Doctoral Program in

Information and Communication Technologies

University of Granada

Automated Detection of Alzheimer's disease and other neurophysiological applications based on EEG

Author

Eduardo Pérez Valero

Supervisors

Christian A. Morillas Gutiérrez

Miguel Ángel López Gordo

Editor: Universidad de Granada. Tesis Doctorales
Autor: Eduardo Pérez Valero
ISBN: 978-84-1117-688-0
URI: https://hdl.handle.net/10481/80006

Abstract

Electrophysiology (EEG) refers to the technique for the acquisition of brain electrical activity using electrodes placed in the scalp. For the past years, hardware and software advancements have promoted the use of wearable EEG systems in multiple research fields such as neuromarketing, sports performance, or the detection of neurological diseases. Simultaneously, artificial intelligence techniques have experienced a remarkable growth, with machine learning approaches widely spreading across research fields. Thereby, researchers in fields like neuroscience and neuroengineering combine EEG feature extraction techniques with machine learning algorithms to study systems based on brain electrical activity for different applications.

In this thesis work, we investigated a series of applications combining EEG acquisition, signal processing, and machine learning. Particularly, we focused on three main topics: attention, stress, and Alzheimer's disease. With respect to attention, we investigated the real-time detection of the attention exerted to mobile visual stimuli, what could help to enhance attention training therapies. For this purpose, we combined EEG and gamification principles to develop a videogame based on brain activity. Regarding stress, we investigated the use of machine learning models to regress the self-perceived stress level during a stress-relax session. This may contribute to improving stress therapies such as those conducted in special needs education schools, where children are often uncapable of communicating verbally. Finally, regarding Alzheimer's disease, we investigated two main topics focused on delivering new automated detection techniques to assist the clinicians and shorten detection times: (a) an automated approach based on EEG activity and machine learning for the real-time detection of early Alzheimer's disease, and (b) the detection of cognitive impairment via a computerized cognitive task evaluating visual dynamics.

To investigate the mentioned applications, we conducted diverse research studies. As a result, we published a series of articles in scientific journals which we have gathered into a compendium for this thesis. In addition to this compendium, in this document we introduce other valuable works conducted in collaboration with other research teams which have not been published yet.

The results originated from the research works presented in this thesis have demonstrated scientific value as indicated by their publication in international scientific journals with high impact factor. Therefore, this work may provide valuable scientific insights which could impact multiple areas in research and society, such as special needs education, stress assessment for professional training or sports performance, and neurological diseases detection.

Resumen

La electrofisiología (EEG) es una técnica para la adquisición de la actividad eléctrica del cerebro mediante el uso de electrodos colocados en el cuero cabelludo. Recientemente, los avances en tecnologías hardware y software han promovido el uso de sistemas EEG vestibles en múltiples campos de investigación como el neuromarketing, el deporte, o la detección de trastornos neurológicos. Simultáneamente, las tecnologías basadas en inteligencia artificial han experimentado un gran crecimiento, con aproximaciones basadas en aprendizaje automático extendiéndose por diversos campos de investigación. De este modo, los investigadores en campos como la neurociencia o la neuroingeniería, combinan técnicas de extracción de características EEG con aprendizaje automático para estudiar sistemas basados en la actividad cerebral para diversas aplicaciones.

En este trabajo de tesis, hemos estudiado una serie de aplicaciones que combinan la adquisición de actividad EEG, el procesamiento de señal, y el aprendizaje automático. Concretamente, nos hemos centrado en tres temas principales: la atención, el estrés, y la enfermedad de Alzheimer. Con respecto a la atención, hemos investigado la detección en tiempo real de la atención ejercida a estímulos visuales móviles. Para ello, hemos combinado principios de gamificación con la estimulación EEG para diseñar un videojuego basado en la actividad cerebral que podría contribuir a mejorar las terapias de entrenamiento de la atención. En el caso del estrés, hemos investigado la predicción cuantitativa del estrés auto percibido durante una sesión de estrés y relajación. Esta aproximación podría contribuir a mejorar las terapias de control de estrés llevadas a cabo en escuelas de educación especial, donde los alumnos a menudo son incapaces de comunicarse verbalmente. Finalmente, en relación con el Alzheimer, hemos investigado dos temas principales relacionados con la generación de nuevos métodos de detección para asistir a los profesionales clínicos y acortar los tiempos de diagnóstico: (a) la detección automática del Alzheimer temprano a través de un sistema EEG vestible y aprendizaje automático, y (b) la detección del deterioro cognitivo mediante una tarea mental que evalúa las dinámicas visuales.

Para investigar estas aplicaciones, hemos llevado a cabo una serie de estudios de campo cuyos resultados han sido publicados en revistas científicas de impacto. Dichos artículos han sido agrupados en un compendio para esta tesis. Además, en este documento reportamos trabajos adicionales realizados en colaboración con otros equipos de investigación y que aún no han sido publicados. Por todo ello, la actividad investigadora recogida en esta tesis ha demostrado tener valor científico. De este modo, este trabajo puede proporcionar resultados relevantes para diversas áreas de la investigación y la sociedad, como la educación especial, la evaluación del estrés, el rendimiento deportivo o la detección de trastornos neurológicos.

Acknowledgments

First, I would like to thank my supervisors, Christian Morillas and Miguel Ángel López, for mentoring me over the last years. They generously accepted me in their research team, trusted me, and inculcated me priceless skills not only limited to research. Likewise, I would like to thank Francisco Pelayo and Jesús Minguillón for their support during my period in the doctoral program. They always helped me and we share fond memories from our time working together. They all had a profound impact on me at a personal and professional level.

Subsequently, I would like to thank all the educational and clinical professionals, psychologists and neurologists, who collaborated in the studies presented in this thesis. I would like to specially thank María José Sánchez at San Rafael special needs education school of Granada, and Rosa Vílchez, Ismael Carrera, and Samuel López at Hospital Universitario Virgen de las Nieves de Granada. My recognition extends also to the volunteers who participated in these studies.

I would like to thank my friends as well, Miguel in particular, for his support throughout this period and for the endless conversations we shared during these years.

Most importantly, I would like to thank my family, my parents and my brother, for their absolute support and blind trust. I know they always try to advise me from love and with the best of intentions. Lastly, I would like to thank Lucía, for her boundless encouragement during this time and for sharing her life with me.

Content index

List of acronyms

AB: Attentional blink.

AD: Alzheimer's disease.

AM: Hypertext markup language.

BAGEN: Batería abreviada Granada de evaluación neuropsicológica.

BCI: Brain-computer interface.

CBNU: Cognitive and behavioral neurology unit.

CSF: Cerebrospinal fluid.

DT: Decision tree.

EEG: Electrophysiology.

ERP: Event related potential.

FIR: Finite impulse response.

HC: Hjorth complexity.

HC: Healthy controls.

HMD: Head-mounted display.

HVN: Hospital Virgen de las Nieves.

ICA: Independent component analysis.

IIR: Infinite impulse response.

ISCEV: International Society for Clinical Electrophysiology of Vision.

LOSO: Leave-one-subject-out.

LR: Logistic regression.

MCI: Mild cognitive impairment.

MEG: Magnetoencephalography.

MIST: Montreal imaging stress test.

MLP: Multi-layer perceptron.

MMSE: Mini mental state examination.

MoCA: Montreal cognitive assessment.

MRI: Magnetic resonance imaging.

MSPE: Mean squared percentage error.

PET: Positron emission tomography.

PSD: Power spectral density.

RF: Random forest.

RP: Relative power.

RSVP: Rapid serial visual presentation.

SE: Spectral entropy.

SEM: Standard error of the mean.

SNR: Signal-to-noise ratio.

SPECT: Single-photon emission computed tomography.

SPSL: Self-perceived stress level.

SSVEP: Steady-state visually evoked potentials.

SVM: Support vector machine.

UGR: University of Granada.

VEP: Visually evoked potentials.

VR: Virtual reality.

List of figures

List of tables

List of equations

Chapter 1: Introduction

In this chapter, we provide a review of the motivation of this thesis, the objectives identified during the elaboration of the research plan, the list of publications derived from this research, and the organization of the present document.

1.1. Motivation

This thesis stems from the research work I performed at the Brain Computer Interfaces (BCI) laboratory of the University of Granada (UGR) over the three year period between 2020 and 2022. During this period, I participated in diverse research projects involving stress assessment, attention detection, and early Alzheimer's disease (AD) diagnosis. These projects are listed below:

- Project XENSORY-UGR (UGR Multimodal Extended Reality Application for Multisensory Interaction) funded by UGR (Ref: PP2021.PP-28).
- Project HPEE-COBE (New Computing Paradigms and Heterogeneous Parallel Architectures for High-Performance and Energy Efficiency of Classification and Optimization Tasks on Biomedical Engineering Applications) funded by the Ministry of Science, Innovation and Universities of the Spanish Government (Ref. PGC2018-098813-B-C31).
- Project DIPRAL (Aid in the Early Diagnosis Of Alzheimer's by Analyzing Brain Activity Based on Low-Cost EEG Recordings) within the framework of the European Regional Development Fund of the Andalusia Operational Programme. (Ref. B-TIC-352-UGR20).

Although the objectives of these projects are different, they all share a common framework: the analysis of brain electrical activity from electrophysiology (EEG) recordings. Briefly, EEG is a neurophysiology technique to record the electrical activity of the brain through electrodes placed at the scalp. This technique has been widely studied for the last century in fields such as medicine and psychology.

However, due to the technical advances in hardware and software over the last two decades, access to EEG systems has grown, capturing the interest of researchers with multiple backgrounds. As a result, today EEG is studied in a wide range of applications including: neuromarketing[1,2], sports performance[3–5], workplace safety[6–8], and detection of neurological diseases[9–11], among others[12–14]. A comprehensive introduction to EEG fundamentals is provided in subsection 2.1.

In parallel, machine learning and signal processing techniques have experienced a striking expansion due to the growing availability of powerful personal computers resulting from the miniaturization of semiconductors and the cost reductions. Signal processing techniques enable the extraction of informative features from raw brain electrical activity, whilst machine learning algorithms enable the development of complex prediction models.

With this in mind, the motivation of this thesis stems from the potential of using wearable EEG systems, signal processing, machine learning, and software development to contribute to scientific research in three main applications: detection of attention, stress, and early AD.

1.2. Objectives

Based on the motivation outlined in the previous section, the main objectives of this thesis work are the following:

- To investigate the combination of EEG activity and the principles of gamification for the real-time detection and potential training of the attention to mobile visual stimuli. This may positively impact attentional training therapies in centers like special needs schools, since scholars may benefit from gamified approaches based on endogenous information such as EEG. For this objective, the implementation of a videogame based on an automated closed-loop BCI architecture and the application of steady state visually evoked potentials (SSVEP) may represent a feasible approach.
- To explore machine learning models for the regression of the self-perceived stress level (SPSL) from EEG activity as an alternative to the classification approaches in the literature, which are typically limited to three levels. For this purpose, a stress-relax session represents an appropriate experimental protocol. In this context, the use of standard techniques to elicit psychosocial stress is required, whilst the use of virtual reality (VR) to simulate the relax stage of the session would be beneficial, since it may contribute to opening the door for ubiquitous immersive relax experiences.

- Healthcare systems congestion often leads to a reduction of medical consulting times and limits the access to certain medical trials. Neurological diseases detection, and AD detection particularly, are affected by these unfavorable circumstances. With this in mind, we defined two main research applications regarding the detection of early AD:
  - An automated online detection approach based on a resting state acquisition using a wearable EEG system and machine learning. Such approach could alleviate the need for trained personnel and shorten detection times. This could contribute to reducing the cost associated to AD detection whilst representing a less stressing alternative from the standpoint of the patients. For this purpose, research into EEG artifact rejection, feature extraction, and stimulation protocols is mandatory.
  - Given cognitive impairment detection is a required step for AD diagnosis, novel automated approaches for cognitive impairment detection could potentially improve current screening protocols. This could be achieved through the implementation of a cognitive task from a software application especially designed for clinicians. Consequently, application development technologies must be investigated to develop a simple and intuitive software which fits into the tight screening protocols conducted by the clinicians. Furthermore, cognitive tasks need to be investigated in order to select a proper evaluation with sound discrimination capability and suitable features, like computerization, auto-scoring, and briefness. This research must be conducted with a direct view towards transferring the potential outcomes to daily-life clinical practice. Derived from this, another objective linked to cognitive impairment detection is the creation of a longitudinal database with patient evaluations and diagnostic information. For this, the software application designed must have multi-user capabilities and high accessibility, in order for multiple clinicians to use it remotely.

1.3. Publications

As a result of the research work reported in this thesis, eight original research articles have been published in journals indexed in the journal citation reports (see Table 1.1). Six of these publications, those where the author of this thesis is the first author, comprise the compendium of articles for this thesis. These works are listed in the next page.

1. **Perez-Valero, E.** et al. An Automated Approach for the Detection of Alzheimer's Disease From Resting State Electroencephalography. Front. Neuroinform. 16, 924547 (2022).

2. **Perez-Valero, E.,** Lopez-Gordo, M. Á., Gutiérrez, C. M., Carrera-Muñoz, I. & Vílchez-Carrillo, R. M. A self-driven approach for multi-class discrimination in Alzheimer's disease based on wearable EEG. Computer Methods and Programs in Biomedicine 220, 106841 (2022).

3. **Perez-Valero, E.,** Vaquero-Blasco, M. A., Lopez-Gordo, M. A. & Morillas, C. Quantitative Assessment of Stress Through EEG During a Virtual Reality Stress-Relax Session. Front. Comput. Neurosci. 15, (2021).

4. **Perez-Valero, E.,** Lopez-Gordo, M. A. & Vaquero-Blasco, M. A. EEG-based multi-level stress classification with and without smoothing filter. Biomedical Signal Processing and Control 69, 102881 (2021).

5. **Perez-Valero, E.,** Lopez-Gordo, M. A., Morillas, C., Pelayo, F. & Vaquero-Blasco, M. A. A Review of Automated Techniques for Assisting the Early Detection of Alzheimer's Disease with a Focus on EEG. Journal of Alzheimer's disease: JAD (2021) doi:10.3233/JAD-201455.

6. **Perez-Valero, E.,** Lopez-Gordo, M. A. & Vaquero-Blasco, M. A. An attention-driven videogame based on steady-state motion visual evoked potentials. Expert Systems n/a, e12682 (2021).

7. Vaquero-Blasco, M. A., **Perez-Valero, E.,** Lopez-Gordo, M. A. & Morillas, C. Virtual Reality as a Portable Alternative to Chromotherapy Rooms for Stress Relief: A Preliminary Study. Sensors 20, 6211 (2020).

8. Vaquero-Blasco, M. A., **Perez-Valero, E.,** Morillas, C. & Lopez-Gordo, M. A. Virtual Reality Customized 360-Degree Experiences for Stress Relief. Sensors 21, 2219 (2021).

The six articles included in the compendium are annexed in Appendices A to F. Additionally, we also presented a conference paper during the International Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC) held in Tenerife in 2022:

- **Perez-Valero, E.,** Minguillon, J., Morillas, C., Pelayo, F. & Lopez-Gordo, M. A. Detection of Alzheimer's Disease Using a Four-Channel EEG Montage. in Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications (eds. Ferrández Vicente, J. M., Álvarez-Sánchez, J. R., de la Paz López, F. & Adeli, H.) 436–445 (Springer International Publishing, 2022). doi:10.1007/978-3-031-06242-1_43.

Table 1.1: Metrics of the journals where we have published the compendium articles.

Journal	IF	Research areas	Quartile
Journal of Alzheimers disease	4.16	Neurosciences	Q2
Frontiers in Neuroinformatics	3.73	Mathematical and Computational Biology	Q2
		Neurosciences	Q2
Computer Methods and Programs in Biomedicine	7.03	Computer Science, Interdisciplinary Applications	Q1
		Computer Science, Theory and Methods	Q1
		Engineering, Biomedical	Q1
		Medical Informatics	Q1
Frontiers in Computational Neuroscience	3.39	Mathematical and Computational Biology	Q2
		Neurosciences	Q3
Biomedical Signal Processing and Control	5.08	Engineering, Biomedical	Q2
Expert Systems	2.81	Computer Science, Artificial Intelligence	Q3
		Computer Science, Theory & Methods	Q2
Sensors	3.84	Chemistry, Analytical	Q2
		Engineering, Electrical and Electronic	Q2
		Instruments and Instrumentation	Q1

1.4. Organization

The remainder of this document is organized as follows: in Chapter 2, we provide background information regarding EEG and machine learning; in Chapter 3, we review the previous works in the literature on the detection of AD through EEG and machine learning; in Chapters 4 to 6, we review the different research works included in this thesis regarding attention detection, stress detection, and early AD detection, respectively; in Chapter 7, we briefly describe the research collaborations conducted in the scope of this thesis; and finally, in Chapter 8, we provide a set of conclusions and report the contributions drawn from this thesis work, including scientific impact, limitations, and future work.

Chapter 2: Fundamentals

In this chapter, we review the background information regarding the topics covered in this thesis work. We survey the basics of EEG, including recording, filtering, and processing, and we explicate the core elements of machine learning, including the main concepts, algorithms, and procedures.

2.1. Background on EEG

EEG is a neurophysiology technique to record the electrical activity of the brain through electrodes located at the scalp (see Figure 2.1). This technique was firstly applied by German neurologist Hans Berger during the late twenties of the 20th century[15]. The electrical activity recorded by the electrodes is the result of the synchronous firing of groups of cortical neurons sharing a common orientation, which results in the addition of the electrical fields generated.

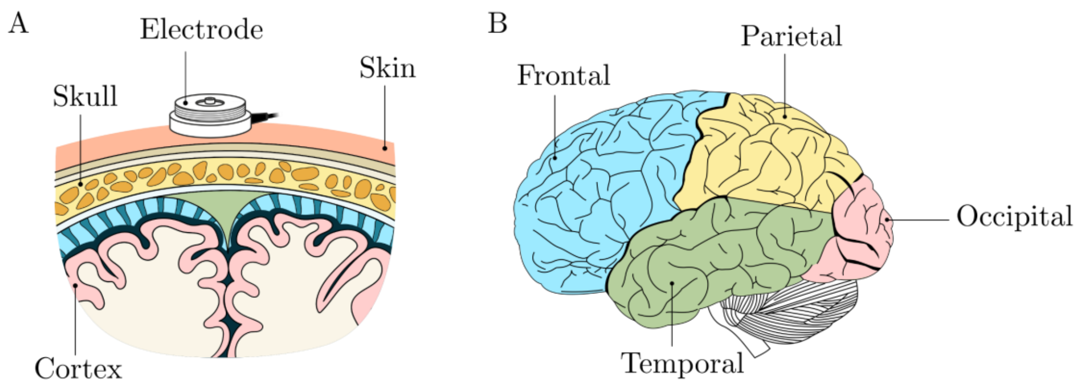


Figure 2.1: EEG acquisition conceptualization. (A) Tissue layers and electrode positioning during an EEG acquisition. Adapted from nagwa.com, and (B) Brain lobes. Adapted from wikipedia.org.

2.1.1. Montages

To conduct an EEG acquisition, the electrodes must be positioned in the scalp according to a particular arrangement referred to as montage. The most extended montages are the 10 – 20 and 10 – 10 International System montages, which were created to standardize the recording methodology. The names of these montages indicate the distance between neighboring electrodes. For instance, in the 10 – 20 International System montage, the distance between adjacent electrodes is 10% and 20% of the front-back and right-left distance of the nasion-to-inion separation, respectively. Within these montages, each electrode is referred by a name consisting of one or two letters (indicating the scalp region) and a number (for the electrodes located in the two hemispheres) or a *z* letter for those located along the mid longitudinal fissure. Figure 2.2 represents the 10-20 International System montage and the nasion-to-inion distance.

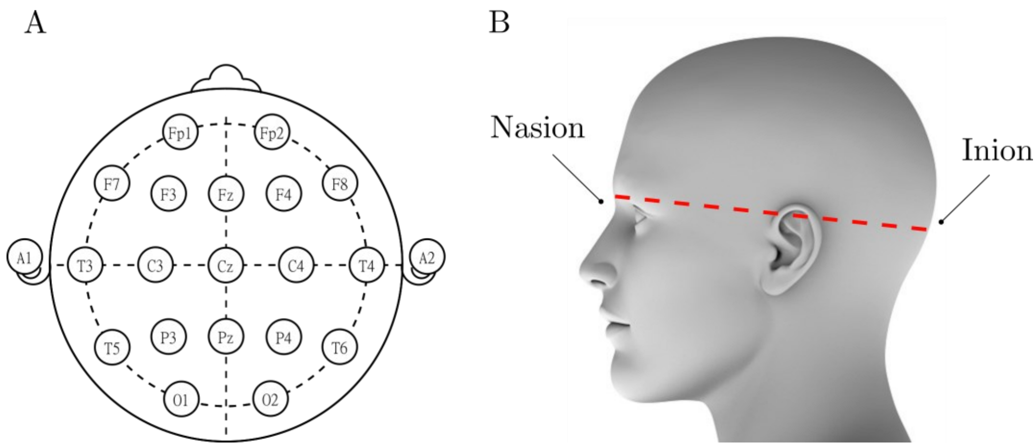


Figure 2.2: 10-20 International System montage. Source: wikipedia.org. (A) Spatial distribution of the channels and (B) Distance from nasion to inion.

At this point, it is important to discern two important concepts: electrode and channel. Electrode refers to the physical sensor positioned in the scalp, whilst channel refers to the measurement of an electrical signal under a particular scalp position. Since EEG measures voltage, a difference in electrical potential between two locations, at least two electrodes are needed for measuring: an active electrode and a reference. Typical montages use an electrode placed in the ear lobe or the mastoid as reference for all measurements. Additionally, a ground electrode is used to minimize the noise captured during measuring. If single-ended measuring is utilized, the ground electrode and the reference electrode are the same, whilst if differential measuring is used, a separate ground electrode is included. Differential measuring is usually more convenient because all that is shared by the two entries of the amplifier is suppressed if the impedances are sufficiently low, what further prevents from introducing electrical noise.

Equation 2.1 and Equation 2.2 represent the theoretical output of single-ended and differential measurings, respectively, using Cz as the active electrode:

$$\mathrm{V} = \mathrm{E_{Cz}} - \mathrm{E_{reference}} + \mathrm{noise}$$

Equation 2.1: Single-ended voltage measuring.

$$\mathrm{V} = \mathrm{E_{Cz}} - \mathrm{E_{ground}} + \mathrm{noise} - \mathrm{E_{reference}} + \mathrm{E_{ground}} - \mathrm{noise} + \mathrm{noise}'$$

Equation 2.2: Differential voltage measuring.

V represents the measured voltage and *E* represents electrical potential at the different positions.

2.1.2. Recording systems

EEG acquisition systems are used to record electrophysiological activity and they can be classified according to different features. Some of the most important features include:

- Connectivity. Depending on the communication capabilities of the system, wireless and wired recording devices can be found. For instance, modern wearable systems very often include a Bluetooth module to forward the electricals signals captured by the electrodes to a computer for visualization and processing.
- Electrode type. Formerly, EEG recording systems required the application of conductive gel to reduce the impedance between the scalp and the sensors in order to increase the signal-to-noise ratio (SNR). This procedure was impractical and spanned the preparation time. Conversely, many current recording systems use dry or water-based electrodes, which are more practical and comfortable from the standpoint of the participants.
- Electrode number. Different montages require different number of electrodes. The higher the number of electrodes, the longer the preparation time, and therefore, the less optimal the montage is for wearable applications. Typical montages include 16, 32, 64, and 128 electrodes.
- Sampling rate. This refers to the number of samples acquired per second by the recording system. Usual values for the sampling rate are 256 Hz, 512 Hz, and 1 kHz.

For the studies conducted for this thesis work, we used two different recording systems: the RABio w8[16], developed by the BCI laboratory at Universidad de

Granada (Granada, Spain), and the Versatile 16 system commercialized by Bitbrain (Zaragoza, Spain).

A B

Figure 2.3: EEG systems used in this thesis. (A) RABio w8, (B) Versatile 16.

A comparison of the main features of these EEG acquisition systems is shown in the table below.

Table 2.1: Comparison of EEG recording systems utilized in this thesis.

Feature	RABio w8	Versatile 16
Maximum sampling rate	1 kHz	256 Hz
Number of channels	8	16
Electrode type	Gel-based	Water-based
Connectivity	Bluetooth	Bluetooth
Recording	Computer	Computer or memory card

2.1.3. Filtering

As outlined in the beginning of this chapter, EEG originates from the simultaneous firing of groups of neurons in the brain cortex. Since the electrical activity generated by these groups of neurons crosses multiple tissue layers before reaching the scalp, the signals captured by the electrodes are highly attenuated. As a result, the amplitude of the EEG recordings is in the range of microvolts. Furthermore, EEG is highly affected by artifacts, noise originated from different sources which contaminates brain activity. Different sources of artifacts include the following:

- Physiological: ocular movements, heart beating, muscle activity, etc.
- Technical: loose sensors, power line interference, body movements, etc.

As a consequence of this, signal processing techniques are crucial to clean artifacts and to extract explainable features from the EEG activity. In this context, digital filtering represents an essential EEG processing step, enabling the suppression of undesirable features of the signal. Most frequently, filters are applied to attenuate a particular range of frequencies from the spectrum of the signal through the convolution of the signal and the filter kernel. For electrophysiology activity, the main frequency bands, also known as rhythms, are listed in Table 2.2.

Table 2.2: Main EEG rhythms and their frequency range.

Rhythm	**Frequency range (Hz)**
Delta	0.1 – 4
Theta	4 – 8
Alpha	8 – 13
Beta	13 – 30
Gamma	30 – 100

Based on the spectral range of these rhythms, EEG signals are typically filtered between 0.1 and 100 Hz using a bandpass filter. In addition, a notch filter is often applied to reject the power line interference, which is located at 50 Hz or 60 Hz depending on the country. Since there are multiple types of digital filters, and filtering is not the main topic of this thesis, here we break digital filters using a basic classification according to their impulse response.

- Finite impulse response (FIR) filter. The filtered signal is estimated using only previous values (b coefficients) of the original signal.

$$y(n) = \sum_{k=0}^{M} b_k \cdot x(n-k)$$

Equation 2.3: Filtered signal estimation for FIR filters.

- Infinite impulse response (IIR) filter. The filtered signal is estimated using previous values of the original signal (b coefficients) and previous values of the filtered signal (a coefficients).

$$y(n) = \sum_{k=0}^{M} b_k \cdot x(n-k) - \sum_{k=1}^{N} a_k \cdot y(n-k)$$

Equation 2.4: Filtered signal estimation for IIR filters.

On the one hand, IIR filters are less numerically stable compared to FIR filters, because they are naturally recursive, and therefore, they can accumulate errors. On the other hand, IIR filters have a much lower number of coefficients compared with FIR filters, hence, they are often considered as more computationally efficient. However, since FIR filters are always stable and easier to control, they are usually preferred unless the application involves strict time constraints.[17]

2.1.4. Features

After filtering, EEG recordings are typically partitioned into segments of a few seconds known as epochs. Then, noisy epochs are discarded by visual inspection or using artifact rejection algorithms. Finally, signal processing techniques are applied to the epochs in order to extract meaningful features. Although there exist a myriad of features, in this subsection we review those relevant for the applications concerning this thesis work.

Depending on the application and the recording condition, different features can be extracted from the EEG recordings. In resting state, which is one of the most extended conditions, the participants typically sit in a chair with their eyes open or closed while their brain activity is recorded. Resting state activity has been studied in applications such as the analysis of brain activity synchronization in AD[18], the assessment of attention-deficit hyperactivity disorder[19], the evaluation of the functional connectivity in endurance athletes[20], etc. Some of the most extended features analyzed during resting state include:

- Power spectral density (PSD). These features represent the power of the EEG signal over a particular range of frequencies. Typically, the relative power is obtained, which represents the proportion of the total power of the signal that is contained in a particular frequency range. PSD features can be estimated using a wide variety of methods based on the Fourier transform. The Fourier transform is one of the basic algorithms in signal processing and enables the decomposition of a signal into a sum of sine waves of different frequencies. Since EEG signals are non-stationary, the PSD is different over different epochs. To cope with this, Welch's method is one of the most extended methods to estimate the PSD. Using this method, the epochs are first windowed to avoid edge effects; then, the spectral density of the windowed epochs is estimated; finally, the spectral density of the different epochs is averaged. Overlapping can also be applied during epoch segmentation, what contributes to reducing the variance inherent to the EEG non-stationarity.
- Complexity. This feature family quantifies the irregularity and predictability of the EEG signals, most frequently using non-linear analysis. An important group of these features are derived from entropy theory (sample entropy, fuzzy entropy, spectral entropy, etc.)[21–23], whilst others are based on more heterogeneous analyses (Lempel-Ziv complexity, Hjorth complexity, fractal dimension, line length, etc.)[24–26].
- Connectivity. While the two previous feature families are univariate (they are estimated for each channel individually), connectivity features measure

the synchronization between pairs of channels. Examples of these features include coherence[27], Granger causality[28], phase lag index[29], correlation[30], mutual information[31], etc.

Alternatively, there exist other features derived from stimulation paradigms, which are also extensively studied in the literature:

- Event related potentials (ERP). This concept refers to the electrical potentials generated in the brain time-locked to a particular auditory, cognitive, or motor event[32]. ERPs are named according to the sign of the fluctuation and to its time distance from the triggering event (latency). For instance, N100 represents a negative fluctuation elicited approximately 100 ms after the stimulus presentation (see Figure 2.4). For ERP analysis, the event is typically triggered over several trials and the potentials elicited during the trials are averaged to enhance the SNR. The averaged signal is used to evaluate sensory and cognitive processes in the brain. Particularly, early potentials (P50, N100, P200, etc.) are considered exogeneous, since they represent sensory processes relying on the physical features of the stimulus (although they can be modulated by higher cognitive functions like attention). In contrast, late potentials (P300, N400, P600, etc.) are considered endogenous, because they reflect complex cognitive processes in the brain. The P300 is arguably the most extensively studied ERP, and represents an index of executive control, particularly working memory and context updating processes. This potential has been widely studied in areas such as dementia detection[33–35], aging[36], brain-computer interface control[37], etc. As mentioned above, ERPs can be elicited using different types of stimuli, which are generated via audio, motor, or visual paradigms. Particularly, visual potentials have been extensively studied in the literature.

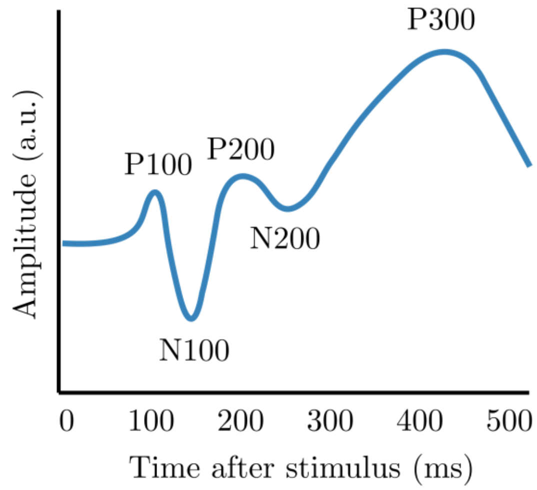


Figure 2.4: Illustration of main EEG event related potentials.

- Visually evoked potentials (VEP). As a response to certain visual stimuli, the visual cortex generates electrical signals referred to as VEP. According to the ISCEV (International Society for Clinical Electrophysiology of Vision), there are two main VEP stimulation classes: pattern and flash[38]. These stimuli are often presented as checkerboards which reverse their pattern or are shown and hidden periodically, while preserving the average luminance.

  When the visual stimulus flickers in a particular range of frequencies (4 Hz to 75 Hz)[39], the response signals generated by the visual cortex are almost sinusoidal. Such response is known as steady-state VEP (SSVEP), and its PSD expands over a narrow window centered at the frequency of the flickering stimulus[40]. Typically, the stimulus used to elicit the SSVEP is a checkerboard reversing its pattern at a particular frequency. These potentials have been widely studied for applications such as detection of visual path affections[41], attention in the workplace[42], febrile seizures[43], and BCI control[44,45].

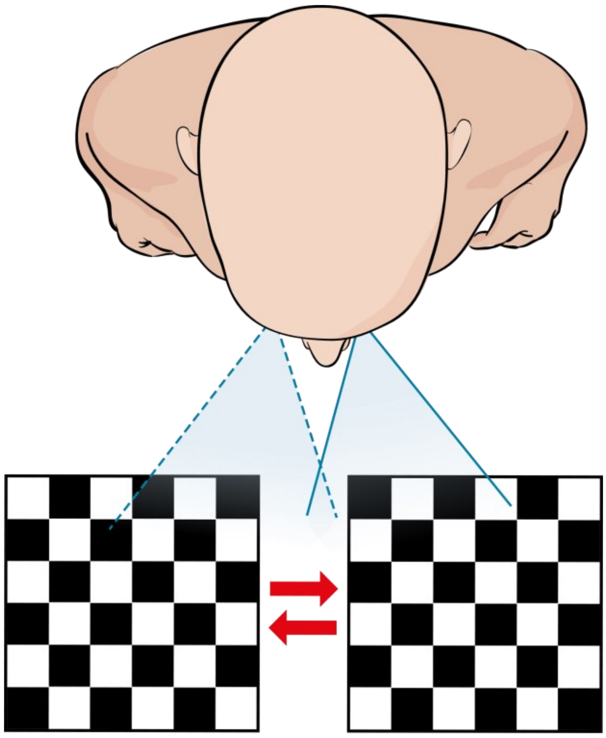


Figure 2.5: VEP reversing checkerboard concept. Source: tidsskriftet.no

2.2. Background on machine learning

Machine learning is a subfield of artificial intelligence dedicated to the implementation of systems that learn patterns from data. Machine learning algorithms are designed to uncover the patterns underlying large data structures, what would be unfeasible by manual processing, and use them to create models. For the past decade, machine learning has spread fast under different aliases like predictive analytics or data mining, owing to the increased accessibility to powerful personal computer systems. Machine learning algorithms are applied to a large number of applications, including fraud detection[46], image classification[47], weather forecast[48], medical imaging analysis[49], neurological disorders classification[50], or financial risk assessment[51], among others[52]. Since these algorithms have been applied in the context of this thesis work, in this subsection we briefly cover the main concepts in the field.

2.2.1. Supervised vs unsupervised

Depending on the data structures used to develop de model, machine learning algorithms can be grouped into supervised an unsupervised learning.

- Supervised. In this kind of learning, the algorithm is input with a feature matrix and a target array. The feature matrix holds the different attributes of the dataset in the columns, and the observations (or samples) in the rows, while the target holds the outcome (or label) corresponding to each observation. Given these two data structures, the machine learning algorithm builds the model in essentially two steps: training and test. Through training, the model coefficients are tuned to fit the underlying mapping function of the algorithm to the data in order to learn its patterns. Alternatively, through test, the algorithm evaluates how accurately the model generalizes to unseen observations. Depending on the target, supervised learning may be applied to classification or regression problems.

- Unsupervised. In this case, the machine learning algorithm is input only with a feature matrix, and hence, no labels are associated with the dataset observations. In this case, the algorithm follows a particular strategy to group the dataset observations and training is not required. The main tasks performed by unsupervised algorithms are clustering (grouping the unlabeled observations into different groups) and dimensionality reduction (removing data features whilst preserving the integrity when the dimensionality of the dataset is too high).

2.2.2. Classification vs regression

As outlined in the previous subsection, depending on the target type, supervised machine learning problems can be divided into classification and regression.

- Classification. In classification problems, the target ranges over a reduced number of categories (see Figure 2.6.A). For instance, a classical machine learning classification problem consists in predicting the survival of the passengers in the Titanic[53]. Here, the features include passenger information like sex, age, cabin number, or boarding pass cost, and the target is a binary variable with possible values of zero (did not survive) and one (survived). Hence, the dataset consist of a feature matrix (with columns indicating the passenger features and rows corresponding to passenger observations) and a target array (indicating if the corresponding passenger survived). The goal of this machine learning problem is to build a model that, given unseen passenger observations, predicts if they survived.

- Regression. In regression problems, the target is a continuous variable (see Figure 2.6.B). For instance, another classical machine learning regression problem consists in predicting Boston housing prices[54]. In this dataset, the features include housing information like number of rooms, neighborhood, or size in square feet, and the target is the house price in dollars. In this case, the dataset consist of a feature matrix (with columns indicating the house features and rows corresponding to house observations) and a target array (indicating the corresponding house price). Accordingly, the goal of this machine learning problem is to build a model that, given unseen house observations, accurately predicts their price.

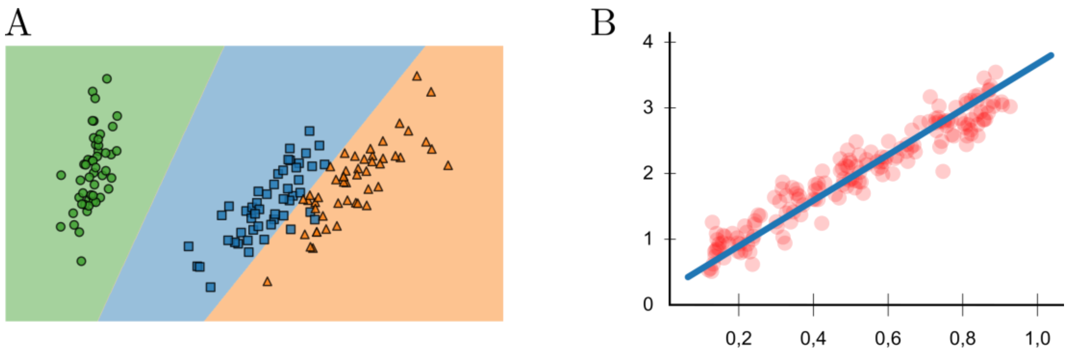


Figure 2.6: Classification and regression concepts. (A) 3-class classification problem and (B) Regression problem.

2.2.3. Overfitting vs underfitting

The ultimate goal of every machine learning system is to maximize generalization. Generalization measures how well the underlying mapping function that the algorithm fits to the training data performs on new data. We can detect when the model does not generalize well through bias and variance.

- Bias. Considering repeated samples of the training set, bias is the difference between the average model prediction and the true value of the observation being predicted.
- Variance. Considering repeated samples of the training set, variance is the variability in the prediction of a particular observation.

Figure 2.7 shows a graphical representation of the bias and variance problems considering a bullseye as the true value of the observation to be predicted.

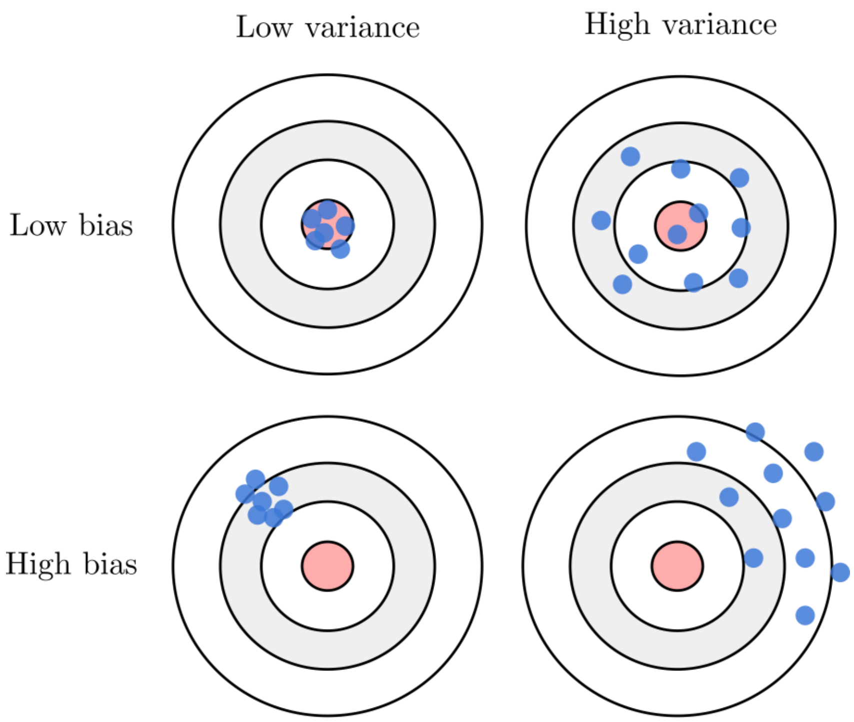


Figure 2.7: Representation of bias and variance. Source: www.endtoend.ai

There are two main causes for bias and variance in machine learning:

- Underfitting. When the model performs bad on the training data, we say the model is underfit. This can occur when the number of training observations is too low or when the model mapping function is too simple, what results in a highly biased model.

- Overfitting. When the model performs well on the training data but generalizes badly (performs bad on the test data), we say the model is overfit. This often occurs when the model mapping function is too complex. In this case, the model memorizes even the slightest variations in the training data (basically noise), but does not learn the underlying pattern of the dataset. Consequently, the model performs accurately in the training set but badly in the test set, what produces high variance.

With this in mind, there is a trade-off between bias and variance in machine learning problems. If the model is oversimplistic, it will not even learn the patterns underlying the data, however, if the model is too complex, it will memorize even the noise and it will generalize badly. Ideally, we need to find a middle point where the model complexity is appropriate so bias and variance are low, and generalization is sound.[52] Figure 2.8 represents the bias-variance tradeoff described in this paragraph.

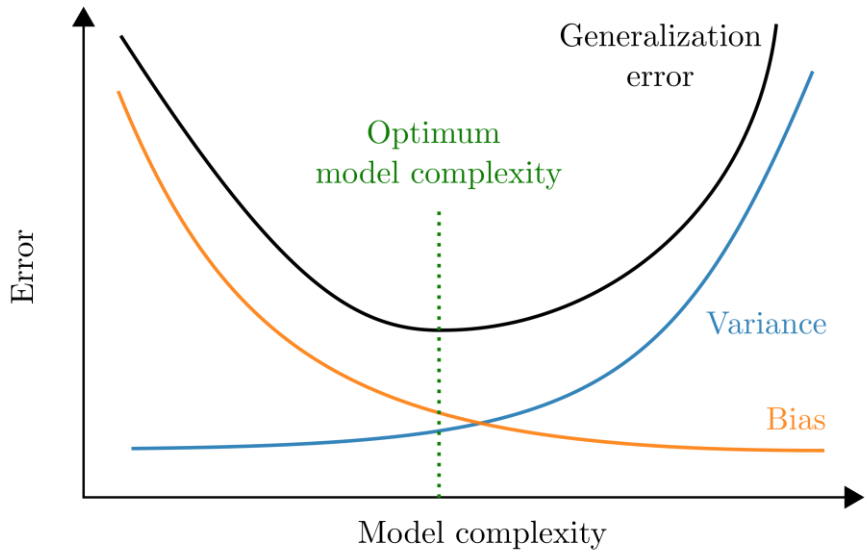


Figure 2.8: Illustration of the bias-variance tradeoff in machine learning.

2.2.4. Cross-validation

A basic approach to evaluate the performance of a machine learning model would be to split the dataset into training and test set (typically 75% and 25% of the data, respectively), then use the training set to tune the model coefficients, and the test set to evaluate its performance. Note that, in this case, the values of the model coefficients and its performance depend on the particular split of the data. This means that, if we splitted the data differently, we would obtain a different model and consequently, a different performance.

To cope with this, cross-validation is extensively used in the machine learning literature. In cross-validation, the dataset is divided into N partitions or folds. Then, the algorithm is trained and evaluated for N iterations. For iteration i, the algorithm is trained using all the data but fold i, which is subsequently used to test the model performance. After completing all the iterations, the performance obtained in all the iterations is averaged and regarded as the general performance of the model. It is important to note that, in every iteration, the model coefficients vary due to the variability in the training set, thus, a common practice is to train the model on the whole dataset once the cross-validation performance is sound. There exist multiple cross-validation strategies, however, the most extended are k-fold, leave-one-out, and leave-one-subject-out. Figure 2.9 shows a basic cross-validation strategy for illustration purposes.

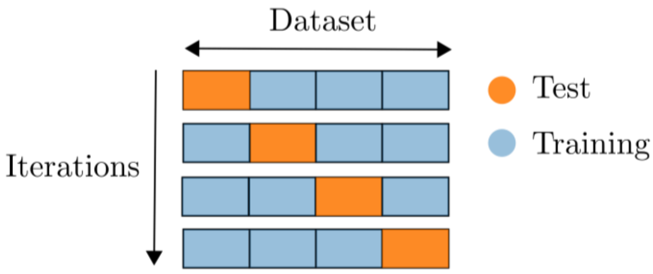


Figure 2.9: Cross-validation example following a 4-fold strategy.

2.2.5. Algorithms

New machine learning algorithms are developed and enhanced ceaselessly, nonetheless, there are some of them whose popularity has been kept at high levels for the last decade[55]. Below, we briefly describe these algorithms without going into detail in terms of their mathematical underlying principles:

- Logistic regression (LR). In spite of its name, which was appointed for historical reasons, this algorithm is designed for classification problems. LR combines the classical polynomial regression with the logistic function, which estimates the probability of an event taking place, to create a decision boundary. This algorithm has been used for applications like social media analysis[56], material degradation assessment[57], or EEG-based seizure detection[58], among others.

- Multi-layer perceptron (MLP). This algorithm is structured as a series of layers: the input layer is designed to enter the features of the dataset; the hidden layers perform non-linear operations to the features; finally, the output layer generates the prediction. Each layer is composed of a number of *perceptrons*, the minimal processing unit of the algorithm, which receives a sum of inputs and performs an operation based on the so-called activation function (see Figure 2.10). Multi-layer perceptron is often referred to as neural network, because the algorithm is inspired by the neural connections in the brain. In this algorithm, the complexity of the model is mainly represented by the number of layers and the number of neurons in each layer. Neural networks are a powerful alternative for classification and regression problems. They have been applied with diverse purposes, such as dementia progression forecasting[59], classification of chronic mental stress[60], or the prediction of ground-reaction forces for athletic performance assessment[61].

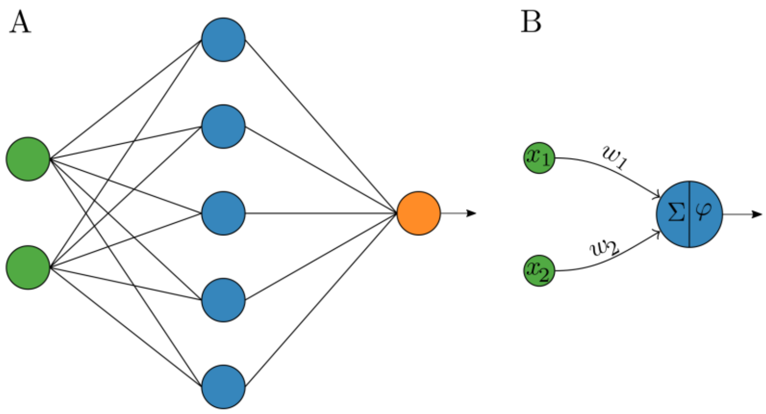


Figure 2.10: Multi-layer perceptron algorithm. (A) Network structure with green, blue, and orange representing the input, hidden, and output layers, respectively. (B) Individual perceptron. Based on source file from wikimedia.org.

- Support vector machine (SVM). This algorithm defines a decision threshold, known as hyperplane, to separate the different classes in a classification problem. For this purpose, the dataset features are projected into a new space, and the observations are compared in that space using a similarity function or kernel. The most extended kernels are the radial basis function and the linear kernel. The use of kernels allows the SVM to generate complex non-linear decision thresholds. SVMs have been utilized for the detection of stress[62,62] and AD[63–65], among other applications[66,67].

- Decision tree (DT). This algorithm constitutes the building-block of the so-called tree-based methods, and can be applied for classification and regression. DT is based on a series of logical queries of the *if* type applied to the dataset features. The most general queries are performed at the top of the tree, and query specificity increases as the tree expands downwards to the lowest tree nodes (known as *leaves*). The goal of the algorithm is to find the set of queries maximizing the classification performance.

- Random forest (RF). RF belongs to the family of *ensemble* methods, which combine different machine learning algorithms to create powerful models. Particularly, RF combines a series of DTs in the following manner: first, each tree is assigned with a number of dataset features (the same feature can be assigned to different trees), thus, the higher the number of features assigned, the more similar the trees will be; then, every tree generates a prediction; finally, the RF outputs the average of the predictions from the ensemble (via majority vote). RF is a very powerful algorithm for classification and regression. This algorithm has been applied to the recognition of driving postures[68], diabetes detection[69], and dementia diagnosis[70], amidst other applications.

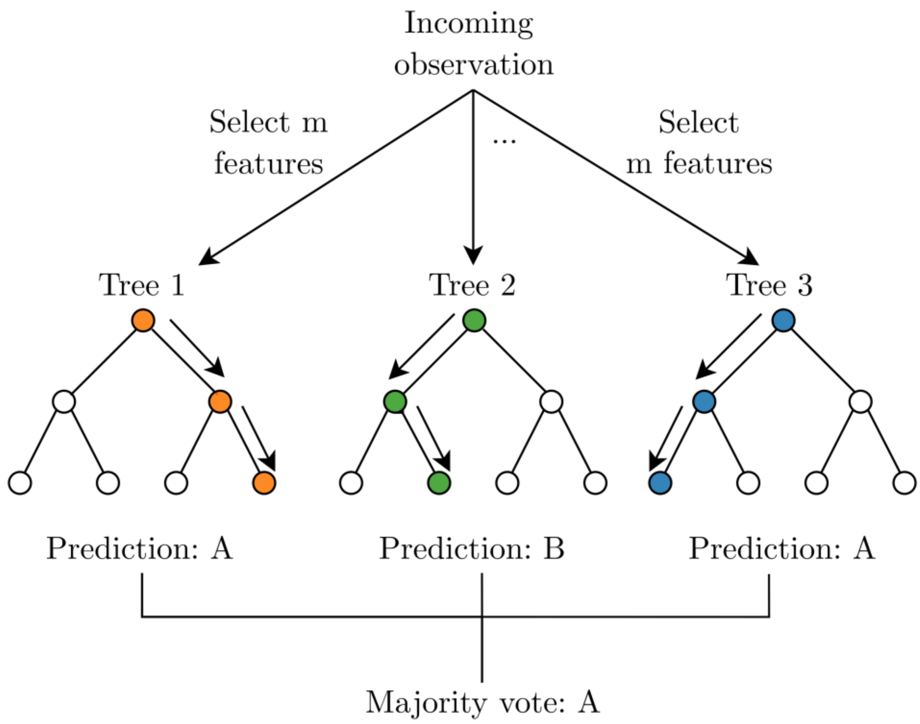


Figure 2.11: Random forest algorithm with three decision trees.

- k-means clustering. k-means arguably represents the most extended algorithm for unsupervised classification. The algorithm is based on an iterative procedure: first, the user defines k, the number of clusters to seek for; then, the algorithm selects k observations in the dataset as centroids, and the rest of the observations are assigned to the nearest centroid, consequently forming k clusters; subsequently, the positions of the centroids are shifted to the average position of their cluster and the whole process is repeated until the position of the centroids remains unchanged (see Figure 2.12).

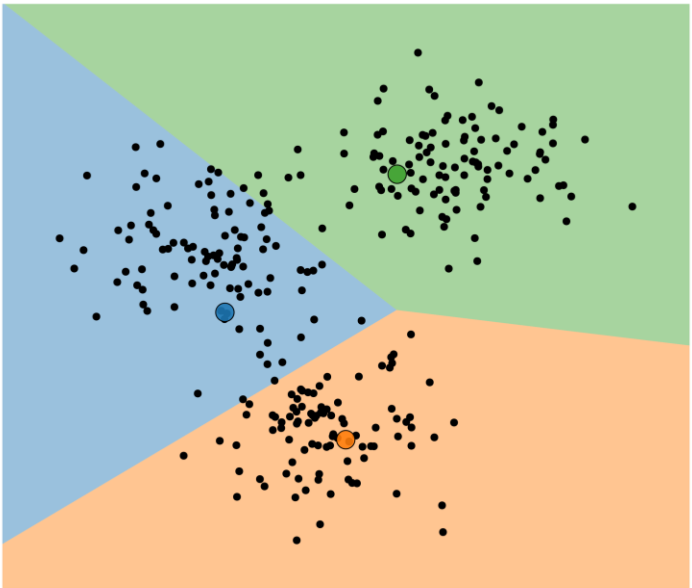


Figure 2.12: k-means clustering algorithm. Black dots show the dataset observations, color dots the centroids, and color regions the decision regions, respectively.

Chapter 3: Previous works in the literature

Although this thesis covers three main applications, namely: detection of attention, stress, and early AD, we reviewed the trends in the literature for the latter because this application had not been investigated in our laboratory before. Consequently, in this chapter we provide a brief survey of the main trends in the application of EEG and machine learning to detect AD. For this purpose, we have been inspired by the review article that we published in the International Journal of Alzheimer's disease in 2021, which is included in Appendix A.

3.1. Basic knowledge on AD

Dementia is an umbrella expression used to describe a series of symptoms related to cognitive impairment, such as loss of orientation, language deterioration, and memory losses. According to the world Alzheimer's report, more than fifty million people live with dementia today, and its prevalence is expected to triple in the next thirty years[71]. In this context, AD, represents more than sixty percent of the dementia cases. AD is a neurodegenerative disease affecting cognitive areas like reasoning, language, orientation, and memory, to the point of disrupting daily life activities and eventually leading to death. Typically, AD is preceded by mild cognitive impairment (MCI), a period characterized by the disruption of certain cognitive areas like memory that does not interfere with daily-life activities and thus, cannot be cataloged as dementia. However, MCI patients transition to AD at a faster rate compared to healthy elderly adults.

Yet little is still known about its etiology, AD is considered a neocortical disconnection syndrome where the disruption of structural and functional integrity of the communication paths between different cortical areas produces serious cognitive losses. In this regard, two main hallmarks have been identified to start forming even before cognitive impairment is notable: amyloid plaques and neurofibrillary tangles. The former are protein deposits accumulating around the neurons after losing their standard structure, and the latter are thickened fibrils encircling the neuron nucleus.

3.2. Early AD detection techniques

Although currently there is no cure for AD, early detection can contribute to controlling the progression of the disease and defer intellectual decline by advancing medical treatment. In this context, multiple techniques are applied in medical practice to detect the disease. These techniques present different levels of accessibility, cost, invasiveness, and accuracy (see Table 3.1). Subsequently, we briefly overview the most extended.

- Neuropsychological tests. They include a series of assessments to evaluate the cognitive areas affected by the disease. The most extended are the mini mental state examination[72] (MMSE) and the Montreal cognitive assessment[73] (MoCA). Even though these tests are simple and quick, they lack detection accuracy, and their score may be disturbed by other pathologies. Nonetheless, they are an essential tool for the detection of cognitive impairment, which is critical for AD diagnosis.

- Biomarkers. These are biological indicators extracted from the cerebrospinal fluid (CSF) and linked to amyloid plaques and neurofibrillary tangles, which are crucial structures related to the disease. Cerebrospinal fluid is located in the spinal cord and the brain, hence, lumbar puncture is required to perform biomarker analyses. However, this procedure is painful and carries known side effects, what is undesirable from the standpoint of the patients[74].

- Medical imaging. This group of techniques is based on the generation of static or functional images of the brain. Some of them achieve this through the reflection of radio pulses, like magnetic resonance imaging (MRI), and others via a radioactive tracer, which is swallowed or inhaled in the case of positron emission tomography (PET) and injected into the bloodstream in the case of single-photon emission computed tomography (SPECT).

- Electrophysiology. The two main techniques in this group are EEG and magnetoencephalography (MEG). Whilst EEG measures the electrical activity in the brain cortex, MEG measures magnetic fields produced by small currents over postsynaptic potentials. Importantly, both techniques provide sub-millisecond temporal resolution.

Table 3.1: Comparison of AD detection techniques.[75]

Technique	Invasiveness	Cost	Accessibility
CSF	Yes	High	Low
MRI	No	High	Low
PET	Yes	High	Low
SPECT	Yes	High	Low
MEG	No	High	Low
EEG	No	Low	High

Presently, the typical procedure for the diagnosis of AD encompasses the performance of neuropsychological tests for cognitive impairment detection, and then incorporates medical trials such as CSF biomarkers or PET analyses for the detection of structures linked to amyloid plaques and neurofibrillary tangles.

3.3. EEG processing for early AD detection

Considering the features listed in the table above, EEG represents a promising alternative for the detection of early AD owing to its reduced cost, high accessibility, and non-invasiveness compared to other techniques. Plus, the spread of EEG has been boosted by the commercialization of affordable wearable devices[76]. With this, the goal of EEG-based early AD detection is to unfold the effects caused by the disease in the electrical activity of the brain. The main effects reported in the literature include:

- Rhythms slowing. This effect refers to a transition of the EEG power from higher frequencies such as alpha and beta to lower frequencies like delta and theta in AD patients compared to healthy elderly adults. This effect was the first to be observed using classical visual inspection techniques.
- Loss of synchronization. Typically, coherence analysis is used to evaluate functional connectivity amid cortical areas. A coherence reduction between close and distant channels in alpha and beta is another established feature associated with AD. This reduction suggests a disconnection pattern among cortical areas and has been correlated with cognitive impairment.[77]
- Complexity reduction. AD produces a reduction in the complexity of the brain electrical activity, which results in more regular and predictable EEG signals for the patients compared to healthy elderly adults. This effect has been arguably attributed to neural death and to the disconnection between brain areas.

- Declining ERPs. As described in subsection 2.1.4, ERPs are markers of sensory and complex cognitive brain processes. Since AD disrupts cognitive functions, these potentials are likewise affected. Particularly, the analysis of the morphological differences in ERPs has evidenced a decreased amplitude and prolonged latency in AD and MCI patients compared to healthy elderly adults.

Taking this into account, works on the detection of AD via EEG and machine learning typically combine the extraction of some of the aforementioned features and the application of classification algorithms to discern between two or three disease groups. The most frequently reported classification problems in the literature are AD vs healthy controls (HC) and MCI vs HC. In these works, EEG is usually recorded in resting state, nonetheless, other tasks have been studied in order to reveal the differences between the disease groups. This includes oddball recognition, language, audio, and writing tasks. Table 3.2 and Table 3.3 show a summary of works from the last decade regarding the discrimination of traditional (HC, MCI, and AD) and alternative disease groups based on EEG activity and machine learning.

Table 3.2: Works from 2010-2020 on the two-class discrimination of AD based on EEG and machine learning. Acronyms in the table: A: accuracy, R: recall, S: specificity, HC: healthy control, LDA: linear discriminant analysis, WT: wavelet transform.

Study	Channels	Cohorts	Protocol	Features (feature selection)	Model	Validation	Performance	
							HC-MCI	HC-AD
Trambaiolli et al. [78] (2011)	19	HC: 19 AD: 16	Resting State	Coherence and spectral peaks (Weka tool)	SVM	Non-cross validation	-	A = 79.9 R = 83.2 S = 76.4
McBride et al. [79] (2013)	32	Not reported	Resting State	EEG reconstruction scores (LOO CV)	SVM	LOO	A = 90.3 R = 87.5 S = 93.3	A = 90.6 R = 88.2 S = 93.3
Aghajani et al. [80] (2013)	128	HC: 17 AD: 17	Resting State	PSD (Singular Value Decomposition)	SVM	LOO	-	A = 84.4 R = 75 S = 93.7
Fiscon et al. [81] (2014)	19	HC: 14 MCI: 37 AD: 49	Resting State	FFT	Decision tree	LOO	A = 90 R = 90 S = 87	A = 87.76 R = 88.57 S = 87.22
McBride et al. [21] (2014)	32	HC: 15 MCI: 16 AD: 17	Resting State	Spectral and complexity parameters (LOO CV)	SVM	LOO	A = 96.8 R = 93.8 S = 100	A = 96.9 R = 100 S = 93.3
Wang et al. [82] (2015)	16	HC: 14 AD: 14	Resting State	PSD and coherence (PCA)	Cluster analysis	None	-	A = 91.4 R = 100 S = 82.9
Morabito et al. [83] (2016)	19	HC: 23 MCI: 56 AD: 63	Resting State	Skewness, mean and std in three frequency bands	CNN	k-fold	A = 85 R = 84 S = 81	A = 85 R = 85 S = 82
Cassani et al. [84] (2017)	7	HC: 24 AD: 35	Resting State	Spectral power, coherence, amplitude modulation (L1 norm)	SVM	LOSO	-	A = 69.5 R = 71.4 S = 66.6
Trambaiolli et al. [85] (2017)	19	HC: 12 AD: 22	Resting State	Spectral peaks in main EEG spectral bands (Filtered Subset Evaluator)	SVM	LOSO	-	A = 91.18 R = 90.91 S = 91.67
Kulkarni et al. [65] (2017)	16	HC: 50 AD: 50	Resting State	Relative Power, WT and complexity	SVM	LOSO	-	A = 96
Ruiz-Gómez et al. [86] (2018)	19	HC: 37 MCI: 37 AD: 37	Resting State	Spectral and nonlinear features (Fast Correlation Based Filter)	MLP, QDA, LDA	LOSO	A = 78.43 R = 82.35 S = 70.59	A = 98.05 R = 97.99 S = 98.21
Fiscon et al. [87] (2018)	19	HC: 23 MCI: 37 AD: 49	Resting State	Fourier Transform and Wavelet Transform	Decision Trees	LOO	A = 91.7 R = 91.7 S = 91.5	A = 83.3 R = 83.3 S = 78.0
Mazaheri et al. [88] (2018)	19-32	HC: 11 MCI: 25	Language task	Oscillatory changes	SVM	LOO	R = 0.80 S = 0.95	-
Khatun et al. [89] (2019)	1	HC: 15 MCI: 8	Audio ERP	Temporal and amplitude of the ERP (Random Forest)	SVM	LOO	A = 87.9 R = 84.8 S = 95	-
Durongbhan et al. [90] (2019)	1	HC: 20 AD: 20	Resting State	FT and WT	KNN	k-fold	-	A = 99.2
Ieracitano et al. [91] (2020)	10	HC: 63 MCI: 63 AD: 63	Resting State	Continuous WT and BiSpectrum	MLP	k-fold	A = 96.24 R = 95.86	A = 96.95 R = 94.91

Table 3.3: Works from 2010-2020 on the discrimination of alternative AD groups based on EEG and machine learning. Acronyms in the table: A: accuracy, R: recall, S: specificity, sMCI: stable MCI, pMCI: progressive MCI, GA: gaussian analysis, PCA: principal component analysis.

Study	Channels	Cohorts	Protocol	Features (feature selection)	Model	Validation	Performance
Chapman et al. [35] (2011)	1	sMCI: 28 pMCI: 15	Number letter test	ERP components (PCA)	Discriminant analysis	LOO	sMCI - pMCI A = 79 R = 80 S = 79
Podgorelec et al. [92] (2012)	2	No AD: 12% AD: 88%	Not Reported	Time and frequency domain parameters (tree-based method)	Evolutionary algorithm	k-fold	No AD - AD A = 86.05 R = 88.89 S = 71.43
Poil et al. [93] (2013)	21	sMCI: 39 pMCI: 25	Resting State	Neurophysiological Biomarker Toolbox (Student's t test and GA)	Logistic regression and GA	Half split	sMCI - pMCI R = 88 S = 82
Buscema et al. [94] (2015)	19	HC: 99 MCI: 46 AD: 127	Resting State	Spatial invariants via MS-ROM (TWIST algorithm)	Multiple classifiers	TWIST algorithm	HC - MCI - AD A = 98.27 R = 98.29 S = 98.21
Houmani et al. [95] (2018)	30	SCI: 22 pAD: 49	Resting State	Epoch based entropy and bump models (Orthogonal Forward Regression)	SVM	LOSO	SCI - pAD A = 91.6 R = 87.8 S = 100
Amezquita et al. [96] (2019)	19	MCI: 37 AD: 37	Resting State	Fractal dimension and Hurst exponent (one-way ANOVA)	EPNN	Random subject selection	MCI - AD A = 90.3 R = 92.1 S = 87.9

Chapter 4: Attention detection

In this chapter, we present the study on attention detection included in the article compendium for this thesis, which is entitled "An attention-driven videogame based on steady-state motion visual evoked potentials" and is included in Appendix B. To do so, we briefly describe the main aspects of this work.

4.1. Objective

For the past years, the BCI laboratory has conducted research works in the field of special needs education in collaboration with San Rafael special needs education school[16,97,98]. This study represents another step in this collaborative history. Concretely, in this research our objective was to investigate the combination of EEG activity and the principles of serious games to evaluate an application for the real-time detection and potential training of the attention to mobile visual stimuli. This could positively impact attention therapies conducted in institutions such as special needs education schools, where some children often struggle to perform basic tasks, such as sustaining the attention.

4.2. Methodology

For this research, we proposed an inclusive videogame based on SSVEPs (see subsection 2.1.4). To implement the game, we designed a cloud architecture including an automated closed-loop BCI system to visually stimulate the participants, process the signals online, use the elicited potentials to detect the attention exerted, and modulate the game course (see Figure 4.1). We presented this videogame in the 1st Telecommunication Meeting held in the Superior Technical School of Computer Science at UGR. Subsequently, to evaluate this approach, we conducted a study whose details are reported next.

- **Participants**

We recruited 20 healthy volunteers (15 males and 5 females) among the student community at UGR (mean age 26.2 ± 9.2). They went through a single experimental session that lasted roughly 40 minutes and was conducted in the BCI laboratory. Only participants that did not suffer from any medical conditions that could affect their EEG activity were considered.

- **Framework**

The system architecture designed consisted of four modules: the stimulation client, the EEG acquisition device, the monitoring client, and the control server. We implemented the whole system using MATLAB R2016a (Nattick, USA). Particularly, for the stimulation and the game interface, we used the Psychtoolbox package, a library designed to implement experiments based on stimulation.

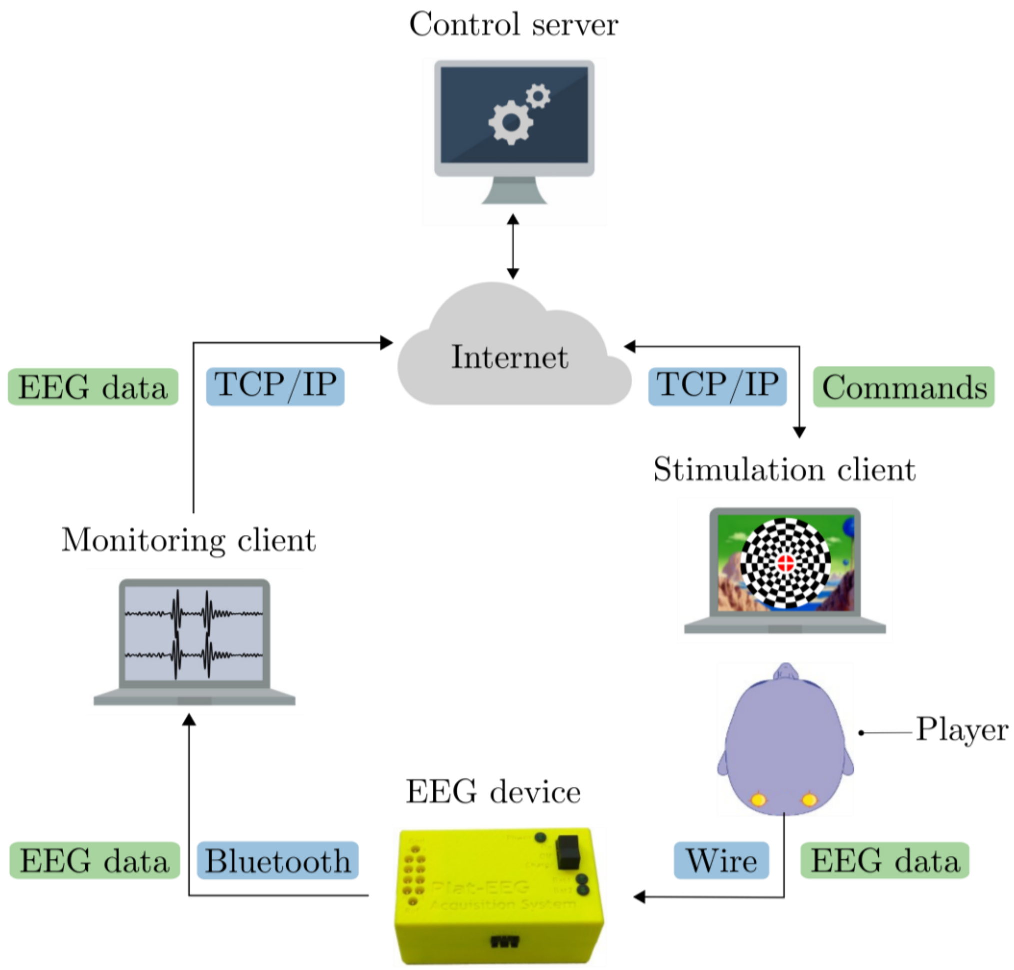


Figure 4.1: SSVEP videogame system architecture. The four modules of the system include: control server, stimulation client, EEG acquisition device, and monitoring client. Blue and green boxes represent the communication channels and the information exchanged. The EEG signals are processed online in the closed-loop.

First, the stimulation client presents the videogame to the player; in parallel, the EEG device records the EEG signals and forwards them to the monitoring client via Bluetooth; this client displays the EEG in real time and sends the raw signals to the control server through a TCP/IP socket; finally, the control server performs EEG signal processing for attention detection and sends game commands to the stimulation client in order to correctly adjust the game course. Thereby, the whole cycle of the videogame is completed in real time.

The game spans five levels. At the start of each level, an enemy avatar appears at a random location of the screen and throws an attack to the player (see Figure 4.2). To stimulate the players, the attack is presented as a ring-shaped checkerboard (see Equation 4.1) with a fixation cross. The attack travels to the center of the screen while increasing its size and reversing its pattern at 15 Hz.

Figure 4.2: SSVEP videogame gameplay. From left to right, each frame shows a further time instant. First, the enemy appears at a random position and throws an attack in the form of a ring-shaped inverting checkerboard. Then, the attack travels towards the center of the screen while growing and inverting its pattern at a 15 Hz rate.

$$\frac{1 + \mathrm{sign}\left(\sin\left(\mathrm{atan}(x, y) \cdot \frac{R}{2}\right)\right) \cdot \mathrm{sign}\left(\sin\left(\sqrt{x^2 + y^2}\right)\right)}{2}$$

Equation 4.1: Formula to generate the ring-shaped checkerboard stimulus. R refers to the number of ring checks; x and y represent the pixel grid coordinates.

The goal of the player is to deflect the attacks by exerting enough attention to the stimulus. If the attention level exerted is not sufficiently high, the attack will keep advancing towards the player and eventually the computer will score a point. Otherwise, the attack will be deflected, and a new enemy avatar will be presented. The first competitor (computer or player) to reach five points wins the level. If the player fails, the level will be repeated; otherwise, the player will advance to the next level. The game is terminated after 15 minutes or after the player beat the five levels.

To increase the difficulty at each level, we set the enemy avatars closer to the player and increased the speed of their attacks. To engage the participants, we based the enemy avatars, music, and scenarios on a renowned TV series.

- **EEG acquisition and processing**

For the EEG acquisition we used the RABio w8 device (see subsection 2.1.2) and a montage including only two channels from the visual cortex (O1 and O2 of the 10-20 International System). For the attention detection procedure performed by the control server, we calculated an attention parameter based on the PSD of the EEG at O1 and O2 (see Equation 4.2). This parameter represented the ratio between the power at the frequency of the SSVEP generated by the enemy attacks (14-16 Hz) and the power at nearby frequencies (12-13 Hz and 17-18 Hz).

$$\mathrm{Attention}_{\mathrm{parameter}} = \frac{\mathrm{P(14-16)Hz}}{\mathrm{P(12-13)Hz} + \mathrm{P(17-18)Hz}}$$

Equation 4.2: Formula for the attention parameter used for attention detection. *P* represents the spectral power over a particular frequency range.

The figure below represents the SSVEP elicited by the videogame stimulation, and justifies the use of the above equation to estimate the SNR in the desired spectral window to detect the attention.

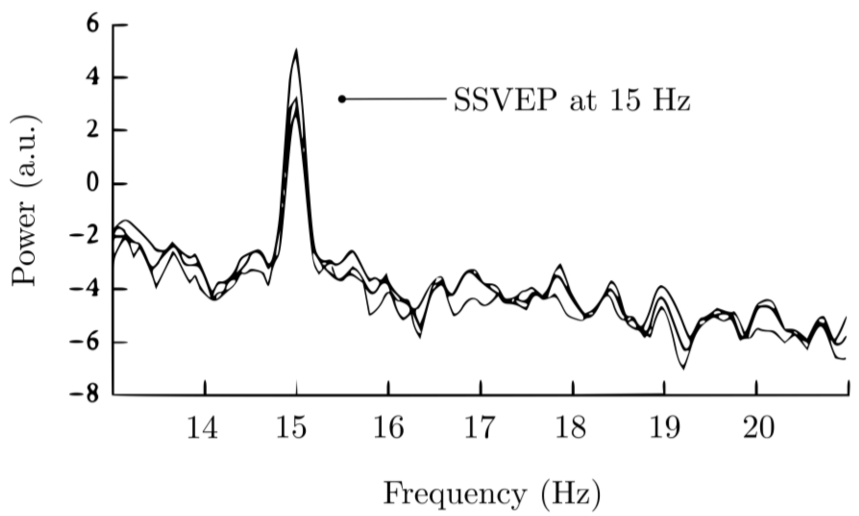


Figure 4.3: Illustration of the SSVEP elicited during gameplay. The pattern reversal of the stimulation checkerboard was inverted at a rate of 15 Hz.

To estimate the attention parameter, the control server filtered the raw two-second EEG epochs using a bandpass Butterworth filter with bandpass 6-75 Hz, then applied a Tuckey window to the filtered signal, and finally calculated the spectral power in the ranges of interest via MATLAB's periodogram. We decided to use a Butterworth filter because the system required strict online processing.

To decide if the estimated attention parameter was high enough to concede the player a point, we compared it with a threshold value estimated during a calibration trial performed before the start of the game. We defined this threshold value as the maximum attention parameter generated by the player during the calibration trial minus three decibels.

- **Analysis**

We assessed participant performance in terms of time played, points scored, points conceded, level success rate, and level point difference. For these metrics, we estimated the grand-average and the standard error of the mean (SEM).

Furthermore, we also conducted surveys between the game levels to gather the self-perceived attention level of the participants and their viewpoint regarding the equity of the scoreboard. After the termination of the game, we conducted a final survey to collect the global opinion of the participants about the whole experience.

We analyzed the participant results in terms of their self-perceived attention, their viewpoint on the scoreboard equity, their gaming experience, and their global game assessment. For this purpose, we applied the Shapiro-Wilk, the Mann-Whitney, and the Kruskal-Wallis hypothesis tests.

4.3. Results

In this subsection, we briefly report the main results from this work. Figure 4.4 shows the evaluation of the game experience reported in the participant surveys.

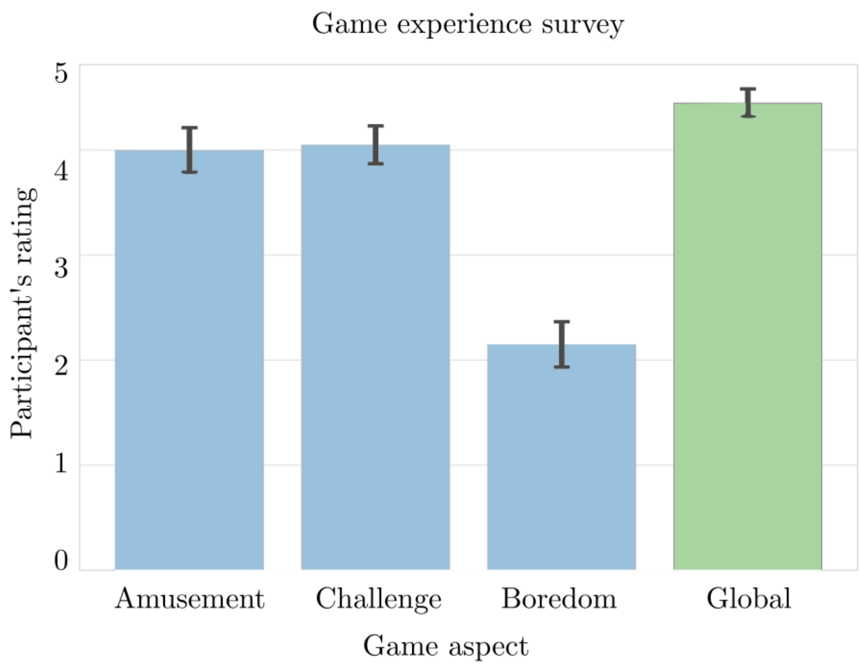


Figure 4.4: Evaluation of the SSVEP game experience over multiple aspects.[99]

Figure 4.5 represents the average total time played and the average number of points scored and conceded by the participants in every level of the game.

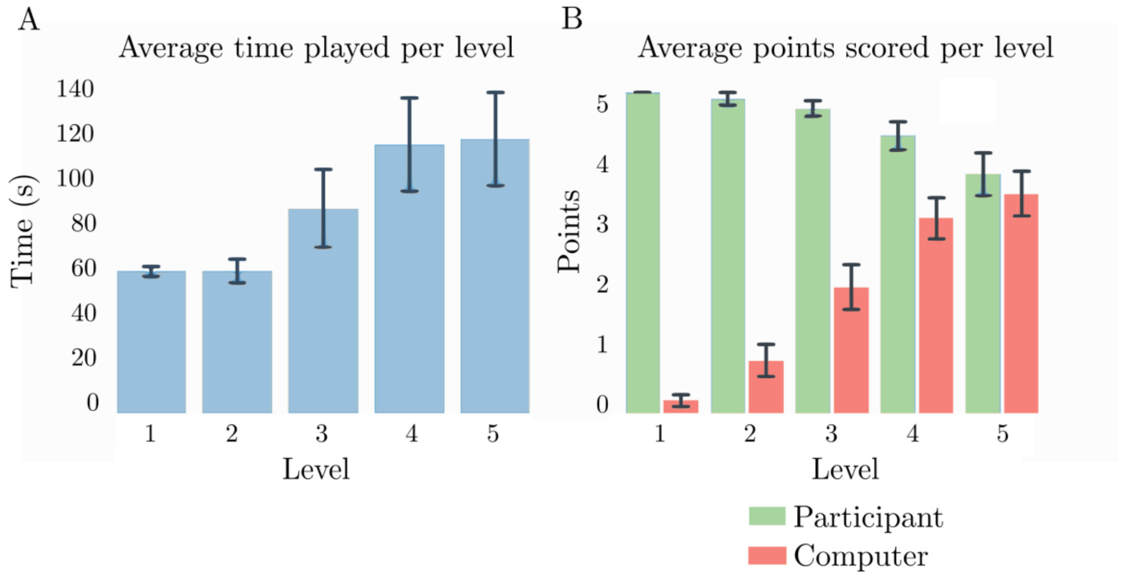


Figure 4.5: Difficulty progression by level for the SSVEP videogame.[99] (A) Total time spent per level and (B) Mean number of points scored and conceded per level. Error bars indicate the SEM.

According to the surveys conducted during the study, the players described the videogame as entertaining and challenging, what is promising towards its application in centers for the treatment of attentional disorders in children. Consequently, the main contribution of this study was the successful implementation of a SSVEP-based videogame for demanding attentional efforts to the players. The obtained results demonstrate:

1. The feasibility of a BCI system based on a cloud architecture to process the EEG activity in real time and modulate the game course according to the brain activity features extracted for the automated detection of visual attention to mobile stimuli.

2. The potential of a ring-shaped mobile flickering stimulus to design gamified experiences as an alternative to traditional SSVEP stimulation approaches.

3. The appropriateness of the combination of game design (avatars, music, scenarios, and level progression) and stimulation design (moving ring-shaped checkerboard simulating an attack) to render an amusing and challenging experience.

Chapter 5: Stress detection

In this chapter, we present the two works focused on stress detection included in the article compendium for this thesis. For additional details about these two works, refer to Appendices C and D, which include the original manuscripts.

5.1 Publication in Appendix C

In this section, we summarize the work "Quantitative Assessment of Stress Through EEG During a Virtual Reality Stress-Relax Session" annexed in Appendix C of the article compendium.

5.1.1. Objective

In this research, our goal was to explore machine learning models for the regression of the self-perceived stress level (SPSL) from EEG activity as an alternative to the classifiers available in the literature, which are limited to two (stress and no-stress) or three (low, medium, high) levels. The results could potentially impact certain research areas leveraging from quantitative stress prediction models, such as neuromarketing, stress relief therapies or industrial workers training. For this work, we analyzed the dataset that we acquired for a previous investigation on the simulation of 360-degree experiences through virtual reality[101] (see subsection 7).

5.1.2. Methodology

For this research, we proposed a stress-relax session based on a test to elicit psychosocial stress and a VR relax experience. Throughout the session, we collected the SPSL of the participants through surveys and we recorded their EEG activity during the whole session. To perform the SPSL regression, we evaluated different regression algorithms.

- **Participants**

We recruited 23 volunteers (14 females, 8 males, and a participant who identified as non-binary) from the student community at UGR (mean age 22.7 ± 5.5). They went through a single experimental session with an approximate 18-minute span which was conducted in the facilities of the BCI laboratory. We only involved participants with no health issues that could influence the experiment outcome.

- **Framework**

We conducted the participants through a stress-relax session while recording their EEG activity. First, for the stress phase, the participants performed the Montreal imaging stress test (MIST), a validated arithmetical test to elicit psychosocial stress[100]. This phase was divided into two sub-phases: a training phase to introduce the participants to the assessment and a test phase to elicit stress.

Subsequently, for the relax phase, the participants used a head-mounted display (HMD) to watch a VR 360-degree relaxing video. Throughout the session, the participants completed eight surveys based on the perceived stress scale to report their SPSL (in the range of 1 to 5).

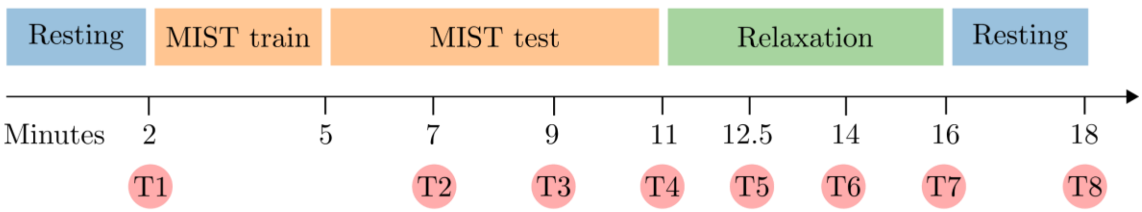


Figure 5.1: Stages of the stress-relax session. First, the participants were comfortably sitting in resting state; then, they performed the training and test phases of the MIST for stress induction; subsequently, they relaxed through the VR 360-degree video experience; finally, they performed a final resting state. The SPSL surveys conducted during the experiment (T1 to T8) are shown inside red circles.

We implemented the MIST through a graphical user interface using MATLAB R2016a (Nattick, USA). The 360-degree video experiences were reproduced in the VR system using Unity engine (San Francisco, USA) and were presented using an Oculus Quest HMD (California, USA)[101]. Surveys T2 to T4 and T5 to T7 were integrated in the MIST interface and the VR experiences, respectively. The remaining surveys were conducted verbally.

- **EEG acquisition and processing**

For the EEG acquisition we used the Versatile 16 system (see subsection 2.1.2) with a montage including channels Fp1, Fp2, F5, and F6 of the extended 10-20 International System. We processed the EEG of the participants individually for each stage of the experiment (see Figure 5.2). First, we applied a notch and a bandpass filter, then we resampled the EEG to match the exact span of the phase being processed. For this, we evaluated different interpolation techniques. Subsequently, we segmented the signals into two-second epochs without overlapping and we rejected those epochs above a predefined threshold of 100 µV. Finally, we extracted seven features from the EEG signals: the channel-averaged PSD in the five main EEG bands, the relative gamma, and the alpha asymmetry, which we smoothed to cope with the EEG time variability.

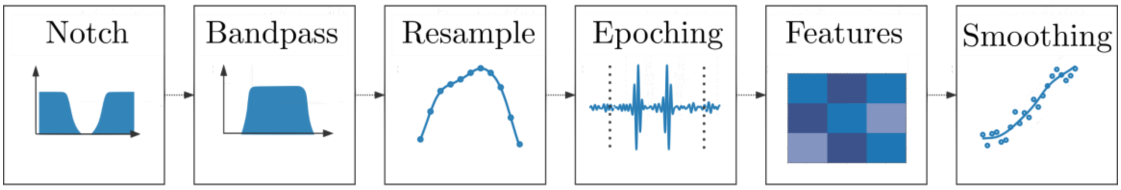


Figure 5.2: EEG processing pipeline for stress prediction. The stages in the pipeline include filtering, resampling, epoching, feature extraction, and feature smoothing.

The processed recordings comprised 12 min: the MIST assessment (6 min), the relaxation phase (5 min), and the central minute of the last resting state (1 min).

- **Analysis**

We used the scikit-learn[102] Python library (version 1.0.2) to implement a separate regression model per participant. For the feature matrix, we used the seven EEG features reported in the previous subsection, whilst for the target array, we interpolated the SPSL surveys. We evaluated multiple regression algorithms: linear regression, RF, MLP, and support vector regressor. We split the participant data into training and test, and we used the training data to find the best hyperparameters for each model via five-fold grid-search cross-validation, and the test set to evaluate the final model. For evaluation, we estimated the mean squared percentage error (MSPE, see Equation 5.1) and the Pearson correlation coefficient.

$$\mathrm{MSPE} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(y_{true}i - y_{pred}i\right)^2}{\frac{1}{n} \cdot \sum_{i=1}^{n} y_{true}i} \cdot 100$$

Equation 5.1: Formula for the mean squared percentage error.

5.1.3. Results

In this subsection, we report the main results obtained in this work. Particularly, Figure 5.3 displays the SPSL regression predictions yielded for all the participants.

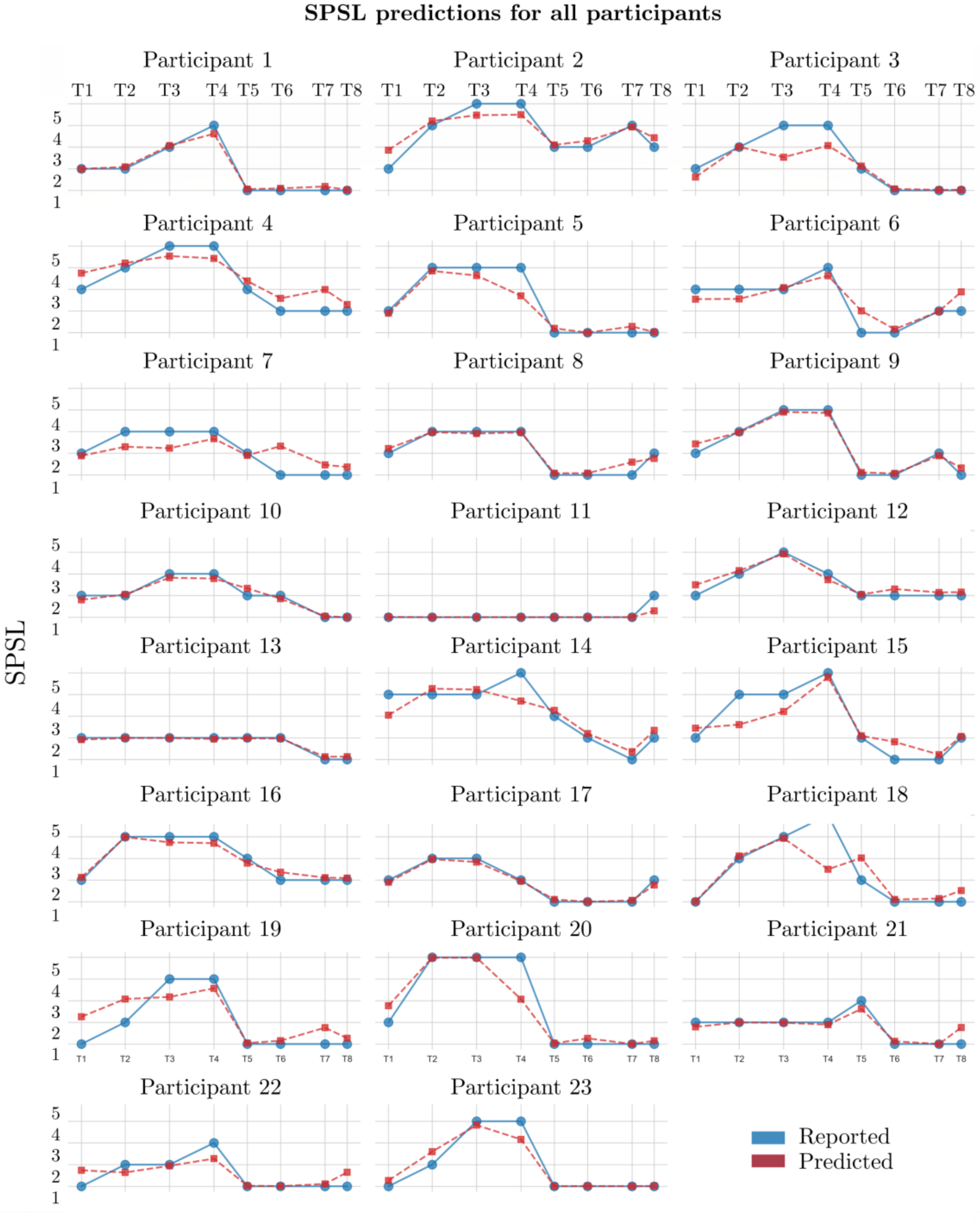


Figure 5.3: True SPSL and predictions for all the participants. *Y* axis displays SPSL, and *X* axis represents the eight surveys. Blue round markers show the SPSL reported by the participant, whilst red square markers refer to the SPSL predictions[103].

Alternatively, Table 5.1 shows the average performance of the different regressors evaluated in this study in terms of the MSPE and the Pearson correlation coefficient for all the interpolation techniques considered.

Table 5.1: SPSL regression performance for the evaluated algorithms. For readability, the table has been split into two sub tables (A) for linear and pchip interpolations, and (B) for spline and nearest interpolations.

(A)	**Linear**		**Pchip**	
Regressor	MSPE	R^2	MSPE	R^2
RR	24.96 ± 3.61	0.76 ± 0.03	24.14 ± 3.35	0.76 ± 0.03
RF	10.62 ± 2.12	0.92 ± 0.02	11.11 ± 2.37	0.91 ± 0.02
MLP	23.09 ± 3.63	0.77 ± 0.03	22.20 ± 3.37	0.78 ± 0.02
SVR	18.97 ± 3.50	0.80 ± 0.04	18.67 ± 3.43	0.80 ± 0.04

(B)	**Spline**		**Nearest**	
Regressor	MSPE	R^2	MSPE	R^2
RR	24.56 ± 3.41	0.76 ± 0.03	23.65 ± 3.26	0.77 ± 0.03
RF	11.24 ± 2.53	0.91 ± 0.02	11.97 ± 2.91	0.90 ± 0.02
MLP	20.14 ± 2.91	0.80 ± 0.02	23.56 ± 3.44	0.73 ± 0.04
SVR	17.84 ± 3.66	0.81 ± 0.04	21.46 ± 3.55	0.80 ± 0.03

According to the results, the RF regressor with linearly interpolated data, which yielded the best results, accurately predicted the SPSL of the participants, as evidenced by averages of 10.62 and 0.92, for the MSPE and the Pearson correlation coefficient, respectively. In light of the results obtained, the main contributions of the work presented in this section are the following:

- Most studies in the literature approach EEG-based stress assessment via classification algorithms, however, this procedure is limited to the detection of two or three stress levels. The principal contribution of this work is the demonstrated feasibility of an intra-participant regressor to accurately predict the quantitative SPSL from EEG activity.

- Furthermore, the classification performance metrics usually reported in the stress literature are very heterogeneous, what hinders their comparison. Alternatively, we proposed the MSPE to quantify the prediction errors and complementing typical regression analysis such as the Pearson correlation.

- Finally, since these findings rely on an 18-minute session and an EEG montage with few electrodes, our approach could potentially contribute to evaluating continuous stress levels in applications such as neuromarketing or industrial workers training, where stress often plays an important role.

5.2 Publication in Appendix D

In this section, we summarize the work "EEG-based multi-level stress classification with and without smoothing filter" annexed in Appendix D.

5.2.1. Objective

The objective of this research was twofold: (a) to evaluate the effects of smoothing window length on the performance of a three-level SPSL classifier, and (b) to examine the feasibility of a two-level and a three-level SPSL classifier from individually processed epochs. For this purpose, we took advantage of the dataset that we acquired for a previous investigation on the simulation of chromotherapy through virtual reality[98] (see subsection 7).

5.2.2. Methodology

In essence, the experimental design for this study was analogous to the study reported in the previous section, with the exception of the relax phase, which in this case was performed inside a chromotherapy room for half of the participants, and through a VR application, for the other half. This study was conducted in collaboration with the San Rafael school for special education located in Granada (Spain), which provided counselling and access to a chromotherapy room.

- **Participants**

We recruited 20 healthy volunteers (14 males and 6 females) among the student community at UGR (mean age 24.2 $\pm$ 4.0). They underwent a single experimental session that lasted roughly 18 minutes.

- **Framework**

The experimental procedure for this study was identical to the one described for the previous study: an initial resting state, the MIST training and test phases, a relax phase, and a final resting state (see Figure 5.1). There are only two differences between the two experimental procedures:

1. Regarding the relax phase, whilst in the study included in Appendix C the relax session was conducted via VR 360-degree video experiences, in this study, this phase was completed using either a real chromotherapy room or a VR simulated one. This relax phase design stems from the objectives of our previous study about chromotherapy[98].

2. Regarding the SPSL surveys, instead of performing eight surveys throughout the experimental session, in this case we only performed four: before the initial resting state, after the MIST, in the middle of the relax session, and before the final resting state.

The MIST and the VR application for the relax phase were implemented following the procedures described for the article presented in subsection 5.1.2.

- **EEG acquisition and processing**

For the EEG acquisition we used the RABio w8 system with a montage including channels Fp1, Fp2, F7, and F8 of the 10-20 International System. To process the recordings, we applied a 2^{nd} order Butterworth bandpass filter (1-50) Hz and a notch filter (50 Hz). Then, we segmented the signals into two-second epochs without overlapping and we discarded those segments above a threshold of 100 µV established upon visual inspection. For the feature extraction, we estimated the PSD in the five main EEG bands using the periodogram. In addition, we estimated the alpha asymmetry and the relative gamma. To calculate the alpha asymmetry, we used the pre-frontal channels, whilst for the rest of the features we applied a channel average. Lastly, we stacked the features from all the participants to create a single data structure and applied a Savitzky-Golay filter with different window lengths to assess the effects of smoothing on classification.

- **Analysis**

We proposed a three-class classification problem to assess the effects of the smoothing filter window length. For the feature matrix, we used the combined smoothed features described in the previous paragraph, whilst for the target array we merged the SPSL surveys into three levels: low stress (SPSL equal to 1 and 2), medium stress (SPSL equal to 3 and 4), and high stress (SPSL equal to 5).

We evaluated the following classifiers: LR, SVM, RF, MLP, and k-nearest neighbors. To tune the hyperparameters we followed the approach described in subsection 5.1.2 (grid-search five-fold cross-validation). To evaluate the classification performance, we estimated accuracy, intra-class precision, recall, and F1-score. Lastly, we estimated the receiver-operating characteristic (ROC) curves to evaluate the performance of the best classifier obtained for each smoothing filter window length. All the machine learning procedures were implemented using the scikit-learn library (version 1.0.2) for Python.

5.2.3. Results

The table below shows the classification performance in terms of precision, recall, F1 score, and accuracy as a function of the smoothing window length for all the classifiers evaluated in the study.

Table 5.2: Classification results for three-SPSL levels per smoothing window length. Classes 1, 2, and 3 refer to low, medium, and high SPSL. clf stands for classifier.

Window length	Best clf	Class 1 Precision	Class 1 Recall	Class 2 Precision	Class 2 Recall	Class 3 Precision	Class 3 Recall	F1 score	Accuracy
Three-level stress classification									
0	RF	0.68	0.70	0.58	0.46	0.29	0.67	0.61	0.61
11	SVM	0.71	0.81	0.72	0.54	0.25	0.33	0.68	0.69
21	MLP	0.74	0.76	0.57	0.54	0.67	0.67	0.67	0.67
31	SVM	0.84	0.86	0.77	0.71	0.75	1	0.81	0.81
41	SVM	0.89	0.86	0.78	0.75	0.6	1	0.83	0.83
51	SVM	0.92	0.95	0.91	0.83	0.75	1	0.91	0.91
61	SVM	0.95	0.95	0.92	0.92	1	1	0.94	0.94
Stress detection (Two-level stress classification)									
0	MLP	0.92	0.85	0.38	0.56	-	-	0.83	0.81

As evidenced in the results, the main conclusions from this work are the following:

- Smoothing window length increases performance in three-class inter-participant SPSL classifiers fed with spectral EEG features at the expense train-test data contamination.
- A feasible two-class inter-participant SPSL classifier can be implemented without applying smoothing. This may contribute to opening the door for real-time applications relying on the prediction of the perceived stress level from EEG (e.g., neuromarketing and stress relief therapies).
- Lastly, we provided a set of recommendations to improve the comparability and discussion among EEG-based stress classification studies. These include reporting intra-class performance metrics as an alternative to global accuracy, processing the epochs of each participant individually to avoid train-test contamination, and considering support features such as galvanic skin response and heart rate variability.

Chapter 6: Early AD detection

In this chapter, we review the two works on the detection of early AD which are included in the article compendium for this thesis (see Appendices E and F). Additionally, we report two further works which are not yet published at the time of writing this thesis: an article on early AD detection using a reduced montage and a study on cognitive impairment detection via a visual dynamics assessment.

6.1 Publications in Appendices E and F

This section resumes the articles “An Automated Approach for the Detection of Alzheimer’s Disease from Resting State Electroencephalography” and “A self-driven approach for multi-class discrimination in Alzheimer’s disease based on wearable EEG”, included in Appendices E and F, respectively. We have grouped these two works because they were derived from the same data capture, and share an important part of the methodology.

6.1.1 Objective

These works originated from the collaboration between the BCI laboratory and the Cognitive and Behavioral Neurology Unit (CBNU) at Hospital Universitario Virgen de las Nieves (HUVN) de Granada (Spain). The objective of both works was to investigate an automated approach for the real-time detection of early AD from EEG activity. Such an approach could contribute to reducing medical waiting lists and costs linked to traditional diagnosis techniques (see Figure 6.1).

6.1.2 Methodology

The experimental session was conducted in the facilities of the HUVN, and comprised a personal interview, a neuropsychological evaluation, and an EEG acquisition under eyes-open resting state conditions.

- **Participants**

The neurologists at the CBNU recruited 26 participants for this data capture (15 females and 11 males). Nonetheless, due to poor signal quality, lack of cooperation, and presence of other disease, we did not analyze the data from five participants. The participants belonged in three groups (MCI, AD, and HC). MCI and AD participants were diagnosed through PET-Amyloid scans and CSF analysis. These techniques were applied to test for the presence of amyloid plaques and total-τ, Aβ42, and τ-phosphorylated fraction, respectively. For the diagnosis, all the participants were also evaluated using the BAGEN[104], a neuropsychological assessment adapted to the local population.

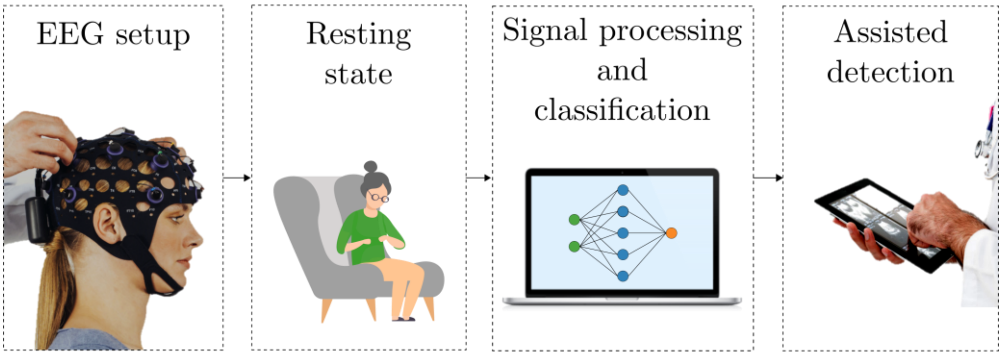


Figure 6.1: Envisioned scenario for the automated early AD detection. The EEG setup involves a wearable wireless EEG acquisition system and is measured during a resting state protocol. The EEG signals are processed in a server and a classification result is sent to the clinician to assist the diagnosis. Based on homeinstead.com.

- **Framework**

For the data capture, we conducted the participants throughout a single experimental session. First, they completed the BAGEN with the assistance of a professional neuropsychologist; then, they performed an EEG acquisition under resting state with eyes open to avoid drowsiness. The session lasted approximately 30 minutes.

- **EEG acquisition and processing**

We performed the EEG acquisition using the Versatile 16 system with a montage including channels Fp1, Fp2, F5, Fz, F6, T7, T8, C3, Cz, C4, P5, Pz, P6, O1, Oz, and O2 of the extended 10-20 International System. The ground electrode was fixed at the AFz position, whilst we set the reference electrode at the left ear lobe. Regarding signal processing, we followed the pipeline shown in Figure 6.2 and described subsequently.

First, we applied a FIR bandpass filter to retain the spectral content in the range 1 – 45 Hz. Then, we segmented the filtered EEG signals into 4-second epochs without overlapping. For artifact rejection, since automated processing was a design constraint, we discarded classical threshold rejection based on visual inspection and we applied the Autoreject rejection algorithm[105]. This algorithm relies on Bayesian optimization and cross-validation to find a separate amplitude threshold per channel and is available as a Python library. The algorithm also marks bad epochs when most channels are artifactual and enables channel interpolation to preserve recording information instead of discarding the data. In addition, we applied independent component analysis (ICA) to reconstruct the EEG signals without blink artifacts[106]. To do so, we used Fp1 channel as electrooculogram proxy, since prefrontal channels are sensitive to blink artifacts.

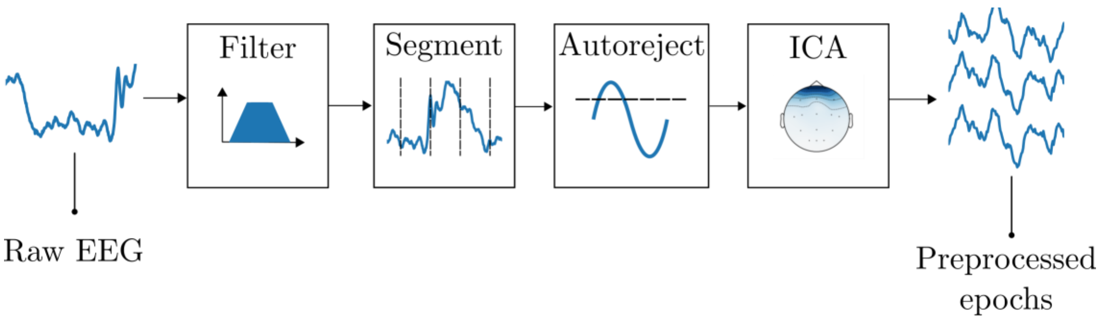


Figure 6.2: Signal processing pipeline for the studies on early AD detection.

Following signal processing, we performed feature extraction using MNE Python library (version 1.0.3), an open-source package for the analysis of neurophysiological data. Particularly, we extracted the relative power (RP) in the five main EEG bands, the Hjorth complexity (HC), and the spectral entropy (SE). We selected these features based on previous knowledge of the effects of AD on EEG activity[28] (see subsection 3.1). Corresponding formulas are reported in Equation 6.1 to Equation 6.3.

$$RP = \frac{\sum_{f_i}^{f_o} P}{\sum_{\forall f} P}$$

Equation 6.1: Formula for the relative power. *P*, *fo*, and *fi* stand for power, higher frequency of the band, and lower frequency of the band, respectively.

$$SE = -\sum_{f} S(f) * log_2 S(f)$$

Equation 6.2: Formula for the spectral entropy. *S*, and *f* stand for normalized power spectrum and frequency, respectively.

$$HC = \frac{\sigma_s''/\sigma_s'}{\sigma_s'/\sigma_s}$$

Equation 6.3: Formula for the Hjorth complexity. σ_s, σ_s', and σ_s'' stand for the standard deviation of the signal, of its first derivative, and of its second derivative.

- **Analysis**

After feature extraction, we performed classification using the scikit-learn Python package for machine learning (version 1.0.2). In the article annexed in Appendix E, we defined two binary classification problems: MCI vs HC and AD vs HC, which represent the most extended approaches in the literature. Then, in the article in Appendix F, we went a step forward and we defined a three-class classification problem: MCI vs AD vs HC.

In both articles, we used grid-search cross-validation under a leave-one-subject-out (LOSO) strategy to tune the hyperparameters of the classifier and assess its performance. Nonetheless, for the article in Appendix E we considered a LR and an SVM, whilst for the article in Appendix F we evaluated an MLP classifier. Lastly, to measure classification performance, we estimated the confusion matrices at the epoch level and the participant level, and the precision, recall, and F1 scores.

6.1.3 Results

The next tables show the classification performance in terms of cross-validation confusion matrices for the classification problems evaluated in the articles included in Appendices E and F.

Table 6.1 and Table 6.2 display the epoch-level and the subject-level confusion matrices, respectively, for the two binary classification problems considered in Appendix E. These results were yielded by the SVM, which outperformed the LR. The epoch-level matrix was calculated from the individual epochs of the participants. The subject-level matrix was estimated by considering the participants were correctly classified when most of their epochs were correctly classified.

Table 6.1: Epoch-level confusion matrices for the binary classification problems.

	AD vs HC		**MCI vs HC**	
	Patients	Controls	Patients	Controls
Patients	0.88	0.12	0.88	0.12
Controls	0.17	0.83	0.01	0.99

Table 6.2: Subject-level confusion matrices for the binary classification problems.

	AD vs HC		**MCI vs HC**	
	Patients	Controls	Patients	Controls
Patients	8	0	4	1
Controls	1	7	0	8

Table 6.3 and Table 6.4 display the analogous information for the three-class classification problem assessed in Appendix F.

Table 6.3: Epoch-level confusion matrix for the three-class classification problem.

	MCI vs AD vs HC		
	MCI	AD	HC
MCI	0.82	0.18	0.00
AD	0.05	0.88	0.07
HC	0.00	0.22	0.78

Table 6.4: Subject-level confusion matrix for the three-class classification problem.

	MCI vs AD vs HC		
	MCI	AD	HC
MCI	5	0	0
AD	0	8	0
HC	0	1	7

The results that we obtained are promising, because they suggest that early AD and MCI may be automatically detected in real time using EEG recordings of a few minutes and a commercial acquisition system.

Daily-life neurology practice is often affected by the congestion of the clinical systems, what reduces the span of medical checks and delays the access to medical trials. With this in mind, for our approach, we utilized a wearable EEG acquisition system, we registered EEG activity in resting state, and we automatized signal processing, feature extraction, and classification. This presents a series of advantages compared to traditional medical techniques in terms of cost, availability, and patient comfort. For instance, additional personnel is not required to visually inspect the EEG and the brain activity recording and classification can be completed in a single session.

Hence, our approach may contribute to opening the door for automated detection techniques that support standard medical procedures. Nonetheless, further research is required before these applications are transferred to practice for assisting clinicians.

6.2 Four-channel montage for early AD detection

As a step further towards investigating automated approaches for the real-time detection of early AD using wearable EEG systems, in this work we examined the classification of early AD vs HC using a reduced montage with only four channels. We shared the results of this research in the IWINAC conference held in Tenerife (Spain) in 2022. Following this contribution, the conference editors invited us to extend our conference paper into a research article for possible publication in the *International Journal of Neural Systems*. At the time of writing this, the article is under review. For this reason, we did not include it in the article compendium.

6.2.1. Objective

The goal of this work was to investigate a reduced montage with only four channels to detect early AD using resting state EEG. The database, signal processing, and feature extraction procedures utilized for this study were exactly those reported for the articles in Appendices E and F (see subsection 6.1). As described above, we studied only early AD and HC groups, therefore, we analyzed the data from 16 participants (8 early AD and 8 HC).

6.2.2. Methodology

To extend our previous contribution, we dedicated the first part of this work to perform a statistical comparison of the different features extracted from the EEG activity between the early AD and the HC groups. To do so, for each extracted feature, we compared the distributions of the two groups using a t-test for independent samples. Subsequently, to find a reduced four-channel montage, we calculated a relative univariate feature score per channel by summing the p-value from the group comparisons corresponding to all the features (see Equation 6.4). Then, we plotted this score in a topographical map. For the reduced montage, we selected the four channels according to the univariate feature score and the appropriateness of the channel location for a minimal wearable design.

$$Score\,(x) = \frac{-\log_{10} p_{value}\,(x)}{\max\,(-\log_{10} p_{values})}$$

Equation 6.4: Formula for the univariate feature score. x represents a particular feature (relative power, sample entropy, or Hjorth complexity) for a particular channel.

Finally, we applied grid-search cross-validation under a LOSO strategy to evaluate the performance of a LR, an SVM, and an MLP for the classification of early AD vs HC using the original sixteen-channel montage and the proposed four-channel montage.

6.2.3. Results

Figure 6.3 displays the results of the statistical comparison between early AD and HC in terms of EEG relative power across the sixteen channels used for recording.

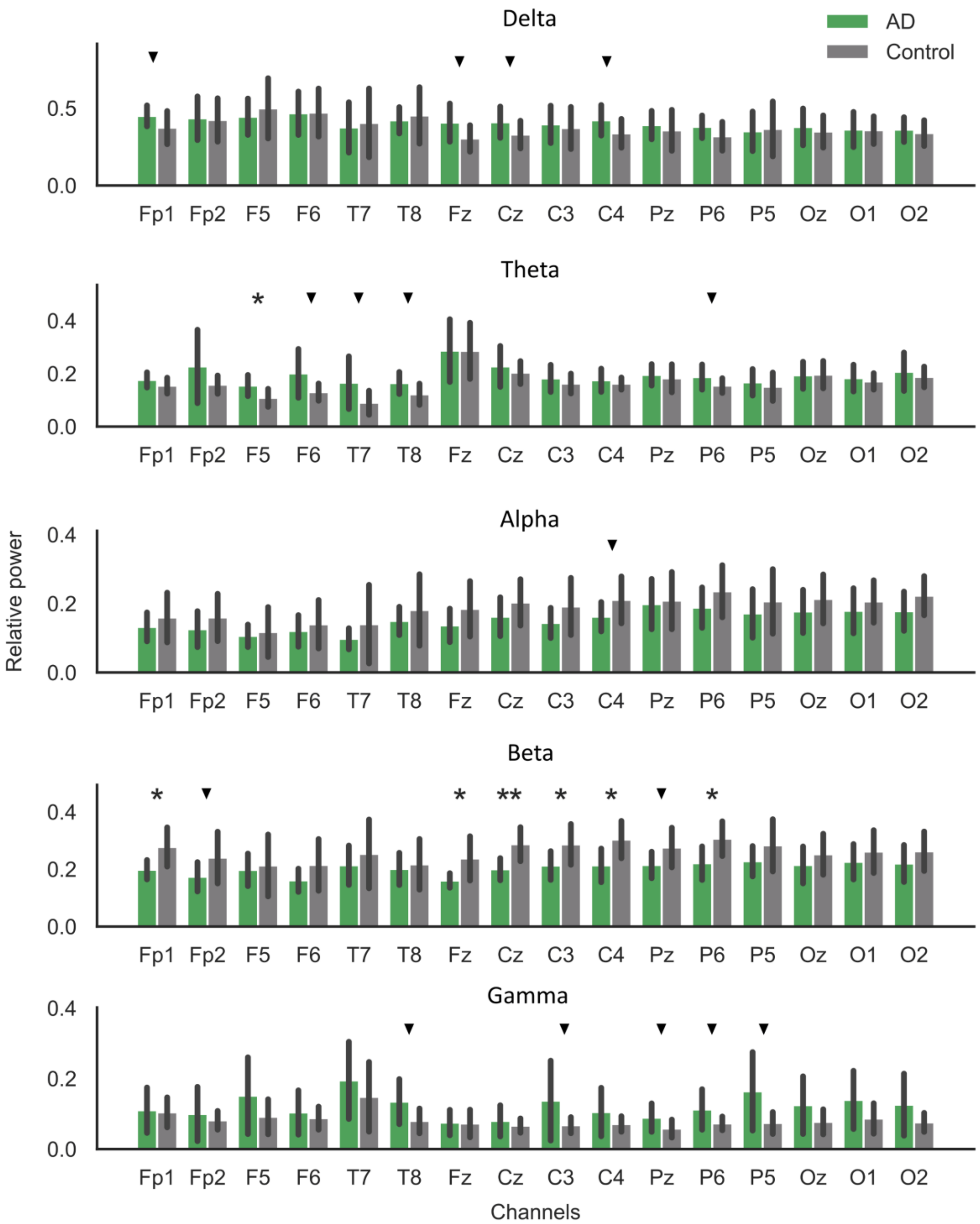


Figure 6.3: Statistical group comparison in terms of relative power. A chart is included for each of the main EEG bands. *X* axis shows channels, and *Y* axis shows relative power. ▼ indicates p-value ≤ 0.15, * indicates p-value ≤ 0.05, and ** indicates p-value ≤ 0.01.

The results presented in the previous figure, suggest the generation of a slowing pattern for the early AD group compared to the controls, what is consistent with the literature (see subsection 3.3). This translates into an increase of the relative power in the delta and theta bands, coupled with a decrease in the alpha and beta bands. The results obtained for the alpha band are less statistically solid, what may have been caused by the eyes open condition followed during recording, which mainly affects this band, and the reduced sample size. With regard to gamma, our results suggest an increase of the relative power for the early AD group, which is consistent with related studies[107,108]. Nonetheless, this band remains substantially less studied in the literature.

Figure 6.4 displays the results of the statistical comparison between early AD and HC in terms of EEG complexity across the sixteen channels used for recording.

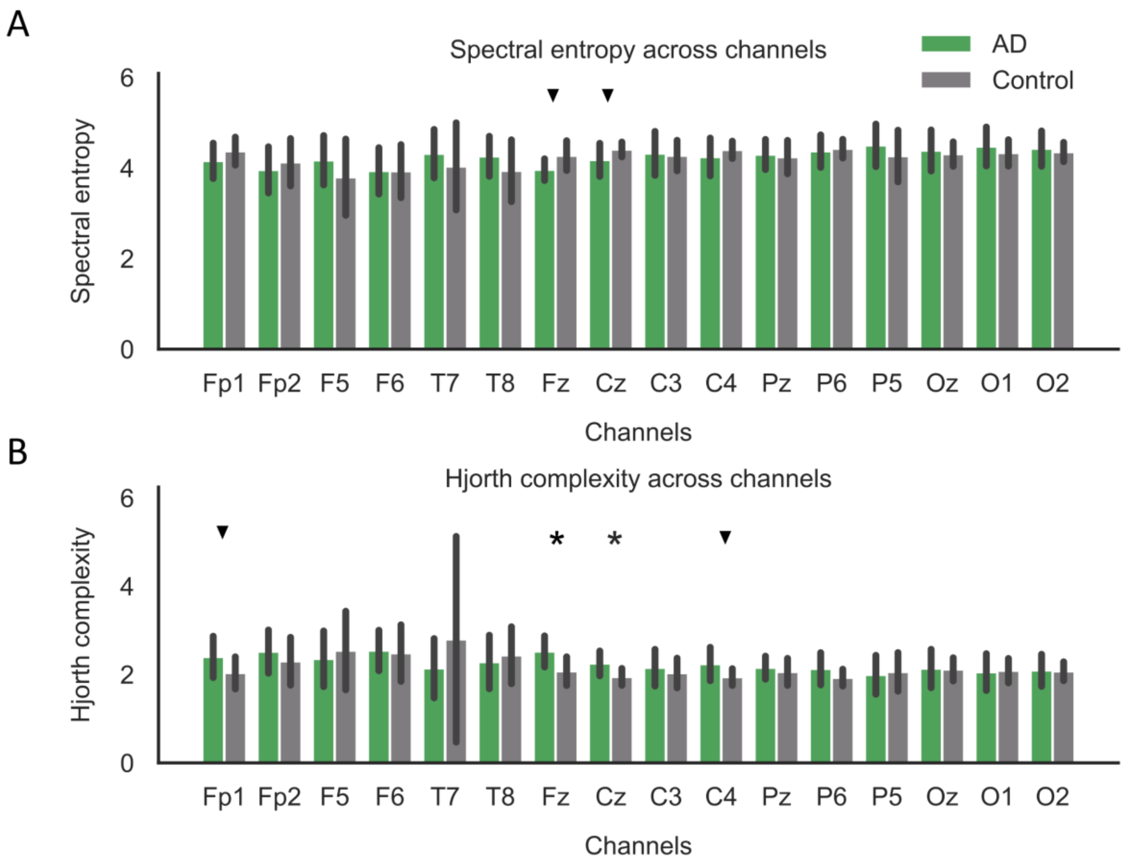


Figure 6.4: Statistical group comparison in terms of EEG complexity. (A) Spectral entropy, and (B) Hjorth complexity. ▼ indicates p-value ≤ 0.15 and * indicates p-value ≤ 0.05.

Since spectral entropy and Hjorth complexity increase with complexity and regularity, respectively, the results displayed in the figure above suggest a loss of EEG complexity for the early AD group compared to the controls. This agrees with the prevailing view in the literature (see subsection 3.3).

Figure 6.5 shows the accumulated feature score per channel estimated from the statistical comparison of all the EEG features across the two study groups.

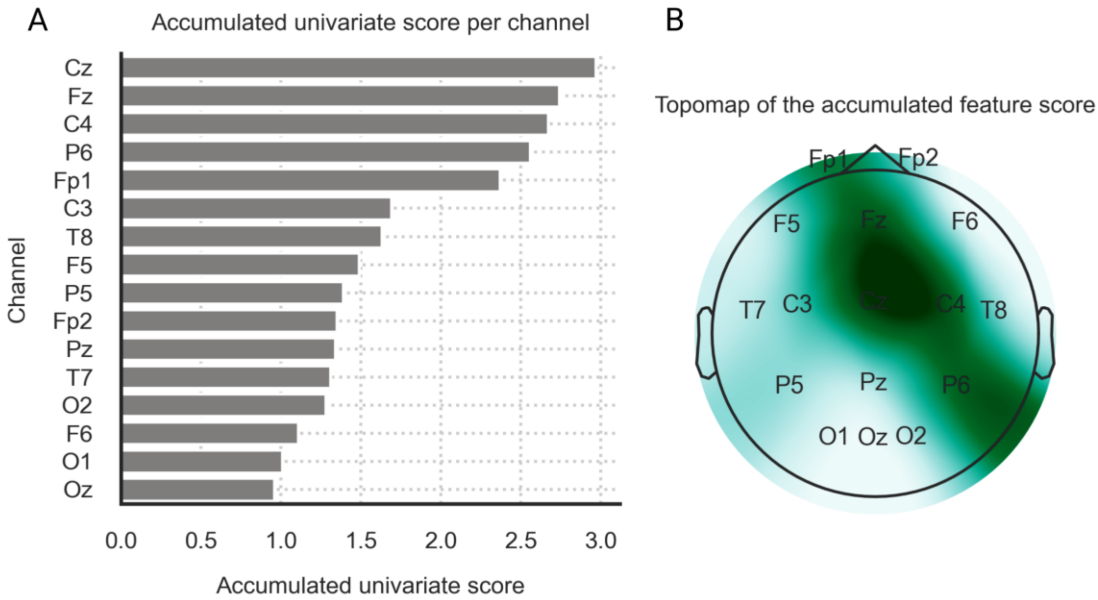


Figure 6.5: Accumulated feature score analysis. (A) Accumulated feature score per channel, and (C) Topographic representation of the accumulated feature score.

Based on these results, we selected channels Cz, Fz, and C4 for the reduced montage because they were ranked top three in terms of accumulated feature score. Then, we added a left hemisphere channel because performance was not sound enough when including only the three mentioned right hemisphere channels. With a view to maximizing the wearability of the proposed montage, we selected C3 (which was also highly ranked according to Figure 6.5.A) as the fourth channel for the montage.

Finally, Table 6.5 presents the confusion matrices at the epoch and subject levels resulting from the cross-validation predictions for the original sixteen-channel and the reduced four-channel montages.

Table 6.5: Subject and epoch confusion matrices for the two montages studied.

Level	Group	Montage: Original AD	Original Control	Reduced AD	Reduced Control
Epoch	AD	0.92	0.08	0.83	0.17
	Control	0.17	0.83	0.12	0.88
Subject	AD	8	0	7	1
	Control	1	7	1	7

According to the results presented in the previous table, the performance yielded by the two montages was very similar, with slightly better results for the original sixteen-channel montage. Particularly, the average cross-validation accuracies were 0.87 ± 0.23 and 0.86 ± 0.20 for the original and the reduced montage. These results were obtained with a LR and a MLP classifier, respectively.

In light of the results obtained, this study represents a preliminary further step towards automated early AD detection techniques based on minimal EEG montages. With this in mind, we foresee a scenario where primary care clinicians can setup a plug-and-play EEG headband and effortlessly perform a resting state acquisition and apply a self-driven detection pipeline. This may represent a valuable supporting tool preceding the application of standard clinical trials. As outlined in the conclusions of the previous works in this chapter, these approaches could contribute to reducing screening costs and to enhancing the availability of EEG-based early AD analysis, especially in scenarios such as congested healthcare systems and rural environments.

6.3 Cognitive impairment detection via visual dynamics

The final work performed in collaboration with the neurologists from the CBNU at HVNU refers to the development of a web application for the detection of cognitive impairment. This application has been already implemented and is currently being utilized by the neurologists at HUVN to evaluate their patients. The analysis of the results obtained will be conducted once a large scale database is generated.

6.3.1 Objective

An essential aspect for the detection of early AD is cognitive impairment screening. For this, clinicians apply neuropsychological evaluations to assess different cognitive areas affected by AD, like orientation, language, and memory. As explained previously in this chapter, the clinical checks conducted in primary care and neurology services are very limited in time. Therefore, reliable and brief evaluations are required for cognitive impairment screening.

With this in mind, the neurologists at CBNU were interested in an application that could be used within minutes to test the patients during clinical checks in the neurology consulting facilities. Consequently, the application should be analogous to a brief cognitive assessment[109] such as the Fototest[110–112] or the Eurotest[113], which are internationally accepted. For such application, they proposed the following features:

- Implementation of a short cognitive task to evaluate cognitive impairment.
- Visualization of patient information and performance results.
- Accessibility from any electronic device.
- Printable reports within the app.
- Different user profiles identified by username and password.
- Easy to use.

6.3.2 Methodology

Following the requirements listed above, we implemented a web application based on the rapid serial visual presentation (RSVP), a cognitive evaluation to assess visual attention and dynamics, which are reportedly compromised in AD patients[114,115].

- **The RSPV**

For the RSVP, participants are presented a fast stream of characters including several distractors and two targets, as illustrated in Figure 6.6. This methodology is repeated for a number of trials. For each trial, the aim is to identify the two red numbers within the stream. The number of intervening distractors (those between the two targets) is varied across trials to challenge visual attention.

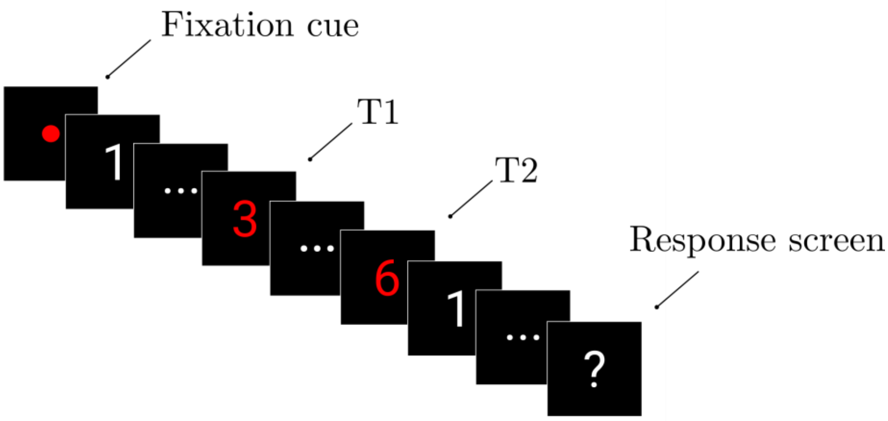


Figure 6.6: Illustration of a RSVP trial. Distractors and targets are displayed in white and red, respectively. T1 and T2 stand for the first and second target.

According to previous studies, AD patients experience two notable effects when they are evaluated using the RSVP: attentional blink (AB) and masking (AM). AB corresponds to the inability to remember T2 after correctly reporting T1. Alternatively, AM refers to the inability to remember T1 after correctly reporting T2. Conversely, healthy older adults do not experience AM. Both effects are illustrated in the results section (see Figure 6.12).

Considering the methodology behind the RSVP, this cognitive evaluation could potentially be auto-graded and also self-administered. These features are advantageous towards lowering the costs linked with cognitive impairment detection and shortening the screening procedures. Additionally, this test may be combined with other evaluations to enhance the discriminability of cognitive impairment procedures, and could also be used to follow up patients under clinical treatment.

With this in mind, and based on the features listed in subsection 6.3.1, the neurologists at CBNU considered the RSVP as a potential assessment tool for detecting cognitive impairment. Therefore, we implemented this task as the main feature of our app.

- **App workflow**

We designed the app with the following user workflow in mind:

1. Clinicians sign in with their email using the app login form.
2. A provisional password is automatically sent to them via email.
3. After login in, they can access the following features:
    a. Test their patients using the RSVP.
    b. Visualize RSVP performance and compare it with mean results.
    c. Store and modify patient information (diagnosis, education, sex, etc.).
    d. Download RSVP reports in PDF format.

Additionally, the clinicians who use the app only have access to their patients, who are identified using a randomly generated nine-character alpha numeric code to preserve their identities anonymous.

- **App implementation**

To implement the application, we used open-source software (see Figure 6.7). Particularly, we implemented the application backend and fronted using Flask, a Python framework for app development. Using this framework, we programmed the fronted (what the user sees) using HTML (hypertext markup language) and JavaScript, and the backend (what the user does not see) using Python. Furthermore, we implemented the RSVP using PsychoPy toolbox for Python (2020 version), an open-source software library for designing psychology experiments[116]. Since the application is web-based, the RSVP had to be converted from Python to JavaScript for web browser compatibility.

To store the patient results and the user credentials, we deployed a MySQL database server. The database in this server included three tables: *clinicians*, *patients*, and *recordings*. *Clinicians* table stored the credentials of the clinicians (email and hashed password for login); *patients* table stored information about the patients (diagnosis, age, sex, etc.); finally, *recordings* table stored information about a specific RSVP evaluation (participant, clinician, time, score, etc.) (see Figure 6.8). No information that could be used to identify the patients was stored in the database.

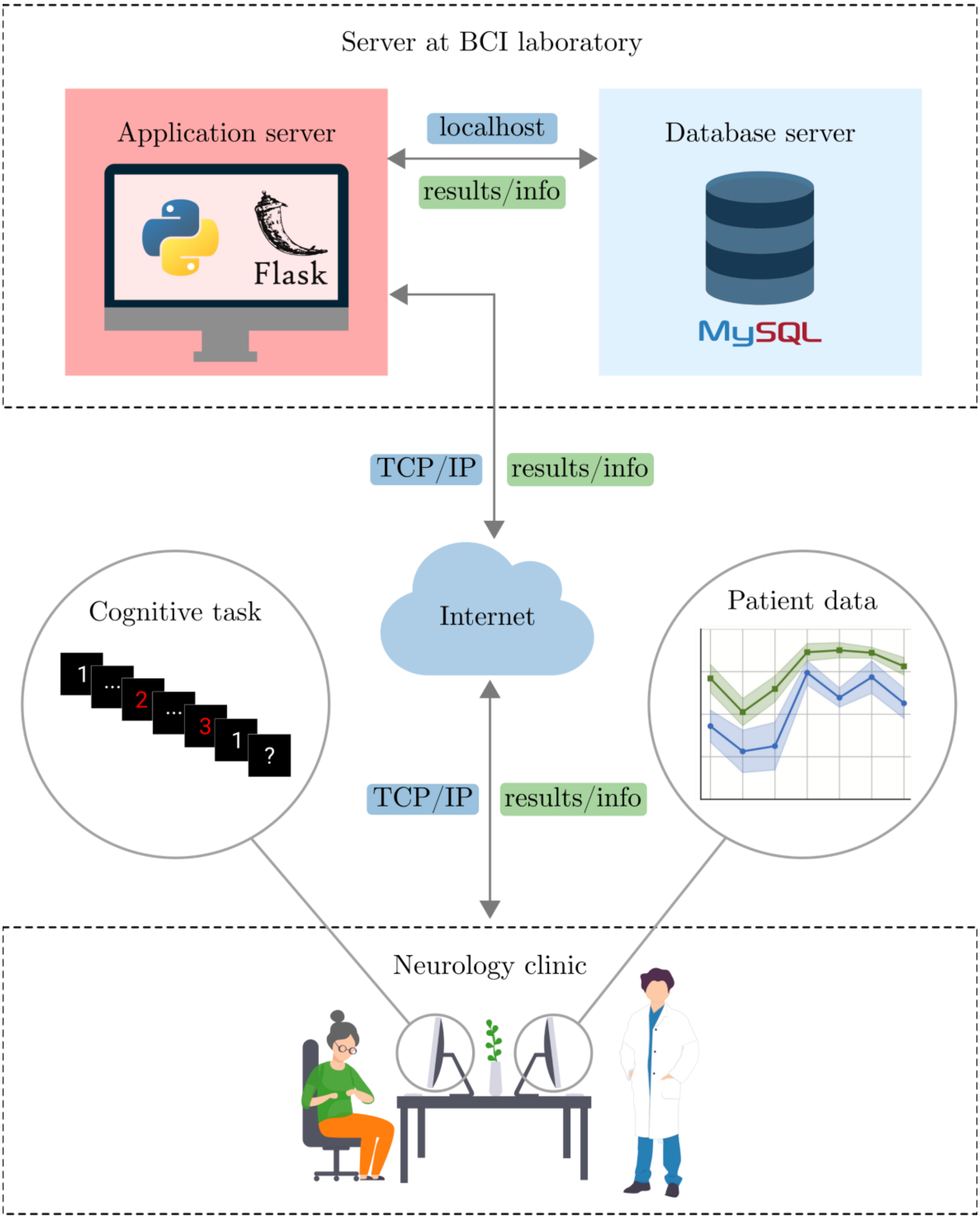


Figure 6.7: Envisioned scenario for the cognitive impairment detection web app. The clinician uses the web application to test the patient and to visualize patient results and information. The data is transmitted through the Internet to the server located at the BCI laboratory for storing and management.

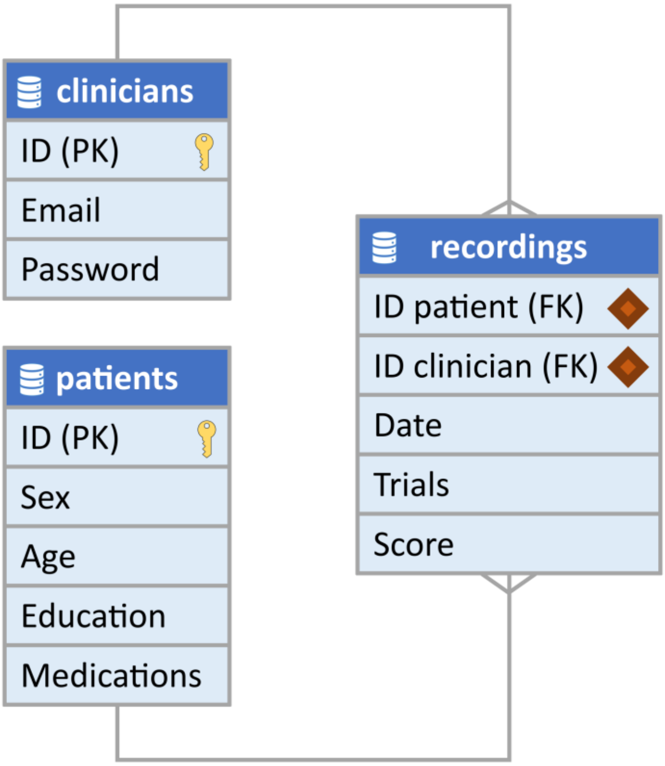


Figure 6.8: Diagram for the RSVP MySQL database tables. *Clinicians* table holds the login information of the clinicians. *Patients* table store the information of the patients without including personal details. *Recordings* table holds the information regarding the RSVP assessments. Each recording is linked to the other tables by the ID of the patient who was evaluated and the ID of the clinician who conducted the evaluation.

For scalability and safety, we developed both the application server and the database server in a Virtual Box virtual machine installed in a server at the BCI laboratory. This virtual machine was connected to the Internet through the physical server using port forwarding.

6.3.3 Results

At the time of writing this thesis document, the first version of the web application for the detection of cognitive impairment is being used by the neurologists of the CBNU at HUVN. They have incorporated the app to their consulting protocol, and they apply the RSVP test to the patients who voluntarily accept to collaborate on this project. They access the service from their consulting room using an Android tablet. The web app is available at http://necolab.ugr.es.

Subsequently, we show some screenshots corresponding to the different modules of the current version of the application and the results of a preliminary data capture conducted at HUVN.

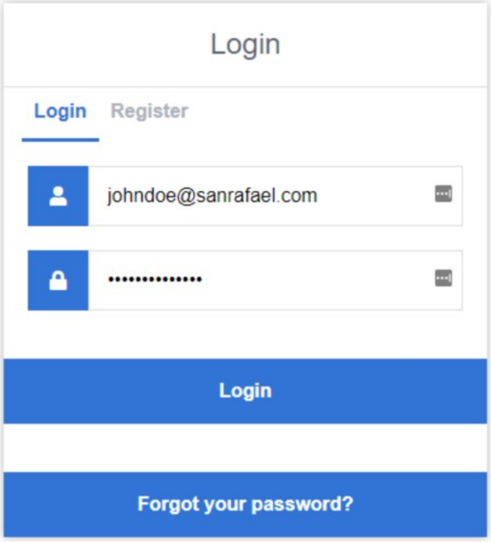


Figure 6.9: Login screen for the cognitive impairment detection app.

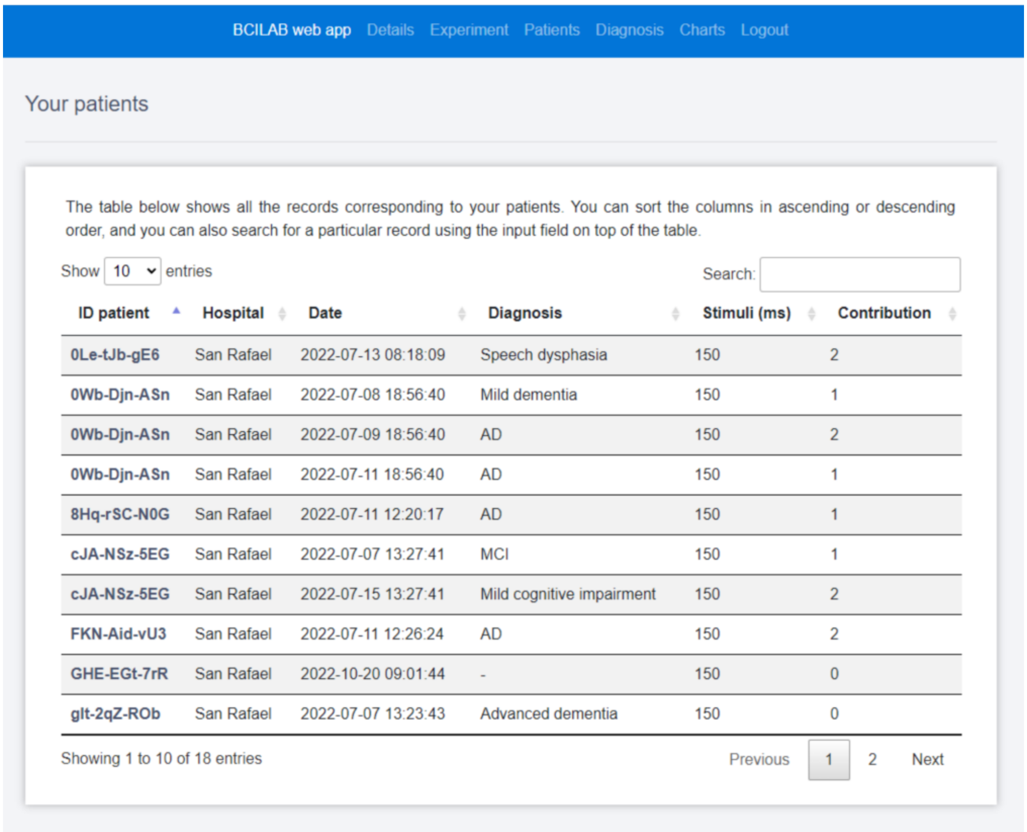


Figure 6.10: Visualization of a clinician's patients' information in the web app.

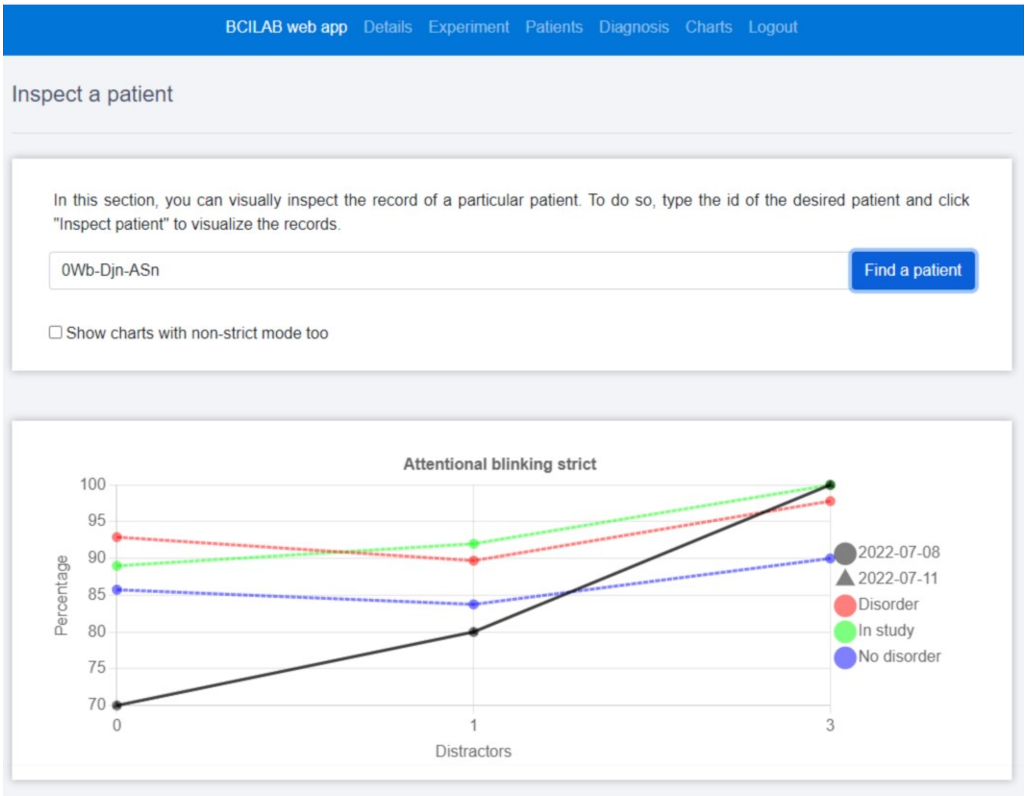


Figure 6.11: Visualization of the patients' RSVP performance in the web app.

To preliminarily evaluate the RSVP approach for cognitive impairment detection, we tested the participants of the study reported in section 6.1. To do so, we grouped the early AD and MCI patients into the same group (PAT, standing for cognitive impairment patients), and we compared their results with the controls. Figure 6.12 represents the preliminary results obtained in terms of AB and AM.

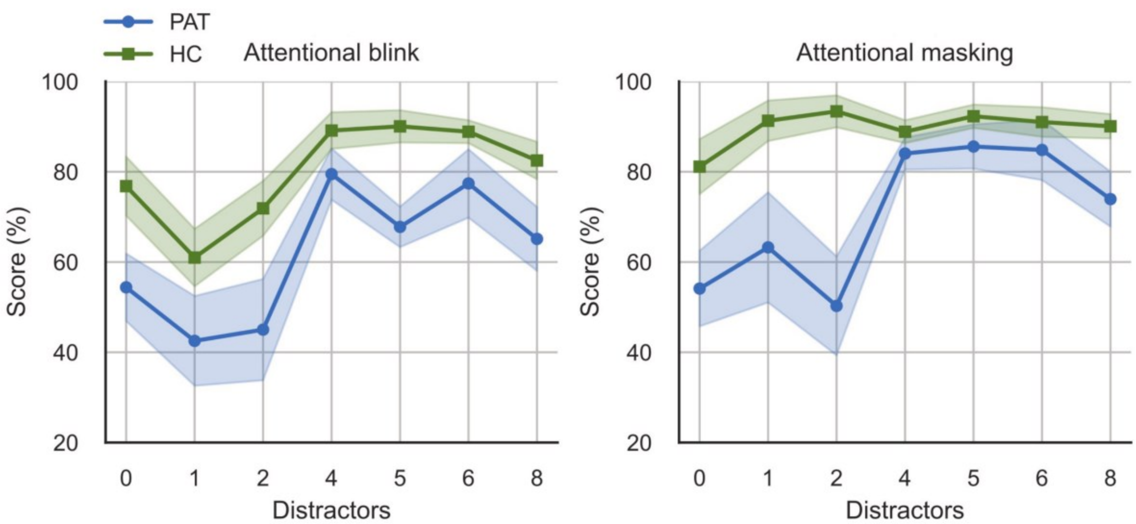


Figure 6.12: Attentional blink and masking effects. *PAT* stands for patients and *HC* for healthy controls.

Finally, Figure 6.13 shows the statistical group comparison between the results of the patients and controls for the RSVP (AB and AM) and for other two cognitive tests conducted during the neuropsychological evaluation of the participants (Clock-drawing test and Phototest).

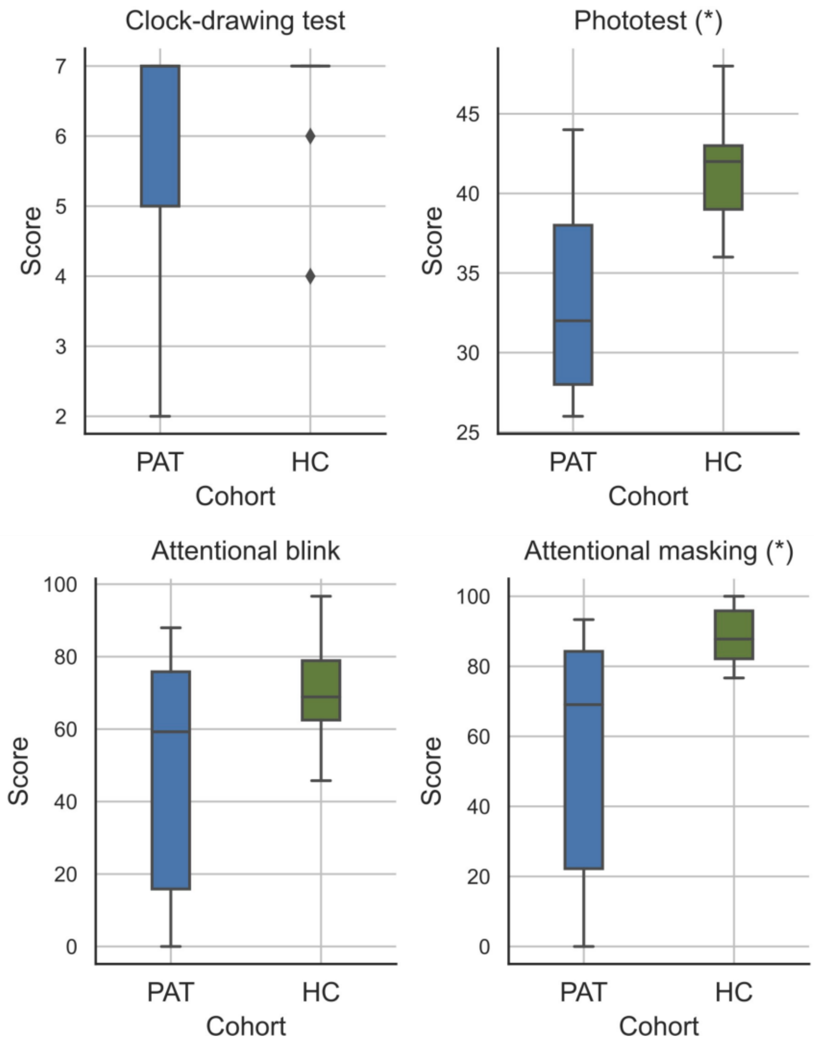


Figure 6.13: Group comparison for different cognitive tests. The clock-drawing test and the Phototest were conducted as part of the neuropsychological evaluation of the participants. The lower charts show the average AB and AM across distractors 0 to 2 of the RSVP. * represents p-value ≤ 0.05. *PAT* and *HC*, stands for patients and controls.

According to the preliminary results obtained, only the Phototest and the average AM yielded statistically significant differences between patients and controls. Although these results are promising, we intend to augment our database with the data collected by the neurologists during their clinical practice. Furthermore, we have included some modifications to the RSVP implementation following the advice of the neurologists. Such modifications affected the trial length, the target color, and the task duration. All these adjustments have been included for the RSVP task implemented in our web application, what may contribute to increasing the discrimination capability of the test.

With regard to the creation of a longitudinal database, which we defined as another objective of this work (see subsection 1.2), the web application is successfully contributing to such task according to the number of recordings stored already. When this number is acceptably large, we will perform a more thorough analysis to evaluate the feasibility of the RSVP for the detection of cognitive impairment. Furthermore, during the app design, we focused on its practical transfer to the clinical practice. Therefore, we envision a concurrent use by a large number of clinicians in different locations in neurology and primary care services. With this in mind, if the results obtained are positive, this approach may help assisting clinicians in early AD screening, as cognitive impairment could be diagnosed remotely and efficiently.

Chapter 7: Collaborations

In this chapter, we provide a brief reporting of the research collaborations conducted during this thesis work. These collaborations were focused on two main topics: the assessment of stress based on EEG and VR, and the analysis of cardiorespiratory fitness through EEG complexity.

7.1 Stress assessment through EEG and VR

In addition to the articles about stress assessment included in the compendium, we published two additional articles as a result of the collaboration between members of the BCI laboratory and the psychologists at San Rafael special needs education school: "Virtual Reality as a Portable Alternative to Chromotherapy Rooms for Stress Relief: A Preliminary Study"[98] and "Virtual Reality Customized 360-Degree Experiences for Stress Relief"[101].

The main objective of these two works was to investigate the feasibility of VR technology to represent a viable alternative to stress relief approaches currently used in practice.

- Particularly, the objective of the first work was to evaluate the feasibility of a VR application to replicate the stress relief therapies conducted in chromotherapy rooms. For this purpose, we simulated a chromotherapy room, a chamber specifically designed to promote relaxation by a combination of ambient light colors, music, and scents[98] (see Figure 7.1).
- Alternatively, the objective of the second work was to evaluate an alternative to chromotherapy based on 360-degree immersive experiences. To do so, during the relax stage of the study, the participants viewed a selection of four 360-degree relaxing videos using a VR HMD.

The data captures collected for these two studies were used for the original works presented in Chapter 5: of this thesis, which are included in Appendices C and D. Therefore, for a description of the experimental protocol designed for the two studies referred in this section, review subsection 5.1.2 and 5.2.2.

Below, we provide a summary of the analysis performed in both works. For a comprehensive description of these works, refer to the published articles[98,101].

- First, to measure the stress level of the participants throughout the session, we processed the EEG activity to estimate the relative gamma time series, and we collected the SPSL of the participants via surveys. The signal processing was conducted in MATLAB R2016a (Nattick, USA) and the SPSL surveys were embedded in the MIST and the VR applications.
- Importantly, the relative gamma signal is directly or inversely related with the stress level according to studies in the literature[16,117,118]. Consequently, we inverted this signal for some of the participants based on visual inspection.
- Finally, we compared the stress levels across the different stages of the experiment to evaluate the feasibility of the relax session for stress relief. We compared the stress levels in terms of the relative gamma time series and the SPSL surveys across the different stages. For this purpose, we applied correlation analysis and the Wilcoxon signed rank non-parametric statistical test, respectively.

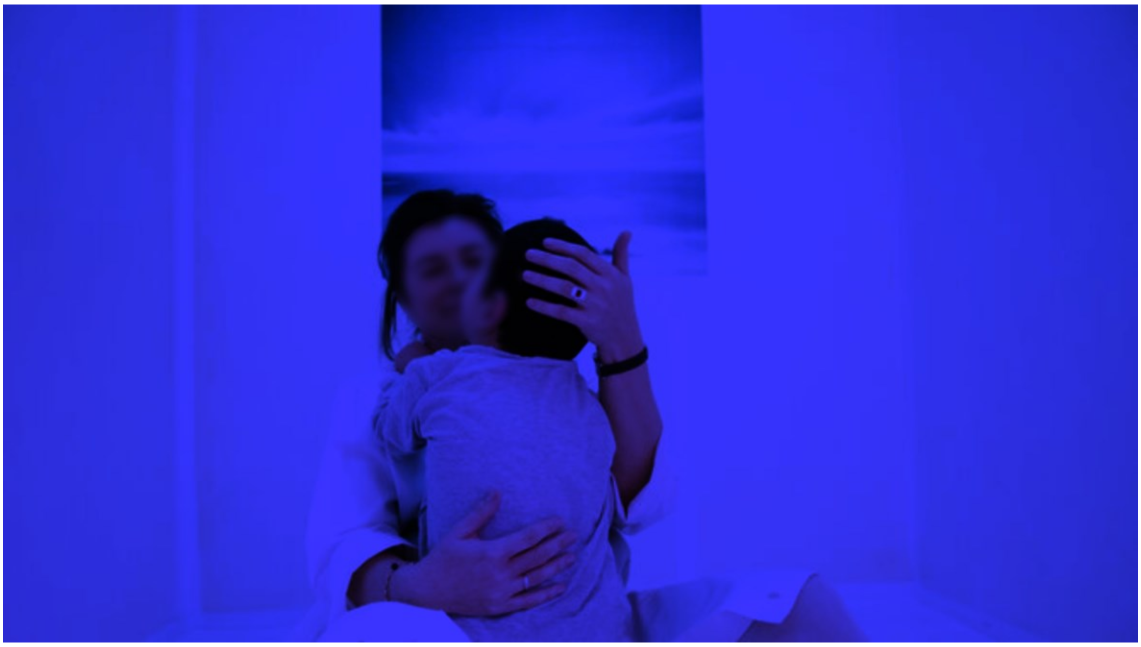


Figure 7.1: A therapist conducts a chromotherapy session. Source: diariodesevilla.es.

7.2 Cardiorespiratory fitness analysis via EEG

The last work completed in this thesis corresponds to a collaboration of the BCI laboratory with members of the department of Physical and Sports Education at UGR[119,120], which has not been published at the time of writing this document.

Recently, researchers from the department of Physical and Sports Education studied the effects of physical fitness and physical activity in the working memory of overweighted kids[121]. They conducted a group of 110 kids through a delayed non-match to sample (DNMS, see Figure 7.2) test while they recorded their EEG activity and multiple fitness variables. Subsequently, they estimated the DNMS behavioral performance (accuracy and mean reaction time) and the P300 (amplitude and latency), which is an accepted marker of working memory and processing speed.

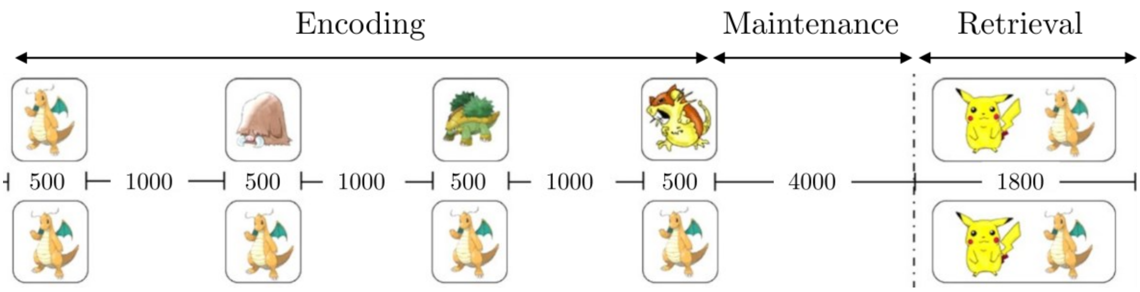


Figure 7.2: Delayed non-match to sample task. Participants are presented a stream of four images during the encoding phase; then during the retrieval phase they are shown two images and asked to report the one which was not present in the original stream. The task includes two cognitive loads: high (upper stream) and low (lower stream). The numbers between the two streams represent time in milliseconds.

With this in mind, we presented them a proposal to analyze the EEG of the participants in terms of complexity during the maintenance phase, which is the stage of the task prior to retrieval when the participants hold the stimuli in mind. For this purpose, we estimated the line length, a measure of complexity previously applied in the EEG literature[25,122]. Our proposal had two main objectives.

- To assess the differences in EEG complexity, P300, and DNMS behavioral performance across cardiorespiratory fitness levels.
- To analyze the correlation between the EEG complexity and P300 amplitude and latency.

To measure cardiorespiratory fitness, we grouped the participants according to the number of laps completed during a shuttle run test. Concretely, we split the participants into two levels of cardiorespiratory fitness using the median separately for both sexes based on previous studies in the literature[123].

With respect to the EEG data, the recordings were acquired using a 64-channel montage, yet we only considered eight regions of interest including three to four channels each (frontal, temporal, parietal, and occipital for both hemispheres). We used the MNE Python library (version 1.0.3) to process the EEG recordings and to extract the EEG complexity. Then, we calculated the average complexity across the maintenance phase of the DNMS trials for all the participants.

To evaluate the differences between the two cardiorespiratory fitness groups in terms of EEG complexity, P300 amplitude and latency, and behavioral performance, we applied the Shapiro-Wilk test and the Mann-Whitney test. Finally, to assess the correlation between the P300 metrics and the EEG complexity we applied the Spearman correlation coefficient along its corresponding p-value.

Table 7.1 displays the correlation between line length, P300 amplitude, and P300 latency with the quantitative cardiorespiratory fitness level for the two mental loads and the eight scalp regions of interest.

Table 7.1: Correlation between EEG metrics and fitness for all regions. The first value in each cell corresponds to the Spearman correlation coefficient, whilst the second value refers to its p-value. Green cells represent the strongest correlation for the region.

	Low load			**High load**		
Region	**Line length**	**P3 amplitude**	**P3 latency**	**Line length**	**P3 amplitude**	**P3 latency**
frontal l	-0.316, 0.004	-0.171, 0.127	0.188, 0.092	-0.29, 0.007	-0.085, 0.432	0.185, 0.087
frontal r	-0.27, 0.015	-0.189, 0.091	0.1, 0.372	-0.237, 0.027	-0.155, 0.153	0.383, 0.0
temporal l	-0.319, 0.004	-0.257, 0.021	0.099, 0.381	-0.318, 0.003	0.024, 0.822	-0.041, 0.703
temporal r	-0.3, 0.007	0.005, 0.968	0.227, 0.042	-0.268, 0.012	-0.049, 0.655	0.148, 0.172
parietal l	-0.274,0.013	0.143, 0.203	-0.36, 0.001	-0.274, 0.01	0.052, 0.631	-0.264, 0.013
parietal r	-0.241,0.031	0.259, 0.02	-0.181, 0.106	-0.214, 0.047	0.211, 0.05	-0.024, 0.824
occipital l	-0.277,0.012	0.195, 0.081	-0.269, 0.015	-0.258, 0.016	0.127, 0.243	-0.253, 0.018
occipital r	-0.157, 0.162	0.234, 0.035	-0.083, 0.463	-0.145, 0.179	0.202, 0.061	-0.082, 0.451

Figure 7.3 represents the statistical comparison between the two fitness groups in terms of the line length for the low and high mental loads and all the regions of interest.

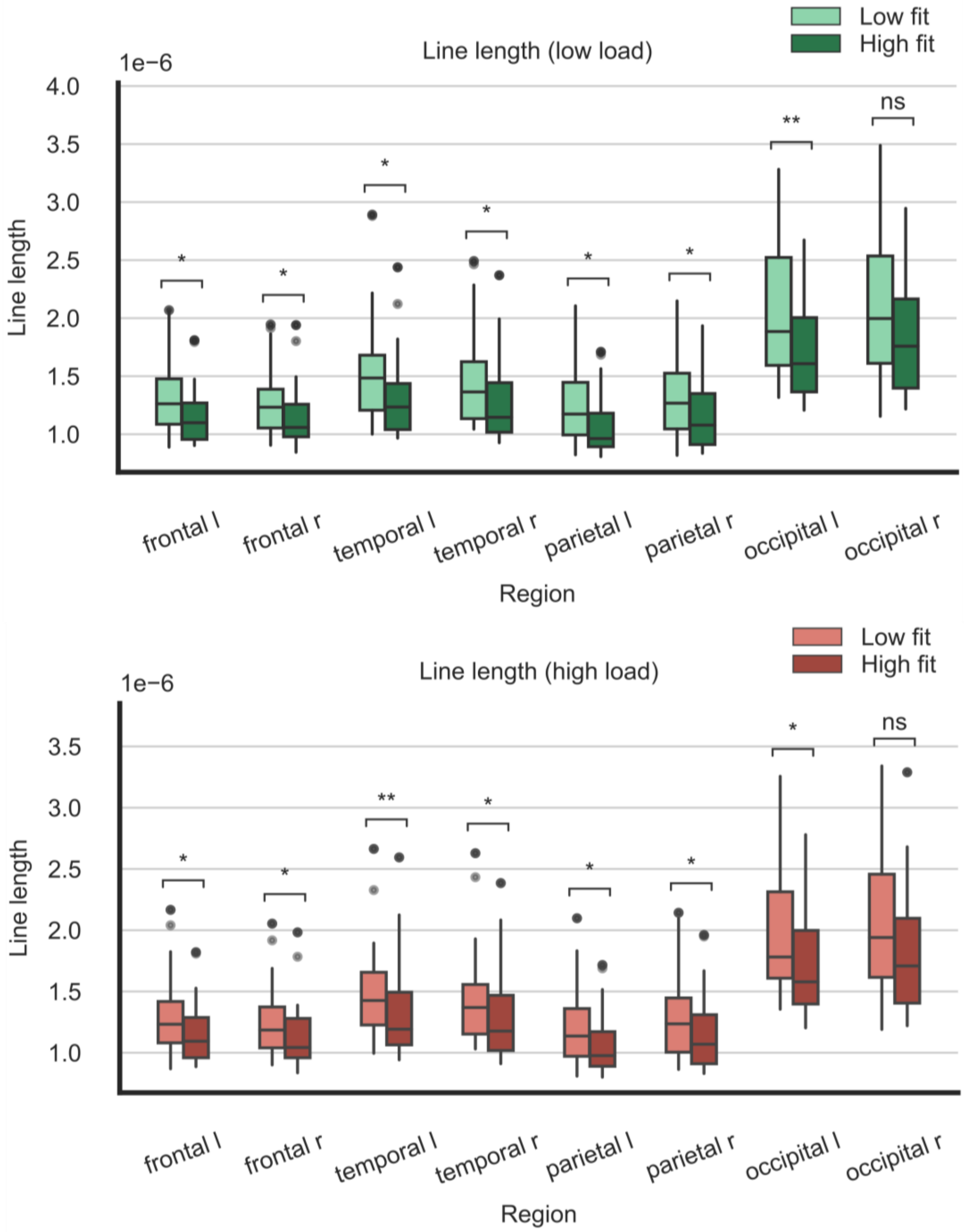


Figure 7.3: Comparison of fitness groups by line length for the mental loads. ns, *, and ** stand for not significant, p ≤ 0.05, and p ≤ 0.01, respectively.

The group comparison presented above suggest a reduced EEG complexity in terms of line length for the fitter group. These results are consistent across all the scalp regions analyzed.

As described in the beginning of this subsection, the results of this work have not been published at the time of writing this thesis document. Nonetheless, we have drawn the main conclusions from the results obtained.

- The results obtained in this work suggest an increased cognitive efficiency in terms of executive control for the fitter group: higher fit participants yielded lower EEG complexity as well as higher and faster P300. This hypothesis is in line with similar works on the links between cardiorespiratory fitness and EEG[123,124].
- Plus, when analyzing the correlation with the quantitative cardiorespiratory fitness level, we found that the line length exhibited higher correlation compared to other well-established EEG metrics such as P300 amplitude and latency.
- In conclusion, in this work, we found statistical evidence of a reduced EEG complexity for higher fit participants compared to their lower fit peers during the maintenance phase of the DNMS task. Particularly, we observed this pattern throughout all the scalp areas studied and for the two cognitive loads evaluated by the task.

Chapter 8: Conclusions

In this final chapter, we briefly review the main contributions derived from this thesis work, we highlight its limitations, and we propose future work to build upon the research presented in this document.

8.1. Contributions

In this document, we presented a multidisciplinary research work which combined the study of EEG activity and the application of signal processing and machine learning algorithms to address three main applications: attention detection, stress detection, and early AD detection.

In view of the results presented in Chapter 4: to Chapter 7: , we can state that the objectives proposed in the first chapter have been successfully completed:

- We investigated the combination of EEG activity and the fundamentals of gamification for the real-time detection of attention to a mobile visual stimulus. To do so, we designed a videogame based on SSVEP as part of an automated closed-loop BCI within a cloud architecture. According to the results, we delivered a challenging and entertaining experience for the participants while using their EEG activity for attention detection. Consequently, these results may contribute to developing ludic attention training therapies in fields like special education, where children may benefit from a gamified experience. Furthermore, the mobile ring-shaped stimuli designed for this work could represent a valuable alternative to classical SSVEP stimulation paradigms.
- We investigated the regression of the self-perceived stress level from EEG activity as an alternative to classical approaches in the literature, which are typically limited to three levels. For this purpose, we implemented a machine learning regression model to predict the SPSL via EEG activity during a stress-relax session. This session was conducted using a validated

methodology to induce psychosocial stress and a VR relaxing application. The results obtained could potentially be transferred to daily-life applications, such as relax therapies or neuromarketing, which may benefit from the quantitative assessment of perceived stress from brain activity.

- We investigated the automated real-time detection of early AD through machine learning algorithms and wearable EEG. First, we performed the preliminary evaluation of an automated signal processing and machine learning pipeline to detect early AD from resting state EEG activity. Subsequently, we went a step further and investigated a reduced montage with only four channels. The results obtained may contribute to opening the door for self-driven approaches based on EEG that help shorten diagnosis times. This is especially relevant in the case of low and medium income countries and rural environments, where patients often experience long waiting lists and complex medical equipment is lacking.
- We investigated the detection of cognitive impairment from visual dynamics. To do so, we designed and implemented an application for the assessment of cognitive impairment based on the RSVP task. We designed the application with a special focus on transferring the proposed approach to daily neurology clinical practice. For this purpose, we followed the guidelines of the neurologists who collaborated with us in this research. This application is currently contributing to the creation of a large scale database of patient evaluations. If the results yielded in the forthcoming future are positive, this application could enrich the framework of cognitive impairment detection techniques used during clinical practice, which are essential for early AD detection.

To accomplish these achievements, we collaborated with multidisciplinary research teams, including neurologists, psychologists, sport scientists, etc, and we published the research articles included in the article compendium for this thesis (see Appendices A to F). Furthermore, at the time of writing this, we are working on the publication of other research articles originated from these collaborations (see subsections 6.2 and 7.2).

8.2. Limitations and future work

In the context of attention detection, traditional stimulation protocols like the checkerboard with inverting pattern may be soon replaced by more immersive audiovisual experiences thanks to the advances in VR technology. Furthermore, VR systems are often more appealing to kids and young adults, what highlights the importance of implementing this technology during research. Consequently, without disregarding the importance of classical stimulation, which in some cases represents the only alternative, VR techniques for attention detection should be investigated.

Regarding AD, we believe EEG-based research should also focus on studying the prodromal stage of the disease, when symptoms have not emerged yet and the individuals cannot be catalogued as patients. For this purpose, large scale longitudinal studies are required. We propose a follow up resting state EEG evaluation of participants with AD family history and the analysis of changes in their EEG activity. Such analysis should be based on (but not be limited to) spectral power, complexity, and connectivity. In the future, we envision a scenario where the clinician may store a plug-and-play EEG system in the consulting room and use it to detect AD long before the symptoms of the disease have emerged. For this scenario to materialize, there is a need for researchers in the medical and technical fields to collaborate tightly and combine the latest advances in clinical investigations with breakthrough algorithms and technologies like deep learning.

Bibliography

1. Al-Nafjan, A. Feature selection of EEG signals in neuromarketing. *PeerJ Comput. Sci.* 8, e944 (2022).
2. Ali, A. *et al.* EEG Signals Based Choice Classification for Neuromarketing Applications. in *A Fusion of Artificial Intelligence and Internet of Things for Emerging Cyber Systems* (eds. Kumar, P., Obaid, A. J., Cengiz, K., Khanna, A. & Balas, V. E.) 371–394 (Springer International Publishing, 2022). doi:10.1007/978-3-030-76653-5_20.
3. Cheng, M.-Y. *et al.* QEEG markers of superior shooting performance in skilled marksmen: An investigation of cortical activity on psychomotor efficiency hypothesis. *Psychology of Sport and Exercise* 102320 (2022) doi:10.1016/j.psychsport.2022.102320.
4. Jiang, P. & He, H. EEG Signal Analysis based on machine learning in psychological balance detection of athletes. *Evol. Intel.* (2022) doi:10.1007/s12065-022-00770-7.
5. Fang, Q., Fang, C., Li, L. & Song, Y. Impact of sport training on adaptations in neural functioning and behavioral performance: A scoping review with meta-analysis on EEG research. *Journal of Exercise Science & Fitness* 20, 206–215 (2022).
6. Lang, X. *et al.* The effects of extreme high indoor temperature on EEG during a low intensity activity. *Building and Environment* 219, 109225 (2022).
7. Huang, D., Wang, X., Liu, J., Li, J. & Tang, W. Virtual reality safety training using deep EEG-net and physiology data. *Vis Comput* 38, 1195–1207 (2022).

8. Ramos, P. M. S., Maior, C. B. S., Moura, M. C. & Lins, I. D. Automatic drowsiness detection for safety-critical operations using ensemble models and EEG signals. *Process Safety and Environmental Protection* 164, 566–581 (2022).

9. Cao, M. *et al.* Virtual intracranial EEG signals reconstructed from MEG with potential for epilepsy surgery. *Nat Commun* 13, 994 (2022).

10. Gunnarsdottir, K. M. *et al.* Source-sink connectivity: a novel interictal EEG marker for seizure localization. *Brain* 145, 3901–3915 (2022).

11. Formaggio, E. *et al.* Oscillatory EEG-TMS Reactivity in Parkinson Disease. *Journal of Clinical Neurophysiology* (2022) doi:10.1097/WNP.0000000000000881.

12. Chowdary, M. K., Anitha, J. & Hemanth, D. J. Emotion Recognition from EEG Signals Using Recurrent Neural Networks. *Electronics* 11, 2387 (2022).

13. Eskine, K. E. Evaluating the three-network theory of creativity: Effects of music listening on resting state EEG. *Psychology of Music* 03057356221116141 (2022) doi:10.1177/03057356221116141.

14. Tuncer, T., Dogan, S., Baygin, M. & Rajendra Acharya, U. Tetromino pattern based accurate EEG emotion classification model. *Artificial Intelligence in Medicine* 123, 102210 (2022).

15. Berger, H. Über das Elektrenkephalogramm des Menschen. *Archiv f. Psychiatrie* 87, 527–570 (1929).

16. Minguillon, J., Perez, E., Lopez-Gordo, M. A., Pelayo, F. & Sanchez-Carrion, M. J. Portable System for Real-Time Detection of Stress Level. *Sensors* 18, 2504 (2018).

17. Ifeachor, E. C. & Jervis, B. W. *Digital Signal Processing: A Practical Approach.* (Pearson Education, 2002).

18. Babiloni, C. *et al.* Brain neural synchronization and functional coupling in Alzheimer's disease as revealed by resting state EEG rhythms. *International Journal of Psychophysiology* 103, 88–102 (2016).

19. Woltering, S., Jung, J., Liu, Z. & Tannock, R. Resting state EEG oscillatory power differences in ADHD college students and their peers. *Behav Brain Funct* 8, 60 (2012).

20. Raichlen, D. A. *et al.* Differences in Resting State Functional Connectivity between Young Adult Endurance Athletes and Healthy Controls. *Frontiers in Human Neuroscience* 10, (2016).

21. McBride, J. C. *et al.* Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease. *Computer Methods and Programs in Biomedicine* 114, 153–163 (2014).

22. Minguillon, J., Lopez-Gordo, M. A. & Pelayo, F. Detection of attention in multi-talker scenarios: A fuzzy approach. *Expert Systems with Applications* 64, 261–268 (2016).

23. Richman, J. S. & Moorman, J. R. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology* 278, H2039–H2049 (2000).

24. Simons, S. & Abásolo, D. Distance-Based Lempel–Ziv Complexity for the Analysis of Electroencephalograms in Patients with Alzheimer's Disease. *Entropy* 19, 129 (2017).

25. Esteller, R., Echauz, J., Tcheng, T., Litt, B. & Pless, B. Line length: an efficient feature for seizure onset detection. in *2001 Conference Proceedings of the 23rd*

Annual International Conference of the IEEE Engineering in Medicine and Biology Society vol. 2 1707–1710 vol.2 (2001).

26. Smits, F. M. *et al.* Electroencephalographic Fractal Dimension in Healthy Ageing and Alzheimer's Disease. *PLoS ONE* 11, e0149587 (2016).
27. Handayani, N., Haryanto, F., Khotimah, S. N., Arif, I. & Taruno, W. P. Coherence and phase synchrony analyses of EEG signals in Mild Cognitive Impairment (MCI): A study of functional brain connectivity. *Polish Journal of Medical Physics and Engineering* 24, 1–9 (2018).
28. Cassani, R., Estarellas, M., San-Martin, R., Fraga, F. J. & Falk, T. H. Systematic Review on Resting-State EEG for Alzheimer's Disease Diagnosis and Progression Assessment. *Disease Markers* vol. 2018 e5174815 https://www.hindawi.com/journals/dm/2018/5174815/ (2018).
29. Stam, C. J., Nolte, G. & Daffertshofer, A. Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum Brain Mapp* 28, 1178–1193 (2007).
30. Watanabe, Y. *et al.* Analysis for Alzheimer's disease using cross-correlation of EEG data. in *2017 10th Biomedical Engineering International Conference (BMEiCON)* 1–5 (2017). doi:10.1109/BMEiCON.2017.8229145.
31. Abásolo, D., Escudero, J., Hornero, R., Gómez, C. & Espino, P. Approximate entropy and auto mutual information analysis of the electroencephalogram in Alzheimer's disease patients. *Med Biol Eng Comput* 46, 1019–1028 (2008).
32. Horvath A. *et al.* EEG and ERP biomarkers of Alzheimer's disease: a critical review. *Frontiers in bioscience (Landmark edition)* 23, 183–220 (2018).

33. Hedges, D. *et al.* P300 Amplitude in Alzheimer's Disease: A Meta-Analysis and Meta-Regression. *Clinical EEG and Neuroscience* (2014) doi:10.1177/1550059414550567.

34. Bobes, M. A. *et al.* ERP generator anomalies in presymptomatic carriers of the Alzheimer's disease E280A PS-1 mutation. *Human Brain Mapping* 31, 247–265 (2010).

35. Chapman, R. M. *et al.* Brain ERP components predict which individuals progress to Alzheimer's disease and which do not. *Neurobiology of Aging* 32, 1742–1755 (2011).

36. van Dinteren, R., Arns, M., Jongsma, M. L. A. & Kessels, R. P. C. Combined frontal and parietal P300 amplitudes indicate compensated cognitive processing across the lifespan. *Frontiers in Aging Neuroscience* 6, 294 (2014).

37. Bulat, M., Karpman, A., Samokhina, A. & Panov, A. Playing a P300-BCI VR game based leads to changes in cognitive function of healthy adults. *bioRxiv* 2020.05.28.118281 (2020) doi:10.1101/2020.05.28.118281.

38. International Society for Clinical Electrophysiology of Vision *et al.* ISCEV standard for clinical visual evoked potentials: (2016 update). *Documenta Ophthalmologica* 133, 1–9 (2016).

39. Beverina, F., Palmas, G., Silvoni, S., Piccione, F. & Giove, S. User adaptive BCIs: SSVEP and P300 based interfaces. 25.

40. Kelly, S. P., Lalor, E. C., Finucane, C., McDarby, G. & Reilly, R. B. Visual Spatial Attention Control in an Independent Brain-Computer Interface. *IEEE Transactions on Biomedical Engineering* 52, 1588–1596 (2005).

41. Brecelj, J. A VEP study of the visual pathway function in compressive lesions of the optic chiasm. Full-field versus half-field stimulation. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 84, 209–218 (1992).

42. Gregori Grgič, R., Calore, E. & de'Sperati, C. Covert enaction at work: Recording the continuous movements of visuospatial attention to visible or imagined targets by means of Steady-State Visual Evoked Potentials (SSVEPs). *Cortex* 74, 31–52 (2016).

43. Sheppard, E. *et al.* Children with a history of atypical febrile seizures show abnormal steady state visual evoked potential brain responses. *Epilepsy & Behavior* 27, 90–94 (2013).

44. Li, Y., Pan, J., Wang, F. & Yu, Z. A Hybrid BCI System Combining P300 and SSVEP and Its Application to Wheelchair Control. *IEEE Transactions on Biomedical Engineering* 60, 3156–3166 (2013).

45. Lopez-Gordo, M. A., Prieto, A., Pelayo, F. & Morillas, C. Customized stimulation enhances performance of independent binary SSVEP-BCIs. *Clinical Neurophysiology* 122, 128–133 (2011).

46. Perols, J. Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms. *AUDITING: A Journal of Practice & Theory* 30, 19–50 (2011).

47. Peña, J. M. *et al.* Object-Based Image Classification of Summer Crops with Machine Learning Methods. *Remote Sensing* 6, 5019–5041 (2014).

48. Scher, S. & Messori, G. Predicting weather forecast uncertainty with machine learning. *Quarterly Journal of the Royal Meteorological Society* 144, 2830–2841 (2018).

49. Padilla, P. *et al.* NMF-SVM Based CAD Tool Applied to Functional Brain Images for the Diagnosis of Alzheimer's Disease. *IEEE Transactions on Medical Imaging* 31, 207–216 (2012).

50. Cura, O. K., Akan, A., Yilmaz, G. C. & Ture, H. S. Detection of Alzheimer's Dementia by Using Signal Decomposition and Machine Learning Methods. *Int J Neural Syst* 32, 2250042 (2022).

51. Mashrur, A., Luo, W., Zaidi, N. A. & Robles-Kelly, A. Machine Learning for Financial Risk Management: A Survey. *IEEE Access* 8, 203203–203223 (2020).

52. Domingos, P. A few useful things to know about machine learning. *Commun. ACM* 55, 78–87 (2012).

53. Singh, A., Saraswat, S. & Faujdar, N. Analyzing Titanic disaster using machine learning algorithms. in *2017 International Conference on Computing, Communication and Automation (ICCCA)* 406–411 (2017). doi:10.1109/CCAA.2017.8229835.

54. Xu, J. A Novel Deep Neural Network based Method for House Price Prediction. in *2021 International Conference of Social Computing and Digital Economy (ICSCDE)* 12–16 (2021). doi:10.1109/ICSCDE54196.2021.00012.

55. Bonaccorso, G. *Machine Learning Algorithms: Popular algorithms for data science and machine learning, 2nd Edition.* (Packt Publishing Ltd, 2018).

56. Indra, S. T., Wikarsa, L. & Turang, R. Using logistic regression method to classify tweets into the selected topics. in *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* 385–390 (2016). doi:10.1109/ICACSIS.2016.7872727.

57. Yan, J. & Lee, J. Degradation Assessment and Fault Modes Classification Using Logistic Regression. *Journal of Manufacturing Science and Engineering* 127, 912–914 (2004).

58. Alkan, A., Koklukaya, E. & Subasi, A. Automatic seizure detection in EEG using logistic regression and artificial neural network. *Journal of Neuroscience Methods* 148, 167–176 (2005).

59. Albright, J. Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 5, 483–491 (2019).

60. Khosrowabadi, R., Quek, C., Ang, K. K., Tung, S. W. & Heijnen, M. A Brain-Computer Interface for classifying EEG correlates of chronic mental stress. in *The 2011 International Joint Conference on Neural Networks* 757–762 (2011). doi:10.1109/IJCNN.2011.6033297.

61. Komaris, D.-S. *et al.* Predicting Three-Dimensional Ground Reaction Forces in Running by Using Artificial Neural Networks and Lower Body Kinematics. *IEEE Access* 7, 156779–156786 (2019).

62. Lotfan, S., Shahyad, S., Khosrowabadi, R., Mohammadi, A. & Hatef, B. Support vector machine classification of brain states exposed to social stress test using EEG-based brain network measures. *Biocybernetics and Biomedical Engineering* 39, 199–213 (2019).

63. Segovia, F., Górriz, J. M., Ramírez, J., Salas-González, D. & Álvarez, I. Early diagnosis of Alzheimer's disease based on Partial Least Squares and Support Vector Machine. *Expert Systems with Applications* 40, 677–683 (2013).

64. Khedher, L. *et al.* Independent Component Analysis-Support Vector Machine-Based Computer-Aided Diagnosis System for Alzheimer's with Visual Support. *Int. J. Neur. Syst.* 27, 1650050 (2017).

65. Kulkarni, N. N. & Bairagi, V. K. Extracting Salient Features for EEG-based Diagnosis of Alzheimer's Disease Using Support Vector Machine Classifier. *IETE Journal of Research* 63, 11–22 (2017).

66. Liu, Y.-H. *et al.* Emotion Recognition from Single-Trial EEG Based on Kernel Fisher's Emotion Pattern and Imbalanced Quasiconformal Kernel Support Vector Machine. *Sensors* 14, 13361–13388 (2014).

67. Martínez-Almeida Nistal, I., Lampreave Acebes, P., Martínez-de-la-Casa, J. M. & Sánchez-González, P. Validation of virtual reality system based on eye-tracking technologies to support clinical assessment of glaucoma. *Eur J Ophthalmol* 31, 3080–3086 (2021).

68. Zhao, C. H., Zhang, B. L., He, J. & Lian, J. Recognition of driving postures by contourlet transform and random forests. *IET Intelligent Transport Systems* 6, 161–168 (2012).

69. Benbelkacem, S. & Atmani, B. Random Forests for Diabetes Diagnosis. in *2019 International Conference on Computer and Information Sciences (ICCIS)* 1–4 (2019). doi:10.1109/ICCISci.2019.8716405.

70. Dauwan, M. *et al.* Random forest to differentiate dementia with Lewy bodies from Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 4, 99–106 (2016).

71. Patterson, C. *World Alzheimer report 2018.* https://apo.org.au/node/260056 (2018).

72. Arevalo-Rodriguez, I. *et al.* Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev* 7, CD010783 (2021).

73. Nasreddine, Z. S. *et al.* The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 53, 695–699 (2005).

74. Holtzman, D. M. CSF biomarkers for Alzheimer's disease: Current utility and potential future use. *Neurobiol Aging* 32, S4–S9 (2011).

75. Perez-Valero, E., Lopez-Gordo, M. A., Morillas, C., Pelayo, F. & Vaquero-Blasco, M. A. A Review of Automated Techniques for Assisting the Early Detection of Alzheimer's Disease with a Focus on EEG. *Journal of Alzheimer's disease: JAD* (2021) doi:10.3233/JAD-201455.

76. Minguillon, J., Lopez-Gordo, M. A. & Pelayo, F. Trends in EEG-BCI for daily-life: Requirements for artifact removal. *Biomedical Signal Processing and Control* 31, 407–418 (2017).

77. Jeong, J. EEG dynamics in patients with Alzheimer's disease. *Clinical Neurophysiology* 115, 1490–1505 (2004).

78. Trambaiolli, L. R. *et al.* Improving Alzheimer's disease diagnosis with machine learning techniques. *Clin EEG Neurosci* 42, 160–165 (2011).

79. McBride, J. *et al.* Scalp EEG signal reconstruction for detection of mild cognitive impairment and early Alzheimer's disease. in *2013 Biomedical Sciences and Engineering Conference (BSEC)* 1–4 (2013). doi:10.1109/BSEC.2013.6618497.

80. Aghajani, H., Zahedi, E., Jalili, M., Keikhosravi, A. & Vahdat, B. V. Diagnosis of Early Alzheimer's Disease Based on EEG Source Localization and a Standardized Realistic Head Model. *IEEE Journal of Biomedical and Health Informatics* 17, 1039–1045 (2013).

81. Fiscon, G. *et al.* Alzheimer's disease patients classification through EEG signals processing. in *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)* 105–112 (2014). doi:10.1109/CIDM.2014.7008655.

82. Wang, R. *et al.* Power spectral density and coherence analysis of Alzheimer's EEG. *Cogn Neurodyn* 9, 291–304 (2015).

83. Morabito, F. C. *et al.* Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings. in *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)* 1–6 (2016). doi:10.1109/RTSI.2016.7740576.

84. Cassani, R. *et al.* Towards automated electroencephalography-based Alzheimer's disease diagnosis using portable low-density devices. *Biomedical Signal Processing and Control* 33, 261–271 (2017).

85. Trambaiolli, L. R., Spolaôr, N., Lorena, A. C., Anghinah, R. & Sato, J. R. Feature selection before EEG classification supports the diagnosis of Alzheimer's disease. *Clinical Neurophysiology* 128, 2058–2067 (2017).

86. Ruiz-Gómez, S. J. *et al.* Automated Multiclass Classification of Spontaneous EEG Activity in Alzheimer's Disease and Mild Cognitive Impairment. *Entropy* 20, 35 (2018).

87. Fiscon, G. *et al.* Combining EEG signal processing with supervised methods for Alzheimer's patients classification. *BMC Med Inform Decis Mak* 18, 35 (2018).

88. Mazaheri, A. *et al.* EEG oscillations during word processing predict MCI conversion to Alzheimer's disease. *NeuroImage: Clinical* 17, 188–197 (2018).

89. Khatun, S., Morshed, B. I. & Bidelman, G. M. A Single-Channel EEG-Based Approach to Detect Mild Cognitive Impairment via Speech-Evoked Brain Responses. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27, 1063–1070 (2019).

90. Durongbhan, P. *et al.* A Dementia Classification Framework Using Frequency and Time-Frequency Features Based on EEG Signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27, 826–835 (2019).

91. Ieracitano, C., Mammone, N., Hussain, A. & Morabito, F. C. A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. *Neural Networks* 123, 176–190 (2020).

92. Podgorelec, V. Analyzing EEG Signals with Machine Learning for Diagnosing Alzheimer's Disease. *Elektron. Elektrotech.* 18, 61–64 (2012).

93. Poil, S.-S. *et al.* Integrative EEG biomarkers predict progression to Alzheimer's disease at the MCI stage. *Front Aging Neurosci* 5, (2013).

94. Buscema, M. *et al.* An improved I-FAST system for the diagnosis of Alzheimer's disease from unprocessed electroencephalograms by using robust invariant features. *Artificial Intelligence in Medicine* 64, 59–74 (2015).

95. Houmani, N. *et al.* Diagnosis of Alzheimer's disease with Electroencephalography in a differential framework. *PLoS ONE* 13, e0193607 (2018).

96. Amezquita-Sanchez, J. P., Mammone, N., Morabito, F. C., Marino, S. & Adeli, H. A novel methodology for automated differential diagnosis of mild cognitive impairment and the Alzheimer's disease using EEG signals. *Journal of Neuroscience Methods* 322, 88–95 (2019).

97. Minguillon, J., Lopez-Gordo, M. A., Renedo-Criado, D. A., Sanchez-Carrion, M. J. & Pelayo, F. Blue lighting accelerates post-stress relaxation: Results of a preliminary study. *PLOS ONE* 12, e0186399 (2017).

98. Vaquero-Blasco, M. A., Perez-Valero, E., Lopez-Gordo, M. A. & Morillas, C. Virtual Reality as a Portable Alternative to Chromotherapy Rooms for Stress Relief: A Preliminary Study. *Sensors* 20, 6211 (2020).

99. Perez-Valero, E., Lopez-Gordo, M. A. & Vaquero-Blasco, M. A. An attention-driven videogame based on steady-state motion visual evoked potentials. *Expert Systems* n/a, e12682 (2021).

100. Dedovic, K. *et al.* The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *J Psychiatry Neurosci* 30, 319–325 (2005).

101. Vaquero-Blasco, M. A., Perez-Valero, E., Morillas, C. & Lopez-Gordo, M. A. Virtual Reality Customized 360-Degree Experiences for Stress Relief. *Sensors* 21, 2219 (2021).

102. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830 (2011).

103. Perez-Valero, E., Vaquero-Blasco, M. A., Lopez-Gordo, M. A. & Morillas, C. Quantitative Assessment of Stress Through EEG During a Virtual Reality Stress-Relax Session. *Front. Comput. Neurosci.* 15, (2021).

104. Lendínez González, A., Carnero Pardo, C., Iribar Ibabe, C., Zunzunegui Pastor, M. V. & González Maldonado, R. Análisis de los ítems que componen la 'Batería

Abreviada Granada de Evaluación Neuropsicológica'. *Geriátrika (Madr.)* 141–154 (2000).

105. Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F. & Gramfort, A. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage* 159, 417–429 (2017).

106. Echtioui, A. *et al.* Automated EEG Artifact Detection Using Independent Component Analysis. in *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* 1–5 (2020). doi:10.1109/ATSIP49331.2020.9231574.

107. van Deursen, J. A., Vuurman, E. F. P. M., van Kranen-Mastenbroek, V. H. J. M., Verhey, F. R. J. & Riedel, W. J. 40-Hz steady state response in Alzheimer's disease and mild cognitive impairment. *Neurobiology of Aging* 32, 24–30 (2011).

108. van Deursen, J. A., Vuurman, E. F. P. M., Verhey, F. R. J., van Kranen-Mastenbroek, V. H. J. M. & Riedel, W. J. Increased EEG gamma band activity in Alzheimer's disease and mild cognitive impairment. *J Neural Transm* 115, 1301–1311 (2008).

109. Olazarán, J. *et al.* Aplicación práctica de los test cognitivos breves. *Neurología* 31, 183–194 (2016).

110. Carnero-Pardo, C. *et al.* Utilidad diagnóstica del Test de las Fotos (Fototest) en deterioro cognitivo y demencia. *Neurología* 22, 860–869 (2007).

111. Carnero-Pardo, C., Sáez-Zea, C., Montiel-Navarro, L., Feria-Vilar, I. & Gurpegui, M. Estudio normativo y de fiabilidad del fototest. *Neurología* 26, 20–25 (2011).

112. Carnero-Pardo, C. *et al.* Utilidad diagnóstica y validez predictiva del uso conjunto de Fototest y Mini-Cog en deterioro cognitivo. *Neurologia* doi:10.1016/j.nrl.2021.01.017.

113. Barrios, L. J. *et al.* Estado del Arte en Neurotecnologías para la Asistencia y la Rehabilitación en España: Tecnologías Auxiliares, Trasferencia Tecnológica y Aplicación Clínica. *Revista Iberoamericana de Automática e Informática Industrial RIAI* 14, 355–361 (2017).

114. Peters, F. *et al.* Abnormal temporal dynamics of visual attention in Alzheimer's disease and in dementia with Lewy bodies. *Neurobiol Aging* 33, 1012.e1–10 (2012).

115. Kavcic, V. & Duffy, C. J. Attentional dynamics and visual perception: mechanisms of spatial disorientation in Alzheimer's disease. *Brain* 126, 1173–1181 (2003).

116. Bridges, D., Pitiot, A., MacAskill, M. R. & Peirce, J. W. *The timing mega-study: comparing a range of experiment generators, both lab-based and online.* https://osf.io/d6nu5 (2020) doi:10.31234/osf.io/d6nu5.

117. Lutz, A., Greischar, L. L., Rawlings, N. B., Ricard, M. & Davidson, R. J. Long-term meditators self-induce high-amplitude gamma synchrony during mental practice. *PNAS* 101, 16369–16373 (2004).

118. Steinhubl, S. R., Muse, E. D. & Topol, E. J. Can Mobile Health Technologies Transform Health Care? *JAMA* 310, 2395–2396 (2013).

119. Mora-Gonzalez, J. *et al.* Physical fitness and brain source localization during a working memory task in children with overweight/obesity: The ActiveBrains project. *Developmental Science* 24, e13048 (2021).

120. Cadenas-Sánchez, C. *et al.* An exercise-based randomized controlled trial on brain, cognition, physical health and mental health in overweight/obese children (ActiveBrains project): Rationale, design and methods. *Contemporary Clinical Trials* 47, 315–324 (2016).

121. Mora-Gonzalez, J. *et al.* Fitness, physical activity, working memory, and neuroelectric activity in children with overweight/obesity. *Scandinavian Journal of Medicine & Science in Sports* 29, 1352–1363 (2019).

122. Koolen, N. *et al.* Line length as a robust method to detect high-activity events: automated burst detection in premature EEG recordings. *Clin Neurophysiol* 125, 1985–1994 (2014).

123. Hogan, M. J. *et al.* The effects of cardiorespiratory fitness and acute aerobic exercise on executive functioning and EEG entropy in adolescents. *Frontiers in Human Neuroscience* 9, (2015).

124. Hogan, M. J. *et al.* Electrophysiological entropy in younger adults, older controls and older cognitively declined adults. *Brain Research* 1445, 1–10 (2012).

Appendix A

A review of automated techniques for assisting the early detection of Alzheimer´s disease with a focus on EEG

Journal of Alzheimer's disease

A review of automated techniques for assisting the early detection of Alzheimer´s disease with a focus on EEG

Eduardo Perez-Valero[a,b], Miguel A. Lopez-Gordo[c,d], Christian Morillas[a,b], Francisco Pelayo[a,b], and Miguel A. Vaquero-Blasco[a,c]

[a]Research Centre for Information and Communications Technologies (CITIC), University of Granada, Spain.

(C/ Periodista Rafael Gómez Montero, 2, 18014 Granada)

[b]Department of Computer Architecture and Technology, University of Granada, Spain.

(C/ Periodista Daniel Saucedo Aranda, s/n, 18014 Granada)

[c]Department of Signal Theory, Telematics and Communications, University of Granada, Spain.

(C/ Periodista Daniel Saucedo Aranda, s/n, 18014 Granada)

[d]Nicolo Association, Churriana de la Vega, Spain.

***Correspondence:**

Corresponding author: malg@ugr.es, Tel.: +34-958-249-721

Keywords: Alzheimer's disease, early diagnosis, Machine Learning, Electroencephalography.

Abstract — In this paper, we review state-of-the-art approaches that apply Signal Processing (SP) and Machine Learning (ML) to automate the detection of Alzheimer's disease (AD) and its prodromal stages. In the first part of the document, we describe the economic and social implications of the disease, traditional diagnosis techniques, and the fundaments of automated AD detection. Then, we present Electroencephalography (EEG) as an appropriate alternative for the early detection of AD, owing to its reduced cost, portability, and non-invasiveness. We also describe the main time and frequency domain EEG features that are employed in AD detection. Subsequently, we examine some of the main studies of the last decade that aim to provide an automatic detection of AD and its previous stages by means of SP and ML. In these studies, brain data was acquired using multiple medical techniques such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and EEG. The main

aspects of each approach, namely feature extraction, classification model, validation approach, and performance metrics are compiled and discussed. Lastly, a set of conclusions and recommendations for future research on AD automatic detection are drawn in the final section of the paper.

INTRODUCTION

AD is an irreversible neurodegenerative disorder that disrupts cognitive functions and eventually leads to death [1]. More than one hundred years after the first case of AD was reported, its etiology remains unknown and a definitive diagnosis can only be established upon analysis of the brain at autopsy [2].

According to the latest World Alzheimer Report, published in 2018, more than 50 million people live with dementia [3], and 60% of those cases correspond to AD. The widespread prevalence of AD and its terminal prognosis represent a heavy economic burden for healthcare and social systems. Indeed, AD is the third most expensive disorder in the United States after cancer and coronary heart disease [4]. However, this cost is not equally distributed worldwide, as in 2010 it was estimated that low, middle, and high income countries accounted for 1%, 10%, and 89% of the global costs, but 14%, 40%, and 46% of the prevalence [5].

Conventional techniques to detect AD are costly and distressing. Nonetheless, early accurate detection is crucial to control the progression of the disease and postpone intellectual decline [6–8]. For this reason, automated and affordable diagnosis techniques have become an important subject of research.

The aim of this work is to provide a narrative review of state-of-the-art studies that combined Signal Processing (SP) and Machine Learning (ML) for the early detection of AD. Furthermore, we present EEG as a suitable technique for AD detection, since it is generally considered a fast, inexpensive, and accessible technique to gather brain data. The rest of the document is organized as follows: first, we briefly present traditional AD detection techniques. Then, we describe the role of SP and ML to assist AD diagnosis. Subsequently, we focus on the potential of EEG to support AD detection. Next, we discuss studies that aimed to detect AD and its prodromal stages through SP and ML. Finally, we draw conclusions and we provide a set of recommendations for future research on AD automatic detection.

Traditional diagnosis

AD begins with a period devoid of symptoms whilst the pathological process advances [9]. Then, minor memory losses emerge while cognitive functions remain unaffected. This stage is known as Mild Cognitive Impairment

(MCI), and is considered a transitional state between normal aging and AD [10].

The disruption of the healthy functioning of the brain present in AD is caused by a continuous neurodegenerative process defined by two main hallmarks, amyloid plaques and neurofibrillary tangles [11,12]. The former are deposits of proteins that lose their standard structure and aggregate into plaques around brain cells. The latter are thickened fibrils that surround the nucleus of neurons. In Fig. 1 we have represented the evolution of these hallmarks along with the decay of neuronal integrity. To postpone this decay, classic diagnosis methods aim to detect AD impairments through a diverse set of psychological and medical techniques. Below, we briefly outline these techniques.

- *Cognitive tests.* These assessments evaluate the performance of the patient in different cognitive areas, such as memory, orientation, language, and mood. For instance, general cognitive status is mainly evaluated using the Mini Mental State Examination (MMSE) [13] and the Montreal Cognitive Assessment (MoCA) [14]. Although these tests are extensively applied, limitations such as lack of sensitivity and high variability have been reported in literature [15,16].
- *Biomarkers.* In the case of AD, cerebrospinal fluid (CSF) is the most suitable alternative to extract biomarkers from [17]. CSF is a fluid located in the brain and the spinal cord that is extracted via lumbar puncture, a costly, scarce, and painful clinical technique [18]. This procedure evidences the necessity to find approaches that are more comfortable and accessible from the point of view of the patient. Refer to [19] for an in-depth study about AD biomarkers.
- *Medical imaging.* Techniques in this field are applied to render static or functional images of the internal organs of the body. This is achieved through diverse procedures that include strong radio pulses in the case of Magnetic Resonance Imaging (MRI), a radioactive tracer injected into the bloodstream in Single Photon Emission Computed Tomography (SPECT), or a radiotracer that can be swallowed, inhaled or injected in the case of Positron Emission Tomography (PET).
- *Neurophysiology.* This medical specialty studies the functioning of the nervous system. The two main AD detection techniques in this field are EEG and Magnetoencephalography (MEG). EEG monitors the electrical activity in the brain

cortex via electrodes placed on the scalp [20,21]. Alternatively, MEG registers the magnetic fields generated by small currents during postsynaptic potentials using a magnetometer covering the skull [22]. Compared to medical imaging techniques, EEG and MEG provide higher temporal resolution, and conversely to EEG, MEG is not affected by distortion, conductivity changes, or attenuation due to human tissues, as it registers primary electrical activity. However, MEG is a more expensive and less accessible technique compared to EEG.

Table 1 summarizes the main aspects of the detection techniques that we described in this subsection. The results yielded by these techniques rely on time consuming tasks, such as semi quantitative analysis and visual inspection [23]. As a consequence, many studies have taken advantage of SP and ML to develop intelligent models to detect AD and its prodromal stages from clinical trials. These models are appropriate to assist clinicians in the challenging task of early AD detection.

Signal Processing and Machine Learning for AD

Recently, AD detection has been approached as a supervised classification problem through SP and ML. In this kind of task, a model is trained to discriminate brain data from different cohorts (usually controls, MCI, and AD). Data may be presented as a cluster of images (MRI, SPECT, PET, etc.), signals (MEG and EEG), or biomarkers. Before model training, features are extracted from brain data through a process known as feature extraction. Ideally, these features should be independent and discriminative to facilitate classification. Depending upon the acquisition technique, different SP techniques are applied to extract features from brain data. In the case of medical imaging techniques, images are typically segmented into voxels and features are extracted from them [24,25]. With respect to neurophysiology techniques, features like complexity, coherence, or spectral power are usually extracted from the time and frequency domains. Once feature extraction is completed, training is the subsequent stage in model development. In this process, the classifier learns patterns from the extracted features to obtain generalization capabilities to discriminate unseen data. After training, the performance of the classifier is evaluated through multiple metrics such as accuracy, sensibility, and sensitivity. In the context of AD, multiple classifiers have been successfully applied to discriminate AD data collected through several acquisition techniques. For instance, support vector machines (SVM) [24] and convolutional neural networks (CNN) [26] have been used to classify subjects into healthy controls, MCI, and AD patients from MRI scans. Similarly, principal component analysis (PCA), and artificial neural networks (ANN) [23,27] have been utilized for the discrimination of those cohorts from SPECT data. Additionally, other studies have addressed the classification of AD cohorts from MEG [28,29] and EEG data [30–32], and the suitability of these techniques to assist clinical diagnosis has been already demonstrated [33].

In the following section we disclose the most widespread EEG features and processing techniques for AD detection.

EEG PROCESSING FOR ADDETECTION

The suitability of EEG for AD detection is based on its reduced cost, notable accessibility, and non-invasiveness compared to other techniques. Investigations in this field aim to obtain features that correlate with the effects that AD creates in the electrical activity of the brain. These effects include EEG slowing, loss of synchronization, and complexity reduction. For a comprehensive review of the EEG abnormalities caused by AD, refer to [34]. To extract these features, EEG is processed in the time and the frequency domains. Moreover, authors like Ieracitano et al. have stressed the relevance of combining multi-domain features to enhance the discrimination performance of the classifiers [35]. Other authors, like Cassani et al., have evidenced the substantial impact of artifact removal algorithms in EEG-based AD diagnosis. Thereby, in [36] the authors proved that wavelet enhanced independent component analysis (wICA) outperformed other automated artifact removal algorithms in AD early detection. Over the next two subsections we report some of the most frequent EEG analysis techniques used in AD detection.

Time domain analysis

In this subsection we describe the most extended EEG time domain features for AD detection.

ERP. Event related potentials (ERP) are electric potentials elicited in the brain as a response to an auditory or visual stimulus. ERP have been utilized to study the morphological differences between AD cohorts, since the potentials in MCI and AD exhibit higher latency and smaller amplitude compared to HC [37]. For this purpose, multiple studies have assessed different cognitive functions like memory [38,39], oddball recognition [32,40], and perception [41] for the detection of patients at different stages of AD.

Complexity. Brain activity in MCI and AD patients has lower complexity and irregularity compared to healthy subjects [42,43]. Multiple metrics have been used in literature to assess EEG complexity. Most of them are computed from nonlinear analysis, and include Lempel-Ziv complexity, fuzzy entropy, Tsallis entropy, Higuchi fractal dimension, and mutual information [44–46].

Frequency domain analysis

The frequency domain represents the principal source of EEG features for AD detection. Frequency analysis is based on power spectral density (PSD), a measure of the power content of a signal with respect to the frequency. In [47], Cassani et al. recommend the use of PSD features as baseline. In this subsection we report the most widespread EEG spectral features for AD research.

Coherence. This metric evaluates the synchrony between two signals. More formally, coherence refers to the normalized covariance of two signals in the spectral domain. In the case of EEG, this metric is usually estimated between pairs of electrodes, and quantifies the functional connectivity of the brain. A high coherence represents high synchrony, what indicates strong functional connectivity between brain areas [30]. In literature, authors have reported an increase of coherence in delta and theta rhythms, and a decrease in alpha and beta rhythms in MCI and AD [30,48]. Other measures of synchrony that have been successfully used to discriminate MCI subjects include the correlation coefficient, Granger causality, and stochastic event synchrony [49].

Relative power. This metric quantifies the proportion of power that is contained in an EEG sub band, compared to the total power of the signal [50]. Studies on relative power have revealed a slowing of brain electrical activity in MCI and AD patients compared to healthy aged [51]. This means that the relative power in the fast rhythms (alpha and beta) decreases, while the relative power in the slow rhythms (delta and theta) increases [7,37,52,53]. Rhythms are typically utilized to analyze the EEG in a particular sub band via filtering, since the activity in each of the different frequency bands is particularly relevant to certain cognitive mechanisms [54–58].

Fourier transform. This transform is used to decompose a time signal into a sum of sines and cosines of different frequencies and infinite length. The Fourier transform has been widely applied in literature [59,60]. Nonetheless, since EEG is a non-stationary signal, the frequency content of brain activity changes continuously. This implies that the spectral content of the EEG cannot be tracked by the Fourier

Transform, as the sines and cosines used in this transform are infinite length. Conversely, the wavelet transform (WT) represents a suitable alternative to address this issue. This transform decomposes a signal into the combination of functions (wavelets) of finite length and different frequencies. Thus, this technique allows the assessment of the spectral content of the EEG over time. This kind of analysis is also generalized in AD literature [35,59,61].

Amplitude modulation. This methodology combines temporal and frequency analysis. To do so, the EEG is filtered in the sub bands of interest and the temporal envelope of each signal is estimated. Then, the temporal dynamics of the envelopes are quantified by performing a second filtering into the modulation sub bands. Amplitude modulation represents an alternative to traditional nonlinear and frequency analysis. This analysis has been validated in the discrimination of AD and healthy cohorts, as well as the characterization of AD severity [62,63]. Additionally, other studies have presented the changes in the envelope of the slow rhythms as potential indicators for early diagnosis [62].

Lastly, despite the reliability for AD detection of the features outlined in this subsection, authors like Durongbhan et al. have evidenced the potential of time-frequency-based features (such as the continuous WT) to improve the results yielded by frequency-based features [64]. Additionally, other authors have recommended the use of both time and frequency metrics to enrich the feature space [45,65,66].

DISCUSSION ON RECENT APPROACHES

In this section, we discuss some of the main studies of the last decade that aimed to automatically discriminate AD cohorts through SP and ML.

We summarized the main features of these proposals in Table 2. In particular, *Source* refers to the brain data acquisition technique (for MEG and EEG, the number of channels is indicated in brackets). *Cohorts* represents the different groups of subjects that were considered. *Protocol* describes the procedure followed by the subjects during data acquisition. *Features* stands for the features used to develop the classification model. If reported, feature selection method is also indicated in brackets in this column. *Model* refers to the classifier utilized to discriminate the cohorts. *Validation* indicates the model assessment technique. Finally, *Performance* reports the evaluation of the classifier in terms of accuracy, sensitivity, specificity, or area under receiver operating curve (AUC). We did not include other unusual performance metrics in the table to preserve the homogeneity of the report. Formulae for accuracy, sensitivity, and specificity are provided in equations 1-3, where TP, TN, FP, and FN stand for True Positives, True Negatives, False Positives, and False Negatives, respectively. Over the next paragraphs, we compare the works in Table 2 in terms of these features.

$$A\ (accuracy) = (TP + TN)/(TP + FP + TN + FN) \quad (1)$$

$$R\ (sensitivity) = TP/(TP + FN) \quad (2)$$

$$S\ (specificity) = TN/(TN + FP) \quad (3)$$

The studies that we compiled in Table 2 assessed AD detection through predominant clinical trials, including medical imaging (MRI, SPECT, and PET), and neurophysiology techniques (MEG and EEG). Regarding medical imaging, MRI-based studies are more extended in literature than those based on SPECT and PET. This may be due to the invasiveness of the latter techniques compared to MRI. Likewise, in the case of neurophysiology approaches, studies using EEG are more widespread than those using MEG. In this case, the lower cost and higher accessibility of EEG may be the reason behind.

With respect to cohorts, the most common classification problem among the works in Table 2 considered three cohorts: HC, MCI, and AD. This problem has been typically splitted into two binary classification problems: HC vs MCI and HC vs AD [24,67,68]. However, it has also been addressed as a multiclass problem [26,69]. Less extended approaches considered MCI vs AD [70] and SCI (subjective cognitive impairment) vs AD [71]. Alternatively, in [72] and [73] the authors trained a model based on data from HC and AD subjects, and then they evaluated it for sMCI (stable MCI) and pMCI (progressive MCI) discrimination. This type of approach is explained in detail in [74]. In this context, note that we have splitted Table 2 into two parts based on the classification problem addressed by the studies. In Table 2 (a), we have gathered studies that reported at least one of the two main classification problems in AD literature (HC vs MCI and HC vs AD). Alternatively, in Table 2 (b), we have included works that studied the classification of other AD cohorts (indicated in brackets under *Performance* column), and works that approached the discrimination of HC, MCI, and AD as a multiclass classification problem.

With respect to the protocol followed, electrophysiology techniques are more flexible than medical imaging techniques, since the latter require the patient to remain completely still. For instance, different authors have recorded MEG [28] and EEG [40], [77] while the participants completed diverse cognitive tests. Likewise, in [40] the authors studied the differences in audio ERP of healthy aging people and MCI subjects. However, the majority of MEG and EEG studies in Table 2 reported data acquisition during resting state protocol. Through this protocol acquisitions become simpler and less stressing for the participant. For an extensive review about resting state EEG for AD detection, refer to [47].

In the case of electrophysiology studies, electrical setup is an additional important aspect. While most of MEG studies reported in Table 2 used more than 300 channels, all EEG studies except [76] used 32 channels or less. Indeed, the number of channels was reduced below 10 in the case of [35,40–42,64,66]. In this respect, the selection of a particular reference montage in EEG approaches strongly influences the performance of a classifier. In [77], the authors assessed the performance of five different EEG montages in the discrimination of HC vs AD patients from spectral features. The results yielded a better classification performance for the Counterpart Bipolar montage, what evidences the importance of electrode montage selection.

Regarding feature extraction, in MEG and EEG works, the most recurrent features include spectral power, coherence [30,31,42,45,65,76], and connectivity [29,78–80], as it has been proved that neural networks in AD patients tend to be less complex and efficient [81,82]. Conversely, in medical imaging studies, investigators tend to extract information from image voxels. Alternatively, other works in this field have combined multiple techniques to obtain multi-domain features [67,83,84]. In this context, feature selection is appropriate to avoid overfitting, what becomes especially relevant in medical imaging studies, where authors typically extract a high number of features through image processing. Likewise, as evidenced by Trambaiolli et al. in [85], feature selection plays an important role in EEG-based studies. In this work, the authors demonstrated that the right choice of the feature selection method removed up to 82% of the EEG spectral features and increased the accuracy of the classifier more than 17%.

In relation to model selection, SVM was the most recurrent classifier among those studies in Table 2 that addressed a binary discrimination of AD cohorts. In the case of multiclass classification studies, more complex classifiers like ensemble methods and deep networks have been tested with diverse results [26,35,69].

Regarding model validation, the studies gathered in this section used mainly three validation approaches: leave-one-subject-out cross-validation (LOSO CV), k-fold cross-validation (k-fold CV), and leave-one-out cross-validation (LOO CV). In this respect, it is mandatory to prevent the models from using highly correlated data for development and evaluation, as this would introduce a positive bias in the predictions. LOSO CV (covered in [42,45,61,71,85]) overcomes this issue by ensuring that the information from one subject is utilized only in the training or the test set. This is not feasible in LOO CV, as the data from a subject is included in both sets. With regard to k-fold CV, a proper experimental design is required to prevent data leakage. Considering this, we deem LOSO CV as the most straight-forward and clear model validation approach for AD classification.

The last aspect that we considered in Table 2 is performance. Most of the works collected disclosed accuracy, sensitivity, and specificity. Other works only reported accuracy [28,69,80,84]. However, accuracy does not assess the rates of false positives and false negatives, that are particularly important in clinical-related classification problems. In all, performance report is consistent over the works that we collected, as only [28,35,61,67,69,75,80,84,86] did not provide accuracy, sensitivity, and specificity jointly. Moreover, it is important to acknowledge the dependency of metrics like accuracy and AUC with group balance [87], an issue that can be confronted via additional metrics like precision-recall curves (PRC), sensitivity, and specificity. With regard to performance results, the heterogeneity of the studies gathered in Table 2 hinder comparisons, however, accuracy, sensitivity, and specificity are generally much higher than random choice (in most cases above 80% for binary classification problems). The fact that many different classifiers achieved similar performance demonstrates the widespread knowledge that classifiers are generally not as important as the quality of the features. Hence, developing discriminative features that enable the detection of AD in its early stages acquires paramount importance.

In terms of data availability, many of the medical imaging studies gathered in Table 2 reported the use of a public database to obtain brain data. The use of public medical imaging databases like ADNI or OASIS [88] is a common practice in AD literature (see [89] for a detailed report about querying and organization of AD data). On the other hand, in EEG studies, due to the high accessibility of this technique, researchers generally acquire their own data.

To sum up, in general the proposals compiled in this section yielded positive results in terms of AD cohort discrimination. The studies achieved similar performance using different medical techniques, namely MRI, SPECT, PET, MEG, and EEG. In MEG and EEG studies, some cognitive tasks were monitored, but resting state was the most recurrent protocol during data acquisition. Regarding feature extraction, multiple approaches were conducted based on the source of the brain data. With respect to classification, SVM was the most extended classifier for binary classification, and we concluded that LOSO CV is the most suitable validation alternative based on its intrinsic capability to avoid data leakage.

Lastly, after discussing the works in Table 2, it is notable that the performance of AD classifiers based on EEG data can equal or even outperform the performance of approaches based on medical imaging. Moreover, imaging-based studies rely on image processing techniques for feature extraction, that are computationally demandant compared to EEG signal processing techniques. In addition, EEG time resolution is higher relative to medical imaging, what enables investigators to study dynamical changes in brain activity. EEG is also non-invasive and far more accessible and inexpensive than medical imaging and MEG. These aspects are crucial, since as stated in the introduction of this work, AD prevalence increases every year, especially in developing countries that lack awareness and technological means. Therefore, a reliable, inexpensive, and non-invasive AD diagnosis technique is required. In this context, EEG represents an appropriate technique to support clinicians in AD detection.

CONCLUSION

In this survey, we discussed state-of-the-art approaches to assist the diagnosis of AD. The works compiled in this paper used diverse medical techniques to acquire brain activity, and feature extraction to derive features from it. In these works, authors trained automated models to discriminate between AD cohorts. The three main cohorts considered in the studies that we reviewed were HC, MCI, and AD subjects. With regard to the classifiers, even though all the approaches that were discussed could discriminate between AD groups with reliable performance, SVM was the most extended alternative. However, the heterogeneity of the disease, along with the discrepancies in the validation approach among the studies, make classification results difficult to compare. In respect to model validation, we consider that LOSO CV is the most suitable approach, since it does not require a particular experimental design to avoid data leakage. Regarding the suitability for assisting the early diagnosis of AD, we believe EEG is the most appropriate technique, since it is unexpensive, highly available, and non-invasive. Its accessibility facilitates the performance of longitudinal studies to monitor the evolution of AD, and it also allows the acquisition of brain data during the performance of cognitive tasks. Regarding electrode setup, authors should consider comparing different EEG montages, since the right choice of the electrode configuration could enhance the discrimination performance of the classifiers [77]. Despite resting state was the most frequent protocol for EEG recording, different cognitive tasks should be studied, as they may provide insights on how AD affects certain cognitive areas. Since AD is expected to affect a large part of the population in the following years, especially in low-income countries that lack medical resources, EEG represents a suitable technique to assist clinical professionals on early diagnosis. For this purpose, we believe that future studies should prioritize the study of new features that support the early detection of AD.

Acknowledgments

This work was supported by project PGC2018-098813- B-C31 (Spanish "Ministerio de Ciencia, Innovación y Universidades"'), by European Regional Development Funds (ERDF) and by the Nicolo Association for the R&D in Neurotechnologies for disability.

Conflict of interest statement

The authors have no conflict of interest to report.

References

[1] Cummings JL, Cole G (2002) Alzheimer Disease. *JAMA* **287**, 2335–2338.

[2] Castellani RJ, Rolston RK, Smith MA (2010) Alzheimer Disease. *Dis Mon* **56**, 484–546.

[3] World Alzheimer Report 2018 - The state of the art of dementia research: New frontiers. *NEW FRONTIERS* 48.

[4] World Alzheimer report 2016: improving healthcare for people living with dementia: coverage, quality and costs now and in the future, Last updated September 2016, Accessed on September 2016.

[5] Wimo A, Jönsson L, Bond J, Prince M, Winblad B (2013) The worldwide economic impact of dementia 2010. *Alzheimer's & Dementia* **9**, 1-11.e3.

[6] Riemenschneider M, Lautenschlager N, Wagenpfeil S, Diehl J, Drzezga A, Kurz A (2002) Cerebrospinal fluid tau and beta-amyloid 42 proteins identify Alzheimer disease in subjects with mild cognitive impairment. *Arch Neurol* **59**, 1729–1734.

[7] Rodrigues PM, Freitas D, Teixeira JP, Bispo B, Alves D, Garrett C (2018) Electroencephalogram Hybrid Method for Alzheimer Early Detection. *Procedia Computer Science* **138**, 209–214.

[8] Kim D, Kim K (2018) Detection of Early Stage Alzheimer's Disease using EEG Relative Power with Deep Neural Network. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 352–355.

[9] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR, Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster MV, Phelps CH (2011) Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **7**, 280–292.

[10] Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH, Berg L (2001) Mild Cognitive Impairment Represents Early-Stage Alzheimer Disease. *Arch Neurol* **58**, 397–405.

[11] Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT (2011) Neuropathological Alterations in Alzheimer Disease. *Cold Spring Harb Perspect Med* **1**, a006189.

[12] Perl DP (2010) Neuropathology of Alzheimer's Disease. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine* **77**, 32–42.

[13] Minati L, Edginton T, Grazia Bruzzone M, Giaccone G (2009) Reviews: Current Concepts in Alzheimer's Disease: A Multidisciplinary Review. *Am J Alzheimers Dis Other Demen* **24**, 95–121.

[14] Freitas S, Simões MR, Alves L, Santana I (2013) Montreal Cognitive Assessment: Validation Study for Mild Cognitive Impairment and Alzheimer Disease. *Alzheimer Disease & Associated Disorders* **27**, 37–43.

[15] Kuslansky G, Katz M, Verghese J, Hall CB, Lapuerta P, LaRuffa G, Lipton RB (2004) Detecting dementia with the Hopkins Verbal Learning Test and the Mini-Mental State Examination. *Archives of Clinical Neuropsychology* 16.

[16] Mendiondo MS, Ashford JW, Kryscio RJ, Schmitt FA (2000) Modelling Mini Mental State Examination changes in Alzheimer's disease. *Statistics in Medicine* **19**, 1607–1616.

[17] Perrin RJ, Fagan AM, Holtzman DM (2009) Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. *Nature* **461**, 916–922.

[18] The CSF tap test in normal pressure hydrocephalus: evaluation time, reliability and the influence of pain - Virhammar - 2012 - European Journal of Neurology - Wiley Online Library.

[19] Lovestone S, Francis P, Kloszewska I, Mecocci P, Simmons A, Soininen H, Spenger C, Tsolaki M, Vellas B, Wahlund L-O, Ward M, AddNeuroMed Consortium (2009) AddNeuroMed--the European collaboration for the discovery of novel biomarkers for Alzheimer's disease. *Ann N Y Acad Sci* **1180**, 36–46.

[20] Berger H (1929) Über das Elektrenkephalogramm des Menschen. *Archiv f Psychiatrie* **87**, 527–570.

[21] Kirschstein T, Köhling R (2009) What is the Source of the EEG? *Clin EEG Neurosci* **40**, 146–149.

[22] Sutherling WW (2012) Magnetoencephalography. In *Aminoff's Electrodiagnosis in Clinical Neurology* Elsevier, pp. 219–229.

[23] López M, Ramírez J, Górriz JM, Álvarez I, Salas-Gonzalez D, Segovia F, Chaves R, Padilla P, Gómez-Río M (2011) Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease. *Neurocomputing* **74**, 1260–1271.

[24] Khedher L, Ramírez J, Górriz JM, Brahim A, Segovia F (2015) Early diagnosis of Alzheimer′s disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing* **151**, 139–150.

[25] Sørensen L, Igel C, Pai A, Balas I, Anker C, Lillholm M, Nielsen M (2016) Differential diagnosis of mild cognitive impairment and Alzheimer's disease usingstructural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry. *NeuroImage: Clinical* **13**,.

[26] Farooq A, Anwar S, Awais M, Alnowami M (2017) Artificial intelligence based smart diagnosis of alzheimer's disease and mild cognitive impairment. In *2017 International Smart Cities Conference (ISC2)*, pp. 1–4.

[27] Illán IA, Górriz JM, López MM, Ramírez J, Salas-Gonzalez D, Segovia F, Chaves R, Puntonet CG (2011) Computer aided diagnosis of Alzheimer's disease using component based SVM. *Applied Soft Computing* **11**, 2376–2382.

[28] Amezquita-Sanchez JP, Adeli A, Adeli H (2016) A new methodology for automated diagnosis of mild cognitive impairment (MCI) using magnetoencephalography (MEG). *Behavioural Brain Research* **305**, 174–180.

[29] Maestú F, Peña J-M, Garcés P, González S, Bajo R, Bagic A, Cuesta P, Funke M, Mäkelä JP, Menasalvas E, Nakamura A, Parkkonen L, López ME, del Pozo F, Sudre G, Zamrini E, Pekkonen E, Henson RN, Becker JT (2015) A multicenter study of the early detection of synaptic dysfunction in Mild Cognitive Impairment using Magnetoencephalography-derived functional connectivity. *NeuroImage: Clinical* **9**, 103–109.

[30] Wang R, Wang J, Yu H, Wei X, Yang C, Deng B (2015) Power spectral density and coherence analysis of Alzheimer's EEG. *Cogn Neurodyn* **9**, 291–304.

[31] Trambaiolli LR, Lorena AC, Fraga FJ, Kanda PAM, Anghinah R, Nitrini R (2011) Improving Alzheimer's disease diagnosis with machine learning techniques. *Clin EEG Neurosci* **42**, 160–165.

[32] Cecchi M, Moore DK, Sadowsky CH, Solomon PR, Doraiswamy PM, Smith CD, Jicha GA, Budson AE, Arnold SE, Fadem KC (2015) A clinical trial to validate event-related potential markers of Alzheimer's disease in outpatient settings. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **1**, 387–394.

[33] Lehmann C, Koenig T, Jelic V, Prichep L, John RE, Wahlund L-O, Dodge Y, Dierks T (2007) Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). *Journal of Neuroscience Methods* **161**, 342–350.

[34] Jeong J (2004) EEG dynamics in patients with Alzheimer's disease. *Clinical Neurophysiology* **115**, 1490–1505.

[35] Ieracitano C, Mammone N, Hussain A, Morabito FC (2020) A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. *Neural Networks* **123**, 176–190.

[36] Cassani R, Falk TH, Fraga FJ, Kanda PAM, Anghinah R (2014) The effects of automated artifact removal algorithms on electroencephalography-based Alzheimer's disease diagnosis. *Front Aging Neurosci* **6**, 55.

[37] Dauwels J, Vialatte F, Cichocki A Diagnosis of Alzheimer's Disease from EEG Signals: Where Are We Standing? 43.

[38] Fraga FJ, Mamani GQ, Johns E, Tavares G, Falk TH, Phillips NA (2018) Early diagnosis of mild cognitive impairment and Alzheimer's with event-related potentials and event-related desynchronization in N-back working memory tasks. *Computer Methods and Programs in Biomedicine* **164**, 1–13.

[39] Missonnier P, Gold G, Fazio-Costa L, Michel J-P, Mulligan R, Michon A, Ibáñez V, Giannakopoulos P (2005)

Early Event-Related Potential Changes During Working Memory Activation Predict Rapid Decline in Mild Cognitive Impairment. *J Gerontol A Biol Sci Med Sci* **60**, 660–666.

[40] Khatun S, Morshed BI, Bidelman GM (2019) A Single-Channel EEG-Based Approach to Detect Mild Cognitive Impairment via Speech-Evoked Brain Responses. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27**, 1063–1070.

[41] Chapman RM, McCrary JW, Gardner MN, Sandoval TC, Guillily MD, Reilly LA, DeGrush E (2011) Brain ERP components predict which individuals progress to Alzheimer's disease and which do not. *Neurobiology of Aging* **32**, 1742–1755.

[42] Cassani R, Falk TH, Fraga FJ, Cecchi M, Moore DK, Anghinah R (2017) Towards automated electroencephalography-based Alzheimer's disease diagnosis using portable low-density devices. *Biomedical Signal Processing and Control* **33**, 261–271.

[43] Fan M, Yang AC, Fuh J-L, Chou C-A (2018) Topological Pattern Recognition of Severe Alzheimer's Disease via Regularized Supervised Learning of EEG Complexity. *Front Neurosci* **12**, 685.

[44] Yang S, Bornot JMS, Wong-Lin K, Prasad G (2019) M/EEG-Based Bio-Markers to Predict the MCI and Alzheimer's Disease: A Review From the ML Perspective. *IEEE Transactions on Biomedical Engineering* **66**, 2924–2935.

[45] Ruiz-Gómez SJ, Gómez C, Poza J, Gutiérrez-Tobal GC, Tola-Arribas MA, Cano M, Hornero R (2018) Automated Multiclass Classification of Spontaneous EEG Activity in Alzheimer's Disease and Mild Cognitive Impairment. *Entropy* **20**, 35.

[46] Al-Nuaimi AHH, Jammeh E, Sun L, Ifeachor E (2018) Complexity Measures for Quantifying Changes in Electroencephalogram in Alzheimer's Disease. *Complexity* **2018**, 1–12.

[47] Disease Markers, Systematic Review on Resting-State EEG for Alzheimer's Disease Diagnosis and Progression Assessment, Last updated October 4, 2018, Accessed on October 4, 2018.

[48] Maestú F, Cuesta P, Hasan O, Fernandéz A, Funke M, Schulz PE (2019) The Importance of the

Validation of M/EEG With Current Biomarkers in Alzheimer's Disease. *Front Hum Neurosci* **13**, 17.

[49] A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG - ScienceDirect.

[50] Woltering S, Jung J, Liu Z, Tannock R (2012) Resting state EEG oscillatory power differences in ADHD college students and their peers. *Behav Brain Funct* **8**, 60.

[51] Benz N, Hatz F, Bousleiman H, Ehrensperger MM, Gschwandtner U, Hardmeier M, Ruegg S, Schindler C, Zimmermann R, Monsch AU, Fuhr P (2014) Slowing of EEG Background Activity in Parkinson's and Alzheimer's Disease with Early Cognitive Dysfunction. *Front Aging Neurosci* **6**,.

[52] Fernández A, Gil Gregorio P, Maestú F (2012) Actividad espontánea electroencefalográfica y magnetoencefalográfica como marcador de la enfermedad de Alzheimer y el deterioro cognitivo leve. *Revista Española de Geriatría y Gerontología* **47**, 27–32.

[53] Moretti DV (2015) Theta and alpha EEG frequency interplay in subjects with mild cognitive impairment: evidence from EEG, MRI, and SPECT brain modifications. *Front Aging Neurosci* **7**,.

[54] Kanda PAM, Oliveira EF, Fraga FJ (2017) EEG epochs with less alpha rhythm improve discrimination of mild Alzheimer's. *Computer Methods and Programs in Biomedicine* **138**, 13–22.

[55] Minguillon J, Lopez-Gordo MA, Pelayo F (2016) Stress Assessment by Prefrontal Relative Gamma. *Frontiers in Computational Neuroscience* **10**,.

[56] Lotfan S, Shahyad S, Khosrowabadi R, Mohammadi A, Hatef B (2019) Support vector machine classification of brain states exposed to social stress test using EEG-based brain network measures. *Biocybernetics and Biomedical Engineering* **39**, 199–213.

[57] Tan H, Wade C, Brown P (2016) Post-Movement Beta Activity in Sensorimotor Cortex Indexes Confidence in the Estimations from Internal Models. *J Neurosci* **36**, 1516–1528.

[58] Bohbot VD, Copara MS, Gotman J, Ekstrom AD (2017) Low-frequency

theta oscillations in the human hippocampus during real-world and virtual navigation. *Nature Communications* **8**, 1–7.

[59] Fiscon G, Weitschek E, Felici G, Bertolazzi P, De Salvo S, Bramanti P, De Cola MC (2014) Alzheimer's disease patients classification through EEG signals processing. In *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pp. 105–112.

[60] Fiscon G, Weitschek E, Cialini A, Felici G, Bertolazzi P, De Salvo S, Bramanti A, Bramanti P, De Cola MC (2018) Combining EEG signal processing with supervised methods for Alzheimer's patients classification. *BMC Med Inform Decis Mak* **18**, 35.

[61] Kulkarni NN, Bairagi VK (2017) Extracting Salient Features for EEG-based Diagnosis of Alzheimer's Disease Using Support Vector Machine Classifier. *IETE Journal of Research* **63**, 11–22.

[62] Fraga FJ, Falk TH, Kanda PAM, Anghinah R (2013) Characterizing Alzheimer's Disease Severity via Resting-Awake EEG Amplitude Modulation Analysis. *PLOS ONE* **8**, e72240.

[63] Trambaiolli LR, Falk TH, Fraga FJ, Anghinah R, Lorena AC (2011) EEG spectro-temporal modulation energy: A new feature for automated diagnosis of Alzheimer's disease. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3828–3831.

[64] Durongbhan P, Zhao Y, Chen L, Zis P, De Marco M, Unwin ZC, Venneri A, He X, Li S, Zhao Y, Blackburn DJ, Sarrigiannis PG (2019) A Dementia Classification Framework Using Frequency and Time-Frequency Features Based on EEG Signals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27**, 826–835.

[65] McBride JC, Zhao X, Munro NB, Smith CD, Jicha GA, Hively L, Broster LS, Schmitt FA, Kryscio RJ, Jiang Y (2014) Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease. *Computer Methods and Programs in Biomedicine* **114**, 153–163.

[66] Podgorelec V (2012) Analyzing EEG Signals with Machine Learning for Diagnosing Alzheimer's Disease. *Elektron Elektrotech* **18**, 61–64.

[67] Munteanu CR, Fernandez-Lozano C, Mato Abad V, Pita Fernández S, Álvarez-Linera J, Hernández-Tamames JA, Pazos A (2015) Classification of mild cognitive impairment and Alzheimer's Disease with machine-learning techniques using 1H Magnetic Resonance Spectroscopy data. *Expert Systems with Applications* **42**, 6205–6214.

[68] Morabito FC, Campolo M, Ieracitano C, Ebadi JM, Bonanno L, Bramanti A, Desalvo S, Mammone N, Bramanti P (2016) Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings. In *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, pp. 1–6.

[69] Nanni L, Lumini A, Zaffonato N (2018) Ensemble based on static classifier selection for automated diagnosis of Mild Cognitive Impairment. *Journal of Neuroscience Methods* **302**, 42–46.

[70] Amezquita-Sanchez JP, Mammone N, Morabito FC, Marino S, Adeli H (2019) A novel methodology for automated differential diagnosis of mild cognitive impairment and the Alzheimer's disease using EEG signals. *Journal of Neuroscience Methods* **322**, 88–95.

[71] Houmani N, Vialatte F, Gallego-Jutglà E, Dreyfus G, Nguyen-Michel V-H, Mariani J, Kinugawa K (2018) Diagnosis of Alzheimer's disease with Electroencephalography in a differential framework. *PLoS ONE* **13**, e0193607.

[72] Tong T, Gao Q, Guerrero R, Ledig C, Chen L, Rueckert D, Initiative ADN (2017) A Novel Grading Biomarker for the Prediction of Conversion From Mild Cognitive Impairment to Alzheimer's Disease. *IEEE Trans Biomed Eng* **64**, 155–165.

[73] Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J (2015) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage* **104**, 398–412.

[74] Cheng B, Liu M, Zhang D, Munsell BC, Shen D (2015) Domain Transfer Learning for MCI Conversion Prediction. *IEEE Trans Biomed Eng* **62**, 1805–1817.

[75] Mazaheri A, Segaert K, Olichney J, Yang J-C, Niu Y-Q, Shapiro K, Bowman H (2018) EEG oscillations during word processing predict MCI conversion to Alzheimer’s disease. *NeuroImage: Clinical* **17**, 188–197.

[76] Aghajani H, Zahedi E, Jalili M, Keikhosravi A, Vahdat BV (2013) Diagnosis of Early Alzheimer’s Disease Based on EEG Source Localization and a Standardized Realistic Head Model. *IEEE Journal of Biomedical and Health Informatics* **17**, 1039–1045.

[77] International Journal of Alzheimer’s Disease, Does EEG Montage Influence Alzheimer’s Disease Electroclinic Diagnosis?, Last updated May 2, 2011, Accessed on May 2, 2011.

[78] Pusil S, López ME, Cuesta P, Bruña R, Pereda E, Maestú F (2019) Hypersynchronization in mild cognitive impairment: the ‘X’ model. *Brain* **142**, 3936–3950.

[79] Pusil S, Dimitriadis SI, López ME, Pereda E, Maestú F (2019) Aberrant MEG multi-frequency phase temporal synchronization predicts conversion from mild cognitive impairment-to-Alzheimer’s disease. *NeuroImage: Clinical* **24**, 101972.

[80] Dimitriadis SI, López ME, Bruña R, Cuesta P, Marcos A, Maestú F, Pereda E (2018) How to Build a Functional Connectomic Biomarker for Mild Cognitive Impairment From Source Reconstructed MEG Resting-State Activity: The Combination of ROI Representation and Connectivity Estimator Matters. *Front Neurosci* **12**,.

[81] Vecchio F, Miraglia F, Marra C, Quaranta D, Vita MG, Bramanti P, Rossini PM (2014) Human Brain Networks in Cognitive Decline: A Graph Theoretical Analysis of Cortical Connectivity from EEG Data. *JAD* **41**, 113–127.

[82] Yu M, Gouw AA, Hillebrand A, Tijms BM, Stam CJ, van Straaten ECW, Pijnenburg YAL (2016) Different functional connectivity and network topology in behavioral variant of frontotemporal dementia and

Alzheimer's disease: an EEG study. *Neurobiology of Aging* **42**, 150–162.

[83] Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D (2014) Early diagnosis of Alzheimer's disease with deep learning. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pp. 1015–1018.

[84] Li F, Tran L, Thung K-H, Ji S, Shen D, Li J (2015) A Robust Deep Model for Improved Classification of AD/MCI Patients. *IEEE Journal of Biomedical and Health Informatics* **19**, 1610–1616.

[85] Trambaiolli LR, Spolaôr N, Lorena AC, Anghinah R, Sato JR (2017) Feature selection before EEG classification supports the diagnosis of Alzheimer's disease. *Clinical Neurophysiology* **128**, 2058–2067.

[86] Poil S-S, de Haan W, van der Flier WM, Mansvelder HD, Scheltens P, Linkenkaer-Hansen K (2013) Integrative EEG biomarkers predict progression to Alzheimer's disease at the MCI stage. *Front Aging Neurosci* **5**,.

[87] Saito T, Rehmsmeier M (2015) The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS One* **10**,.

[88] LaMontagne PJ, Benzinger TLS, Morris JC, Keefe S, Hornbeck R, Xiong C, Grant E, Hassenstab J, Moulder K, Vlassenko A, Raichle ME, Cruchaga C, Marcus D (2019) *OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease*, Radiology and Imaging.

[89] Arisi I, Bertolazzi P, Cappelli E, Conte F, Cumbo F, Fiscon G, Sonnessa M, Taglino F (2018) An ontology-based approach to improve data querying and organization of Alzheimer's Disease data. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2732–2734.

[90] Górriz JM, Segovia F, Ramírez J, Lassl A, Salas-Gonzalez D (2011) GMM based SPECT image classification for the diagnosis of Alzheimer's disease. *Applied Soft Computing* **11**, 2313–2325.

[91] Padilla P, Lopez M, Gorriz JM, Ramirez J, Salas-Gonzalez D, Alvarez I (2012) NMF-SVM Based CAD Tool Applied to Functional Brain Images for the Diagnosis of Alzheimer's Disease.

IEEE Transactions on Medical Imaging **31**, 207–216.

[92] Segovia F, Górriz JM, Ramírez J, Salas-González D, Álvarez I (2013) Early diagnosis of Alzheimer's disease based on Partial Least Squares and Support Vector Machine. *Expert Systems with Applications* **40**, 677–683.

[93] McBride J, Zhao X, Munro N, Jiang Y, Smith C, Jicha G (2013) Scalp EEG signal reconstruction for detection of mild cognitive impairment and early Alzheimer's disease. In *2013 Biomedical Sciences and Engineering Conference (BSEC)*, pp. 1–4.

[94] Buscema M, Vernieri F, Massini G, Scrascia F, Breda M, Rossini PM, Grossi E (2015) An improved I-FAST system for the diagnosis of Alzheimer's disease from unprocessed electroencephalograms by using robust invariant features. *Artificial Intelligence in Medicine* **64**, 59–74.

[95] Lu D, Popuri K, Ding GW, Balachandar R, Beg MF (2018) Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. *Medical Image Analysis* **46**, 26–34.

Table 1: Comparison of the medical techniques described in the Introduction section.

Technique	Invasive	Costly	Accessibility
MRI	No	Yes	Low
SPECT	Yes	Yes	Low
PET	Yes	Yes	Low
EEG	No	No	High
MEG	No	Yes	Low

Table 2: Summary of some of the main AD automated detection approaches over the past decade. In Table 2 (a), we included works that addressed at least one of the two main AD classification problems in literature (HC vs MCI and HC vs AD). We included two *Performance* columns, one for each of these two problems. In Table 2 (b), we gathered the works that analyzed a different classification task (in those cases only a *Performance* column is reported, and the specific task is indicated in brackets in that column). The studies that analyzed the HC vs MCI vs AD multiclass problem are also included in Table 2 (b). A, R, S and AUC, stand for accuracy, sensitivity, specificity, and area under the receiver operating curve, respectively.

Table 2 (a): Studies that addressed the HC vs MCI and the HC vs AD classification tasks.

Study	Source	Cohorts	Protocol	Features (feature selection)	Model	Validation	Performance (HC vs MCI)	Performance (HC vs AD)
López et al. [23] (2011)	PET	HC: 52 MCI: 114 AD: 53	-	Voxel data (LDA and PCA)	SVM, ANN	LOO CV	A = 79.52 R = 93.86 S = 48.08	A = 89.52 R = 86.68 S = 85.72
Górriz et al. [90] (2011)	SPECT	NOR: 43 AD: 30	-	Gaussian Mixture Model	SVM	k-fold CV	-	A = 84.51 R = 76.67 S = 90.24
Trambaiolli et al. [31] (2011)	EEG (19)	HC: 19 AD: 16	Resting State	Coherence and spectral peaks (Weka tool)	SVM	Non-cross validation	-	A = 79.9 R = 83.2 S = 76.4
Padilla et al. [91] (2012)	SPECT	NOR: 41 AD: 56	-	Voxel data (Non-Negative Matrix Factorization)	SVM	LOO CV	-	A = 91.42 R = 90.56 S = 92.30
Segovia et al. [92] (2013)	SPECT	HC: 41 AD: 56	-	Partial Least Squares (Out-Of-Bag Error)	SVM	LOO CV	-	A = 91.75 R = 92.68 S = 91.07
McBride et al. [93] (2013)	EEG (32)	Not reported	Resting State	EEG reconstruction scores (LOO CV)	SVM	LOO CV	A = 90.3 R = 87.5 S = 93.3	A = 90.6 R = 88.2 S = 93.3
Aghajani et al. [76] (2013)	EEG (128)	HC: 17 AD: 17	Resting State	PSD (Singular Value Decomposition)	SVM	LOO CV	-	A = 84.4 R = 75 S = 93.7
Liu et al. [83] (2014)	MRI, PET	HC: 77 MCI: 169 AD: 65	-	MRI volumes and PET patterns (alteration of the number of neurons)	DNN	k-fold CV	A = 76.92 R = 74.29 S = 78.13	A = 91.4
Fiscon et al. [59] (2014)	EEG (19)	HC: 14 MCI: 37 AD: 49	Resting State	FFT	Decision tree	LOO CV	A = 90 R = 90 S = 87	A = 87.76 R = 88.57 S = 87.22
McBride et al. [65] (2014)	EEG (32)	HC: 15 MCI: 16 AD: 17	Resting State	Spectral and complexity parameters (LOO CV)	SVM	LOO CV	A = 96.8 R = 93.8 S = 100	A = 96.9 R = 100 S = 93.3
Munteanu et al. [67] (2015)	MRI	HC: 99 MCI: 94 AD: 67	-	Voxel volumes (Filtering)	ANN, Forest	k-fold CV	AUC = 0.81	AUC = 0.89
Khedher et al. [24] (2015)[a]	MRI	HC: 229 MCI: 401 AD: 188	-	Voxel gray and white matter (Voxel selection)	SVM	k-fold CV	A = 81.89 R = 82.16 S = 81.62	A = 88.49 R = 91.27 S = 85.11
Moradi et al. [73] (2015)	MRI	HC: 231 AD: 200	-	Voxel data (ElasticNet)	SVM, LDS	k-fold CV	-	A = 81.7 R = 86.7 S = 73.6
Wang et al. [30] (2015)	EEG (16)	HC: 14 AD: 14	Resting State	PSD and coherence (PCA)	Cluster analysis	None	-	A = 91.4 R = 100 S = 82.9

Maestú et al. [29] (2015)	MEG (306)	NC: 82 MCI: 102	Resting State	Connectivity via Mutual Information (Clinical Data Partitioning)	Multiple classifiers	Bootstrap	A = 83 R = 100 S = 69	-
Morabito et al. [68] (2016)	EEG (19)	HC: 23 MCI: 56 AD: 63	Resting State	Skewness, mean and std in three frequency bands	CNN	k-fold CV	A = 85 R = 84 S = 81	A = 85 R = 85 S = 82
Amezquita et al. [28] (2016)	MEG (148)	HC: 19 MCI: 18	Sternberg test	Permutation entropy (One-way ANOVA)	EPNN	CV	A = 98.4	-
Cassani et al. [42] (2017)	EEG (7)	HC: 24 AD: 35	Resting State	Spectral power, coherence, amplitude modulation (L1 norm)	SVM	LOSO CV	-	A = 69.5 R = 71.4 S = 66.6
Trambaiolli et al. [85] (2017)	EEG (19)	HC: 12 AD: 22	Resting State	Spectral peaks in the main EEG spectral bands (Filtered Subset Evaluator)	SVM	LOSO CV	-	A = 91.18 R = 90.91 S = 91.67
Kulkarni et al. [61] (2017)	EEG (16)	HC: 50 AD: 50	Resting State	Relative Power, WT and complexity	SVM	LOSO CV	-	A = 96
Tong et al. [72] (2017)	MRI	HC: 229 AD: 191	-	Voxel intensities (ElasticNet)	SVM, RF	k-fold CV	A = 80.7 R = 86.7 S = 72.6	A = 73 R = 73 S = 46
Dimitriadis et al. [80] (2018)	MEG (306)	HC: 30 MCI: 24	Resting State	Functional connectivity (MRMR)	SVM	k-fold CV	A = 98	-
Ruiz-Gómez et al. [45](2018)	EEG (19)	HC: 37 MCI: 37 AD: 37	Resting State	Spectral and nonlinear features (Fast Correlation Based Filter)	MLP, QDA, LDA	LOSO CV	A = 78.43 R = 82.35 S = 70.59	A = 98.05 R = 97.99 S = 98.21
Fiscon et al. [60] (2018)	EEG (19)	HC: 23 MCI: 37 AD: 49	Resting State	Fourier Transform and Wavelet Transform	Decision Trees	LOO CV	A = 91.7 R = 91.7 S = 91.5	A = 83.3 R = 83.3 S = 78.0
Mazaheri et al. [75] (2018)	EEG (19-32)	HC: 11 MCI: 25	Language task	Oscillatory changes	SVM	LOO CV	R = 0.80 S = 0.95	-
Khatun et al. [40] (2019)	EEG (1)	HC: 15 MCI: 8	Audio ERP	Temporal and amplitude of the ERP (Random Forest)	SVM	LOO CV	A = 87.9 R = 84.8 S = 95	-
Durongbhan et al. [64] (2019)	EEG (1)	HC: 20 AD: 20	Resting State	FT and WT	KNN	k-fold CV	-	A = 99.2
Ieracitano et al. [35] (2020)	EEG (10)	HC: 63 MCI: 63 AD: 63	Resting State	Continuous WT and BiSpectrum	MLP	k-fold CV	A = 96.24 R = 95.86	A = 96.95 R = 94.91

Study	**Source**	**Cohorts**	**Protocol**	**Features (feature selection)**	**Model**	**Validation**	**Performance**
Chapman et al. [41] (2011)	EEG (1)	sMCI: 28 pMCI: 15	Number letter test	ERP components (PCA)	Discriminant analysis	LOO CV	(sMCI vs pMCI) A = 79 R = 80 S = 79
Podgorelec et al. [66] (2012)	EEG (2)	No AD: 12% AD: 88%	Not Reported	Time and frequency domain parameters (tree-based method)	Evolutionary algorithm	k-fold CV	(No AD vs AD) A = 86.05 R = 88.89 S = 71.43
Poil et al. [86] (2013)	EEG (21)	sMCI: 39 pMCI: 25	Resting State	Neurophysiological Biomarker Toolbox (Student's t test and GA)	Logistic regression and GA	Half split CV	(sMCI vs pMCI) R = 88 S = 82
Li et al. [84] (2015)	MRI, PET, CSF	HC: 52 MCI: 99 AD: 51	-	Multi-domain features (PCA and Lasso)	DNN with dropout	k-fold CV	(HC vs MCI vs AD) A = 77.4
Buscema et al. [94] (2015)	EEG (19)	HC: 99 MCI: 46 AD: 127	Resting State	Spatial invariants via MS-ROM (TWIST algorithm)	Multiple classifiers	TWIST algorithm	(HC vs MCI vs AD) A = 98.27 R = 98.29 S = 98.21
Farooq et al. [26] (2017)[a]	MRI	HC > 1000 MCI > 1000 AD > 1000	-	Voxel gray matter (None)	DNN	Not reported	(HC vs MCI vs AD) A = 98 R = 94-99 S = 94-99
Nanni et al. [69] (2018)[b]	MRI	HC: 100 MCI: 200 AD: 100	-	Physical measurements (Fischer score)	SVM	k-fold CV	(HC vs MCI vs AD) A = 0.57
Houmani et al. [71] (2018)	EEG (30)	SCI: 22 pAD: 49	Resting State	Epoch based entropy and bump models (Orthogonal Forward Regression)	SVM	LOSO CV	(SCI vs possible AD) A = 91.6 R = 87.8 S = 100

Lu et al. [95] (2018)	PET	sMCI: 409 pMCI: 112	-	Metabolism features in regions of interest (None)	DNN	k-fold CV	(sMCI vs pMCI) A = 81.55 R = 73.33 S = 83.83
Pusil et al. [79] (2019)	MEG (306)	sMCI: 27 pMCI: 27	Resting State	ROI connectivity (centroid selection)	SVM	k-fold CV	(sMCI vs pMCI) A = 100 R = 100 S = 100
Pusil et al. [78] (2019)	MEG (306)	sMCI: 27 pMCI: 27 HC: 27	Resting State	Connectivity, temporal lobe volume and test scores (None)	Logistic regression	LOO CV	(sMCI vs pMCI vs HC) A = 96.2 R = 92.6 S = 100
Amezquita et al. [70] (2019)	EEG (19)	MCI: 37 AD: 37	Resting State	Fractal dimension and Hurst exponent (one-way ANOVA)	EPNN	Random subject selection	(MCI vs AD) A = 90.3 R = 92.1 S = 87.9

Table 2 (b): Studies that addressed the classification of other AD cohorts or considered a 3-label multiclass problem.

[a]These studies analyzed images from an image database, thus the figures reported in the table refer to images in each cohort. [b]This work utilized a database that stored data from four classes (HC, non-converters MCI, converters MCI and AD) thus, for the sake of clarity, the two MCI categories have been collapsed in the table.

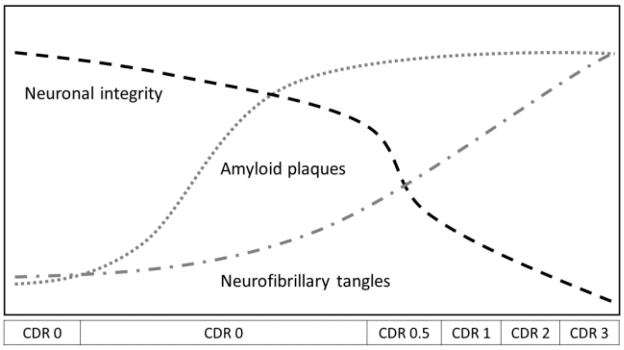


Fig. 1. Progression of the stages of AD regarding the evolution of the hallmark biomarkers and neuronal integrity. The dotted line refers to the appearance of amyloid plaques. The dashed and dotted line corresponds to the accumulation of neurofibrillary tangles and the dashed line represents neuronal integrity. Vertical and horizontal axes correspond to intensity and time. Each stage is assigned a Clinical Dementia Rating (CDR) value according to the severity of the disease in that stage. The first stage corresponds to the Non-AD period and has a CDR score of 0. The second stage refers to Preclinical AD and has a CDR score of 0. In this phase, amyloid plaques start forming in the brain for more than a decade, starting the neuronal decay. The last four stages represent MCI, mild AD, moderate AD and severe AD and have CDR scores of 0.5, 1, 2 and 3 respectively. This graph has been inspired by the work of Perrin et al [17].

Appendix B

An Attention-driven Videogame based on Steady-State Motion Visual Evoked Potentials

Expert Systems

An Attention-driven Videogame based on Steady-State Motion Visual Evoked Potentials

Eduardo Perez-Valero [1, 2], Miguel Angel Lopez-Gordo [1, 3 *] and Miguel A. Vaquero-Blasco [1, 3]

[1] Research Centre for Information and Communications Technologies (CITIC), University of Granada, Calle Periodista Rafael Gómez Montero, 2, 18014 Granada (Spain).
[2] Department of Computer Architecture and Technology, University of Granada, Calle Periodista Daniel Saucedo Aranda, s/n, 18014 Granada (Spain).
[3] Department of Signal Theory, Telematics and Communications, University of Granada, Calle Periodista Daniel Saucedo Aranda, s/n, 18014 Granada (Spain).
Correspondence: malg@ugr.es; Tel.: +34-958-249-721

Abstract. In Neuroscience, steady-state visually evoked potentials (SSVEPs) have been generally used for the characterization of dynamic visual processes along the afferent pathway. In Neuroengineering, SSVEP-based brain-computer interfaces (SSVEP-BCI) have been used in applications such as communications or entertainment for the detection of the overt attention to static-flickering stimuli or other structured stimulation that involves motion (SSMVEP-BCI). In this work, we propose an attention-driven videogame controlled by an SSVEP-BCI with mobile-flickering stimuli. In this game, enemy avatars launch attacks that are presented as mobile ring-shaped checkerboards that flicker at 15 Hz to the player, who must deflect them by exerting attention. We detected the attention of the participants based on the power spectral density of the elicited SSVEP and the adjacent EEG background noise. Twenty volunteers participated in this study. After completion, according to a game experience survey, the participants described the game as amusing and challenging. We also conducted a level survey that revealed a significant difference between the levels of attention exerted by the participants when they passed the levels with respect to when they missed (p-value = $17 \cdot 10^{-5}$). Additionally, the detection accuracy of the SSVEP-BCI had a baseline level of 84 %. The results suggest that mobile flickers can be robustly detected within few seconds by means of an SSVEP-BCI. This principle could be used as a serious game to play or train the attention and the visual tracking capabilities to mobile target stimuli in special needs education schools or in centers dealing with attention disorders.

Keywords: SSVEP, attention, videogame, EEG, brain-computer interface

1. Introduction

Visually evoked potentials (VEP) are electrophysiological signals generated by the visual cortex as a response to, typically, the pattern reversal of a visual stimulus. These responses can be elicited by presenting a subject a contrast reversal paradigm at low frequencies (typically below 2 Hz) such as the described in (International Society for Clinical Electrophysiology of Vision et al., 2016). Steady-state visually evoked potentials (SSVEP) are quasi-sinusoidal responses to visual stimuli that flicker in a range of frequencies (usually above 4 Hz) [2]. The power spectral density (PSD) of an SSVEP extends over a narrow band whose central frequency matches that of the flickering stimulus (Kelly, Lalor, Finucane, McDarby, & Reilly, 2005). Both VEPs and SSVEPs have been used in academic and clinical institutions for the assessment of the afferent visual pathway (e.g., glaucoma) and for visual processing such as in the assessment of covert attention at work, migraine with aura, patients with febrile seizures and others (Ahluwalia & Vold, n.d.; Gregori Grgič, Calore, & de'Sperati, 2016; Sheppard et al., 2013; Shibata, Yamane, Otuka, & Iwata, 2008).

In Neuroengineering, brain-computer interfaces based on SSVEP (SSVEP-BCI) detect both overt and covert visual attention that are correlates of the endogenous volition within the frame of a paradigm or application. A BCI is a device that provides the brain with a new, non-muscular communication and control channel (Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002), and thus allows subjects to interact with an external device by means of their neural activity. Currently, there exist many kinds of BCIs and applications, such as those intended for communication, emotion and motor imagery classification (Hassanpour, Moradikia, Adeli, Khayami,

& Shamsinjead Babaki, 2019; Miguel Angel Lopez, Pomares, Damas, Prieto, & de la Plaza Hernandez, 2007; Lopez-Gordo & Pelayo, 2013; Lopez-Gordo, Pelayo, Fernandez, & Padilla, 2015; Minguillon, Lopez-Gordo, Renedo-Criado, Sanchez-Carrion, & Pelayo, 2017; Vaquero-Blasco, Perez-Valero, Lopez-Gordo, & Morillas, 2020). Among them, one of the most popular is the SSVEP-BCI. The reason behind is that this BCI barely requires training, cognitive effort and its performance, in terms of accuracy and information transmission rate (ITR), is typically high. Furthermore, due to its nature, most of the energy of the SSVEP extends over a narrow spectral band, and thus a high signal-to-noisy ratio can be achieved, what is ideal detection-wise and at the same time permits the voluntary modulation of its energy by means of visual attention (Russo, Teder-Sälejärvi, & Hillyard, 2003; Walter, Quigley, Andersen, & Mueller, 2012). To review various examples of health-related and daily life SSVEP-BCI applications, refer to (Brennan et al., 2015; Lim, Lee, Hwang, Kim, & Im, 2015; Yin, Zhou, Jiang, Yu, & Hu, 2015). In other fields, such as gaming, BCIs have been used in some studies as an alternative to classic controllers. For instance, Mind the sheep (Gürkök, Nijholt, Poel, & Obbink, 2013), The Mindgame (Finke, Lenhardt, & Ritter, 2009), Tetris (Pires, Torres, Casaleiro, Nunes, & Castelo-Branco, 2011) and others (Krepki, Blankertz, Curio, & Müller, 2007).

In the above-mentioned applications, the flickering stimuli have a fixed size and are located at a static position of the screen, and the participant stares at one of them while the SSVEP-BCI performs an online detection. In the last years, researchers and developers have been working on a variant called steady-state motion visual evoked potentials (SSMVEP-BCI). In this paradigm, stimuli move up and down, expand, contract or oscillate to generate motion VEPs that repeat at a certain rate (Guitong et al., 2018; Han, Xu, Xie, Chen, & Zhang, 2018; P. Stawicki, Rezeika, Saboor, & Volosyak, 2019). The mayor advantage of SSMVEP-BCIs is that they produce less visual fatigue seemingly without impacting their performance in terms of accuracy and ITR. See (Piotr Stawicki & Volosyak, 2020) for a recent comparison of diverse SSMVEP approaches.

In this work, we propose an attention-driven videogame that uses mobile flickers. An earlier version of this game was first demonstrated in a technologic event celebrated at the University of Granada in 2018 (Lopez-Gordo, Perez, & Minguillon, 2019). The videogame is designed to detect the overt attention that players exert on attacks that enemy avatars fire in a battle scenario. For this purpose, we texturized the enemy attacks as mobile ring-shaped checkerboards that reverse their contrast at a constant frequency and expand their size as they approach to the player, thus eliciting mobile SSVEPs in their electroencephalographic (EEG) activity. We based the attention detection on a combination of SSVEP amplitude and the EEG background noise level of the spectral bins adjacent to the frequency of stimulation. As it is indicated in (Miguel A. Lopez, Pelayo, Madrid, & Prieto, 2009), SSVEP power is a valuable biomarker in BCI applications. During gameplay, if the detection process assesses that participants are paying sufficient attention to the attack of an enemy, that attack is immediately deflected and the players will score a point, otherwise, the attack will reach the foreground of the virtual scenario and the player will concede a point. The game involves the use of gamification principles such as continuous progress feedback, immediate success feedback and autonomy support. In this studio, the participants described the game as an exciting and challenging entertainment that kept their attention focused for moderate periods of time while experiencing an increasing level of difficulty. Our proposal arises as an alternative that could be used as a serious game to play and train the attentional and visual tracking capabilities with direct application in Special Needs Education or in attention disorders.

2. Methods

2.1. Participants

Twenty healthy volunteers (15 males and 5 females) participated in the study (mean age 26.20 ± 9.23). They were asked to carefully read and sign an informed consent prior to the onset of the experiment. Each participant went through a single experiment session that lasted approximately forty minutes. The entire data capture of the experiment was completed in eight days. During the experiment, participants were comfortably sitting and they were asked to move as little as possible as their interaction with the game was restricted to their brain activity. They were recruited from the students and staff at the University of Granada (Spain) and their relatives. Only those participants who did not present medical conditions and were not taking any medications that could interfere with the study were considered. All the participants presented normal or corrected sight and reported no issues during the study visual stimulation.

2.2. Game description

The game spans five levels. In each level, the screen is conceived as a virtual 3D scenario where enemy avatars appear (see Figure 1). In this scenario, horizontal (x) and vertical (y) coordinates stretch from (x = 0, y = 0) in the left top corner of the screen to (x = 10, y = 10) in the right bottom corner of it. The depth (z) spans from 10 (background) to zero (foreground).

Once a level starts, an enemy avatar shows up in a random position of the virtual scenario and launches an attack in the form of a ring-shaped checkerboard stimulus with a fixation cross. Immediately after, the stimulus advances from its initial position toward the center position of the foreground of the scenario while it augments in size and reverses its pattern at a rate of 15 Hz. The goal of the participants is to prevent the enemy attack from reaching the center of the foreground (coordinates x = 5, y = 5, z = 0). To this effect, they must visually pursue the mobile stimulus by maintaining their attention on the fixation cross (see white cross at the center of the checkerboard in Figure 1). If the participants exert sufficient visual attention, the enemy attack will be deflected and they will score a point. Otherwise, the mobile stimuli will get to the center of the foreground of the virtual scenario and the computer will score the point. The first contender (participant or computer) to reach five points wins the level. If the participants miss a level, that level will be repeated, else the participants will head to the next level.

Figure 1: Progression of the mobile visual stimulus. The stimulus moves from the initial location of the enemy avatar (left-most image) in the virtual scenario, with coordinates (x = 1, y = 1, z = 10) in this example, to the center of its foreground (right-most image), with coordinates (x = 5, y = 5, z = 0). Through this progression, the checkerboard inverts its pattern at a constant rate of 15 Hz and augments its size. If the participants apply enough visual attention to the fixation cross, the enemy attack will be neutralized and they will score a point. Otherwise the attack will reach the center of the foreground and the computer will score a point. Notice that we have added the coordinate system of the virtual scenario in the bottom right corner of the left-most image of the figure.

To deliver a challenging experience throughout the game, we designed each level to be harder than the previous by setting the initial position of the enemies closer to the foreground of the virtual scenario and by boosting the speed of the mobile stimuli. The end of the game takes place when the participants beat the five levels or when fifteen minutes have passed since the start of the game. For a complete description of the stages of the experiment, refer to Figure 2.

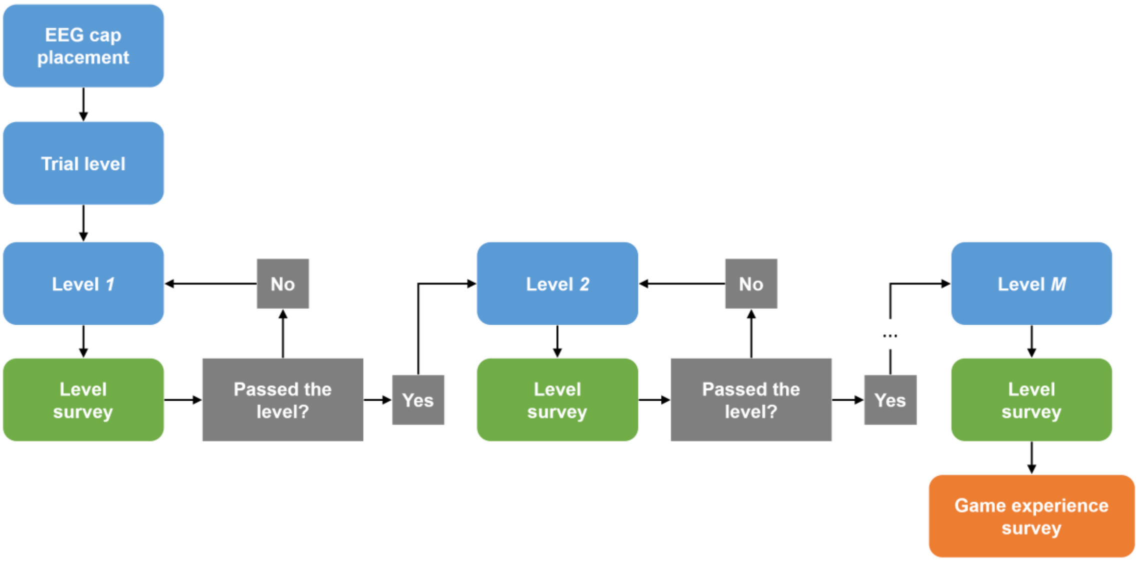


Figure 2: Flow diagram of the experiment. First, the participants are equipped with the EEG cap for the acquisition of their brain activity. Next, the participants play a trial level to familiarize with the game and to calibrate the attention detection threshold parameter. Once this level is over, the gameplay begins. The game spans five levels. The winner of each level is the first contender to reach five points in that level. If the participants pass a level, they will tackle the next one. Otherwise they will play the same level again. The game is over when the participants complete the five levels or when fifteen minutes have elapsed since the start of the gameplay. Note that the right-most blue text box stands for Level *M.* We named it like this to stress the fact that the game does not necessarily end after Level 5 if fifteen minutes have passed. The green textboxes represent the level surveys that we used to gather information about the participant at the end of every level. The orange text box represents the game experience survey that we conducted at the end of the game to evaluate its main aspects.

2.3. System design

The game system that we developed in this study is composed by four modules: a client to present the game to the participant (stimulation client), an EEG acquisition system, a client to monitor the brain activity (monitoring client) and a control server to coordinate the system (see Figure 3). The stimulation client presents the mobile visual stimuli and the rest of the game elements to the participants. Further details about the visual stimulation are provided in the following subsection. With regard to the brain activity of the participants, we placed two electrodes in the visual cortex area (O1 and O2 positions of the 10-20 International System) and we used the RABio w8 for the acquisition of the EEG. Our team at the University of Granada designed and built this device, and we have successfully used it in previous studies (Minguillon et al., 2017; Minguillon, Perez, Lopez-Gordo, Pelayo, & Sanchez-Carrion, 2018; Vaquero-Blasco et al., 2020). This BCI system enables the acquisition of different kinds of bio signals (EEG in the case of this study) and their transmission to the monitoring client via Bluetooth. The monitoring client runs a MATLAB graphic user interface for real time visualization of the EEG of the participants. It also sends the raw EEG to the control server through a TCP/IP socket, every second, for the attention detection process. The control server manages the attention detection process through signal processing of the raw EEG that receives from the monitoring client. Once this procedure is completed, the control server decides if one of the players (participant or computer) scored a point and guides the flow of the game. Subsequently, the control server sends a command to the stimulation client to continue the game through the right course.

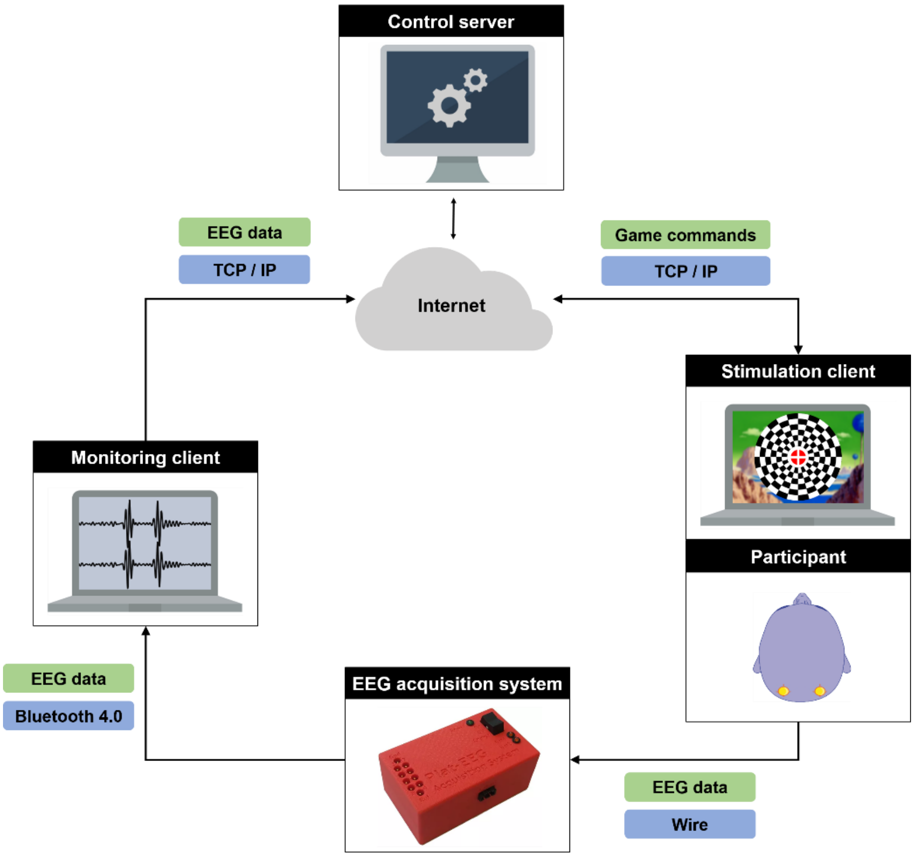


Figure 3: Diagram of the system designed for this study. The four modules of the system and the participant are displayed in the figure. The stimulation client presents the game to the participant, whose EEG is obtained by an acquisition system that transmits the brain signals via Bluetooth to the monitoring client. This client sends the raw EEG data to the server in one-second intervals through a TCP/IP socket. The server processes the EEG and performs the attention detection process. Depending on the outcome of this procedure, the server sends commands to the stimulation client via another TCP/IP socket to continue the flow of the game. The blue text boxes in the figure indicate the communication channel employed in each segment of the system. The green text boxes illustrate the type of information that is transmitted.

2.4. Visual stimulation

For the stimulation client we used a 13-inch laptop with an Intel i7-5500U CPU, 8 GB RAM and NVIDIA 920M graphic card. The screen resolution was set to 1920 x 1080 pixels and the refresh rate of the screen was adjusted to 60 Hz. This client was placed 60 cm from the participant and we used it to present the mobile stimuli and the rest of the game elements (enemy avatars, background, scoreboard, etc.) via MATLAB R2016a (MathWorks, Natick, USA) using Windows 10 (Microsoft, Redmond, USA). Each mobile stimulus consisted of a ring-shaped checkerboard (see Figure 1) that flickered at a constant rate of 15 Hz (four times the screen frame period) to elicit an SSVEP of the same frequency in the brain activity of the participants. Alongside this inversion, we programmed the checkerboard to move toward the center of the foreground of our virtual scenario (represented by coordinates x = 5, y = 5 z = 0) and to simultaneously augment its size with the intention to create the illusion of its approximation to the participant. To achieve an optimal precision in the presentation of the mobile stimuli, we developed the game interface using Psychtoolbox, a MATLAB free software toolbox for visual and neuroscience experiments. The formula that we applied to generate the ring-shaped checkerboard that models the enemy attacks is reported in Equation 1.

$$\frac{1 + \mathrm{sign}\left(\sin\left(\mathrm{atan}(x,y) \cdot {}^{R}\!/_{2}\right)\right) \cdot \mathrm{sign}\left(\sin\left(\sqrt{x^2 + y^2}\right)\right)}{2} \qquad (1)$$

x and *y* stand for the ring pixel grid coordinates and *R* refers to the number of checks in a ring.

2.5. Attention detection

To parametrize the attention exerted by the participants during gameplay we calculated a power spectral density (PSD) parameter derived from the brain activity in positions O1 and O2. This parameter was defined as the ratio between the power at the frequency of the SSVEP generated by the mobile visual stimuli, (14-16) Hz, and the power at the surrounding frequencies, (12-13) Hz and (17-18) Hz. This ratio was measured in dB and its formula is reported in Equation 2.

$$\text{Attention}_{\text{parameter}} = \frac{\text{P}[14-16]\text{ Hz}}{\text{P}[12-13]\text{ Hz} + \text{P}[17-18]\text{ Hz}}\ (\text{dB}) \qquad (2)$$

P represents the power in the frequency range in brackets.

To determine if the attention expended by the participants was sufficient to score a point, we defined a threshold for the attention parameter described in Equation 2. For calibration purposes, all the participants played a trial version of the first level of the game (see Figure 2) while we registered the attention parameter that they produced. Once the trial was over, we adjusted the threshold to the maximum value of the attention parameter that we registered minus three decibels to assure that the game resulted demanding enough. During the gameplay, we computed the value of the attention parameter using two-second epochs of the EEG signal without overlapping, and then compared the value obtained with the threshold. This procedure was referred as attention detection process in the previous subsections. If the value was higher than the threshold, the point was conceded to the participant, otherwise the game continued until the enemy attack reached the center of the foreground or until the attention parameter was higher than the threshold.

For the calculation of the attention parameter, we processed each two-second EEG epoch as follows: first, we applied a 2nd order bandpass Butterworth filter with a bandpass between 6 Hz and 75 Hz. Next, we filtered the signal with a 2nd order Butterworth notch filter to remove electric coupling (stopband from 47 Hz to 53 Hz). Then, we applied a Tukey window to the filtered signal with a cosine-tapered section ratio of 0.3 to obtain a smoother spectrum. After that, we standardized the processed signal using MATLAB's *zscore* function and lastly, we estimated the PSD in the bands of interest by means of the periodogram (see Equation 2). This signal processing procedure is represented in Figure 4.

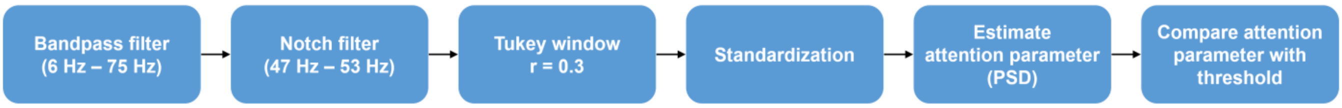


Figure 4: Block diagram of the signal processing applied to the raw two-second EEG epochs. First, we applied a bandpass filter to withdraw undesired spectral content and a notch filter to remove the electric coupling. Then, we used a Tukey window to flatten the edges of the signal and obtain a smoother spectrum. Next, we standardized the processed signal and estimated the attention parameter based on the PSD. Finally, we compared the obtained value of the attention parameter with the established threshold to decide if the participant deserved a point.

To assess the robustness of the detection, we requested one of the participants to play a special level of the game. Concretely, this level spanned one hundred points. Out of these one hundred points, fifty of the enemy attacks were presented as the already described ring-shaped checkerboards that inverted their pattern at 15 Hz, while the other fifty stimuli flickered at a rate of 10 Hz (the different flickering frequencies were visually unnoticeable to the participants). The order in what the two different kinds of stimuli appeared was randomized using a normal distribution to avoid any acknowledgement from the participant. After the conclusion of this test, we expected the attention detection process to classify all the 10 Hz stimuli as points for the computer (absence of false positives) and the majority of the 15 Hz stimuli as points for the participant (this last quantity would depend on the attention exerted by the participant and the quality of the bio signals). We assumed this since the attention detection process was designed to measure the attention that the participants exert on the 15 Hz flickering stimulus, and thus the 10 Hz stimulus should not raise any false points for the participant.

2.6. Surveys

We conducted a two-question survey at the end of each level (level survey). Our purpose was to gather the self-perceived level of attention exerted by the participants along with their viewpoint about the equity the result of the level. The two questions we asked were:

1. "In a scale from 1 (minimum) to 5 (maximum), what was the level of attention you applied during this level?"
2. "In a scale from 1 (minimum) to 5 (maximum), what was the equity of the scoreboard result in this level?".

We also conducted a four-question survey at the end of the game (game experience survey). Our intention was to evaluate the game according to three main aspects, amusement, challenge and boredom, and also to harvest the global opinion of the participants about the game. To this effect, we asked the following four questions:

1. "In a scale from 1 (minimum) to 5 (maximum), what level of amusement did you experience throughout the game?"
2. "In a scale from 1 (minimum) to 5 (maximum), what level of challenge did you experience throughout the game?"
3. "In a scale from 1 (minimum) to 5 (maximum), what level of boredom did you experience throughout the game?"
4. "In a scale from 1 (minimum) to 5 (maximum), how much did you enjoy the whole experience?

2.7. Statistical analysis

To assess the performance of the participants in each level we calculated the grand average and the standard error of the mean (SEM) of the following metrics: total time played, points scored by the participant and points scored by the computer. We also estimated the success rate per level (as the proportion between the times that a level was passed and the times that it was attempted) and the point difference per level (as the absolute value of the difference between the points scored by the participants and the points scored by the computer).

To understand the perception of the participants relative to the attention they exerted and the equity of the scoreboard result, we calculated the grand average and the SEM of the ratings provided during the level survey that we conducted at the end of each level. With the aim to evaluate the amusement, boredom and challenge that the participants experienced, as well as their global assessment on the game, we followed the same procedure with the ratings that they provided during the game experience survey.

We applied the Wilcoxon-Mann-Whitney test to assess if there were statistical differences between the responses that the participants provided to the question about the attention that they exerted during the level survey when they passed the levels and when they missed. We utilized this test since the distributions did not pass the Shapiro-Wilk normality test and the samples were not paired.

Finally, we splitted the participants into three groups according to their gaming experience: no experience (0 gaming hours per week), medium experience (1 to 5 gaming hours per week) and high experience (6 or more gaming hours per week), and we carried out a Kruskal-Wallis test to assess if there were significant differences between the point difference they obtained during the game. We applied this test since the distributions did not pass the Shapiro-Wilk normality test. For all tests performed in this study the significance level, α, was set to 0.05.

3. Results

Figure 5 shows the average total time that the participants spent in each level of the game.

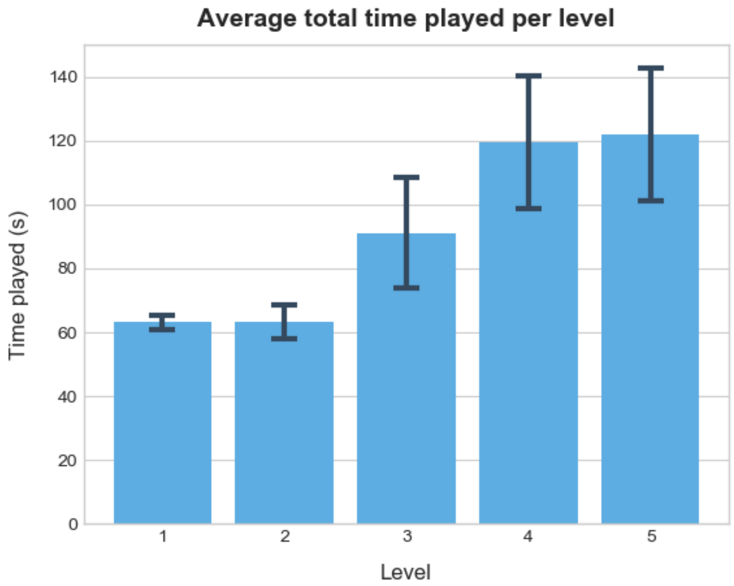


Figure 5: Average total time consumed by participants in each level of the game. For each participant, the total time spent in a level was calculated as the sum of the time consumed in each of the occasions that the participant faced that particular level. The error bars indicate the standard error of the mean (SEM).

Figure 6 displays the average number of points scored in each level by the participants and by the computer.

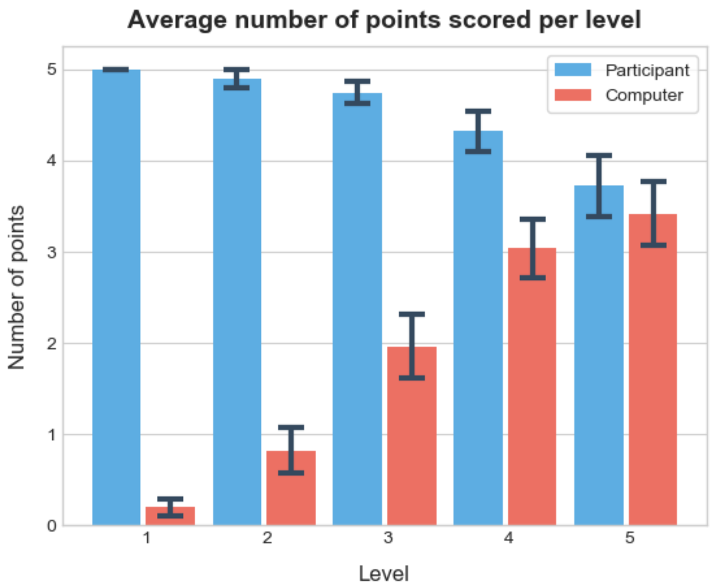


Figure 6: Average number of points scored by the participants and the computer in each level. The blue bars correspond to the points scored by the participants, and the red bars refer to the points scored by the computer. The error bars indicate the standard error of the mean (SEM). Notice that, since all the participants passed the first level on their first attempt, there is no error bar for the bar relative to the points scored by the participants in that level.

Figure 7 represents the success rate for each level of the game alongside the 95% confidence interval.

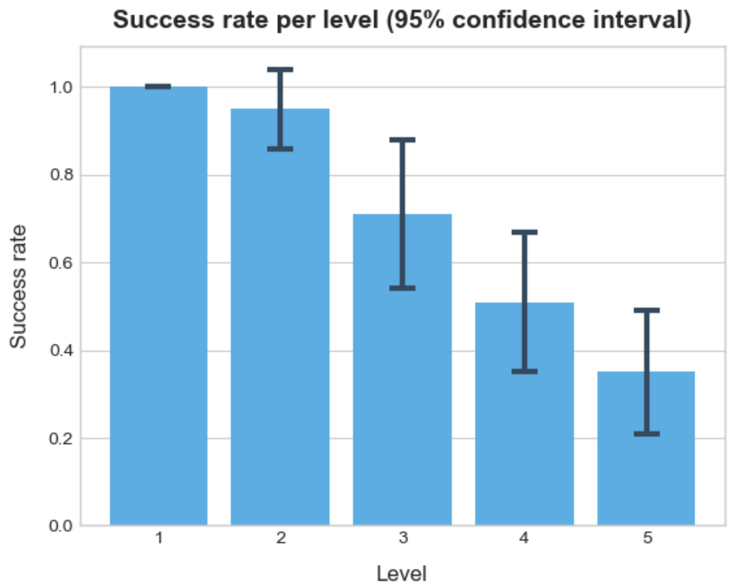
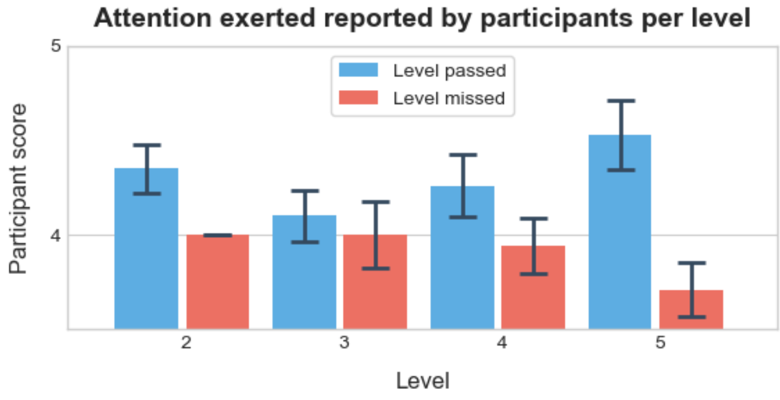
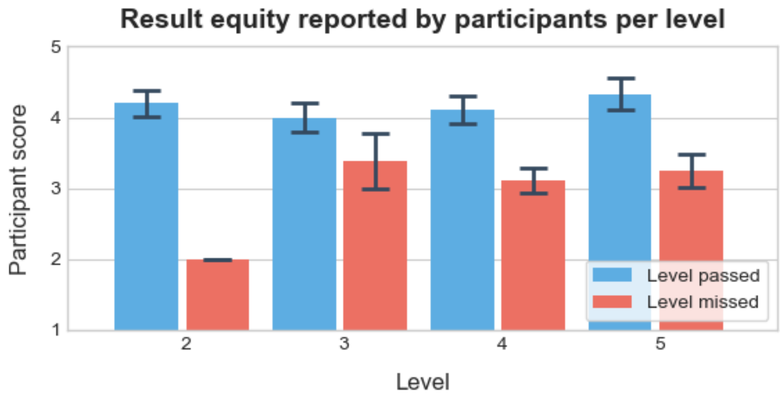


Figure 7: Success rate for each level of the game. We calculated the success rate of each level as the proportion between the times that the level was passed and the times that it was attempted. The error bars enclose the 95% confidence interval. Notice that, since all the participants passed the first level in their first attempt, there is no error bar for the bar relative to the success rate of that level.

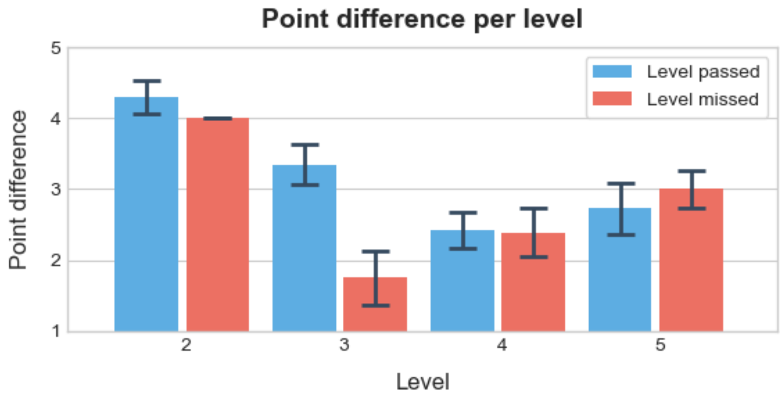
Figure 8 displays, for levels two to five, the average attention exerted and the average result equity reported by participants in the level surveys, as well as the average point difference in the result. We have splitted the data in the cases where the participants passed the level and the cases where they missed it. Since none of the participants missed the first level, we did not include it in the graph.



(a)



(b)



(c)

Figure 8: For each level in the game, (a) average attention exerted and (b) average scoreboard result equity reported by participants in the level surveys, and (c) result point difference achieved. The blue bars refer to the cases in which the participants passed the level, meanwhile the red bars correspond to the cases where the participants missed the level. We calculated the result point difference as the absolute value of the difference between the points scored by the participants and the points scored by the computer.

Figure 9 presents the average ratings provided by the participants in the game experience survey to evaluate the different aspects of the game.

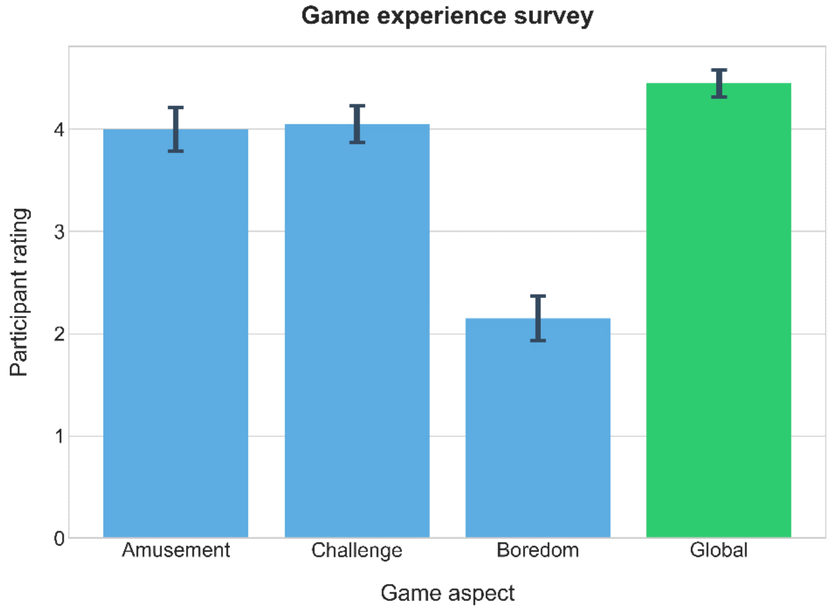


Figure 9: Game experience survey scores. The blue bars refer to the scores that the participants provided to evaluate the three aspects considered in the game experience survey, namely, amusement, challenge and boredom. The green bar represents the score that the participants rendered to evaluate the global experience. The error bars indicate the standard error of the mean (SEM).

Table 1 presents the results of the special level that we engineered to assess the robustness of the attention detection.

Table 1: Results from the robustness assessment level. In this level, one of the participants played for one hundred points. In half of those points, the enemy attacks were presented as the already mentioned checkerboard that inverted its pattern at a rate of 15 Hz. For the other half, the attacks were presented as checkerboards that flickered at 10 Hz. The order in which these two kinds of stimuli were presented was randomized to avoid participant acknowledgement and the different flickering frequencies were visually unnoticeable to the participants. The cases we considered as misclassifications are shaded in red and those we considered as correct classifications are shaded in green.

	Participant points	**Computer points**
10 Hz stimuli	0	50
15 Hz stimuli	34	16

Table 2 shows the outcomes of the statistical tests that we carried out in this study. For the first test (first row of the table), we compared the scores reported by the participants in the level surveys (attention exerted and scoreboard result equity) when they passed the levels with respect to when they missed. To this end we applied the Wilcoxon-Mann-Whitney test. For the last test (second row of the table), we splitted the participants into three groups according to their gaming experience (null, medium and high experience), and compared their performance in terms of the result point difference they achieved during the game. In this case, we employed the Kruskal-Wallis test. Asterisks (*) denote significant differences between the samples compared. For all tests, the significance level was set to 0.05 ($\alpha = 0.05$).

Table 2: *P*-values obtained from the statistical tests performed in the study.

Samples compared	***p*-value**
Attention reported when passing vs when missing	$17 \cdot 10^{-5}$ *
Point difference for the three groups of participants	0.33

4. Discussion

The goal of this study was to assess the feasibility of the SSVEP-based game that we implemented to train the attention. To attain this, we designed our application to be attractive, in such a way that the participants would find it enjoyable, and challenging, and they would feel a progressive demand throughout the game that kept their involvement. After twenty participants played the game, the preliminary results we obtained support the appropriateness of our game to train the attention.

Figure 5 represents the total time that participants spent, in average, in each level. This chart clearly shows that as the levels elapse, the amount of time the participants need to pass a level is higher, what is indicative of a progressively higher difficulty. Similarly, Figure 6 displays the average number of points scored by the participants and the computer, and Figure 8 (c) shows the point difference obtained when passing and when missing, in each level. From these graphs it can be appreciated that as the game progresses, the participants tend to score less points while the computer tends to score more, what again supports the fact that the game gradually augments the demand on the participants. Finally, Figure 7 shows the success rate of the participants in each level. While for the first level the success rate is 100%, this rate progressively decreases over the gameplay, and reaches a value around 38% for the last level. This confirms the potential of the game to increasingly challenge the participants.

Figure 8 (a) shows the self-perceived attention exerted by the participants in each level when they passed and when they missed. From this graph it is noticeable that when the participants passed a level, they applied a higher level of attention to the stimuli than when they missed. We evaluated this behavior through the comparison of the attention that the participants reported when they passed against the attention they provided when they missed according to the level surveys. To this end, we applied the Wilcoxon-Mann-Whitney test, and found significant differences between the two samples (see Table 2). This evidences that the attention expended by the participants acted as the game joystick. Furthermore, according to Figure 8 (b), when the participants passed a level as well as when they missed, they reported a fairly high rating for the equity of the scoreboard result (above 3 out of 5 for all cases but one). These two facts support the engaging capability of our application, since the players were aware that their attention was the only responsible for the result.

When we compared the performance of the participants, grouped into three clusters by their gaming experience, the Kruskal-Wallis test did not yield any significant differences between the samples. This result suggests that the performance is not influenced by the gaming expertise of the player, and thus the game could be equally exploited by a broad audience.

In relation to the attention detection process, Table 1 shows the classification of the one hundred points that comprised the special level described in the last paragraph of subsection 2.5. As we anticipated, none of the 50 mobile visual stimuli that flickered at a frequency of 10 Hz were classified as points for the participant. Conversely, from the 50 mobile visual stimuli that flickered at a frequency of 15 Hz, 34 were classified as points for the participant and 16 as points for the computer. The outcome of this test is therefore two-fold: first, the system correctly assesses the attention of the participants to the mobile visual stimuli flickering at 15 Hz that models the enemy attacks, and thus the decisions made by the server are not random; and second, the accuracy of the system is in the range of (0.84, 1) since 84 points were correctly classified, and from those 16 points that were classified as points for the computer despite corresponding to 15 Hz stimuli, a fraction of them could obey to a lack of attention from the participant or to the low quality of the EEG.

With reference to the game experience survey, whose results are reported in Figure 9, the participants found the game very amusing and challenging (mean ratings of 4 out of 5) and they experienced little boredom (mean rating close to 2 out of 5). Finally, the participants widely enjoyed the global experience that the game offered (mean rating higher than 4 out of 5). In light of these results, we can state that our approach has a clear recreational component and it motivates the participants to train their attention capability, thus it can be considered as a serious game for attention training.

Lastly, although there are other works in the recent literature that explored the use of SSVEPs with a moving stimulus (the so-called SSMVEP), they are mostly constrained to spelling tasks (Han et al., 2018; P. Stawicki et al., 2019; Piotr Stawicki & Volosyak, 2020) or to testing the suitability of a particular kind of stimulus to elicit an SSVEP (Guitong et al., 2018; Yan et al., 2017; Zheng et al., 2020). For this reason, only an indirect comparison between these works and the study that we present in this paper is feasible. In this context, the works that we just referred aim to obtain quality SSVEPs while maximizing the comfort of the participant. To this extent, many of these studies have tested flicker-free visual stimuli with remarkable success. In these works, the traditional checkerboard or grating pattern inversion was replaced by different combinations of stimuli and movements such as circles with pendulum motion (P. Stawicki et al., 2019), checkerboard expansion-contraction (Yan et al., 2017) or square ring motion reversal (Guitong et al., 2018). However, the movements described in these studies are normally restricted to one axis and the participants need to focus on a fixed point of the screen (usually corresponding to the center of the stimulus or a fixation cue). Conversely, the stimulus we engineered in this work reproduces a three-axis movement. Indeed, to the best of our knowledge, this is the first stimulus that performs a movement in the two axes of the screen while at the same time augments its size. In addition, in our work the participants must follow the stimulus fixation cue while it travels, what enhances the visual tracking demands. With respect to classification, a frequent and successful approach in the SSMVEP and SSVEP literature is to apply the canonical correlation analysis (CCA), as the goal of these kind of works is often to identify what of the stimuli presented in the screen the participant is focusing. In the case of this study, we developed a much simpler detector based on signal processing and a spectral power threshold, since we only considered one type of stimulus.

5. Conclusion

In this work, we proposed an attention-driven videogame to train the attention capabilities of the participants. We modeled the enemy attacks as mobile visual stimuli flickering at 15 Hz that elicited SSVEPs of the same frequency in the brain activity of the participants. We designed the game levels to present an increasing difficulty by placing the enemies closer to the foreground of a virtual scenario and boosting the speed of the stimuli as the levels progressed. The results we obtained from the level surveys, as well as those yielded from the performance of the participants, backup the appropriateness of our approach to demand attentional efforts from the players, and thus to train their attention. In the game experience survey, our application was described as amusing and challenging, what may support its potential to be applied in centers dedicated to the treatment of attentional disorders. However, further studies must assess the acceptance of this videogame by players suffering from this kind of issues. In the future, we plan to include some additional features to increase the usability of the game. Among others, we will: (i) add more than one mobile visual stimulus at a time, (ii) implement a collaborative gameplay that allows several participants to play at the same time, and (iii) explore the use of other kinds of SSMVEP stimuli to minimize visual fatigue.

Acknowledgements

This research was funded by the Ministry of Economy and Competitiveness (Spain) grant number [TIN2015-67020P], the Junta of Andalucía (Spain) grant number [P11-TIC-7983], the Spanish Nation-al Youth Guarantee Implementation Plan, the Association Nicolo for the R+D in Neurotechnology for the disability. The authors would like to thank all the volunteers who participated in the study.

Conflict of interest

The authors declare no conflicts of interest.

References

1. Ahluwalia, M. A., & Vold, S. D. (n.d.). *Visual Evoked Potentials and Glaucoma*. 2.
2. Brennan, C., McCullagh, P., Lightbody, G., Galway, L., Feuser, D., González, J. L., & Martin, S. (2015). Accessing Tele-Services Using a Hybrid BCI Approach. In I. Rojas, G. Joya, & A. Catala (Eds.), *Advances in Computational Intelligence* (Vol. 9094, pp. 110–123). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-19258-1_10
3. Finke, A., Lenhardt, A., & Ritter, H. (2009). The MindGame: A P300-based brain–computer interface game. *Neural Networks*, *22*(9), 1329–1333. https://doi.org/10.1016/j.neunet.2009.07.003
4. Gregori Grgič, R., Calore, E., & de'Sperati, C. (2016). Covert enaction at work: Recording the continuous movements of visuospatial attention to visible or imagined targets by means of Steady-State Visual Evoked Potentials (SSVEPs). *Cortex*, *74*, 31–52. https://doi.org/10.1016/j.cortex.2015.10.008
5. Guitong, L., Zhimin, Z., Xiaoke, C., Yangting, L., Tengyu, Z., Haijun, N., & Yubo, F. (2018). Study of Steady State Motion Visual Evoked Potential-based Visual Stimulation Paradigm. *Proceedings of the 12th International Convention on Rehabilitation Engineering and Assistive Technology*, 17–20. Midview City, SGP: Singapore Therapeutic, Assistive & Rehabilitative Technologies (START) Centre.
6. Gürkök, H., Nijholt, A., Poel, M., & Obbink, M. (2013). Evaluating a multi-player brain–computer interface game: Challenge versus co-experience. *Entertainment Computing*, *4*(3), 195–203. https://doi.org/10.1016/j.entcom.2012.11.001
7. Han, C., Xu, G., Xie, J., Chen, C., & Zhang, S. (2018). Highly Interactive Brain–Computer Interface Based on Flicker-Free Steady-State Motion Visual Evoked Potential. *Scientific Reports*, *8*(1), 5835. https://doi.org/10.1038/s41598-018-24008-8
8. Hassanpour, A., Moradikia, M., Adeli, H., Khayami, R., & Shamsinjead Babaki, P. (2019). A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals. In *Expert Systems* (Vol. 36). https://doi.org/10.1111/exsy.12494
9. International Society for Clinical Electrophysiology of Vision, Odom, J. V., Bach, M., Brigell, M., Holder, G. E., McCulloch, D. L., … Tormene, A. P. (2016). ISCEV standard for clinical visual evoked potentials: (2016 update). *Documenta Ophthalmologica*, *133*(1), 1–9. https://doi.org/10.1007/s10633-016-9553-y
10. Kelly, S. P., Lalor, E. C., Finucane, C., McDarby, G., & Reilly, R. B. (2005). Visual Spatial Attention Control in an Independent Brain-Computer Interface. *IEEE Transactions on Biomedical Engineering*, *52*(9), 1588–1596. https://doi.org/10.1109/TBME.2005.851510
11. Krepki, R., Blankertz, B., Curio, G., & Müller, K.-R. (2007). The Berlin Brain-Computer Interface (BBCI) – towards a new communication channel for online control in gaming applications. *Multimedia Tools and Applications*, *33*(1), 73–90. https://doi.org/10.1007/s11042-006-0094-3
12. Lim, J.-H., Lee, J.-H., Hwang, H.-J., Kim, D. H., & Im, C.-H. (2015). Development of a hybrid mental spelling system combining SSVEP-based brain–computer interface and webcam-based eye tracking. *Biomedical Signal Processing and Control*, *21*, 99–104. https://doi.org/10.1016/j.bspc.2015.05.012
13. Lopez, Miguel A., Pelayo, F., Madrid, E., & Prieto, A. (2009). Statistical Characterization of Steady-State Visual Evoked Potentials and Their Use in Brain–Computer Interfaces. *Neural Processing Letters*, *29*(3), 179–187. https://doi.org/10.1007/s11063-009-9102-8
14. Lopez, Miguel Angel, Pomares, H., Damas, M., Prieto, A., & de la Plaza Hernandez, E. M. (2007). Use of Kohonen Maps as Feature Selector for Selective Attention Brain-Computer Interfaces. In J. Mira & J. R. Álvarez (Eds.), *Bio-inspired Modeling of Cognitive Tasks* (pp. 407–415). Springer Berlin Heidelberg.

15. Lopez-Gordo, M. A., & Pelayo, F. (2013). A BINARY PHASE-SHIFT KEYING RECEIVER FOR THE DETECTION OF ATTENTION TO HUMAN SPEECH. *International Journal of Neural Systems*, *23*(04), 1350016. https://doi.org/10.1142/S0129065713500160
16. Lopez-Gordo, M. A., Pelayo, F., Fernandez, E., & Padilla, P. (2015). Phase-shift keying of EEG signals: Application to detect attention in multitalker scenarios. *Signal Processing*, *117*, 165–173. https://doi.org/10.1016/j.sigpro.2015.05.004
17. Lopez-Gordo, M. A., Perez, E., & Minguillon, J. (2019). Gaming the Attention with a SSVEP-Based Brain-Computer Interface. *Understanding the Brain Function and Emotions*, 51–59. Springer, Cham. https://doi.org/10.1007/978-3-030-19591-5_6
18. Minguillon, J., Lopez-Gordo, M. A., Renedo-Criado, D. A., Sanchez-Carrion, M. J., & Pelayo, F. (2017). Blue lighting accelerates post-stress relaxation: Results of a preliminary study. *PLOS ONE*, *12*(10), e0186399. https://doi.org/10.1371/journal.pone.0186399
19. Minguillon, J., Perez, E., Lopez-Gordo, M. A., Pelayo, F., & Sanchez-Carrion, M. J. (2018). Portable System for Real-Time Detection of Stress Level. *Sensors*, *18*(8), 2504. https://doi.org/10.3390/s18082504
20. Pires, G., Torres, M., Casaleiro, N., Nunes, U., & Castelo-Branco, M. (2011, November). *Playing Tetris with non-invasive BCI*. 1–6. IEEE. https://doi.org/10.1109/SeGAH.2011.6165454
21. Russo, F. D., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2003). Steady-State VEP and Attentional Visual Processing. In *The Cognitive Electrophysiology of Mind and Brain* (pp. 259–274). Elsevier. https://doi.org/10.1016/B978-012775421-5/50013-3
22. Sheppard, E., Birca, A., Carmant, L., Lortie, A., Vannassing, P., Lassonde, M., & Lippé, S. (2013). Children with a history of atypical febrile seizures show abnormal steady state visual evoked potential brain responses. *Epilepsy & Behavior*, *27*(1), 90–94. https://doi.org/10.1016/j.yebeh.2012.12.023
23. Shibata, K., Yamane, K., Otuka, K., & Iwata, M. (2008). Abnormal visual processing in migraine with aura: A study of steady-state visual evoked potentials. *Journal of the Neurological Sciences*, *271*(1–2), 119–126. https://doi.org/10.1016/j.jns.2008.04.004
24. Stawicki, P., Rezeika, A., Saboor, A., & Volosyak, I. (2019). Investigating Flicker-Free Steady-State Motion Stimuli for VEP-Based BCIs. *2019 E-Health and Bioengineering Conference (EHB)*, 1–4. https://doi.org/10.1109/EHB47216.2019.8969970
25. Stawicki, Piotr, & Volosyak, I. (2020). Comparison of Modern Highly Interactive Flicker-Free Steady State Motion Visual Evoked Potentials for Practical Brain–Computer Interfaces. *Brain Sciences*, *10*(10). https://doi.org/10.3390/brainsci10100686
26. Vaquero-Blasco, M. A., Perez-Valero, E., Lopez-Gordo, M. A., & Morillas, C. (2020). Virtual Reality as a Portable Alternative to Chromotherapy Rooms for Stress Relief: A Preliminary Study. *Sensors*, *20*(21), 6211. https://doi.org/10.3390/s20216211
27. Walter, S., Quigley, C., Andersen, S. K., & Mueller, M. M. (2012). Effects of overt and covert attention on the steady-state visual evoked potential. *Neuroscience Letters*, *519*(1), 37–41. https://doi.org/10.1016/j.neulet.2012.05.011
28. Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, *113*(6), 767–791. https://doi.org/10.1016/S1388-2457(02)00057-3
29. Yan, W., Xu, G., Li, M., Xie, J., Han, C., Zhang, S., … Chen, C. (2017). Steady-State Motion Visual Evoked Potential (SSMVEP) Based on Equal Luminance Colored Enhancement. *PLOS ONE*, *12*(1), e0169642. https://doi.org/10.1371/journal.pone.0169642
30. Yin, E., Zhou, Z., Jiang, J., Yu, Y., & Hu, D. (2015). A Dynamically Optimized SSVEP Brain Computer Interface (BCI) Speller. *IEEE Transactions on Biomedical Engineering*, *62*(6), 1447–1456. https://doi.org/10.1109/TBME.2014.2320948

31. Zheng, X., Xu, G., Wu, Y., Wang, Y., Du, C., Wu, Y., … Han, C. (2020). Comparison of the performance of six stimulus paradigms in visual acuity assessment based on steady-state visual evoked potentials. *Documenta Ophthalmologica*, *141*(3), 237–251. https://doi.org/10.1007/s10633-020-09768-x

Appendix C

Quantitative assessment of stress through EEG during a Virtual Reality stress-relax session

Frontiers in

Computational Neuroscience

Quantitative assessment of stress through EEG during a Virtual Reality stress-relax session

Eduardo Perez-Valero[1,2], Miguel A. Vaquero-Blasco[2,3], Miguel A. Lopez-Gordo[2,3*], and Christian Morillas[1,2]

[1]Department of Computer Architecture and Technology, University of Granada, Spain

[2]Research Centre for Information and Communications Technologies, University of Granada, Spain

[3]Department of Signal Theory, Telematics and Communications, University of Granada, Spain

Correspondence:
Corresponding Author
malg@ugr.es

Keywords: EEG, stress, regression, machine learning, virtual reality.

Abstract

Recent studies have addressed stress level classification via electroencephalography (EEG) and machine learning. These works typically use EEG-based features, like power spectral density (PSD), to develop stress classifiers. Nonetheless, these classifiers are usually limited to the discrimination of two (stress and no stress) or three (low, medium, and high) stress levels. In this study we propose an alternative for quantitative stress assessment based on EEG and regression algorithms. To this aim, we conducted a group of 23 participants (mean age 22.65 ± 5.48) over a stress-relax experience while monitoring their EEG. First, we stressed the participants via the Montreal imaging stress task (MIST), and then we led them through a 360-degree virtual reality (VR) relaxation experience. Throughout the session, the participants reported their self-perceived stress level (SPSL) via surveys. Subsequently, we extracted spectral features from the EEG of the participants and we developed individual models based on regression algorithms to predict their SPSL. We evaluated stress regression performance in terms of the mean squared percentage error (MSPE) and the correlation coefficient (R^2). The results yielded from this evaluation (MSPE = 10.62 ± 2.12, R2 = 0.92 ± 0.02) suggest that our approach predicted the stress level of the participants with remarkable performance. These results may have a positive impact in diverse areas that could benefit from stress level quantitative prediction. These areas include research fields like neuromarketing, and training of professionals such as surgeons, industrial workers, or firefighters, that often face stressful situations.

Introduction

Psychological stress is one of the most frequent human affections and has a considerable impact in modern society. According to the 2014 American Psychological Association stress report, about 75% of the population has suffered symptoms linked to stress [1], and employers spend 300 billion dollars each year in concept of stress related therapies and missed work. Main causes behind stress include job pressure, money, and health condition. In this context, since mental stress is highly correlated with medical issues such as headaches, depression, or insomnia [2], stress detection and quantification have recently become active subjects of research.

Recent studies have focused on the early detection of stress in order to prevent more severe health issues. For that purpose, studies generally address stress level detection through multiple biomarkers. For instance, several works have used EEG-based features to develop accurate machine learning stress-level classifiers [3]–[6]. In these works, researchers typically extract time-frequency features from EEG recordings and split them into training and test sets. Then, they train a classifier using the training set and validate it on the test set. These approaches often achieve remarkable performance, for instance, in [3] and [4], the authors obtained accuracies of 80.32% and 95.06% for a two-level and three-level stress classifier, respectively. Nevertheless, these works restrict the target space to two or three stress levels, corresponding to low, medium, and high stress. Additionally, performance report is very heterogeneous across studies. Indeed, some of these works only report classification accuracy, overlooking intra-class performance, and they typically develop a single prediction model for all the participants, disregarding participant distinctiveness of brain activity. Alternatively, other works have addressed stress assessment through regression [7]–[9]. Conversely to stress classification, these approaches do not aim to detect two or three stress levels, but to examine stress as a continuous variable. However, many of these works only assessed the linear relationship between certain biomarkers and SPSL [8], [10], while others presented preliminary quantitative predictions of stress level that yielded middling results (for instance, in [7] the authors obtained a correlation coefficient of 0.64, and in [11] the authors did not provide a scoring metric for their regression predictions, although graphic representation of the predicted and actual stress values showed poor performance).

In this paper, our aim is to develop an individualized quantitative stress prediction model using EEG activity and regression algorithms. To achieve this, we developed an individual regression model for each of the participants that predicted their SPSL from their EEG spectral features. To this aim, we conducted a group of participants through a stress-relax procedure. First, we stressed them by means of the Montreal imaging stress task (MIST) [12], a well-established methodology to elicit psychosocial stress [13]–[15], and then we conducted them through a virtual reality (VR) relaxation experience [16]. We recorded the brain activity of the participants across the session with an EEG acquisition system to obtain relevant EEG biomarkers. These biomarkers have been widely used in literature to evaluate stress, and include spectral power in different bands [17], [18], asymmetry in the Alpha band [19], [20] and Relative Gamma [21]–[24]. Additionally, we asked the participants to report their self-perceived stress level (SPSL) throughout the session according to an adaptation of the perceived stress scale (PSS) [25]. Finally, we used the SPSL surveys and the EEG biomarkers extracted from each participant to train and validate different regression algorithms, namely Ridge regression (RR), random forest (RF), multi-layer perceptron (MLP), and support vector regression (SVR), that are usually applied in stress prediction literature [8], [11], [18], [26].

To sum up, the objective of this work is to provide an individualized quantitative stress prediction model from EEG biomarkers and machine learning regression algorithms. To validate our proposal, we obtained regression performance metrics and we compared our approach with other approximations in literature. Lastly, we believe this work can impact fields like neuromarketing, gaming, and training of certain professionals dealing with stressful situations, as in all of these areas stress assessment may contribute positively.

Materials and Methods

2.1. Participants

Twenty-three healthy volunteers (14 females, 8 males, and 1 non-binary) were recruited two weeks prior to the onset of the study. The age of the participants ranged from 18 to 40 (mean age 22.65 ± 5.48). We only considered participants without health issues and mental disorders. Participants belonged to the community of the University of Granada, and they were not rewarded in any way for their participation. Before the beginning of the study, we briefed the participants about the different phases of the experimental procedure, and we required them not to consume any stimulant nor relaxant the day before the session. Every participant took part in a single 18-minute experimental session. The entire data capture spanned for approximately three weeks. Since the information derived from the participants was limited to the recording of their brain activity, ethical approval was not required for the present study in compliance with local institutional requirements. Nonetheless, all the participants signed an informed consent prior to the start of the study.

2.2. Experimental procedure

First, we asked the participants to carefully read and sign the informed consent. Then, we equipped them with an EEG cap to acquire their brain activity, and we briefed them about the tasks they had to perform during the experimental session. These tasks are presented in Figure 1.

Along the first phase, we conducted the participants through a two-minute resting state period. During this period, they remained relaxed with their eyes closed. We designed this phase to prompt a baseline relaxation level among the participants. After that, we stressed the participants via the MIST, a test designed to elicit psychosocial stress via arithmetical operations and social pressure. Subsequently, we led the participants through a five-minute relaxation phase. During this phase, the participants selected their preferred 360-degree virtual experience from a group of four experiences, and underwent it using a VR head mounted display (HMD). Available experiences included a cascade, an aurora borealis, a solitary beach, and a space trip. Throughout this phase, participants remained alone in the room while researchers monitored them from the outside. Finally, during the last phase of the experimental session, participants performed another two-minute resting state period with their eyes closed.

During the entire session, participants were equipped with an EEG acquisition system that recorded their EEG activity. Upon conclusion of the data capture, we extracted EEG features from these recordings to predict the stress level of the participants. Additionally, we conducted several surveys to obtain the SPSL of the participants throughout the experimental session (T1-T8 in Figure 1).

2.3. Experimental setup

Regarding the MIST, we implemented this test as a graphical interface in MATLAB R2016a (MathWorks). The MIST comprised a series of arithmetical operations including combinations of additions, multiplications, and divisions. To complete this phase, we briefed the participants to use only their dominant hand to work out the operations via the touchscreen of the laptop that displayed the MIST interface. We divided the MIST phase into a three-minute training and a six-minute test. Training was designed for the participants to familiarize with the MATLAB interface. During this phase, the participants did not have an imposed time limit to solve each operation, and they were not required to reach a baseline success rate. Conversely, during test, the participants had an imposed time limit to solve each operation that was displayed as a progress bar in the MATLAB interface. Moreover, after each operation, the interface displayed an additional progress bar reporting their success rate. To induce more stress on the participants, researchers required the participants to reach an unfeasible success rate during the test. Additionally, researchers joined the room in three occasions during the test to verbally stress the participant. In relation to the SPSL surveys conducted during this phase (T2-T4), we integrated them in the MIST interface so the participants could answer using the touchscreen of the laptop.

With respect to the relaxation phase, we implemented each virtual experience using a 360-degree video and Unity engine, and then we utilized the Oculus Quest HMD to reproduce the virtual experiences. In the same way we did for the MIST phase, we embedded surveys T5-T7 in the virtual experiences so the participants were able to answer through the HMD controller. Before the start of the relaxation phase, the participants selected one of the four different virtual experiences available, namely a cascade, an aurora borealis, a solitary beach, and a space trip.

In relation to the EEG acquisition, we used the Versatile EEG system, a semi-dry acquisition device designed by Bitbrain (Spain), that works at a sampling rate of 256 Hz. For the electric setup, we considered four electrodes located at positions Fp1, Fp2, F5, and F6 of the 10-20 International System based on previous successful studies about emotions assessment [27]–[29]. The electrodes were referenced at the left ear lobe and grounded to an extra electrode equidistant from Fpz and Fz. For all the participants, we acquired an independent EEG recording for each phase of the experimental session.

Lastly, regarding the SPSL surveys, we implemented a customized version of the perceived stress scale (PSS) to minimize the time required for the participants to answer. In particular, each survey comprised a single question: “What is your level of stress in a scale from 1 to 5, if 1 is the minimum level and 5 is the maximum level?”. We performed the surveys at eight key points of the experimental session to gather a suitable sample of the SPSL of the participants. Concretely, we rendered the surveys as follows: T1, at the end of the initial resting state; T2-T3, 120 and 240 seconds past the onset of the MIST test; T4, at the end of the MIST test; T5-T6, 90 and 180 seconds after the start of the relaxation phase; T7, at the end of the relaxation phase; T8, at the end of the second resting state. We did not conduct a survey after the MIST training as we implemented this phase exclusively to brief the participants about the use of the MIST interface.

2.4. Signal processing

Since we acquired a separate EEG recording for every phase of the study, we applied the processing pipeline described in Figure 2 to each of these recordings. Additionally, we withdrew the recordings corresponding to the first resting state, since we intended this phase to induce a baseline relaxation state only, and the MIST training, as we designed this phase only to instruct the participants on the use of the MATLAB interface. In total, the processed recordings spanned for 12 minutes corresponding to the MIST test (six minutes), the relaxation phase (five minutes), and the final resting state (we considered only the central minute of this phase).

To process the EEG recordings, first we applied a notch filter with stopband 48-52 Hz to remove electric coupling. Then, we applied a 4^{th} order zero-phase shift Butterworth bandpass filter with bandpass 2-48 Hz to retain the relevant EEG spectral content. Furthermore, it is worth to note that the duration of the EEG recordings of the stages of the study differed by a few seconds across participants. The reason behind is that we marked the onset and the ending of the EEG recordings manually. To overcome this issue, we resampled the filtered signals at a sampling rate such that the length of the resampled signals matched the theoretical length of the stages described in Section 2.2. Thereafter, we splitted the filtered signals into 2-second epochs. We tagged the epochs above an amplitude of 75 µV as outliers and zeroed them. We selected this amplitude limit after visual inspection of the signals in accordance to prior studies [30], [31]. Then, we detrended and standardized every epoch. To detrend, we removed the best straight-line fit for the epoch from each epoch value. To standardize, we subtracted the epoch mean from each epoch value and divided it by the epoch standard deviation. Subsequently, we carried out a feature extraction phase. In this phase, we estimated the channel-averaged power spectral density (PSD) in Delta (1-4 Hz), Theta (4-8 Hz), Alpha (8-13 Hz), Beta (13-25 Hz), and Gamma (25-45 Hz) bands, the relative Gamma (RG), and the Alpha asymmetry (AA). Prior to the estimation of the spectral power, we applied a Tukey window to ease edge effects. Formulae for RG and AA are presented in Equations 1-2. Lastly, we applied a moving average filter with a 30-second window to smooth the extracted features. Figure 2 represents the main stages of the signal processing pipeline described in this paragraph. It is worth to note that, as a result of resampling the processed features had the same length for all the participants. This length was equal to 360 samples, corresponding to the 12-minute EEG recording splitted into 2-second epochs (see Equation 3). Consequently, we obtained a feature matrix with size 360 x 7 (360 samples, 7 features) per participant.

$$RG = \frac{P_{Gamma}}{P_{Alpha} + P_{Theta}} \quad (1)$$

$$AA = P_{Alpha\,(F6)} - P_{Alpha\,(F5)} \quad (2)$$

$$360\ epochs = \frac{60_{\frac{s}{min}} \cdot 12_{min}}{2_{\frac{s}{epoch}}} \quad (3)$$

RG, P, AA, F6, F5, s, and min stand for Relative Gamma, Power, Alpha asymmetry, F6 electrode positions, F5 electrode position, seconds, and minutes.

With regard to stress assessment, as outlined in subsection 2.3., we performed eight SPSL surveys at key instants of the experimental session. Since we planned to develop participant-level models, we had only eight target samples per participant at our disposal (T1-T8), in contrast to the 360-sample feature matrix. To overcome this, we interpolated the SPSL surveys of the participants to match the number of samples of the feature matrix. Particularly, we implemented four interpolation methods, namely, linear, pchip (cubic), spline, and nearest interpolation. We applied this procedure based on the hypothesis that emotions cannot drastically change during a small period of time [32]. As an example, Figure 3 illustrates the different SPSL interpolation methods.

2.5. Model development and evaluation

To develop the SPSL regression models we used Scikit-learn toolbox for Python 3. We developed a separate model per each participant. To this end, for each participant, we considered a feature matrix holding 360 samples of the seven spectral features that we estimated, and a target array carrying 360 SPSL samples (eight true SPSL survey answers plus the interpolated data). See Figure 4 for a visual representation of these structures.

The procedure that we followed to develop and evaluate the regression models is described next. First, we extracted the training and test sets from feature matrix (X) and target array (y) of the participant. For the training set we selected the samples that corresponded to the SPSL points that we interpolated (352 samples). And for the test set we utilized the samples that corresponded to the SPSL answers provided by the participants (T1-T8). Then, we standardized the columns of the feature matrix to normalize the range of the features so they contributed proportionately. Next, we carried out the following procedure for each of the four regressors that we considered in the study, namely RR, RF, MLP, and SVR: we performed grid search cross-validation on the training set to find the best set of hyperparameters for the regressor. During this procedure, we explored multiple combinations of hyperparameters and applied 5-fold cross validation to each combination using the training set. After ensuring cross-validation results were not subject to overfitting nor underfitting, we obtained the best set of hyperparameters and we trained the regressor with those hyperparameters and using all the data in the training set. Finally, we evaluated the regressor on the test set in terms of the mean squared percentage error (MSPE) and the Pearson correlation coefficient (R^2). We examined the MSPE as this metric provides a comprehensible measure of the error that does not depend on the amplitude of the target (see Equation 4). In addition, we examined R^2 to assess the correlation between the real stress values and the predictions. We followed the procedure described in this paragraph to train and evaluate a model for each participant. For a visual representation of this procedure, refer to Figure 5.

$$MSPE = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(y_{test_i} - y_{pred_i} \right)^2}{\frac{1}{n} \cdot \sum_{i=1}^{n} y_{test_i}} \cdot 100 \quad (4)$$

Notice that, since we considered four regressors, we explored different hyperparameters for each of them during grid search cross-validation. The list of hyperparameters that we examined for each regressor is provided in Table 1. For reproducibility, we set random_state variable to zero for all Scikit-learn functions.

Results

Table 2 presents the performance yielded by the different regressors for all the SPSL interpolation methods that we examined in this study.
Figure 6 displays the predictions for each of the participants yielded by the best SPSL regression model.

Figure 7 represents the average absolute error of the predictions at each survey instant for the best regression model.

Discussion

The goal of this study was to provide a quantitative assessment of the stress level through individualized regression algorithms. To this end, we gathered the SPSL of a group of participants throughout a stress-relax procedure, and we evaluated different regressors based on EEG spectral features. Instead of fitting a single regressor to the data from all the participants, we fitted a regressor per participant. The main reason behind this argument is the well-known inter and intra-participant particularities observed in EEG due to non-stationary brain activity. The results that we obtained indicate that regression models can quantitatively predict stress level with noteworthy performance. On the one side, the high correlation coefficient (0.92) obtained by the best-performing regressor (random forest) evidences the ability of this model to track the progression of the SPSL throughout a stress-relax session. Moreover, the low MSPE (10.62) yielded by this regressor demonstrates its capability to produce quantitative predictions of the SPSL with remarkable performance, what fulfills our expectations in regards to stress assessment.

According to Table 2, random forest regression from linearly interpolated SPSL data yielded the best performance in terms of MSPE (10.62) and R^2 (0.92). Figure 6 evidences that, in general, random forest regression accurately predicted the SPSL values for all the participants. Indeed, this regression model predicted very different SPSL patterns with remarkable performance, from steady progressions such as participant 13, to more intricate progressions such as participants 9 and 5. The notable performance of this model is further evidenced in Figure 7. This figure indicates that the mean absolute error among the different surveys was lower than 0.6 for all surveys. It is also notable that the higher error occurred at T4, that corresponds to the last survey of the MIST test phase. This may be explained by the fact that participants often reached their peak SPSL at the end of the MIST, while for the rest of the surveys the SPSL followed a steadier progression.

The results that we obtained in this study support other works that assessed the feasibility of predicting stress from EEG features. Most part of these studies addressed stress prediction as a classification task. Nevertheless, this approach is limited to the detection of two or three classes (typically low stress, medium stress, and high stress) [17], [33], [34]. Furthermore, performance is usually non-uniformly reported among studies, with some authors often reporting only classification accuracy, hence disregarding class balance [35], [36]. As a result, the outcomes of these studies are hardly comparable. Alternatively, we approached stress prediction as a regression task to broad the target space to continuous values of stress, and provide a more interpretable performance metric (prediction percentage error). In this context, other authors have evaluated stress level regression from features such as EEG spectral power or heart rate variability (HRV). For instance, in [7], authors assessed self-perceived stress predictions through linear regression of an HRV spectral feature. However, they did not evaluate the linearity between the predictor and the target, and restricted their analysis to a scatter plot. As a result, the model yielded poor performance for some of the participants. Furthermore, they computed accuracy after scaling the prediction error by the maximum value of the PSS scale (40), although some of the participants did not reach a level of stress even close to this value, what led to unrealistic high accuracy for some of them. In contrast, we normalized the error by the average value of the SPSL, what results in a more realistic performance report (See Equation 4). Other authors have assessed stress level regression from EEG features, nevertheless, they often report their results in terms of R^2, hence disregarding the quantitative error between the predictions and the real stress levels. For instance, in [10] and [11], the authors used a regression model to predict stress level, however, they only reported performance in terms of R^2 (0.13 for the first study, and close to 0.80 for the second). Conversely, we supported the correlation score provided by R^2 with an error score like the MSPE (although other metrics like the mean squared error of the mean absolute error may be considered). Additionally, in [11], although the authors applied a non-linear regression model (random forest), they did not report a procedure for estimating the hyperparameters of the regressor. Such a procedure is crucial to optimize the prediction capabilities of the model. We applied grid search cross-validation to this end, but other methods like random search also achieve remarkable performance at hyperparameter optimization [37].

It is worth noting that we designed the SPSL surveys to minimize the interferences during the experimental session. However, a longer session that included a higher number of surveys may improve stress predictions, as the dataset would rely on interpolated data to a lower extent compared to the present study. In addition, although we believe our results are good enough to support our conclusions, increasing the number of participants would likely reduce the SEM of the predictions.

Finally, multiple fields could take advantage of the stress regression approach that we propose in this study, as it only requires an 18-minute calibration to gather data to train the participant-individualized model and few frontal electrodes. For instance, emotion prediction based on psychophysiological measures is a conventional target in neuromarketing [38], [39]. In this scenario, an unobtrusive approach like the one presented in this paper may provide insights about the decisions made by consumers and their reactions to certain products. Alternatively, for professionals working under stressful conditions such as firefighters, industrial workers, or surgeons, stress quantitative assessment may be useful during training [40], [41]. Thereby, individual stress models of the trainees could provide the instructors valuable stress biofeedback that may be used to support education techniques. Moreover, our approach could be also accommodated in gaming, where ubiquitous biofeedback is gaining popularity recently [42], [43]. In that direction, quantitative predictions of the stress felt by the players may be utilized to anticipate their actions or modulate game difficulty. Lastly, further studies must evaluate the use of different electric setups and alternative EEG features, such as coherence among different brain areas and synchronization, as these aspects may enhance the prediction capabilities of the stress regression models.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Conceptualization, E. P.-V. and M.A.L.-G.; methodology, E.P.-V, M.A.V.-B., and M.A.L.-G.; software, E. P.-V and M.A.V.-B.; validation, E.P.-V., M.A.V.-B., M.A.L.-G., and C.M.; formal analysis, E.P.-V. and M.A.V.-B.; investigation, E.P.-V. and M.A.V.-B.; resources, C.M. and M.A.L-G.; data curation, E. P.-V.; writing—original draft preparation, E. P.-V and M.A.V.-B.; writing—review and editing, E.P.-V., M.A.V.-B., M.A.L.-G. and C.M.; visualization, E.P.-V.; supervision, M.A.L.-G. and C.M.; project administration, M.A.L.-G. and C.M. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the project PGC2018-098813-B-C31 (the Spanish Ministry of Science, Innovation and Universities, by European Regional Development Funds.

Acknowledgments

The authors would like to thank all the volunteers that participated in the study.

References

[1] N. B. Anderson *et al.*, “AMERICAN PSYCHOLOGICAL ASSOCIATION,” p. 23.

[2] K. Marksberry, “Stress Effects,” *The American Institute of Stress*. https://www.stress.org/stress-effects (accessed Mar. 02, 2021).

[3] H. Jebelli, S. Hwang, and S. Lee, “EEG-based workers’ stress recognition at construction sites,” *Automation in Construction*, vol. 93, pp. 315–324, Sep. 2018, doi: 10.1016/j.autcon.2018.05.027.

[4] A. Asif, M. Majid, and S. M. Anwar, “Human stress classification using EEG signals in response to music tracks,” *Computers in Biology and Medicine*, vol. 107, pp. 182–196, Apr. 2019, doi: 10.1016/j.compbiomed.2019.02.015.

[5] M. Zubair and C. Yoon, “Multilevel mental stress detection using ultra-short pulse rate variability series,” *Biomedical Signal Processing and Control*, vol. 57, p. 101736, Mar. 2020, doi: 10.1016/j.bspc.2019.101736.

[6] S. Lotfan, S. Shahyad, R. Khosrowabadi, A. Mohammadi, and B. Hatef, “Support vector machine classification of brain states exposed to social stress test using EEG-based brain network measures,” *Biocybernetics and Biomedical Engineering*, vol. 39, no. 1, pp. 199–213, Jan. 2019, doi: 10.1016/j.bbe.2018.10.008.

[7] J. Park, J. Kim, and S. Kim, “Prediction of Daily Mental Stress Levels Using a Wearable Photoplethysmography Sensor,” in *TENCON 2018 - 2018 IEEE Region 10 Conference*, Oct. 2018, pp. 1899–1902, doi: 10.1109/TENCON.2018.8650109.

[8] D. A. Dimitriev, E. V. Saperova, O. S. Indeykina, and A. D. Dimitriev, “Heart rate variability in mental stress: The data reveal regression to the mean,” *Data in Brief*, vol. 22, pp. 245–250, Feb. 2019, doi: 10.1016/j.dib.2018.12.014.

[9] R. Ahuja and A. Banga, “Mental Stress Detection in University Students using Machine Learning Algorithms,” *Procedia Computer Science*, vol. 152, pp. 349–353, Jan. 2019, doi: 10.1016/j.procs.2019.05.007.

[10] S. Muhammad Umar Saeed, S. Muhammad Anwar, and M. Majid, “Quantification of Human Stress Using Commercially Available Single Channel EEG Headset,” *IEICE Trans. Inf. & Syst.*, vol. E100.D, no. 9, pp. 2241–2244, 2017, doi: 10.1587/transinf.2016EDL8248.

[11] D. Das, T. Bhattacharjee, S. Datta, A. D. Choudhury, P. Das, and A. Pal, “Classification and quantitative estimation of cognitive stress from in-game keystroke analysis using EEG and GSR,” in *2017 IEEE Life Sciences Conference (LSC)*, Dec. 2017, pp. 286–291, doi: 10.1109/LSC.2017.8268199.

[12] K. Dedovic, R. Renwick, N. K. Mahani, V. Engert, S. J. Lupien, and J. C. Pruessner, “The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain,” *J Psychiatry Neurosci*, p. 7.

[13] S. Freitas, M. R. Simões, L. Alves, and I. Santana, “Montreal Cognitive Assessment: Validation Study for Mild Cognitive Impairment and Alzheimer Disease,” *Alzheimer Disease & Associated Disorders*, vol. 27, no. 1, pp. 37–43, 2013, doi: 10.1097/WAD.0b013e3182420bfe.

[14] K. Dedovic, R. Renwick, N. K. Mahani, V. Engert, S. J. Lupien, and J. C. Pruessner, “The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain,” *J Psychiatry Neurosci*, vol. 30, no. 5, pp. 319–325, Sep. 2005.

[15] A. Asif, M. Majid, and S. M. Anwar, “Human stress classification using EEG signals in response to music tracks,” *Computers in Biology and Medicine*, vol. 107, pp. 182–196, Apr. 2019, doi: 10.1016/j.compbiomed.2019.02.015.

[16] M. A. Vaquero-Blasco, E. Perez-Valero, M. A. Lopez-Gordo, and C. Morillas, “Virtual Reality Customized 360-degree Experiences for Stress Relief,” *Sensors*, vol. Manuscript in preparation, 2021.

[17] B. R. Schlink, S. M. Peterson, W. D. Hairston, P. König, S. E. Kerick, and D. P. Ferris, “Independent Component Analysis and Source Localization on Mobile EEG Data Can Identify Increased Levels of Acute Stress,” *Front. Hum. Neurosci.*, vol. 11, 2017, doi: 10.3389/fnhum.2017.00310.

[18] J. W. Ahn, Y. Ku, and H. C. Kim, “A Novel Wearable EEG and ECG Recording System for Stress Assessment,” *Sensors*, vol. 19, no. 9, Art. no. 9, Jan. 2019, doi: 10.3390/s19091991.

[19] R. Düsing, M. Tops, E. L. Radtke, J. Kuhl, and M. Quirin, “Relative frontal brain asymmetry and cortisol release after social stress: The role of action orientation,” *Biological Psychology*, vol. 115, pp. 86–93, Mar. 2016, doi: 10.1016/j.biopsycho.2016.01.012.

[20] C. W. E. M. Quaedflieg, F. T. Y. Smulders, T. Meyer, F. Peeters, H. Merckelbach, and T. Smeets, “The validity of individual frontal alpha asymmetry EEG neurofeedback,” *Social Cognitive and Affective Neuroscience*, vol. 11, no. 1, pp. 33–43, Jan. 2016, doi: 10.1093/scan/nsv090.

[21] M. A. Vaquero-Blasco, E. Perez-Valero, M. A. Lopez-Gordo, and C. Morillas, “Virtual Reality as a Portable Alternative to Chromotherapy Rooms for Stress Relief: A Preliminary Study,” p. 15, 2020, doi: 10.3390/s20216211.

[22] J. Minguillon, M. A. Lopez-Gordo, D. A. Renedo-Criado, M. J. Sanchez-Carrion, and F. Pelayo, “Blue lighting accelerates post-stress relaxation: Results of a preliminary study,” *PLOS ONE*, vol. 12, no. 10, p. e0186399, Oct. 2017, doi: 10.1371/journal.pone.0186399.

[23] S. R. Steinhubl *et al.*, “Cardiovascular and nervous system changes during meditation,” *Front. Hum. Neurosci.*, vol. 9, Mar. 2015, doi: 10.3389/fnhum.2015.00145.

[24] A. Lutz, L. L. Greischar, N. B. Rawlings, M. Ricard, and R. J. Davidson, “Long-term meditators self-induce high-amplitude gamma synchrony during mental practice,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 46, pp. 16369–16373, Nov. 2004, doi: 10.1073/pnas.0407401101.

[25] S. Cohen, T. Kamarck, and R. Mermelstein, “A global measure of perceived stress,” *J Health Soc Behav*, vol. 24, no. 4, pp. 385–396, Dec. 1983.

[26] R. Khosrowabadi, C. Quek, K. K. Ang, S. W. Tung, and M. Heijnen, “A Brain-Computer Interface for classifying EEG correlates of chronic mental stress,” in *The 2011 International Joint Conference on Neural Networks*, Jul. 2011, pp. 757–762, doi: 10.1109/IJCNN.2011.6033297.

[27] L. Brown, B. Grundlehner, and J. Penders, “Towards wireless emotional valence detection from EEG,” *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2011, pp. 2188–2191, 2011, doi: 10.1109/IEMBS.2011.6090412.

[28] M. A. Vaquero-Blasco, E. Perez-Valero, M. A. Lopez-Gordo, and C. Morillas, “Virtual Reality as a Portable Alternative to Chromotherapy Rooms for Stress Relief: A Preliminary Study,” *Sensors*, vol. 20, no. 21, Art. no. 21, Jan. 2020, doi: 10.3390/s20216211.

[29] R. Jenke, A. Peer, and M. Buss, “Feature Extraction and Selection for Emotion Recognition from EEG,” *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 327–339, Jul. 2014, doi: 10.1109/TAFFC.2014.2339834.

[30] C. Jeunet *et al.*, “Uncovering EEG Correlates of Covert Attention in Soccer Goalkeepers: Towards Innovative Sport Training Procedures,” *Scientific Reports*, vol. 10, no. 1, Art. no. 1, Feb. 2020, doi: 10.1038/s41598-020-58533-2.

[31] R. Cassani, T. H. Falk, F. J. Fraga, M. Cecchi, D. K. Moore, and R. Anghinah, “Towards automated electroencephalography-based Alzheimer’s disease diagnosis using portable low-density devices,” *Biomedical Signal Processing and Control*, vol. 33, pp. 261–271, Mar. 2017, doi: 10.1016/j.bspc.2016.12.009.

[32] J. B. Taylor, *My stroke of insight: a brain scientist’s personal journey*. Bloomington, IN: Jill Bolte Taylor, 2006.

[33] O. Attallah, “An Effective Mental Stress State Detection and Evaluation System Using Minimum Number of Frontal Brain Electrodes,” *Diagnostics*, vol. 10, no. 5, Art. no. 5, May 2020, doi: 10.3390/diagnostics10050292.

[34] J. Wijsman, B. Grundlehner, H. Liu, H. Hermens, and J. Penders, “Towards mental stress detection using wearable physiological sensors,” in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2011, pp. 1798–1801, doi: 10.1109/IEMBS.2011.6090512.

[35] H. Jebelli, S. Hwang, and S. Lee, “EEG-based workers’ stress recognition at construction sites,” *Automation in Construction*, vol. 93, pp. 315–324, Sep. 2018, doi: 10.1016/j.autcon.2018.05.027.

[36] S. Lotfan, S. Shahyad, R. Khosrowabadi, A. Mohammadi, and B. Hatef, “Support vector machine classification of brain states exposed to social stress test using EEG-based brain network measures,” *Biocybernetics and Biomedical Engineering*, vol. 39, no. 1, pp. 199–213, Jan. 2019, doi: 10.1016/j.bbe.2018.10.008.

[37] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” *J. Mach. Learn. Res.*, vol. 13, no. null, pp. 281–305, Feb. 2012.

[38] S. Kumar, “Neuromarketing: The New Science of Advertising,” *Universal Journal of Management*, vol. 3, pp. 524–531, Dec. 2015, doi: 10.13189/ujm.2015.031208.

[39] U. Garczarek-Bąk, A. Szymkowiak, P. Gaczek, and A. Disterheft, “A comparative analysis of neuromarketing methods for brand purchasing predictions among young adults,” *J Brand Manag*, vol. 28, no. 2, pp. 171–185, Mar. 2021, doi: 10.1057/s41262-020-00221-7.

[40] A. Oskooei, S. M. Chau, J. Weiss, A. Sridhar, M. R. Martínez, and B. Michel, “DeStress: Deep Learning for Unsupervised Identification of Mental Stress in Firefighters from Heart-Rate Variability (HRV) Data,” in *Explainable AI in Healthcare and Medicine: Building a Culture of Transparency and Accountability*, A. Shaban-Nejad, M. Michalowski, and D. L. Buckeridge, Eds. Cham: Springer International Publishing, 2021, pp. 93–105.

[41] K. L. Chrouser, J. Xu, S. Hallbeck, M. B. Weinger, and M. R. Partin, “The influence of stress responses on surgical performance and outcomes: Literature review and the development of the surgical stress effects (SSE) framework,” *Am J Surg*, vol. 216, no. 3, pp. 573–584, Sep. 2018, doi: 10.1016/j.amjsurg.2018.02.017.

[42] H. A. Osman, H. Dong, and A. E. Saddik, "Ubiquitous Biofeedback Serious Game for Stress Management," *IEEE Access*, vol. 4, pp. 1274–1286, 2016, doi: 10.1109/ACCESS.2016.2548980.

[43] B. Kerous, F. Skola, and F. Liarokapis, "EEG-based BCI and video games: a progress report," *Virtual Reality*, vol. 22, no. 2, pp. 119–135, Jun. 2018, doi: 10.1007/s10055-017-0328-x.

Table 1: Hyperparameters explored using grid search cross-validation for the different regressors examined in the study. RBF stands for Radial basis function.

Regressor	Hyperparameters	Values
Ridge regression	polynomial degree	1, 2, 3
	regularization penalty	10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10, 10^{2}
Random forest	number of estimators	100, 200, 300
	maximum number of features	5, 6, 7
	maximum depth	6, 7
Multi-layer perceptron	number of hidden layers	1, 2
	number of neurons per hidden layer	3, 4
	activation function	relu, tanh
	regularization penalty	10^{-3}, 10^{-2}, 10^{-1}
Support vector regression	kernel type	linear, poly, RBF, sigmoid
	regularization parameter	2^{-5}, 2^{-3}, 2^{-1}, 2, 2^{3}
	epsilon	10^{-3}, 10^{-2}, 10^{-1}

Table 2: Performance of the different regressors evaluated in the study. Columns display the average mean squared percentage error and Pearson correlation coefficient for each of the four interpolation methods applied. For both metrics we have included also the standard error of the mean (SEM). Rows represent the different regressors that we evaluated.

	Linear		**Pchip**		**Spline**		**Nearest**	
	MSPE	R^2	MSPE	R^2	MSPE	R^2	MSPE	R^2
RR	24.96 ± 3.61	0.76 ± 0.03	24.14 ± 3.35	0.76 ± 0.03	24.56 ± 3.41	0.76 ± 0.03	23.65 ± 3.26	0.77 ± 0.03
RF	**10.62 ± 2.12**	**0.92 ± 0.02**	11.11 ± 2.37	0.91 ± 0.02	11.24 ± 2.53	0.91 ± 0.02	11.97 ± 2.91	0.90 ± 0.02
MLP	23.09 ± 3.63	0.77 ± 0.03	22.20 ± 3.37	0.78 ± 0.02	20.14 ± 2.91	0.80 ± 0.02	23.56 ± 3.44	0.73 ± 0.04
SVR	18.97 ± 3.50	0.80 ± 0.04	18.67 ± 3.43	0.80 ± 0.04	17.84 ± 3.66	0.81 ± 0.04	21.46 ± 3.55	0.80 ± 0.03

Figure 1: Experimental procedure of the study. Initially, we conducted the participants through a two-minute eyes-closed resting state phase. Secondly, we trained the participants to learn how to use the MIST interface. Then, we conducted them through the MIST, a test designed to induce psychosocial stress via arithmetical operations and social pressure. Subsequently, the participants relaxed for five minutes via a VR experience. Finally, they completed another two-minute eyes-closed resting state period. The entire procedure spanned for approximately 18 minutes. Throughout the session, we performed multiple surveys to acquire the SPSL of the participants (T1-T8).

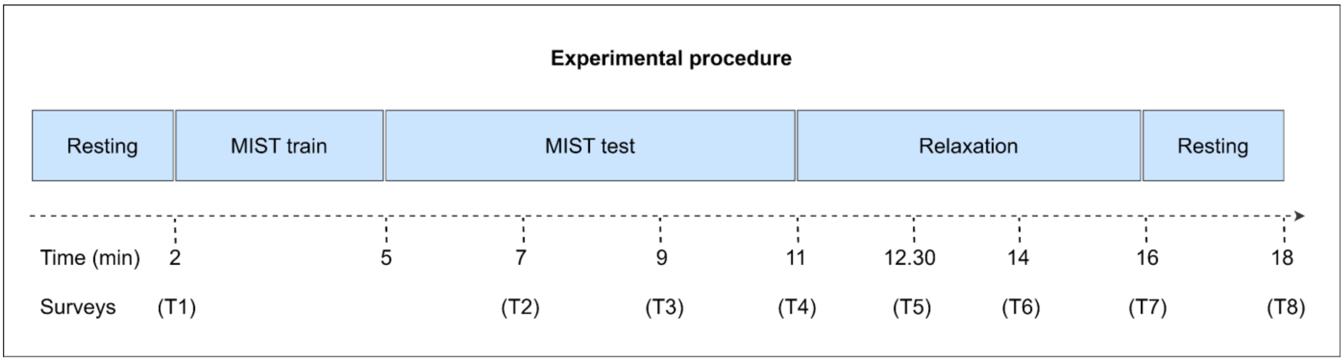


Figure 2: EEG processing pipeline. First, we applied a notch filter to dismiss the power line artifact. Next, we used a bandpass filter to retain only the desired EEG spectral content. Then, we resampled the filtered signal to match the expected span of the phase under processing. Subsequently, we divided the EEG signals into 2-second epochs. We performed artifact removal, detrending, and standardization at epoch level, and finally we extracted spectral features and smoothed them through a moving-average filter.

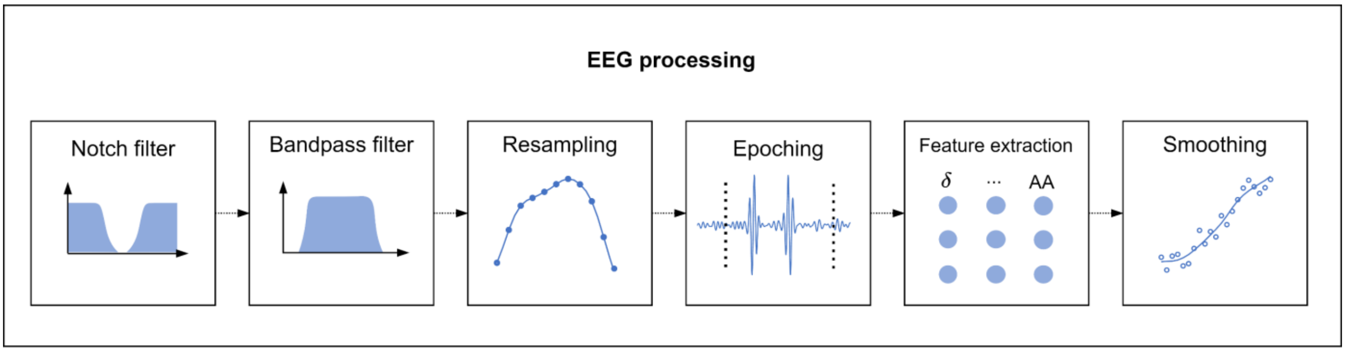


Figure 3: SPSL interpolation example. Blue dots represent the responses provided by the participant to the SPSL surveys (T1-T8). Dashed blue line corresponds to the interpolation of the SPSL values. Each graph represents one of the four interpolation schemes applied. **(A)** Linear interpolation, **(B)** pchip interpolation, **(C)** spline interpolation, and **(D)** nearest interpolation.

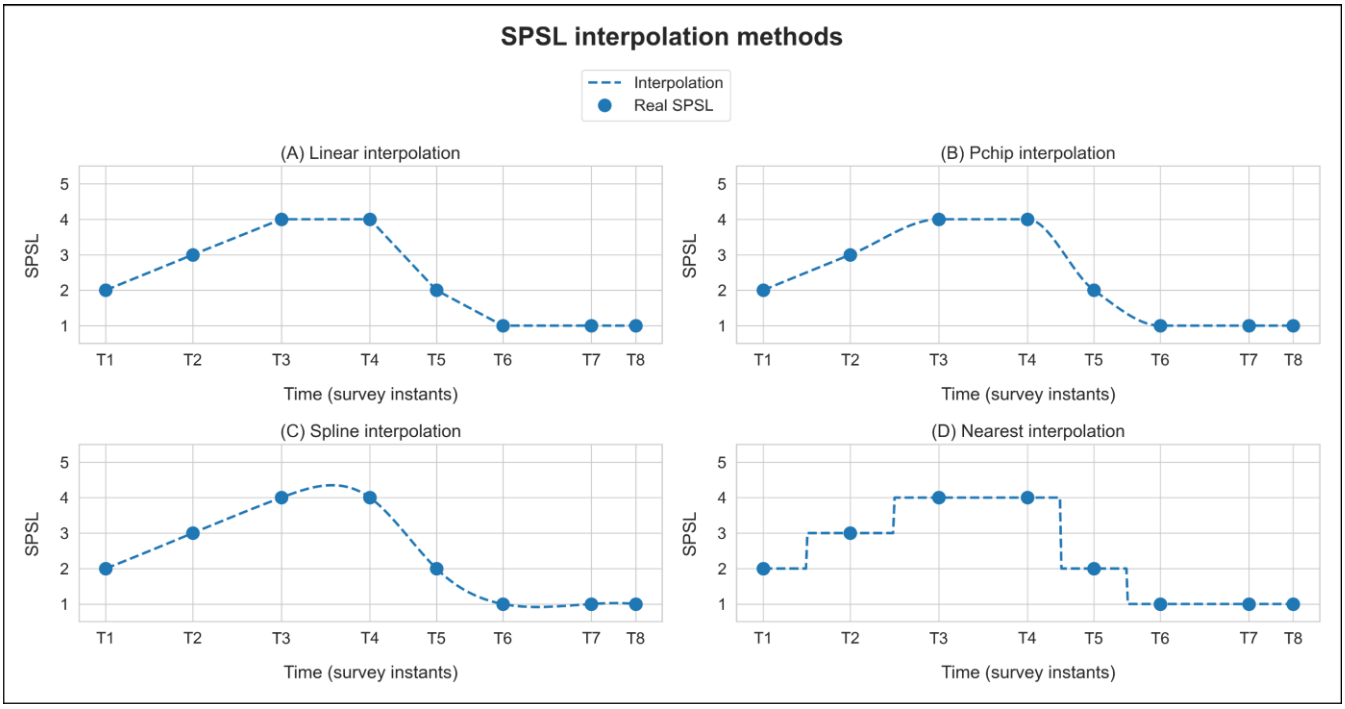


Figure 4: Participant's data structures utilized for model development. Feature matrix X corresponds to the seven spectral features that we estimated from the EEG recordings. Target array y refers to the SPSL values that we obtained from the SPSL surveys of the participants through interpolation.

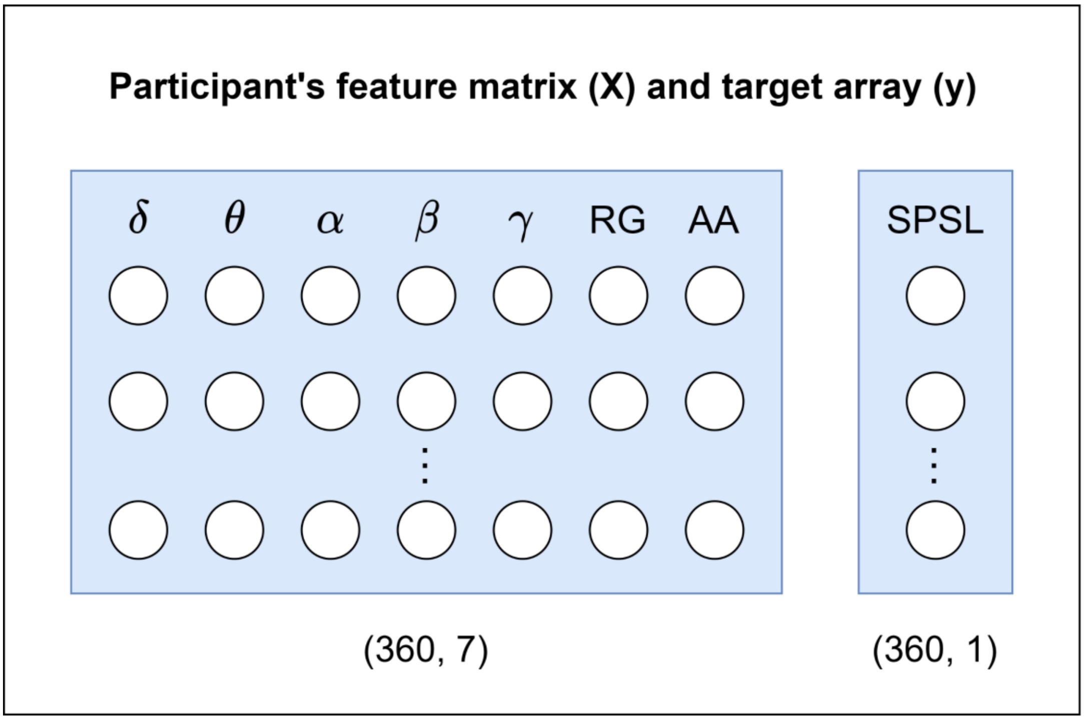


Figure 5: Model development pipeline. First, we extracted feature matrix (X) and target array (y) from EEG recordings and SPSL surveys, respectively. Later, for every participant, we splitted these data structures into training and test set. Then, we performed grid search cross-validation to find the best set of hyperparameters for the four regressors considered. Subsequently, we refitted the regressor with best hyperparameters on the training set, and we assessed the model on the test set.

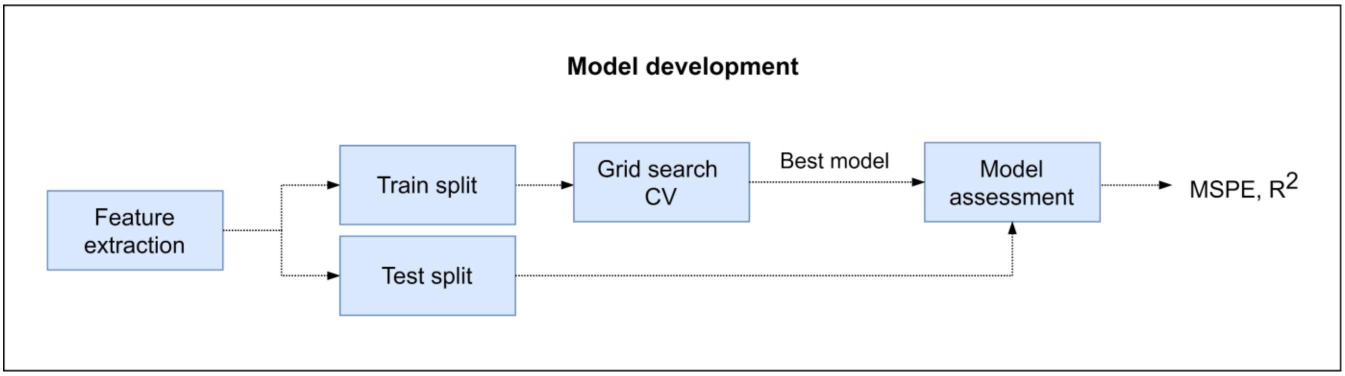


Figure 6: SPSL values predicted by the best regression model (random forest with linearly interpolated data) for each participant. Red square markers represent predicted SPSL values in T1-T8. Blue round markers represent SPSL values provided by the participant in T1-T8. MSPE and R2 for each participant are reported in brackets in the graph titles. For simplicity, we have displayed Y axis only in the first graph of each row, and we have used the same scale for all the participants.

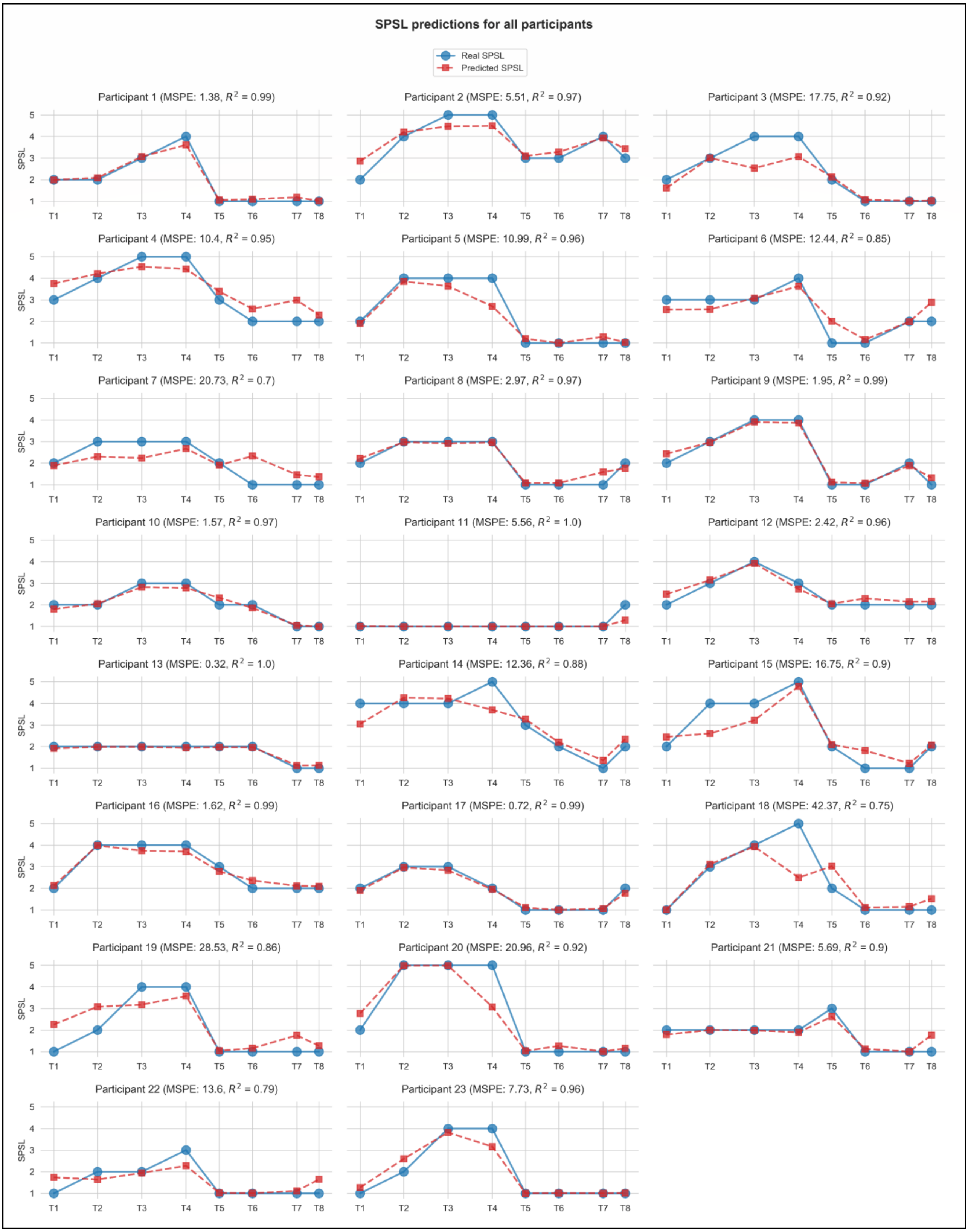
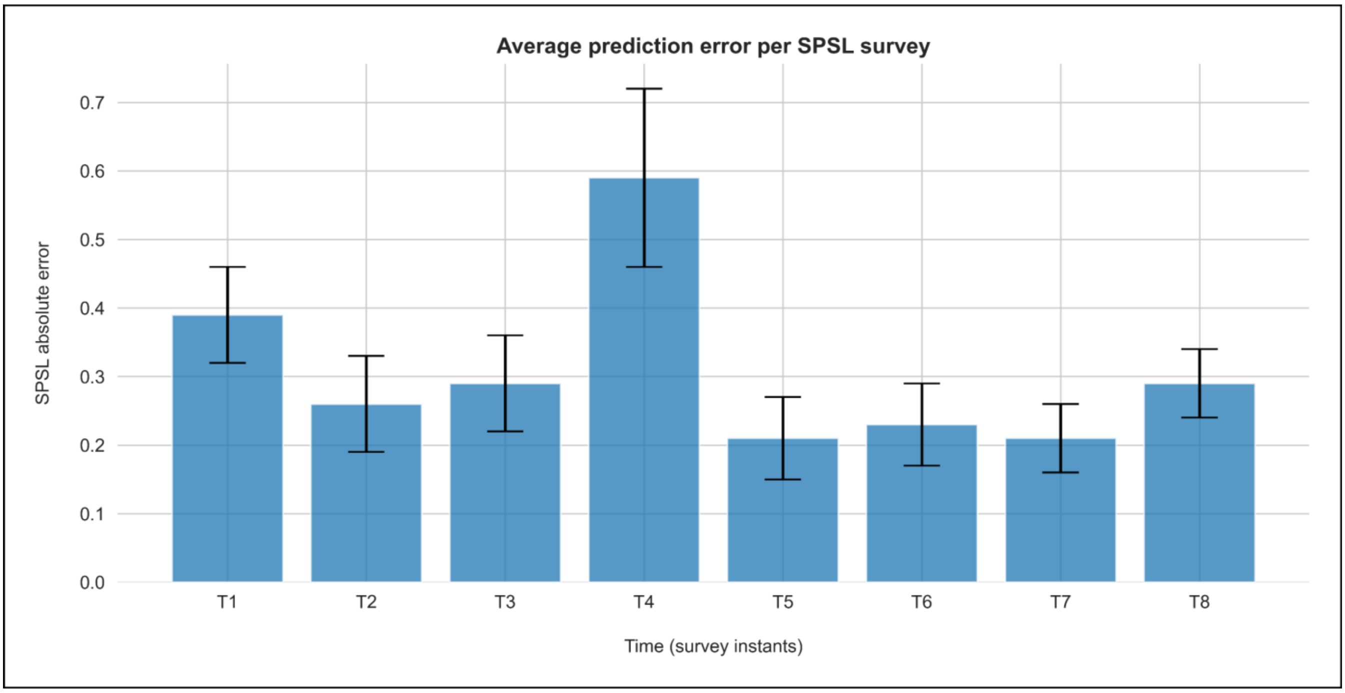


Figure 7: Absolute error of SPSL predictions on each survey for the best regression model (random forest with linearly interpolated data). Each bar represents the average absolute error of the SPSL predictions across the participants. Error bars indicate the SEM.



Appendix D

EEG-based multi-level stress classification with and without smoothing filter

Biomedical signal processing and control

EEG-based Multi-level Stress Classification with and without Smoothing Filter

Eduardo Perez-Valero[1,2], Miguel A. Lopez-Gordo[2,3,4*], and Miguel A. Vaquero-Blasco[2,3]

[1]Department of Computer Architecture and Technology, University of Granada

[2]Research Centre for Information and Communications Technologies, University of Granada

[3]Department of Signal Theory, Telematics and Communications, University of Granada

[4]Nicolo Association, Churriana de la Vega, Spain

*corresponding author (e-mail: malg@ugr.es, tlf: +34958249721)

Abstract. Recently, multi-level stress assessment has become an active research subject. In this context, researchers typically develop models based on machine learning classifiers and features extracted from biosignals like electrocardiogram (ECG) or electroencephalogram (EEG). For that purpose, EEG power spectral density (PSD) is a recurrent feature owing to its high responsiveness and remarkable performance. However, PSD is usually smoothed to cope with its bursty nature, what may cause data leakage and hence call into question classification performance. In this study, our aim was twofold: first, to examine the effect of EEG-PSD smoothing in three-level stress classification, and second, to evaluate the practical viability of a two-level stress detector without smoothing. To this end, we conducted participants through a stress-relax session while recording their EEG. Then, we estimated the EEG-PSD and used the stress reported by the participants as labels for classification. Initially, we developed a three-level stress classifier and examined the effect of smoothing on its performance. We found that classification performance was directly proportional to smoothing intensity (F1-score 0.61-0.94), and also that when smoothing was not applied to features, classification performance was insufficient for practical applicability (AUC < 0.7). We link this behavior to train-test contamination due to smoothing. Subsequently, we attempted two-level stress classification without smoothing. In this case, performance met the criteria for practical applicability (AUC = 0.76). This suggest that performance enhancement in three-level stress classification was caused by data leakage produced by smoothing, and hence, to render realistic stress classifiers each epoch should be processed individually.

Keywords: stress, classification, EEG, signal smoothing, data leakage.

1. Introduction

Currently, mental stress is an unavoidable concern that affects people on a global scale. According to the American Psychological Association [1], main sources of stress include health care, climate change, and safety. Mental stress can be triggered by several aspects of daily life, such as work, routine, and restless periods, and is usually linked to psychophysiological symptoms like headaches or fatigue [2], although other important health issues may appear [3]. In terms of economic impact, annual expenses associated to work miss and stress-related health issues are estimated around USD 300 billion. These facts justify the growing interest in stress early detection and classification.

According to literature, stress can be detected through several physiological markers, like galvanic skin response (GSR) [4]–[6], electroencephalography (EEG) [7]–[10], electrocardiography (ECG) [11]–[15], and cortisol [16]. With regard to EEG, brain activity is usually acquired and processed to obtain markers linked with stress, such as power in Theta, Alpha, and Beta bands, relative Gamma [17]–[19], coherence, and asymmetry. This technique has been widely applied in emotion recognition and mental illness screening studies [20]–[23]. In the case of ECG and GSR, stress level assessment is derived from heartbeat and skin conductance, respectively. In this context, cortisol is considered a gold standard for stress assessment, as the level of this hormone that is secreted into saliva increases when people are stressed. Among these techniques, EEG is the most generally cited approach for the evaluation of stress. This may be due to its high responsiveness and temporal resolution [19], what supports the implementation of real-time solutions [24].

Current research trends dwell on stress assessment through the development of classification models that are built on biomarkers extracted from EEG, ECG, GSR, etc. With regard to EEG, classification is accomplished through the combination of brain activity features and classifiers such as support vector machines (SVM) [8], [25], [26], artificial neural networks (ANN) [27], random forest (RF) [28], [29], etc. [30], [31]. In this context, classification performance report is paramount, since practical stress classifiers require a certain level of efficiency. Although main approaches typically consider the classification of two levels of stress (stress and no stress), there are proposals for the discrimination of three or even more stress levels (typically low, medium, and high stress). For instance, in [32], authors aimed to evaluate the effect of music on the stress level of the participants from their EEG activity. They proposed a three-level stress classification considering low, medium, and high stress classes. The highest accuracy that they obtained for the three-level classification was around 95%. In the case of [25], the authors also achieved a remarkable performance for a five-level discrimination (the minimum accuracy obtained for one class was 90.26%). These examples support the potential of multi-level stress classification. Nevertheless, two important aspects are occasionally obviated when the results of a multi-level stress classifier are reported: a detailed explanation of the EEG processing and a meticulous report of the classifier performance. Usual EEG processing involves smoothing and filtering (e.g., Butterworth, Chebyshev, and Savitzky–Golay filters) what may produce data leakage when the data are splitted using common procedures such as grid search cross validation or k-fold cross validation. For instance, if multiple epochs are smoothed out together, information from adjacent epochs is shared. Thus, when most reproduced cross validation methods are applied, epochs in the train and test sets may contain correlated information, what leads to unrealistic good performances. Therefore, to render reliable and reproducible results, non-adequate signal processing must be avoided. Additionally, when classifier performance is reported, authors often provide only accuracy, hence disregarding the performance obtained in the discrimination of each stress level. To provide a comprehensive interpretation of the model performance, other metrics such as precision, recall, and F1 score should be considered.

In this paper, we examined the effect that smoothing filter window length has on the performance of a three-level stress classifier. We also evaluated a three-level and a two-level stress classifier from individually processed epochs and assessed their practical applicability. To this effect, we computed the EEG-PSD during a stress-relax session, and we performed two and three-level classifications of the stress perceived by a group of participants. We considered different performance metrics aside from accuracy, namely precision and recall for each class, and weighted F1 score. Finally, we discussed our results and examined other studies in literature.

2. Material and methods

2.1. Participants

Twenty healthy volunteers (14 males, 6 females, mean age 24.20 ± 4.03) participated in the study. They belonged to the community of the University of Granada (mostly students and staff members) and were recruited via email distribution lists. The participants did not suffer from any health condition neither mental disorder. They were asked to sign an informed consent, and to avoid any stimulant or relaxant the day before the experiment. Furthermore, participants were not rewarded in any way for their participation in the study. Each participant took part in a single study session that lasted roughly 18 minutes. The entire data capture was completed in approximately two weeks.

2.2. Experimental design

Once we briefed the participants about the different phases of the study, we equipped them with an EEG acquisition system to record their EEG activity. The timeline of the experiment is illustrated in Figure 1.

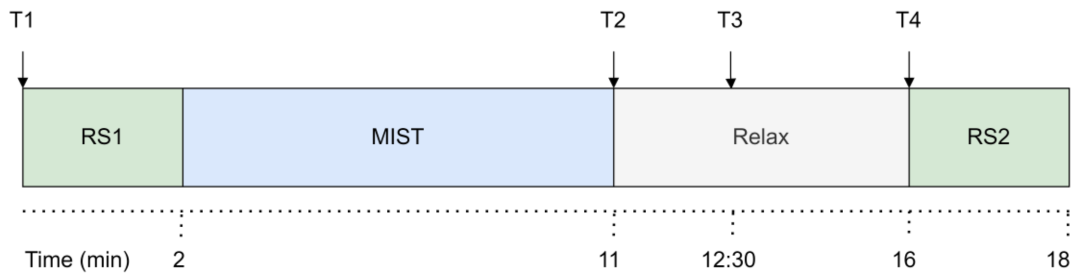


Figure 1. Timeline of the experimental process. First the participants completed a two-minute resting state period (RS1). Then, they performed the Montreal imaging stress task (MIST). Subsequently, they went through a relax session (Relax). Finally, the participants completed another two-minute resting state period (RS2). Several surveys (T1-T4) were fulfilled by the participants to grade their self-perceived stress level (SPSL). The total span of the experiment was 18 minutes.

First, the participants completed a two-minute eyes-closed resting state period. Then, they performed the Montreal imaging Stress task (MIST), a test specifically designed to induce psychosocial stress [33] that has been widely validated in literature [34]–[36]. Subsequently, participants were randomly separated into two groups. The first group was conducted through a relaxation program that used a loop of three ambient light colors (blue, magenta, and green) inside a chromotherapy room. The second group experienced an immersive virtual reality app that simulated the chromotherapy room program. Finally, the participants completed another two-minute resting state period. Throughout the experiment, we required the participants to report their self-perceived stress level (SPSL) via surveys. These surveys are referred as T1-T4 in Figure 1.

2.3. Experimental setup

Regarding the MIST, we implemented the test using MATLAB R2016a (MathWorks, USA) as a graphical interface controlled with the touchscreen of a laptop. To complete the test, the participants sat in a chair and were asked to use only their dominant hand. The MIST lasted nine minutes, including a three-minute training period and a six-minute test period. To record the

EEG activity, we used the RABio w8 acquisition system, that works at a sampling rate of 500 Hz. Our team at the University of Granada developed this device, and we have successfully utilized it in previous works [24], [37]. We located the electrodes at Fp1, Fp2, F7, F8, Fz, Cz, O1, and O2 positions of the 10–20 International System. Nevertheless, we only considered frontal and pre-frontal positions (Fp1, Fp2, F7, and F8) for further analysis in this study. We defined this electrode montage according to previous successful studies on stress assessment [17]–[19], [24]. We referenced and grounded the electrodes to the left ear lobe.

In relation to the SPSL surveys, we conceived an adaptation of the perceived stress scale (PSS) to minimize the time required to answer. Only two questions were asked to the participants: "Is your current stress level higher or lower than the last time that we asked?" (possible answers "lower", "higher", and "equal") and "What is your stress level in a scale from 1 to 5?" (possible answers 1 to 5, where 1 is the minimum level of stress and 5 is the maximum).

2.4. Signal processing

Upon completion of the data capture, we processed the EEG data of the participants offline. First, we filtered the EEG signals using a zero-phase shift 2nd order Butterworth filter with bandpass 1-50 Hz. Then, we applied a notch filter at 50 Hz to remove electric coupling. Thereafter, we splitted the EEG signals into two-second epochs without overlapping. To reject artifacts, we zeroed the epochs above a prearranged threshold of 100 µV. We selected this threshold following visual inspection and in compliance with previous EEG studies [38], [39]. Subsequently, we detrended and z-scored each epoch, and estimated the channel-averaged PSD in five frequency bands (see Table 1) using the periodogram (no overlapping). For RS1 and RS2 (see Figure 1), we processed only the central minute, hence the total time span of the processed signals was approximately 16 minutes.

Table 1. Frequency bands where we obtained the PSD of the EEG.

Band	Range (Hz)
Delta	1 - 4
Theta	4 - 8
Alpha	8 - 13
Beta	13 - 25
Gamma	25 - 45

In addition, we calculated the Alpha asymmetry as the difference between the Alpha power at prefrontal electrodes (Fp1 and Fp2), and the relative Gamma as indicated in Equation 1. Furthermore, according to previous studies on stress assessment by relative Gamma [17]–[19], we inverted the gamma power for participants S01, S04, S06, S09, S10, S12, S13, S15, S18, and S19 after visual inspection.

$$\mathrm{RG} = \frac{\mathrm{P_{Gamma}}}{\mathrm{P_{Alpha}} + \mathrm{P_{Theta}}} \qquad (1)$$

Overall, we derived seven EEG-PSD signals for each participant. To equal the span of the signals from all the participants, we resampled them to 480 samples (what corresponds to the aforementioned 16-minute span divided into two-second epochs). Lastly, we smoothed the seven signals using a 2nd order Savitzky-Golay filter. This procedure is widely applied in EEG literature to cope with the high temporal variability of brain electrical activity. For this filter, we examined window lengths of 11, 21, 31, 41, 51, and 61 epochs in order to assess the effect of this parameter on stress level classification performance. We also examined the case where each epoch was processed individually (referred as 0 window length throughout the paper). To illustrate the smoothing effect, Figure 2 shows the application of a Savitzky-Golay filter with different window lengths to the EEG Delta power.

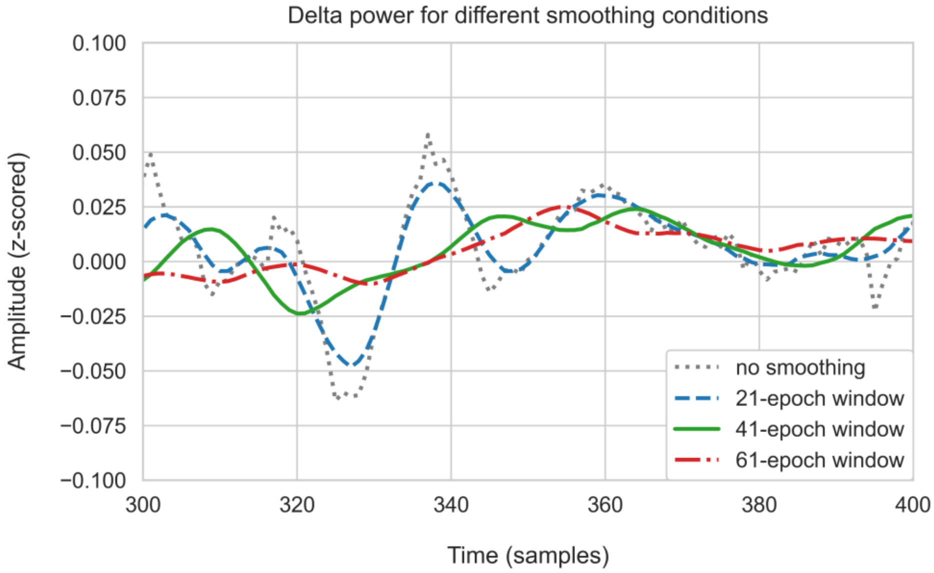


Figure 2. Effect of smoothing Delta power using a 2nd order Savitzky-Golay filter. The curves in the graph correspond to smoothing windows of length 0, 21, 41, and 61 epochs (dotted gray, dashed blue, solid green, and dashed-dotted red, respectively).

To conclude this subsection, it is worth to mention that EEG signals are affected by contamination sources such as MEG power and EMG power of forehead muscles. Decoupling of these signals would likely increase stress level classification performance. However, this procedure is out of the scope of this study, nonetheless the interested reader may refer to [33] and [34] for a comprehensive understanding about this topic.

2.5. Feature extraction

For each of the seven PSD signals extracted, we partitioned the minute preceding the SPSL surveys (T1-T4) into five segments, each one enclosing six epochs. Then, we averaged those six epochs for each segment, and we assigned it a label equal to the SPSL survey under consideration. Consequently, for each participant, we obtained a data matrix with 140 samples (4 surveys x 5 segments x 7 PSD signals), and one label per sample (values in the range 1-5 from the SPSL surveys). For classification, we combined the data matrices from all the participants. Figure 3 represents the segmentation procedure described in this paragraph for one of the seven PSD signals.

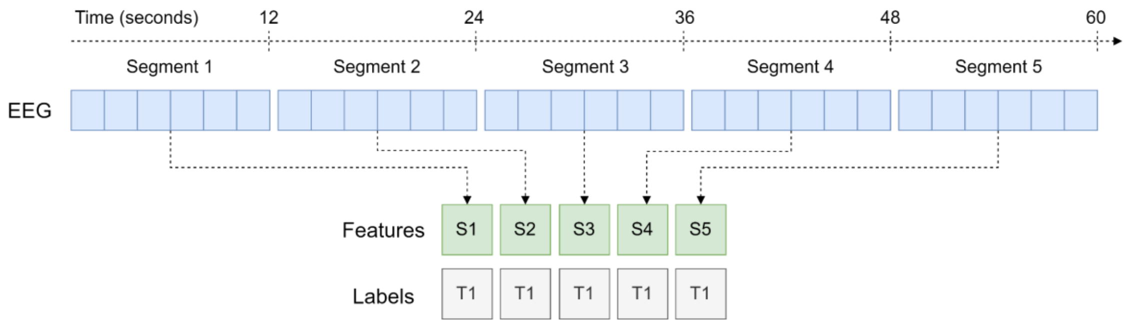


Figure 3. PSD segmentation. We partitioned the minute preceding each SPSL survey into five segments each one holding six epochs (in blue). Then, we averaged each segment (in green) and we assigned it a label (in gray) corresponding to the SPSL of the corresponding survey. This figure is referred to the first survey that we conducted in the study (T1).

2.6. Classification

As stated in the Introduction, in this study we examined a three-level classification task and a two-level stress detection. For the three-level classification task, we aimed to predict three stress levels (low stress, medium stress, and high stress). To this end, we merged the SPSL survey answers into three classes: low stress (labels 1 and 2), medium stress (labels 3 and 4), and high stress (label 5). For two-level stress detection, we intended to discern between two states (stress and no stress). Thus, in this case we combined the SPSL survey answers into two classes, namely, no stress (labels 1, 2, and 3) and stress (labels 4 and 5).

In regard to classification, first we splitted feature matrix *X* and target array *y* into stratified train and hold-out sets (80% train, 20% hold-out). We used the train set to find the best combination of hyperparameters for multiple classifiers, and the hold-out set to evaluate the classifier that achieved the best performance. Subsequently, we applied standardization and oversampling to the feature matrix. We implemented oversampling through SMOTE (Synthetic Minority Oversampling Technique), as the high stress class (label 5) had a lower presence in the data compared to the rest. Essentially, this technique synthetizes new instances from the minority class by interpolating real instances of that class. After interpolation, we applied grid search cross-validation (GSCV) to find the combination of classifier and hyperparameters that best predicted SPSL from spectral features. The classifiers considered in this study included logistic regression (LR), support vector machine (SVM), random forest (RF), k-nearest neighbors (KNN), and multi-layer perceptron (MLP). We configured GSCV to perform a 5-fold cross-validation over all the possible combinations of hyperparameters for each classifier. For this procedure, we selected the weighted F1 score as scoring metric. In multiclass problems, this score is calculated per class and then averaged to obtain a general metric for the classifier performance. We selected the weighted approximation of the F1 score to account for class imbalance.

Lastly, as stated in subsection 2.4, we considered different window lengths for the Savitzky-Golay filter (0, 11, 21, 31, 41, 51, and 61 epochs) to assess the effect that this parameter had on three-level stress classification performance. To sum up, Figure 4 represents the procedures described through this section.

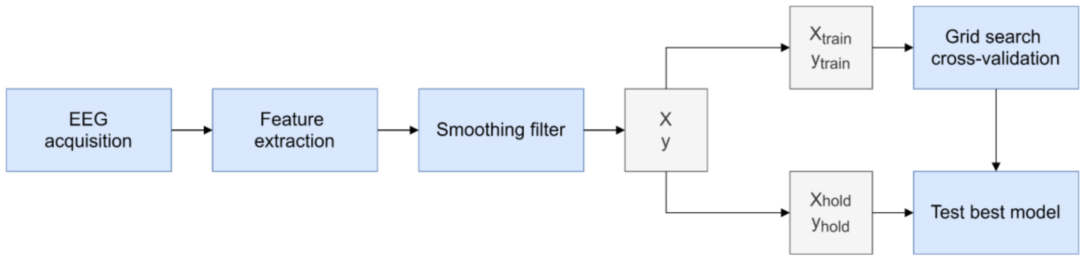


Figure 4. Stress classification pipeline. First, we performed the EEG acquisition using RABio w8 system. Then, we estimated the PSD in different frequency bands, and we applied a Savitzky-Golay filter to smooth spectral data. Subsequently, we splitted the minute preceding the SPSL surveys (T1-T4) into five segments, and we estimated the average spectral features in each of them. Then, we reshaped the spectral features and survey answers to build feature matrix *X* and target array *y*. After that, we splitted these two data structures into training and hold-out sets. Thereafter, for each classifier analyzed in the study, we performed GSCV using only the training set. Finally, we assessed the best classifier on the hold-out set.

3. Results

3.1 Cross-validation

Table 2 presents GSCV results for the three-level stress classification task. In this table, we report the F1 score obtained by each of the five classifiers for each of the smoothing window lengths examined in this study.

Table 2. Results obtained during GSCV for the three-level stress classification task. The left most column represents the classifiers. The rest of the columns display the average F1 score (± std) obtained by each classifier for the different smoothing window lengths reported in epochs (0-61). Shadowed, the best classifier for each smoothing window length.

	F1 score per smoothing window length						
Model	**0**	**11**	**21**	**31**	**41**	**51**	**61**
LR	0.45 ± 0.02	0.45 ± 0.03	0.48 ± 0.04	0.52 ± 0.02	0.55 ± 0.07	0.54 ± 0.05	0.56 ± 0.05
SVM	0.55 ± 0.04	0.63 ± 0.04	0.63 ± 0.05	0.72 ± 0.05	0.78 ± 0.07	0.84 ± 0.02	0.89 ± 0.05
RF	0.60 ± 0.05	0.62 ± 0.03	0.62 ± 0.08	0.67 ± 0.07	0.66 ± 0.08	0.74 ± 0.06	0.76 ± 0.05
KNN	0.54 ± 0.03	0.63 ± 0.05	0.63 ± 0.04	0.68 ± 0.03	0.74 ± 0.06	0.83 ± 0.04	0.82 ± 0.03
MLP	0.59 ± 0.02	0.61 ± 0.08	0.64 ± 0.06	0.68 ± 0.06	0.74 ± 0.08	0.75 ± 0.08	0.78 ± 0.03

Table 3 shows the GSCV results for the stress detection task. In this case, we did not apply a smoothing filter. The left column corresponds to the classifiers, and the right column displays the F1 score obtained by each of the models in the stress detection task.

Table 3. Results obtained during GSCV for the stress detection task. Left column indicates the models analyzed. Right column represents the average F1 score (± std) for each of the models analyzed. Shadowed, the classifiers that obtained the highest performance.

Model	**F1 score**
LR	0.65 ± 0.06
SVM	0.84 ± 0.02
RF	0.84 ± 0.03
KNN	0.78 ± 0.04
MLP	0.84 ± 0.02

Figure 5 represents a graphic arrangement of Table 2, and is intended to facilitate its overview. It displays the F1 score obtained by the classifiers during GSCV for each of the smoothing window lengths considered in the three-level stress classification task.

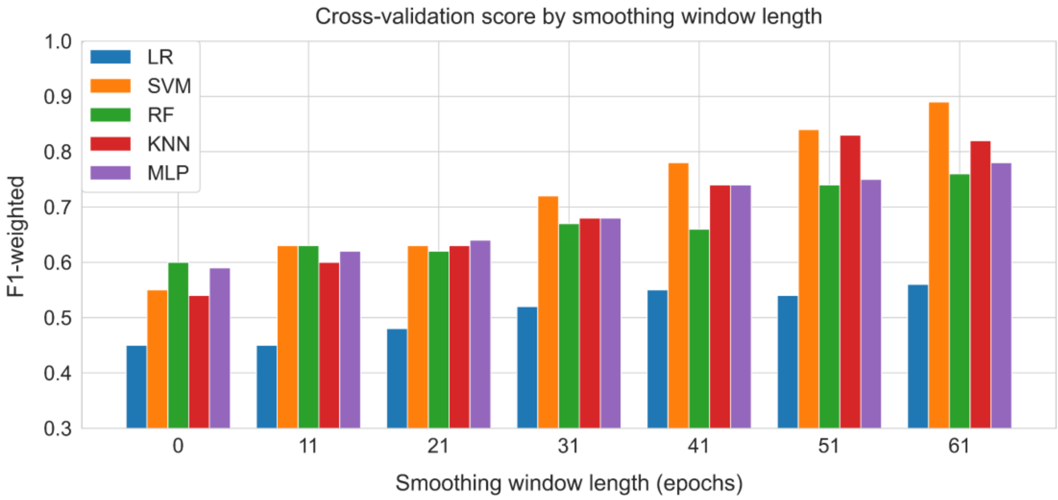


Figure 5. GSCV score yielded by each classifier for the three-level stress classification task. Y axis corresponds to the F1 score, and X axis refers to the window length of the Savitzky-Golay filter used during EEG-PSD smoothing. Each bar in the graph represents a classifier: LR (blue), SVM (orange), RF (green), KNN (red), and MLP (purple)

3.2 Test

Table 4 presents the performance of the best model yielded by GSCV on the hold-out set. The top part of the table corresponds to three-level classification and the second part refers to stress detection. The first column represents the different smoothing window lengths analyzed. For each of these lengths, the rest of the columns indicate the best classifier and its performance in terms of class precision and recall, F1 score, and accuracy.

Table 4. Performance metrics on the hold-out set for three-level classification (top) and stress detection (bottom). Reported metrics include precision (P) and recall (R) for each class, F1 score, and accuracy.

Smoothing window length	Best model	Class 1 (P)	Class 1 (R)	Class 2 (P)	Class 2 (R)	Class 3 (P)	Class 3 (R)	F1 score	Accuracy
Three-level stress classification									
0	RF	0.68	0.70	0.58	0.46	0.29	0.67	0.61	0.61
11	SVM	0.71	0.81	0.72	0.54	0.25	0.33	0.68	0.69
21	MLP	0.74	0.76	0.57	0.54	0.67	0.67	0.67	0.67
31	SVM	0.84	0.86	0.77	0.71	0.75	1	0.81	0.81
41	SVM	0.89	0.86	0.78	0.75	0.6	1	0.83	0.83
51	SVM	0.92	0.95	0.91	0.83	0.75	1	0.91	0.91
61	SVM	0.95	0.95	0.92	0.92	1	1	0.94	0.94
Stress detection									
0	MLP	0.92	0.85	0.38	0.56	-	-	0.83	0.81

Figure 6 illustrates the receiver operating characteristic (ROC) curves of the best model for each smoothing window length considered in three-level stress classification. Each graph includes three curves, that represent the binary classification problem for each class versus the rest.

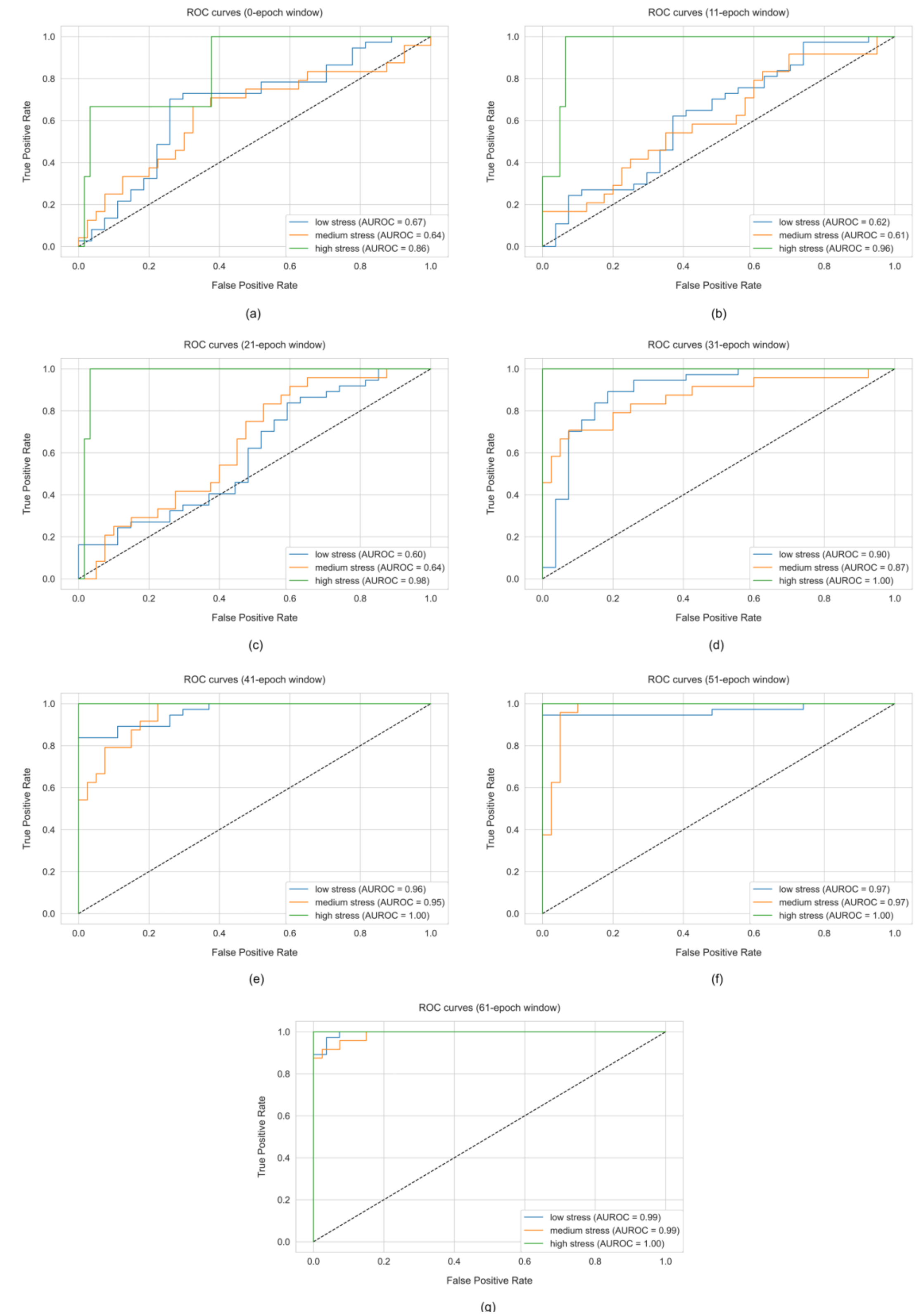


Figure 6. ROC curves for three-level stress classification task. Smoothing window lengths of 0 (a), 11 (b), 21 (c), 31 (d), 41 (e), 51 (f), and 61 (g) epochs. Each graph includes three curves, representing the binary problem for a class versus the rest of the classes, for low stress (blue), medium stress (orange), and high stress (green). The area under the curve (AUC) is reported in brackets. Black dashed line represents the chance level.

Figure 7 depicts the ROC curve for the best model in the stress detection task.

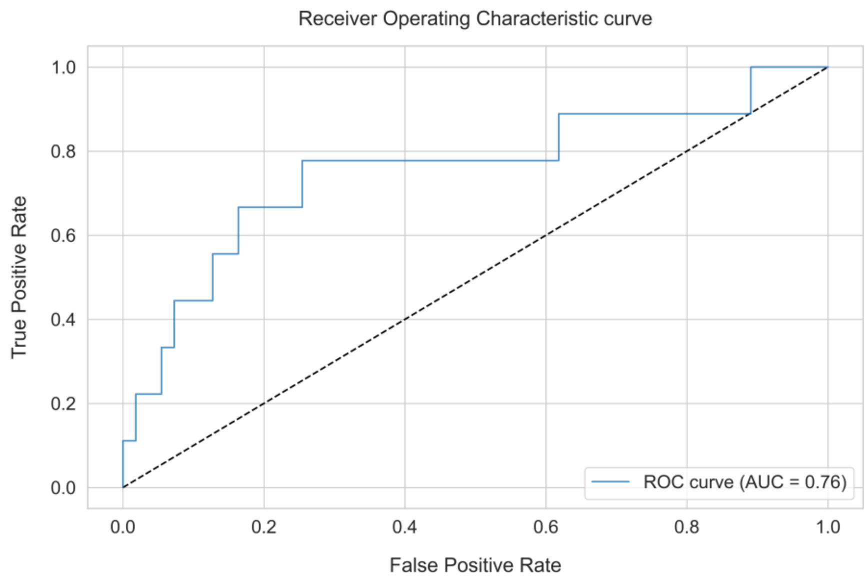


Figure 7. ROC curve for the binary classification problem (stress vs. no stress). AUC is reported in brackets. Black dashed line represents the chance level.

Figure 8 displays the F1 score obtained on the hold-out set by the best model for each of the smoothing window lengths analyzed in the three-level stress classification. We included this figure to illustrate the enhancement in performance as the window length is increased.

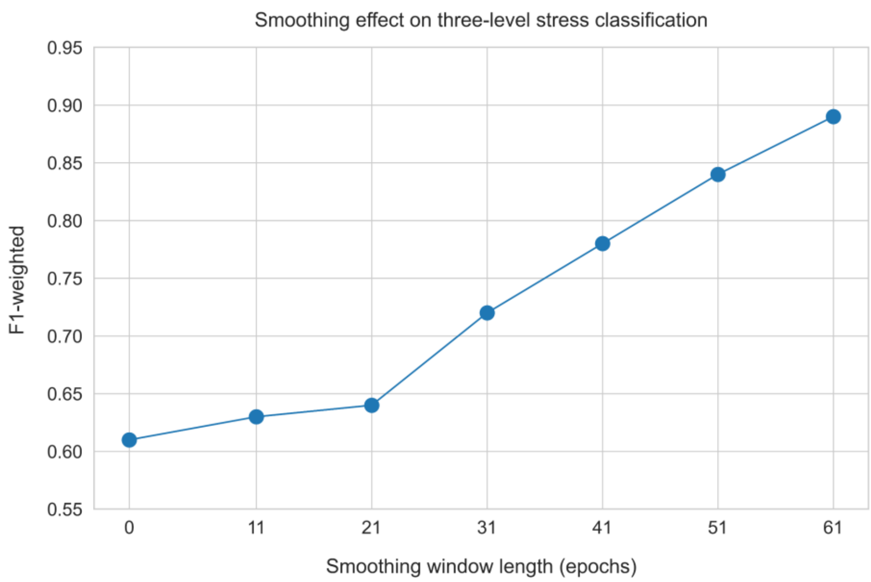


Figure 8. Smoothing effect on three-level stress classification performance. X axis refers to the length of the smoothing filter window in epochs. Y axis represents the weighted F1 score yielded on the hold-out set by the best model.

4. Discussion

In this study we have assessed the effect of smoothing filter window length on the performance of a multi-level stress classifier. Moreover, we have examined the practical appropriateness of a three-level stress classifier and a stress detector when each EEG-PSD segment is processed individually. With respect to three-level stress classification, for the non-smoothed case (zero epochs), the best model yielded an accuracy of 0.61. Alternatively, for the longest smoothing window length (61 epochs), the accuracy obtained by the best model was 0.94. In regard to the stress detection task, the best performing model yielded an accuracy of 0.81. These results evidence that, (i) the smoothing of the EEG-PSD improves classification performance, possibly due to train-test contamination; (ii) although the performance of our three-level stress classifier with no smoothing is not good enough for practical implementations, stress can be successfully detected by means of a two-level classification from non-smoothed EEG-PSD signals.

4.1 Cross-validation results

According to the results presented in Table 2, with scarcely smoothed data (filter window lengths 0-31 epochs) all the classifiers performed similarly, with the exception of LR. However, for longer windows (lengths 31-61 epochs), SVM performed noticeably better than the rest of the classifiers. These findings suggest that SVM outperforms other EEG-based multi-level classifiers when intense smoothing is applied to EEG data. This is coherent with other results reported in the literature for multi-level stress classification [8], [25], [26], [28]. With respect to stress detection, SVM, RF, and MLP yielded the highest performance, while KNN reached reasonable performance and LR discriminated worst. For the stress detection task, we processed each epoch individually, what prevents from spreading correlated information between adjacent epochs, and enhances the timing capabilities of the system. For instance, for a 61-epoch window, classification requires a 1-minute delay as a result of the collection of the 30 epochs posterior to the central sample. Conversely, for the non-smoothed approach, we used segments of 6 consecutive epochs, and hence, classification can be performed with a latency of 6 seconds and 12-second time resolution.

4.2 Test results

Table 4 shows the results of the test phase. In literature, it is frequent to report just the accuracy of three-level stress classifiers [8], [26]. Nevertheless, to offer a comprehensive performance report, other relevant metrics such as precision and recall are required. To this end, in Table 4 we have reported accuracy, F1 score, and intra-class precision and recall. For both the three-level classification and stress detection tasks, the performance of the best classifier for all smoothing window lengths yielded results similar to those obtained in the cross-validation phase (confront the F1 score in Tables 2 and 3 with the same metric in Table 4).

To gain insights about the prediction capacities of the classifiers, we computed the ROC curves. Figure 6 shows the ROC curves for all the smoothing window lengths considered in the three-level stress classification task. If we consider the clinical environment criterion in regard to AUC [42]: acceptable (AUC 0.7-0.8); excellent (AUC 0.8-0.9), we could interpret the results of Table 4 and Figure 6 as follows:

- Three-level stress classification: a smoothing window length of 31 epochs is enough to achieve an excellent prediction capability of the three classes (low-medium-high AUC 0.90-0.87-1.00, respectively). With lengths shorter than 31 epochs, the quality of our classifier is below acceptable for low and medium stress classes (AUC between 0.60 and 0.64). Regardless the smoothing window length, the detection of high stress level is always excellent (AUC between 0.86 and 1).

- Stress detection: the detection of stress from non-smoothed data can be described as acceptable (AUC 0.76).

In view of the results, we can conclude that our methodological approach is reproducible and the stress detection we proposed in this study constitutes a feasible alternative for practical implementations of a two-level stress classifier. With regard to three-level stress classification, our non-smoothed classification approach is close to acceptable for low and medium levels of stress (AUC 0.67 and 0.64, respectively), and excellent for high level of stress (AUC 0.86). The reason behind the remarkable performance in the discrimination of the high stress class might be justified by more energetic EEG-PSD features than the other stress levels, although this assumption requires further analysis. Consequently, our non-smoothed three-level classification approach is not appropriate for practical implementations of multi-level stress classifiers. For wider smoothing windows (31-61 epochs), the performance of the three-level stress classifiers improves considerably. However, as stated before, when smoothing is applied, the classification may incur in train-test contamination, as the information of adjacent EEG segments is shared, and diminish timing capabilities.

4.3 Comparative discussion

In literature, many studies have reported high performance for multi-level stress classification. However, a direct comparison with our results is not trivial. In some of those studies, authors only reported accuracy [8], [26] and did not provide an analysis of the intra class performance (e.g., confusion matrix, precision, recall, or ROC curves). This is not an appropriate way to report the performance of a three-level stress classifier. For instance, a classifier with a very high sensibility and low specificity for a high-probability class, and null sensibility and specificity for a low-probability class, could obtain a very high accuracy. In addition, although segmentation and filtering greatly influence classification performance, theses aspects of the EEG processing are often overlooked. In [32], the authors declared that they applied a 75% overlapping method (unspecified) to the EEG signals to obtain the PSD, and reached an accuracy of 98.76%. Other studies reported accuracies up to 93.6% and 94.3% [25], [26], but the authors only indicated that the EEG spectral bands were denoised without providing additional details. Figures 5 and 8 demonstrate that three-level stress classification performance increases as the smoothing window is enlarged. Therefore, a comparative discussion with other studies that do not accurately report the EEG processing is inherently unviable.

5. Conclusions

In this work, we evaluated the performance of a three-level stress classifier for different smoothing filter window lengths, and we also approached three-level and two-level stress classification from non-smoothed EEG-PSD data. To this effect, we recorded the brain activity of participants during a stress-relax session and we extracted EEG-PSD features. We used self-perceived stress level surveys as labels for classification. We assessed multiple stress classifiers whose hyperparameters were derived via GSCV. To evaluate the effect of smoothing on three-level stress classification, we considered smoothing filter window lengths ranging from 0 to 61 epochs, and we estimated multiple performance metrics (accuracy, F1 score, and intra-class precision and recall). The results we obtained do not support the appropriateness of a three-level stress classifier from non-smoothed EEG data for practical applications. Our results also evidence that data smoothing increases classification performance at the expense of train-test contamination and poorer timing capabilities. According to our results, it is feasible to classify three stress levels from individually processed EEG segments with remarkable performance (AUC 0.67, 0.64, and 0.86 for low, medium, and high stress classes). In terms of stress detection, we implemented a two-level stress classifier without smoothing of adjacent segments that achieved close to excellent performance (AUC 0.76). This supports the implementation of our two-level stress approach in practical stress detection solutions. In this regard, an EEG stress classifier could be extended to real-time applications, avoiding the need to enquiry the participants about their SPSL. These applications may include stress relief therapies or neuromarketing, as in these environments stress is considered an important concern, and there is still a lack of stress anticipation solutions. Lastly, we provide the following recommendations for future EEG-based stress classification studies: (i) performance of three and two-level stress classifiers could be further enhanced if the EEG spectral features were combined with other features, such as galvanic skin response or heart rate variability; (ii) each EEG segment should be processed individually, so the content of adjacent segments is not combined; (iii) an appropriate classifier performance report should include metrics aside from accuracy, like intra-class precision and recall, F1 score, or AUC. In the future, reliable three-level stress classifiers could be used in brand new scenarios such as offices, schools, or even at home.

Acknowledgements

This work was supported by the project PGC2018-098813-B-C31 (the Spanish Ministry of Science, Innovation and Universities, by European Regional Development Funds and by the Nicolo Association for the R&D in Neurotechnology for disability. The authors would also like to thank Dr. Maria Jose Sanchez Carrion from the School for Special Education San Rafael of Granada, Hospitaller Order of St. John of God, for their support and for providing access to their chromotherapy room. We also want to thank Prof. Alexander Bertrand, from the Department of Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, (Belgium) for his suggestions and valuable comments for this manuscript.

References

[1] "Stress in America™ 2019," p. 9.

[2] N. B. Anderson *et al.*, "AMERICAN PSYCHOLOGICAL ASSOCIATION," p. 23.

[3] S. Cohen *et al.*, "Chronic stress, glucocorticoid receptor resistance, inflammation, and disease risk," *Proc Natl Acad Sci U S A*, vol. 109, no. 16, pp. 5995–5999, Apr. 2012, doi: 10.1073/pnas.1118355109.

[4] A. Fernandes, R. Helawar, R. Lokesh, T. Tari, and A. V. Shahapurkar, "Determination of stress using Blood Pressure and Galvanic Skin Response," in *2014 International Conference on Communication and Network Technologies*, Sivakasi, India, Dec. 2014, pp. 165–168, doi: 10.1109/CNT.2014.7062747.

[5] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, "Stress detection from speech and Galvanic Skin Response signals," in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, Porto, Portugal, Jun. 2013, pp. 209–214, doi: 10.1109/CBMS.2013.6627790.

[6] M. V. Villarejo, B. G. Zapirain, and A. M. Zorrilla, "A Stress Sensor Based on Galvanic Skin Response (GSR) Controlled by ZigBee," *Sensors*, vol. 12, no. 5, pp. 6075–6101, May 2012, doi: 10.3390/s120506075.

[7] H. Díaz M., F. M. Cid, J. Otárola, R. Rojas, O. Alarcón, and L. Cañete, "EEG Beta band frequency domain evaluation for assessing stress and anxiety in resting, eyes closed, basal conditions," *Procedia Computer Science*, vol. 162, pp. 974–981, 2019, doi: 10.1016/j.procs.2019.12.075.

[8] H. Jebelli, S. Hwang, and S. Lee, "EEG-based workers' stress recognition at construction sites," *Automation in Construction*, vol. 93, pp. 315–324, Sep. 2018, doi: 10.1016/j.autcon.2018.05.027.

[9] A. C. Marshall and N. R. Cooper, "The association between high levels of cumulative life stress and aberrant resting state EEG dynamics in old age," *Biological Psychology*, vol. 127, pp. 64–73, Jul. 2017, doi: 10.1016/j.biopsycho.2017.05.005.

[10] J. Minguillon, M. A. Lopez-Gordo, D. A. Renedo-Criado, M. J. Sanchez-Carrion, and F. Pelayo, "Blue lighting accelerates post-stress relaxation: Results of a preliminary study," *PLoS ONE*, vol. 12, no. 10, p. e0186399, Oct. 2017, doi: 10.1371/journal.pone.0186399.

[11] L. Chen, Y. Zhao, P. Ye, J. Zhang, and J. Zou, "Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers," *Expert Systems with Applications*, vol. 85, pp. 279–291, Nov. 2017, doi: 10.1016/j.eswa.2017.01.040.

[12] P. Karthikeyan, M. Murugappan, and S. Yaacob, "Detection of human stress using short-term ecg and hrv signals," *J. Mech. Med. Biol.*, vol. 13, no. 02, p. 1350038, Mar. 2013, doi: 10.1142/S0219519413500383.

[13] N. Keshan, P. V. Parimi, and I. Bichindaritz, "Machine learning for stress detection from ECG signals in automobile drivers," in *2015 IEEE International Conference on Big Data (Big Data)*, Oct. 2015, pp. 2661–2669, doi: 10.1109/BigData.2015.7364066.

[14] S. Pourmohammadi and A. Maleki, "Stress detection using ECG and EMG signals: A comprehensive study," *Computer Methods and Programs in Biomedicine*, vol. 193, p. 105482, Sep. 2020, doi: 10.1016/j.cmpb.2020.105482.

[15] M. N. Rastgoo, B. Nakisa, F. Maire, A. Rakotonirainy, and V. Chandran, "Automatic driver stress level classification using multimodal deep learning," *Expert Systems with Applications*, vol. 138, p. 112793, Dec. 2019, doi: 10.1016/j.eswa.2019.07.010.

[16] M. Annerstedt *et al.*, "Inducing physiological stress recovery with sounds of nature in a virtual reality forest — Results from a pilot study," *Physiology & Behavior*, vol. 118, pp. 240–250, Jun. 2013, doi: 10.1016/j.physbeh.2013.05.023.

[17] A. Lutz, L. L. Greischar, N. B. Rawlings, M. Ricard, and R. J. Davidson, "Long-term meditators self-induce high-amplitude gamma synchrony during mental practice," *Proceedings of the*

National Academy of Sciences, vol. 101, no. 46, pp. 16369–16373, Nov. 2004, doi: 10.1073/pnas.0407401101.

[18] S. R. Steinhubl *et al.*, “Cardiovascular and nervous system changes during meditation,” *Front. Hum. Neurosci.*, vol. 9, Mar. 2015, doi: 10.3389/fnhum.2015.00145.

[19] J. Minguillon, M. A. Lopez-Gordo, and F. Pelayo, “Stress Assessment by Prefrontal Relative Gamma,” *Front. Comput. Neurosci.*, vol. 10, Sep. 2016, doi: 10.3389/fncom.2016.00101.

[20] A. R. Aslam, T. Iqbal, M. Aftab, W. Saadeh, and M. A. B. Altaf, “A10.13uJ/classification 2-channel Deep Neural Network-based SoC for Emotion Detection of Autistic Children,” in *2020 IEEE Custom Integrated Circuits Conference (CICC)*, Mar. 2020, pp. 1–4, doi: 10.1109/CICC48029.2020.9075952.

[21] A. R. Aslam and M. A. B. Altaf, “An On-Chip Processor for Chronic Neurological Disorders Assistance Using Negative Affectivity Classification,” *IEEE Trans Biomed Circuits Syst*, vol. 14, no. 4, pp. 838–851, 2020, doi: 10.1109/TBCAS.2020.3008766.

[22] H. Jebelli, S. Hwang, and S. Lee, “EEG-based workers’ stress recognition at construction sites,” *Automation in Construction*, vol. 93, pp. 315–324, Sep. 2018, doi: 10.1016/j.autcon.2018.05.027.

[23] A. R. Aslam and M. A. B. Altaf, “An 8 Channel Patient Specific Neuromorphic Processor for the Early Screening of Autistic Children through Emotion Detection,” in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2019, pp. 1–5, doi: 10.1109/ISCAS.2019.8702738.

[24] J. Minguillon, E. Perez, M. Lopez-Gordo, F. Pelayo, and M. Sanchez-Carrion, “Portable System for Real-Time Detection of Stress Level,” *Sensors*, vol. 18, no. 8, p. 2504, Aug. 2018, doi: 10.3390/s18082504.

[25] M. Zubair and C. Yoon, “Multilevel mental stress detection using ultra-short pulse rate variability series,” *Biomedical Signal Processing and Control*, vol. 57, p. 101736, Mar. 2020, doi: 10.1016/j.bspc.2019.101736.

[26] S. Lotfan, S. Shahyad, R. Khosrowabadi, A. Mohammadi, and B. Hatef, “Support vector machine classification of brain states exposed to social stress test using EEG-based brain network measures,” *Biocybernetics and Biomedical Engineering*, vol. 39, no. 1, pp. 199–213, Jan. 2019, doi: 10.1016/j.bbe.2018.10.008.

[27] H. Sharma, R. Raj, and M. Juneja, “EEG signal based classification before and after combined Yoga and Sudarshan Kriya,” *Neuroscience Letters*, vol. 707, p. 134300, Aug. 2019, doi: 10.1016/j.neulet.2019.134300.

[28] Z. Halim and M. Rehan, “On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning,” *Information Fusion*, vol. 53, pp. 66–79, Jan. 2020, doi: 10.1016/j.inffus.2019.06.006.

[29] Y.-W. Kim *et al.*, “Riemannian classifier enhances the accuracy of machine-learning-based diagnosis of PTSD using resting EEG,” *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 102, p. 109960, Aug. 2020, doi: 10.1016/j.pnpbp.2020.109960.

[30] O. Attallah, “An Effective Mental Stress State Detection and Evaluation System Using Minimum Number of Frontal Brain Electrodes,” *Diagnostics*, vol. 10, no. 5, p. 292, May 2020, doi: 10.3390/diagnostics10050292.

[31] M. Shim, M. J. Jin, C.-H. Im, and S.-H. Lee, “Machine-learning-based classification between post-traumatic stress disorder and major depressive disorder using P300 features,” *NeuroImage: Clinical*, vol. 24, p. 102001, 2019, doi: 10.1016/j.nicl.2019.102001.

[32] A. Asif, M. Majid, and S. M. Anwar, “Human stress classification using EEG signals in response to music tracks,” *Computers in Biology and Medicine*, vol. 107, pp. 182–196, Apr. 2019, doi: 10.1016/j.compbiomed.2019.02.015.

[33] K. Dedovic, R. Renwick, N. K. Mahani, V. Engert, S. J. Lupien, and J. C. Pruessner, “The Montreal Imaging Stress Task: using functional imaging to investigate the effects of

perceiving and processing psychosocial stress in the human brain," *J Psychiatry Neurosci*, p. 7.

[34] A. Brugnera, C. Zarbo, M. P. Tarvainen, P. Marchettini, R. Adorni, and A. Compare, "Heart rate variability during acute psychosocial stress: A randomized cross-over trial of verbal and non-verbal laboratory stressors," *International Journal of Psychophysiology*, vol. 127, pp. 17–25, May 2018, doi: 10.1016/j.ijpsycho.2018.02.016.

[35] L. Han, Q. Zhang, X. Chen, Q. Zhan, T. Yang, and Z. Zhao, "Detecting work-related stress with a wearable device," *Computers in Industry*, vol. 90, pp. 42–49, Sep. 2017, doi: 10.1016/j.compind.2017.05.004.

[36] A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel, and A. S. Malik, "Machine Learning Framework for the Detection of Mental Stress at Multiple Levels," *IEEE Access*, vol. 5, pp. 13545–13556, 2017, doi: 10.1109/ACCESS.2017.2723622.

[37] M. A. Vaquero-Blasco, E. Perez-Valero, M. A. Lopez-Gordo, and C. Morillas, "Virtual Reality as a Portable Alternative to Chromotherapy Rooms for Stress Relief: A Preliminary Study," *Sensors*, vol. 20, no. 21, Art. no. 21, Jan. 2020, doi: 10.3390/s20216211.

[38] C. Jeunet *et al.*, "Uncovering EEG Correlates of Covert Attention in Soccer Goalkeepers: Towards Innovative Sport Training Procedures," *Scientific Reports*, vol. 10, no. 1, Art. no. 1, Feb. 2020, doi: 10.1038/s41598-020-58533-2.

[39] R. Cassani, T. H. Falk, F. J. Fraga, M. Cecchi, D. K. Moore, and R. Anghinah, "Towards automated electroencephalography-based Alzheimer's disease diagnosis using portable low-density devices," *Biomedical Signal Processing and Control*, vol. 33, pp. 261–271, Mar. 2017, doi: 10.1016/j.bspc.2016.12.009.

[40] R. B. Malmo and H. P. Malmo, "On electromyographic (EMG) gradients and movement-related brain activity: significance for motor control, cognitive functions, and certain psychopathologies," *International Journal of Psychophysiology*, vol. 38, no. 2, pp. 143–207, Nov. 2000, doi: 10.1016/S0167-8760(00)00113-6.

[41] J. Wijsman, B. Grundlehner, H. Liu, H. Hermens, and J. Penders, "Towards mental stress detection using wearable physiological sensors," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2011, pp. 1798–1801, doi: 10.1109/IEMBS.2011.6090512.

[42] J. N. Mandrekar, "Receiver Operating Characteristic Curve in Diagnostic Test Assessment," *Journal of Thoracic Oncology*, vol. 5, no. 9, pp. 1315–1316, Sep. 2010, doi: 10.1097/JTO.0b013e3181ec173d.

Appendix E

An Automated Approach for the Detection of Alzheimer's Disease From Resting State Electroencephalography

Frontiers in Neuroinformatics

An automated approach for the detection of Alzheimer's disease from resting state electroencephalography

Eduardo Perez-Valero [1,3*], Christian A. Morillas [1,3], Miguel A. Lopez-Gordo [2,3], Ismael Carrera-Muñoz [4], Samuel López-Alcalde [5] and Rosa M. Vílchez-Carrillo [4]

[1]*Department of Computers Architecture and Technology, University of Granada, Granada, Spain*

[2]*Department of Signal Theory, Telematics, and Communications, University of Granada, Granada, Spain*

[3]*Brain Computer Interface Laboratory, Research Center for Information and Comunications Technologies, University of Granada, Granada, Spain*

[4]*Cognitive Neurology Group, Hospital Universitario Virgen de las Nieves, Granada, Spain*

[5]*Hospital Universitario San Rafael, Granada, Spain*

Correspondence*:
M. A. López Gordo
malg@ugr.es

ABSTRACT

Early detection is crucial to control the progression of Alzheimer's disease and to postpone intellectual decline. Most current detection techniques are costly, inaccessible, or invasive. Furthermore, they require laborious analysis, what delays the start of medical treatment. To overcome this, researchers have recently investigated AD detection based on electroencephalography, a non-invasive neurophysiology technique, and machine learning algorithms. However, these approaches typically rely on manual procedures such as visual inspection, that requires additional personnel for the analysis, or on cumbersome EEG acquisition systems. In this paper, we performed a preliminary evaluation of a fully-automated approach for AD detection based on a commercial EEG acquisition system and an automated classification pipeline. For this purpose, we recorded the resting state brain activity of twenty-six participants from three groups: mild AD, mild cognitive impairment (MCI-non-AD), and healthy controls. First, we applied automated data-driven algorithms to reject EEG artifacts. Then, we obtained spectral, complexity, and entropy features from the preprocessed EEG segments. Finally, we assessed two binary classification problems: mild AD vs controls, and MCI-non-AD vs controls, through leave-one-subject-out cross-validation. The preliminary results that we obtained are comparable to the best reported in literature, what suggests that AD detection could be automatically detected through automated processing and commercial EEG systems. This is promising, since it may potentially contribute to reducing costs related to AD screening, and to shortening detection times, what may help to advance medical treatment.

Keywords: Alzheimer's disease, EEG, machine learning, disease detection, classification

1 INTRODUCTION

Alzheimer's disease (AD) is a neurogenerative disease that represents 60%-70% of the dementia cases Amezquita-Sanchez et al., 2019). According to the 2018 World Alzheimer Report, more than fifty million people suffer from dementia and this prevalence is likely to triple by 2050 (Patterson, 2018). AD patients experience a progressive decline in multiple cognitive areas, such as memory, orientation, and reasoning, to the extent of interfering with their daily-life activities. Although this disease was initially documented in 1906, its etiology is still uncertain, and a final diagnosis can only be performed at brain autopsy. However, according to the literature, two major hallmarks begin to form before the impairment is noticeable: amyloid plaques and neurofibrillary tangles (Serrano-Pozo et al., 2011). Amyloid plaques are deposits of proteins that lose their normal structure and accumulate around the neurons. Likewise, neurofibrillary tangles are thickened fibrils that surround the neuron nucleus. In this context, mild cognitive impairment (MCI) is often referred as a condition between normal aging and AD (Petersen et al., 2001). MCI patients experience unexpected memory losses for age but these do not interfere with their daily-life activities. Nonetheless, MCI patients transition to AD at a faster rate than healthy individuals of the same age. Although presently there is no cure for AD, early detection can help control the progression of the disease and postpone intellectual decline. This, alongside the high prevalence of the disease, evidence the need for non-invasive accessible detection techniques.

The purpose of AD detection techniques is to reveal the physical and cognitive symptoms linked to the disease. Traditionally, this is accomplished through neuropsychological tests and medical procedures. Neuropsychological tests are designed to evaluate cognitive areas affected early in the AD course, such as memory, language, and orientation (Carnero-Pardo et al., 2007; Ciesielska et al., 2016; Mat´ıas-Guiu et al., 2017). Although neuropsychological tests are extensively applied, previous works have evidenced their lack of sensitivity and high variability (Kuslansky et al., 2004; Mendiondo et al., 2000). Conversely, medical procedures aim to unfold the damages produced to specific brain structures. Among these procedures, cerebrospinal fluid (CSF) analysis is the most reliable. However, this fluid is extracted via lumbar puncture, an invasive medical procedure with reported side effects (Virhammar et al., 2012). Alternatively, medical imaging techniques are designed to render functional or static images of the brain. The most extended techniques include: magnetic resonance imaging (MRI) (Farooq et al., 2017; Hojjati et al., 2017), positron emission tomography (PET) (Li et al., 2015; Lu et al., 2018), and single photon emission computed tomography (SPECT) (Bi and Wang, 2019; Górriz et al., 2011). Although these techniques yield accurate results, they present some drawbacks: they usually involve long waiting lists, their analysis is often based on visual inspection, and some of them require invasive procedures.

In this context, for the past decade researchers have studied the role of electroencephalography (EEG) for AD detection. EEG is a neurophysiology technique to measure the electrical activity of the brain through electrodes placed on the scalp. EEG is portable, non-invasive, and affordable compared to medical imaging procedures. Consequently, it represents a promising approach for the detection of neurological diseases. Indeed, researchers have recently combined EEG signal processing and machine learning algorithms to discriminate AD and MCI patients from age matched controls (Aghajani et al., 2013; Cassani et al., 2017; Durongbhan et al., 2019; Fiscon et al., 2018; Ieracitano et al., 2020; Kashefpoor et al., 2016; Khatun et al., 2019; McBride et al., 2013; Ruiz-Gómez et al., 2018; Trambaiolli et al., 2017, 2011; Perez-Valero et al., 2022; Wang et al., 2015). These works typically analyze the EEG signals in terms of spectral content, complexity, and synchronization, since previous studies have found these features are affected in AD patients. For instance, in (Gallego-Jutgla` et al., 2015), the authors classified AD patients and controls based on frequency and power features using linear discriminant analysis. Likewise, in (Wang et al., 2015), researchers performed cluster analysis on power spectral density (PSD) and coherence features to differentiate AD patients and controls. Additionally, the authors of (McBride et al., 2015) and (McBride et al., 2013) implemented a support vector machine (SVM) to classify early stage AD and MCI patients based on coherence and entropy measures, respectively.

These works demonstrate the scientific community has made unquestionable progress towards EEG-based AD detection. Furthermore, recent advances in computational neuroscience have provided new insights about the collective behavior of the brain by attempting to explain the mechanisms driving the interaction between neuronal and synaptic processes, what may contribute to deepening our understanding of how neurodegeneration evolves over the course of AD (Jones et al., 2022; Caligiore et al., 2020). Indeed, in recent studies, researchers endeavor to bridge the gap in efficiency and cognitive skills between actual models and their biological equivalents (Yang et al., 2022a,b). In this context, to understand the complexity of the human brain, models are required to account for the morphological features responsible for neural dynamics and the high-level scale of the human neural network. For this purpose, approaches such as (Yang et al., 2020), based on a reduced compartmental model, have been proposed recently. Furthermore, to implement large-scale neuromorphic systems, online learning and fault-tolerant operation is mandatory (Vu et al., 2019; Yang et al., 2021). With regard to actual AD detection approaches, there are still methodological decisions that hinder the application of automated detection models. This includes the use of clinical EEG acquisition systems, EEG artifact cleaning through visual inspection and annotations, and complex electrode montages. To tackle this, in this paper, we wanted to perform a preliminary evaluation of a automated EEG classification approach for AD. To simplify the electrode setup, we used a commercial EEG acquisition device with only sixteen electrodes. Furthermore, to automatically process the acquired signals, we applied the Autoreject algorithm and independent component analysis (ICA). Then, we extracted spectral and complexity features from all the EEG channels. Finally, to assess the general performance of our model, we implemented leave-one-subject-out (LOSO) cross-validation. In order to evaluate our approach, we conducted a preliminary study in collaboration with the Cognitive and Behavioral Neurology Unit (CBNU) at Hospital Universitario Virgen de las Nieves de Granada (Spain) to discriminate MCI-non-AD patients, mild AD patients, and healthy controls. Although the sample size that we studied is reduced, the results that we obtained are promising, as our automated approach yielded classification results that are comparable to the best in literature. We believe the results derived from this work may contribute to opening the door to forthcoming ubiquitous, affordable, and accurate detection of mild AD and MCI-non-AD, nonetheless, further studies must validate the conclusions drawn in this study.

This paper is structured as follows: first, we report the methods and materials that we employed throughout the study; then, we describe the results that we obtained; subsequently, we discuss these results, and we compare our approach with analogous studies; finally, we draw conclusions, evaluate the impact and limitations of this research, and provide some guidelines for future studies.

2 MATERIALS AND METHODS

Throughout the following subsections, we describe the main aspects of the study methodology, including participants, experimental procedure and setup, EEG signal processing, feature extraction, and classification.

2.1. Participants

Twenty-six volunteers participated in the study. However, due to poor signal quality (participants 21 and 26), lack of cooperation (participants 9 and 18), and presence of a condition that may have impacted the

analysis (participant 12), we excluded the data from five participants. Consequently, we only studied the recordings from twenty-one participants. The head of the CBNU at Hospital Universitario Virgen de las Nieves de Granada recruited the participants the week before the experiment. Table 1 displays the group, age, and sex distributions of the participants. Thirteen of the participants were patients of the neurology unit at Hospital Universitario Virgen de las Nieves. Due to memory complaints, all these patients had undergone one of the following two medical trials during the past year: (a) measurement of Aβ42, p-tau, and total tau in CSF, or (b) β-amyloid PET (florbetaben PET). We cataloged the participants whose medical trials yielded positive biomarkers as mild AD, otherwise, we cataloged them as MCI-non-AD. The remaining participants were healthy controls. We conducted the study following a protocol approved by the local ethics committee at Hospital Universitario Virgen de las Nieves de Granada. Additionally, the participants signed an informed consent before the experiment onset, and CBNU personnel monitored them throughout the study session.

2.2 Experimental procedure and setup

First, we asked the participants to read and sign the informed consent. Then, we briefed them on the study details and the basics of EEG acquisition. It is worth to note that the participants of this study were recruited with two objectives in mind: (1) go through a cognitive test, and (2) participate in the study reported in this paper. The cognitive test consisted of two tasks of approximately ten minutes whose details are out of the scope of this paper. In parallel, we recorded three minutes of the eye-open resting state activity of the participants in three key moments: before the first task, after the first task, and after the second task. To avoid edge effects, we only considered the central two-minute window of each recording. For the analysis reported in this paper, we concatenated the three recordings, thus, we ended up with an effective six-minute EEG recording per participant.

For the acquisition, we used the Versatile system by Bitbrain (Zaragoza, Spain), a commercial wearable device consisting of an EEG cap with semi-dry electrodes and a Bluetooth acquisition module that works at a fixed sampling rate of 256 Hz. This sampling rate has been previously used in related studies (Amezquita-Sanchez et al., 2019; Garn et al., 2017; Gouw et al., 2017; Ieracitano et al., 2019). For the electrode montage, we selected sixteen channels located at positions Fp1, Fp2, F5, Fz, F6, T7, T8, C3, Cz, C4, P5, Pz, P6, O1, Oz, and O2 of the extended 10-20 International System, and we referenced them to the left ear lobe. We used this setup to uniformly cover the scalp according to previous AD detection studies (Fiscon et al., 2018; Kulkarni and Bairagi, 2017; Wang et al., 2015). Figure 1 shows the electrode montage selected for this study alongside the Versatile EEG system.

2.3 Signal preprocessing

We applied the procedure described in this subsection to the six-minute EEG recordings of the participants. First, we filtered the raw EEG using a 1690 order FIR filter with 1-45 Hz bandpass and zero phase-shift. We used a FIR filter instead of an IIR filter because our study did not involve high throughput constraints (Ifeachor and Jervis, 2002), and we prioritized filter stability and control. Then, we segmented the filtered EEG into four-second epochs without overlapping. According to the systematic review on resting state EEG for AD diagnosis by Cassani et al. (Cassani et al., 2018), this epoch length lays within one of the most common epoch length ranges considered in literature (see Table 17 in the aforementioned review article).

Regarding artifact rejection, in order to keep our approach automated, we applied the Autoreject algorithm. Autoreject is a data-driven artifact rejection algorithm that combines cross-validation and Bayesian optimization to find a separate threshold per channel. It also marks bad trials when most channels display high-amplitude artifacts, and allows channel interpolation. Therefore, bad data segments can be repaired instead of discarded. In terms of performance, Autoreject has been validated against multiple datasets, and reportedly performed equally or better than common artifact processing approaches. We refer the interested reader to (Jas et al., 2017), where the developers of the algorithm provide an exhaustive interpretation of this validation. Additionally, we also implemented blink artifact rejection through independent component analysis (ICA). ICA enables the decomposition of a signal made from the combination of multiple sources into those sources and a mixing matrix (Echtioui et al., 2020; Schlink et al., 2017). Since EEG represents the combination of internal neurological sources, ICA is usually applied to identify and remove artifactual components such as blinks. Blink components display high variance and present a spatial distribution towards the frontotemporal region of the head (Shahbakhti et al., 2022). Therefore, to identify them, we followed an automated electro-oculogram proxy approach. To this end, we determined the ICA component whose correlation with Fp1 time series was the highest, and we reconstructed the EEG without that component. Figure 2 represents the signal processing pipeline described in this subsection.

2.4 Feature extraction

After preprocessing, we used MNE Python toolbox to extract three features per channel from each of the preprocessed four-second EEG epochs: relative power (RP) in the five main EEG bands, Hjorth complexity (HC), and spectral entropy (SE). Consequently, we extracted a total of 112 features (80 for the RP, 16 for the HC, and 16 for the SE) per epoch. We considered these features since they have been already applied in analogous studies (Houmani et al., 2018; McBride et al., 2014; Poil et al., 2013; Trambaiolli et al., 2017; Wang et al., 2015). RP represents the fraction of the total power of the signal that is contained in a particular frequency band. To estimate this parameter, we applied the default MNE Python settings, that include the use of Welch's method for the calculation of the PSD. On the other hand, HC is one of the three Hjorth parameters (activity, mobility, and complexity), and is calculated as the ratio between the mobility of the first derivative of the signal and the mobility of the signal itself (Päivinen et al., 2005). Finally, SE is defined as the Shannon entropy of the power spectrum of the signal. SE represents the uniformity of the power spectrum distribution, and, hence, the irregularity of the EEG. SE is minimal for a pure sine wave and maximal for white noise. (Inouye et al., 1991). The formulas for the extracted features are presented in Eq. (1) to Eq. (3).

$$RP = \frac{\sum_{f_i}^{f_o} P}{\sum_{\forall f} P} \tag{1}$$

$$HC = \frac{\sigma_{s''}/\sigma_{s'}}{\sigma_{s'}/\sigma_{s}} \tag{2}$$

$$SE = -\sum_{f} S(f) * \log_2 S(f) \tag{3}$$

In Eq.(1), fi and fo represent the lower and upper frequencies of an EEG band, and the denominator represents the total power of the EEG signal. We estimated RP for the main EEG frequency bands: delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (> 30 Hz). In Eq.(2), σs, σs′ , and

$\sigma s''$ represent the standard deviations of the signal epoch under analysis, of its first, and of its second derivative, respectively. Finally, in Eq.(3), f represents frequency, and S represents the normalized power spectrum.

To create the feature matrix, we vertically concatenated the features extracted for each participant. Consequently, we ended up with a feature matrix with 112 columns (features) and as many rows as the total number of epochs from all the participants (see Figure 3). Then, we averaged every S consecutive rows in the feature matrix following to the approach described in (Fraga et al., 2013). We followed this approach with two objectives in mind: to enhance the signal to noise ratio of the features, and to reduce the size of the feature matrix. We evaluated four values for S: 6, 8, 10, and 12, and we reported the best for each classification problem (MCI-non-AD vs NC and mild-AD vs NC) in the Results section. Therefore, for each classification problem, we utilized a feature matrix (RP, HC, and SE) and a target array holding the cohort corresponding to each row (epoch) in the feature matrix.

2.5 Classification

To solve the two binary classification problems examined in this study, we used scikit-learn Python module (Pedregosa et al., 2011) to implement a three-step classification pipeline including a feature scaler, a feature selector, and a classifier. The feature scaler normalizes each feature in the range between zero and one, in a way that those with higher order of magnitude are not favored during classification. Subsequently, the feature selector selects the most relevant features according to a particular strategy. In our case, we applied the chi-square test included as part of scikit-learn feature selection module to find the features most related to the target. We selected this method due to its intrinsic speed. We refer the interested reader to (Chandrashekar and Sahin, 2014) for a comprehensive overview of feature selection strategies. Finally, for the classifier, we evaluated two algorithms: SVM with a radial basis function kernel and logistic regression (LR). To find the best hyperparameters for the classification pipeline we implemented grid search. For the SVM, we explored different values for the regularization parameter (C) and the kernel coefficient (γ). The C parameter refers to the strength of the regularization that is applied to avoid overfitting. In particular, C influences the selection of the SVM hyperplane margin and its amplitude is inversely proportional to the regularization strength. On the other hand, the γ parameter defines the kernel function used by the SVM algorithm to define a notion of similarity between input data and transform the data in the feature space. Likewise, for the LR, we explored different values for the regularization parameter (C) and the weights associated with the classes. In the case of logistic regression, the regularization parameter controls a penalty term added to the loss function in order to avoid overfitting by penalizing extreme parameter weights. On the other hand, the class weight parameter allow the algorithm to associate different weights to the classes based on their support. We examined the default option using no class weights, and the balanced option that adjusts the weights inversely proportional to class frequencies. Figure 3 illustrates the pipeline described in this subsection.

To evaluate the generalization ability of the classifier, we performed cross-validation using a LOSO strategy. Under this strategy, data is split into as many folds as participants. For each fold, the training set includes data from all the participants but one, whose data is reserved for the test set. Hence, information from a participant is never in the training and the test set simultaneously, what prevents from positive bias. Furthermore, according to the comprehensive review by Cassani et al. (Cassani et al., 2018), this cross-validation strategy is the most extended across the studies that aim to classify AD groups from resting state EEG.

For the sake of clarity, it is important to note that, during cross-validation, the parameters corresponding to the three stages of the classification pipeline (feature scaling, feature selection, and classification) are estimated using the training set, and subsequently, those parameters are applied on the test set. This avoids overfitting and prevents from artificial positive bias.

3 RESULTS

In this section, we present the results that we obtained for the two binary discrimination problems evaluated in this study: mild AD vs NC and MCI-non-AD vs NC, based on the 16-channel resting state EEG recordings that we acquired (see Section 2.2).

Figure 4 shows the average F1-score obtained during cross-validation as a function of the number of consecutive epochs averaged after feature extraction (see Section 2.4). To create this figure, we evaluated the classification pipeline on four different feature matrices (considering 6, 8, 10, and 12 consecutive epochs for the average).

Table 2 presents the hyperparameter values that we inspected through grid-search. We have also included the combination of hyperparameters that yielded the best cross-validation results. As shown in the right-most column of this table, the SVM classifier yielded the best performance for the two binary discrimination problems that we considered. Note that, in this table, we have not reported the best hyperparameters for the LR because the SVM outperformed this classifier on the two classification problems.

Table 3 holds the confusion matrices for the mild AD vs NC and the MCI-non-AD vs NC problems. As outlined in the previous paragraph, these results were yielded by the SVM classifier. The values outside the parenthesis represent the epoch-level confusion matrix, whilst the values inside the parenthesis represent the participant-level confusion matrix. To estimate the epoch-level matrix, we gathered the predictions and the true labels from all the cross-validation test sets, and we computed the percentage of correct predictions per class. Alternatively, for the participant-level matrix, we evaluated how well the classifier grouped each participant into one of the two groups for each binary classification problem. Consequently, if the classifier correctly guessed more than 50% of the epochs from a participant, we considered it correctly classified the participant; otherwise, we considered it missclassified the participant.

Table 4 shows the per-cohort cross-validation metrics yielded by the classification pipeline for the mild AD vs NC and the MCI-non-AD vs NC problems. We derived these metrics from the epoch-level confusion matrices described above.

4 DISCUSSION

The goal of this study was to perform a preliminary evaluation of an automated approach for the discrimination of AD cohorts. For this purpose, we recorded the EEG of a group of volunteers, and we proposed an approach based on a commercial EEG device with a reduced montage and an automated classification pipeline. The preliminary results that we obtained from a reduced sample suggest that such an approach can precisely discriminate mild AD and MCI-non-AD participants from healthy controls.

In Figure 4, we examined the epoch average procedure that we applied to the feature matrix. According to this figure, performance increased with the number of averaged epochs and then plateaued for both classification problems, what supports the application of the epoch average step described in Section 2.4. Regarding the classification problem of mild AD vs NC, in view of the results presented in Table 3, our approach yielded promising results, both at the epoch and the participant levels. More specifically, according to the results presented in Table 4, recall for the mild AD cohort was the highest performance metric (0.88). This is particularly important because it evidences the capability of the classifier to detect the participants affected by the disease, what is crucial in studies that aim to discriminate clinical cohorts. Similarly, for the classification of MCI-non-AD vs NC, the confusion matrices presented in Table 3 illustrate the sound performance of the classifier both at the epoch and the participant levels. According to the metrics reported in Table 4, the average performance of the classifier was higher for the MCI-non-AD vs NC problem than for the mild AD vs NC problem, as evidenced by average F1-scores of 0.96 and 0.86, respectively. An analogous study also reported better discrimination performance for the MCI vs NC classification problem (Fiscon et al., 2018), although the reasons behind this result remain still unexplained. An evaluation of the presented approach on a larger sample size could contribute to elucidating this circumstance.

In terms of recording conditions, most studies in literature typically use between 17 and 32 electrodes (Cassani et al., 2018). However, to minimize participant distress, we selected only sixteen electrodes. A similar number of electrodes was considered in related works (Chen et al., 2018; Kulkarni and Bairagi, 2017; Yu et al., 2020; Wang et al., 2015). With respect to epoch length, since EEG signals are non-stationary, we utilized an intermediate length of four seconds (Cassani et al., 2018), comparable to the duration selected in analogous works (Coronel et al., 2017; Durongbhan et al., 2019; Mammone et al., 2017). Regarding artifact rejection, works on AD detection traditionally have applied manual epoch selection for artifact removal (Azami et al., 2017; Chen et al., 2018; Mammone et al., 2017; Ruiz-Gómez et al., 2018; Simons et al., 2015), what may be counterproductive from the early detection standpoint. Alternatively, since we wanted to perform a preliminary evaluation of an automated classifier, we decided to perform artifact processing through Autoreject and automated ICA. The rationale behind the use of Autoreject was the interpolation of bad data spans supported by the algorithm, and its capability to perform data-driven rejection of artifactual segments. Regarding ICA, multiple works have used this analysis in a semi-automated way to identify and remove artifactual components (Vecchio et al., 2017). Alternatively, in other works, researchers implemented an automated version of ICA (Echtioui et al., 2020). We followed the latter approach to remove blink artifacts and comply with the automation constraints of our approach. Particularly, we identified the blink artifact component through correlation with the Fp1 channel time series. We selected this channel because it has the position that most resembles an electro-oculogram channel. With respect to cross-validation, the three main strategies adopted in literature are leave-one-out (Simons and Abásolo, 2017; Yang et al., 2019), k-fold (Durongbhan et al., 2019; Ieracitano et al., 2020), and leave-one-subject-out (Fiscon et al., 2018; Ruiz-Gómez et al., 2018; Trambaiolli et al., 2017). We selected LOSO because with this cross-validation strategy, data from a participant is not in the training and the test set simultaneously. This prevents the classifier from positive bias and thus, from rendering unrealistically good predictions. Lastly, in terms of performance, our results are comparable to similar works in literature (Fiscon et al., 2018; Kulkarni and Bairagi, 2017; Mazaheri et al., 2018; Trambaiolli et al., 2017), or even superior in some cases (Amezquita-Sanchez et al., 2019; Cassani et al., 2017; Liu et al., 2020; Morabito et al., 2016).

The preliminary results that we obtained hint at the potential of fully automated AD discrimination approaches. The use of wearable commercial EEG acquisition systems may enhance the accessibility of electrophysiology-based screening. Additionally, automated processing techniques such as those considered in the approach presented in this paper may contribute to reducing the personnel required for AD screening practice. Jointly, this could contribute to reducing the costs associated to AD screening and to speeding-up medical treatment. Consequently, we believe this preliminary study along with further research, may help to open the door for fast and portable EEG-based AD detection approaches that could integrate into the clinical ecosystem in the near future. Nonetheless, this preliminary study also presents some limitations. First, the sample size that we considered is too small to yield definitive conclusions. Therefore, future works must evaluate the presented approach on a larger sample to validate the conclusions drawn in this study. Furthermore, the participants that we recruited for this study do not represent the typical patient who attend the neurology services. Often, patients suffer from additional pathologies accompanied by symptoms that could overlap with those from dementia. Obviously, the inclusion of such participants would represent an important challenge for researchers, nonetheless, we believe the study of typical patients is crucial in order to potentially transfer these approaches into daily medical practice. Finally, although an automated AD detection approach could bring benefits in terms of economical costs and waiting times, it would also require the clinicians to be properly trained to use a commercial EEG acquisition system (wear the EEG headset, monitor channel impedance, and control the graphical user interface of the device, among other tasks).

5 CONCLUSIONS

In this paper, we preliminarily evaluated the feasibility of an automated pipeline based on EEG activity for the discrimination of mild AD and MCI-non-AD. For this purpose, we recorded the resting state activity of a reduced sample of volunteers using a commercial EEG device, and we designed our approach to automatically perform signal processing, feature extraction, and classification. We applied the Autoreject algorithm and ICA for artifact rejection. Then, we estimated the relative power, the Hjorth complexity, and the spectral entropy of the preprocessed epochs. Lastly, we assessed two binary classifiers (SVM and LR) for the discrimination of the three cohorts of interest via leave-one-subject-out cross-validation. The results that we obtained are comparable to the best reported in literature, what is promising towards the implementation of automated AD detection approaches based on commercial acquisition devices. Since we analyzed a reduced sample size, further studies must evaluate the presented approach on larger samples to validate the conclusions yielded in this paper. Nonetheless, the results obtained in this work are promising, as they suggest that AD detection could be performed automatically using a few minutes of resting state EEG activity and a commercial acquisition device. With this in mind, we believe the future inclusion of this kind of approaches into AD screening practice could contribute to reducing the costs linked to AD detection, and also enable the early detection of the disease, what could potentially advance medical treatments. Nonetheless, further research is also required to investigate the feasibility of automated EEG-based approaches in participants that suffer from additional pathologies whose symptoms overlap with those of dementia. This is often the case in daily-life neurological practice, hence, such cases must be comprehensively studied before future patients can take advantage from fast, accurate, and affordable detection techniques that enhance their standards of living.

ACKNOWLEDGMENTS

This work was supported by project PGC2018-098813-B-C31 funded by the Spanish Ministry of Science, Innovation and Universities, by European Regional Development Funds, and by project B-TIC-352-UGR20, co-funded by the Operative Program FEDER 2014-2020 and the Economy, Universities and Science Office of the Andalusian Regional Government. The authors would like to especially thank the Hospital Universitario Virgen de las Nieves de Granada and the participants involved for their collaboration in this study.

REFERENCES

Aghajani, H., Zahedi, E., Jalili, M., Keikhosravi, A., and Vahdat, B. V. (2013). Diagnosis of Early Alzheimer's Disease Based on EEG Source Localization and a Standardized Realistic Head Model. IEEE Journal of Biomedical and Health Informatics 17, 1039–1045. doi:10.1109/JBHI.2013.2253326

Amezquita-Sanchez, J. P., Mammone, N., Morabito, F. C., Marino, S., and Adeli, H. (2019). A novel methodology for automated differential diagnosis of mild cognitive impairment and the Alzheimer's disease using EEG signals. Journal of Neuroscience Methods 322, 88–95. doi:10.1016/j.jneumeth.2019.04.013

Azami, H., Aba´solo, D., Simons, S., and Escudero, J. (2017). Univariate and Multivariate Generalized Multiscale Entropy to Characterise EEG Signals in Alzheimer's Disease. Entropy 19, 31. doi:10.3390/e19010031

Bi, X. and Wang, H. (2019). Early Alzheimer's disease diagnosis based on EEG spectral images using deep learning. Neural Networks 114, 119–135. doi:10.1016/j.neunet.2019.02.005

Caligiore, D., Silvetti, M., D'Amelio, M., Puglisi-Allegra, S., and Baldassarre, G. (2020). Computational Modeling of Catecholamines Dysfunction in Alzheimer's Disease at Pre-Plaque Stage. Journal of Alzheimer's Disease 77, 275–290. doi:10.3233/JAD-200276. Publisher: IOS Press

Carnero-Pardo, C., Sa´ez-Zea, C., Montiel Navarro, L., Del Saz, P., Feria-Vilar, I., Pe´rez-Navarro, M. J., et al. (2007). Utilidad diagno´stica del Test de las Fotos (Fototest) en deterioro cognitivo y demencia. Neurologiıa 22, 860–869

Cassani, R., Estarellas, M., San-Martin, R., Fraga, F. J., and Falk, T. H. (2018). Systematic Review on Resting-State EEG for Alzheimer's Disease Diagnosis and Progression Assessment. Publication Title: Disease Markers

Cassani, R., Falk, T. H., Fraga, F. J., Cecchi, M., Moore, D. K., and Anghinah, R. (2017). Towards automated electroencephalography-based Alzheimer's disease diagnosis using portable low-density devices. Biomedical Signal Processing and Control 33, 261–271. doi:10.1016/j.bspc.2016.12.009

Chandrashekar, G. and Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering 40, 16–28. doi:10.1016/j.compeleceng.2013.11.024

Chen, Y., Cai, L., Wang, R., Song, Z., Deng, B., Wang, J., et al. (2018). DCCA cross-correlation coefficients reveals the change of both synchronization and oscillation in EEG of Alzheimer disease patients. Physica A: Statistical Mechanics and its Applications 490, 171–184. doi:10.1016/j.physa.2017.08.009

Ciesielska, N., Sokołowski, R., Mazur, E., Podhorecka, M., Polak-Szabela, A., and Kedziora-Kornatowska, K. (2016). Is the Montreal Cognitive Assessment (MoCA) test better suited than the Mini-Mental State Examination (MMSE) in mild cognitive impairment (MCI) detection among people aged over 60? Meta-analysis. Psychiatria Polska 50, 1039–1052. doi:10.12740/PP/45368

Coronel, C., Garn, H., Waser, M., Deistler, M., Benke, T., Dal-Bianco, P., et al. (2017). Quantitative EEG Markers of Entropy and Auto Mutual Information in Relation to MMSE Scores of Probable Alzheimer's Disease Patients. Entropy 19, 130. doi:10.3390/e19030130

Durongbhan, P., Zhao, Y., Chen, L., Zis, P., De Marco, M., Unwin, Z. C., et al. (2019). A Dementia Classification Framework Using Frequency and Time-Frequency Features Based on EEG Signals. IEEE Transactions on Neural Systems and Rehabilitation Engineering 27, 826–835. doi:10.1109/TNSRE. 2019.2909100

Echtioui, A., Zouch, W., Ghorbel, M., Slima, M. B., Hamida, A. B., and Mhiri, C. (2020). Automated EEG Artifact Detection Using Independent Component Analysis. In 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) (2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)), 1–5. doi:10.1109/ATSIP49331. 2020.9231574

Farooq, A., Anwar, S., Awais, M., and Rehman, S. (2017). A deep CNN based multi-class classification of Alzheimer's disease using MRI. In 2017 IEEE International Conference on Imaging Systems and Techniques (IST) (2017 IEEE International Conference on Imaging Systems and Techniques (IST)), 1–6. doi:10.1109/IST.2017.8261460

Fiscon, G., Weitschek, E., Cialini, A., Felici, G., Bertolazzi, P., De Salvo, S., et al. (2018). Combining EEG signal processing with supervised methods for Alzheimer's patients classification. BMC Medical Informatics and Decision Making 18, 35. doi:10.1186/s12911-018-0613-y

Fraga, F. J., Falk, T. H., Trambaiolli, L. R., Oliveira, E. F., Pinaya, W. H. L., Kanda, P. A. M., et al. (2013). Towards an EEG-based biomarker for Alzheimer's disease: Improving amplitude modulation analysis features. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (2013 IEEE International Conference on Acoustics, Speech and Signal Processing), 1207–1211. doi:10.1109/ ICASSP.2013.6637842

Gallego-Jutglà, E., Solé-Casals, J., Vialatte, F.-B., Elgendi, M., Cichocki, A., and Dauwels, J. (2015). A hybrid feature selection approach for the early diagnosis of Alzheimer's disease. Journal of Neural Engineering 12, 016018. doi:10.1088/1741-2560/12/1/016018

Garn, H., Coronel, C., Waser, M., Caravias, G., and Ransmayr, G. (2017). Differential diagnosis between patients with probable Alzheimer's disease, Parkinson's disease dementia, or dementia with Lewy bodies and frontotemporal dementia, behavioral variant, using quantitative electroencephalographic features. Journal of Neural Transmission 124, 569–581. doi:10.1007/s00702-017-1699-6

Gouw, A. A., Alsema, A. M., Tijms, B. M., Borta, A., Scheltens, P., Stam, C. J., et al. (2017). EEG spectral analysis as a putative early prognostic biomarker in nondemented, amyloid positive subjects. Neurobiology of Aging 57, 133–142. doi:10.1016/j.neurobiolaging.2017.05.017

Górriz, J. M., Segovia, F., Ramírez, J., Lassl, A., and Salas-Gonzalez, D. (2011). GMM based SPECT image classification for the diagnosis of Alzheimer's disease. Applied Soft Computing 11, 2313–2325. doi:10.1016/j.asoc.2010.08.012

Hojjati, S. H., Ebrahimzadeh, A., Khazaee, A., and Babajani-Feremi, A. (2017). Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM. Journal of Neuroscience Methods 282, 69–80. doi:10.1016/j.jneumeth.2017.03.006

Houmani, N., Vialatte, F., Gallego-Jutglà, E., Dreyfus, G., Nguyen-Michel, V.-H., Mariani, J., et al. (2018). Diagnosis of Alzheimer's disease with Electroencephalography in a differential framework. PLoS ONE 13. doi:10.1371/journal.pone.0193607

Ieracitano, C., Mammone, N., Bramanti, A., Hussain, A., and Morabito, F. C. (2019). A Convolutional Neural Network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings. Neurocomputing 323, 96–107. doi:10.1016/j.neucom.2018.09.071

Ieracitano, C., Mammone, N., Hussain, A., and Morabito, F. C. (2020). A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. Neural Networks 123, 176–190. doi:10.1016/j.neunet.2019.12.006

Ifeachor, E. C. and Jervis, B. W. (2002). Digital Signal Processing: A Practical Approach (Pearson Education)

Inouye, T., Shinosaki, K., Sakamoto, H., Toi, S., Ukai, S., Iyama, A., et al. (1991). Quantification of EEG irregularity by use of the entropy of the power spectrum. Electroencephalography and Clinical Neurophysiology 79, 204–210. doi:10.1016/0013-4694(91)90138-T

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., and Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. NeuroImage 159, 417–429. doi:10.1016/j.neuroimage.2017.06.030

Jones, D., Lowe, V., Graff-Radford, J., Botha, H., Barnard, L., Wiepert, D., et al. (2022). A computational model of neurodegeneration in Alzheimer's disease. Nature Communications 13, 1643. doi:10.1038/s41467-022-29047-4

Kashefpoor, M., Rabbani, H., and Barekatain, M. (2016). Automatic Diagnosis of Mild Cognitive Impairment Using Electroencephalogram Spectral Features. Journal of Medical Signals and Sensors 6, 25–32

Khatun, S., Morshed, B. I., and Bidelman, G. M. (2019). A Single-Channel EEG-Based Approach to Detect Mild Cognitive Impairment via Speech-Evoked Brain Responses. IEEE Transactions on Neural Systems and Rehabilitation Engineering 27, 1063–1070. doi:10.1109/TNSRE.2019.2911970

Kulkarni, N. N. and Bairagi, V. K. (2017). Extracting Salient Features for EEG-based Diagnosis of

Alzheimer's Disease Using Support Vector Machine Classifier. IETE Journal of Research 63, 11–22. doi:10.1080/03772063.2016.1241164

Kuslansky, G., Katz, M., Verghese, J., Hall, C. B., Lapuerta, P., LaRuffa, G., et al. (2004). Detecting dementia with the Hopkins Verbal Learning Test and the Mini-Mental State Examination. Archives of Clinical Neuropsychology , 16

Li, F., Tran, L., Thung, K.-H., Ji, S., Shen, D., and Li, J. (2015). A Robust Deep Model for Improved Classification of AD/MCI Patients. IEEE Journal of Biomedical and Health Informatics 19, 1610–1616. doi:10.1109/JBHI.2015.2429556

Liu, L., Zhao, S., Chen, H., and Wang, A. (2020). A new machine learning method for identifying Alzheimer's disease. Simulation Modelling Practice and Theory 99, 102023. doi:10.1016/j.simpat.2019.102023

Lu, D., Popuri, K., Ding, G. W., Balachandar, R., and Beg, M. F. (2018). Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. Medical Image Analysis 46, 26–34. doi:10.1016/j.media.2018.02.002

Mammone, N., De Salvo, S., Ieracitano, C., Marino, S., Marra, A., Corallo, F., et al. (2017). A Permutation Disalignment Index-Based Complex Network Approach to Evaluate Longitudinal Changes in Brain-Electrical Connectivity. *Entropy* 19, 548. doi:10.3390/e19100548

Mat´ıas-Guiu, J. A., Valles-Salgado, M., Rognoni, T., Hamre-Gil, F., Moreno-Ramos, T., and Mat´ıas-Guiu, J. (2017). Comparative Diagnostic Accuracy of the ACE-III, MIS, MMSE, MoCA, and RUDAS for Screening of Alzheimer Disease. *Dementia and Geriatric Cognitive Disorders* 43, 237–246. doi:10.1159/000469658

Mazaheri, A., Segaert, K., Olichney, J., Yang, J.-C., Niu, Y.-Q., Shapiro, K., et al. (2018). EEG oscillations during word processing predict MCI conversion to Alzheimer's disease. *NeuroImage: Clinical* 17, 188–197. doi:10.1016/j.nicl.2017.10.009

McBride, J., Zhao, X., Munro, N., Jicha, G., Smith, C., and Jiang, Y. (2015). Discrimination of Mild Cognitive Impairment and Alzheimer's Disease Using Transfer Entropy Measures of Scalp EEG. *Journal of healthcare engineering* 6, 55–70. doi:10.1260/2040-2295.6.1.55

McBride, J., Zhao, X., Munro, N., Smith, C., Jicha, G., and Jiang, Y. (2013). Resting EEG discrimination of early stage Alzheimer's disease from normal aging using inter-channel coherence network graphs. *Annals of Biomedical Engineering* 41, 1233–1242. doi:10.1007/s10439-013-0788-4

McBride, J. C., Zhao, X., Munro, N. B., Smith, C. D., Jicha, G. A., Hively, L., et al. (2014). Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease. *Computer Methods and Programs in Biomedicine* 114, 153–163. doi:10.1016/j.cmpb.2014.01.019

Mendiondo, M. S., Ashford, J. W., Kryscio, R. J., and Schmitt, F. A. (2000). Modelling Mini Mental State Examination changes in Alzheimer's disease. *Statistics in Medicine* 19, 1607–1616. doi:10.1002/(SICI) 1097-0258(20000615/30)19:11/12\textless1607::AID-SIM449\textgreater3.0.CO;2-O

Morabito, F. C., Campolo, M., Ieracitano, C., Ebadi, J. M., Bonanno, L., Bramanti, A., et al. (2016). Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings. In *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)* (2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)), 1–6. doi:10.1109/RTSI.2016.7740576

Patterson, C. (2018). World Alzheimer report 2018 Journal Abbreviation: The state of the art of dementia research: new frontiers

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830

Perez-Valero, E., Lopez-Gordo, M. A., Gutierrez, C. M., Carrera-Munoz, I., and Vilchez-Carrillo, R. M. (2022). A self-driven approach for multi-class discrimination in Alzheimer's disease based on wearableEEG. *Computer Methods and Programs in Biomedicine* 220, 106841. doi:10.1016/j.cmpb.2022.106841

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., et al. (2001). Current Concepts in Mild Cognitive Impairment. *Archives of Neurology* 58, 1985–1992. doi:10.1001/archneur. 58.12.1985

Poil, S.-S., de Haan, W., van der Flier, W. M., Mansvelder, H. D., Scheltens, P., and Linkenkaer-Hansen, K. (2013). Integrative EEG biomarkers predict progression to Alzheimer's disease at the MCI stage. *Frontiers in Aging Neuroscience* 5. doi:10.3389/fnagi.2013.00058

Päivinen, N., Lammi, S., Pitkänen, A., Nissinen, J., Penttonen, M., and Grönfors, T. (2005). Epileptic seizure detection: A nonlinear viewpoint. *Computer Methods and Programs in Biomedicine* 79, 151–159. doi:10.1016/j.cmpb.2005.04.006

Ruiz-Gómez, S. J., Gómez, C., Poza, J., Gutiérrez-Tobal, G. C., Tola-Arribas, M. A., Cano, M., et al. (2018). Automated Multiclass Classification of Spontaneous EEG Activity in Alzheimer's Disease and Mild Cognitive Impairment. Entropy 20, 35. doi:10.3390/e20010035

Schlink, B. R., Peterson, S. M., Hairston, W. D., König, P., Kerick, S. E., and Ferris, D. P. (2017). Independent Component Analysis and Source Localization on Mobile EEG Data Can Identify Increased Levels of Acute Stress. Frontiers in Human Neuroscience 11. doi:10.3389/fnhum.2017.00310

Serrano-Pozo, A., Frosch, M. P., Masliah, E., and Hyman, B. T. (2011). Neuropathological Alterations in Alzheimer Disease. Cold Spring Harbor Perspectives in Medicine 1, a006189.

doi:10.1101/cshperspecta006189. 01616

Shahbakhti, M., Beiramvand, M., Rejer, I., Augustyniak, P., Broniec-Wojcik, A., Wierzchon, M., et al. (2022). Simultaneous Eye Blink Characterization and Elimination From Low-Channel Prefrontal EEG Signals Enhances Driver Drowsiness Detection. IEEE journal of biomedical and health informatics 26, 1001–1012. doi:10.1109/JBHI.2021.3096984

Simons, S., Abasolo, D., and Escudero, J. (2015). Classification of Alzheimer's disease from quadratic sample entropy of electroencephalogram. Healthcare Technology Letters 2, 70–73. doi:10.1049/htl.2014.0106

Simons, S. and Abásolo, D. (2017). Distance-Based Lempel–Ziv Complexity for the Analysis of Electroencephalograms in Patients with Alzheimer's Disease. Entropy 19, 129. doi:10.3390/e19030129

Trambaiolli, L., Spolaôr, N., Lorena, A., Anghinah, R., and Sato, J. (2017). Feature selection before EEG classification supports the diagnosis of Alzheimer's disease. Clinical Neurophysiology 128, 2058–2067. doi:10.1016/j.clinph.2017.06.251

Trambaiolli, L. R., Lorena, A. C., Fraga, F. J., Kanda, P. A. M., Anghinah, R., and Nitrini, R. (2011). Improving Alzheimer's disease diagnosis with machine learning techniques. Clinical EEG and neuroscience 42, 160–165. doi:10.1177/155005941104200304

Vecchio, F., Miraglia, F., Piludu, F., Granata, G., Romanello, R., Caulo, M., et al. (2017). "Small World" architecture in brain connectivity and hippocampal volume in Alzheimer's disease: a study via graph theory from EEG data. Brain Imaging and Behavior 11, 473–485. doi:10.1007/s11682-016-9528-3

Virhammar, J., Cesarini, K. G., and Laurell, K. (2012). The CSF tap test in normal pressure hydrocephalus: evaluation time, reliability and the influence of pain. European Journal of Neurology 19, 271–276. doi:10.1111/j.1468-1331.2011.03486.x

Vu, T. H., Ikechukwu, O. M., and Ben Abdallah, A. (2019). Fault-Tolerant Spike Routing Algorithm and Architecture for Three Dimensional NoC-Based Neuromorphic Systems. IEEE Access 7, 90436–90452. doi:10.1109/ACCESS.2019.2925085. Conference Name: IEEE Access

Wang, R., Wang, J., Yu, H., Wei, X., Yang, C., and Deng, B. (2015). Power spectral density and coherence analysis of Alzheimer's EEG. Cognitive Neurodynamics 9, 291–304. doi:10.1007/s11571-014-9325-x

Yang, S., Bornot, J. M. S., Wong-Lin, K., and Prasad, G. (2019). M/EEG-Based Bio-Markers to Predict the MCI and Alzheimer's Disease: A Review From the ML Perspective. IEEE Transactions on Biomedical Engineering 66, 2924–2935. doi:10.1109/TBME.2019.2898871

Yang, S., Deng, B., Wang, J., Li, H., Lu, M., Che, Y., et al. (2020). Scalable Digital Neuromorphic Architecture for Large-Scale Biophysically Meaningful Neural Network With Multi-Compartment Neurons. IEEE Transactions on Neural Networks and Learning Systems 31, 148–162. doi:10.1109/

TNNLS.2019.2899936. Conference Name: IEEE Transactions on Neural Networks and Learning Systems

Yang, S., Gao, T., Wang, J., Deng, B., Azghadi, M. R., Lei, T., et al. (2022a). SAM: A Unified Self Adaptive Multicompartmental Spiking Neuron Model for Learning With Working Memory. *Frontiers in Neuroscience*

Yang, S., Tan, J., and Chen, B. (2022b). Robust Spike-Based Continual Meta-Learning Improved by Restricted Minimum Error Entropy Criterion. Entropy 24, 455. doi:10.3390/e24040455. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute

Yang, S., Wang, J., Deng, B., Azghadi, M. R., and Linares-Barranco, B. (2021). Neuromorphic Context-Dependent Learning Framework With Fault-Tolerant Spike Routing. IEEE Transactions on Neural Networks and Learning Systems , 1–15doi:10.1109/TNNLS.2021.3084250. Conference Name: IEEE Transactions on Neural Networks and Learning Systems

Yu, H., Lei, X., Song, Z., Liu, C., and Wang, J. (2020). Supervised Network-Based Fuzzy Learning of EEG Signals for Alzheimer's Disease Identification. IEEE Transactions on Fuzzy Systems 28, 60–71. doi:10.1109/TFUZZ.2019.2903753

Table 1. Group, sex, and age distributions of the participants engaged for this study. The values reported in the Age column represent the mean ± the standard deviation.

Group	Females	Males	Age
NC	7	1	67 ± 3.5
MCI-non-AD	0	5	73.4 ± 7.1
mild AD	5	3	68.8 ± 4.9

Table 2. Hyperparameter values inspected using grid-search cross-validation. "Best" column indicates the values of each hyperparameter that resulted in the best performance. We have not reported the best values for the LR because the SVM algorithm yielded superior results.

Hyperparameter	Range	Best	
		Mild AD vs NC	MCI-non-AD vs NC
% Features to keep	[10, 25, 50, 75, 100]	100	25
C (SVM)	[10^{-4}, 10^{-3}, ..., 10^{4}]	10^2	10^1
γ (SVM)	[scale, auto]	auto	scale
C (LR)	[10^{-4}, 10^{-3}, ..., 10^{4}]	-	-
class weight (LR)	[none, balanced]	-	-

Table 3. Confusion matrices for the two binary classification problems. They were estimated from the predictions yielded by the SVM, which was the best performing classifier for both problems. The values outside and inside the parenthesis represent the results for the epoch-level and the participant-level classification, respectively. The highlighted cells denote the true positives and true negatives.

	Mild AD vs NC		MCI-non-AD vs NC	
	PAT	HC	PAT	HC
PAT	0.88 (8)	0.12 (0)	0.88 (4)	0.12 (1)
HC	0.17 (1)	0.83 (7)	0.01 (0)	0.99 (8)

Table 4. Classification metrics for the mild AD vs NC and MCI-non-AD vs NC problems. The right-most column indicates the average cross-validation F1-score ± the standard error of the mean.

Problem	Cohort	Precision	Recall	F1-score
Mild AD vs NC	Mild AD	0.83	0.88	0.86 ± 0.06
	NC	0.88	0.83	
MCI-non-AD vs NC	MCI-non-AD	0.98	0.88	0.96 ± 0.03
	NC	0.92	0.99	

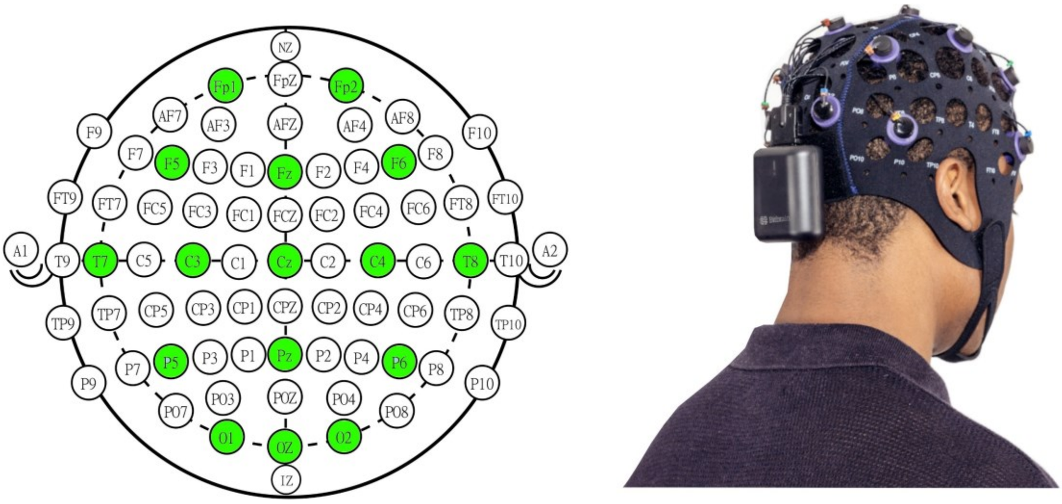


Figure 1. (Left) Sixteen-channel montage used for the present study (in green) in the extended 10-20 International System. We selected this montage to evenly cover the scalp area. (Right) Versatile semi-dry EEG acquisition system utilized for the data capture. The system includes a Bluetooth acquisition module placed below the occipital area and an EEG headset with semi-dry electrodes. Please, note that the right panel image does not represent the actual montage used in this study and it was included to illustrate the acquisition system and headset.

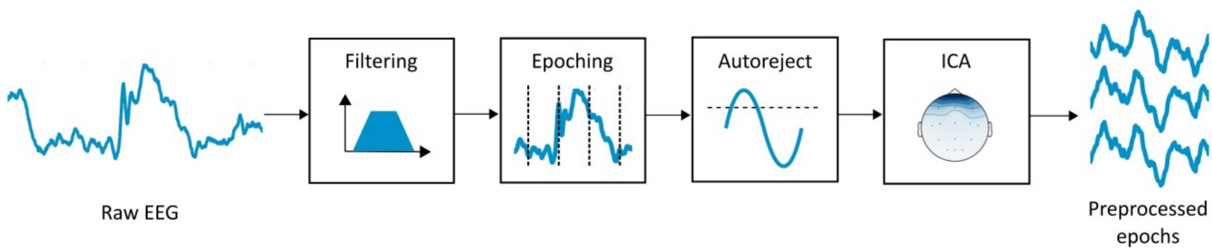


Figure 2. EEG signal processing pipeline. First, we applied a FIR filter with bandpass 1-45 Hz to remove the power line interference and retain the spectral content in the desired frequency range. Then, we segmented the filtered EEG into four-second epochs without overlapping. Finally, we performed automated artifact removal through Autoreject algorithm and ICA.

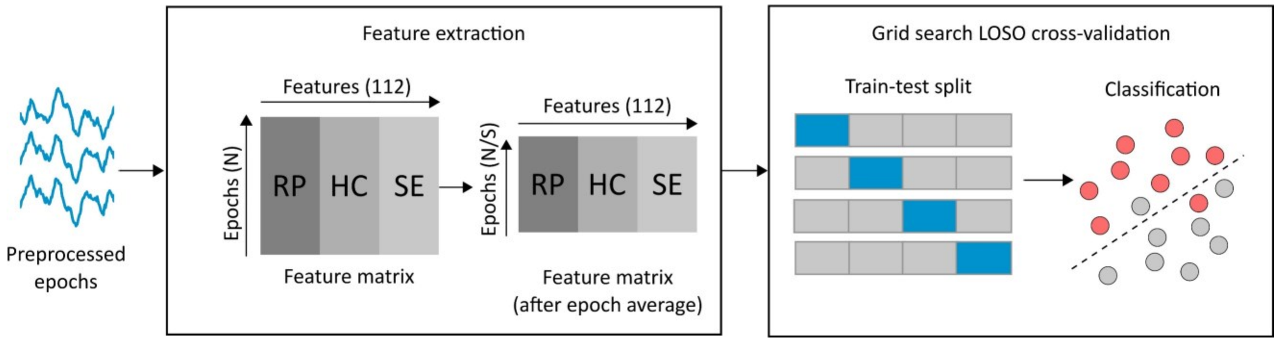


Figure 3. EEG-based feature extraction and classification pipeline. First, we performed feature extraction on the four-second clean epochs from all the participants. This yielded a feature matrix with 112 columns (features) and as many columns as the total number of epochs (N). Next, we averaged every S consecutive epochs to enhance the SNR of the features (we evaluated values of 6, 8, 10, and 12 for S). Then, we implemented a grid search cross-validation procedure to find the best hyperparameters for two possible classifiers (SVM and LR). We selected a LOSO strategy for cross-validation. Finally, we estimated the classification performance metrics across all the participants.

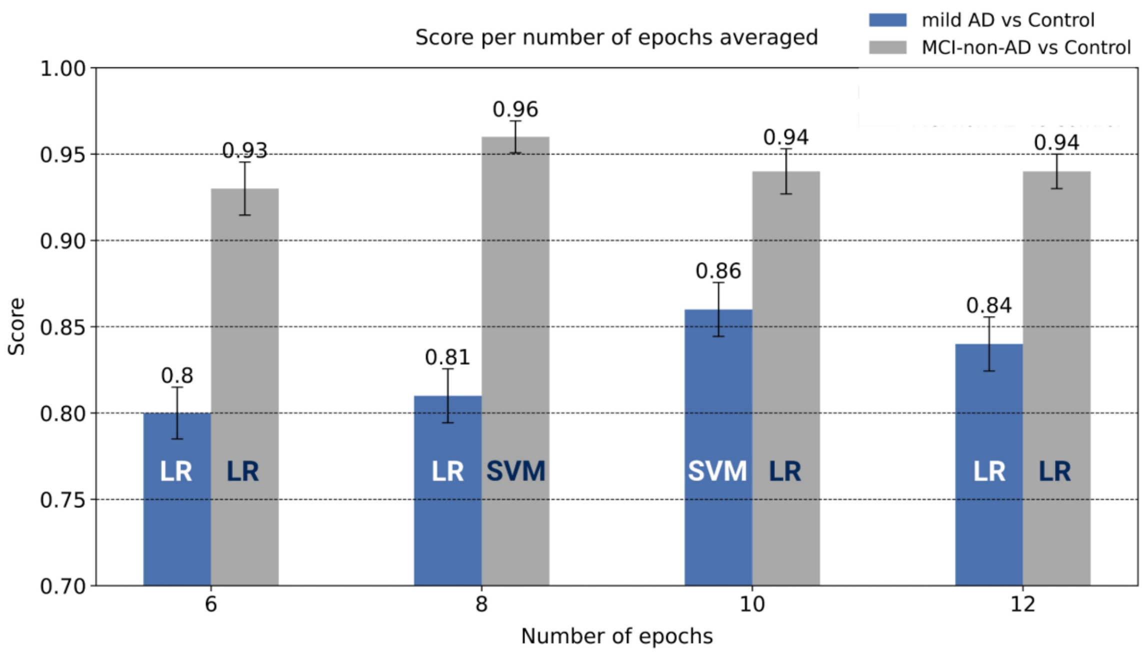


Figure 4. Comparison of the results yielded for the two binary classification problems as a function of the number of epochs averaged during processing. Error bars indicate the standard error of the mean. For each case, we have also included the name of the best performing classifier. For the mild AD vs Control problem and the MCI-non-AD vs Control problem, SVM with 10 and 8 epochs used for the average, yielded the best performance, respectively.

Appendix F

A self-driven approach for multi-class discrimination in Alzheimer's disease based on wearable EEG

Computer Methods and Programs in Biomedicine

A Self-driven Approach For Multi-class Discrimination In Alzheimer's Disease Based On Wearable EEG.

Eduardo Perez-Valero[1, 3], Miguel Ángel Lopez-Gordo[2, 3], Christian Morillas Gutiérrez[1, 3], Ismael Carrera-Muñoz[4] and Rosa M. Vílchez-Carrillo[4]

1. Department of Computer Architecture and Technology, University of Granada, Spain.
2. Department of Signal Theory, Telematics and Communications, University of Granada, Spain.
3. Brain-Computer Interfaces Laboratory, Research Centre for Information and Communications Technologies, University of Granada, Spain.
4. Cognitive Neurology Group, Neurology Unit, Hospital Universitario Virgen de las Nieves, Granada, Spain.

Correspondence: edu@ugr.es

Abstract—**Early detection is critical to control Alzheimer's disease (AD) progression and postpone cognitive decline. Traditional medical procedures such as magnetic resonance imaging are costly, involve long waiting lists, and require complex analysis. Alternatively, for the past years, researchers have successfully evaluated AD detection approaches based on machine learning and electroencephalography (EEG). Nonetheless, these approaches frequently rely upon manual processing or involve non-portable EEG hardware. These aspects are suboptimal regarding automated diagnosis, since they require additional personnel and hinder portability. In this work, we report the preliminary evaluation of a self-driven AD multi-class discrimination approach based on a commercial EEG acquisition system using sixteen channels. For this purpose, we recorded the EEG of three groups of participants: mild AD, mild cognitive impairment (MCI) non-AD, and controls, and we implemented a self-driven analysis pipeline to discriminate the three groups. First, we applied automated artifact rejection algorithms to the EEG recordings. Then, we extracted power, entropy, and complexity features from the preprocessed epochs. Finally, we evaluated a multi-class classification problem using a multi-layer perceptron through leave-one-subject-out cross-validation. The preliminary results that we obtained are comparable to the best in literature (0.88 F1-score), what suggests that AD can potentially be detected through a self-driven approach based on commercial EEG and machine learning. We believe this work and further research could contribute to opening the door for the detection of AD in a single consultation session, therefore reducing the costs associated to AD screening and potentially advancing medical treatment.**

Keywords—**Alzheimer's disease, automated detection, EEG, machine learning.**

1. INTRODUCTION

The term dementia refers to a group of neurological pathologies dominated by a progressive loss of cognitive functions [1]. Among these pathologies, Alzheimer's disease (AD) represents between 60% and 70% of the cases, and affects more than fifty million people worldwide [2]. AD patients experience a decline in cognitive areas such as reasoning, memory, and orientation. Although this disease was firstly observed in 1906, its etiology remains unclear, and a definitive diagnosis can only be established upon brain autopsy. Nonetheless, researchers have identified two main hallmarks that start to form before the impairment is notable: amyloid plaques and neurofibrillary tangles [3], [4]. Amyloid plaques are deposits of proteins that lose their standard structure and aggregate around the neurons. Alternatively, neurofibrillary tangles are thickened fibrils that encircle the nucleus of the neurons. In this context, mild cognitive impairment (MCI) is considered as a transitional stage between normal aging and AD. MCI patients experience minor memory losses, but they do not interfere with their daily-life activities. However, previous studies have found MCI patients progress to AD faster than healthy individuals of the same age [5]. Consequently, early detection is crucial to control the progression of the disease and defer the intellectual decline [6].

The purpose of AD detection techniques is to expose the cognitive and physical effects produced by the disease. Traditionally, this is accomplished through neuropsychological tests and medical procedures. Neuropsychological tests are evaluations designed to assess the intellectual areas affected in the AD course, like memory, orientation, and language. However, previous works have reported their lack of sensitivity and high variability [7], [8]. Alternatively, medical procedures are designed to unfold the physical damages produced in specific brain areas. Among them, cerebrospinal fluid (CSF) analysis is the most reliable, since it allows the study of biomarkers linked to amyloid plaques and neurofibrillary tangles [9]. However, this fluid is obtained via lumbar puncture, a costly and invasive medical procedure not feasible for all patients. Conversely, medical imaging techniques enable the representation of static and functional images of the brain. These images are rendered using different procedures such as magnetic resonance imaging (MRI) [10], positron emission tomography (PET) [11], and single photon emission computed tomography (SPECT) [12]. Even though these techniques are accurate, they lack temporal resolution, involve long waiting lists, and their analysis is traditionally based on visual inspection. Therefore, these procedures represent an unsuitable alternative towards early detection.

Alternatively, for the past decade researchers have explored the use of electroencephalography (EEG) for AD detection [13]–[16]. EEG is a neurophysiology technique to acquire brain electrical signals through electrodes placed on the scalp. This technique is affordable, portable, and non-invasive. Hence, it represents a promising alternative for the detection of neurological diseases. In this regard, researchers have recently approached AD detection through the combination of EEG processing and machine learning algorithms. As the authors of [17] outline in their comprehensive review, these studies typically target two main goals: detection of AD cohorts and analysis of AD progression. Detection studies attempt to classify well defined clinical cohorts such as AD and MCI. Conversely, progression studies attempt to identify the participants who convert to AD after a longitudinal evaluation. In AD detection studies, such as the one presented in this paper, researchers typically extract features to characterize the three main AD effects on EEG activity: slowing, complexity reduction, and loss of synchronization [17]–[20]. These effects are quantified using multiple features, including: spectral and wavelet analysis (slowing) [21], [22], entropy and information theory (complexity) [23]–[26], and coherence (synchronization) [15], [27]. For this purpose, the usual methodological stages considered in state-of-the-art works include pre-processing, time-frequency and non-linear analysis, feature extraction, and classification. For instance, in [28], the authors investigated a feature selection system in the context of two binary classification tasks: MCI vs controls, and mild AD vs controls. Their results were promising as they obtained classification rates of 0.95 and 1, respectively, using synchrony and relative power features. Similarly, the authors of [21] applied cluster analysis using power and coherence measures to a similar classification task, yielding a classification accuracy of 0.91. More recently, in [29], the authors estimated statistical coefficients from a time-frequency map to perform two classification tasks: AD vs controls, and AD vs MCI, yielding accuracies of 0.96 and 0.87, respectively. In [30], the authors evaluated the classification of AD patients, frontotemporal dementia patients (FTD), and controls, from resting state EEG activity. The authors obtained 0.86 accuracy for the FTD vs controls task using a random forest classifier. Alternatively, they obtained 0.79 accuracy for the AD vs controls task using a decision tree classifier.

Although the study of AD progression is not the goal of the present work, we believe it is worth to note that multiple works have focused on this problem for the past years. Typically, the authors of these works perform an EEG acquisition follow-up from participants at risk of suffering AD. In [31], for instance, they performed an initial and a three-month follow-up evaluation of a group of MCI patients using high-density EEG. Then, the authors successfully applied a convolutional neural network (CNN) and power spectral density (PSD) features to classify the recordings into AD and MCI. They also analyzed which channels and frequencies resulted more active in the AD progression. Although deep learning algorithms, such as the CNN, have been successfully applied to AD cohort discrimination and progression analysis, these algorithms have not yet been applied on large EEG datasets [17].

The works referred before attempted to discriminate AD cohorts through binary classification. Conversely, a few authors have approached AD detection as a three-class classification problem. This kind of approach is essential for the simplification of the detection protocol. The first attempt was presented in [32], where the authors extracted complexity and spectral features to discriminate MCI, AD, and controls using a pairwise classification method. In [25], the authors trained a multilayer perceptron (MLP) to discriminate the three cohorts from spectral and non-linear features, yielding an accuracy of 0.63. In [33], the authors tackled a three-class classification task via deep learning. They proposed the application of a CNN to classify MCI patients, AD patients, and controls from 2D grayscale images obtained from the PSD of the EEG. Alternatively, although the most studied three-class classification task corresponds to MCI vs AD vs controls, the authors of [34] successfully analyzed the mild AD vs moderate AD vs controls task. More recently, in [35], the authors proposed a new Wavelet based analysis referred as lacsogram. They used cepstrum and lacsogram distance features to classify MCI patients, AD patients, and controls, and they obtained a classification accuracy of 0.96. Likewise, in [36], the authors successfully evaluated a similar classification task under different resting conditions.

In general, the works discussed in the previous paragraph approached three-class classification of AD cohorts with success. Nonetheless, they present certain features that hinder the self-driven detection of the disease. For instance, the need for visual inspection of the EEG signals, hardware-based artifact processing, or the use of clinical EEG acquisition systems that require trained personnel to be operated. In contrast, self-driven AD classification could contribute to reducing the costs associated to screening and help to advance medical treatment. In this context, the use of commercial EEG acquisition devices and the implementation of self-driven EEG processing pipelines may have the potential to speed up the analysis, foster reproducibility, and would not require additional personnel to be operated. Considering this, in this paper we report the preliminary results of a fully self-driven approach for AD three-class discrimination based on a commercial EEG acquisition device and self-driven processing. To evaluate our approach, we conducted a study alongside the Cognitive and Behavioral Neurology Unit (CBNU) at Hospital Universitario Virgen de las Nieves de Granada (Spain). We recorded the resting state brain activity of a group of MCI patients, AD patients, and controls, using a sixteen-channel wearable acquisition device. Then, we performed artifact rejection through automated independent component analysis (ICA) and Autoreject, a self-driven artifact rejection algorithm. We evaluated a multi-layer perceptron (MLP) for the three-class discrimination of the three study groups via leave-one-subject-out (LOSO) cross-validation. The preliminary results that we obtained are comparable to the best in literature. Although the sample size that we considered is reduced, our results suggest that AD and MCI could potentially be detected using a self-driven multiclass approach based on commercial wearable EEG. This is promising from the standpoint of screening protocols, since it could potentially enable the inclusion of AD detection techniques that are ubiquitous, affordable, and accurate. Nonetheless, further research must validate the conclusions yielded in this work.

2. MATERIALS AND METHODS

In this section, we review the main aspects of the study methodology. First, we describe the cohorts involved in this research. Then, we disclose the experimental procedure, and we report the processing steps that we applied. Finally, we describe the feature extraction procedure and the implementation of the three-class classification.

2.1. Participants

The head of the CBNU recruited twenty-six participants for this study. Table 1 shows the participants group, sex, and age distributions. The participants labeled as mild AD and MCI-non-AD were CBNU patients at Hospital Universitario Virgen de las Nieves. These patients were diagnosed from PET- Amyloid or CSF analysis.

CSF analysis was carried out by two different laboratories during the recruitment period, and the reference cutoff value for the patients was that established by the corresponding laboratory. This cutoff value was based on a model of AD patients versus non-AD patients with no age stratification. CSF was obtained via lumbar puncture using a 20-gauge needle and syringe. The samples were collected in polypropylene collection tubes and sent immediately to the laboratory, where an ELISA Innotest assay was used to determine levels of Aβ42, total-τ, and τ-phosphorylated fraction. The results of the analysis were codified as normal or pathological.

PET-Amyloid was analyzed using 18F-florbetaben (FBB) by trained nuclear medicine specialists who had completed the learning curve for accredited PET-FBB scan interpretation and were blinded to the clinical situation of the patient. They reported the scans as positive (loss of grey-white matter contrast; regional cortical tracer uptake in any cortical target region: lateral temporal, frontal, posterior cingulate precuneus, or parietal), or negative (good grey-white matter contrast; no tracer uptake in target regions) for amyloid plaque presence. Cases with doubtful imaging results were discussed by both specialists to achieve a consensus.

These medical analysis were performed under the highest clinical guarantees, and we considered them as the gold standard for the patients diagnosis. Consequently, we labeled the patients as mild AD (CSF pathological results or PET positive amyloid plaque presence) and MCI-non-AD (CSF normal results or PET negative amyloid plaque presence). The rest of the participants were healthy age matched controls who did not report any neurological condition and were not CBNU patients. To conduct the study, we followed a protocol approved by the ethics committee at Hospital Universitario Virgen de las Nieves de Granada. Furthermore, the participants signed an informed consent before the start of the experiment, and clinical personnel supervised them throughout the entire session.

Table 1

Group, sex, and age distribution of the participants involved in the study. The age column includes the mean age ± the standard deviation.

Group	Females	Males	Age
MCI-non-AD	0	6	72.8 ± 6.5
mild AD	7	4	68.3 ± 4.6
control	8	1	66.7 ± 3.4

2.2. Setup and experimental procedure

The objective behind the engagement of the participants was the acquisition of their resting state EEG activity for our study. Nonetheless, taking advantage of the fact that the participants attended the consulting room for the acquisition, they performed a cognitive test consisting of two cognitive tasks. The details regarding such test are not reported in this work since they are out of the scope of this study, and they did not interfere with the EEG acquisition in any manner.

Before the start of the experiment, we requested the participants to carefully read and sign the informed consent. Then, we acquired three minutes of their eye-open resting state EEG activity in three instants: prior to the first cognitive task, after this task, and after the second cognitive task. To prevent edge effects, we only studied the central two-minute segment of each recording. To perform our analysis, we concatenated the three recordings associated to each participant, hence, we considered a total EEG time of six minutes per participant.

To record the brain electrical activity, we used the Versatile wireless wearable system by Bitbrain. This commercial device includes a Bluetooth acquisition module working at 256 Hz, and an EEG headset with semi-dry electrodes. For the electrode setup, we considered sixteen sensors located at positions Fp1, Fp2, F5, Fz, F6, T7, T8, C3, Cz, C4, P5, Pz, P6, O1, Oz, and O2 of the extended 10-20 International System, and we referenced them to the left ear lobe. We selected this sensor montage to evenly cover the scalp following setups from analogous studies [21], [37], [38]. Fig. 1 represents the electrode setup selected in this study along with the Versatile system.

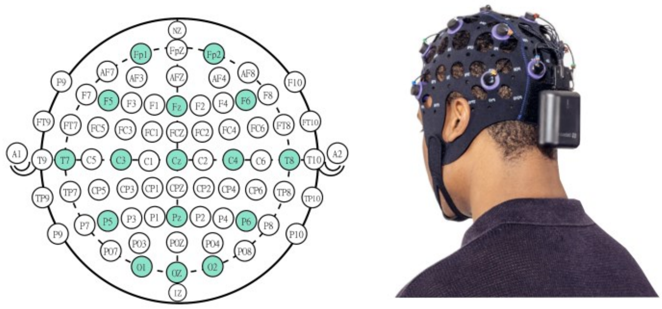


Fig. 1: (Left) Electrode setup selected for this study. We considered sixteen electrodes from the extended 10-20 International System to uniformly cover the scalp. (Right) Versatile semi-dry acquisition system used for the EEG data capture. The system operates at a fixed sampling rate of 256 Hz and provides support for sixteen electrodes.

2.3. Signal processing

First, we filtered the raw EEG signals with a 1690 order bandpass FIR filter with 1-45 Hz passband and zero phase-shift. We selected a FIR filter rather than an IIR filter because our analysis did not entail high throughput restrictions, and we gave priority to filter control and stability [38]. Then, we split the filtered signals into four-second epochs without overlapping.

Subsequently, we implemented automated artifact rejection in two steps: Autoreject and ICA. First, we applied the Autoreject algorithm to automatically find an artifact threshold per channel and identify and reject bad data spans. Autoreject is an automated data-driven artifact rejection algorithm based on Bayesian optimization and cross-validation. The algorithm is available as a Python module and its application is straightforward. In terms of performance, Autoreject has been evaluated against several datasets and achieved equal or better performance compared to typical approaches. We refer the interested reader to [40], where the developers of the algorithm provide a comprehensive review of its implementation and evaluation. Subsequently, we applied ICA to correct blink artifacts. ICA enables the decomposition of a signal conformed by multiple sources into those sources and a mixing matrix. This is particularly useful to process EEG signals, as they represent the combination of internal brain sources. Based on this, ICA is typically used to detect and remove artifactual components such as blinks. These components display high variance and a spatial distribution towards the frontotemporal area of the scalp. Therefore, to identify the blink artifact component, we used Fp1 as an electro-oculogram proxy, and automatically identified the component that showed the highest correlation with this sensor. Then, we removed the blink component from the source matrix, and we reconstructed the EEG signals. Fig. 2 shows the signal processing stage described in this subsection.

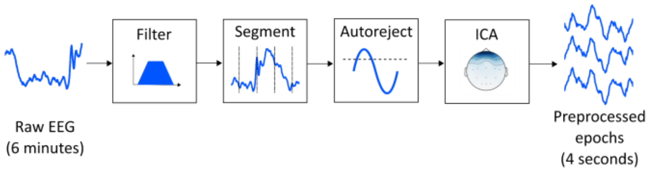


Fig. 2: Self-driven signal processing pipeline implemented in this study. Initially, we applied a 1-45 Hz bandpass FIR filter to remove the power line interference and retain the desired spectral content. Then, we segmented the filtered signals into four-second epochs without overlapping. Finally, we implemented automated artifact removal in two stages: Autoreject algorithm and ICA. The input and output signals represent the raw EEG and the preprocessed epochs of a single participant.

2.4. Feature extraction

After applying the EEG processing pipeline, we extracted three features per EEG channel from each preprocessed epoch: relative power (RP) in the five main EEG bands, spectral entropy (SE), and Hjorth complexity (HC). We examined these features because they have been already validated in analogous studies [21], [32], [41]–[43]. RP represents the fraction of the total signal power that is contained in a frequency range. Alternatively, SE represents the Shannon entropy of the power spectrum of the signal. SE describes the uniformity of the power spectrum distribution, and, hence, the irregularity of the EEG. Therefore, SE is minimal for a pure sine wave and maximal for white noise [44]. Finally, HC is one of the three Hjorth parameters (activity, mobility, and complexity), and is derived as the ratio of the mobility of the first derivative of the signal to the mobility of the signal itself [45]. The mathematical representation of the three features described in this paragraph are presented in Equations 1-3.

$$RP = \frac{\sum_{f_i}^{f_o} P}{\sum_{\forall f} P} \quad (1)$$

$$SE = -\sum_{f} S(f) * \log_2 S(f) \quad (2)$$

$$HC = \frac{\sigma_s''/\sigma_s'}{\sigma_s'/\sigma_s} \quad (3)$$

In Equation 1, f_i and f_o represent the lower and upper frequency bounds of a particular frequency band. For this study, we estimated RP in the five main EEG bands: delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (> 30 Hz). In Equation 2, f represents the frequencies in the frequency range of the signal, and S represents its normalized power spectrum. Finally, in Equation 3, σ_s, σ_s', and σ_s'' represent the variance of the signal, the variance of its first derivative, and the variance of its second derivative.

After the feature extraction from the preprocessed epochs, we concatenated the features to create the feature matrix. Then, following the approach presented in [46], we averaged the features of every S adjacent rows in the feature matrix to increase the signal-to-noise ratio and reduce the size of the dataset. We examined four values for S: 6, 8, 10, and 12, and we kept the one that yielded the best performance.

2.5. Classification

Following feature extraction, we implemented a three-step classification pipeline including feature scaling, feature selection, and classification. First, the feature scaler normalizes the features between zero and one. This technique is extensively applied in machine learning to enhance the performance of non-tree-based models. Then, the feature selector gathers the most relevant features according to a predefined strategy. We suggest the interested reader to review [47] for a comprehensive review of frequent feature selection strategies. In this study, we applied a method based on the chi-squared test to identify the features most related to the target. We selected this strategy owing to its intrinsic speed. Lastly, for the last stage of the pipeline, we selected a MLP as this algorithm inherently allows multilabel classification.

To find the best combination of hyperparameters for the elements in the pipeline we applied grid-search. For the feature matrix averaging procedure described the last paragraph of Section 2.4, we evaluated different values for the number of rows to average. For the feature selection stage, we evaluated different values for the percentage of features selected. For the MLP, we evaluated different values for the architecture of the network, the L2 penalty (Alpha), and the activation function. Table 2 shows the different value ranges we considered during grid search for all the mentioned hyperparameters. We refer the interested reader to the scikit-learn documentation for a comprehensive description of the hyperparameters reported in this table. The best combination of hyperparameters is reported in the Results section.

To evaluate the classification performance during the grid search, we applied cross-validation using a built-in function from scikit-learn Python module (GridSearchCV). This function evaluates the classification pipeline using different sets of hyperparameters and a cross-validation scheme. Since this study involves the discrimination of clinical groups, we followed a LOSO strategy. Thereby, data is split into as many folds as participants. Then, for each fold, the training set holds data from all the participants except one, whose data is reserved for the test set. Thus, information from a participant is never in the training set and the test set at the same time. This prevents from positive bias and represents a robust alternative to evaluate the general performance of the model. Fig. 3 displays the stages described in this section.

Table 2
List of hyperparameters evaluated during grid search cross-validation[a].

Stage	Hyperparameter	Value range
Epoch averaging	Epochs averaged	6, 8, 10, 12
Feature selection	Feature percentage	10, 25, 50, 75, 100
Classification	Layer sizes	3-3-3, 4-4-4, 5-5-5
	Alpha	10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}
	Activation	relu, tanh

[a]Note that for the layer sizes parameter, each figure between the parenthesis refers to a hidden layer, and its value represents the number of neurons in the layer. For instance, 4-4-4 represents a MLP with three hidden layers and four neurons in each layer.

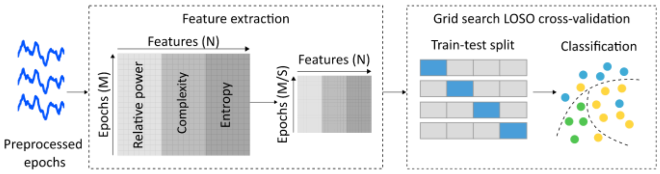


Fig. 3: Self-driven feature extraction and classification pipeline. First, we performed feature extraction on the preprocessed epochs. This procedure yielded a feature matrix with M rows (epochs) and N features. Then, we averaged every S consecutive epochs to improve the signal-to-noise ratio (we assessed S values of 6, 8, 10, and 12). Subsequently, we performed grid search cross-validation to obtain the best set of hyperparameters for the classification pipeline.

3. RESULTS

In this section, we report the results we obtained for the three-class classification problem examined in this study. Owing to lack of cooperation (S09 and S18), presence of additional diseases (S12), and poor signal quality (S21 and S26), we excluded the data from five participants. Hence, we considered data from twenty-one participants in our analysis.

Fig. 4 represents the confusion matrix at the epoch level. To create this matrix, we gathered the true epoch labels and classifier predictions from the test sets used during cross-validation, and we estimated the distribution of the predictions across the three classes.

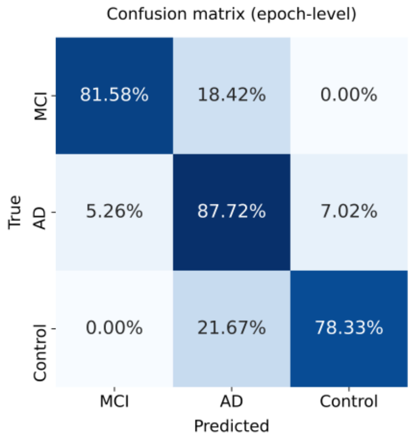


Fig. 4: Epoch-level confusion matrix for the discrimination of MCI-non-AD, mild AD, and controls.

Fig. 5 shows the confusion matrix at the subject level. In this case, we evaluated the ability of the classifier to discriminate the participants via majority vote. This is, if the model correctly classified most of the epochs for a given participant, we considered it correctly discriminated the participant. Otherwise, we considered it misclassified the participant.

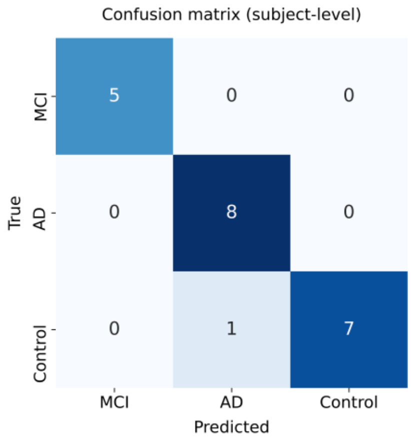


Fig. 5: Subject-level confusion matrix for the discrimination of MCI-non-AD, mild AD, and controls.

Table 3 presents a class report with the scores obtained by the model during cross-validation in terms of precision and recall. To create this table, we followed the procedure described for the epoch-level confusion matrix. We also included the average cross-validation F1-score.

Table 3
Classification report for the mild AD vs MCI-non-AD vs control discrimination problem[a].

Cohort	Precision	Recall	F1-score
mild AD	0.91	0.82	0.88 ± 0.05
MCI-non-AD	0.71	0.88	
control	0.92	0.78	

[a]We estimated the metrics displayed in this table after gathering all the test true labels and classifier predictions from the cross-validation. The right-most column represents the average cross-validation F1-score ± the standard error of the mean (SEM).

Lastly, the best set of hyperparameters found via grid search is reported in Table 4.

Table 4
Best set of hyperparameters found via grid search cross-validation.

Hyperparameter	Best value
Epochs averaged	12
Feature percentage	100
MLP Layer sizes	5-5-5
MLP Alpha	10^{-6}
MLP Activation	relu

4. DISCUSSION

The purpose of this study was to evaluate the potential of a fully self-driven approach for AD three-class discrimination based on a commercial EEG acquisition device and automated processing. To this end, we recorded the resting state activity of a group of participants from these cohorts, and we developed a classification pipeline built on signal processing and machine learning. We used a commercial wearable EEG acquisition device that does not require trained experts to be operated. Thus, no further personnel aside from a neurologist is required to perform the data acquisition. The preliminary results that we obtained suggest that self-driven multiclass discrimination of AD cohorts from commercial EEG can be successfully performed. This may contribute to opening the door for the future implementation of affordable and portable early AD screening techniques. Nonetheless, further research is required to validate the preliminary results reported in this work.

In terms of epoch-level performance, the confusion matrix in Fig. 4 suggests that remarkable classification performance can be achieved following the approach presented in this work. Specifically, classification accuracy yielded by our classifier was 81.58%, 87.72%, and 78.33% for the MCI-non-AD, mild AD, and control epochs analyzed. Notably, the classifier yielded the best results for the mild AD and MCI-non-AD groups. This points at its ability to identify the patient cohorts, what is essential in the discrimination of clinical groups. This is further noted in the participant-level confusion matrix displayed in Fig. 5. According to this figure, our approach only misclassified a control participant. Naturally, the same conclusions are drawn from the cross-validation metrics reported in Table 3. As shown in this table, the classifier yielded an average F1-score of 0.88 ± 0.05.

According to the preliminary results obtained in this work, we believe a self-driven multi-class classification approach could present some advantages compared to the approaches reported in literature. For clarity, Table 5 in the appendix summarizes the main aspects of related approaches. First, it is important to acknowledge that a direct comparison between these works is not entirely feasible due to the heterogeneity of the cohorts reported. Whilst most studies considered AD, MCI, and controls [16], [32], [36], [48], other studies evaluated different clinical groups. For instance, [34] examined mild AD, AD, and controls, and [35] evaluated MCI, mild AD (ADM), advanced AD (ADA), and controls. In this sense, we believe the cohorts that we considered (MCI-non-AD, mild AD, and controls) are the most suitable for early detection research. Another fundamental difference among these studies is the diagnosis procedure. Most previous works diagnosed the participants based on neuropsychological tests: [34], [35] applied the mini mental state examination (MMSE), [16], [36] used the Montreal cognitive assessment (MoCA), [36], [48] applied the recommendations of the Diagnostic and Statistical Manual of Mental Disorders (DSMMD), [25] followed the guidelines of the National Institute of Aging and Alzheimer's Association (NIA-AA), and [32] applied a battery of cognitive tests and other evaluations. Conversely, we identified the participants based on the diagnosis yielded from a CSF analysis and a PET scan, medical procedures with higher reliability compared to neuropsychological examinations. In terms of electrode setup, most of these works selected 19-21 sensors [16], [25], [34]–[36], [48], with only [32] using a considerably higher number of sensors (32). In this sense, we believe the use of a low number of electrodes promotes participant comfort and simplifies the montage. Regarding the acquisition system, we used the device with the highest degree of portability compared to the rest of approaches gathered in Table 5, as most of them utilized medical-oriented systems with reduced mobility.

Regarding artifact processing, [24] and [48] performed epoch manual selection based on visual inspection. This is unsuitable from the early detection standing point because it involves additional personnel and delays the analysis. Alternatively, [34] and [16] used the built-in capabilities of their acquisition hardware, the Nihon Kohden EEG system, to perform artifact rejection, what hinders the flexibility of their proposal. Instead, [36] used EEGLAB toolbox for MATLAB to perform artifact rejection. However, the authors do not specify if they used a manual or an automated procedure. Finally, [43] and [35] did not specify the artifact removal techniques applied. For this preliminary study, we applied Autoreject and automated ICA. We believe this is a more appropriate alternative for artifact rejection because the analysis remains self-driven, what is essential for early detection, and also fosters the reproducibility of the results yielded.

With respect to performance, a direct comparison is hindered by the heterogeneity of the cohorts and the metrics used to evaluate the results. The F1-score reported in these works ranges from 0.81 to 0.89 [32], [34], [36], [48]. Other authors reported accuracy values between 0.63 and 0.94 [16], [25], [48]. Alternatively, [35] reported excellent results in terms of the area under the curve (AUC). As per the results reported in Table 3, a self-driven approach based on a commercial EEG device, such as the one introduced in this work, can achieve a performance that is comparable to the best results in literature.

5. CONCLUSIONS

In this study, we evaluated the potential of a fully self-driven approach for AD three-class discrimination using a commercial EEG acquisition system. For this purpose, we conducted a study involving participants from three cohorts: MCI-non-AD, mild AD, and healthy age-matched controls. First, we recorded their eye-open resting state brain activity, and then we implemented a self-driven pipeline including artifact rejection, feature extraction, and classification. For artifact rejection, we applied Autoreject, a data-driven algorithm, and ICA. Then, we extracted the relative power, Hjorth complexity, and spectral entropy from the clean epochs. Finally, we performed grid search cross-validation under a LOSO strategy.

In terms of performance, the preliminary results that we obtained are comparable to those reported in analogous studies, or even better in some cases. These results are promising since self-driven detection could potentially outperform existing approaches, promote early detection, contribute to reducing costs, and encourage the reproducibility of results. This may help reduce detection times associated with other traditional medical procedures like clinical EEG and nuclear medicine procedures. In this regard, early detection is paramount to control neural loss and defer cognitive decline. This is especially important in the case of middle-aged adults, since the impact of the disease in their social life and workplace is higher compared to older adults.

Alternatively, whilst the median number of participants in the studies discussed in Table 5 is forty-seven, we only considered twenty-one. Therefore, in future studies, we must evaluate the approach described in this paper on a larger sample to reinforce the conclusions drawn in this work. In addition, we are aware the participants involved in this and similar studies, do not strictly represent the typical patients who attend to neurology services. These patients, usually suffer from other pathologies apart from dementia. Hence, to effectively transfer EEG-based self-driven approaches into the clinical ecosystem, we must evaluate their performance on real-life participants. This way, in the future, patients may benefit from accurate, fast, and affordable detection approaches that help advance their medical treatments.

ACKNOWLEDGEMENTS

This research was supported by project PGC2018-098813-B-C31 funded by the Spanish Ministry of Science, Innovation and Universities, by European Regional Development Funds, and by project B-TIC-352-UGR20, co-funded by the Operative Program FEDER 2014-2020 and the Economy, Universities and Science Office of the Andalusian Regional Government. The authors would like to especially thank the Cognitive and Behavioral Neurology Unit at Hospital Universitario Virgen de las Nieves de Granada and the participants involved in the study for their collaboration in this research.

Appendix of the article

Table 5

Comparison of studies on multi-class discrimination of AD cohorts. The columns indicate, from left to right, the authors, the cohorts involved, the diagnosis procedure to label the cohorts, the EEG acquisition device, the number of sensors utilized during the acquisition, the artifact rejection technique applied, and the performance metrics yielded for the multi-class discrimination of the cohorts.

Study	Cohorts	Diagnosis	EEG device	Sensors	Artifact rejection	Performance
McBride et al. **[32]**(2014)	17 early AD 16 MCI 15 HC	Cognitive tests and other evaluations	Neuroscan II	32	Not reported	F1-score = 0.83 (subject-level)
Ruiz-Gómez et al. **[25]**(2018)	37 AD 37 MCI 37 HC	NIA-AA	XLTEK Natus medical	19	Visual inspection	Accuracy = 0.63 (subject -level)
Tzimourtra et al. **[34]**(2019)	8 mild AD 6 AD 10 HC	MMSE	Nihon Kohden EEG 2100	19	Hardware	F1-score = 0.85 (epoch-level)
Ieracitano et al. **[48]**(2020)	63 AD, 63 MCI 63 HC	DSMMD	Not disclosed	19	Visual inspection	F1-score = 0.81 (epoch-level)
Oltu et al. **[16]**(2021)	8 AD 16 MCI 11 HC	MoCA	Nihon Kohden EEG 1200	19	Hardware	Accuracy = 0.94 (epoch-level)
Rodrigues et al. **[35]**(2021)	11 HC 8 MCI 11 ADM 8 ADA	MMSE	Not disclosed	19	Not specified	AUC = 0.95 (epoch-level)
Sharma et al. **[36]**(2021)	16 AD 16 MCI 15 HC	DSMMD	SOMNOscreen EEG 32	21	EEGLAB	F1-score = 0.85 (epoch-level)
Our approach	8 mild AD 5 MCI-non-AD 8 HC	CSF/PET	Versatile EEG 16	16	Autoreject and ICA	F1-score = 0.88 (epoch-level) Accuracy = 0.95 (subject -level)

REFERENCES

[1] M. J. Moore, C. W. Zhu, and E. C. Clipp, “Informal Costs of Dementia CareEstimates From the National Longitudinal Caregiver Study,” *J Gerontol B Psychol Sci Soc Sci*, vol. 56, no. 4, pp. S219–S228, Jul. 2001, doi: 10.1093/geronb/56.4.S219.

[2] “World Alzheimer Report 2018 - The state of the art of dementia research: New frontiers,” *NEW FRONTIERS*, p. 48.

[3] D. P. Perl, “Neuropathology of Alzheimer’s Disease,” *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, vol. 77, no. 1, pp. 32–42, 2010, doi: 10.1002/msj.20157.

[4] A. Serrano-Pozo, M. P. Frosch, E. Masliah, and B. T. Hyman, “Neuropathological Alterations in Alzheimer Disease,” *Cold Spring Harb Perspect Med*, vol. 1, no. 1, p. a006189, Jan. 2011, doi: 10.1101/cshperspect.a006189.

[5] R. A. Sperling *et al.*, “Toward defining the preclinical stages of Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease,” *Alzheimer’s & Dementia*, vol. 7, no. 3, pp. 280–292, May 2011, doi: 10.1016/j.jalz.2011.03.003.

[6] M. Riemenschneider, N. Lautenschlager, S. Wagenpfeil, J. Diehl, A. Drzezga, and A. Kurz, “Cerebrospinal fluid tau and beta-amyloid 42 proteins identify Alzheimer disease in subjects with mild cognitive impairment,” *Arch. Neurol.*, vol. 59, no. 11, pp. 1729–1734, Nov. 2002, doi: 10.1001/archneur.59.11.1729.

[7] G. Kuslansky *et al.*, “Detecting dementia with the Hopkins Verbal Learning Test and the Mini-Mental State Examination,” *Archives of Clinical Neuropsychology*, p. 16, 2004.

[8] M. S. Mendiondo, J. W. Ashford, R. J. Kryscio, and F. A. Schmitt, “Modelling Mini Mental State Examination changes in Alzheimer’s disease,” *Statistics in Medicine*, vol. 19, no. 11–12, pp. 1607–1616, 2000, doi: 10.1002/(SICI)1097-0258(20000615/30)19:11/12<1607::AID-SIM449>3.0.CO;2-O.

[9] R. J. Perrin, A. M. Fagan, and D. M. Holtzman, “Multimodal techniques for diagnosis and prognosis of Alzheimer’s disease,” *Nature*, vol. 461, no. 7266, pp. 916–922, Oct. 2009, doi: 10.1038/nature08538.

[10] S. H. Hojjati, A. Ebrahimzadeh, A. Khazaee, and A. Babajani-Feremi, “Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM,” *Journal of Neuroscience Methods*, vol. 282, pp. 69–80, Apr. 2017, doi: 10.1016/j.jneumeth.2017.03.006.

[11] S. R. Meikle, F. J. Beekman, and S. E. Rose, “Complementary molecular imaging technologies: High resolution SPECT, PET and MRI,” *Drug Discovery Today: Technologies*, vol. 3, no. 2, pp. 187–194, Jun. 2006, doi: 10.1016/j.ddtec.2006.05.001.

[12] J. M. Górriz, F. Segovia, J. Ramírez, A. Lassl, and D. Salas-Gonzalez, “GMM based SPECT image classification for the diagnosis of Alzheimer’s disease,” *Applied Soft Computing*, vol. 11, no. 2, pp. 2313–2325, Mar. 2011, doi: 10.1016/j.asoc.2010.08.012.

[13] F. J. Fraga, G. Q. Mamani, E. Johns, G. Tavares, T. H. Falk, and N. A. Phillips, “Early diagnosis of mild cognitive impairment and Alzheimer’s with event-related potentials and event-related desynchronization in N-back working memory tasks,” *Computer Methods and Programs in Biomedicine*, vol. 164, pp. 1–13, Oct. 2018, doi: 10.1016/j.cmpb.2018.06.011.

[14] P. A. M. Kanda, E. F. Oliveira, and F. J. Fraga, “EEG epochs with less alpha rhythm improve discrimination of mild Alzheimer’s,” *Computer Methods and Programs in Biomedicine*, vol. 138, pp. 13–22, Jan. 2017, doi: 10.1016/j.cmpb.2016.09.023.

[15] C. T. Briels, D. N. Schoonhoven, C. J. Stam, H. de Waal, P. Scheltens, and A. A. Gouw, “Reproducibility of EEG functional connectivity in Alzheimer’s disease,” *Alzheimer’s Research & Therapy*, vol. 12, no. 1, p. 68, Jun. 2020, doi: 10.1186/s13195-020-00632-3.

[16] B. Oltu, M. F. Akşahin, and S. Kibaroğlu, “A novel electroencephalography based approach for Alzheimer’s disease and mild cognitive impairment detection,” *Biomedical Signal Processing and Control*, vol. 63, p. 102223, Jan. 2021, doi: 10.1016/j.bspc.2020.102223.

[17] K. D. Tzimourta *et al.*, “Machine Learning Algorithms and Statistical Approaches for Alzheimer’s Disease Analysis Based on Resting-State EEG Recordings: A Systematic Review,” *Int J Neural Syst*, vol. 31, no. 5, p. 2130002, May 2021, doi: 10.1142/S0129065721300023.

[18] C. Babiloni *et al.*, “Brain neural synchronization and functional coupling in Alzheimer’s disease as revealed by resting state EEG rhythms,” *International Journal of Psychophysiology*, vol. 103, pp. 88–102, May 2016, doi: 10.1016/j.ijpsycho.2015.02.008.

[19] A. H. H. Al-Nuaimi, E. Jammeh, L. Sun, and E. Ifeachor, “Complexity Measures for Quantifying Changes in Electroencephalogram in Alzheimer’s Disease,” *Complexity*, vol. 2018, pp. 1–12, 2018, doi: 10.1155/2018/8915079.

[20] N. Benz *et al.*, "Slowing of EEG Background Activity in Parkinson's and Alzheimer's Disease with Early Cognitive Dysfunction," *Front. Aging Neurosci.*, vol. 6, 2014, doi: 10.3389/fnagi.2014.00314.

[21] R. Wang, J. Wang, H. Yu, X. Wei, C. Yang, and B. Deng, "Power spectral density and coherence analysis of Alzheimer's EEG," *Cogn Neurodyn*, vol. 9, no. 3, pp. 291–304, Jun. 2015, doi: 10.1007/s11571-014-9325-x.

[22] V. Bairagi, "EEG signal analysis for early diagnosis of Alzheimer disease using spectral and wavelet based features," *Int. j. inf. tecnol.*, vol. 10, no. 3, pp. 403–412, Sep. 2018, doi: 10.1007/s41870-018-0165-5.

[23] C. Jiang, Y. Li, Y. Tang, and C. Guan, "Enhancing EEG-Based Classification of Depression Patients Using Spatial Information," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 566–575, 2021, doi: 10.1109/TNSRE.2021.3059429.

[24] C. Coronel *et al.*, "Quantitative EEG Markers of Entropy and Auto Mutual Information in Relation to MMSE Scores of Probable Alzheimer's Disease Patients," *Entropy*, vol. 19, no. 3, Art. no. 3, Mar. 2017, doi: 10.3390/e19030130.

[25] S. J. Ruiz-Gómez *et al.*, "Automated Multiclass Classification of Spontaneous EEG Activity in Alzheimer's Disease and Mild Cognitive Impairment," *Entropy*, vol. 20, no. 1, p. 35, Jan. 2018, doi: 10.3390/e20010035.

[26] M. Şeker, Y. Özbek, G. Yener, and M. S. Özerdem, "Complexity of EEG Dynamics for Early Diagnosis of Alzheimer's Disease Using Permutation Entropy Neuromarker," *Comput Methods Programs Biomed*, vol. 206, p. 106116, Jul. 2021, doi: 10.1016/j.cmpb.2021.106116.

[27] A. H. Meghdadi *et al.*, "Resting state EEG biomarkers of cognitive decline associated with Alzheimer's disease and mild cognitive impairment," *PLOS ONE*, vol. 16, no. 2, p. e0244180, Feb. 2021, doi: 10.1371/journal.pone.0244180.

[28] E. Gallego-Jutglà, J. Solé-Casals, F.-B. Vialatte, M. Elgendi, A. Cichocki, and J. Dauwels, "A hybrid feature selection approach for the early diagnosis of Alzheimer's disease," *J. Neural Eng.*, vol. 12, no. 1, p. 016018, Feb. 2015, doi: 10.1088/1741-2560/12/1/016018.

[29] C. Ieracitano, N. Mammone, A. Bramanti, S. Marino, A. Hussain, and F. C. Morabito, "A Time-Frequency based Machine Learning System for Brain States Classification via EEG Signal Processing," in *2019 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2019, pp. 1–8. doi: 10.1109/IJCNN.2019.8852240.

[30] A. Miltiadous *et al.*, "Alzheimer's Disease and Frontotemporal Dementia: A Robust Classification Method of EEG Signals and a Comparison of Validation Methods," *Diagnostics*, vol. 11, no. 8, Art. no. 8, Aug. 2021, doi: 10.3390/diagnostics11081437.

[31] F. C. Morabito, C. Ieracitano, and N. Mammone, "An explainable Artificial Intelligence approach to study MCI to AD conversion via HD-EEG processing," *Clin EEG Neurosci*, p. 15500594211063662, Dec. 2021, doi: 10.1177/15500594211063662.

[32] J. C. McBride *et al.*, "Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease," *Computer Methods and Programs in Biomedicine*, vol. 114, no. 2, pp. 153–163, Apr. 2014, doi: 10.1016/j.cmpb.2014.01.019.

[33] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, and F. C. Morabito, "A Convolutional Neural Network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings," *Neurocomputing*, vol. 323, pp. 96–107, Jan. 2019, doi: 10.1016/j.neucom.2018.09.071.

[34] K. D. Tzimourta *et al.*, "EEG Window Length Evaluation for the Detection of Alzheimer's Disease over Different Brain Regions," *Brain Sciences*, vol. 9, no. 4, Art. no. 4, Apr. 2019, doi: 10.3390/brainsci9040081.

[35] P. M. Rodrigues, B. C. Bispo, C. Garrett, D. Alves, J. P. Teixeira, and D. Freitas, "Lacsogram: A New EEG Tool to Diagnose Alzheimer's Disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3384–3395, Sep. 2021, doi: 10.1109/JBHI.2021.3069789.

[36] N. Sharma, M. H. Kolekar, and K. Jha, "EEG based dementia diagnosis using multi-class support vector machine with motor speed cognitive test," *Biomedical Signal Processing and Control*, vol. 63, p. 102102, Jan. 2021, doi: 10.1016/j.bspc.2020.102102.

[37] N. N. Kulkarni and V. K. Bairagi, "Extracting Salient Features for EEG-based Diagnosis of Alzheimer's Disease Using Support Vector Machine Classifier," *IETE Journal of Research*, vol. 63, no. 1, pp. 11–22, Jan. 2017, doi: 10.1080/03772063.2016.1241164.

[38] G. Fiscon *et al.*, "Combining EEG signal processing with supervised methods for Alzheimer's patients classification," *BMC Med Inform Decis Mak*, vol. 18, no. 1, p. 35, Dec. 2018, doi: 10.1186/s12911-018-0613-y.

[39] E. C. Ifeachor and B. W. Jervis, *Digital Signal Processing: A Practical Approach*. Pearson Education, 2002.

[40] M. Jas, D. A. Engemann, Y. Bekhti, F. Raimondo, and A. Gramfort, "Autoreject: Automated artifact rejection for MEG and EEG data," *NeuroImage*, vol. 159, pp. 417–429, Oct. 2017, doi: 10.1016/j.neuroimage.2017.06.030.

[41] L. R. Trambaiolli, N. Spolaôr, A. C. Lorena, R. Anghinah, and J. R. Sato, "Feature selection before EEG classification supports the diagnosis of Alzheimer's disease," *Clinical Neurophysiology*, vol. 128, no. 10, pp. 2058–2067, Oct. 2017, doi: 10.1016/j.clinph.2017.06.251.

[42] N. Houmani *et al.*, "Diagnosis of Alzheimer's disease with Electroencephalography in a differential framework," *PLoS One*, vol. 13, no. 3, Mar. 2018, doi: 10.1371/journal.pone.0193607.

[43] S.-S. Poil, W. de Haan, W. M. van der Flier, H. D. Mansvelder, P. Scheltens, and K. Linkenkaer-Hansen, "Integrative EEG biomarkers predict progression to Alzheimer's disease at the MCI stage," *Front Aging Neurosci*, vol. 5, Oct. 2013, doi: 10.3389/fnagi.2013.00058.

[44] T. Inouye *et al.*, "Quantification of EEG irregularity by use of the entropy of the power spectrum," *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 3, pp. 204–210, Sep. 1991, doi: 10.1016/0013-4694(91)90138-T.

[45] N. Päivinen, S. Lammi, A. Pitkänen, J. Nissinen, M. Penttonen, and T. Grönfors, "Epileptic seizure detection: A nonlinear viewpoint," *Computer Methods and Programs in Biomedicine*, vol. 79, no. 2, pp. 151–159, Aug. 2005, doi: 10.1016/j.cmpb.2005.04.006.

[46] F. J. Fraga *et al.*, "Towards an EEG-based biomarker for Alzheimer's disease: Improving amplitude modulation analysis features," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 1207–1211. doi: 10.1109/ICASSP.2013.6637842.

[47] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.

[48] C. Ieracitano, N. Mammone, A. Hussain, and F. C. Morabito, "A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia," *Neural Networks*, vol. 123, pp. 176–190, Mar. 2020, doi: 10.1016/j.neunet.2019.12.006.