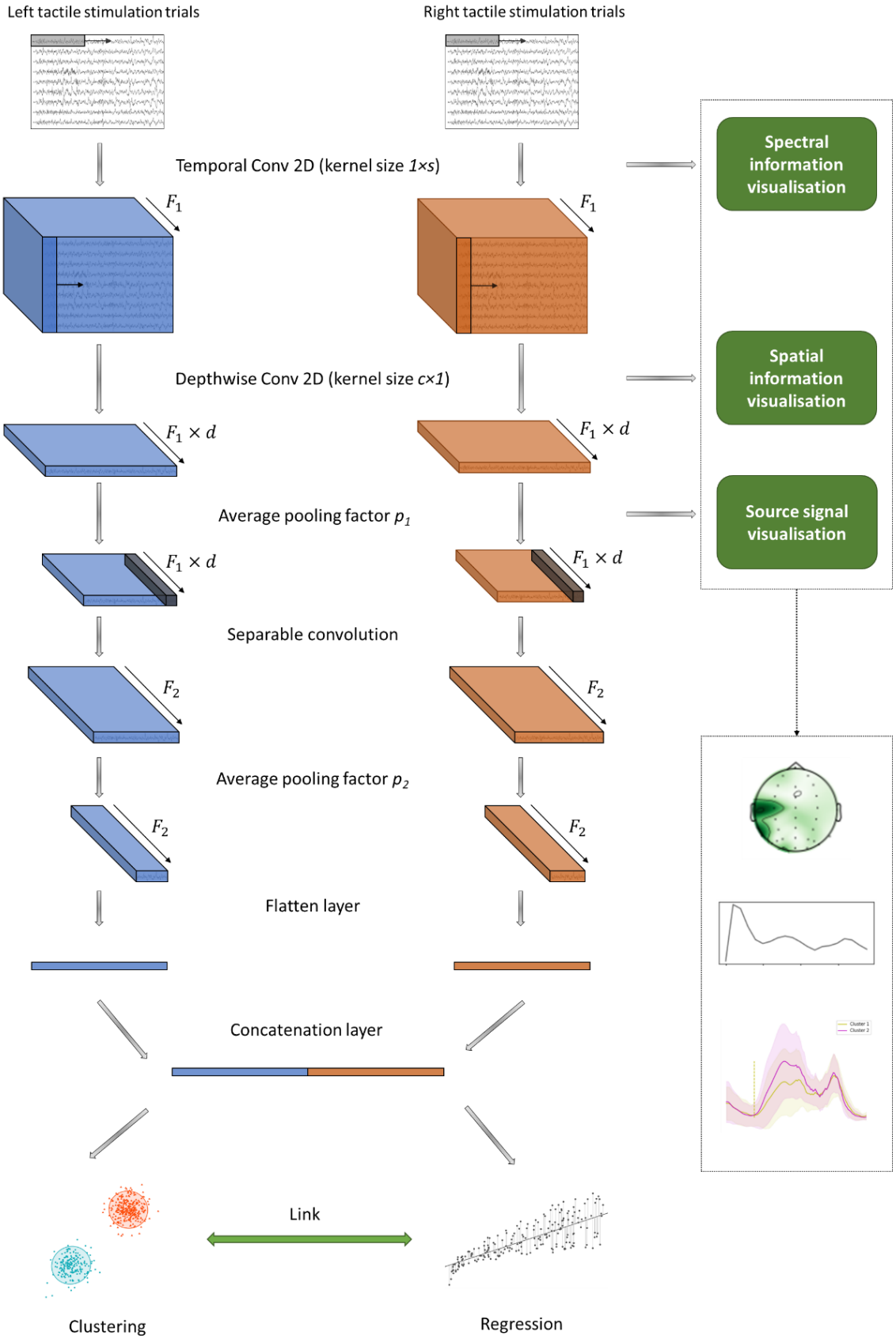

Figure 4.4 Overview of the proposed deep model for biomarker discovery

size of $1 \times s$ acting as a frequency filter, followed by another depthwise 2D convolutional layer [47] with a kernel size of $c \times 1$ acting as a spatial filter, which together simulates the behaviour of the well-known FBCSP algorithm (explanation on that has been provided in section 4.3). The kernel size for the depthwise convolutions is equal to the number of electrodes in the dataset to design a spatial filter that combines all the electrodes to find the sources of brain activities. However, the number of convolutional filters in the first layer (F_1) and the second layer (d) are the hyperparameters of the system, which determine the number of frequency bands and spatial filters in each frequency range, respectively. The goal of the later separable convolutional and flatten layers is to find a low-dimensional representation of the input multi-channel EEG signal. Then, low-dimensional representations derived from both types of trials are concatenated to form a more extensive feature vector, which serves as an input to the final multi-task learning structure of the network.

The output of the network consists of a supervised and an unsupervised branch. The supervised branch is a fully-connected layer with one unit and a "Relu" activation function to predict the continuous CDS scores, while the unsupervised branch is a fully-connected layer with two units and a "softmax" activation function to generate cluster assignments. The idea was to learn a representation that separates the dataset into two patient and control clusters (clustering branch) guided by the CDS scores (regression branch). Because on the one hand, I mentioned that the CDS scores are subjective and imprecise, so I should rely on more than just them to find the electrophysiological biomarkers. On the other hand, clustering can be accomplished by discriminative yet nonmeaningful or confounded features. Therefore by defining an appropriate loss function, I 1) guide the network to find and extract features that represent two distinct neural patterns; 2) make sure they represent patterns of individuals with a high and low level of depersonalisation.

4.2.3 Proposed Loss Function

In order to achieve my objectives, I proposed a loss function as follows:

$$\mathcal{L}_{total} = w_{regression}\mathcal{L}_{regression} + w_{clustering}\mathcal{L}_{clustering} + w_{link}\mathcal{L}_{link} \quad \text{s.t.} \; w_{regression} + w_{clustering} + w_{link} = 1 \tag{4.1}$$

where $\mathcal{L}_{regression}$ is the loss function associated with the regression branch, forcing the network to predict the continuous CDS scores. I used Mean Squared Error (MSE) for this purpose which can be formulated as follows:

$$\mathcal{L}_{regression} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4.2}$$

where n is the number of samples per batch, y_i is the reported CDS score, and $\hat{y}_i$ is the network-predicted score from the regression branch. The $\mathcal{L}_{clustering}$ in (4.1) denotes the unsupervised loss function used to help the network find a low-dimensional representation that separates data points into two distinct clusters and their corresponding cluster assignments. For this purpose, I used information maximisation [261, 262], which is simply defined as an estimate of the mutual information between the low-dimensional representation of the input data and cluster assignments. Let $E \in (e_1, \dots, e_n)$, where $e_i \in \mathbb{R}^d$, denotes a d-dimensional random variable representing the concatenated low-dimensional representation of left and right tactile stimulus inputs. By defining $Z \in \{0,1\}$ as a random variable expressing cluster assignments, one can estimate mutual information between E and Z as follows:

$$I(E;Z) = H(Z) - H(Z|E) \tag{4.3}$$

where $H(.)$ and $H(.|.)$ are entropy and conditional entropy, respectively, and can be calculated as follows on a batch:

$$H(Z) = -\sum_z \left[\left(\frac{1}{n}\sum_{i=1}^{n} p(z|e_i)\right)\left(\log\left(\frac{1}{n}\sum_{i=1}^{n} p(z|e_i)\right)\right)\right] \quad (4.4)$$

$$H(Z|E) = -\frac{1}{n}\sum_{i=1}^{n}\left[\sum_z p(z|e_i)\log\left(p(z|e_i)\right)\right] \quad (4.5)$$

where z is an instance of the random variable Z. Thus, the clustering loss function based on information maximisation can be finally defined as follows:

$$\mathcal{L}_{clustering} = -(H(Z) - H(Z|E)) \quad (4.6)$$

By this definition, a lower loss value would be subject to an increase in marginal entropy $H(Z)$, which encourages the cluster assignments toward class balance and avoids trivial solutions, and a decrease in conditional entropy $H(Z|E)$, which ensures having high confidence in each cluster assignment.

Using the weighted sum of $\mathcal{L}_{regression}$ and $\mathcal{L}_{clustering}$ as the loss function, the supervised and unsupervised tasks in my model could be achieved independently due to their independent units, making the network highly prone to overfit on uninformative features. In other words, finding clusters representing my desired groups of people with low and high levels of depersonalisation was not guaranteed. To tackle this problem, I introduced a third term in my loss function called $\mathcal{L}_{link}$, which bridges the gap between the supervised and unsupervised branches and ensures the finding of meaningful desirable neural patterns to distinguish the patients from the control group. To achieve that, I first used a smooth logistic function in the form of $f(x) = \frac{1}{1+e^{-(\frac{Thr}{10}+0.1x)}}$ (see **Figure 4.5-right**) to scale the predicted scores from the supervised task to numbers between 0 and 1. The *Thr* in the indicated logistic function determines the turning point of the function and is the threshold on CDS scores to evaluate subjects as control or DPD patients. As mentioned earlier, there is no globally agreed threshold on CDS scores for DPD classification. While researchers often choose 50 in

their studies, clinicians prefer to use 70. Therefore, I performed a greedy search by sweeping all the possible values for the threshold from 20 to 100 to find the one that results in the lowest loss value, with the idea that an optimum threshold would be the one that performs subjects' separability with high confidence while making an accurate prediction of the CDS scores. **Figure 4.5-left** shows the sweep result, illustrating the mean and standard deviation (as shadow) of the clustering loss values over ten iterations for each threshold value. I found an optimum threshold of 56, which lies between the common threshold in the literature and clinician preference, and used that value in the rest of my analysis.

The scaled scores following the logistic function were then compared to the cluster assignments using the cross-entropy loss function. The idea was based on the fair assumption that participants with extreme scores (too low or too high) should be assigned to their corresponding cluster with higher confidence. So assume $\hat{y}_i$ the network predicted score from the regression branch for the i-th data point and $f: \hat{y}_i \rightarrow s_i$ the optimised logistic function, where s_i donates the scaled scores ($s_i \in [0,1]$). The scaled scores can form a vector $\vec{s} = \begin{bmatrix} 1-s_i \\ s_i \end{bmatrix}$ showing how likely each input data belongs to each group. Similarly, in my binary

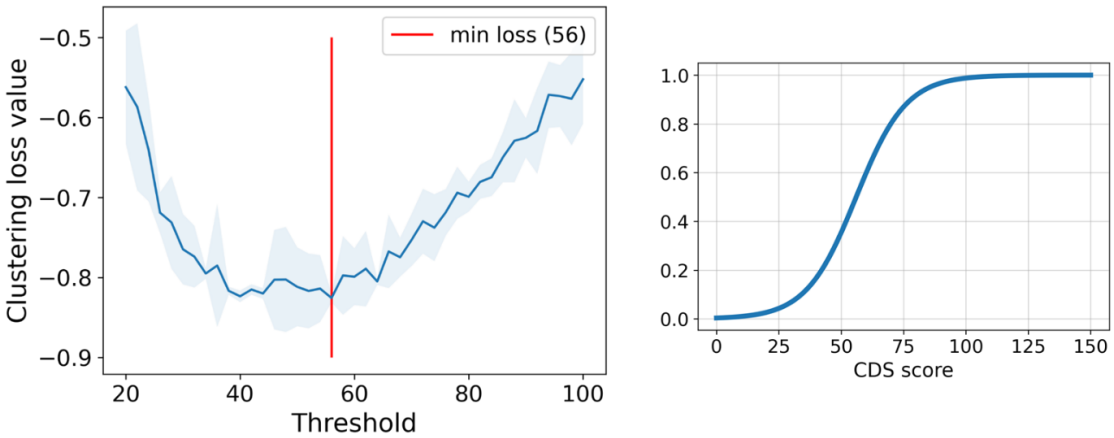

Figure 4.5 Minimum clustering loss value for different CDS thresholds (left) and the smooth logistic function defined to transform CDS predictions to a value between 0 and 1 with a turning point of optimum threshold (right)

problem, the output of the clustering branch for the input i is a vector $\vec{z} = \begin{bmatrix} z_{i1} \\ z_{i2} \end{bmatrix}$ (where $z_{i1} + z_{i2} = 1$) containing the cluster assignments. As a result, the $\mathcal{L}_{link}$ can be defined as follows:

$$\mathcal{L}_{link} = -\frac{1}{n}\sum_{i=1}^{n}[(1 - s_i).logz_{i1} + (s_i).\log(1 - z_{i1})] \quad (4.7)$$

In sum, the proposed loss function (4.1) can be trained using gradient descent to minimise the CDS prediction error while forming two clusters and guarantee getting clusters representing my two groups of participants with low and high levels of depersonalisation. Notice that my proposed deep learning model does not require a validation set as the $\mathcal{L}_{link}$ causes a trade-off between the regressions and the clustering losses, preventing them from overfitting. Therefore, I only used early stopping in my model to terminate the learning process once I no longer see improvements in the total loss, meaning the supervised and unsupervised tasks have reached an equilibrium point.

4.3 Explainability and Visualisation Technique

The goal of my study was to find potential electrophysiological biomarkers for DPD. For that, I needed to dig into the learning process of my deep model by visualising the spectral, spatial, and temporal information that the model used to make a decision.

4.3.1 Spectral Information

As depicted in **Figure 4.4**, the first 2D convolutional layers in my model were applied over the time axis of multi-channel EEG with varying kernel sizes. Adding an extra dimension to the input signal preserved for convolutional channels, these initial 2D convolutional layers act as frequency filters and extract the signal in different informative sub-bands. To further elaborate on this, assume the input signal in each electrode $X[n] \in \mathbb{R}^{1\times S}$, where s is the number of samples. Since a convolutional layer is simply the dot product of the input signal and a kernel, the output $X'[n] \in R^{1\times S}$ of the initial vanilla 2D convolutional layer with a

kernel K of size $2l + 1$ and "same" padding [47] over each EEG channel can be calculated as follows:

$$X'[n] = \sum_{i=-l}^{l} X[i]K[i - n] \tag{4.8}$$

which represents the convolution between the EEG time series of each electrode and the reverse of the kernel $(X[n] * K[-n])$. We also know that the convolution in the time domain acts as multiplication in the frequency domain. Therefore, taking the Fourier transform of convolutional kernels, one can find out the frequency sub-bands selected by the network during the training phase. However, since the deep learning algorithms are sensitive to even tiny discriminative features, the Fourier transform of convolutional kernels can be highly varied. So for the sake of visualisation, I use Savitzky-Golay filtering [263] to smooth the Fourier transforms. Savitzky-Golay is a type of digital filtering consisting of a series of least-square polynomial approximations applied on fixed-length time windows swept over the time series. For a sequence of samples $Y[n]$, the mean-squared approximation error for a time window centred at c can be calculated as follows:

$$E = \sum_{n=c-l}^{c+l} (\sum_{m=0}^{p} a_m n^m - Y[n])^2 \tag{4.9}$$

where $a_m,\ m = 0, \dots, p$ are polynomial coefficients, p is the polynomial order, and l is the half-width of the filter window. To minimise the mean-squared approximation error and find the optimal polynomial coefficients, one should take the derivative of (4.9) with respect to all the polynomial coefficients and set them equal to 0, which yields a set of $p + 1$ normal equations [264]. Accordingly, the polynomial coefficients can be found as follows:

$$A = (D^T D)^{-1} D^T Y \tag{4.10}$$

where $A = [a_0, a_1, \dots, a_m, \dots, a_p]^T$ is the vector of polynomial coefficients. The matrix $D = \{d_{n,m}\}$ is called the design matrix with the size of $2l + 1$ by $p + 1$ and the elements as follows:

$$d_{n,m} = n^m, \qquad c - l \leq n \leq c + l, \qquad m = 0, \dots, p \tag{4.11}$$

The Savitzky-Golay smoothing process can also be explained as a shift-invariant discrete convolution process, which is why it is referred to as Savitzky-Golay filtering (more details can be found in [264]).

4.3.2 Spatial Information

Besides batch normalisation and activation function, my model also contains depthwise convolutional layers after the first 2D temporal convolutions. The depthwise convolutional layer with "valid" padding acts as spatial filtering. It linearly combines signals in different electrodes to transform the electrode domain to the source domain and find discriminative sources. Note that this convolutional layer is applied separately in each convolutional channel, as each convolutional channel represents the signal in a different frequency sub-band. Also, the nonlinear activation function that follows this layer further improves spatial filtering by giving it nonlinearity. Thus, for an input EEG signal $X[n] \in R^{c \times s}$ with c channels and s number of samples in each channel, the transformation from the electrode domain to the source domain $S[n] \in R^{c' \times s}$, with c' being the number of sources, can be formulated as follows:

$$S[n] = WX[n] \tag{4.12}$$

where W is a $c' \times c$ matrix containing spatial filters; however, W in (4.12) represents the unmixing matrix and does not represent the spatial locations of sources. Therefore, one must find and visualise the mixing matrix W^{-1} to find the correct location of sources learned during the training phase.

4.3.3 Temporal Information

In addition to spectral and spatial information, I also aimed to visualise the temporal information in both the source and electrode domains. Since the first two convolutional layers of my model behave as frequency and spatial filters, they map the input multi-channel EEG data from the electrode domain to the source domain. Hence after the training process and learning the optimal network weights, I was able to extract and depict the corresponding source activity for each input sample. Furthermore, I utilised the signal of the closest electrodes to each source to investigate the neural patterns in the electrode domain.

4.4 Motor Imagery EEG-ITNet

In order to show the explainability and reliability of my visualisation technique, I applied the same approach to a novel convolutional network, which I called EEG-ITNet, designed for motor imagery classification. In this section, I introduce EEG-ITNet as a promising motor imagery-based BCI system, and in the next chapter, I provide results supporting the effectiveness of my visualisation technique to bring explainability to similar deep models.

4.4.1 Network Architecture

The general architecture of EEG-ITNet is depicted in **Figure 4.6** and consists of 4 main blocks: inception block, temporal convolution (TC) block, dimension reduction (DR) block, and classification block.

1) *Inception block*: The learning process starts with three parallel sets of layers. Similar to my DPD biomarker discovery system, each set includes a 2D convolutional layer along the time axis, which acts as frequency filtering, followed by a 2D depthwise convolutional layer acting as spatial filtering. Adding inception modules with varying convolutional kernel sizes eliminates the need for a fixed-length kernel [265] and allows the network to learn filters representing various frequency sub-bands. I will later show that the longer the kernel size, the

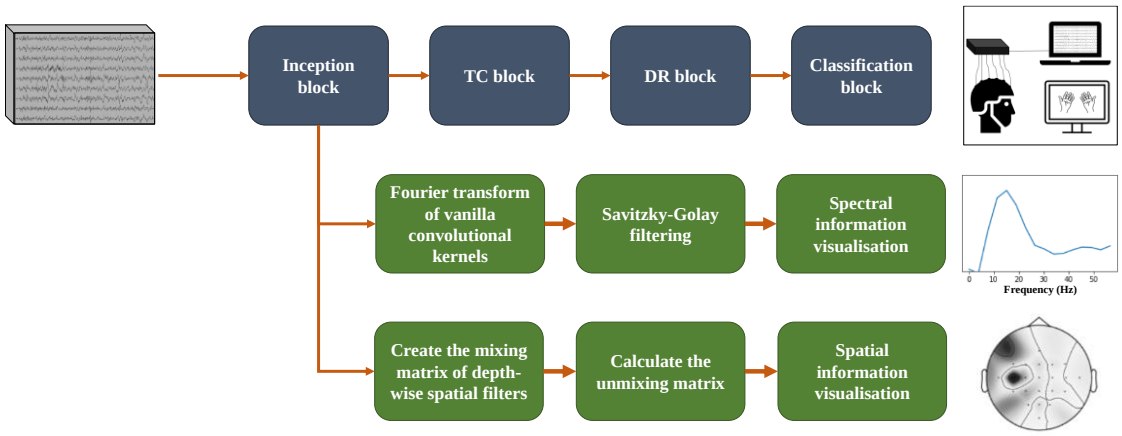

Figure 4.6 General schematic of EEG-ITNet

more likely to learn features in low-frequency components. The kernel size for the depthwise convolution again equals the number of electrodes in the dataset to design a spatial filter that combines all the electrodes to find the sources of brain activity. Hence, the tensor obtained after the inception modules represent the signals of sources in different frequency sub-bands. Finally, this block ends with a nonlinear activation function and dropout to allow the network to learn more complex nonlinear spatial information and avoid overfitting, respectively.

2) *Temporal convolution (TC) block*: After extracting sources in different informative frequency sub-bands, the TCN architecture is applied to extract the discriminative temporal features while taking the history of the time series into account. The TC block consists of several residual blocks, and each is formed by depthwise causal convolutional layers with leading zero padding, followed by activation function and dropout. Using depthwise causal convolution followed by batch normalisation instead of weight normalisation is a modification I made in the original TCN structure (described in section 2.4.2.2). Since the output of the inception block represents signals in the source domain, depthwise causal convolution has been used to ensure that the temporal information of each source is extracted separately. With this modification, my observation showed more robust and accurate performance than the conventional TCN. Depending on the number of previous time steps intended to be considered, the number of residual blocks, kernel length, and dilation base can

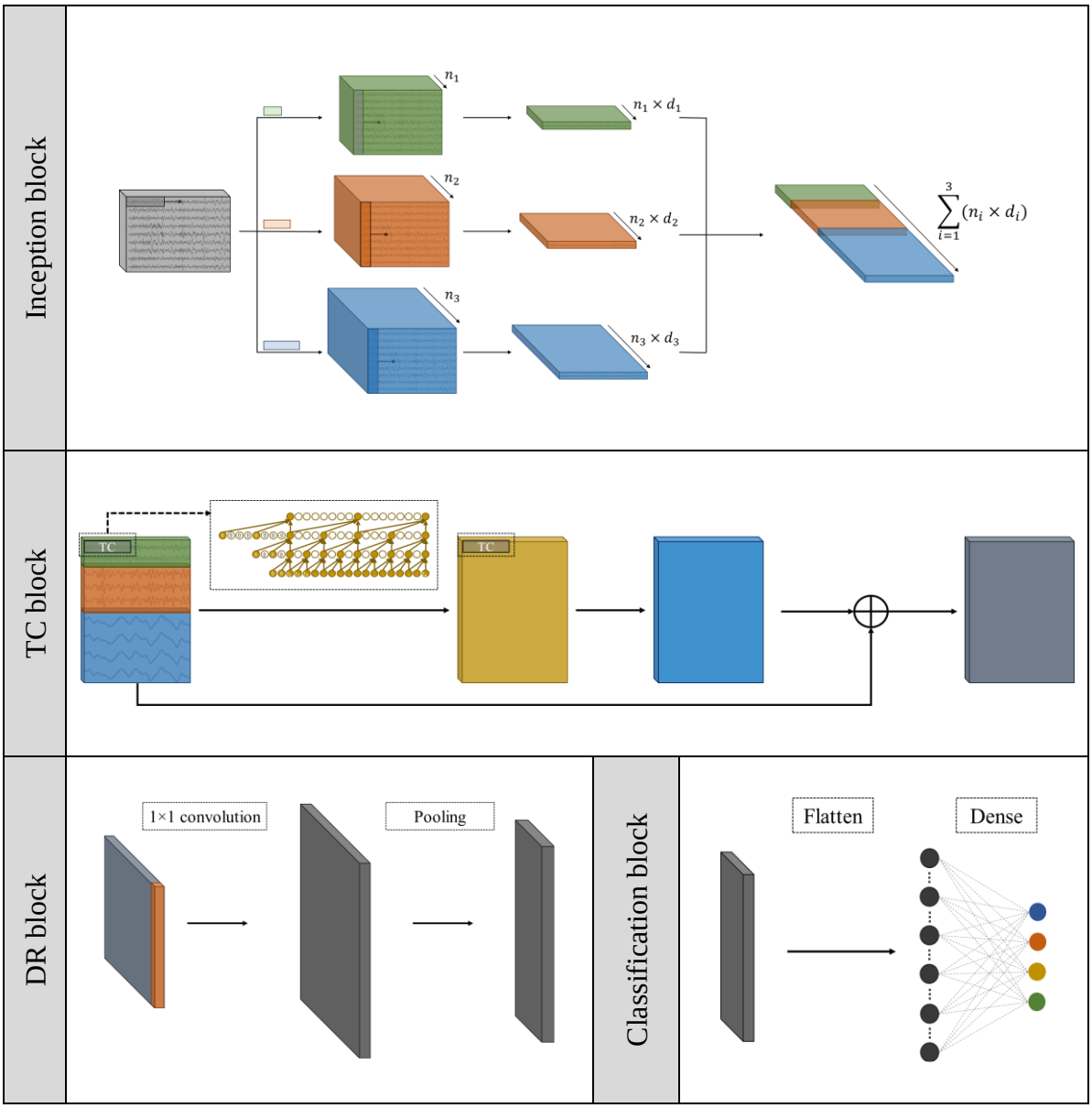

Figure 4.7 Details of different blocks in EEG-ITNet architecture

be chosen based on (2.7). An average pooling layer also precedes this block to reduce the data dimensions and avoid overfitting.

3) *Dimension reduction (DR) block*: The output of the TC block essentially contains temporal information extracted from sources with various frequency spectrums. Thus, I utilised a 1×1 convolutional layer to combine these temporal features and control the number of final features used to perform the classification task. The number of convolutional filters is a hyperparameter and needs to be adjusted to avoid overfitting. As well as an activation

function and a dropout layer, this block also ends with an average pooling layer to further reduce the tensor dimension. Notice that these sets of layers also appeared in my score-guided model in **Figure 4.4**, with the same purpose of reducing input dimension before the concatenation layer and going through clustering and regressions branches.

4) *Classification block*: This block is the final piece of EEG-ITNet and contains a fully connected layer with a "softmax" activation function that follows a flatten layer. Although I call it the classification layer, it can be easily modified depending on the desired output and the problem set.

4.4.2 Dataset and Evaluation Scenarios

Dataset 2a of BCI competition IV and OpenBMI motor imagery dataset [266] were used to evaluate the performance of EEG-ITNet exhaustively. The description of each dataset is as follows:

4.4.2.1 Motor Imagery Datasets

1) *BCI competition IV dataset 2a*: It is a multi-class motor imagery dataset containing EEG recordings of 9 participants during imagination of left hand, right hand, both feet, and tongue movements. The dataset has been collected in two separate sessions, each consisting of 288 trials with an equal number of trials for each class. The actual labels were provided for the data in the first session, while the second session was reserved for testing the classification algorithms in the competition. Each trial in the paradigm started with 2 seconds of preparation time, followed by a cue that lasted for 1.25s, representing the imagination class (**Figure 4.8**). The imagination period continued for 4s after the cue onset and was terminated by an inter-trial break. The EEG signals were recorded using 22 electrodes and sampled with 250Hz. Since I aimed to propose an end-to-end neural network, the signals were only downsampled to 125Hz, scaled to have zero mean and unit variance, and epoched to a 3-

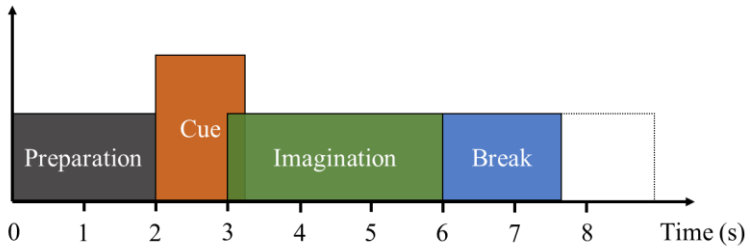

Figure 4.8 Timing scheme of the motor imagery dataset 2a from BCI competition IV

seconds time window after the cue onset. It should also be noted that the 3 EOG channels available in the dataset were excluded from my analysis.

2) *OpenBMI motor imagery dataset*: Collected by Lee et al. [266], OpenBMI is a more extensive EEG dataset containing EEG recordings from 54 subjects across multiple sessions performing three different BCI tasks: motor imagery, ERP, and Steady-state Visually Evoked Potential (SSVEP). In this study, I used EEG recordings from the first session of the motor imagery task, which represent neural activities while performing the imagery task of grasping with either left or right hand. The experiment consists of balanced train and test datasets with 100 trials each. The experimental paradigm of this dataset is depicted in **Figure 4.9**. It starts with 3s of preparation period followed by 4s of imagination period based on the movement direction of the fixation cross. 62 Ag/AgCl electrodes were used to record the signals at 1kHz sampling frequency. Similar to BCI competition IV dataset 2a, the dataset was only downsampled to 125Hz, standardised to have zero mean and unit variance, and epoched to a 3-seconds time window after the cue onset. I should note that similar to what the authors

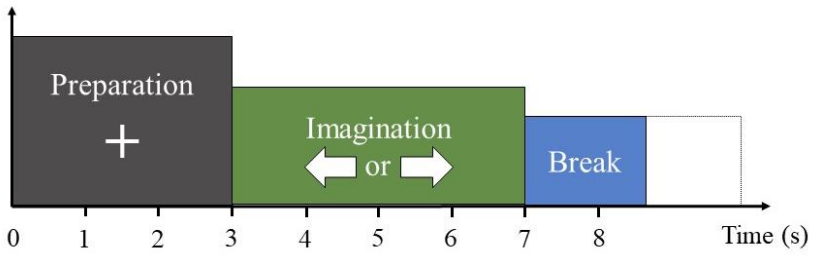

Figure 4.9 Timing scheme of the OpenBMI motor imagery dataset

proposed in the original paper [266], only 20 electrodes located on the motor cortex (FC-5/3/1/2/4/6, C5/3/1/z/2/4/6, and CP-5/3/1/z/2/4/6) were selected for the analysis.

4.4.2.2 Evaluation Scenarios

The performance of EEG-ITNet was evaluated in three different scenarios, corresponding to within-subject, cross-subject, and cross-subject with fine-tuning, detailed below. Note that since the data from the second sessions were used for testing the system and the labels were unknown to the participants in the competition, I did not use them as training or validation sets in my analysis.

1) *Within-subject*: For each subject independently, the EEG recordings of their first session were divided into train and validation sets to find the best parameters of the network. Then, the trained network was assessed on the EEG recordings of the test session, and the results were reported for each subject.

2) *Cross-subject*: The network parameters were learned using the training set of all other subjects rather than the test subjects themselves, meaning the system had never seen any data from the test subject. The subjects preserved for training were divided to form my training and validation sets. Similar to the within-subject case, unique network parameters were obtained for each individual subject in this case.

3) *Cross-subject with fine-tuning*: In the final scenario, the labelled EEG recordings of the training session of each target subject were used to fine-tune the parameters of the system developed for each target subject in the second scenario.

4.5 Classification using Time Series Analysis

To show the importance of biomarker discovery for DPD, I investigated the possibility of performing a classification task between people with a low and high level of depersonalisation

using an electrophysiological biomarker from ERP signals. I used a separate DPD dataset for this purpose, which will be referred to as the auxiliary DPD dataset in my thesis.

4.5.1 Auxiliary DPD Dataset

The dataset collected by [37] was used in my initial investigation to show the importance of biomarker discovery for DPD and the potential of EEG to act as a supporting tool in the diagnostic process using a reliable electrophysiological biomarker. The dataset consists of 29 subjects who initially took the self-rating CDS questionnaire to quantify their level of depersonalisation. According to the authors of [37], I used a threshold of 50, which resulted in 15 subjects being evaluated as a group of participants with a low level of depersonalisation and 14 sex and age-matched subjects as individuals with a high level of depersonalisation for the auxiliary DPD dataset. Since it was my preliminary study evaluating EEG signals as a diagnostic tool for DPD, the subjects with poor SNR or close CDS scores to the threshold were excluded from my analysis. Therefore, the auxiliary DPD dataset eventually contained 10 subjects with high CDS scores and 10 with low CDS scores, referred to as DPD patients and control individuals, respectively.

Figure 4.10 shows the animated schematics of the two distinct categories of trials in the auxiliary DPD dataset. Each trial started with 700ms pre-stimuli in which subjects saw a picture of their own faces and a pencil. Then for 200ms, they received tactile stimulation while they saw a picture of themselves being touched or not being touched (half of the trials for each condition) by the pencil. In no-touch trials, the pencil was next to the cheek, so the perceived distance that the pencil travelled would be the same across trials. Each trial ended with 800ms post-stimuli in which subjects again saw their own faces and the pencil. The tactile stimulation was delivered to either left or right cheek (half of the trials for each) to cancel the possible effects of anatomical congruency between the viewed and felt touch. The EEG signals were recorded using a 64-channels cap at 500Hz sampling frequency.

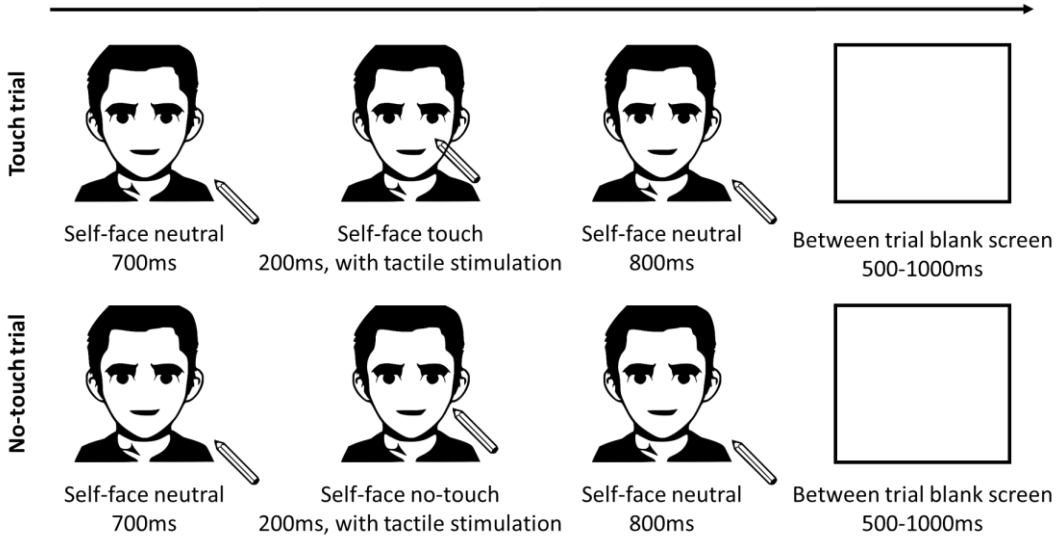

Figure 4.10 Two types of trial design in the experiment of the auxiliary DPD dataset

4.5.2 Extracted Features and Classification Method

In the next chapter, I will present the results of my score-guided biomarker discovery system on my main DPD dataset and will confirm the P45 component of ERP (also found by Adler et al. [37]) as an EEG biomarker for DPD. P45 is an early component of SEP, which appears as a positive peak around 45ms after the stimulus onset and represents the autonomic (implicit) information processing [267]. Later, on my auxiliary DPD dataset, I will demonstrate a difference between the average SEPs over touch and no-touch trials in the control group over P45, missing in DPD patients. Finally, I will use that biomarker to perform a classification task between DPD and control groups using the classification procedure depicted in **Figure 4.11**.

I first calculated the difference signal between touch and no-touch trials for the two groups, which was a positive signal for the control group while a fluctuated signal around zero for the DPD group in the time window associated with P45 (more details in the next chapter). Then I used leave-one-subject-out cross-validation so that in each iteration, one subject was left for testing and the others for training. I calculated the average difference

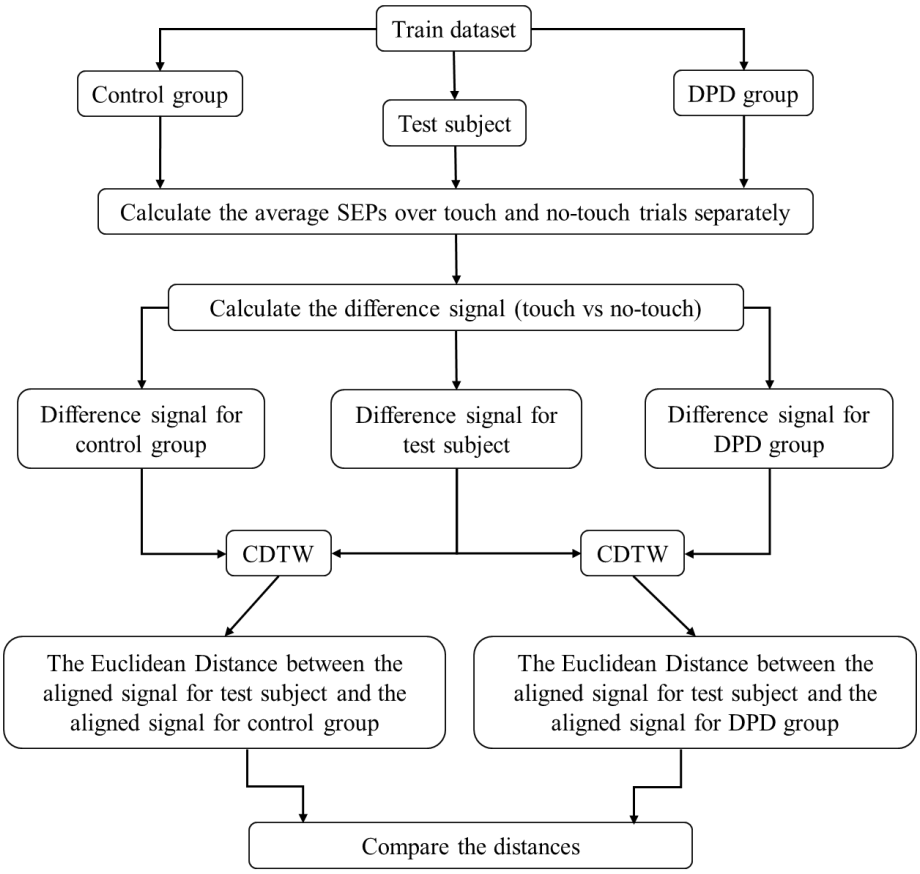

Figure 4.11 Flowchart of the classification procedure

SEPs (touch vs no-touch) for the DPD patients and the control subjects in the training sets and for the test subjects separately. Because the exact timing of the P45 component is not often clearly apparent, I considered a time window between 20 and 50ms after the stimulus to analyse the P45 component. To further address the possible distortion and time shift in the ERP components, including P45, I then used CDTW (explained in detail in section 2.3.1) to find the similarity between the average SEP of the test subject with each group. For instance, **Figure 4.12** shows an example of the extracted signal for control subjects, DPD patients, and a single test subject. It also shows the aligned signals after applying CDTW. For this example, the similarity measure between the signal of the test subject and the control group and the DPD group was 12.4742 and 18.9643, respectively. Therefore, the test subject was assigned to the control group.

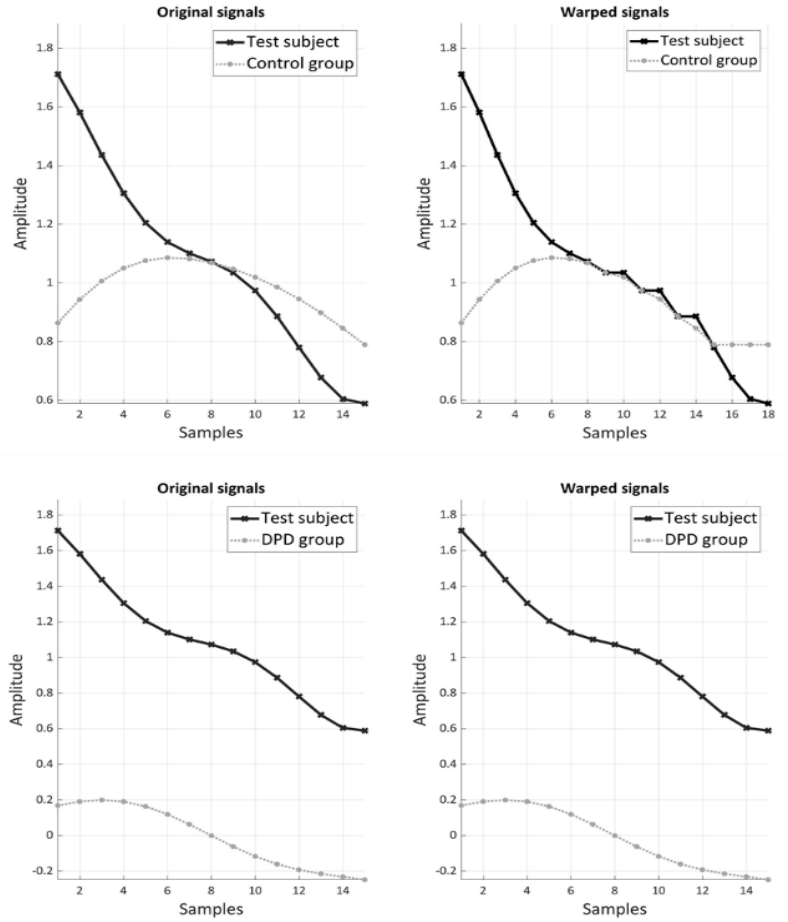

Figure 4.12 Signal alignment for a test subject (C19) in comparison with the average SEP of the control group and the DPD group

5 Applications

This chapter presents all the results and findings from my PhD study, along with discussions supporting the proposed methods' effectiveness. The chapter begins with the evaluation of EEG-ITNet, which I proposed as part of my study to show the usefulness of my visualisation technique on motor imagery datasets. As the primary goal of my PhD research, I will then present the outputs of my score-guided biomarker discovery system, with an exhaustive evaluation of the effectiveness of each step taken to find reliable electrophysiological biomarkers for DPD. I will also discuss the connection between EEG-ITNet learned features with the neural correlates underlying motor imagery tasks as well as the association of discovered DPD biomarkers with its symptoms from a cognitive neuroscientific point of view. Finally, in the last section of this chapter, I present the results from my time series analysis to perform a classification task on a separate DPD dataset using one of the discovered EEG biomarkers.

5.1 Explainability in Motor Imagery BCIs

This section compares the performance of EEG-ITNet to other similar end-to-end deep learning models in motor imagery classification tasks, containing details on the learning process and hyperparameter selection. More importantly, the section comprises a visual representation of the learning process and provides interpretable outputs of the spectral and spatial information learned and used by EEG-ITNet to make decisions in the classification task. Neuroscientific pieces of evidence have been provided supporting the results.

5.1.1 Comparison Models

I selected three similar existing end-to-end architectures for an exhaustive evaluation of the EEG-ITNet performance, including EEGNet, EEG-TCNet, and EEG-Inception. This section explains each of these networks and their implementation in my research.

1) *EEGNet*: As a compact convolutional neural network, EEGNet showed promising results for analysing different EEG datasets [249]. Like DeepConvNet [247], EEGNet starts with temporal and spatial convolutions, followed by a separable convolution.

2) *EEG-TCNet*: Proposed as an extension of EEGNet, EEG-TCNet applied a TCN structure after feature extraction layers of EEGNet to further improve its performance [52].

3) *EEG-Inception*: Designed as a classification tool for ERP-based spellers, EEG-Inception is a CNN architecture that comprises two inception modules and an output module [265]. Each inception module consists of three parallel sets of layers, including 2D vanilla convolutions.

Since my dataset and preprocessing stages are similar to Lawhern et al. [249], I used the same implementation of EEGNet. The EEG-Inception was also implemented based on table IV in [265] as they used a dataset with the same sampling frequency as mine (after downsampling). For EEG-TCNet, the authors introduced their network as an extension to the shallow EEGNet. So to reproduce EEG-TCNet to be compatible with my preprocessing stages, I used the same implementation of EEGNet provided in [249] followed by a TCN structure. The parameters of TCN were selected based on the first formula in [52] as follows: $K_T = 3, L = 2, F_T = 16, p_t = 0.3$. Accordingly, **Table 5.1** summarises a comparison of the key factors of EEG-ITNet with the above deep models.

5.1.2 Training Process and Hyperparameters Selection

The inception block in EEG-ITNet consists of three 2D convolutional layers with 2, 4, and 8 filters and kernel sizes of 16, 32, and 64 samples, respectively, followed by three depthwise

Property / Work	Inception	Dilation	Visualisation	Interpretability	Deepness (# layers)	Parametric complexity
EEGNet 8,2 [14]	No	No	Yes	Partial	Low (7)	Low (~2k)
EEG-TCNet [23]	No	Yes	No	No	Medium (17)	High (~5k)
EEG-Inception [15]	Yes	No	No	No	High (23)	Very high (~15k)
EEG-ITNet	Yes	Yes	Yes	Yes	Very high (31)	Medium (~3k)

Table 5.1 Comparison of the key factors of eeg-itnet with other end-to-end architectures

convolutional layers with a depth of 1. Since the motor imagery input signal's sampling rate is 125Hz, the largest kernel was selected with the size of 64 samples as it can capture frequency components as low as $125/64 \approx 2$Hz of the input signal.

For example, **Figure 5.4** contains the learned frequency filters with the 16, 32, and 64-sample kernels for subject 3 in BCI competition IV dataset 2a. As can be seen, the larger the filter size, the more it is prone to capture lower frequencies. The reason behind selecting more filters for larger kernel sizes is the fact that motor imagery is often associated with activities in the lower alpha (8–12Hz) and beta (13–30Hz) bands [268], rather than high-frequency bands, e.g. gamma (30–100 Hz) band. Besides, the lack of samples in short filter sizes makes it difficult to calculate their Fourier transform and visualise them. Notwithstanding, gamma activity can still be observed during motor imagery with open eyes due to changes in subjects' spatial attention toward the target limb [269].

The inception block is followed by an average pooling layer with a pool size of 4. Afterwards, to extract temporal features in the TC block, I decided to use four residual blocks, each consisting of two convolutional layers, with a dilation base of 2 capable of covering the whole signal history. For this purpose, the TC block requires convolutional layers with a filter size of 4, which can be calculated based on (2.7). Finally, the number of filters in the 1×1 convolutional layer used as a dimension reduction technique was selected to be 14, followed by another average pooling with a pool size of 4. The dropout rate of 0.4 and 0.2 was also picked throughout the network for within-subject and cross-subject scenarios, respectively.

In all scenarios, I first used 10-fold cross-validation to select the best model parameters, with the help of early stopping with the patience of 100 (15 in cross-subject) to avoid overfitting. Then I trained the networks for a maximum of 50 extra epochs with the learning

rate of 10^{-4} on the combined training and validation sets to benefit from the information of all the labelled samples. The number of extra epochs needs to be carefully investigated to ensure no overfitting. For this purpose, I extracted the curves representing train and test losses versus a range of extra epochs used in the training process, concatenated to the loss patterns in the selected best fold of cross-validation. For instance, **Figure 5.1** shows this curve for subject 3 in BCI competition IV dataset 2a while training EEG-ITNet in the cross-subject scenario. Notice that the sudden jump in the curves is due to the fact that in each fold, I split all labelled data to form train and validation sets, while during extra epochs, I use all labelled data as the train set, resulting in a different range of values for the losses. It should be reminded that in the competition, the participants did not have access to the test data labels to evaluate the performance of their proposed algorithm. However, in my investigation, for the sake of fair comparison, I selected the best test accuracy for each deep learning model and each subject, meaning the number of extra epochs varies over models and subjects.

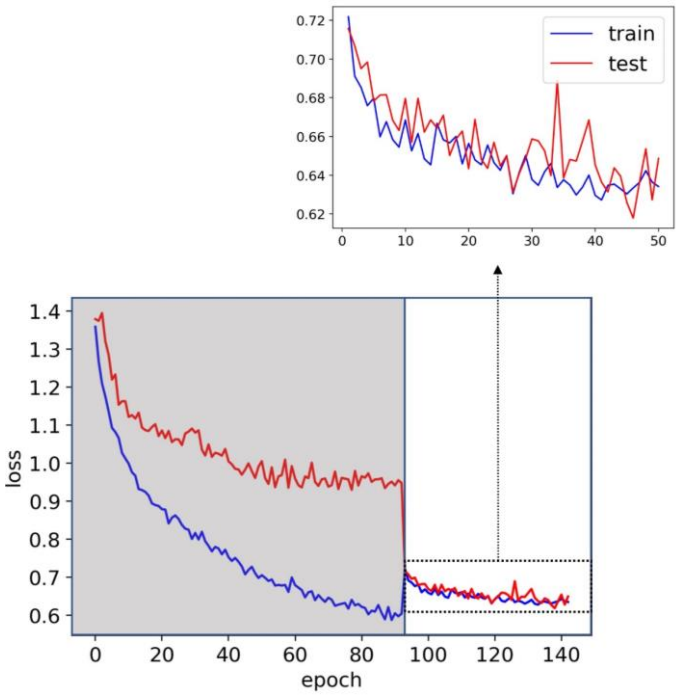

Figure 5.1 The effect of the number of extra epochs in the training of EEG-ITNet for subject 3 from BCI competition IV dataset 2a in the cross-subject scenario. Left (grey box) - training process in the selected best fold of cross-validation. Right (white box) - training process for 50 extra epochs using all labelled data

5.1.3 Performance Evaluation

Table 5.2 summarises the classification accuracies in the within-subject case for EEG-ITNet and my comparison models on BCI competition IV dataset 2a. The performance of EEG-ITNet has been superior to other classification algorithms in terms of mean classification accuracy over all the subjects. Nevertheless, a one-sided Wilcoxon signed-rank test with a significance level of 0.05 was used to verify the significance of improvement [270]. The Wilcoxon signed-rank test is a non-parametric statistical test, especially suitable for small sample sizes. Non-parametric statistical tests generally do not require large sample sizes and make fewer assumptions about the data distribution, including normality. The performance of EEG- ITNet is very promising, reaching the highest accuracy for five out of nine subjects and the second best in three cases.

	EEG-Inception	**EEGNet 8,2**	**EEG-TCNet**	**EEG-ITNet**
S1	77.43 (4th)	81.94 (3rd)	82.29 (2nd)	**84.38 (1st)**
S2	54.51 (4th)	56.94 (3rd)	**64.24 (1st)**	62.85 (2nd)
S3	82.99 (4th)	**90.62 (1st)**	88.89 (3rd)	89.93 (2nd)
S4	**72.22 (1st)**	67.01 (3rd)	60.76 (4th)	69.1 (2nd)
S5	73.26 (2nd)	72.57 (4th)	72.92 (3rd)	**74.31 (1st)**
S6	**64.24 (1st)**	58.68 (3rd)	62.5 (2nd)	57.64 (4th)
S7	82.64 (3rd)	76.04 (4th)	83.33 (2nd)	**88.54 (1st)**
S8	77.78 (4th)	81.25 (2nd)	79.51 (3rd)	**83.68 (1st)**
S9	76.39 (3rd)	78.12 (2nd)	76.39 (3rd)	**80.21 (1st)**
Average	73.50	73.69	74.54	**76.74**
Std	9.11	11.12	10.09	11.48
p-value	**0.043***	**0.010***	0.055	-

***Significant at level of 0.05**

Table 5.2 Performance evaluation for within-subject case in terms of classification accuracy for BCI competition IV dataset 2a

Table 5.3 shows the performance of EEG-ITNet in the cross-subject scenario and its comparison with other end-to-end architectures on BCI competition IV dataset 2a. It also contains the p-value needed to reject the null hypothesis (no difference in the classification accuracies obtained by EEG-ITNet and each comparison model). In this scenario, each model is trained without using any data from the target subject. Therefore, due to the dynamic nature of EEG signals and their variation from one subject to another, a significantly lower performance has been observed in this case. However, my proposed model has reached the highest mean accuracy in the cross-subject case, with statistically significant improvement over EEG-Inception, EEGNet, and EEG-TCNet. **Figure 5.2** also shows this comparison and the data distribution of the results reported in **Table 5.3** using a box plot.

Finally, for each subject, their EEG recordings from the training session were used to fine-tune the parameters of the systems designed for them in the cross-subject scenario. The

	EEG-Inception	**EEGNet 8,2**	**EEG-TCNet**	**EEG-ITNet**
S1	66.32	68.75	69.1	**71.88**
S2	48.26	50	52.08	**62.85**
S3	73.61	80.21	**81.94**	**81.94**
S4	56.6	59.38	61.81	**65.62**
S5	**65.62**	64.24	60.42	63.19
S6	**56.25**	48.26	51.39	**56.25**
S7	73.61	72.57	76.39	**80.21**
S8	70.49	77.43	74.31	**78.12**
S9	61.11	55.56	58.68	**64.93**
Average	63.54	64.04	65.12	**69.44**
Std	8.69	11.59	10.86	8.98
p-value	**0.009***	**0.008***	**0.006***	**-**

***Significant at level of 0.05**

Table 5.3 Performance evaluation for cross-subject case in terms of classification accuracy for BCI competition IV dataset 2a

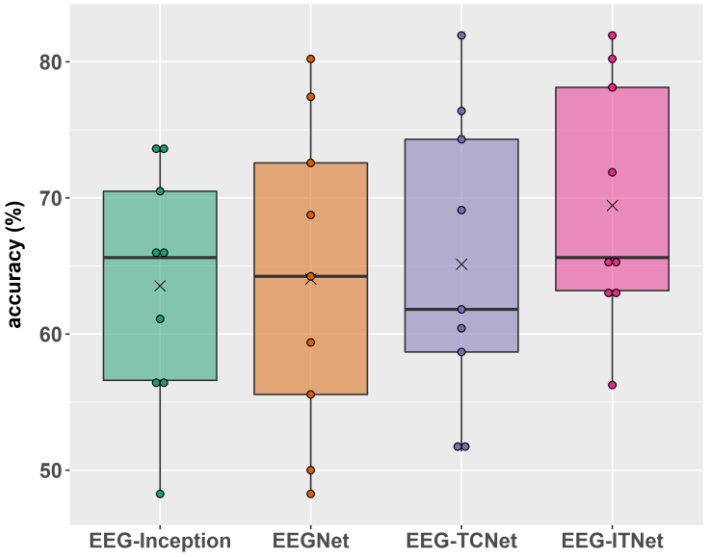

Figure 5.2 Box plot of the classification performance on BCI competition IV dataset 2a in the cross-subject scenario

classification results of this case can be found in **Table 5.4**, which confirm the higher accuracy of EEG-ITNet compared to other deep learning models with statistical significance.

	EEG-Inception	**EEGNet 8,2**	**EEG-TCNet**	**EEG-ITNet**
S1	77.43	84.38	**86.46**	84.03
S2	54.86	54.86	64.93	**65.28**
S3	87.85	**92.36**	90.28	92.01
S4	72.57	67.01	71.53	**73.96**
S5	74.65	66.67	73.26	**75.35**
S6	**66.32**	61.46	58.68	64.93
S7	79.17	79.86	80.56	**84.72**
S8	83.33	82.99	82.99	**84.72**
S9	79.17	75.69	73.61	**83.68**
Average	75.04	73.92	75.81	**78.74**
Std	9.76	12.19	10.24	9.4
p-value	**0.008***	**0.010***	**0.022***	**-**

***Significant at level of 0.05**

Table 5.4 Performance evaluation for cross-subject with fine tuning case in terms of classification accuracy for BCI competition IV dataset 2a

I have also investigated the improvement in the classification results compared to the within-subject scenario in **Figure 5.3**. Comparing the improvement of shallow EEGNet (5 out of 9 subjects with an average improvement of 0.23%) with deep EEG-ITNet (7 out of 9 subjects with an average improvement of 2%) proves the ability of the proposed deeper architecture to handle extra information from other subjects more effective than a shallow one. The improvement is also more significant for complex networks (in terms of the number of parameters) such as EEG-Inception, which showed improvement in 7 out of 9 subjects with an average improvement of 1.54%. However, deep EEG-ITNet is more favoured than the deep and complex EEG-Inception in this area, as it has more potential to be visualised and interpreted.

Because of the high number of subjects in the OpenBMI motor imagery dataset, results have been summarised in **Table 5.5**. The table shows superior performance for EEG-ITNet in all of our cases compared to other deep models in terms of mean classification accuracy. In

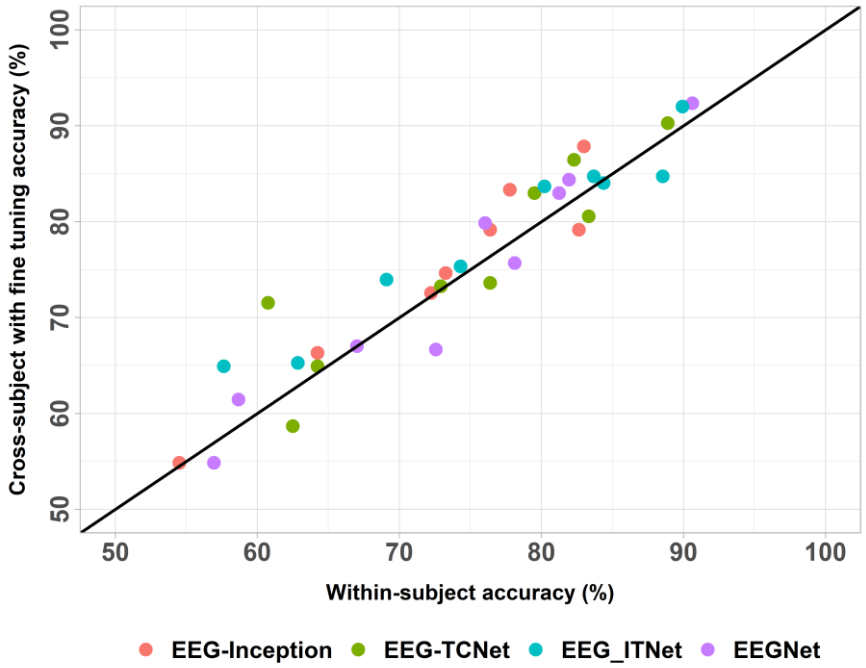

Figure 5.3 Classification improvement in cross-subject with fine-tuning case compared to the within-subject scenario for different deep learning models on BCI competition IV dataset 2a

	Within-subject			Cross-subject			Cross-subject with fine tuning		
	acc	t (n)	*p-value*	acc	t (n)	*p-value*	acc	t (n)	*p-value*
EEG-Inception	69.3	3.88	**<0.001***	71.15	3.71	***<0.001****	75.11	1.52	0.067
EEGNet 8,2	69.61	4.23	**<0.001***	71.2	3.35	***<0.001****	74.04	3.35	**<0.001***
EEG-TCNet	68.35	5.23	**<0.001***	70.83	4.55	***<0.001****	73.85	4.32	**<0.001***
EEG-ITNet	**71.91**	-	-	**73.52**	-	-	**76.19**	-	-

acc: Classification accuracy, t: Test statistic T, n: Degree of freedom = 53
***Significant at level of 0.05**

Table 5.5 Summary of classification performance in different scenarios for OpenBMI motor imagery dataset

order to evaluate the significance of improvement, I performed right-tailed paired t-test after checking the test assumptions. I confirmed data normality by Shapiro–Wilk test [271].

5.1.4 Network Visualisation

Figure 5.4 shows the spectral and spatial filters learned for subject 3 from BCI competition IV dataset 2a in the inception block of EEG-ITNet using the visualisation techniques proposed in section 4.3. This has been selected as an example to show how each set of frequency response and scalp topo-map represents the power spectrum and spatial location of a group of sources activating uniquely among motor imagery classes. Notice that the sign of spatial coefficients does not carry any information about oscillatory synchronisation or desynchronisation processes. The network is trained to perform a classification task by finding discriminative neural patterns among the four motor imagery classes. The spatial activations' location and the absolute value of coefficients are important only as they represent the occurrence and significance of distinct neural activities in at least one of the motor imagery classes. So in order to avoid confusion, the spatial patterns are plotted in grayscale symmetric to zero.

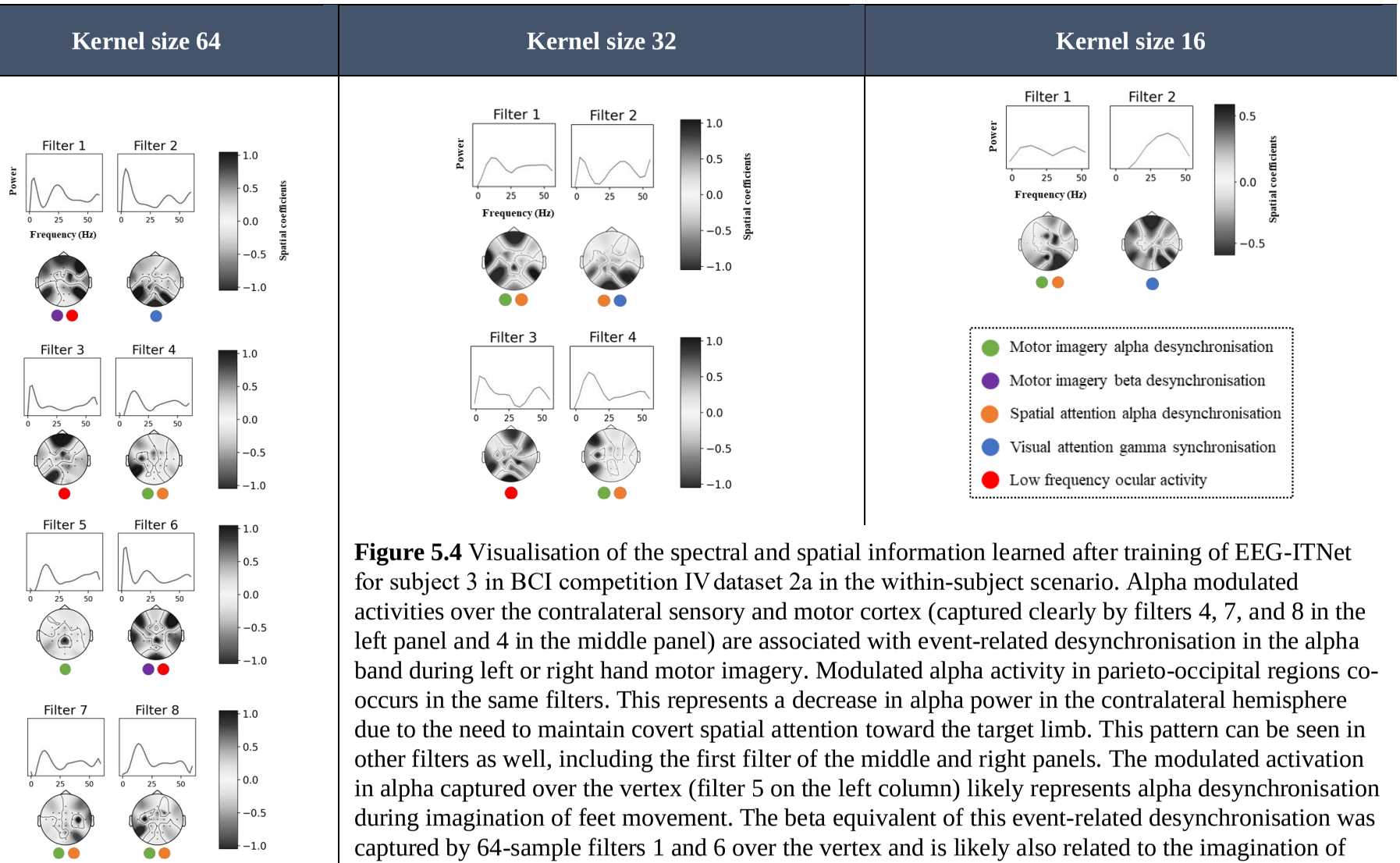

Figure 5.4 Visualisation of the spectral and spatial information learned after training of EEG-ITNet for subject 3 in BCI competition IV dataset 2a in the within-subject scenario. Alpha modulated activities over the contralateral sensory and motor cortex (captured clearly by filters 4, 7, and 8 in the left panel and 4 in the middle panel) are associated with event-related desynchronisation in the alpha band during left or right hand motor imagery. Modulated alpha activity in parieto-occipital regions co-occurs in the same filters. This represents a decrease in alpha power in the contralateral hemisphere due to the need to maintain covert spatial attention toward the target limb. This pattern can be seen in other filters as well, including the first filter of the middle and right panels. The modulated activation in alpha captured over the vertex (filter 5 on the left column) likely represents alpha desynchronisation during imagination of feet movement. The beta equivalent of this event-related desynchronisation was captured by 64-sample filters 1 and 6 over the vertex and is likely also related to the imagination of foot movement. The low-frequency activations (< 5Hz) over frontal areas such as in filters 1, 3, and 6 of the left panel and filter 3 in the middle likely represent eye blinks and movements

5.1.5 Conclusion

As part of my PhD research, I proposed EEG-ITNet, an explainable CNN architecture based on inception modules and causal convolutions with dilation. Inception modules were proposed to eliminate the need to use fix-length kernels, allowing the network to learn input data patterns at different scales. In addition, causal convolutions with dilation were employed as an alternative to RNNs to learn and extract temporal dependencies and information in EEG time series. The series of depthwise causal convolutions with dilation in the form of residual blocks extracted informative discriminative features to perform the classification task in such a way that allowed EEG-ITNet to outperform well-known existing end-to-end neural networks in different scenarios. Another significance of EEG-ITNet is its less complex structure (in terms of the number of trainable parameters) compared to other existing end-to-end architectures, including EEG-Inception and EEG-TCNet.

This part of my research aimed to apply my feature visualisation technique to EEG-ITNet and examine the potential of my approach in bringing explainability to this model, which has similar initial layers as my score-guided biomarker discovery system. I discussed the association between the network's learned spectral and spatial patterns with expected neural patterns during motor imagery tasks. Among other end-to-end comparison models, only the authors of EEGNet have attempted to explain and visualise the learned features. However, in my study, the Fourier transform combined with Savitzky-Golay filtering offered more interpretability and accuracy for network visualisation (**Figure 5.5**). For instance, the weights of the convolutional layers applied along the time axis in EEGNet were translated as representing a single frequency component each. However, it is unlikely for the neural patterns of motor movements or mental tasks to be generated from a single frequency component. That is why the reader may find the smoothed power spectrum of sources generated in EEG-ITNet more intuitive and easier to interpret. Besides, the authors of

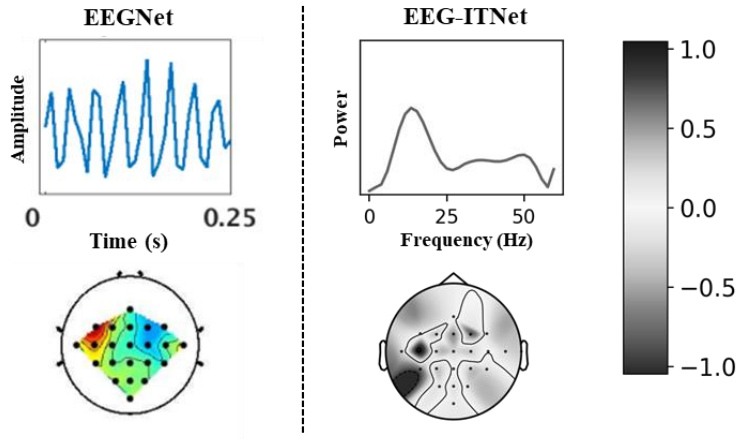

Figure 5.5 Comparison of the outcome of EEGNet (left) and EEG-ITNet (right) feature visualisation techniques. In EEGNet, frequency information can be calculated based on the number of detectable cycles in the learned temporal kernel window, resulting in a single frequency component. So, for example, the above kernel from EEGNet represents 32Hz as it contains 8 cycles in 0.25 seconds.

EEGNet have made no attempt to explain the spatial filters from a neuroscientific viewpoint. Hence, we discussed the validity of the extracted features and supported them from a neuroscientific perspective. My discussion was based on a fair assumption that motor imagery in our selected experiment relies on lateralised changes in the spatial attention toward the target limb and the imagination of motor execution.

5.2 DPD Biomarker Finding

In this section, I extensively evaluate the significance of my score-guided biomarker discovery model to find and propose reliable biomarkers for DPD. I start the section by showing the impact of each component added to my model to improve its performance. The section also describes how to interpret learned features to exploit EEG biomarkers.

5.2.1 The effect of $\mathcal{L}_{link}$

I proposed a third term in my loss function for my multi-task learning scenario called $\mathcal{L}_{link}$ as I argued that only using the weighted sum of $\mathcal{L}_{regression}$ and $\mathcal{L}_{clustering}$ as the loss function would make the network highly prone to overfit and result in finding uninformative and

meaningless features in my case. To show the effect of $\mathcal{L}_{link}$, I investigated and compared the low-dimensional representation of the multi-channel EEG signals obtained from two networks with and without the $\mathcal{L}_{link}$ term in their loss function in **Figure 5.6**. For the sake of visualisation, I used the two strongest principal components of the learned low-dimensional representations. **Figure 5.6** shows how $\mathcal{L}_{link}$ forces the network to find clusters representing the two groups of people with a low and high level of depersonalisation. Furthermore, **Figure 5.7** shows the clustering loss values in the training process of those two networks and confirms the effect of $\mathcal{L}_{link}$ on preventing overfitting.

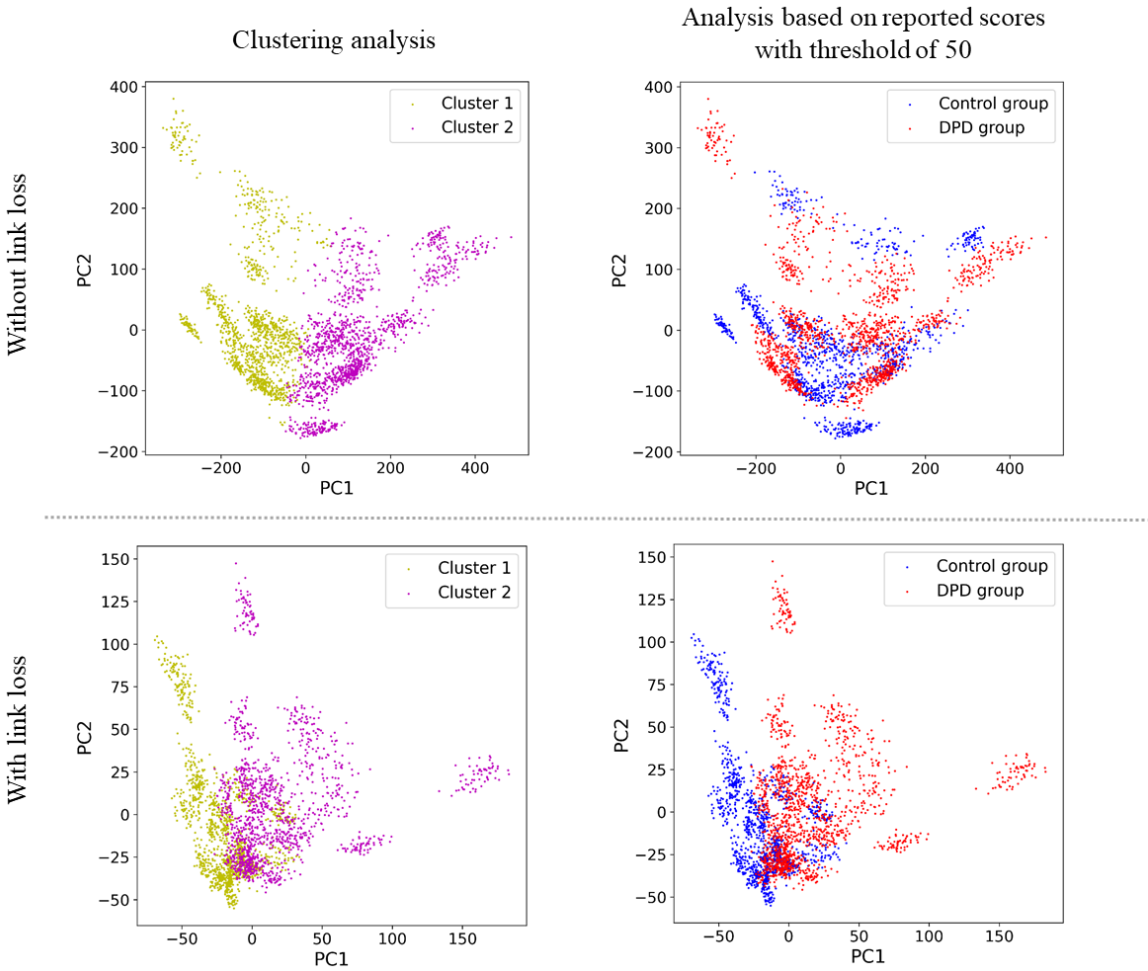

Figure 5.6 The scatter plots of the two strongest principal components of the learned low-dimensional representation of the input multi-channel EEG signals. The illustrations have been derived after training without (top) and with (bottom) link loss. The grouping is based on the output of the clustering layer (left) or reported CDS scores (right).

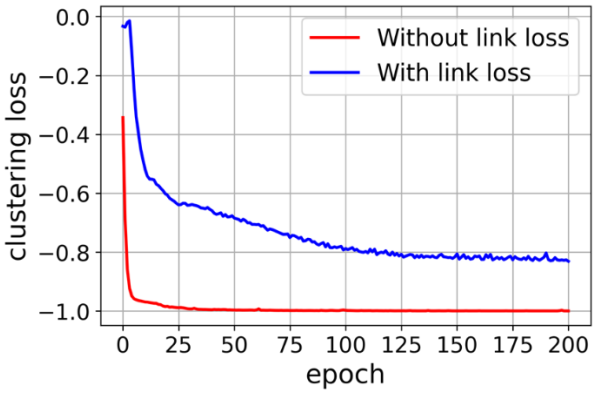

Figure 5.7 The clustering loss value during the training process with and without the link loss term.

One of the objectives of my research was to address the unreliability of CDS scores as a diagnostic metric. The outcome of my proposed deep multi-task learning model with the parameters in **Table 5.6** formed two clusters representing the DPD and the control groups.

Hyperparameter name	Hyperparameter value
Number of filters in temporal convolution (F_1)	4
Number of filters in depthwise convolution (d)	2
Number of filters in separable convolution	8
Temporal convolution kernel size ($1 \times s$)	1×128
Separable convolution kernel size	1×32
First average pooling factor (p_1)	2
Second average pooling factor (p_2)	4
Dropout rate	0.2
Batch size	32
Learning rate	10^{-4}
Early stopping patience	20
Number of generated trials using resampling-average method	60
$w_{regression}$	1.058×10^{-4}
$w_{clustering}$	3.333×10^{-4}
w_{link}	6.666×10^{-4}

Table 5.6 Selected hyperparameters for the proposed deep model

After comparing the subjects assigned to each cluster with their original classification based on the reported CDS scores, I noticed that the discrepancy between my clustering analysis and original grouping was two DPD subjects that my clustering analysis assigned to the cluster representing the control group. In other words, the classification of the subjects based on my clustering analysis differed slightly from the one based on reported CDS scores. Accordingly, to investigate my method's superiority and accuracy over the conventional diagnosis based on CDS scores, I calculated a point-biserial correlation coefficient [273] between the outcome of clustering and original classification with some additional data available from the participants, such as their depression and anxiety scores. **Table 5.7** shows that the outcome of clustering analysis shows a higher correlation with some psychological factors that are highly comorbid with depersonalisation [260] compared to the original classification based on the reported CDS scores.

Psychological factors	Grouping based on reported CDS scores		Grouping based on clustering analysis	
	r_{pb}	*p-value*	r_{pb}	*p-value*
Total CDS score	0.78	< 0.001	0.79	< 0.001
Anomalous body experience factor of CDS	0.66	< 0.001	0.67	< 0.001
Emotional numbing factor of CDS	0.73	< 0.001	0.72	< 0.001
PHQ-9 depression test score	0.29	0.061	0.31	0.041
Cognitive anxiety factor of STICSA	0.42	0.005	0.49	< 0.001
Somatic anxiety factor of STICSA	0.28	0.072	0.33	0.029
Self-object differentiation subscale of the Operationalized Psychodynamic Diagnosis-Structure Questionnaire (OPD-SQ) score	0.56	< 0.001	0.6	< 0.001

Table 5.7 Point-biserial correlation analysis for grouping based on reported CDS scores and our proposed learning method with participants' psychological factors.

5.2.2 Biomarker Extraction

In order to find potential biomarkers for DPD diagnosis, I aimed to visualise spectral, spatial, and temporal information obtained after training the proposed deep model in **Figure 5.8**. Notice that my deep learning pipeline consists of two parallel branches, with the same type of layers, to simultaneously analyse trials associated with synchronous visual-tactile stimulation to the participant's left and right hand. Using the visualisation technique I proposed in 4.3, I illustrated the power spectrum and the spatial location of sources that contributed the most during my multi-task learning process to simultaneously perform the clustering task and predict CDS scores. Therefore, any sources in **Figure 5.8** can serve as potential electrophysiological biomarkers. Notice that the spatial activities depicted for each source in

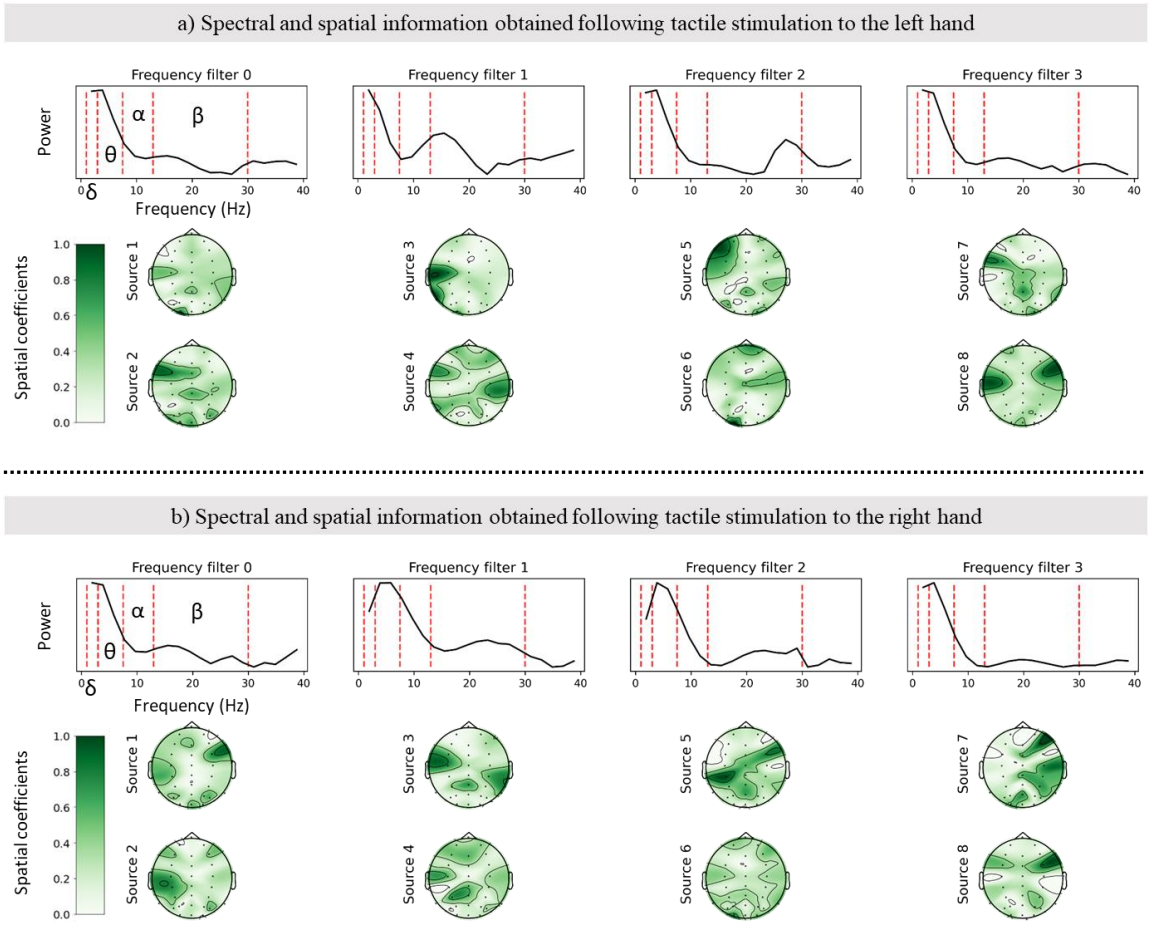

Figure 5.8 Spectral and spatial information of potential electrophysiological biomarkers obtained from analysing synchronous visual-tactile stimulation to the left (a) and right (b) hands.

Figure 5.8 can represent the combination of more than one source. That is why spatial activities can be observed in multiple spatial locations for each source. The power spectrum displayed above every two sources indicates the response of the frequency filter learned to extract those sources, with the red dashed lines separating EEG waves, including delta, theta, alpha, and beta.

Each of the sources in **Figure 5.8** can serve as a potential electrophysiological biomarker. Since each of those topoplots might represent the spatial activities of a group of sources, I may choose sources that were sparse and showed stronger spatial activity for further investigation. Therefore, for the left touch trials (**Figure 5.8-a**), I focused on sources 3 and 8. By extracting the output of depthwise 2D convolutional layer (see **Figure 4.4**) from our network after the training process, we can also display the temporal activity of those sources for their corresponding inputs.

Accordingly, **Figure 5.9** shows the spectral and spatial features of sources 3 and 8 derived by analysing synchronous visual-tactile stimulation to the participant's left hand along with their average temporal activities taken on all the trials for each group of people with a low and high level of depersonalisation. As participants' grouping differed based on my clustering analysis and reported CDS scores, I present the average temporal activities for both. The shaded region shows the standard deviation of activities. Any time window with non-zero temporal activity indicates the temporal characteristic of the potential biomarker. In other words, by looking at the illustrations in **Figure 5.9-a** for source 3, one may propose a hypothesis that there is an EEG biomarker for depersonalisation (higher activation for DPD patients based on CDS score and for Cluster 1) during synchronous visual-tactile stimulation to the participant's left hand with a **delta** (and **high alpha/low beta**) component over the left centro-temporal lobe (around **channel C3**) in the time window encompassing the **P300** ERP component.

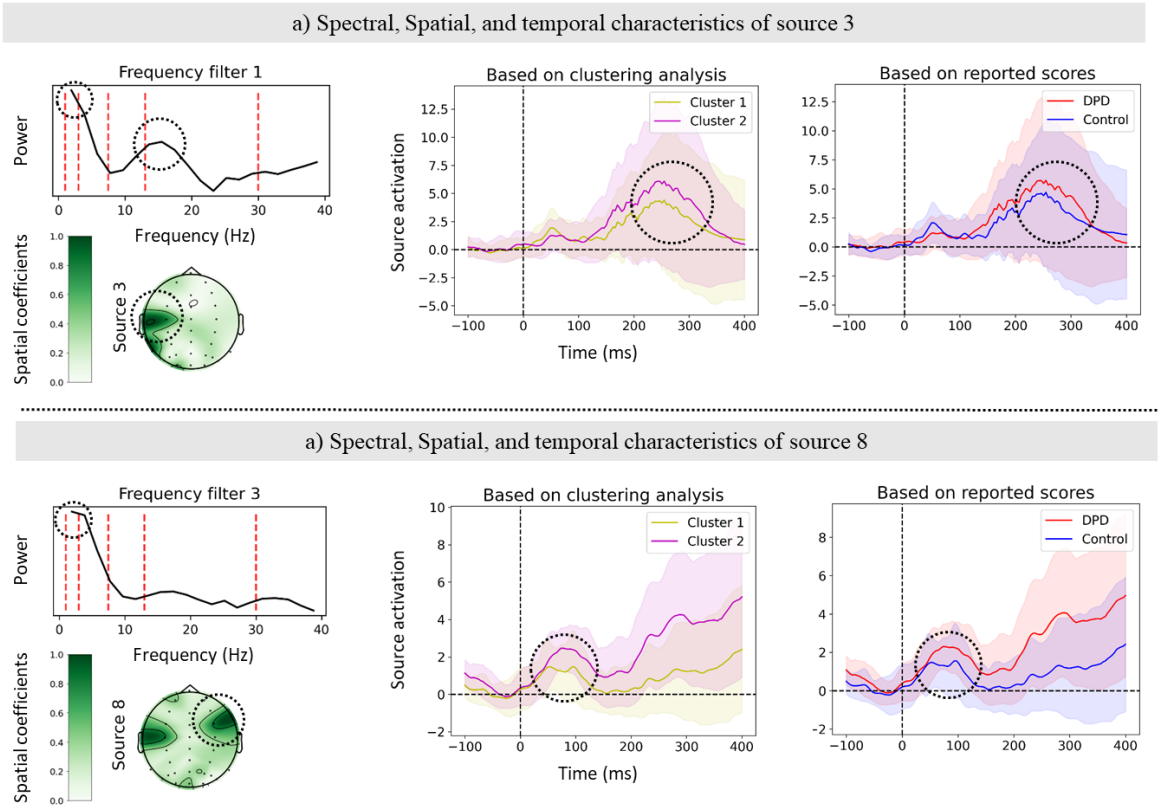

Figure 5.9 Spectral, Spatial, and temporal characteristics of sources 3 and 8 obtained from analysing synchronous visual-tactile stimulation to the participant's left hand. The temporal response slightly differs based on clustering and reported CDS scores grouping. Dashed circles indicate characteristics of potential biomarkers.

In order to statistically investigate the proposed hypothesis, I first visualised the average ERP responses to synchronous visual-tactile stimulation to the participant's left hand over channel C3, which is the closest electrode to the spatial characteristics of source 3 in **Figure 5.9**. **Figure 5.10** shows the average ERPs based on my clustering analysis and reported CDS scores. Both plots in **Figure 5.10** clearly show a difference around the P300 component. Nevertheless, after checking the test assumptions, such as normal distribution and homogeneity of variance, I performed an independent samples t-test on the P300 average amplitude (average ERP in the time window of 260-360ms post-stimulus) to investigate the significance level. The result was significant, with 95% confidence for both groupings, revealing a stronger effect for grouping based on clustering analysis ($t(41)=2.57$, $p=0.014$,

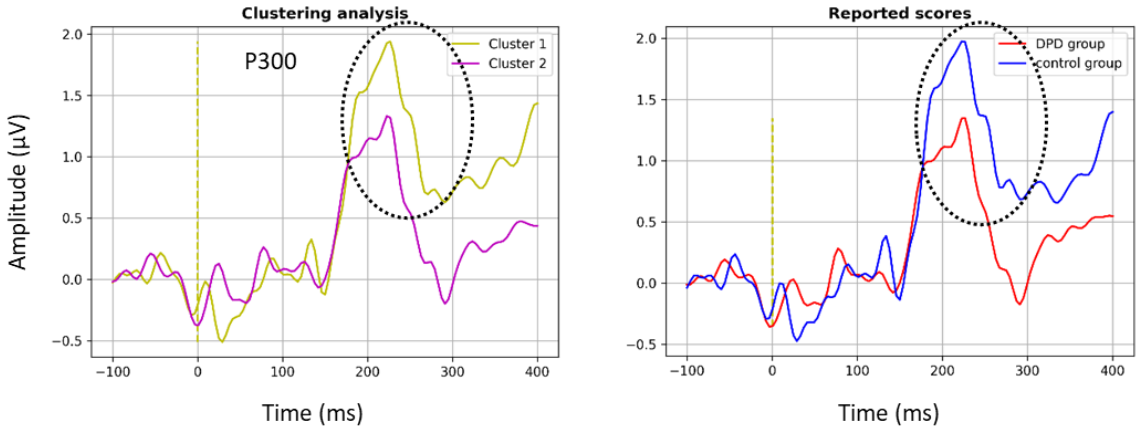

Figure 5.10 Average ERP responses to synchronous visual-tactile stimulation to the participant's left hand over channel C3. Clusters 1 and 2 represent DPD and control groups, respectively.

Cohen's d=0.80, 95% CI [0.14, 1.13]) than based on reported CDS scores (*t*(41)=2.23, *p*=0.032, *Cohen's d*=0.70, 95% CI [0.05, 1.08]).

Similarly, one can propose other hypotheses by looking at spectral, spatial, and temporal characteristics of sources obtained in our learning process. As another example, looking at source 8 in **Figure 5.9**, we can assume that there is a potential biomarker in response to synchronous visual-tactile stimulation to the participant's left hand with delta component over channel C3 or C4/FC4 in the time window around early P45, P100 and later P300 ERP components. Notice that source 8 in **Figure 5.9** represents the combination of two sources; one of them (P300 component over C3) overlaps with source 3 investigated earlier. As a result, one can form their second hypothesis as an **early P45/P100** component cluster in the **delta** range activating over channel **C4/FC4** contralateral to the left-hand stimulation. In order to explore the second hypothesis further, I again aimed to visualise the time responses in the electrode domain over C4/FC4, and the results can be seen in **Figure 5.11**. I also statistically evaluated the second hypothesis and confirmed a significant difference in P45 average amplitude (time window of 40-70ms) with 95% confidence between DPD and control groups over channel C4/FC4.

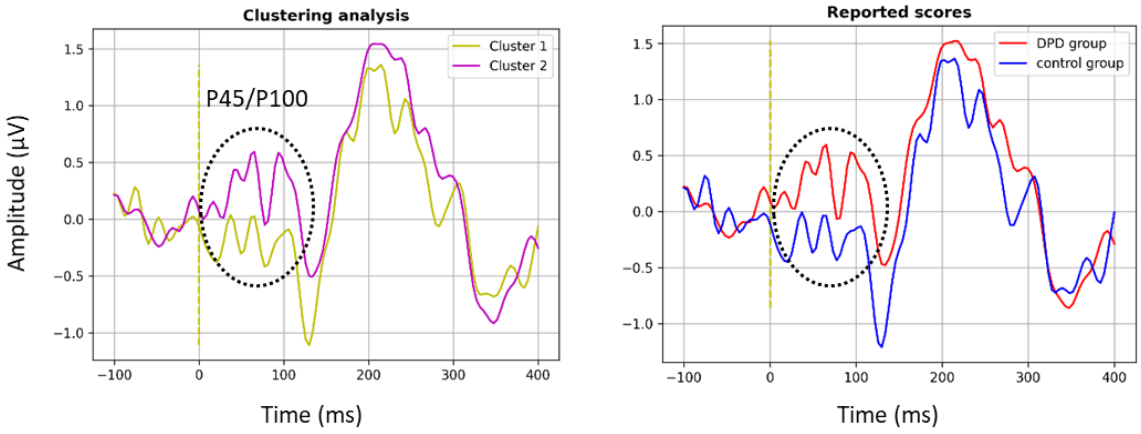

Figure 5.11 Average ERP responses to synchronous visual-tactile stimulation to the participant's left hand over channel C4/FC4. Clusters 1 and 2 represent DPD and control groups, respectively.

I should note that the group differences in the proposed biomarkers might be statistically more significant if we perform the t-test on the filtered ERP components based on the spectral characteristics of biomarkers. However, I did not investigate it as the length of each trial was only 500ms, resulting in high signal distortion following a bandpass filter on each trial. However, since my deep processing pipeline can be used for any similar psychological disorder assessed by clinical assessment scores, a bandpass filter based on the spectral map of learned features is highly recommended if a more extended time window is available.

The activity of P45 over the contralateral somatosensory cortex concluded from my second hypothesis (using source 8 in **Figure 5.8-a**) was also captured by source 3 in **Figure 5.8-b** for synchronous visual-tactile stimulation to the participant's right hand, confirming the above component as a promising biomarker for DPD study and diagnosis. Therefore, using the same approach, more hypotheses can be inferred and investigated as potential electrophysiological biomarkers for DPD. The summary of discovered biomarkers and their evaluation can be found in **Table 5.8**. The statistical results reported in this table are based on the clustering analysis grouping. A summarised comparison of all discovered biomarkers between the control group and DPD patients is also illustrated using boxplots in **Figure 5.12**.

Stimulated hand	Electrode	ERP component (Time window)	t-statistic	p-value	Effect size (Cohen's d)	95% confidence interval
left	C3	P300 (260-360ms)	2.57	**0.014***	0.80	[0.14, 1.13]
left	C4/FC4	P45 (40-70ms)	-2.78	**0.008***	0.87	[-0.99, -0.16]
right	FC4	P200 (180-280ms)	2.29	**0.027***	0.71	[0.09, 1.46]
right	C3/CP3	P45 (40-70ms)	-2.40	**0.021***	0.75	[-0.83, -0.07]
right	PO8	P200 (180-280ms)	-3.00	**0.005***	0.93	[-2.75, -0.53]

* Show significance at 0.05 level
Degree of freedom = 41

Table 5.8 The summary of all the biomarkers identified using the proposed deep multi-task learning model for our DPD dataset and their statistical evaluation between DPD and control clusters.

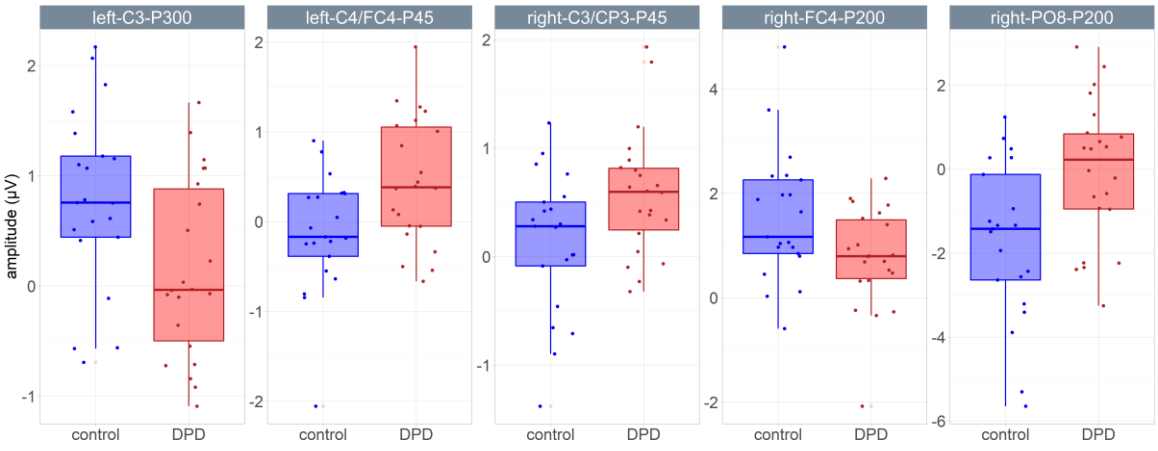

Figure 5.12 The component amplitude comparison between the control and DPD groups using boxplots for each biomarker discovered in **Table 5.8**.

5.3 DPD Diagnosis

In the final stage of my PhD research, I showed how we could employ EEG biomarkers found by my score-guided biomarker discovery system to perform a classification task between individuals with low and high levels of depersonalisation. Notice that the dataset used to train my network (main DPD dataset) was different from the dataset I used to perform the classification task (auxiliary DPD dataset).

5.3.1 Illustration of Abnormal P45 Activation

One of the biomarkers detected by my network was P45 activation over the contralateral somatosensory cortex (see **Table 5.8**). Adler et al. [37] also showed a difference in the activation of the P45 component in DPD patients compared with the control group. Accordingly, **Figure 5.13** shows the average SEPs over touch (thick line) and no-touch trials

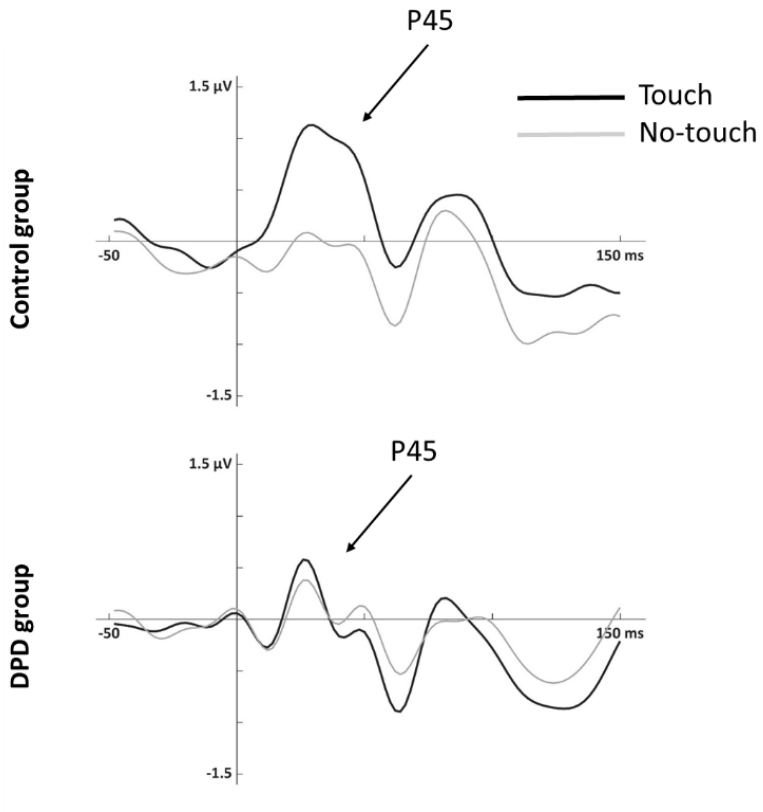

Figure 5.13 Average SEPs for touch and no-touch trials in a cluster of electrodes located in the somatosensory cortex (electrodes marked in **Figure 5.14-B**)

(thin line) for DPD patients and the control group over the somatosensory cortex (centro-parietal regions depicted in grey in **Figure 5.14-B**). Since the P45 activation found by my network was contralateral to tactile stimulation, for stimulation to the left cheek (right cheek), I calculated SEP on the average of selected electrodes in the right (left) hemisphere. Then I obtained the average SEP response over both hemispheres. As can be seen in **Figure 5.13**, there is a difference in the activation of P45 components in DPD patients compared with the control group. As I discussed in section 3.1, the lack of P45 activation during touch trials represents impairment in early information processing in DPD patients [148, 201]. Furthermore, since there is a link between P45 and the sense of body ownership [37, 214, 215], the observed lack of P45 can be associated with feelings of disembodiment in the DPD group [37].

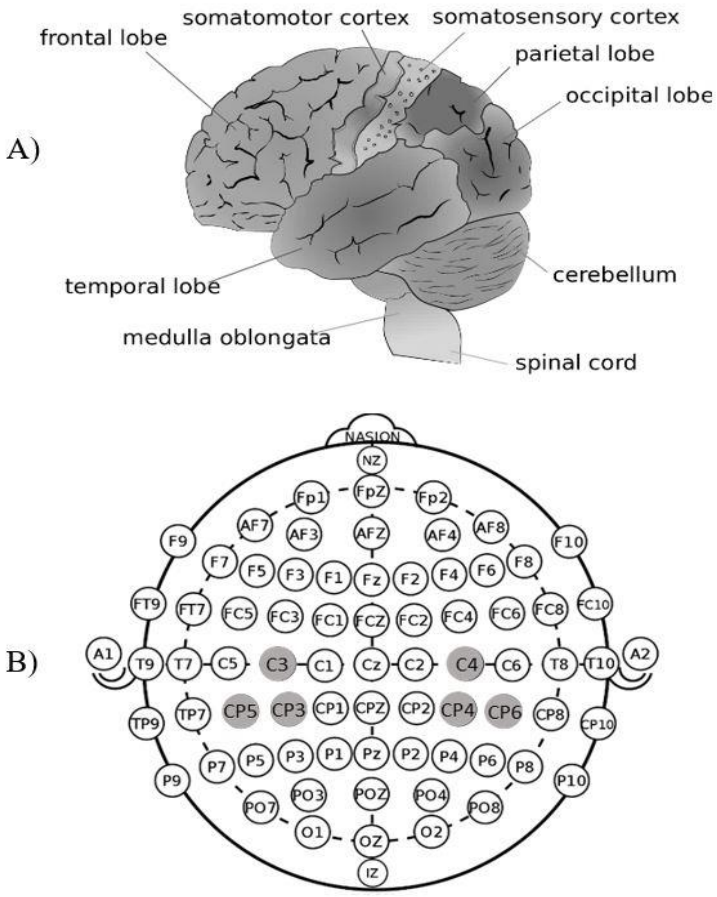

Figure 5.14 Different brain regions (A) and the schematic of 64-channels cap (B) - filled in grey channels are the ones used to calculated SEPs

5.3.2 Classification Results

Following the classification algorithms discussed in section 4.5.2, **Table 5.9** shows the classification results using leave-one-subject-out cross-validation. **Table 5.9** also contains the

ID	True class	CDS score	Similarity measure with the control group	Similarity measure with the DPD group	Result
C10	Ctrl.	10	3.9061	3.9663	Ctrl.
C15	Ctrl.	9	1.4856	18.0678	Ctrl.
C17	Ctrl.	18	4.1491	29.6794	Ctrl.
C19	Ctrl.	19	12.4742	18.9643	Ctrl.
C2	Ctrl.	19	0.1501	9.7086	Ctrl.
C20	Ctrl.	12	0.1976	9.8148	Ctrl.
C4	Ctrl.	4	4.2445	4.9838	Ctrl.
C66	Ctrl.	19	7.6045	28.8936	Ctrl.
C70	Ctrl.	11	19.0275	48.8561	Ctrl.
C71	Ctrl.	17	10.1552	35.2177	Ctrl.
D11	DPD	71	26.3729	0.7552	DPD
D14	DPD	199	11.6940	5.3986	DPD
D15	DPD	101	19.7699	1.0860	DPD
D18	DPD	72	83.6372	34.2489	DPD
D3	DPD	91	23.1570	1.1153	DPD
D5	**DPD**	99	8.6665	9.4076	**Ctrl.**
D6	DPD	140	67.6014	24.0648	DPD
D9	DPD	97	4.4002	2.3384	DPD
D13	**DPD**	111	2.6640	10.8761	**Ctrl.**
D69	**DPD**	98	3.0716	26.8930	**Ctrl.**
Number of subjects correctly classified				17	
Total number of subjects				20	
Accuracy % (Kappa value)				**85% (0.7)**	

Table 5.9 Participants CDS score and the classification results using proposed electrophysiological biomarker (P45)

similarity measures between the signals for each subject and the two classes. I was able to reach 85% classification accuracy on the auxiliary DPD dataset. In addition, I asserted that the signals have a distortion and time shift. Thus, to show the effectiveness of using CDTW in my scenario, I also excluded CDTW and calculated the number of correctly classified subjects based on the Euclidean distances (Euclidean distances can be calculated since all the time series have the same number of samples) on original SEP signals without alignment. In addition, one other possible approach would be to consider the peak value in the chosen time window as the P45 component. If the value for the signal peak in the time window was closer to the signal peak of the control group, it would be assigned to the control group; otherwise, the DPD group. My proposed classification procedure is compared with the above methods in **Table 5.10**.

5.3.3 Conclusion

As an initial investigation, I aimed to explore the discovered P45 component of ERP as a potential feature to discriminate between individuals with high and low DPD symptoms. I first showed that although P45 was extracted from my score-guided biomarker discovery system trained on our main DPD dataset, the lack of P45 was also evident in the auxiliary DPD dataset. I used P45 as a feature to perform the classification task using time series analysis. In this regard, CDTW was applied to tackle the problem of possible time shifts and distortion in signals around P45. I reached 85% accuracy (Kappa value of 0.7), which shows

Classification strategy	**Number of correctly classified subjects**	**Classification accuracy (%)**
Using peak amplitude	15	75
Without signal alignment	16	80
Proposed approach	**17**	**85**

Table 5.10 Comparison of my classification approach with other possible approaches

the importance and effectiveness of EEG signals and a reliable biomarker to act as a diagnostic tool.

.

6 Discussion

In this chapter, I provide neuroscientific evidence supporting the validity of discovered neural patterns for both motor imagery and DPD datasets in my analysis.

6.1 Motor Imagery Biomarkers

As shown in **Figure 5.4**, larger filter sizes capture relatively lower frequencies, while smaller filters tend to learn higher frequency components. To explain the learned filters, one should consider the dataset to be composed of two tasks, including motor imagery and sustained spatial attention toward the target limb. Motor imagery is associated with a decrease in alpha and beta power (event-related desynchronisation) over the sensory and motor cortex in the contralateral hemisphere [268, 274, 275]. This modulated alpha activity has been clearly captured by filters 4, 5, 7, and 8 in the left panel of **Figure 5.4** and filter 4 in the middle panel. Notice that among these filters, spatial activation is located at the vertex only in filter 5 of the left panel, which is likely to be associated with the imagination of bilateral foot movement. In other cases, the filters represent modulated neural patterns during left or right hand motor imagery located in the contralateral hemisphere. The beta equivalent to the alpha desynchronisation seen over the vertex (filter 5 of the 64-sample kernel) during motor imagery was also captured by 64-sample filters 1 and 6, which may be similarly related to bilateral foot movement imagination.

The spatial activation in the parieto-occipital regions in the alpha band witnessed by filters 4, 7, and 8 of the left panel may be associated with the need to maintain spatial attention toward the target limb. Studies have shown that covert spatial attention can lead to alpha desynchronisation in the contralateral hemisphere alongside an increase in alpha power ipsilaterally over the visual cortex [276-278]. So, for instance, during right-hand motor imagery, when the subjects direct their visual attention to the right visual hemifield and start imagining movement, alpha desynchronisation happens contralaterally in motor and visual

cortical regions that are associated with motor imagery and visual attention, respectively. Increased gamma activity has also been reported during visual attention contralateral to the visual hemifield [269], which would explain the increase in the power spectrum between 30-40Hz in most filters, yet more detectable in the second filter of each inception module. Gamma-band activity can also be explained, considering that the time window used for the analysis contains the period in which subjects attended to the visual cue. Please also note that visualisations >60Hz are not valid based on the Nyquist theorem, as we have downsampled our signals to 125Hz.

Finally, spatial activations in low frequencies (< 5Hz) over frontal regions most likely represent eye blinks and eye movements. This pattern can be seen in filters 1, 3, and 6 of the left panel and filter 3 in the middle. While eye blinks are unlikely to play a functional role, there might be a relationship between subjects' eye movement and how they attended to and performed each motor imagery task in the experiment, which is captured as a discriminative pattern by the neural network.

6.2 DPD Biomarkers

6.2.1 Contralateral P45

P45 findings are straightforward to interpret. They reflect the activity in the somatosensory cortex contralateral to tactile stimulation. This activity is heightened in DPD patients relative to controls (see **Figure 5.12**), suggesting enhanced visual-tactile processing following negative emotional primes as a biomarker for DPD. Why might tactile processing be enhanced in such circumstances? Adler et al. [37] showed that P45 visual-tactile processing is suppressed rather than enhanced for self-related stimulation conditions (self-face observation) in those with high levels of DPD compared to those with low levels. My findings of DPD-based P45 enhancements following negative emotional primes may thus be seen as

contradictory. However, the literature on DPD suggests that the stimulation conditions may have tapped into a different mechanism (other-related stimulation) in our scenario because we did not show participants' own body parts but photographs of other people's hands only during tactile stimulation. Farmer et al. [39] reported enhanced visual-tactile processing for other faces in those with high vs low levels of DPD. It may be that when the visual stimulus is more indicative of "other" than of "self", visual-tactile integration is enhanced in DPD relative to control groups, and this is what has been extracted by my current analyses. The underlying reasons for this remain yet unexplored in research. Still, it is not unlikely that they reflect a strategy of emotional over-activation or over-attunement in those with frequent depersonalisation symptoms, given the well-established links between DPD and childhood trauma (typically from emotional abuse or neglect [4, 20]).

6.2.2 Ipsilateral P200/P300 over Sensory-motor Processing Regions

The differences between the groups in P200 (right hand stimulation) and later P300 (left hand stimulation), with sources in ipsilateral sensory-motor processing regions, are likely related to those reported by Adler et al. [37] at frontocentral P200. In [37], other-related visual tactile processing was reduced in those with high levels of DPD compared to those with low levels. My analysis also showed less activation among DPD patients compared to the control group. Although my results are derived from an experiment that was not designed to directly manipulate and measure self-other distinction, the P200/P300 findings may still be speculated as a biomarker of DPD that reflects reduced self-other differentiation, as argued by Adler et al. [37]. It is frequently reported that in DPD, the mirror image feels like a stranger to the observer despite the full realisation that they are looking at themselves. This phenomenology may be underlain by less distinct self and other processing mechanisms operating in this time range. It is feasible to propose that this may be a consequence of the earlier over-attunement with the other that was seen in the time range of P45. Interestingly, the relevant right-

hemispheric biomarker emerges earlier in processing (P200) than its left-hemispheric equivalent (P300), but the left-hemispheric group differences are markedly stronger. It may be speculated that this may relate to the potential left hemispheric abnormalities that have been documented in DPD [24, 231], possibly reflecting an aberration of the typical pattern of hemispheric differences in emotional processing [233], whereby the left hemisphere predominantly processes positive emotions, and the right hemisphere predominantly processes negative emotions [234].

6.2.3 P200 over Occipital-temporal Cortex

P200 findings over the right occipital-temporal cortex (PO8) may be related to aberrant visual processing in DPD relative to controls. The identified time range of the effect and its spatial source is in line with ERP components related to the recognition of familiar faces and bodies in occipitotemporal regions, where typically a smaller P200 is obtained for familiar relative to unfamiliar shapes (e.g., [279]). Enhanced processing for the DPD group relative to controls in this time range may thus reflect greater unfamiliarity during synchronous visual-tactile stimulation. DPD is typically marked by feelings of disembodiment, where one's own hands and face may not feel like they belong to one's self, and derealisation, where one's surroundings and reality in general may appear dreamlike, intangible and unfamiliar. It is conceivable that the relatively heightened P200 processing in higher-level visual association regions for the DPD group may be a biomarker of this phenomenological experience.

7 Conclusion, Limitations, and Future Work

DPD affects 1-2% of the population, comparable to schizophrenia and OCD. Yet, it takes seven to 12 years on average to be accurately diagnosed. Therefore, a correct diagnosis of DPD is an urgent matter in the area of psychological disorders, and there is a need to find diagnostic markers highly specific to DPD to distinguish it from other alternative diagnoses. I argued that understanding the potential of electrophysiological tools for identifying DPD symptoms can help diagnose DPD quickly and effectively. I appealed to researchers interested in the phenomena of self-awareness in health and disease to consider using these tools more frequently. Accordingly, in the first stage of my PhD, I provided a systematic review to describe research targeting the neural correlates of core DPD symptoms that have used electrophysiological techniques. I aimed to investigate the diagnostic potential of these relatively inexpensive and convenient neuroimaging tools. I reviewed the EEG power spectrum, components of the ERP, as well as vestibular and heartbeat-evoked potentials as likely electrophysiological biomarkers to study DPD symptoms. I argued that acute anxiety- or trauma-related impairments in the integration of interoceptive and exteroceptive signals play a key role in the formation of DPD symptoms and that future research needs analysis methods that can take this integration into account. Besides, literature has reported theta abnormalities in DPD related to emotion, attention/inhibition and working memory. Therefore, future studies should consider a relatively prolonged time window to analyse low-frequency EEG components such as theta, which may affect stimulus design. Practically, the time window should contain at least three cycles of the target frequency.

My research then focused on employing deep learning techniques to design a more powerful biomarker discovery system for DPD. I focused on developing an explainable end-to-end deep learning model that can extract rich, informative neural patterns from EEG signals and exploit neural patterns specific to DPD symptoms to help the community better understand the disorder. I discussed why my DPD scenario, or generally analysing any

mental disorders assessed based on clinical assessment scores, should be seen as a multi-task learning problem to reduce the impact of uncertainty in CDS scores. Besides, I argued that the literature often relies on experts' knowledge of the disorder and is based on hypothesis testing to find DPD biomarkers. As a result, I proposed a multi-input multi-output deep learning structure, which was designed to find the best separability in the dataset, guided by clinical assessment score. Furthermore, I proposed a method to visualise and explain the learning and decision-making process in deep neural networks designed for EEG analysis and described how it could be applied to exploit multiple reliable EEG biomarkers for DPD. Finally, I summarised and interpreted the obtained biomarkers, including P45 contralateral to tactile stimulation, ipsilateral P200/P300 over sensory-motor processing regions, and P200 over the occipital-temporal cortex, from a cognitive neuroscientific point of view and provided references in the literature supporting my arguments. It must be remembered that transient depersonalisation is a common phenomenon during life span and could be early signs of chronic type risk. Hence, developing a system to track the depersonalisation state and its severity could be of great importance to help with the prevention of the chronic type. The potential biomarkers and analytics presented in my work can help to find a solution for online tracking of a depersonalisation state.

The limitation in my results on the potential neural signatures of depersonalisation symptoms was the need for more discussion on the spectral information derived during the learning process. Although I visualised the frequency responses of the filters trained in my model, I did not investigate them further for the ERP dataset. Since each trial was only 500ms long, there needed to be more samples to calculate the Fourier transform accurately. Besides, applying a band-pass filter on such a short window to focus on a specific frequency band would result in severe signal distortion and, therefore, the unreliability of results. That is why I encourage future studies to consider a relatively prolonged time window in order to

investigate also spectral information and some low-frequency components of EEG (such as theta), which were flagged as potential biomarkers in the literature.

To indicate my visualisation technique's effectiveness, I also applied it to a novel end-to-end deep learning model for motor imagery BCI called EEG-ITNet. I comprehensively explained and supported the validity of network illustration from a neuroscientific perspective and discussed the superior interpretability of my visualisation technique over other attempts in the literature, such as the one from the authors of EEGNet. In addition, using inception modules and causal convolutions with dilation, EEG-ITNet showed capabilities in extracting rich spectral, spatial, and temporal information from multi-channel EEG signals with less complexity (in terms of the number of trainable parameters) than other existing end-to-end architectures, such as EEG-Inception and EEG-TCNet. I thoroughly evaluated EEG-ITNet on two motor imagery BCI datasets (dataset 2a from BCI competition IV and OpenBMI motor imagery dataset) in three different scenarios: within-subject, cross-subject, and cross-subject with fine-tuning. EEG-ITNet showed statistically significant improvement in performing the classification task in most scenarios (up to 5.9%) compared to other end-to-end architectures. I argued that the performance of EEG-ITNet may be improved even further by finding the optimum hyperparameters for each scenario. In addition, finding a way to visualise and explain temporal features and dependencies learned inside its TC block would be of great importance and can further improve the interpretability of EEG-ITNet. Besides, I was interested in further investigating the effect of network deepness on its performance. These could be one's focus in future studies. Nevertheless, I believe that the concepts introduced in my motor imagery study can help to develop more robust, interpretable, and high-accuracy BCI systems.

In my proposed EEG analytics, the deep learning model is no longer recognised as a black box, and its learning process can be explained. In addition, it can be modified and applied to

any psychological and mental disorders currently indicated based on clinical assessment scores to exploit electrophysiological biomarkers that can help clinicians with a more accurate diagnosis. The input to the network can be ERP, EEG recording during a mental task, or resting state EEG, and one can employ it to extract and interpret neural patterns in the same way with just a few modifications in the network parameters and layers. However, it should be noted that the goal of my PhD research was not to perform a classification task or develop a diagnostic tool, as the true labels in my scenario were based on CDS scores which I argued are only partially reliable. This was a critical limitation in the main DPD dataset I used for my research, as individuals with a high level of depersonalisation were determined only according to the outcome of CDS rather than being clinically diagnosed with DPD.

Nevertheless, to show the importance of finding a reliable electrophysiological biomarker in the diagnostic process, I demonstrated the significance of advanced signal processing algorithms to perform a more accurate classification of individuals with high and low levels of depersonalisation symptoms using an electrophysiological biomarker. For this purpose, I used contralateral P45 components of SEP over the somatosensory cortex discovered in my analysis and achieved 85% accuracy (Kappa value of 0.7) in a classification task. Indeed a more extensive dataset with clinically approved labels of participants is needed to confirm P45 as an accurate and reliable biomarker for DPD diagnosis.

Bibliography

[1] B. Consortium *et al.*, "Analysis of shared heritability in common disorders of the brain," *Science,* vol. 360, no. 6395, p. eaap8757, 2018.

[2] E. C. Hunter, M. Sierra, and A. S. David, "The epidemiology of depersonalisation and derealisation," *Soc. Psychiatry Psychiatr. Epidemiol.,* vol. 39, no. 1, pp. 9-18, 2004.

[3] W. E. Lee, C. H. Kwok, E. C. Hunter, M. Richards, and A. S. David, "Prevalence and childhood antecedents of depersonalization syndrome in a UK birth cohort," *Soc. Psychiatry Psychiatr. Epidemiol.,* vol. 47, no. 2, pp. 253-261, 2012.

[4] M. Michal, M. E. Beutel, J. Jordan, M. Zimmermann, S. Wolters, and T. Heidenreich, "Depersonalization, mindfulness, and childhood trauma," *The Journal of nervous and mental disease,* vol. 195, no. 8, pp. 693-696, 2007.

[5] C. M. Michel and M. M. Murray, "Towards the utilization of EEG as a brain imaging tool," *Neuroimage,* vol. 61, no. 2, pp. 371-385, 2012.

[6] M. Sierra and G. E. Berrios, "The Cambridge Depersonalisation Scale: A new instrument for the measurement of depersonalisation," *Psychiatry Res.,* vol. 93, no. 2, pp. 153-164, 2000.

[7] H. Merckelbach, T. Giesbrecht, D. van Heugten-van der Kloet, J. d. Jong, T. Meyer, and K. Rietman, "The overlap between dissociative symptoms and symptom over-reporting," *The European Journal of Psychiatry,* vol. 29, no. 3, pp. 165-172, 2015.

[8] S. Vieira, W. H. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications," *Neurosci. Biobehav. Rev.,* vol. 74, pp. 58-75, 2017.

[9] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.,* vol. 42, pp. 60-88, 2017.

[10] M. L. Phillips and M. Sierra, "Depersonalization disorder: a functional neuroanatomical perspective," *Stress,* vol. 6, no. 3, pp. 157-165, 2003.

[11] E. Hunter, M. L. Phillips, T. Chalder, M. Sierra, and A. David, "Depersonalisation disorder: a cognitive–behavioural conceptualisation," *Behav. Res. Ther.,* vol. 41, no. 12, pp. 1451-1467, 2003.

[12] D. J. Stein and D. Simeon, "Cognitive-affective neuroscience of depersonalization," *CNS spectrums,* vol. 14, no. 9, pp. 467-471, 2009.

[13] E. Shilony and F. K. Grossman, "Depersonalization as a defense mechanism in survivors of trauma," *J. Trauma. Stress,* vol. 6, no. 1, pp. 119-128, 1993.

[14] M. Sierra and G. E. Berrios, "Depersonalization: neurobiological perspectives," *Biol. Psychiatry,* vol. 44, no. 9, pp. 898-908, 1998.

[15] J. Dixon, "Depersonalization phenomena in a sample population of college students," *The British journal of psychiatry,* vol. 109, no. 460, pp. 371-375, 1963.

[16] A. N. Tibubos, J. Grammes, M. E. Beutel, M. Michal, G. Schmutzer, and E. Brähler, "Emotion regulation strategies moderate the relationship of fatigue with depersonalization and derealization symptoms," *J. Affect. Disord.,* vol. 227, pp. 571-579, 2018.

[17] D. van Heugten–van der Kloet, T. Giesbrecht, and H. Merckelbach, "Sleep loss increases dissociation and affects memory for emotional stimuli," *J. Behav. Ther. Exp. Psychiatry,* vol. 47, pp. 9-17, 2015.

[18] H. Kaplan, B. Sadock, and J. Grebb, "Substance related disorders," *Kaplan HI, Sadock BJ. Kaplan and Sadock's synopsis of psychiatry: behavioral sciences, clinical psychiatry. 8th ed. Baltimore: Williams & Wilkins,* pp. 419-26, 1998.

[19] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders* (BMC Med). 2013, pp. 133-137.

[20] D. Simeon, O. Guralnik, J. Schmeidler, B. Sirof, and M. Knutelska, "The role of childhood interpersonal trauma in depersonalization disorder," *Am. J. Psychiatry,* vol. 158, no. 7, pp. 1027-1033, 2001.

[21] R. Mathew, W. Wilson, N. Chiu, T. Turkington, T. R. DeGrado, and R. Coleman, "Regional cerebral blood flow and depersonalization after tetrahydrocannabinol adrninistration," *Acta Psychiatr. Scand.,* vol. 100, no. 1, pp. 67-75, 1999.

[22] N. Medford *et al.,* "Chronic depersonalization following illicit drug use: a controlled analysis of 40 cases," *Addiction,* vol. 98, no. 12, pp. 1731-1736, 2003.

[23] D. Simeon, "Depersonalisation disorder," *CNS drugs,* vol. 18, no. 6, pp. 343-354, 2004.

[24] E. Hollander, J. L. Carrasco, L. S. Mullen, S. Trungold, C. M. DeCaria, and J. Towey, "Left hemispheric activation in depersonalization disorder: a case report," *Biol. Psychiatry,* vol. 31, no. 11, pp. 1157-1162, 1992.

[25] M. Sierra and A. S. David, "Depersonalization: a selective impairment of self-awareness," *Conscious. Cogn.,* vol. 20, no. 1, pp. 99-108, 2011.

[26] D. J. Stein, J. Schmeidler, and E. Hollander, "Feeling unreal: 30 cases of DSM-III-R depersonalization disorder," *Am. J. Psychiatry,* vol. 154, pp. 1107-1113, 1997.

[27] M. V. Lambert, C. Senior, W. D. Fewtrell, M. L. Phillips, and A. S. David, "Primary and secondary depersonalisation disorder:: a psychometric study," *J. Affect. Disord.,* vol. 63, no. 1-3, pp. 249-256, 2001.

[28] E. C. Hunter, J. Charlton, and A. S. David, "Depersonalisation and derealisation: assessment and management," *BMJ,* vol. 356, p. j745, 2017.

[29] D. Baker *et al.*, "Depersonalisation disorder: clinical features of 204 cases," *The British Journal of Psychiatry,* vol. 182, no. 5, pp. 428-433, 2003.

[30] M. Michal *et al.*, "A case series of 223 patients with depersonalization-derealization syndrome," *BMC Psychiatry,* vol. 16, no. 1, p. 203, 2016.

[31] D. Simeon, M. Knutelska, D. Nelson, and O. Guralnik, "Feeling unreal: a depersonalization disorder update of 117 cases," *The Journal of clinical psychiatry,* 2003.

[32] M. Michal, M. E. Beutel, and T. G. Grobe, "How often is the Depersonalization-Derealization Disorder (ICD-10: F48. 1) diagnosed in the outpatient health-care service?," *Zeitschrift fur Psychosomatische Medizin und Psychotherapie,* vol. 56, no. 1, pp. 74-83, 2010.

[33] N. Medford, M. Sierra, D. Baker, and A. S. David, "Understanding and treating depersonalisation disorder," *Advances in psychiatric Treatment,* vol. 11, no. 2, pp. 92-100, 2005.

[34] O. Devinsky, F. Putnam, J. Grafman, E. Bromfield, and W. H. Theodore, "Dissociative states and epilepsy," *Neurology,* vol. 39, no. 6, pp. 835-835, 1989.

[35] M. V. Lambert, M. Sierra, M. L. Phillips, and A. S. David, "The spectrum of organic depersonalization: a review plus four new cases," *The Journal of neuropsychiatry and clinical neurosciences,* vol. 14, no. 2, pp. 141-154, 2002.

[36] C. Armour, A. A. Contractor, P. A. Palmieri, and J. D. Elhai, "Assessing latent level associations between PTSD and dissociative factors: Is depersonalization and derealization related to PTSD factors more so than alternative dissociative factors?," *Psychol. Inj. Law,* vol. 7, no. 2, pp. 131-142, 2014.

[37] J. Adler, N. Schabinger, M. Michal, M. E. Beutel, and H. Gillmeister, "Is that me in the mirror? Depersonalisation modulates tactile mirroring mechanisms," *Neuropsychologia,* vol. 85, pp. 148-158, 2016.

[38] N. Kanayama, A. Sato, and H. Ohira, "The role of gamma band oscillations and synchrony on rubber hand illusion and crossmodal integration," *Brain Cogn.,* vol. 69, no. 1, pp. 19-29, 2009.

[39] H. Farmer *et al.*, "The Detached Self: Investigating the Effect of Depersonalisation on Self-Bias in the Visual Remapping of Touch," 2019.

[40] S. P. van den Broek, F. Reinders, M. Donderwinkel, and M. Peters, "Volume conduction effects in EEG and MEG," *Electroencephalogr. Clin. Neurophysiol.,* vol. 106, no. 6, pp. 522-534, 1998.

[41] S. Sur and V. Sinha, "Event-related potential: An overview," *Industrial psychiatry journal,* vol. 18, no. 1, p. 70, 2009.

[42] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust.*, vol. 26, no. 1, pp. 43-49, 1978.

[43] R. Bellman, "Dynamic programming princeton university press princeton," *New Jersey Google Scholar*, 1957.

[44] M. E. Munich and P. Perona, "Continuous dynamic time warping for translation-invariant curve alignment with applications to signature verification," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, vol. 1: IEEE, pp. 108-115.

[45] B. Li, Z. Liu, X. Gao, and Y. Lin, "N400 extraction from a few trials of EEG data using spatial and temporal-frequency pattern analysis," *Journal of Neural Engineering*, vol. 16, no. 6, p. 066035, 2019.

[46] V. Vovk, "The fundamental nature of the log loss function," in *Fields of Logic and Computation II*: Springer, 2015, pp. 307-318.

[47] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[48] G. Li, C. H. Lee, J. J. Jung, Y. C. Youn, and D. Camacho, "Deep learning for EEG data analytics: A survey," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 18, p. e5199, 2020.

[49] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.

[50] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[51] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 156-165.

[52] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. 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)*, 2020: IEEE, pp. 2958-2965.

[53] Y. K. Musallam *et al.*, "Electroencephalography-based motor imagery classification using temporal convolutional network fusion," *Biomedical Signal Processing and Control*, vol. 69, p. 102826, 2021.

[54] M. Michal *et al.*, "Depersonalization disorder: disconnection of cognitive evaluation from autonomic responses to emotional stimuli," *PLoS One*, vol. 8, no. 9, p. e74331, 2013.

[55] K.-M. Monde, S. Ketay, T. Giesbrecht, A. Braun, and D. Simeon, "Preliminary physiological evidence for impaired emotion regulation in depersonalization disorder," *Psychiatry Res.*, vol. 209, no. 2, pp. 235-238, 2013.

[56] M. Sierra, C. Senior, M. L. Phillips, and A. S. David, "Autonomic response in the perception of disgust and happiness in depersonalization disorder," *Psychiatry Res.*, vol. 145, no. 2-3, pp. 225-231, 2006.

[57] H. Dewe, D. G. Watson, and J. J. Braithwaite, "Uncomfortably numb: new evidence for suppressed emotional reactivity in response to body-threats in those predisposed to sub-clinical dissociative experiences," *Cogn. Neuropsychiatry,* vol. 21, no. 5, pp. 377-401, 2016.

[58] T. Giesbrecht, H. Merckelbach, K. van Oorsouw, and D. Simeon, "Skin conductance and memory fragmentation after exposure to an emotional film clip in depersonalization disorder," *Psychiatry Res.*, vol. 177, no. 3, pp. 342-349, 2010.

[59] E.-L. Jay, M. Sierra, F. Van den Eynde, J. C. Rothwell, and A. S. David, "Testing a neurobiological model of depersonalization disorder using repetitive transcranial magnetic stimulation," *Brain stimulation,* vol. 7, no. 2, pp. 252-259, 2014.

[60] C. H. Röder, M. Michal, G. Overbeck, V. G. van de Ven, and D. E. Linden, "Pain response in depersonalization: a functional imaging study using hypnosis in healthy subjects," *Psychother. Psychosom.*, vol. 76, no. 2, pp. 115-121, 2007.

[61] D. Simeon, O. Guralnik, E. A. Hazlett, J. Spiegel-Cohen, E. Hollander, and M. S. Buchsbaum, "Feeling unreal: a PET study of depersonalization disorder," *Am. J. Psychiatry,* vol. 157, no. 11, pp. 1782-1788, 2000.

[62] D. Simeon, D. S. Kozin, K. Segal, B. Lerch, R. Dujour, and T. Giesbrecht, "De-constructing depersonalization: further evidence for symptom clusters," *Psychiatry Res.*, vol. 157, no. 1-3, pp. 303-306, 2008.

[63] E. Lemche *et al.*, "Limbic and prefrontal responses to facial emotion expressions in depersonalization," *Neuroreport,* vol. 18, no. 5, pp. 473-477, 2007.

[64] E. Lemche *et al.*, "Cerebral and autonomic responses to emotional facial expressions in depersonalisation disorder," *The British Journal of Psychiatry,* vol. 193, no. 3, pp. 222-228, 2008.

[65] A. Mancini-Marïe, C. Fahim, S. Potvin, M. Beauregard, and E. Stip, "Quetiapine: focus on emotional numbing in depersonalization disorder: an fMRI case report," *Eur. Psychiatry,* vol. 21, no. 8, pp. 574-577, 2006.

[66] N. Medford, B. Brierley, M. Brammer, E. T. Bullmore, A. S. David, and M. L. Phillips, "Emotional memory in depersonalization disorder: a functional MRI study," *Psychiatry Research: Neuroimaging,* vol. 148, no. 2-3, pp. 93-102, 2006.

[67] M. L. Phillips *et al.*, "Depersonalization disorder: thinking without feeling," *Psychiatry Research: Neuroimaging,* vol. 108, no. 3, pp. 145-160, 2001.

[68] N. Medford, M. Sierra, A. Stringaris, V. Giampietro, M. J. Brammer, and A. S. David, "Emotional experience and awareness of self: functional MRI studies of depersonalization disorder," *Front. Psychol.,* vol. 7, p. 432, 2016.

[69] A. Etkin, T. Egner, D. M. Peraza, E. R. Kandel, and J. Hirsch, "Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala," *Neuron,* vol. 51, no. 6, pp. 871-882, 2006.

[70] E. R. Paul, M. Farmer, R. Kämpe, H. R. Cremers, and J. P. Hamilton, "Functional Connectivity Between Extrastriate Body Area and Default Mode Network Predicts Depersonalization Symptoms in Major Depression: Findings From an A Priori Specified Multinetwork Comparison," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging,* 2019.

[71] M. Derome *et al.,* "Resting-state networks of adolescents experiencing depersonalization-like illusions: Cross-sectional and longitudinal findings," *Schizophr. Bull.,* vol. 44, no. suppl_2, pp. S501-S511, 2018.

[72] M. Sierra, D. Baker, N. Medford, and A. S. David, "Unpacking the depersonalization syndrome: an exploratory factor analysis on the Cambridge Depersonalization Scale," *Psychol. Med.,* vol. 35, no. 10, pp. 1523-1532, 2005.

[73] Y. Aderibigbe, R. Bloch, and W. Walker, "Prevalence of depersonalization and derealization experiences in a rural population," *Soc. Psychiatry Psychiatr. Epidemiol.,* vol. 36, no. 2, pp. 63-69, 2001.

[74] M. Sierra, N. Medford, G. Wyatt, and A. S. David, "Depersonalization disorder and anxiety: a special relationship?," *Psychiatry Res.,* vol. 197, no. 1-2, pp. 123-127, 2012.

[75] M. Harper and M. Roth, "Temporal lobe epilepsy and the phobic anxiety-depersonalization syndrome: I: A comparative study," *Compr. Psychiatry,* 1962.

[76] C. Maslach and S. E. Jackson, "The measurement of experienced burnout," *Journal of organizational behavior,* vol. 2, no. 2, pp. 99-113, 1981.

[77] R. A. Lanius, B. Brand, E. Vermetten, P. A. Frewen, and D. Spiegel, "The dissociative subtype of posttraumatic stress disorder: Rationale, clinical and neurobiological evidence, and implications," *Depress. Anxiety,* vol. 29, no. 8, pp. 701-708, 2012.

[78] K. Davison, "Episodic depersonalization: observations on 7 patients," *The British Journal of Psychiatry,* vol. 110, no. 467, pp. 505-513, 1964.

[79] T. Dietl, C. Bien, H. Urbach, C. Elger, and M. Kurthen, "Episodic depersonalization in focal epilepsy," *Epilepsy Behav.,* vol. 7, no. 2, pp. 311-315, 2005.

[80] M. Roth, "The phobic anxiety-depersonalization syndrome and some general aetiological problems in psychiatry," *J. Neuropsychiatr.,* vol. 1, p. 293, 1960.

[81] R. A. Lanius *et al.*, "Emotion modulation in PTSD: Clinical and neurobiological evidence for a dissociative subtype," *Am. J. Psychiatry*, vol. 167, no. 6, pp. 640-647, 2010.

[82] J. W. Hopper, P. A. Frewen, B. A. Van der Kolk, and R. A. Lanius, "Neural correlates of reexperiencing, avoidance, and dissociation in PTSD: Symptom dimensions and emotion dysregulation in responses to script-driven trauma imagery," *J. Trauma. Stress*, vol. 20, no. 5, pp. 713-725, 2007.

[83] M. Locatelli, L. Bellodi, G. Perna, and S. Scarone, "EEG power modifications in panic disorder during a temporolimbic activation task: relationships with temporal lobe clinical symptomatology," *The Journal of neuropsychiatry and clinical neurosciences*, 1993.

[84] E. B. Raimo, R. A. Roemer, M. Moster, and Y. Shan, "Alcohol-induced depersonalization," *Biol. Psychiatry*, vol. 45, no. 11, pp. 1523-1526, 1999.

[85] K. Hayashi, M. Makino, M. Hashizume, K. Nakano, and K. Tsuboi, "Electroencephalogram abnormalities in panic disorder patients: a study of symptom characteristics and pathology," *Biopsychosoc. Med.*, vol. 4, no. 1, p. 9, 2010.

[86] M. Iacoboni, R. P. Woods, M. Brass, H. Bekkering, J. C. Mazziotta, and G. Rizzolatti, "Cortical mechanisms of human imitation," *Science*, vol. 286, no. 5449, pp. 2526-2528, 1999.

[87] P. Molenberghs, R. Cunnington, and J. B. Mattingley, "Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies," *Neurosci. Biobehav. Rev.*, vol. 36, no. 1, pp. 341-349, 2012.

[88] P. Molenberghs, R. Cunnington, and J. B. Mattingley, "Is the mirror neuron system involved in imitation? A short review and meta-analysis," *Neurosci. Biobehav. Rev.*, vol. 33, no. 7, pp. 975-980, 2009.

[89] S. Morales, L. C. Bowman, K. R. Velnoskey, N. A. Fox, and E. Redcay, "An fMRI study of action observation and action execution in childhood," *Dev. Cogn. Neurosci.*, vol. 37, p. 100655, 2019.

[90] R. M. Le Bel, J. A. Pineda, and A. Sharma, "Motor–auditory–visual integration: the role of the human mirror neuron system in communication and communication disorders," *J. Commun. Disord.*, vol. 42, no. 4, pp. 299-304, 2009.

[91] C. Hu *et al.*, "Distinct and common aspects of physical and psychological self-representation in the brain: A meta-analysis of self-bias in facial and self-referential judgements," *Neurosci. Biobehav. Rev.*, vol. 61, pp. 197-207, 2016.

[92] I. Molnar-Szakacs and L. Q. Uddin, "Self-processing and the default mode network: interactions with the mirror neuron system," *Front. Hum. Neurosci.*, vol. 7, p. 571, 2013.

[93] O. Turjman, "On the role of mirror neurons in the sense of self," *Journal of Consciousness Exploration & Research,* vol. 7, no. 4, 2016.

[94] M. L. Filippetti, S. Lloyd-Fox, M. R. Longo, T. Farroni, and M. H. Johnson, "Neural mechanisms of body awareness in infants," *Cereb. Cortex,* vol. 25, no. 10, pp. 3779-3787, 2014.

[95] V. Gallese and C. Sinigaglia, "The bodily self as power for action," *Neuropsychologia,* vol. 48, no. 3, pp. 746-755, 2010.

[96] S. Ketay, H. K. Hamilton, B. W. Haas, and D. Simeon, "Face processing in depersonalization: An fMRI study of the unfamiliar self," *Psychiatry Research: Neuroimaging,* vol. 222, no. 1-2, pp. 107-110, 2014.

[97] M. Botvinick and J. Cohen, "Rubber hands ‘feel’touch that eyes see," *Nature,* vol. 391, no. 6669, p. 756, 1998.

[98] H. H. Ehrsson, C. Spence, and R. E. Passingham, "That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb," *Science,* vol. 305, no. 5685, pp. 875-877, 2004.

[99] M. Tsakiris and P. Haggard, "The rubber hand illusion revisited: visuotactile integration and self-attribution," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 31, no. 1, p. 80, 2005.

[100] B. Lenggenhager, T. Tadi, T. Metzinger, and O. Blanke, "Video ergo sum: manipulating bodily self-consciousness," *Science,* vol. 317, no. 5841, pp. 1096-1099, 2007.

[101] E. Nakul, N. Orlando-Dessaints, B. Lenggenhager, and C. Lopez, "Measuring perceived self-location in virtual reality," *Sci. Rep.*, vol. 10, no. 1, pp. 1-12, 2020.

[102] A. Pomés and M. Slater, "Drift and ownership toward a distant virtual body," *Front. Hum. Neurosci.,* vol. 7, p. 908, 2013.

[103] O. Blanke, M. Slater, and A. Serino, "Behavioral, neural, and computational principles of bodily self-consciousness," *Neuron,* vol. 88, no. 1, pp. 145-166, 2015.

[104] H. H. Ehrsson, N. P. Holmes, and R. E. Passingham, "Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas," *J. Neurosci.,* vol. 25, no. 45, pp. 10564-10573, 2005.

[105] J.-P. Noel, O. Blanke, and A. Serino, "From multisensory integration in peripersonal space to bodily self-consciousness: from statistical regularities to statistical inference," *Ann. NY Acad. Sci,* vol. 1426, pp. 146-165, 2018.

[106] M. Tsakiris, "My body in the brain: a neurocognitive model of body-ownership," *Neuropsychologia,* vol. 48, no. 3, pp. 703-712, 2010.

[107] G. A. Calvert and T. Thesen, "Multisensory integration: methodological approaches and emerging principles in the human brain," *Journal of Physiology-Paris,* vol. 98, no. 1-3, pp. 191-205, 2004.

[108] K. O. Bushara, J. Grafman, and M. Hallett, "Neural correlates of auditory–visual stimulus onset asynchrony detection," *J. Neurosci.,* vol. 21, no. 1, pp. 300-304, 2001.

[109] K. O. Bushara, T. Hanakawa, I. Immisch, K. Toma, K. Kansaku, and M. Hallett, "Neural correlates of cross-modal binding," *Nat. Neurosci.,* vol. 6, no. 2, p. 190, 2003.

[110] N. O'Sullivan *et al.*, "I Am There… but Not Quite: An Unfaithful Mirror That Reduces Feelings of Ownership and Agency," *Perception,* vol. 47, no. 2, pp. 197-215, 2018.

[111] J. J. Braithwaite, D. G. Watson, and H. Dewe, "Predisposition to out-of-body experience (OBE) is associated with aberrations in multisensory integration: Psychophysiological support from a “rubber hand illusion” study," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 43, no. 6, p. 1125, 2017.

[112] A. D. Craig, "How do you feel--now? The anterior insula and human awareness," *Nature reviews neuroscience,* vol. 10, no. 1, 2009.

[113] M. Tsakiris, A. T.-. Jiménez, and M. Costantini, "Just a heartbeat away from one's body: interoceptive sensitivity predicts malleability of body-representations," *Proceedings of the Royal Society B: Biological Sciences,* vol. 278, no. 1717, pp. 2470-2476, 2011.

[114] H.-D. Park and O. Blanke, "Coupling Inner and Outer Body for Self-Consciousness," *Trends Cogn. Sci.,* 2019.

[115] S. S. Khalsa and R. C. Lapidus, "Can interoception improve the pragmatic search for biomarkers in psychiatry?," *Frontiers in psychiatry,* vol. 7, p. 121, 2016.

[116] A. D. Craig and A. Craig, "How do you feel--now? The anterior insula and human awareness," *Nature reviews neuroscience,* vol. 10, no. 1, 2009.

[117] E. Lemche *et al.*, "Interoceptive–reflective regions differentiate alexithymia traits in depersonalization disorder," *Psychiatry Research: Neuroimaging,* vol. 214, no. 1, pp. 66-72, 2013.

[118] N. Medford, "Emotion and the unreal self: depersonalization disorder and de-affectualization," *Emot. Rev.,* vol. 4, no. 2, pp. 139-144, 2012.

[119] A. Sierk *et al.*, "White matter network alterations in patients with depersonalization/derealization disorder," *Journal of psychiatry & neuroscience: JPN,* vol. 43, no. 5, p. 347, 2018.

[120] M. Sierra, S. Nestler, E.-L. Jay, C. Ecker, Y. Feng, and A. S. David, "A structural MRI study of cortical thickness in depersonalisation disorder," *Psychiatry Research: Neuroimaging*, vol. 224, no. 1, pp. 1-7, 2014.

[121] M. Michal *et al.*, "Striking discrepancy of anomalous body experiences with normal interoceptive accuracy in depersonalization-derealization disorder," *PLoS One*, vol. 9, no. 2, p. e89823, 2014.

[122] L. Sedeño *et al.*, "How do you feel when you can't feel your body? Interoception, functional connectivity and emotional processing in depersonalization-derealization disorder," *PLoS One*, vol. 9, no. 6, p. e98769, 2014.

[123] R. Schandry, "Heart beat perception and emotional experience," *Psychophysiology*, vol. 18, no. 4, pp. 483-488, 1981.

[124] W. E. Whitehead, V. M. Drescher, P. Heiman, and B. Blackwell, "Relation of heart rate control to heartbeat perception," *Biofeedback and Self-regulation*, vol. 2, no. 4, pp. 371-392, 1977.

[125] E. Ferentzi, R. Drew, B. T. Tihanyi, and F. Köteles, "Interoceptive accuracy and body awareness–Temporal and longitudinal associations in a non-clinical sample," *Physiol. Behav.*, vol. 184, pp. 100-107, 2018.

[126] M. A. Gray *et al.*, "A cortical potential reflecting cardiac function," *Proceedings of the National Academy of Sciences*, vol. 104, no. 16, pp. 6818-6823, 2007.

[127] H.-D. Park and O. Blanke, "Heartbeat-evoked cortical responses: Underlying mechanisms, functional roles, and methodological considerations," *Neuroimage*, vol. 197, pp. 502-511, 2019.

[128] O. Pollatos and R. Schandry, "Accuracy of heartbeat perception is reflected in the amplitude of the heartbeat-evoked brain potential," *Psychophysiology*, vol. 41, no. 3, pp. 476-482, 2004.

[129] H.-D. Park *et al.*, "Neural sources and underlying mechanisms of neural responses to heartbeats, and their role in bodily self-consciousness: an intracranial EEG study," *Cereb. Cortex*, vol. 28, no. 7, pp. 2351-2364, 2017.

[130] H.-D. Park, S. Correia, A. Ducorps, and C. Tallon-Baudry, "Spontaneous fluctuations in neural responses to heartbeats predict visual detection," *Nat. Neurosci.*, vol. 17, no. 4, p. 612, 2014.

[131] I. García-Cordero *et al.*, "Attention, in and out: scalp-level and intracranial EEG correlates of interoception and exteroception," *Front. Neurosci.*, vol. 11, p. 411, 2017.

[132] P. Montoya, R. Schandry, and A. Müller, "Heartbeat evoked potentials (HEP): topography and influence of cardiac awareness and focus of attention," *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, vol. 88, no. 3, pp. 163-172, 1993.

[133] H.-D. Park, F. Bernasconi, J. Bello-Ruiz, C. Pfeiffer, R. Salomon, and O. Blanke, "Transient modulations of neural responses to heartbeats covary with bodily self-consciousness," *J. Neurosci.,* vol. 36, no. 32, pp. 8453-8460, 2016.

[134] L. Heydrich and O. Blanke, "Distinct illusory own-body perceptions caused by damage to posterior insula and extrastriate cortex," *Brain,* vol. 136, no. 3, pp. 790-803, 2013.

[135] R. Ronchi *et al.,* "Right insular damage decreases heartbeat awareness and alters cardio-visual effects on bodily self-consciousness," *Neuropsychologia,* vol. 70, pp. 11-20, 2015.

[136] A. Schulz *et al.,* "Altered patterns of heartbeat-evoked potentials in depersonalization/derealization disorder: neurophysiological evidence for impaired cortical representation of bodily signals," *Psychosom. Med.,* vol. 77, no. 5, pp. 506-516, 2015.

[137] A. Schulz *et al.,* "Cardiac modulation of startle is altered in depersonalization-/derealization disorder: evidence for impaired brainstem representation of baro-afferent neural traffic," *Psychiatry Res.,* vol. 240, pp. 4-10, 2016.

[138] J. E. Aspell, L. Heydrich, G. Marillier, T. Lavanchy, B. Herbelin, and O. Blanke, "Turning body and self inside out: visualized heartbeats alter bodily self-consciousness and tactile perception," *Psychol. Sci.,* vol. 24, no. 12, pp. 2445-2453, 2013.

[139] K. Suzuki, S. N. Garfinkel, H. D. Critchley, and A. K. Seth, "Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion," *Neuropsychologia,* vol. 51, no. 13, pp. 2909-2917, 2013.

[140] A. Sel, R. T. Azevedo, and M. Tsakiris, "Heartfelt Self: Cardio-Visual Integration Affects Self-Face Recognition and Interoceptive Cortical Processing," *Cereb. Cortex,* vol. 27, no. 11, pp. 5144-5155, Nov 1 2017, doi: 10.1093/cercor/bhw296.

[141] M. Tsakiris, "Looking for myself: current multisensory input alters self-face recognition," *PLoS One,* vol. 3, no. 12, p. e4040, 2008.

[142] L. Heydrich, J. E. Aspell, G. Marillier, T. Lavanchy, B. Herbelin, and O. Blanke, "Cardio-visual full body illusion alters bodily self-consciousness and tactile processing in somatosensory cortex," *Sci. Rep.,* vol. 8, no. 1, p. 9230, 2018.

[143] J. Adler, M. E. Beutel, A. Knebel, S. Berti, J. Unterrainer, and M. Michal, "Altered orientation of spatial attention in depersonalization disorder," *Psychiatry Res.,* vol. 216, no. 2, pp. 230-235, 2014.

[144] O. Guralnik, T. Giesbrecht, M. Knutelska, B. Sirroff, and D. Simeon, "Cognitive functioning in depersonalization disorder," *The Journal of nervous and mental disease,* vol. 195, no. 12, pp. 983-988, 2007.

[145] P. Gerrans, "Depersonalization Disorder, Affective Processing and Predictive Coding," *Review of Philosophy and Psychology,* vol. 10, no. 2, pp. 401-418, 2019.

[146] C. A. Blevins, T. K. Witte, and F. W. Weathers, "Factor structure of the Cambridge Depersonalization Scale in trauma-exposed college students," *Journal of Trauma & Dissociation,* vol. 14, no. 3, pp. 288-301, 2013.

[147] M. Sierra *et al.*, "Autonomic response in depersonalization disorder," *Arch. Gen. Psychiatry,* vol. 59, no. 9, pp. 833-838, 2002.

[148] C. W. Quaedflieg *et al.*, "Early emotional processing deficits in depersonalization: An exploration with event-related potentials in an undergraduate sample," *Psychiatry Research: Neuroimaging,* vol. 212, no. 3, pp. 223-229, 2013.

[149] S. B. Most, M. M. Chun, D. M. Widders, and D. H. Zald, "Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness," *Psychonomic bulletin & review,* vol. 12, no. 4, pp. 654-661, 2005.

[150] D. Simeon, O. Guralnik, M. Knutelska, E. Hollander, and J. Schmeidler, "Hypothalamic-pituitary-adrenal axis dysregulation in depersonalization disorder," *Neuropsychopharmacology,* vol. 25, no. 5, pp. 793-795, 2001.

[151] B. R. Stanton *et al.*, "Basal activity of the hypothalamic–pituitary–adrenal axis in patients with depersonalization disorder," *Psychiatry Res.,* vol. 104, no. 1, pp. 85-89, 2001.

[152] S. Konishi, J. Chikazoe, K. Jimura, T. Asari, and Y. Miyashita, "Neural mechanism in anterior prefrontal cortex for inhibition of prolonged set interference," *Proceedings of the National Academy of Sciences,* vol. 102, no. 35, pp. 12584-12588, 2005.

[153] K. L. Phan, D. A. Fitzgerald, P. J. Nathan, G. J. Moore, T. W. Uhde, and M. E. Tancer, "Neural substrates for voluntary suppression of negative affect: a functional magnetic resonance imaging study," *Biol. Psychiatry,* vol. 57, no. 3, pp. 210-219, 2005.

[154] F. Dolcos, P. Kragel, L. Wang, and G. McCarthy, "Role of the inferior frontal cortex in coping with distracting emotions," *Neuroreport,* vol. 17, no. 15, pp. 1591-1594, 2006.

[155] A. Uusberg, H. Uibo, K. Kreegipuu, and J. Allik, "EEG alpha and cortical inhibition in affective attention," *Int. J. Psychophysiol.,* vol. 89, no. 1, pp. 26-36, 2013.

[156] M. Palmiero and L. Piccardi, "Frontal EEG asymmetry of mood: A mini-review," *Front. Behav. Neurosci.,* vol. 11, p. 224, 2017.

[157] P. Putman, "Resting state EEG delta–beta coherence in relation to anxiety, behavioral inhibition, and selective attentional processing of threatening stimuli," *Int. J. Psychophysiol.,* vol. 80, no. 1, pp. 63-68, 2011.

[158] P. Putman, J. van Peer, I. Maimari, and S. van der Werff, "EEG theta/beta ratio in relation to fear-modulated response-inhibition, attentional control, and affective traits," *Biol. Psychol.,* vol. 83, no. 2, pp. 73-78, 2010.

[159] P. Luu and D. M. Tucker, "Self-regulation and the executive functions: Electrophysiological clues," in *The cognitive electrophysiology of mind and brain*: Elsevier, 2003, pp. 199-II.

[160] H. T. Schupp, M. Junghöfer, A. I. Weike, and A. O. Hamm, "Attention and emotion: an ERP analysis of facilitated emotional stimulus processing," *Neuroreport,* vol. 14, no. 8, pp. 1107-1110, 2003.

[161] M. V. Lambert, C. Senior, M. L. Phillips, M. Sierra, E. Hunter, and A. S. David, "Visual imagery and depersonalisation," *Psychopathology,* vol. 34, no. 5, pp. 259-264, 2001.

[162] E. K. Warrington and M. James, "The visual object and space perception battery," 1991.

[163] D. L. Schacter, "Constructive memory: past and future," *Dialogues Clin. Neurosci.,* vol. 14, no. 1, p. 7, 2012.

[164] V. C. Martin, D. L. Schacter, M. C. Corballis, and D. R. Addis, "A role for the hippocampus in encoding simulations of future events," *Proceedings of the National Academy of Sciences,* vol. 108, no. 33, pp. 13858-13863, 2011.

[165] R. N. Spreng and C. L. Grady, "Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network," *J. Cogn. Neurosci.,* vol. 22, no. 6, pp. 1112-1123, 2010.

[166] C. Papageorgiou, E. Ventouras, N. Uzunoglu, A. Rabavilas, and C. Stefanis, "Changes of P300 elicited during a working memory test in individuals with depersonalization-derealization experiences," *Neuropsychobiology,* vol. 46, no. 2, pp. 70-75, 2002.

[167] H. U. Amin, A. S. Malik, N. Kamel, W.-T. Chooi, and M. Hussain, "P300 correlates with learning & memory abilities and fluid intelligence," *J. Neuroeng. Rehabil.,* vol. 12, no. 1, p. 87, 2015.

[168] A. M. Proverbio, M. E. Vanutelli, and S. Viganò, "Remembering faces: The effects of emotional valence and temporal recency," *Brain Cogn.,* vol. 135, p. 103584, 2019.

[169] T. Curran, "The electrophysiology of incidental and intentionalretrieval: erp old/new effects in lexical decision andrecognition memory," *Neuropsychologia,* vol. 37, no. 7, pp. 771-785, 1999.

[170] T. Curran, "Brain potentials of recollection and familiarity," *M&C,* vol. 28, no. 6, pp. 923-938, 2000.

[171] V. Romei, T. Rihs, V. Brodbeck, and G. Thut, "Resting electroencephalogram alpha-power over posterior sites indexes baseline visual cortex excitability," *Neuroreport,* vol. 19, no. 2, pp. 203-208, 2008.

[172] R. Keogh, J. Bergmann, and J. Pearson, "Cortical excitability controls the strength of mental imagery," *elife,* vol. 9, p. e50232, 2020.

[173] H. Dewe, D. G. Watson, K. Kessler, and J. J. Braithwaite, "The depersonalized brain: New evidence supporting a distinction between depersonalization and derealization from discrete patterns of autonomic suppression observed in a non-clinical sample," *Conscious. Cogn.,* vol. 63, pp. 29-46, 2018.

[174] M. Sierra, F. Lopera, M. Lambert, M. Phillips, and A. David, "Separating depersonalisation and derealisation: the relevance of the “lesion method”," *J. Neurol. Neurosurg. Psychiatry,* vol. 72, no. 4, pp. 530-532, 2002.

[175] L. Heydrich, G. Marillier, N. Evans, M. Seeck, and O. Blanke, "Depersonalization- and derealization-like phenomena of epileptic origin," *Annals of clinical and translational neurology,* 2019.

[176] N. A. Merriman, J. Ondřej, E. Roudaia, C. O’Sullivan, and F. N. Newell, "Familiar environments enhance object and spatial memory in both younger and older adults," *Exp. Brain Res.,* vol. 234, no. 6, pp. 1555-1574, 2016.

[177] R. E. Mruczek and D. L. Sheinberg, "Context familiarity enhances target processing by inferior temporal cortex neurons," *J. Neurosci.,* vol. 27, no. 32, pp. 8533-8545, 2007.

[178] H. Park, F. Leal, C. Abellanoza, and J. D. Schaeffer, "The formation of source memory under distraction," *Behavioral and Brain Functions,* vol. 10, no. 1, p. 40, 2014.

[179] J. Polich and A. Kok, "Cognitive and biological determinants of P300: an integrative review," *Biol. Psychol.,* vol. 41, no. 2, pp. 103-146, 1995.

[180] V. Wise, A. C. McFarlane, C. R. Clark, and M. Battersby, "Event-related potential and autonomic signs of maladaptive information processing during an auditory oddball task in panic disorder," *Int. J. Psychophysiol.,* vol. 74, no. 1, pp. 34-44, 2009.

[181] S. Lithfous, A. Dufour, F. Blanc, and O. Després, "Allocentric but not egocentric orientation is impaired during normal aging: an ERP study," *Neuropsychology,* vol. 28, no. 5, p. 761, 2014.

[182] J.-M. Hopf, E. Vogel, G. Woodman, H.-J. Heinze, and S. J. Luck, "Localizing visual discrimination processes in time and space," *J. Neurophysiol.,* vol. 88, no. 4, pp. 2088-2095, 2002.

[183] S. J. Luck and S. A. Hillyard, "Electrophysiological correlates of feature analysis during visual search," *Psychophysiology,* vol. 31, no. 3, pp. 291-308, 1994.

[184] E. K. Vogel and S. J. Luck, "The visual N1 component as an index of a discrimination process," *Psychophysiology,* vol. 37, no. 2, pp. 190-203, 2000.

[185] K. Gramann, J. Onton, D. Riccobon, H. J. Mueller, S. Bardins, and S. Makeig, "Human brain dynamics accompanying use of egocentric and allocentric reference frames during navigation," *J. Cogn. Neurosci.,* vol. 22, no. 12, pp. 2836-2849, 2010.

[186] S. Khan and R. Chang, "Anatomy of the vestibular system: a review," *NeuroRehabilitation,* vol. 32, no. 3, pp. 437-443, 2013.

[187] J. Ventre-Dominey, "Vestibular function in the temporal and parietal cortex: distinct velocity and inertial processing pathways," *Front. Integr. Neurosci.,* vol. 8, p. 53, 2014.

[188] B. Lenggenhager, C. Lopez, and O. Blanke, "Influence of galvanic vestibular stimulation on egocentric and object-based mental transformations," *Exp. Brain Res.,* vol. 184, no. 2, pp. 211-221, 2008.

[189] E. R. Ferrè, C. Lopez, and P. Haggard, "Anchoring the self to the body: vestibular contribution to the sense of self," *Psychol. Sci.,* vol. 25, no. 11, pp. 2106-2108, 2014.

[190] K. J. Renaud, "Vestibular function and depersonalization/derealization symptoms," *Multisensory research,* vol. 28, no. 5-6, pp. 637-651, 2015.

[191] F. Y. P. Sang, K. Jauregui-Renaud, D. A. Green, A. M. Bronstein, and M. A. Gresty, "Depersonalisation/derealisation symptoms in vestibular disease," *J. Neurol. Neurosurg. Psychiatry,* vol. 77, no. 6, pp. 760-766, 2006.

[192] K. Jáuregui-Renaud, F. Y. P. Sang, M. A. Gresty, D. A. Green, and A. M. Bronstein, "Depersonalisation/derealisation symptoms and updating orientation in patients with vestibular disease," *J. Neurol. Neurosurg. Psychiatry,* vol. 79, no. 3, pp. 276-283, 2008.

[193] R. Tschan, J. Wiltink, J. Adler, M. E. Beutel, and M. Michal, "Depersonalization experiences are strongly associated with dizziness and vertigo symptoms leading to increased health care consumption in the German general population," *The Journal of nervous and mental disease,* vol. 201, no. 7, pp. 629-635, 2013.

[194] M. Ertl and R. Boegle, "Investigating the vestibular system using modern imaging techniques—A review on the available stimulation and imaging methods," *J. Neurosci. Methods,* vol. 326, p. 108363, 2019.

[195] M. Ertl, M. Moser, R. Boegle, J. Conrad, P. zu Eulenburg, and M. Dieterich, "The cortical spatiotemporal correlate of otolith stimulation: Vestibular evoked potentials by body translations," *Neuroimage,* vol. 155, pp. 50-59, 2017.

[196] S. Kammermeier, A. Singh, S. Noachtar, I. Krotofil, and K. Bötzel, "Intermediate latency evoked potentials of cortical multimodal vestibular areas: acoustic stimulation," *Clin. Neurophysiol.,* vol. 126, no. 3, pp. 614-625, 2015.

[197] S. Kammermeier, A. Singh, and K. Bötzel, "Intermediate latency-evoked potentials of multimodal cortical vestibular areas: galvanic stimulation," *Front. Neurol.,* vol. 8, p. 587, 2017.

[198] S. Gale *et al.*, "Oscillatory neural responses evoked by natural vestibular stimuli in humans," *J. Neurophysiol.,* vol. 115, no. 3, pp. 1228-1242, 2016.

[199] T. P. Gutteling and W. Medendorp, "Role of alpha-band oscillations in spatial updating across whole body motion," *Front. Psychol.,* vol. 7, p. 671, 2016.

[200] O. Guralnik, J. Schmeidler, and D. Simeon, "Feeling unreal: Cognitive processes in depersonalization," *Am. J. Psychiatry,* vol. 157, no. 1, pp. 103-109, 2000.

[201] N. Schabinger, H. Gillmeister, S. Berti, M. Michal, M. E. Beutel, and J. Adler, "Detached and distracted: ERP correlates of altered attentional function in depersonalisation," *Biol. Psychol.,* vol. 134, pp. 64-71, Apr 2018, doi: 10.1016/j.biopsycho.2018.02.014.

[202] M. I. Posner and Y. Cohen, "Components of visual orienting," *Attention and performance X: Control of language processes,* vol. 32, pp. 531-556, 1984.

[203] S. A. Hillyard and L. Anllo-Vento, "Event-related brain potentials in the study of visual selective attention," *Proceedings of the National Academy of Sciences,* vol. 95, no. 3, pp. 781-787, 1998.

[204] M. I. Posner, C. R. Snyder, and B. J. Davidson, "Attention and the detection of signals," *J. Exp. Psychol. Gen.,* vol. 109, no. 2, p. 160, 1980.

[205] J. Holdstock, "The effect of attention on the P300 deflection elicited by novel sounds," *J. Psychophysiol.,* vol. 9, pp. 18-31, 1995.

[206] J. Polich, "P300 clinical utility and control of variability," *J. Clin. Neurophysiol.,* vol. 15, no. 1, pp. 14-33, 1998.

[207] T. W. Picton, "The P300 wave of the human event-related potential," *J. Clin. Neurophysiol.,* vol. 9, no. 4, pp. 456-479, 1992.

[208] T. Turan, E. Esel, F. Karaaslan, M. Basturk, A. Oguz, and I. Yabanoglu, "Auditory event-related potentials in panic and generalised anxiety disorders," *Prog. Neuropsychopharmacol. Biol. Psychiatry,* vol. 26, no. 1, pp. 123-126, 2002.

[209] S. J. Luck *et al.*, "A roadmap for the development and validation of event-related potential biomarkers in schizophrenia research," *Biol. Psychiatry,* vol. 70, no. 1, pp. 28-34, 2011.

[210] G. Knyazev, J. Y. Slobodskoj-Plusnin, and A. Bocharov, "Event-related delta and theta synchronization during explicit and implicit emotion processing," *Neuroscience,* vol. 164, no. 4, pp. 1588-1600, 2009.

[211] E. Kirmizi-Alsan, Z. Bayraktaroglu, H. Gurvit, Y. H. Keskin, M. Emre, and T. Demiralp, "Comparative analysis of event-related potentials during Go/NoGo and

CPT: decomposition of electrophysiological markers of response inhibition and sustained attention," *Brain Res.,* vol. 1104, no. 1, pp. 114-128, 2006.

[212] O. Jensen and C. D. Tesche, "Frontal theta activity in humans increases with memory load in a working memory task," *Eur. J. Neurosci.,* vol. 15, no. 8, pp. 1395-1399, 2002.

[213] S. Raghavachari *et al.,* "Gating of human theta oscillations by a working memory task," *J. Neurosci.,* vol. 21, no. 9, pp. 3175-3183, 2001.

[214] N. Otsuru *et al.,* "Sensory incongruence leading to hand disownership modulates somatosensory cortical processing," *Cortex,* vol. 58, pp. 1-8, 2014.

[215] S. Rigato, A. J. Bremner, H. Gillmeister, and M. J. Banissy, "Interpersonal representations of touch in somatosensory cortex are modulated by perspective," *Biol. Psychol.,* p. 107719, 2019.

[216] J. Adler and H. Gillmeister, "Bodily self-relatedness in vicarious touch is reflected at early cortical processing stages," *Psychophysiology,* vol. 56, no. 12, p. e13465, 2019.

[217] F. H. Petzschner, L. A. Weber, K. V. Wellstein, G. Paolini, C. T. Do, and K. E. Stephan, "Focus of attention modulates the heartbeat evoked potential," *Neuroimage,* vol. 186, pp. 595-606, 2019.

[218] E. Al *et al.,* "Heart-Brain Interactions Shape Somatosensory Perception and Evoked Potentials," *bioRxiv,* p. 750315, 2019.

[219] B. Lenggenhager, C. Lopez, T. Metzinger, and J. M. Windt, "Vestibular contributions to the sense of body, self, and others," 2015.

[220] C. C. Duncan *et al.,* "Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400," *Clin. Neurophysiol.,* vol. 120, no. 11, pp. 1883-1908, 2009.

[221] B. Gangadhar, J. Ancy, N. Janakiranaiah, and C. Umapathy, "P300 amplitude in non-bipolar, melancholic depression," *J. Affect. Disord.,* vol. 28, no. 1, pp. 57-60, 1993.

[222] D. H. Mathalon, J. M. Ford, and A. Pfefferbaum, "Trait and state aspects of P300 amplitude reduction in schizophrenia: a retrospective longitudinal study," *Biol. Psychiatry,* vol. 47, no. 5, pp. 434-449, 2000.

[223] A. K. Seth, K. Suzuki, and H. D. Critchley, "An interoceptive predictive coding model of conscious presence," *Front. Psychol.,* vol. 2, p. 395, 2012.

[224] A. K. Seth and M. Tsakiris, "Being a beast machine: the somatic basis of selfhood," *Trends Cogn. Sci.,* vol. 22, no. 11, pp. 969-981, 2018.

[225] M. I. Garrido, J. M. Kilner, K. E. Stephan, and K. J. Friston, "The mismatch negativity: a review of underlying mechanisms," *Clin. Neurophysiol.,* vol. 120, no. 3, pp. 453-463, 2009.

[226] S. Dehaene, M. I. Posner, and D. M. Tucker, "Localization of a neural system for error detection and compensation," *Psychol. Sci.,* vol. 5, no. 5, pp. 303-305, 1994.

[227] C. Pfeiffer and M. De Lucia, "Cardio-audio synchronization drives neural surprise response," *Sci. Rep.,* vol. 7, no. 1, pp. 1-10, 2017.

[228] T. Sueyoshi, F. Sugimoto, J. i. Katayama, and H. Fukushima, "Neural correlates of error processing reflect individual differences in interoceptive sensitivity," *Int. J. Psychophysiol.,* vol. 94, no. 3, pp. 278-286, 2014.

[229] Y. Tan, J. Vandeput, J. Qiu, O. Van den Bergh, and A. von Leupoldt, "The error-related negativity for error processing in interoception," *Neuroimage,* vol. 184, pp. 386-395, 2019.

[230] A. Yoris *et al.,* "The inner world of overactive monitoring: neural markers of interoception in obsessive–compulsive disorder," *Psychol. Med.,* vol. 47, no. 11, pp. 1957-1970, 2017.

[231] A. M. Jiménez-Genchi, "Repetitive transcranial magnetic stimulation improves depersonalization: a case report," *CNS spectrums,* vol. 9, no. 5, pp. 375-376, 2004.

[232] C. Spitzer, C. Willert, H.-J. Grabe, T. Rizos, B. Möller, and H. J. Freyberger, "Dissociation, hemispheric asymmetry, and dysfunction of hemispheric interaction: A transcranial magnetic stimulation approach," *The Journal of neuropsychiatry and clinical neurosciences,* vol. 16, no. 2, pp. 163-169, 2004.

[233] G. Gainotti, "The role of the right hemisphere in emotional and behavioral disorders of patients with frontotemporal lobar degeneration: an updated review," *Front. Aging Neurosci.,* vol. 11, p. 55, 2019.

[234] E. K. Silberman and H. Weingartner, "Hemispheric lateralization of functions related to emotion," *Brain Cogn.,* vol. 5, no. 3, pp. 322-353, 1986.

[235] A. Mantovani, D. Simeon, N. Urban, P. Bulow, A. Allart, and S. Lisanby, "Temporo-parietal junction stimulation in the treatment of depersonalization disorder," *Psychiatry Res.,* vol. 186, no. 1, pp. 138-140, 2011.

[236] P. Mowlaee, J. Kulmer, J. Stahl, and F. Mayer, "Fundamentals of Phase-Based Signal Processing," *Single Channel Phase-Aware Signal Processing in Speech Communication: Theory and Practice,* p. 33, 2016.

[237] A. P. Burgess, "Towards a unified understanding of event-related changes in the EEG: the firefly model of synchronization through cross-frequency phase modulation," *PLoS One,* vol. 7, no. 9, 2012.

[238] C. Lopez, E. Nakul, N. Preuss, M. Elzière, and F. W. Mast, "Distorted own-body representations in patients with dizziness and during caloric vestibular stimulation," *J. Neurol.,* vol. 265, no. 1, pp. 86-94, 2018.

[239] O. Blanke, T. Landis, L. Spinelli, and M. Seeck, "Out-of-body experience and autoscopy of neurological origin," *Brain,* vol. 127, no. 2, pp. 243-258, 2004.

[240] C. Lopez and M. Elzière, "Out-of-body experience in vestibular disorders–A prospective study of 210 patients with dizziness," *Cortex,* vol. 104, pp. 193-206, 2018.

[241] C. Lopez, P. Halje, and O. Blanke, "Body ownership and embodiment: vestibular and multisensory mechanisms," *Neurophysiologie Clinique/Clinical Neurophysiology,* vol. 38, no. 3, pp. 149-161, 2008.

[242] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain–computer interfaces for communication and rehabilitation," *Nature Reviews Neurology,* vol. 12, no. 9, pp. 513-525, 2016.

[243] S. N. Abdulkader, A. Atia, and M.-S. M. Mostafa, "Brain computer interfacing: Applications and challenges," *Egyptian Informatics Journal,* vol. 16, no. 2, pp. 213-230, 2015.

[244] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *Journal of neural engineering,* vol. 15, no. 3, p. 031005, 2018.

[245] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of neural engineering,* vol. 16, no. 5, p. 051001, 2019.

[246] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *Journal of neural engineering,* vol. 14, no. 1, p. 016003, 2016.

[247] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.,* vol. 38, no. 11, pp. 5391-5420, 2017.

[248] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, 2008: IEEE, pp. 2390-2397.

[249] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of neural engineering,* vol. 15, no. 5, p. 056013, 2018.

[250] Z.-G. Zhang, J.-L. Yang, S.-C. Chan, K. D.-K. Luk, and Y. Hu, "Time-frequency component analysis of somatosensory evoked potentials in rats," *BioMedical Engineering OnLine,* vol. 8, no. 1, pp. 1-10, 2009.

[251] C. A. E. Kothe and T.-P. Jung, "Artifact removal techniques with signal reconstruction," ed: Google Patents, 2016.

[252] M. A. Klados, C. Papadelis, C. Braun, and P. D. Bamidis, "REG-ICA: a hybrid methodology combining blind source separation and regression techniques for the rejection of ocular artifacts," *Biomedical Signal Processing and Control,* vol. 6, no. 3, pp. 291-300, 2011.

[253] P. Ablin, J.-F. Cardoso, and A. Gramfort, "Faster independent component analysis by preconditioning with Hessian approximations," *IEEE Transactions on Signal Processing,* vol. 66, no. 15, pp. 4040-4049, 2018.

[254] P. Ablin, J.-F. Cardoso, and A. Gramfort, "Faster ICA under orthogonal constraint," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018: IEEE, pp. 4464-4468.

[255] H. Gillmeister, J. Adler, D. Savva, H. Li, and C. Parapadakis, "Atypical multisensory and emotional body perception in adults with symptoms of Depersonalisation-Derealisation Disorder.," (in preparation).

[256] R. C. Kessler, M. Gruber, J. M. Hettema, I. Hwang, N. Sampson, and K. A. Yonkers, "Co-morbid major depression and generalized anxiety disorders in the National Comorbidity Survey follow-up," *Psychol. Med.,* vol. 38, no. 3, pp. 365-374, 2008.

[257] D. Simeon, "Depersonalisation disorder: a contemporary overview," *CNS drugs,* vol. 18, pp. 343-354, 2004.

[258] S. Sole, "Dissociative symptoms and the quality of structural integration in borderline personality disorder," UCL (University College London), 2014.

[259] W. E. Mehling, C. Price, J. J. Daubenmier, M. Acree, E. Bartmess, and A. Stewart, "The multidimensional assessment of interoceptive awareness (MAIA)," *PLoS One,* vol. 7, no. 11, p. e48230, 2012.

[260] A. Salami, J. Andreu-Perez, and H. Gillmeister, "Symptoms of depersonalisation/derealisation disorder as measured by brain electrical activity: A systematic review," *Neurosci. Biobehav. Rev.,* vol. 118, pp. 524-537, 2020.

[261] J. Bridle, A. Heading, and D. MacKay, "Unsupervised classifiers, mutual information and'phantom targets," *Adv. Neural Inf. Process. Syst.,* vol. 4, 1991.

[262] A. Krause, P. Perona, and R. Gomes, "Discriminative clustering by regularized information maximization," *Adv. Neural Inf. Process. Syst.,* vol. 23, 2010.

[263] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical chemistry,* vol. 36, no. 8, pp. 1627-1639, 1964.

[264] R. W. Schafer, "What is a Savitzky-Golay filter?[lecture notes]," *IEEE Signal processing magazine,* vol. 28, no. 4, pp. 111-117, 2011.

[265] E. Santamaría-Vázquez, V. Martínez-Cagigal, F. Vaquerizo-Villar, and R. Hornero, "EEG-Inception: A Novel Deep Convolutional Neural Network for Assistive ERP-

Based Brain-Computer Interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 28, no. 12, pp. 2773-2782, 2020.

[266] M.-H. Lee *et al.,* "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience,* vol. 8, no. 5, p. giz002, 2019.

[267] R. C. Josiassen, C. Shagass, R. A. Roemer, S. Slepner, and B. Czartorysky, "Early cognitive components of somatosensory event-related potentials," *Int. J. Psychophysiol.,* vol. 9, no. 2, pp. 139-149, 1990.

[268] E. Formaggio, S. F. Storti, R. Cerini, A. Fiaschi, and P. Manganotti, "Brain oscillatory activity during motor imagery in EEG-fMRI coregistration," *Magnetic Resonance Imaging,* vol. 28, no. 10, pp. 1403-1412, 2010.

[269] M. M. Müller, T. Gruber, and A. Keil, "Modulation of induced gamma band activity in the human EEG by attention and visual information processing," *International Journal of Psychophysiology,* vol. 38, no. 3, pp. 283-299, 2000.

[270] R. Woolson, "Wilcoxon signed-rank test," *Wiley encyclopedia of clinical trials,* pp. 1-3, 2007.

[271] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika,* vol. 52, no. 3/4, pp. 591-611, 1965.

[272] A. Salami, J. Andreu-Perez, and H. Gillmeister, "EEG-ITNet: An Explainable Inception Temporal Convolutional Network for Motor Imagery Classification," *IEEE Access,* vol. 10, pp. 36672-36685, 2022.

[273] R. F. Tate, "Correlation between a discrete and a continuous variable. Point-biserial correlation," *The Annals of mathematical statistics,* vol. 25, no. 3, pp. 603-607, 1954.

[274] Y. Y. Chen, K. J. Lambert, C. R. Madan, and A. Singhal, "Mu oscillations and motor imagery performance: A reflection of intra-individual success, not inter-individual ability," *Human Movement Science,* vol. 78, p. 102819, 2021.

[275] X. Zhang, Y. Guo, B. Gao, and J. Long, "Alpha frequency intervention by electrical stimulation to improve performance in mu-based bci," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 28, no. 6, pp. 1262-1270, 2020.

[276] A. Bollimunta, J. Mo, C. E. Schroeder, and M. Ding, "Neuronal mechanisms and attentional modulation of corticothalamic alpha oscillations," *Journal of Neuroscience,* vol. 31, no. 13, pp. 4935-4943, 2011.

[277] Y. Bagherzadeh, D. Baldauf, D. Pantazis, and R. Desimone, "Alpha synchrony and the neurofeedback control of spatial attention," *Neuron,* vol. 105, no. 3, pp. 577-587. e5, 2020.

[278] S. M. Doesburg, J. J. Green, J. J. McDonald, and L. M. Ward, "From local inhibition to long-range integration: a functional dissociation of alpha-band synchronization

across cortical scales in visuospatial attention," *Brain research,* vol. 1303, pp. 97-110, 2009.

[279] M. L. Itz, S. R. Schweinberger, and J. M. Kaufmann, "Effects of caricaturing in shape or color on familiarity decisions for familiar and unfamiliar faces," *PLoS One,* vol. 11, no. 2, p. e0149796, 2016.

Appendix (List of Abbreviations)

No.	Abbreviation	Definition
1	ACC	Anterior Cingulate Cortex
2	AI	Artificial Intelligence
3	ANN	Artificial Neural Network
4	ASR	Artefact Subspace Reconstruction
5	AWGN	Additive White Gaussian Noise
6	BCI	Brain-computer Interface
7	CDS	Cambridge Depersonalization Scale
8	CDTW	Continuous Dynamic Time Warping
9	CNN	Convolutional Neural Network
10	DPD	Depersonalisation/derealisation Disorder
11	DSM	Diagnostic and Statistical Manual of Mental Disorders
12	DTW	Dynamic Time Warping
13	EEG	Electroencephalogram
14	ERN	Error-related Negativity
15	FBCSP	Filter Bank Common Spatial Pattern
16	FIR	Finite Impulse Response
17	fMRI	functional Magnetic Resonance Imaging
18	GSR	Galvanic Skin Response
19	HEP	Heartbeat-Evoked Potentials
20	ICA	Independent Component Analysis
21	LMS	Least Mean Square
22	MMN	Mismatch Negativity
23	MSE	Mean Squared Error

24	OCD	Obsessive-compulsive Disorder
25	PET	Positron Emission Tomography
26	PHQ9	Patient Health Questionnaire-9
27	PICARD	Preconditioned ICA
28	PTSD	Post-traumatic Stress Disorder
29	ReLU	Rectified Linear Unit
30	RNN	Recurrent Neural Networks
31	rTMS	repetitive Transcranial Magnetic Stimulation
32	SEP	Somatosensory Evoked Potential
33	SNR	Signal-to-noise Ratio
34	SSVEP	Steady-state Visually Evoked Potential
35	STICSA	State-trait Inventory for Cognitive and Somatic Anxiety
36	TCN	Temporal Convolutional Network