UNIVERSIDADE D
COIMBRA

Ana Rita Paulete Correia

Assessing Emotional States from Biosignals in Autistic and Neurotypically Developed Individuals

Dissertação no âmbito do Mestrado Integrado em Engenharia Biomédica com especialização em Informática Clínica e Bioinformática orientada pelo Professor Doutor Marco António Machado Simões e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia.

Outubro de 2021

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

Assessing emotional states from biosignals in autistic and neurotypically developed individuals

Dissertation to obtain a master's degree in Biomedical Engineering with specialization in Clinical Informatics and Bioinformatics, supervised by Professor Marco António Machado Simões, and presented to the Physics Department of the Faculty of Sciences and Technology of the University of Coimbra.

Ana Rita Paulete Correia

Outubro de 2021

Supervised by: Marco Simões, Ph.D.

This work was developed in collaboration with:

CIBIT – Coimbra Institute for Biomedical Imaging and Translational Research

ICNAS – Institute for Nuclear Sciences Applied to Health

CISUC - Center for Informatics and Systems of the University of Coimbra

Agradecimentos

Esta dissertação é o culminar dos cinco anos mais desafiantes e enriquecedores da minha existência, que devem todo o seu brilho a cada uma das pessoas com quem os partilhei. Aproveito o momento para expressar a enorme gratidão que sinto, por, algures na aleatoriedade do universo, os nossos caminhos se terem cruzado.

Em primeiro lugar, a toda a equipa do projeto BioHab, o meu sincero agradecimento por toda a partilha de conhecimento e, principalmente, por me terem mostrado a beleza e a arte subjacentes na Investigação. Depois de repararmos nelas, é impossível não ver. Deixo um agradecimento especial à Investigadora Catarina Duarte por toda a disponibilidade e por toda a experiência que partilhou comigo ao longo das várias horas de aquisição de dados. Tenho-a como um exemplo a seguir.

Aos amigos de uma vida, por serem sempre a melhor escapatória. Obrigada à Cláudia, à Raquel e ao Rafael, por trazerem sempre ao de cima o meu lado mais autêntico. À Maria Miguel e à Ana Beatriz, por serem sempre Casa e me apoiarem incondicionalmente.

Às que partilharam comigo estes anos de academia, Carolina, Leonie, Lúcia e Joana, obrigada pelo suporte e pelas trocas de desabafos, mas, sobretudo, por me lembrarem sempre que não nos devemos levar demasiado a sério, e por encontrarem o *comic relief* em todas as situações. Um agradecimento especial à Beatriz, por ter partilhado comigo os momentos mais marcantes desta jornada e por nunca me deixar baixar os braços.

Ao Professor Marco, por todas as oportunidades que me proporcionou, pela orientação sempre tão próxima e disponível, mas, sobretudo, por ter sido uma fonte de inspiração desde o primeiro dia, quero deixar o meu mais sincero e profundo agradecimento. Tenho toda esta experiência como um privilégio, que se deve maioritariamente à Pessoa brilhante que a possibilitou.

Aos grandes responsáveis pelas minhas vitórias: Maria Helena e António Manuel. Não há palavras que possam expressar a gratidão que sinto pela forma como me ensinaram a ver e estar no mundo. Tudo o que sou e alcancei, devo-o a vós. À minha irmã, Inês, por ser a minha maior admiradora e por acreditar em mim, quando eu própria duvido, e a toda a restante família, por me lembrar sempre que o segredo está na simplicidade das coisas, deixo o meu eterno agradecimento.

A todos, por confiarem no processo, obrigada.

Abstract

The increase in Autism Spectrum Disorder's (ASD) prevalence over the last decades has accentuated the need to develop new and improved rehabilitation techniques that can reach a larger part of this population. This quest has resulted in the development of computerized alternatives, frequently in the form of gamified tasks that, being low cost and portable, serve as a complement to traditional therapy, optimizing and consolidating its effects, or, in some cases, being the only type of therapy that the patient can have access to. These new rehabilitation alternatives, however, still bear some limitations. ASD is often associated with hypersensitivity, which means that the presentation of the wrong stimuli can have a negative effect on the subject and lead to his disengagement from the computerized therapy, compromising its efficacy. That is why the next step must be to make the games capable to adapt to their users and their emotional responses. Although there are different ways through which one expresses emotion, none of them is as reliable as the autonomic nervous system's (ANS) response, which cannot be simulated.

Motivated by these issues, this dissertation focused upon the physiological measures of the ANS and their relationship with the individual's emotional state, with a special focus on the ASD case. However, considering the association of this disorder with emotional dysregulation, the identification of the subject's true emotional state is a real concern and should not be dependent upon other people's interpretation, or even his own. That is why, as part of a bigger goal, our set of physiological signals was acquired simultaneously with functional Magnetic Resonance Imaging (fMRI), while the participants completed a task designed to elicit different emotions classified in the two dimensions of valence and arousal, with the prospect of combining biosignals and fMRI to create models capable of assessing the true emotional state of their users based on the physiological response, having as ground truth the activity of targeted brain regions.

Focusing on the information contained in the physiological signals, to assess whether the simultaneous acquisition of fMRI and biosignals is feasible, we first investigated their quality and the interference posed by the Magnetic Resonance (MR) environment, by comparing them with recordings obtained outside the MR scanner, thus free of its influence, for the same experimental task. Then, the signals were pre-processed and segmented according to the events of interest, which were labelled in two different ways: according to the subject's perception of his emotional response and using the pre-established emotional connotation of the stimuli, both based on the two-dimensional space of valence and arousal. A total of 70 features were then extracted from the physiological signals, for each segment. The resulting dataset was used to train and test four different machine learning algorithms: Minimum Distance Classifier, K-Nearest Neighbours and Support Vector Machines with linear and Radial Basis Function kernels.

The simultaneous acquisition of physiological signals and fMRI was validated, with the comparison between the recordings acquired inside and outside of the scanner not suggesting any

major quality differences that can be attributed to the MR. This opens the door to a wide range of applications involving the understanding of the relationships between brain response and its translation in the peripheral nervous system. Some issues related to the data acquisition setup, however, did compromise the quality of some of the biosignals, resulting in information loss and consequently also compromising the discriminative power of the extracted features. That said, overall, the classification outcomes were suboptimal, underlining the challenging nature of automatic emotion assessment. These results call for further exploration of the measured signals and the features extracted from them, to refine and improve their discriminative power. Moreover, there is the possibility that the stimuli employed were not sufficiently strong to offset significant emotional and physiological reactions in the participants. Hence, it would be important to try different stimuli with varying emotional classifications and intensities.

Keywords: Autism Spectrum Disorder, Physiological Signals, Biosignals, Automatic Emotion Assessment

Resumo

O aumento da prevalência da Perturbação do Espectro do Autismo (PEA) nas últimas décadas acentuou a necessidade de desenvolver novas e melhoradas técnicas de reabilitação que sejam acessíveis à maior parte desta população. Essa procura resultou no desenvolvimento de alternativas computorizadas, frequentemente na forma de tarefas gamificadas que, sendo portáteis e de baixo custo, servem como complemento à terapia tradicional, otimizando e consolidando os seus efeitos, sendo, em alguns casos, o único tipo de terapia à qual o paciente tem acesso. No entanto, estas novas alternativas de reabilitação, apresentam ainda algumas limitações. A PEA está muitas vezes associada à hipersensibilidade, o que significa que a apresentação do tipo de estímulos errado pode afetar negativamente o sujeito e levar à rejeição desta terapia, comprometendo a sua eficácia. É por isso que o próximo passo deve passar por tornar os dispositivos que albergam estes jogos educativos sensíveis aos seus utilizadores e às suas respostas emocionais. Embora existam diferentes canais através dos quais expressamos as nossas emoções, nenhum deles será tão fiável quanto a resposta do sistema nervoso autónomo (SNA), não passível de ser simulada.

Motivada por estas questões, esta dissertação foca nas medidas fisiológicas do SNA e na sua relação com o estado emocional do indivíduo, com especial atenção ao caso da PEA. No entanto, considerando a associação desta perturbação com desregulação emocional, a identificação do verdadeiro estado emocional do sujeito é um fator importante e não deve depender da interpretação de outras pessoas, ou mesmo do próprio sujeito. É por isso que, estando este trabalho integrado num projeto mais abrangente, o conjunto de sinais fisiológicos utilizado aqui foi adquirido em simultâneo com ressonância magnética funcional (fMRI), enquanto os participantes completavam uma tarefa desenhada para evocar várias emoções, classificadas no espaço bidimensional de valência e ativação, com a perspetiva de combinar biossinais e fMRI para criar modelos capazes de avaliar o verdadeiro estado emocional dos seus utilizadores com base na resposta fisiológica, tendo como verdade fundamental a atividade cerebral.

Ao focarmo-nos apenas nos sinais fisiológicos, para avaliar se a aquisição simultânea de fMRI e biossinais é viável, a sua qualidade e a influência que o ambiente da Ressonância Magnética (RM) teve sobre eles foi investigada e comparada com as medições obtidas, com a mesma tarefa experimental, fora da RM e, portanto, livre da sua influência. Em seguida, os sinais foram pré-processados e segmentados de acordo com os eventos de interesse, os quais foram rotulados de duas formas diferentes: de acordo com a perceção do sujeito da sua resposta emocional e usando a classificação emocional pré-estabelecida de cada estímulo, ambos baseados no espaço bidimensional de valência e ativação. De seguida, um total de 70 features foram extraídas dos biossinais, para cada segmento. Os dados resultantes foram usados para treinar e testar quatro algoritmos de Machine Leaning: Minimum Distance Classifier, K-Nearest Neighbours e Support Vector Machines com os kernels de Função de Base Radial e Linear.

A aquisição simultânea de sinais fisiológicos e fMRI foi validada, com a comparação entre os sinais adquiridos dentro e fora do scanner não revelando diferenças de qualidade que possam ser atribuídas à RM. Isto abre portas para uma vasta gama de aplicações que envolvem a descodificação das interações entre a resposta cerebral e a sua tradução no sistema nervoso periférico. No entanto, alguns problemas relacionados com as configurações na aquisição de dados comprometeram a qualidade de alguns dos biossinais, resultando na perda de informação e, consequentemente, comprometendo também o poder discriminativo das features extraídas. Posto isto, em geral, os resultados da classificação foram insuficientes, pondo em evidência a natureza desafiante da classificação automática de estados emocionais. Estes resultados exigem uma exploração mais profunda dos sinais e das features extraídas, de modo a refinar e melhorar o seu poder discriminativo. Para além disso, devemos considerar a possibilidade de que os estímulos usados não tenham sido suficientemente fortes para provocar reações emocionais e fisiológicas significativas nos participantes. Por isso, será importante experimentar novos estímulos com diferentes classificações e intensidades emocionais.

List of Contents

List of Figures

List of Tables

List of Abbreviations

ADOS – Autism Diagnostic Observation Schedule

ANS – Autonomic Nervous System

AQ – Autism Spectrum Quotient

ASD – Autism Spectrum Disorder

BART – Balloon Analogue Risk Task

BCI – Brain-Computer Interface

BRV – Breath Rate Variability

C – Cost

CAAV – Chieti Affective Action Videos

CNS – Central Nervous System

ECG – Electrocardiogram

EDA – Electrodermal Activity

EEG – Electroencephalogram

EMG – Electromyography

EQ – Empathy Quotient

ERP – Event-Related Potential

FASTR - FMRI Artifact Slice Template Removal

fMRI – Functional Magnetic Resonance Imaging

FSIQ – Full-Scale Intelligence Quotient

HA – High Arousal

HF – High Frequency

HR – Heart Rate

HRV – Heart Rate Variability

HV – High Valence

HVHA – High Valence and High Arousal

IBI – Interbeat Intervals

IR – Infrared

KNN – K-Nearest Neighbours

LA – Low Arousal

LDA – Linear Discriminant Analysis

LF – Low Frequency

LOSO – Leave One Subject Out

LV – Low Valence

LVHA – Low Valence and High Arousal

MDC – Minimum Distance Classifier

MHD – Magnetohydrodynamic

MR – Magnetic Resonance

MRI – Magnetic Resonance Imaging

NV – Neutral Valence

NVLA – Neutral Valence and Low Arousal

PCA – Principal Component Analysis

PMU – Physiological Measurement Unit

PNS – Parasympathetic Nervous System

PPG – Photoplethysmography

PR – Pulse Rate

PRV – Pulse Rate Variability

PSC – Percent Signal Change

RBF – Radial Basis Function

SAM – Self-Assessment Manikin

SCR – Skin Conductance Response

SNR – Signal-to-Noise Ratio

SNS – Sympathetic Nervous System

SpO2 – Peripheral Oxygen Saturation

SVM – Support Vector Machines

TD – Typically Developed

USA – United States of America

VE – Virtual Environments

VLF – Very Low Frequency

VR – Virtual Reality

1 Introduction

In the last decades, autism has earned much attention from the general public in consequence of a considerable increase in prevalence estimates. As of 2014, 16.8 per 1000 children aged 8 years were diagnosed with Autism Spectrum Disorder (ASD) in the United States of America (USA), which represents an increase of 150% when compared to 2000 estimates (Baio *et al.*, 2018). As for Portugal, in 2007 it was estimated that autism prevalence was 9.2 in every 10000 school-aged children (Oliveira *et al.*, 2007). Such growth may be due to a greater awareness regarding the condition together with an improvement in screening techniques, or it can in fact represent an increase in prevalence. Either way, the numbers must not be ignored. The amount of people that can benefit from new and improved rehabilitation techniques is enormous. That is why continuous research in the field is crucial, representing the possibility of a significant increase in the quality of life of these individuals and their families.

1.1 Motivation

Autism Spectrum Disorder is a neurodevelopmental condition that affects social and communication skills, as well as normal behavioural patterns. People diagnosed with ASD usually favour repetitive routines and interests and feel the most comfortable when alone, not pursuing or enjoying the engagement in social activities (Johnson and Myers, 2007). All these symptoms end up limiting the patients' social inclusion, which further affects their development. Accordingly, early diagnosis and subsequent participation in rehabilitation therapies are crucial to obtain good results and improve the patients' lives. However, as of now, these interventions entail high costs, given that they usually require one therapist per patient. Besides, a recent study conducted in the USA evidences the scarcity of qualified professionals to treat ASD, which potentially leads to overbooking (Zhang and Cummings, 2020). This situation can be very tiresome for the professionals and may consequently jeopardize the patients' experience and outcomes given that the therapist is not in his best form and the time reserved for each subject is small. This translates into less therapy time, which consequently lowers the training outcome. For patients living in remote locations, access to qualified professionals might be even more troublesome given that they might have to travel for several hours for receiving therapy. In addition to this, the Covid-19 pandemic came to highlight the need for telepresence solutions, that are now more relevant than ever with the urgent need to limit personal contact to stop the spread of the disease. Therefore, the development and improvement of viable alternative therapy strategies is necessary to mitigate these issues, guaranteeing that the biggest number of patients have access to therapy while making the most out of every single intervention.

In the literature, the employment of virtual reality (VR) as a therapeutic and rehabilitation tool has been recurrently mentioned as suitable for the ASD population (Blascovich *et al.*, 2002; Goodwin, 2008; Bellani *et al.*, 2011; Parsons and Cobb, 2011; Kandalaft *et al.*, 2013; Didehbani *et al.*, 2016; Simões *et al.*, 2018, 2020). A key property of VR is that it allows for the creation of unlimited different virtual environments (VE) with shared similarities with the real world, as well as to easily modify them so that incremental changes can be applied to a specific scene enabling the repeated practice of a given task in slightly different settings which may promote the generalization of the learned skills to different scenarios and real-world situations (Strickland, 1997; Parsons and Mitchell, 2002; Parsons and Cobb, 2011). By cancelling the stress, the confusing stimuli, and the possible dangers of the outside world, VR provides a sheltered environment to practice social interactions and everyday life situations such as a public transport ride, or even job interviews (Kandalaft *et al.*, 2013; Didehbani *et al.*, 2016; Simões *et al.*, 2018). In addition, this technology allows the inclusion of gaming factors to increase the motivation of its users. Moreover, individuals with autism are commonly particularly interested in computers and digital gadgets due to their structure and consistency. This enthusiasm often represents a greater engagement in this kind of therapy when compared with traditional methods (Strickland, 1997).

This leads to the concept of serious gaming. A serious game is a game with an educational purpose, going beyond the sole purpose of entertainment. These games are a promising tool and have been applied and studied as a rehabilitation alternative in the ASD population mostly targeting emotion recognition and production, and social skills (Grossard *et al.*, 2017). A key advantage of this kind of therapy is its portability. Generally, access to a computer is all that is required to run these games, and nowadays, when more and more people have access to tablets and smartphones, the dissemination of the games is even easier. This allows the patients to practice in the comfort of their homes, without the need for the presence of a therapist which facilitates access to therapy for a greater number of people.

Serious games mostly resort to positive reinforcement mechanisms to cement the concepts they intend to teach (Grossard *et al.*, 2017). However, response to reward is known to be altered in ASD patients (Dichter *et al.*, 2012; Kohls *et al.*, 2013, 2018; Watson *et al.*, 2015) which invalidates the use of most of the typical rewards in this population, especially social rewards. Thus, investigating and comprehending the reward mechanisms and motivational factors in ASD is a fundamental step in the path of perfecting therapeutical interventions, virtual or not.

Recently, researchers have been delving into the concept of biofeedback. Biofeedback consists of recording physiological information and using that information, in real time, usually represented in a clear and understandable way so that it can be easily interpreted by the user, to aid him endogenously alter a specific physiological response (McKee, 2008). In the case of serious games, however, this concept can be used to tailor them to specific clinical populations,

by creating models that, based on real-time physiological signals measurement, assess the emotional state of the user and adapt the task, the environment, or the rewards accordingly. This personalized paradigm might be especially interesting for the ASD population, since the game adapts itself to the user and not the opposite, which maximizes engagement, and consequently optimizes the learning process. However, even though automatic emotion recognition is a growing field, as well as its dissemination through various applications, it has not been sufficiently considered for the case of ASD, whose individuals commonly show signs of emotion dysregulation (Samson *et al.*, 2014). Thus, it is necessary to understand the real emotional response that the individuals of this population present to a particular stimulus, and how it reflects on different autonomic nervous system (ANS) functions, measured through the corresponding biosignals.

This leads to what is probably one of the major concerns when it comes to automatic emotion recognition: finding the ground truth. As already mentioned, autism is associated with emotional dysregulation. This means that the use of self-assessment questionnaire responses or labels based on the general population's emotional perception of a stimulus as ground truth has questionable accuracy. Hereupon, due to its spatial resolution that allows for the precise mapping of brain regions or networks of interest, functional Magnetic Resonance Imaging (fMRI) may just be the ideal true state indicator. While researchers in the field have struggled to map different emotions into specific brain regions, some findings involving the analysis of whole-brain fMRI activity patterns suggest that it is possible to predict an emotional response based on that information with considerable specificity levels, which suggests the involvement of specific brain networks in different emotional responses (Kragel and LaBar, 2016). Sessions involving this imaging technique, however, are quite expensive and nonportable, which limits their applicability and dissemination (Sarracanie *et al.*, 2015). Furthermore, the acquisition of physiological signals in the magnetic resonance (MR) environment represents a tremendous challenge. The static magnetic field and the gradient switch severely contaminate most of the records, which can, in some cases, jeopardize the appearance of changes in the signal related to the emotional state (Niazy *et al.*, 2005; Oster and Clifford, 2017). It is then important to ascertain how feasible it is to acquire multimodal biosignals inside of the MR scanner for the purposes described above. The simultaneous acquisitions of fMRI and multimodal biosignals could therefore be used to find ANS physiological patterns that are representative of the targeted brain regions modulation so that it can be inferred outside the MR scanner and applied as ground truth in automatic emotion recognition algorithms.

1.2 Objectives

The main goals of this dissertation are first to explore the feasibility and efficiency of constructing models, specifically tailored for the ASD population, capable of automatically discriminate between different emotional states, and then, to ascertain how much the MR environment affects the quality of the multimodal physiological signals and the subsequent performance of the classifiers. This last point can be a helpful indicator of whether the simultaneous acquisition of fMRI and biosignals can be useful for different research purposes, namely looking for patterns in the conjoined signals that can be, in some way, representative of brain regions modulation.

To fulfil these goals, an experimental task took place involving the simultaneous acquisition of fMRI and physiological signals in both, autistic and typically developed individuals. Additionally, so that comparison was made possible, a similar experimental task was conducted outside the MR environment while acquiring a smaller set of physiological signals.

1.3 Dissertation Structure

This document is organized into 7 chapters. Following this introduction, the 2nd chapter presents a background on the physiological signals used in this study and how they respond to different emotional states. Subsequently, chapter 3 focuses on the state of the art of the research that has been carried out around them to develop and achieve automatic emotion assessment. The Methods section in chapter 4 describes in detail the work that was carried out in pursuing the main goals of the project, and chapters 5 and 6 present the findings and their analysis and discussion. To finalize, chapter 7 highlights the main conclusions taken out of this study and bears some insights into the paths that could follow the line of work presented in this dissertation that are yet to explore.

2 Background

2.1 Biosignals

The experience of emotion is commonly accompanied by physiological variations that may be perceived by the individual through bodily sensations and manifestations, or that can be so subtle that they go unnoticed even to the one experiencing them. Recent technological advances have allowed the optimization of physiological sensors, that are now smaller and cheaper than ever, which has permitted their dissemination through a wide array of applications that go beyond direct clinical practice, including their use to monitor and find relationships between different physiological responses and emotional states, opening new doors to the knowledge of human physiology and emotion recognition. This section is a summary of the main characteristics of the biosignals explored in the present study and their application in emotion recognition.

2.1.1 Respiration

Pulmonary ventilation, breathing, or respiration, refers to the rhythmic processes through which air is inhaled into the lungs and subsequently exhaled out.

Inhalation or inspiration is known as the active phase of respiration since it depends on the contraction of the respiratory muscles. When inhaling, the diaphragm contracts to pull down the lungs while the intercostal muscles contract to expand the ribcage, these actions create space for the lungs to expand as well, increasing the intrapulmonary volume and consequently lowering the intrapulmonary pressure in comparison to the external pressure. This way, as air flows from higher pressure to lower pressure, the air naturally fills the lungs. As inspiration ceases, begins the passive phase of respiration, known as expiration. Expiration does not depend on any muscle contraction, instead, it relies on the elastic recoil of the lungs and respiratory muscles. As the diaphragm and intercostal muscles relax, the pulmonary volume decreases, increasing the intrapulmonary pressure. As follows, air flows to the exterior, reversely to what happens during inhalation. Expiration persists until the intrapulmonary pressure equals the external pressure, at which point inspiration begins and the process is restarted (De Troyer and Boriek, 2011).

One’s emotional state has a great influence on respiratory rhythm. While it is common to find ourselves breathing faster when we are experiencing excitement, more calm and relaxed states are associated with a slow and steady rhythm. On the other hand, negative and aroused states, such as anger or fear, are usually accompanied by irregular breathing and even occasional cease of respiration (Rainville *et al.*, 2006). A peculiarity of respiratory activity is that it can be voluntarily and consciously controlled, with some studies reporting that besides being influenced by emotions, the voluntary modulation of respiratory pattern seems to be capable of influencing

and altering the self-perceived emotional state, suggesting the existence of a reciprocal relationship that represents a powerful potential target for biofeedback interventions (Philippot, Chapelle and Blairy, 2002).

Measuring and analysing these breathing fluctuations can thus potentially provide us with valuable information regarding the emotional status of an individual. Respiration can be recorded through a series of different processes, including contact and non-contact-based techniques. Since this study explores the feasibility and practicality of measuring physiological signals simultaneously with fMRI acquisitions, the specifications and limitations of the MR environment render some of the possibilities impracticable. However, there are some alternatives through which the respiratory signal can still safely be acquired inside the MR scanner. For the case of this research, respiration variations were tracked with the aid of a respiratory cushion connected to a pressure hose, and a respiratory belt (Siemens Healthcare). The working principle of such a sensor relies on the rise and fall movement of the rib cage caused by the respiratory muscles. As the chest expands during inspiration the cushion is pressed against the belt which consequently increases the pressure inside the cushion and hose that in turn transmits the pressure variation to a measurement unit which turns it into a voltage. On the other hand, during expiration, the rib cage decreases in volume allowing the cushion to expand between the subject's body and the belt, and the pressure to drop. This gives rise to a biosignal that continuously captures the respiratory rhythm of the individual.

2.1.2 Photoplethysmography

Photoplethysmography (PPG) is an optical-based method used to measure the changes in blood volume that occur at the capillaries. This blood volume change is directly influenced by the quantity of blood that is pumped by the heart into the aortic artery during ventricular depolarization. Hence, the PPG signal carries valuable information about the cardiovascular function.

To measure this biosignal, a probe composed of both, a source emitting light, and a photodetector, is placed on the person's finger, earlobe, or toe. Given the optical properties of tissue and its main constituent, water, that absorb the vast majority of shorter wavelengths, the light source operates either at the red or near-infrared portion of the spectrum. The intensity of reflected light measured at the photodetector is affected by changes in the blood volume and is then converted into a voltage which gives rise to a pulsatile waveform. This waveform, also known as the 'AC' component of the PPG signal, accompanies a baseline ('DC' component) that oscillates slowly with physiological variations related to respiration, the sympathetic nervous system, or thermoregulation (Allen, 2007).

Different studies have focused on exploring the use of PPG for the extraction of the pulse rate variability (PRV), an analogous measure to the heart rate variability (HRV) that is commonly derived from the electrocardiogram (ECG). HRV is the term commonly used to refer to the variations in the interval between consecutive heartbeats and represents one of the most popular markers of autonomic activity (Malik *et al.*, 1996). To infer the HRV, it is first necessary to find the intervals between successive heartbeats. While in the ECG signal the heart rate (HR) is inferred from the identification of successive R peaks, in the PPG, the HR equivalent, denominated as pulse rate (PR), is found by identifying the time between two consecutive characteristic points of the PPG waveform. Some of these characteristic points correspond to the onset of the pulse wave, the systolic peak, or the point corresponding to 50% of the maximum amplitude. The maximum value of the first and second derivative are also commonly used for this effect and seem to consistently provide a good agreement between the HRV and PRV (Fig. 2.1). However, the characteristic point interval that results in the best correlation with the RR interval is still under investigation and no consensus has been reached yet (Elgendi, 2012; Pinheiro *et al.*, 2016).

The finding that is consistent between studies is the fact that PRV presents high correlations with HRV under most circumstances, even though there are signs of compromise in this relationship during stressful situations, physical or mental (Bolanos, Nazeran and Haltiwanger, 2006; Pinheiro *et al.*, 2016).

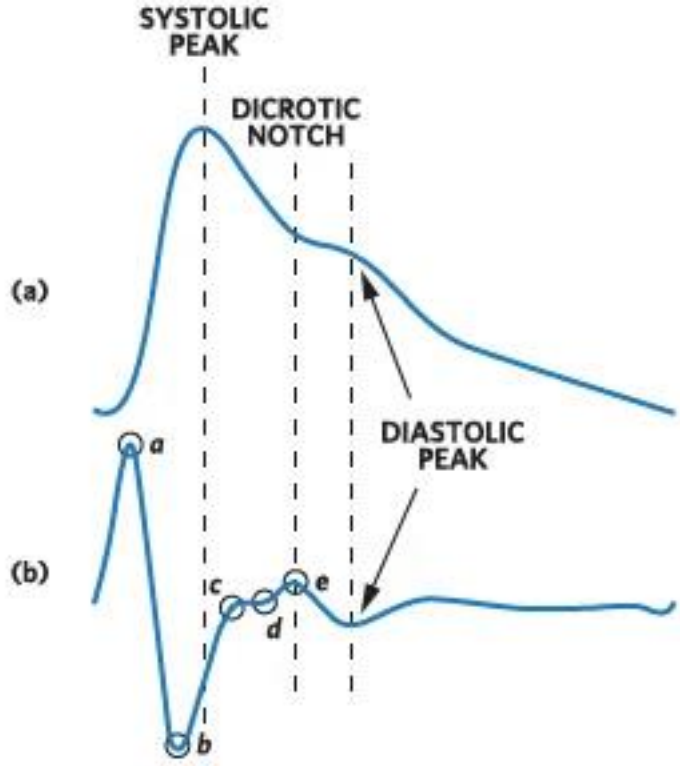

Figure 2.1 - Representation of a photoplethysmogram pulse wave, its main characteristic points, and its second derivative (a) Photoplethysmogram pulse wave (b) Second derivative of the pulse wave in (a). – adapted from Acceleration plethysmogram. (2020). https://www.maximintegrated.com/en/design/blog/what-is-in-photoplethysmography-data-a-look-at-the-interaction-between-sensor-performance-and-algorithms.html

This evidence allied to the convenience of the measurement technique of the PPG, indicates it to be a suitable replacement of ECG for ambulatory monitoring. ECG is a rather sensitive signal to a variety of artifact sources and its acquisition in the MR environment is extremely challenging, since the electrically charged particles in the blood flowing through the magnetic field generate an additional electrical field in the subject's body. This is known as the magnetohydrodynamic (MHD) effect and it produces changes in the ECG, harmless to the individual, but that substantially alter the morphology of the wave, with the most prominent effect being an elevation of the T wave, that frequently overshoots the QRS complex, making the detection of the R peaks a very difficult and possibly inaccurate task (Oster and Clifford, 2017). Hence, PPG might just be the perfect replacement to monitor cardiac activity during magnetic resonance imaging (MRI) and fMRI acquisitions, potentially achieving more accurate estimates of HRV than ECG under the static magnetic field.

2.1.3 Pulse Oximetry

Pulse oximetry is a technique used to measure the peripheral oxygen saturation (SpO2) at a given time. It takes advantage of the optical properties of oxygenated and deoxygenated haemoglobin. While the first absorbs infrared (IR) light, deoxyhaemoglobin absorbs greater amounts of red light instead. The sensor is composed of two light-emitting diodes, emitting red and IR light, which are placed on one side of the probe, and at the opposite side stands a photodiode with the function of detecting the quantity of light that is transmitted through the tissues. By determining the relative amount of light from each wavelength that was absorbed in between the two probe sides, the device can estimate the fraction of haemoglobin that is bound to oxygen in near-real-time (Chan, Chan and Chan, 2013). The sensor's design is small, minimally invasive, and it can be placed on a finger or earlobe with relative ease, making it suitable to use outside the clinical environment, even for populations with greater sensitivity.

The extent to which SpO2 is influenced by an individual's emotional state is still underexplored. However, as mentioned in a previous section, it is well understood that breathing patterns are affected by stressful situations. Now, if the way our body acquires oxygen is through breathing, if a sudden scare will cause us to skip a breath, it is only natural that the quantity of oxygenated blood will momentarily drop. Thus, SpO2 can potentially represent an interesting physiological indicator of emotion, and its application in that context must be further explored.

2.1.4 Electrodermal Activity

Electrodermal activity (EDA) refers to the electrical properties of the skin which are constantly suffering fluctuations that are directly related to sweat secretion and, therefore, sudomotor neurons activity. While most bodily functions are regulated by both divisions of the ANS, sweat gland activity is exclusively modulated by the sympathetic nervous system (SNS), the SNS works in response to physical or psychological stress to make the necessary adaptations for the body to deal with the situation at hand, preparing it for action. The EDA signal is then a direct measure of sympathetic activity, which is why it is so popular and relevant for emotion recognition systems.

According to Ohm's law, which states that a conductive body's electrical resistance (R) equals the voltage (V) to current (I) ratio (R = V/I), the changes in skin conductance (G), i.e. the reciprocal of skin resistance (G = I/V = 1/R), can be measured by applying a small constant voltage between two skin spots and obtaining the resultant current flow (Dawson, Schell and Filion, 2007).

This biosignal is composed of a tonic response component and a phasic response component. The tonic response oscillates slowly and is attributed to the baseline level of EDA, i.e., the background activity of sweat glands. This baseline is variable from individual to individual, being

an indicator of general arousal and alertness, and it can be influenced by a variety of external factors such as temperature or humidity. The phasic component concerns the fast variations in the signal that are known as skin conductance responses (SCRs). An SCR is the result of a sudomotor burst and findings suggest that its' amplitude is linearly dependant on the frequency of the action potentials and the number of activated sweat glands (Benedek and Kaernbach, 2010). SCRs can be observed in reaction to a stimulus that can either be external or self-induced, such as the recollection of a memory. A particularity of this physiological response, representing a limitation for some applications, is its' latency, SCRs can arise from 1 up to 5 seconds after stimulus presentation and last for a few seconds. This factor must be taken into consideration in experimental design, not considering an appropriate inter-stimulus interval can result in SCRs resultant from previous events overlapping with the ones resulting from the current one which can lead to uncertainties.

2.1.5 Electroencephalography

The electroencephalogram (EEG) is the record of the electrical activity of the brain and thus represents a window to the central nervous system. The EEG signal is commonly acquired at the scalp with the aid of surface electrodes, and it is the result of the spatially averaged electrical activity of multiple neurons located at the cerebral cortex, with the main contribution coming from pyramidal cells (Nunez and Srinivasan, 2006). The cerebral cortex is the outer layer of the cerebrum, composed of billions of neurons, and is organized in a folded arrangement. The cerebral fissure crosses the cortex dividing the brain into the left and right hemispheres, which are joined by the corpus callosum underneath the cerebral cortex.

The correlation of the synaptic activity captured in surface EEG with cognition and behaviour is widely recognized, with the biosignal changing its patterns and characteristics as different tasks are performed or as a response to new stimuli. These activities can be classified into two kinds: spontaneous rhythmic activity or event-related potentials (ERPs).

The rhythmic oscillations of brain activity are classified into five principal frequency bands, named delta (0.5 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 13 Hz), beta (13 – 30 Hz), and gamma (> 30 Hz).

Delta

Delta waves are low frequency, high amplitude waves and are most often found at deep sleep stages. Even though their presence is common during infancy, finding these waves in awake adults is atypical and usually a sign of disease (Rangayyan, 2015).

Theta

Theta oscillations are commonly encountered in young children, and, like delta waves, their absolute power in wakeful states seems to be age-dependent. In adults, they are present in the primary stages of sleep and are also believed to show up during the encoding process of new memories and information (Klimesch, 1999). Additionally, evidence suggests that this activity is strongly connected to cognition, being particularly found at frontally located electrodes during mental tasks. A correlation between theta and alpha activity has been consistently pointed out (Nunez and Srinivasan, 2006).

Alpha

The alpha rhythm was the first identified type of rhythmic activity of the brain, and it is dominant at the occipital area in wakeful, resting states, particularly with the eyes closed which earned it the designation of 'idle rhythm'. In fact, at an eye-opening cue or during complex mental tasks, alpha activity is significantly toned down, which leads to the theory that it is connected to the active blockage of sensory information. Although for the beginning of its acknowledgment it was thought to be a resting state oscillation, further exploration of EEG in different contexts and during behavioural tasks, led to the recognition that alpha is present in the course of multiple functions, originating from different sources. Alpha has been shown to be involved in memory processes (Klimesch, 1997), motor activity (Pfurtscheller *et al.*, 1997), as well as visual and auditory sensory processing (Schürmann, Başar-Eroglu and Başar, 1997). What is now known about alpha waves, indicates that they play a much larger role in brain signalling than was once thought.

Beta

Like the alpha rhythm, also beta waves were addressed in Hans Berger's first recognition of the presence of rhythmic activity in the brain back in 1929. They are distinguishable from the first for their higher frequency and smaller amplitude.

Beta waves are well known for their participation in sensorimotor processes. Power in this frequency range increases during posture maintenance in contrast to movement, the imagination of movement, or response to touch. Moreover, beta attenuation has been captured after a pre-go cue presentation, which suggests that it may play a role in motor planning (Kilavik *et al.*, 2013). Additionally, recent research has shown that beta oscillations are related to working memory and seem to display some kind of filtering action against potential distractions that could interfere with the working memory contents (Lundqvist *et al.*, 2018; Schmidt *et al.*, 2019). Thus, as true for all other rhythms mentioned above, also beta waves are involved in a variety of functions, with their characteristics changing in a context-dependent manner.

Gamma

The last and fastest of the brain rhythms is the gamma rhythm. Its frequency range spans from around 30 Hz to 80 Hz, and it is present in a wide array of sensorial and cognitive processes, such as working memory (Lundqvist *et al.*, 2018), perception, and selective attention (Tiitinen *et al.*, 1993). Additionally, some studies suggest that gamma waves arise during high arousal or heightened alertness (Strüber *et al.*, 2000). Again, the gamma rhythm cannot be attributed to a specific function or role in brain signalling, it is instead involved in a multitude of sensory and cognitive functions, with different frequency sub-bands being related to different processes.

Besides the above-described spontaneous rhythmic activity, EEG can exhibit yet another type of behaviour. This behaviour surges as a direct response to a sensory stimulus and is for this reason known as event-related potentials. Since this activity is usually buried in other brain responses, it is acquired over several trials of the same stimulus presentation so that later the ERPs can be obtained and analysed by averaging the waveforms, time-locked with the event cue, which improves the signal to noise ratio (SNR), cancelling and removing the extra activity while preserving the response of interest which is considered constant between trials of the same stimuli. The number of stimulus repetitions necessary to achieve an acceptable SNR is variable, but usually better results are achieved for a greater number of observations, which, to some extent, limits the exploration of ERPs (Nunez and Srinivasan, 2006).

All in all, EEG accounts for a relatively affordable and practical alternative for accessing the brain and the central nervous system. The rhythms and responses recorded through this technique can be studied and used as markers or indicators for behavioural, cognitive, or physiological functions and states. Even though the spatial resolution of scalp EEG is considerably limited, its temporal resolution is high, which makes it an excellent solution for near real-time applications like biofeedback or brain-computer interfaces (BCI).

2.1.6 Eye-Tracking

It is a well-accepted concept that the eyes are powerful sources of information regarding one's state of mind. Findings have shown that pupil size fluctuates with arousal and stress, with pupil diameter increasing during stressful situations and for higher arousal states, even though the positive or negative connotation of the stimulus seems to have no relevant contribution to this response (Partala and Surakka, 2003; Ren *et al.*, 2013; Kassem *et al.*, 2017). Together with the pupil diameter, also the blinking rate is influenced by emotions, with an increase in the number of blinks being positively correlated with stress and higher arousal (Haak *et al.*, 2009). Additionally, eye-related metrics can carry clues about visual attention and cognitive load.

Besides pupil size variations and blinking rate, fixation on a specific target, fixation time, and saccadic movements are indicators of attention levels (McIntire *et al.*, 2014; Maffei and Angrilli, 2019), and cognitive workload (He *et al.*, 2012).

Now, to measure these eye activities, one must resort to an eye-tracking device. Such devices have suffered considerable improvements over time and, nowadays, it is possible to reliably and non-invasively track ocular responses with high resolution, inclusively at considerable distances from the subject. Technology advancements have also lowered the price of this equipment, which has improved their availability for applications outside of the laboratory environment. Most modern eye-trackers rely on video-based techniques, usually involving the use of an infrared or near-infrared light that produces a reflection at the cornea which is used as a reference point to compute the relevant measures (Chennamma and Yuan, 2013).

Out of all the eye metrics, perhaps pupil dilation is the most popular and explored measure, particularly for emotion assessment. The fact that the sphincter and dilator muscles are under the control of the ANS (Hall and Chilcott, 2018), therefore presenting an involuntary and non-manipulable response, maybe the reason behind its good reputation. However, when considering pupil size as an indicator in unpredictable environments or while designing visual emotion eliciting experimental tasks, there is one major concern that must be taken into consideration. The sphincter and dilator muscles of the iris respond fiercely to changes in luminosity. This response can overshadow the ones related to changes in emotional and cognitive state, contaminating the data with unrelated noise. Hence, it is paramount to control and prevent unnecessary luminance changes during the experimental task. If the acquisition takes place at locations where such control is not possible, like outdoor environments, caution should be taken while analysing and interpreting data.

3 State of the Art

3.1 Physiological State Classification

For more than a century, emotion theorists have been debating over the possibility that specific ANS activity patterns accompany each discrete emotion. First, arguing whether different emotions are effectively accompanied by distinct physiological responses, and later, whether the different ANS activity patterns could be utilized to automatically discriminate and classify different emotional states.

The ever-growing interest and development of intelligent systems that can learn from their users, has led researchers to further explore the field of automatic emotion assessment for the optimization of applications that go from healthcare to the automobile or the video game industries. Some of these studies have focused on emotion classification from speech, facial expressions, body movements, or a combination of these (Chen *et al.*, no date; Essa and Pentland, 1997; Wang and Guan, 2004; Sapiński *et al.*, 2019). However, such physical expressions can easily be simulated to appear as if one is experiencing a certain emotion when, in reality, he is not, especially in artificial situations, such as a laboratory setup. In fact, those studies usually rely on databases of subjects expressing emotions "on-demand", lacking ecological validity, which calls into question the real-world applicability of an algorithm based on the classification of physical expressions. Furthermore, these approaches require the recording of the participants either in video or audio, which can be impractical for some applications, and besides can be perceived as invasive. Moreover, if the subjects are aware that they are being recorded at the outset, it can further influence the natural outward expression of emotion.

Emotions, however, involve more than the external bodily expression that is usually associated with each one of them. Damasio and Carvalho (2013) describe them as a set of physiological responses that are mostly triggered by external stimuli and that result in body state changes that may, or may not, be consciously perceived (Damasio and Carvalho, 2013). Those physiological reactions can be measured with the aid of biosensors, and subsequently analysed to look for patterns in the signals, which can lead to correct discrimination and classification of different emotions without the need for a camera or an audio recorder. Besides, those reactions are modulated by the ANS, which means that they constitute unconscious responses that cannot be simulated, mirroring with greater precision the true emotional state of an individual.

Most of the research on automatic emotion classification has focused on classifying ANS activity patterns into basic emotions. The issue with this approach is that no consensus has been reached on a universal and definite set of basic emotions. Another viewpoint, however, is that emotions can be sorted in relation to one or more dimensions. Two dimensions have been consistently mentioned as the most appropriate to differentiate and sort emotional states: valence

and arousal. Valence is an indicator of the pleasantness of the emotion, while arousal reflects its intensity (Winton, Putnam and Krauss, 1984).

Regarding the physiological measures commonly used for emotional state classification, EDA and ECG are possibly the most popular. Healey and Picard (2005) found that metrics derived from those measures presented the highest correlations with stress levels when compared with metrics extracted from electromyography and respiration. Furthermore, several findings have proven that higher EDA levels are related to equally higher arousal states, which evidences the power of this measure as a psychophysiological tool (Winton, Putnam and Krauss, 1984; Lang *et al.*, 1993; Dawson, Schell and Filion, 2007). A drawback of EDA, however, is its lagged response, with a latency of 1-3 seconds, which imposes some limits to the utility of EDA for near-real-time applications, cases when alternative measures may be more appropriate (Dawson, Schell and Filion, 2007). ECG data are used to compute the HR, which evidence suggests is primarily affected by the pleasantness (i.e., valence) of a stimulus rather than its' intensity. Different experiments have reported a relationship between stimulus valence and heart rate (Winton, Putnam and Krauss, 1984; Lang *et al.*, 1993; Legrand *et al.*, 2021). This indicates that measuring cardiac activity is potentially valuable for discriminating between more refined physiological states. Nevertheless, ECG is a quite sensitive signal, easily prone to noise contamination from a variety of sources. Additionally, QRS complexes usually diverge significantly, which makes their detection challenging and sometimes leads to errors in HR computation (Bolanos, Nazeran and Haltiwanger, 2006). These points should be taken into consideration when handling ECG data.

Respiration and skin temperature are also frequently associated with emotional states. A steady and slow respiration rate is usually present in states of low arousal, while fast and irregular respiration characterizes higher arousal (Rainville *et al.*, 2006). Variations in ANS activity linked with emotional change cause blood vessels' constriction to vary which leads to skin temperature fluctuations. Different studies suggest that a decrease in skin temperature is connected to higher activation, while states of calmness are accompanied by higher skin temperature (Genno *et al.*, 1997; Chanel *et al.*, 2011).

A set of measures whose relations to the physiological state have only recently started to be explored are eye-tracking data. Eye movements, gazing targets, and pupil size are some of the measures that can be recorded with eye-tracking devices, in near real-time. While voluntary metrics, such as gaze or eye movement, can be important indicators of cognitive states, like interest or alertness, involuntary metrics like pupil size are probably better suited as emotional state indicators. Evidence suggests that pupil size is sensitive to both, stimulus arousal and valence. Even though dilation invariably occurs for more arousing stimuli, size differences have been detected between negative and positive valences which draws special attention to the potential of this measure (Kassem *et al.*, 2017). The biggest weakness of pupil size measurement, and a probable justification for its delayed adoption, is the major influence of luminosity in pupillary

response in comparison with changes associated with emotional state. This issue limits the use of such measure to controlled settings and stimulus luminance must be taken into consideration when analysing eye-tracking data.

Electromyography (EMG) is often used in automatic emotion assessment studies, particularly facial EMG. It can somewhat be perceived as an alternative to emotion assessment from facial expressions since it measures the level of activity of a chosen muscle. Like HR, facial EMG seems to be more associated with the emotional dimension of valence, rather than arousal or discrete emotions (Larsen *et al.*, 2008). This measure is also associated with stress levels which seem to be directly proportional to muscular activity (Katsis *et al.*, 2008; Lohani, Payne and Strayer, 2019). Some drawbacks of facial EMG, however, are preparation time and the added discomfort, since it requires that electrodes be placed directly on facial muscles. These points may render EMG unsuitable to use in some applications, particularly when involving clinical populations like ASD.

A promising measure, surprisingly not as popular among automatic emotion assessment studies using biosignals, is PPG. Different studies involving ECG and PPG measurements have found significant matches between HR metrics derived from both signals (Bolanos, Nazeran and Haltiwanger, 2006; Pinheiro *et al.*, 2016). The positive results suggest that PPG may even be a more suitable solution than ECG for some applications, given that it is less prone to noise contaminations, and that it is easily measured by placing a sensor on the subject's finger or earlobe, being less intrusive and uncomfortable.

Finally, a measure that is a direct manifestation of the central nervous system (CNS): EEG. One of EEG's main advantages is its high temporal resolution, which makes it a valuable measure for real-time applications. Furthermore, studies have found that, by combining EEG with peripheral physiological measures, the classification accuracies of emotion recognition systems can be improved, suggesting that the combination of peripheral signals with EEG is advantageous for automatic emotion assessment (Chanel *et al.*, 2011). Despite this, the inclusion of EEG in multimodal emotional state recognition systems seems to be underexplored. Instead, it is more common to find studies using either EEG alone or a combination of peripheral signals.

3.1.1 Automatic Emotion Assessment in Autism Spectrum Disorder

While physiology-based automatic emotion assessment has been substantially considered and tested for the typically developed (TD) population, it is underexplored for ASD. To the best of the author's knowledge, there are only three papers describing automatic emotion classification in autistic subjects. By measuring EDA, PPG, skin temperature, EMG, and ECG on children with ASD while they performed computerized cognitive tasks, Liu et al. (2008) successfully attempted to classify emotional states of liking, anxiety, and engagement in this population, achieving an accuracy of 82.9 % with Support Vector Machines (SVM), and possibly being the first to consider

affect recognition as an approach to optimize computerized rehabilitation tools (Liu *et al.*, 2008). Kushki et al. (2015) classified anxiety-related arousal using metrics derived from the ECG and a modified Kalman filter obtaining an average specificity of 92% and sensitivity of 99% (Kushki *et al.*, 2015). More recently, Sarabadani et al. (2020) automatically discriminated positive from negative valence during high and low arousal in ASD obtaining accuracies of 78.1% and 84.5% for high arousal and low arousal, respectively, using K-Nearest Neighbours (KNN), Linear Discriminant Analysis (LDA) and SVM, and combining the outputs of all the classifiers using a Majority Vote to enhance the performance (Sarabadani *et al.*, 2020). The positive outcomes suggest that emotion recognition is a viable but underexplored approach in ASD.

An issue that seems to be overlooked by the existing studies, however, is the evidence of emotion dysregulation in ASD. Several studies have reported discrepancies between physiological indicators of arousal and corresponding self-report ratings in autistic individuals, in comparison to matched control groups (Ben Shalom *et al.*, 2006; Dichter *et al.*, 2010). These findings suggest that either this population experiences emotions differently, or that the individuals' interpretation of such emotions is impaired. It is estimated that about 50% of the ASD population is affected by alexithymia, a condition characterized by one's limitation in interpreting and externalizing his own emotions (Hill, Berthoz and Frith, 2004). This point raises a serious question of whether the use of self-assessment questionnaire answers or labels based on the general population's emotional perception of a stimulus is appropriate as ground truth. The application of fMRI to this population has revealed differences in brain function in comparison to typically developed individuals (Philip *et al.*, 2012). Due to its spatial resolution that allows for the precise mapping of brain regions or networks of interest, fMRI might just be the ideal true state indicator. Sessions involving this imaging technique, however, are quite expensive and nonportable, which limits their applicability and dissemination. The simultaneous acquisition of fMRI and physiological signals, and subsequent exploration to find relationships between ANS physiological patterns and the targeted brain regions modulation, should be investigated under the hypothesis that using brain activation as the ground truth for training emotion recognition algorithms would optimize the outcomes. Furthermore, the knowledge of such relations could be used to infer brain activity outside the MR scanner, using biosignals measurements alone.

4 Methods

To fulfil this dissertation's objectives, which are to assess the viability of constructing machine learning models, tailored for the ASD population, which are efficient to automatically discriminate between different arousal conditions, and to investigate how much the MR environment affects the quality of the biosignals acquired simultaneously with fMRI, an experimental task took place involving the simultaneous acquisition of fMRI and physiological signals in both, autistic and typically developed individuals. Additionally, a similar experimental task was conducted outside the MR environment in a group of typically developed individuals while acquiring a smaller set of physiological signals. Firstly, for each setting and biosignal, the SNR and Percent Signal Change (PSC) was quantified. The data from each setting was then pre-processed to correct for noise interference and artifacts. Events of interest that occurred during the tasks were identified so that signal features corresponding to each observation could be extracted. The features were next analysed to look for signs of significant differences between the studied conditions and groups, and finally, different classifiers were fed with the data, and their performance in automatically discriminating between arousal intensities was assessed.

4.1 Participants

Fifteen individuals diagnosed with ASD (2 females) were recruited to take part in this study with the help of the Portuguese ASD patients' associations from Viseu and Coimbra. The Autism Diagnostic Observation Schedule (ADOS) was employed by an expert to confirm the diagnosis. To form the control group, sixteen TD individuals (4 females) were recruited. All individuals had a full-scale intelligence quotient (FSIQ) score above 70, assessed using the Wechsler Adult Intelligence Scale. Every participant (or their legal representative) signed an informed consent to participate in the study. Every subject completed the entire task. Table 4.1 provides a detailed description of the participants.

Table 4.1 - Demographic description of the ASD and TD groups, including age, Full-Scale Intelligence Quotient (FSIQ), Empathy Quotient (EQ), Autism Spectrum Quotient (AQ), and the Autism Diagnostic Observation Schedule (ADOS-II) total score. Each score is presented in terms of group average and standard error, in brackets. Group differences were assessed with a two-sample T-test, with p-values on the last column. Groups are matched by age.

	ASD	**TD**	***p***
N	15	16	
Age	21.60 (1.19)	23.06 (0.87)	0.32
FSIQ	94.87 (2.63)	111.81 (3.72)	<0.01
EQ	37.13 (3.40)	46.38 (2.57)	0.04
AQ	24.07 (1.20)	15.38 (1.46)	<0.01
ADOS-II	11.07 (0.77)		

4.2 Experimental Task

The experimental task carried out in this study follows a block design and is divided into five task runs. For the first, third, and fifth runs, each block consists of a 15-second video presentation trailed by a self-assessment period and is preceded by a rest period of approximately the same duration. Each of these runs is made up of 10 video trials. The 30 videos (10 videos x 3 runs) were taken from the Chieti Affective Action Videos (CAAV) database and represent different actions, examples include hugging someone, stealing from another, or simply hanging a jacket (Di Crosta *et al.*, 2020). The database comprises videos recorded from both first and third-person perspectives, and each video has a female actor version and a male actor version. For our experiment, we chose to use videos recorded in the first-person perspective and to coincide the gender of the participant with the main actor. Each video in the database is accompanied by the mean rating of valence and arousal given by an evaluation group, comprised of 444 young adults, using a 9-point Self-Assessment Manikin (SAM). To select the appropriate videos for the task, those same ratings were analysed and a quick visual inspection of the two-dimensional space of the ratings revealed that the videos followed a V pattern: the highest levels of arousal can be found in the extreme levels of valence (both positive and negative), while the lowest values of arousal occurred where the valence is null (see Fig. 4.1). Hence, we sorted each video into one of the following 3 categories: low valence and high arousal (LVHA); high valence and high arousal (HVHA); neutral valence and low arousal (NVLA). Thus, 10 videos of each category were selected to integrate the task. For every participant, the order of the 30 videos was randomly shuffled at the begging of the task.

For self-assessment, subjects were asked to rate each video they watched inside the MR also in the 9-point SAM scale. For this purpose, participants used a joystick.

The SAM is a graphic self-assessment instrument that represents an alternative to other verbal and more complex assessment tools. For its simple characteristics and ease of use, SAM has been widely employed to quantify emotional responses to varied types of stimuli in different populations, including children and autistic individuals (Bradley and Lang, 1994; Lang, Bradley and Cuthbert, 1997; Bölte, Feineis-Matthews and Poustka, 2008; Sarabadani *et al.*, 2020). It was then considered appropriate to use in the context of this study to gather the self-assessment ratings from the participants for the emotional dimensions of valence and arousal. Figure 4.2 a) presents a schematic representation of the video trials.

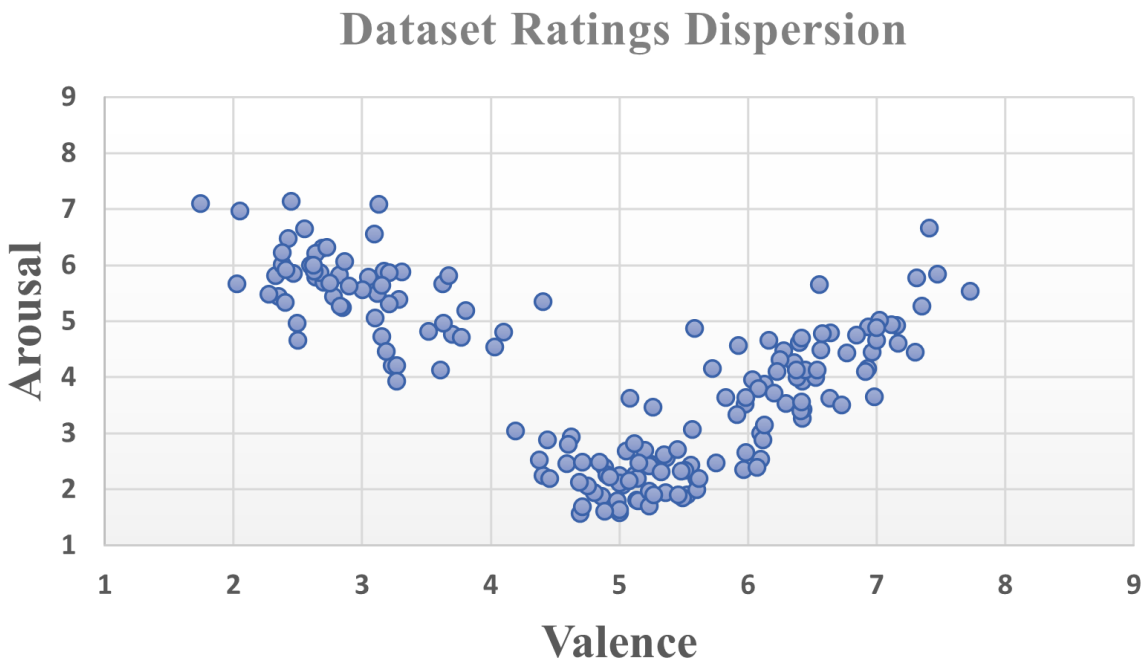

Figure 4.1 - Interaction between the average ratings of valence and arousal given to each video by the validation group. Each point corresponds to a video in the dataset. The cluster on the left corresponds to the videos rated as low valence and high arousal (LVHA category), closer to the bottom and slightly to the right is the cluster corresponding to the neutral videos (NVLA category), finally, to the right, are the data points corresponding to the videos that were rated high in both dimensions (HVHA category).

The second and fourth runs are an adaptation of the Balloon Analogue Risk Task (BART), a behavioural measure of risk-taking, where risky behaviour is rewarded up until a point when taking further risks represents losses for the individual. In the original version of the BART, participants were shown a virtual balloon and were instructed to inflate it. To inflate the balloon, without it popping, translated into a monetary reward wherein the amount of money earned was proportional to the final balloon volume. However, if the balloon popped, the participant would lose all the money accumulated from previous rounds (Lejuez *et al.*, 2002). Our version consists of a further gamified BART with the introduction of competition, in this case, against the computer. The participant only sees his balloon, but knows the computer is performing the same task and his objective is to obtain a larger volume than the computer. Also, the monetary reward was replaced by two different kinds of rewards: social and interest-based rewards. The social-based reward consisted in the presentation of images of people with approving expressions whereas the interest-based reward consisted of images depicting objects related to a topic of interest of the individual. This last kind of reward was personalized for each participant.

So, in the second and fourth task runs, each task block consists of a game trial. A game trial is made up of three consecutive balloon rounds. To win the trial and level up, the participant must inflate the balloon further than the computer at least in two out of the three rounds, knowing that it may burst at any given moment. In case of a win, the participant levels up and is rewarded in one

of the runs with the interest-based reward, and the other, with the social-based reward, described above. In each block, before the game trial, there is a rest period of approximately 15s. Finally, the block closes with the valence and arousal self-assessment period using the 9-point SAM scale. To inflate the balloon and complete the self-assessment, participants used a joystick. Figure 4.2 b) is a schematic representation of the adapted BART rounds.

Before each session, the task was explained and participants were asked to rate some training videos as well as to play a few rounds of the BART (without rewards), to guarantee that they understood the task, that both concepts of valence and arousal were grasped, and that they knew how to operate the joystick.

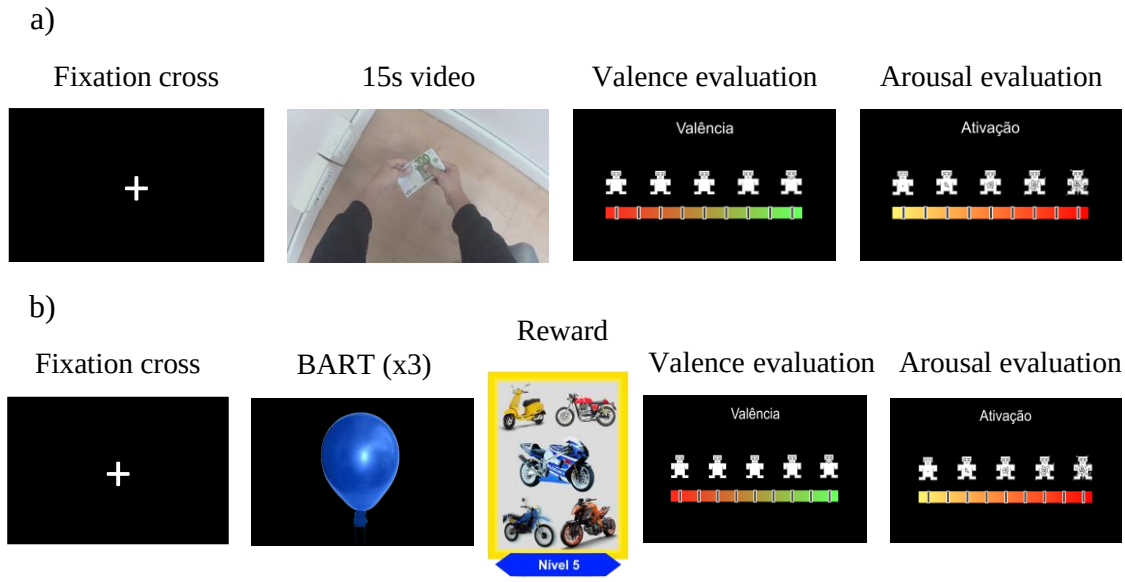

Figure 4.2 - Schematic representation of the experimental task, including stimuli, self-assessment, and structure. a) Stucture of the video trials. Each trial begins with a fixation cross lasting approximately 15s followed by a video of that same length that is finally trailed by the valence and arousal self – assessment period. b) Stucture of the BART adaptation trials. Like the video trials, each BART trial starts with a fixation cross of approximately 15s that is followed by 3 BARTS. To win, the participant must inflate the balloon further than the computer at least in 2 out of the 3 tries. In case of win, the participant levels up and is rewarded in one of the runs with images relative to a topic of his interest and in the other with images of people with approving expressions (social reward). Finally the trial closes with the valence and arousal self-assessment period.

4.3 Data Acquisition

While participants performed the task, and simultaneously with fMRI acquisition, EEG, EDA, SpO2, and HR were acquired using the MP150 system and AcqKnowledge 4.2 software (BIOPAC Systems, Inc.). Respiration and PPG were recorded using the Physiological Measurement Unit (PMU) of the MRI scanner (Siemens Healthcare) and pupil size was registered using the EyeLink 1000 Plus Eye Tracker with the long-range mount (SR Research Ltd.).

Due to the hypersensitivity of the ASD population, we tried to simplify and reduce preparation time as much as possible, thus, EEG was acquired using only 3 electrodes: signal channel placed on the forehead, reference channel on either the right or left earlobe, and the ground electrode on the temporal area. EDA was measured using 2 Ag/AgCl electrodes taped to the proximal phalanges of the index and middle fingers of the participant's non-dominant hand. SpO2, HR, and PPG were measured using two different pulse finger sensors, one connected to the BIOPAC module (SpO2 and HR) and the other to the PMU (PPG). Respiration was measured with a pneumatic respiratory cushion attached to the participant using a respiratory belt. EEG, EDA, SpO2, and HR were recorded with a sampling rate of 5000 Hz, PPG and Respiration were acquired at 400 Hz, and pupil size at 500 Hz.

For results comparison purposes, besides the physiological signals acquired simultaneously with fMRI, a smaller set of signals including PPG, respiration, EDA, and pupil size were recorded in a regular laboratory setting, in a group of typically developed individuals, while the participants watched and rated the same videos that were used in the video runs of the fMRI experimental task. All biosignals, except pupil size, were acquired using the BioNomadix system (BIOPAC Systems, Inc.) at a sampling rate of 5000 Hz. Pupil size was recorded using the EyeLink II eye tracker (SR Research Ltd.) at a sampling rate of 500 Hz.

4.4 Behavioural Analysis

First, to understand how much the self-assessment answers of our two groups for the video runs were in agreement with the database labelling and with each other, a correlation analysis was performed recurring to Pearson's linear correlation coefficient. The mean valence and arousal values for each video, and for each group, were first calculated. The results of each group were then compared to the database results and each other.

Additionally, to verify if the answers reported by the two groups for each video are in accordance with the three categories from which they were retrieved (LVHA, HVHA, NVLA), the average answers from the database and the clinical and control groups were plotted and inspected in the two-dimensional space of valence and arousal.

4.5 Signal-to-Noise Ratio Quantification

To ascertain the overall quality of the acquired biosignals, the average SNR was quantified. To do this, for each subject, the mean of every signal across the 5 runs was calculated. In the end, the SNR was found by averaging the mean values of each subject, and dividing it by the standard deviation, according to equation (4.1), where μ represents the overall mean of the signal and σ represents its standard deviation.

$$SNR = \frac{\mu}{\sigma} \tag{4.1}$$

4.6 Percent Signal Change

To assess if the events in the experimental task induced significant changes in the physiological signals, percent signal change (PSC) was calculated at each event in the video runs, i.e., at each video, for every participant, and every biosignal. A one-sample Wilcoxon Signed Rank test was then performed, first for the PSC values of the events of each participant, for each signal, and then, combining the PSC values of every participant in a group, the same test was applied for each signal. PSC for each event was obtained according to equation (4.2).

$$PSC = \frac{\bar{x}_{event} - \bar{x}_{baseline}}{|\bar{x}_{baseline}|} \tag{4.2}$$

where,

$\bar{x}_{event}$ = average value of the signal during the event of interest

$\bar{x}_{baseline}$ = average value of the signal during the baseline period, immediately preceding the event of interest

4.7 Data Pre-processing

Photoplethysmogram

To reduce possible noise contamination coming from different sources, like movement, baseline drift, or the powerline, the PPG signal was bandpass filtered using a 6th order Butterworth filter with a lower cut-off frequency of 0.5 Hz and a higher cut-off frequency of 20 Hz. The clean PPG signal was then used to compute the PRV by first identifying the PPG pulse peaks, subsequently computing the interbeat intervals (IBI), and finally using a cubic spline interpolation to obtain a uniformly sampled time series with a new sampling frequency of 4 Hz. This last step is recommended for frequency domain analysis of the PRV (Peltola, 2012).

Electrodermal Activity

The electrodermal activity data acquired simultaneously with fMRI were high-pass filtered with a 0.5 Hz cut-off frequency as it was collected and were later low-pass filtered using a 5^{th} order Butterworth filter with a 1 Hz cut-off frequency to edit the signal regarding existing movement artifacts and noise.

As for the same signal, acquired outside the MR environment, a moving average filter with a 1s window was used to smooth the data and get rid of the high-frequency noise.

Electroencephalogram

As expected, the MR gradient switching considerably contaminated the EEG recordings with artifacts. So, to clean the signal of such artifacts, the FMRI Artifact Slice Template Removal (FASTR) algorithm from the FMRIB plug-in for EEGLAB (version 1.21) was used. Feeding the algorithm with the corrupted signal and the events for each slice acquired, it computes an average template for the artifact and subtracts it from the EEG, locked to each slice trigger. The corrected EEG was then low pass filtered using an infinite impulse response filter with a cut-off frequency of 80 Hz.

Peripheral Oxygen Saturation and Heart Rate

To get rid of the random noise that was present in the SpO2 and HR recordings, these signals were smoothed using a moving average filter with a window size corresponding to approximately 12% of the sampling frequency.

Respiration

The respiration signal acquired outside the MR environment was considerably noisier than the one obtained in the scanner. Hence, before analysis, it was low pass filtered using a Chebyshev Type II filter with a cut-off frequency of 1 Hz.

Similar to what we did to obtain the PRV, the respiration signal was used to compute the breath rate variability (BRV) to investigate the respiratory frequency components. The respiration peaks, corresponding to the end of inhalation were identified and the interbreath intervals were subsequently computed. A cubic spline interpolation was applied to obtain a uniformly sampled time series with a new sampling frequency of 4 Hz.

4.8 Feature Extraction

Signal Segmentation

To extract features from the acquired biosignals, the signals from each run were first divided into individual segments time-locked to each event of interest.

For the video runs, where we expect a slower physiological response, evolving as the video progresses and even after it ends, as the information is taken in, every signal was divided into 30-second segments, time-locked to the beginning of each video. This way, each segment includes the 15 seconds of the video, the self-assessment period, and some seconds after.

In another way, considering the fleeting nature of the BART events of interest, which are quickly replaced by others, and to avoid capturing physiological information unrelated to the present event, the signals for the two BART runs were segmented into 10-second epochs whose beginning coincides with the following events: individual balloon trials ('balloon' event), balloon pops ('pop' event), level up during the social reward runs ('win_S' event), stay on the same level during the social reward runs ('loss_S' event), level up during the interest reward runs ('win_I' event), stay on the same level during the interest reward runs ('loss_I' event).

Additionally, for every video and BART run, and for every block, baseline segments, coinciding with the last seconds of the fixation cross, presented at the begging of each block, were defined for every signal. Their duration coincided with the last 5 seconds of the fixation cross, for the video blocks, and the last 10 seconds, for the BART blocks.

Feature Extraction

For each segment, a total of 70 features was extracted from the different biosignals. To account for possible carryover effects, a baseline period previous to each event was considered to normalize the features, so they represent event-induced fluctuations instead of absolute values.

For each signal, a description of the extracted features is given in the following tables (Table 4.2 to Table 4.10).

Table 4.2 - EDA features and their description.

Feature	Code Name	Description
Mean EDA value	meanEDA	Mean value of EDA. Evidence suggests that this metric is significantly correlated with arousal (Lang *et al.*, 1993).
Maximum EDA value	maxEDA	The maximum value of EDA.
Minimum EDA value	minEDA	The minimum value of EDA.
Mean absolute first difference of EDA	meanAbsFdEDA	Approximation of the first derivative, and thus, related to signal changes. Defined as: $\frac{1}{N-1}\sum_{n=1}^{N-1} \mid x_{n+1} - x_n \mid,$ (4.3) where x_n represents a signal sample and N equals the total number of samples in the segment.
Mean absolute second difference of EDA	meanAbsSdEDA	Approximation of the second derivative, and thus, related to signal changes. Defined as: $\frac{1}{N-2}\sum_{n=1}^{N-2} \mid x_{n+2} - x_n \mid.$ (4.4)
Mean Derivative of EDA for negative values only	meanDerivNegEDA	Mean of the negative values of the first derivative. Represents the average decrease rate of the signal.
Decrease time of EDA signal	decTimeEDA	Total decrease time of EDA relative to the total duration of the segment.
Number of EDA falls	nFallsEDA	Number of EDA falls per minute, measured by identification of local minima.
Mean amplitude of SCRs	meanAmpSCR	Mean amplitude of the SCRs. Signal peaks were considered as SCR if their amplitude was higher than the minimum threshold of 0.01 μS. SCRs amplitude is positively correlated to the number of activated sweat glands (Freedman *et al.*, 1994; Nishiyama *et al.*, 2001) and hence may represent a good indicator of SNS activation.
Maximum amplitude of SCRs	maxAmpSCR	The maximum amplitude of the SCRs.
Rate of SCRs	rateSCR	Number of SCRs per minute.
Mean rise duration of SCRs	meanRiseDurSCR	Mean rise time between SCRs onsets and corresponding peaks.
Standard deviation of the rising duration of SCRs	SDRiseDurSCR	Standard deviation of the rise times of SCRs.

Table 4.3 - PPG features and their description.

Feature	Code Name	Description
Mean PPG value	meanPPG	Mean value of the PPG segment.
Maximum PPG value	maxPPG	The maximum value of the PPG segment.
Minimum PPG value	minPPG	The minimum value of the PPG segment.
Standard deviation of the PPG signal	stdPPG	Standard deviation of the PPG segment values.
Mean amplitude of the PPG pulse peaks	meanPeakAmp	The average amplitude of the PPG segment pulse peaks. Pulse amplitude is related to blood volume and skin temperature, and its low-frequency fluctuations have been attributed to the SNS (Shaffer and Ginsberg, 2017).
Standard deviation of the PPG pulse peaks' amplitude	stdPeakAmp	Standard deviation of the PPG segment pulse peaks' amplitude. Pulse amplitude is related to blood volume and skin temperature, and its low-frequency fluctuations have been attributed to the SNS (Shaffer and Ginsberg, 2017).

Table 4.4 - HRV time-domain features and their description.

Feature	Code Name	Description
Mean of Normal-to-Normal (NN) time intervals	NNmean	Average IBI.
Minimum NN time interval	NNmin	Shortest IBI.
Maximum NN time interval	NNmax	Longest IBI.
Standard deviation of NN intervals	SDNN	Standard deviation of the IBI. This metric is a representation of the rhythmic components that induce variability in the HR (Malik *et al.*, 1996), and evidence suggests it is influenced by both, the SNS and the parasympathetic nervous system (PNS) (Shaffer and Ginsberg, 2017).
Mean of Successive Differences between NN time intervals	meanDeltaNN	Average of the differences between successive IBI.
Standard Deviation of Successive Differences between NN time intervals	SDSD	Standard deviation of the differences between successive IBI. This feature reflects the short-term variability of HR (Shaffer and Ginsberg, 2017).
Root Mean Square of Successive Differences between NN time intervals	RMSSD	Root mean square of the differences between successive IBI. Like SDSD, this metric is related to the short-term components of HRV. RMSSD is known for being mainly influenced by the PNS and is then the principal time-domain measure used to assess vagal tone contribution to the HRV (Shaffer and Ginsberg, 2017).
Number of Successive Differences between NN intervals greater than 50ms	NN50	The number of successive IBI that differ in length by more than 50 ms. Like RMSSD, NN50 is mostly modulated by the HF components of the HRV (Shaffer and Ginsberg, 2017).
The ratio between NN50 and the total number of NN time intervals	pNN50	The ratio between the number of successive IBI that differ in length by more than 50 ms and the total number of IBI. Similarly to RMSSD and NN50, pNN50 is mainly influenced by the PNS, and thus shows correlations with the two previous measures (Shaffer and Ginsberg, 2017).

Table 4.5 - HRV frequency-domain features and their description.

Feature	Code Name	Description
Relative power in the Very Low Frequency (VLF) band of the HRV	VLF	Power in the VLF band [0.003 – 0.04] Hz of the HRV, normalized by total power [0.003 – 0.4] Hz. Although the physiological correlates of this band are unsettled, PNS activity seems to influence its power (Shaffer and Ginsberg, 2017).
Relative power in the Low Frequency (LF) band of the HRV	LF	Power in the LF band [0.04 – 0.15] Hz of the HRV, normalized by total power. While some argue that the LF activity is an indicator of sympathetic activation, studies that resulted in LF power decreases during conditions related to SNS activation have suggested that LF contributions are more complex than initially thought (Malik *et al.*, 1996).
Relative power in the High Frequency (HF) band of the HRV	HF	Power in the HF band [0.15 – 0.4] Hz of the HRV, normalized by total power. Contrarily to the uncertain nature of the LF components, it is well established that the HF activity is mostly related to the PNS (Malik *et al.*, 1996). Accordingly, states of stress and anxiety are accompanied by a decrease in HF power, while positive emotions seem to be related to a higher power in the HF band (Shaffer and Ginsberg, 2017).
The ratio between the powers in the LF and HF band	RaLH	Despite some contestation and assuming that the LF activity is mostly related to SNS activation, some believe that this is a marker of sympathovagal balance, with a higher value reflecting a dominance of the SNS, while a lower value reflects parasympathetic dominance (Shaffer and Ginsberg, 2017).
The ratio between the powers in the LF and VLF band	RaLVL	
The ratio between the powers in the HF and VLF band	RaHVL	

Table 4.6 - EEG features and their description.

Feature	Code Name	Description
Relative power in the delta band of the EEG	delta	Power in the delta band [0.5 – 4] Hz of the EEG, normalized by total power [0.5 – 80] Hz.
Relative power in the theta band of the EEG	theta	Power in the theta band [4 – 8] Hz of the EEG, normalized by total power.
Relative power in the alpha band of the EEG	alpha	Power in the alpha band [8 – 13] Hz of the EEG, normalized by total power.
Relative power in the beta band of the EEG	beta	Power in the beta band [13 – 30] Hz of the EEG, normalized by total power.
Relative power in the gamma band of the EEG	gamma	Power in the gamma band [30 – 80] Hz of the EEG, normalized by total power.

Table 4.7 - Respiration features and their description.

Feature	Code Name	Description
Mean value of respiration	meanResp	Average value of the respiration segment.
Maximum value of respiration	maxResp	Highest value of the respiration segment.
Minimum value of respiration	minResp	Lowest value of the respiration segment.
Mean absolute first difference of respiration	meanAbsFdResp	Approximation of the first derivative, and thus, related to signal changes – Equation (4.2).
Mean absolute second difference of respiration	meanAbsSdResp	Approximation of the second derivative, and thus, related to signal changes - Equation (4.3).
Respiration rate	respRate	Number of breaths per minute. This metric was calculated by first identifying the signal peaks, corresponding to the end of inhalation, counting the number of occurrences in the segment, and converting it to breaths per minute.
Mean respiration amplitude	meanRespAmp	Mean amplitude of breath. Calculated by first finding the signal peaks, corresponding to the end of inhalation, and the slope bases, considered as the inhalation onsets,

		computing the differences between the corresponding signal values, and finally averaging the amplitude values. Both, the signal peaks and slope onsets were identified recurring to (MacIntyre, 2021).
Median respiration amplitude	medianRespAmp	Median amplitude of breath.
Standard deviation of respiration amplitude	stdRespAmp	Standard deviation of the breath amplitude.
Mean rise duration of respiration	meanRiseDurResp	Average duration of inhalation in seconds.
Standard deviation of the rise duration of respiration	SDRiseDurResp	Standard deviation of inhalation duration in seconds.
Mean respiration interval	meanRespInter	Mean duration of a respiration cycle, in seconds. Measured by averaging the differences between two successive inhalation peaks.
Median respiration interval	medianRespInter	Median duration of a respiration cycle, in seconds.
Minimum respiration interval	minRespInter	Shortest respiration cycle in the segment.
Maximum respiration interval	maxRespInter	Longest respiration cycle in the segment.
Standard deviation of the respiration intervals	stdRespInter	Standard deviation of the respiration cycles duration.
Absolute power in the HF band of the breath rate variability	respHF	Power in the HF band [0.15 – 0.4] Hz of the breath rate variability, correlated with parasympathetic activity (Soni and Muniyandi, 2019).
Standard deviation of respiration	stdResp	Standard deviation of the respiration value.
Dynamic range of respiration	DRResp	Dynamic range of respiration, measured by finding the difference between the maximum and minimum value of the respiration segment.

Table 4.8 - Pupil size features and their description.

Feature	Code Name	Description
Mean pupil size	meanPupilSize	Mean value of pupil size.
Maximum pupil size	maxPupilSize	Maximum value of pupil size.
Number of blinks	nBlinks	Number of eye blinks occurring during the segment.

Table 4.9 - Instantaneous HR features and their description.

Feature	Code Name	Description
Mean Heart Rate	meanHR	Mean value of the instantaneous HR.
Maximum Heart Rate	maxHR	Maximum value of the instantaneous HR.
Minimum Heart Rate	minHR	Minimum value of the instantaneous HR.
Mean absolute first difference of the Heart Rate	meanAbsFdHR	Approximation of the first derivative, and thus, related to signal changes (Equation 4.2).
Mean absolute second difference of the Heart Rate	meanAbsSdHR	Approximation of the second derivative, and thus, related to signal changes (Equation 4.3).
Standard deviation of the Heart Rate	stdHR	Standard deviation of the instantaneous HR.

Table 4.10 – Pulse oximetry features and their description.

Feature	Code Name	Description
Mean SpO2	meanSpO2	Mean peripheral oxygen saturation in %.
Maximum SpO2	maxSpO2	Maximum peripheral oxygen saturation in %.
Minimum SpO2	minSpO2	Minimum peripheral oxygen saturation in %.

4.9 Statistical Analysis

Following feature extraction, a statistical analysis was conducted intending to inspect the features for significant differences among conditions and groups.

For the video runs the means of every feature for each subject were computed, for the conditions of High Arousal (HA), Low Arousal (LA), High Valence (HV), Neutral Valence (NV), and Low Valence (LV). The HA and LA conditions were obtained by condensing the 3 original ones (LVHA, HVHA, NVLA) and all conditions were defined and analysed twice. One for the database ratings and the other for the self-assessed answer values given by each participant. For the self-assessment, k-means clustering was performed on each participant's answers individually to partition them into 2 and 3 clusters, for the arousal (high/low) and valence (high/neutral/low) conditions, respectively. In some rare cases, the participant did not respond in time, which resulted in some missing values in the answers. For these cases, a linear regression was fitted to the relationship between the existing answers and the database ratings of the corresponding video. The missing answers were then extracted by evaluating the regression at the corresponding points, using the database rating of the video for which an answer is missing as the independent variable. Finally, trials were classified based on the cluster they fell into. Wilcoxon signed-rank tests were then performed to look for statistically significant differences in feature values between HA and

LA, HV and NV, HV and LV and NV and LV, for each group, and Wilcoxon rank-sum tests were applied to look for differences between groups for the five conditions.

As for the BART runs, the primal events of interest were considered winning or losing a play. Additionally, given the evidence of an altered reward processing in ASD, the distinction between winning a social or a personalized interest reward was also considered for analysis.

The means of each feature, for each subject, were then calculated, first for the general events of winning and losing, and then for the particular events of winning a social reward and winning an interest reward.

Wilcoxon signed-rank tests were then performed to look for statistically significant differences in feature values for each group between the winning and losing events and between winning a social reward and winning an interest reward. In addition, Wilcoxon rank-sum tests were applied to look for differences between groups for the four different events selected.

4.10 Classification

To explore the accuracy of automatic emotion assessment in the data acquired with our experimental protocol, 4 classification algorithms were tested for the classification problems summarized in Table 4.11.

As ground truth for the video samples, both the ratings of the CAAV database and the self-assessment answers were considered separately to label the data, to ascertain which case resulted in the best classifier performance. For the BART samples, we first considered the segments corresponding to the events: ‘win_S’, ‘win_I’, ‘loss_S’ and ‘loss_I’, and attempted to classify the observations as a win or a loss, regardless of the reward involved, and then, taking only the win events, each observation was labelled either as winning a social reward or an interest reward, these labels were then used as ground truth for the win observations.

Table 4.11 - Summary of the classification problems considered for analysis.

Code Name	Task	Target	Labels	Ground Truth
A1	Videos	Arousal	High Arousal, Low Arousal	Database average ratings
A2	Videos	Arousal	High Arousal, Low Arousal	Self-assessment answers
A3	Videos	Valence	High Valence, Neutral Valence, Low Valence	Database average ratings
A4	Videos	Valence	High Valence, Neutral Valence, Low Valence	Self-assessment answers
A5	BART	Game outcome	Winning, Losing	Game outcome
A6	BART	Reward type	Social Reward, Interest Reward	Reward type

The considered classifiers were a Euclidean Minimum Distance Classifier (MDC – Euclidean), a K-Nearest Neighbours (KNN), and SVM using a Radial Basis Function (SVM RBF) kernel and a linear (SVM Linear) kernel. Additionally, in an attempt to optimize the classifiers performance, Principal Component Analysis (PCA) was employed on the feature space to reduce redundancy and the classifiers were retrained for the new dimensionality. A brief background on each classifier and PCA will be given next, to help understand what characteristics may optimize the classification results.

Principal Component Analysis

Principal Component Analysis is a multivariate statistical analysis technique that reduces the dimensionality of the feature space, while minimizing redundancy between features, by finding the directions where data presents a greater variance and projecting it onto them, so that the axes of the new feature space, also known as principal components, correspond to such directions. An important requirement that must be met is that the principal components are uncorrelated to each other, and thus, mutually orthogonal.

Now, given that we want to reduce the dimensionality of our dataset, we must keep only the principal components that explain the most variance. The eigenvectors of the data covariance matrix correspond to the principal components, and their magnitude is given by the corresponding eigenvalues which are correlated to the amount of variance they bear. After obtaining the eigenvalues, some methods can help with deciding how many dimensions to keep. Here, the Kaiser Criterion was considered, which states that the principal components to keep correspond to the

ones whose corresponding eigenvalue magnitudes are higher than 1. Finally, the transformation matrix that maps the dataset into the new lower-dimensional space is built using only training data, and then, both training and test data are projected accordingly.

Minimum Distance Classifier - Euclidean Distance

Minimum distance classifiers are linear discriminant approaches based on distance metrics. When considering the Euclidean distance (Equation 4.4) to evaluate the similarity, these classifiers consider a class's mean vector as its representative and assign each object to the class which representative is nearest, i.e., has a greater similarity to the object.

$$D_i(x) = \sqrt{(x - m_{i)}{}^T (x - m_i)} \tag{4.5}$$

where,

D_i = Euclidean distance to the mean vector of class i

x = Data vector

m_i = Mean vector of class i

K-Nearest Neighbours

Contrarily to the MDC, where assumptions are made about the model that describes the pattern distribution, the K-Nearest Neighbours (KNN) classifier is a model-free technique, i.e. a non-parametric learning algorithm. It bases itself on the estimation of the probability density function of the pattern distributions described in Equation 5.

$$p(x) \cong \frac{k}{nV} \tag{4.6}$$

where,

p = Probability density function

x = Data vector

V = A volume surrounding x

n = Total number of points

k = Number of points inside V

The method is applied by fixing k and then determining the minimum V that encloses k patterns. A low-density value is obtained when there is only a small number of points surrounding x, while a high density is obtained when there are many points. As a classifier, k translates to the k - nearest neighbours of x, and x is classified with the class label that appears most frequently on said neighbours.

Since it does not generate a model from the training set, KNN is called a lazy learning algorithm. On the one hand, this kind of algorithms have the advantage of requiring less

computational costs in the training phase, hence taking less time to train. On the other hand, the prediction phase of such classifiers is slower, since the algorithm has to search the entire training data set for the k-nearest points, additionally, it also requires a lot more training observations to accurately make predictions.

For this project, the ideal number of neighbours to use was determined by applying a 50/50 partition on the training set 5 times, varying k from 1 to 20 for each partition and calculating the error between the predicted classes and the actual classes for each case. Finally, the k that obtained the smallest error for the most runs was chosen to retrain and test the classifier.

Support Vector Machines

Support Vector Machines (SVM) represent a machine learning technique based on statistical learning theory that owe their popularity to their good generalization performance. Originally binary classifiers, these algorithms working principle consists in finding the optimal hyperplane, i.e., the one that maximizes the margin between the training samples of the different classes. This margin is the hypervolume in between the hyperplanes constructed through the closest examples of both classes. Such examples are called support vectors and owe their name to the physical intuition that the margins are supported on them.

Even though these principles were thought primarily for linearly separable data, it is still possible to apply them to non-linearly separable problems. The principle behind non-linear SVM is to implicitly transform the dataset to a higher dimensionality space where linear separation becomes possible. These non-linear separations can be made with the help of kernel functions.

For this project, both the Linear kernel and the Gaussian Radial Basis Function (Gaussian RBF) (4.6) were explored.

The Gaussian RBF is one of the most popular kernels for its resemblance to the normal distribution. The kernel function calculates the distance between two feature vectors, i.e., how similar they are to each other. The Gaussian Radial Basis Function is given by:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{4.7}$$

where,

$||x_i - x_j|| =$ Euclidean distance between feature vectors x_i and x_j

$\sigma =$ Variance, and the kernel hyperparameter

Besides the kernel parameter in the Gaussian RBF, another parameter that can be tuned for optimal results is the cost (C), as it is usually named. The cost parameter determines the level of penalization imposed for misclassifications. The higher the cost, the higher the penalty for wrong

classifications and the more the algorithm tries to separate the classes. However, if the value of this parameter is too large, it can lead to overfitting and longer training times, thus it is important to find the value that optimizes the process. Like for KNN, the values of the optimal hyperparameters (σ and C) were determined by applying a random 50/50 partition on the training set 5 times, the error between the predicted classes and the actual classes was computed for each combination of the parameters for the Gaussian RBF and for each value of C, for the case of linear SVM. To classify the test samples, the combination of values and value, respectively, that obtained the smallest error for the most runs was used. After the hyperparameters selection, the classifiers were retrained with all training data with the chosen parameters.

The classifiers were then tested for both intraparticipant and interparticipant classification. For the intraparticipant approach, data from each subject were randomly split using the 70:30 ratio, where 70% of the data were used to train the classifier and the remaining data were used for testing. This process was repeated 30 times to avoid outlier results. As for the interparticipant classification, we employed the Leave One Subject Out (LOSO) method where the data from each participant are used once for testing, while the rest of the participants' data are used to train the classifier. The LOSO approach was selected over the 70:30 split due to the low number of subjects.

Finally, permutation tests were employed to ascertain if the accuracies of the classifiers were significantly greater than the chance level. For each partition, after testing, the true labels of the test set were iteratively shuffled, and accuracies were calculated using the random labels as the predicted classes. The number of times that these random accuracies were greater than the one obtained with the classes predicted by the classifier were then counted, whose percentage corresponds to the statistical significance level of the classifier.

5 Results

5.1 Behavioural Analysis

Overall, the behavioural analysis revealed strong positive correlations between average database ratings and average self-assessment answers for the two dimensions of valence and arousal. We observed, however, that for the arousal dimension, both ASD and TD groups, tend to give lower ratings than those given by the CAAV database validation group, which is not observed for the valence dimension. Additionally, from Figures 5.2, 5.6 and 5.7, we can conclude that participants in the ASD group tended to give extreme valence ratings, avoiding the neutral ones, which is not verified for the arousal dimension.

Figures 5.1 and 5.2 display the relationship between the average database answers and average ASD answers, for the arousal and valence dimensions, respectively. Figures 5.3 and 5.4 are analogous to Figures 5.1 and 5.2 but are in respect to the TD group. Figures 5.5 and 5.6 present the relationship between the average ratings of the two groups, for the arousal and valence dimensions, respectively.

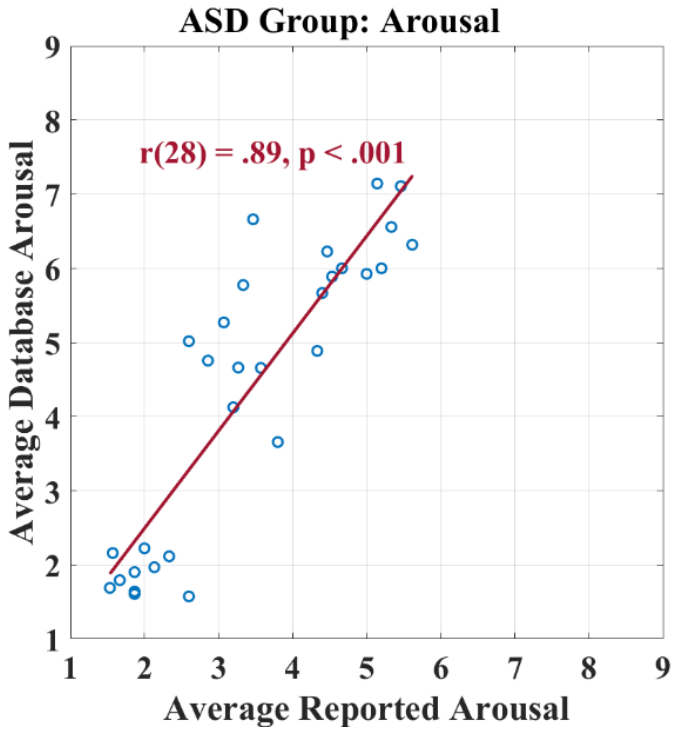

Figure 5.1 - Relationship between the average database answers and the average reported answers of the clinical group (ASD), for the arousal dimension. Each blue circle represents one of the videos chosen to integrate the experimental task. Red line represents the first-degree polynomial that is a best fit to the data. r represents the Pearson's linear correlation coefficient, and p represents the statistical significance.

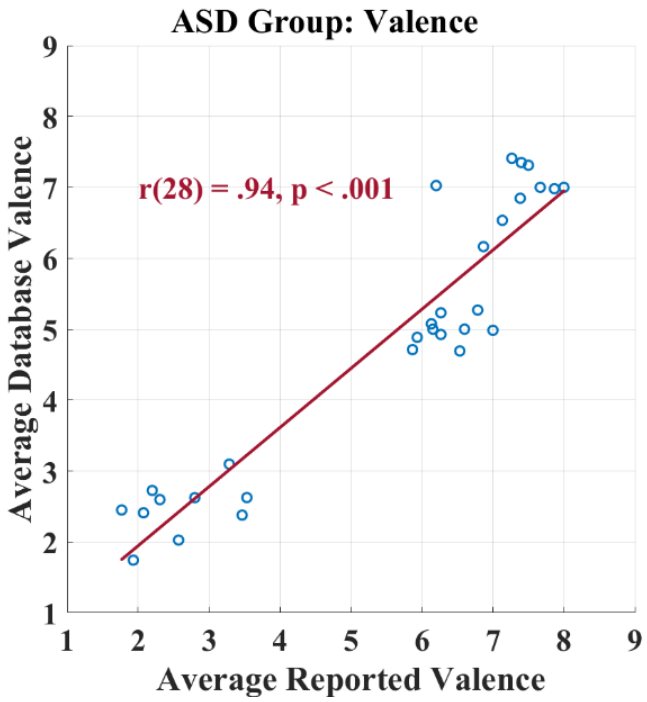

Figure 5.2 - Relationship between the average database answers and the average reported answers of the clinical group (ASD), for the valence dimension. Each blue circle represents one of the videos chosen to integrate the experimental task. Red line represents the first-degree polynomial that is a best fit to the data. r represents the Pearson's linear correlation coefficient, and p represents the statistical significance.

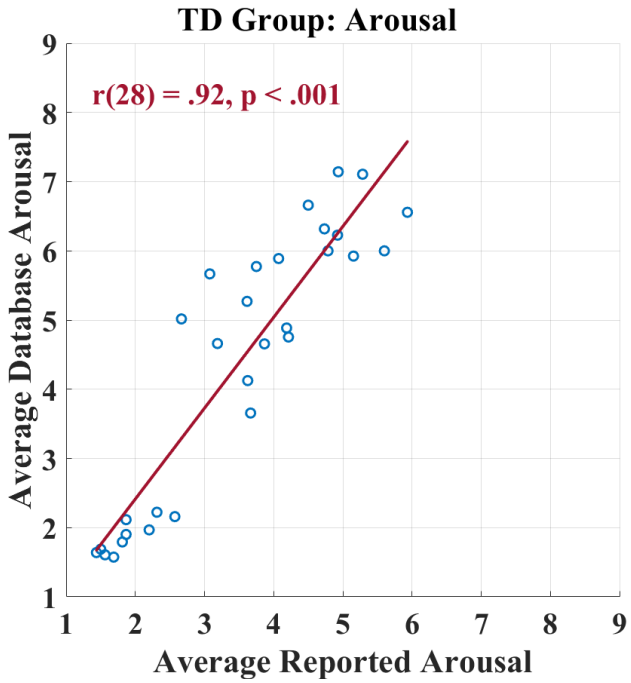

Figure 5.3 - Relationship between the average database answers and the average reported answers of the control group (TD), for the arousal dimension. Each blue circle represents one of the videos chosen to integrate the experimental task. Red line represents the first-degree polynomial that is a best fit to the data. r represents the Pearson's linear correlation coefficient, and p represents the statistical significance.

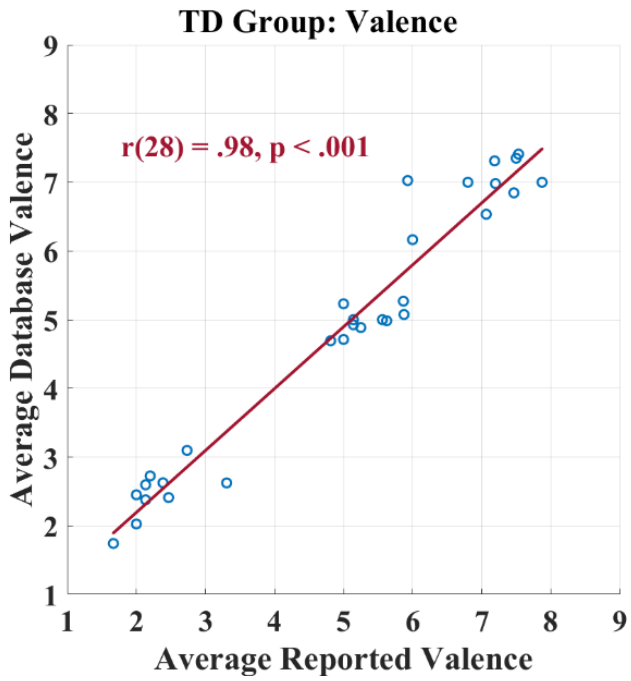

Figure 5.4 - Relationship between the average database answers and the average reported answers of the control group (TD), for the valence dimension. Each blue circle represents one of the videos chosen to integrate the experimental task. Red line represents the first-degree polynomial that is a best fit to the data. r represents the Pearson's linear correlation coefficient, and p represents the statistical significance.

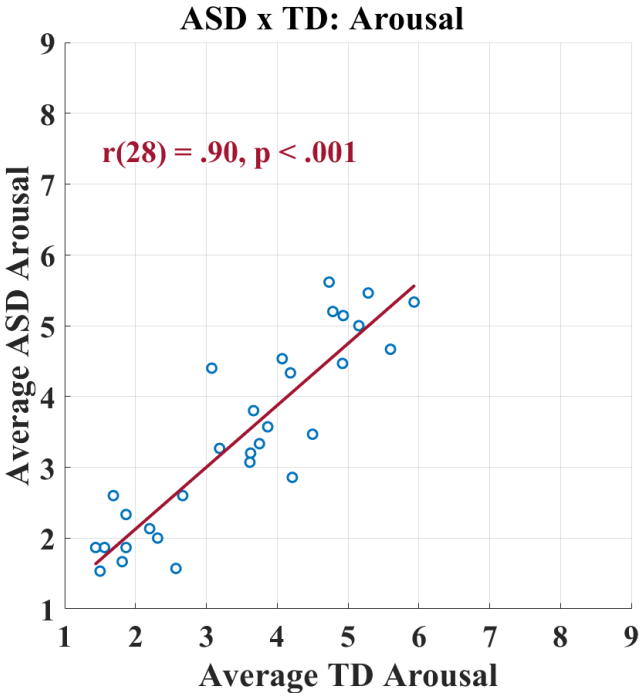

Figure 5.5 - Relationship between the average answers of the clinical and control groups, for the arousal dimension. Each blue circle represents one of the videos chosen to integrate the experimental task. Red line represents the first-degree polynomial that is a best fit to the data. r represents the Pearson's linear correlation coefficient, and p represents the statistical significance.

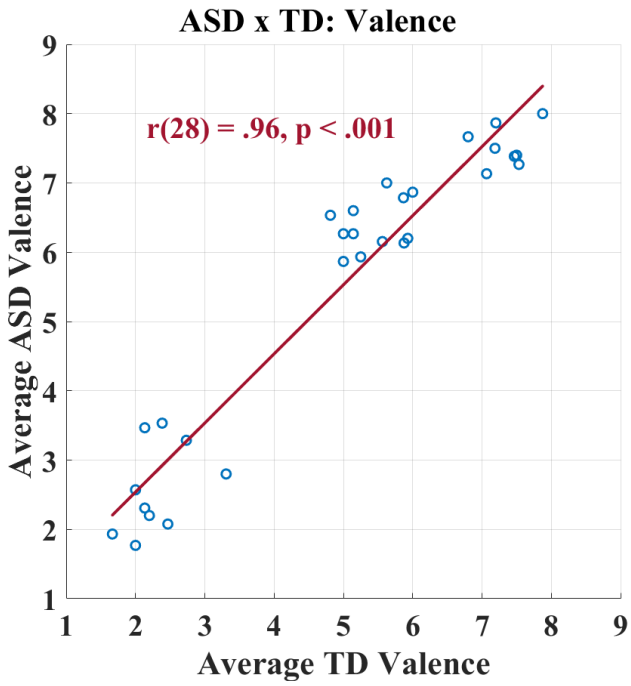

Figure 5.6 - Relationship between the average answers of the clinical and control groups, for the valence dimension. Each blue circle represents one of the videos chosen to integrate the experimental task. Red line represents the first-degree polynomial that is a best fit to the data. r represents the Pearson's linear correlation coefficient, and p represents the statistical significance.

By comparing the mean ratings of the two dimensions of valence and arousal of the three groups, i.e., the CAAV database validation group, the clinical group, and the control group, for the 30 videos selected, three major clusters stand out, as expected (see Fig. 5.7). On the left, fairly separated from the others there is the cluster corresponding to a low valence and high arousal, which encompasses the videos with a negative connotation. On the bottom centre, there is another cluster corresponding to the no valence, low arousal group (an intermediate value of valence is considered neutral, hence the no valence denomination), which includes the videos depicting ordinary everyday actions. Lastly, on the right, slightly above the central cluster, it is possible to identify the third one, with the points corresponding to videos rated higher in both scales, these videos coincide with the HVHA category.

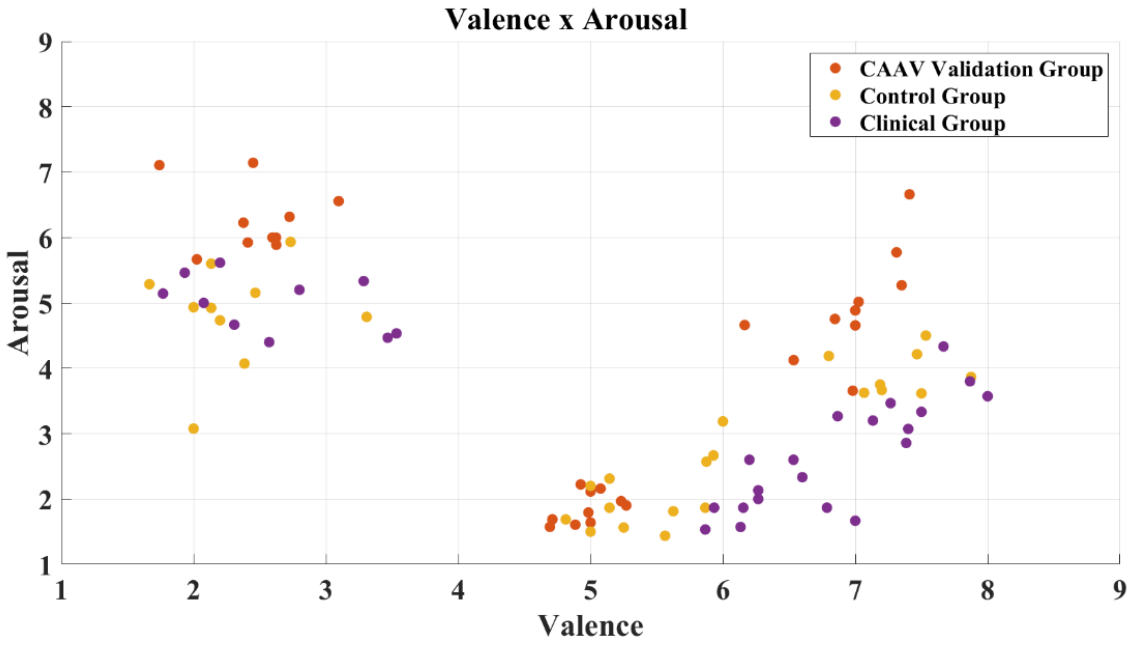

Figure 5.7 - Relationship between the Valence and Arousal ratings for the three groups: CAAV dataset validation group, control group and clinical group. Each coloured dot corresponds to one of the 30 videos selected to integrate the experimental task. On the left is the cluster corresponding to the LVHA category, closer to the bottom and slightly to the right is the cluster corresponding to the NVLA category and, on the far right, sits the cluster that coincides with the HVHA category.

5.2 Signal-to-Noise Ratio Quantification

The values resulting from the SNR calculation of the biosignals acquired in the magnetic resonance environment and outside are presented in table 5.1, alongside the results of an independent samples t-test, to determine if there are statistically significant differences between groups for the biosignals for which there are recordings in the MR environment and outside.

Table 5.1 - Signal-to-Noise ratio results for the biosignals acquired inside the Magnetic Resonance environment. Last three columns display the results for an independent samples t-test where t is the value of the test statistic, df is the degrees of freedom and p is the statistical significance.

Biosignal	MR Environment	Common Environment	Independent Samples T-Test		
	SNR	SNR	t	df	p
Instantaneous Heart Rate	20.96 (8.77)	N/A			
Electrodermal Activity	0.49 (0.96)	1.26 (0.78)	**-2.72**	**24.55**	**0.01**
Electroencephalogram	4.85 (4.22)	N/A			
Photoplethysmogram	3.97 (2.52)	0.80 (0.62)	**6.55**	**37.29**	**<0.01**
Respiration	3.32 (2.92)	34.12 (30.77)	**-3.74**	**13.11**	**<0.01**
Peripheral Oxygen Saturation	208.19 (88.71)	N/A			

For the MR environment, every signal presents a SNR higher than 1, except for EDA, which, for the common environment, presents a SNR just slightly above one. This is probably due more to the quantification method that was applied rather than to high noise contamination given the morphology of this signal, which is subject and circumstance dependent.

For the signals acquired in a common environment, besides EDA, respiration presents a SNR considerably higher than 1, with PPG displaying the worst results.

The SNRs of every signal, for which we had recordings both inside and outside the MR, show statistically significant differences between environments, with higher SNR outside the scanner for EDA and respiration, and higher SNR inside the MR environment for PPG.

5.3 Percent Signal Change

The following tables present the PSC values and the significance levels of the one-sample Wilcoxon Signed Rank tests performed on such values, for each participant individually (Tables 5.3 to 5.5) and for each group (Table 5.6).

Pupil size presents significant changes to baseline, for every participant, for the data acquired simultaneously with fMRI, while it presents a similar behaviour for only 4 out of the 14 participants that completed the task in a common laboratory. This will be discussed in further detail in chapter 6, however, we have reasons to believe that the pronounced changes observed for the MR environment owe their magnitude to existent luminosity variations rather than emotional changes,

while it is safe to believe that the results obtained for the group that performed the task outside the scanner are related to emotion.

Besides pupil size, the physiological signal that presents the most consistent changes is instantaneous heart rate. This is especially true for the ASD group, with only 3 out of the 14 subjects in the group which had instantaneous HR data successfully collected, not presenting significant changes between baseline and events of interest. The changes are less pronounced for the TD group with only a third of the subjects displaying significant results, however, when performing the group analysis, the overall changes in signal between events and the baseline seem to be significant.

Furthermore, respiration and EEG also present prominent changes for a considerable number of subjects in both MR setting groups. However, when performing group statistics, only EEG, for the control group, returned significant results. Reversely, even though only 3 subjects in the ASD group present significant changes in PPG signal, when joining the PSC of the events of the whole group, the result seems to be significant.

EDA presents promising results for the data acquired outside the MR with 75% of the subjects displaying significant results, which also translates into significant intragroup changes.

Overall, only 3 subjects present significant variations for SpO2, hence, it does not seem reasonable to attribute any weight to these results, especially given the lack of literature relating this measure with emotional processes.

Table 5.2 – Average Percent Signal Change values, in percentage, and corresponding standard deviations, in brackets, and p-values for one-sample Wilcoxon Signed Rank Test performed for every biosignal, for each participant in the ASD group. Last row corresponds to the total number of subjects presenting a statistically significant result for each biosignal.

Participant	Respiration		PPG		SpO2		HR		EDA		EEG		Pupil Size	
	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p
A01	0.88 (10.54)	0.50	**-2.59 (9.21)**	**0.03**	0.06 (0.45)	0.84	**-3.66 (3.55)**	**<0.01**	**58.77 (167.59)**	**0.04**	**-1.20 (2.28)**	**0.01**	**-34.21 (17.39)**	**<0.01**
A02	**-12.65 (19.63)**	**<0.01**	-0.02 (6.22)	0.74	-0.16 (0.46)	0.07	**-2.12 (4.37)**	**0.03**	-8.17 (32.80)	0.10	0.02 (0.46)	0.36	**-35.96 (5.87)**	**<0.01**
A03	0.75 (11.94)	0.75	1.21 (4.32)	0.19	0.03 (0.25)	0.28	-0.47 (1.47)	0.19	38.64 (146.30)	0.07	-0.12 (1.79)	0.55	**-5.15 (124.52)**	**<0.01**
A04	**10.82 (21.01)**	**0.01**	-0.13 (2.05)	0.67	0.02 (0.25)	0.79	-0.58 (3.41)	0.25	-0.09 (4.77)	0.98	**0.14 (0.28)**	**0.02**	123.22 (284.21)	0.08
A05	**6.24 (11.79)**	**0.01**	-0.52 (1.55)	0.07	-0.19 (0.62)	0.09	**-1.72 (2.54)**	**<0.01**	193.38 (992.32)	0.56	**0.26 (0.54)**	**0.02**	**-44.51 (8.65)**	**<0.01**
A06	8.53 (24.62)	0.11	-1.45 (4.43)	0.12	**-0.15 (0.31)**	**0.02**	**-4.09 (4.72)**	**<0.01**	10.01 (49.86)	0.60	-0.17 (0.57)	0.07	**-45.13 (3.57)**	**<0.01**
A07	0.47 (11.01)	0.83	**-3.08 (6.01)**	**0.03**	**0.08 (0.33)**	**0.03**	**-3.83 (4.19)**	**<0.01**	**-8.95 (12.46)**	**<0.01**	0.16 (0.80)	0.05	**-45.18 (19.76)**	**<0.01**
A08	-1.31 (10.61)	0.28	**-1.34 (2.49)**	**0.03**	-0.01 (0.04)	0.57	0.32 (1.91)	0.30	-2.20 (14.42)	0.50	**1.55 (3.66)**	**0.01**	**-36.23 (6.72)**	**<0.01**
A10	-1.27 (6.44)	0.47	-0.59 (3.30)	0.50	0.07 (0.43)	0.67	**-2.44 (3.32)**	**<0.01**	12.56 (51.34)	0.67	-0.02 (0.18)	0.43	**-26.31 (13.58)**	**<0.01**
A11	-2.44 (13.57)	0.12	-0.59 (3.55)	0.08	-0.03 (0.15)	0.68	**-1.69 (2.31)**	**<0.01**	-0.22 (9.23)	0.19	-0.03 (0.10)	0.16	**-40.09 (37.56)**	**<0.01**
A12	**-7.25 (13.78)**	**0.01**	0.09 (0.28)	0.10	0.04 (0.46)	0.72	**-1.76 (2.00)**	**<0.01**	**-4.51 (10.35)**	**0.04**	**-0.47 (0.99)**	**0.03**	**-48.43 (5.08)**	**<0.01**
A13	-1.23 (11.25)	0.34	1.29 (3.83)	0.09	0.03 (0.27)	0.77	**-1.30 (1.91)**	**<0.01**	15.54 (73.07)	0.83	**0.95 (3.13)**	**0.02**	**-38.22 (8.48)**	**<0.01**
A14	**15.75 (39.05)**	**0.01**	2.21 (8.60)	0.46	-0.02 (0.50)	0.94	**-3.79 (3.82)**	**<0.01**	91.11 (449.03)	0.89	-0.35 (4.82)	0.94	**-29.59 (71.09)**	**<0.01**
A15	-3.76 (9.79)	0.12	0.38 (2.54)	0.56	-0.06 (0.40)	0.67			0.50 (1.93)	0.37	0.07 (2.18)	0.94	**-28.88 (47.76)**	**<0.01**
A16	1.73 (6.34)	0.26	-0.32 (2.04)	0.39	-0.04 (0.33)	0.16	**-1.34 (2.80)**	**0.03**	0.59 (6.34)	0.67	-0.05 (0.82)	0.72	**11.49 (108.18)**	**0.04**
Total	5		3		2		11		3		6		14	

Table 5.3 - Average Percent Signal Change values, in percentage, and corresponding standard deviations, in brackets, and p-values for one-sample Wilcoxon Signed Rank Test performed for every biosignal, for each participant in the TD group that performed the task in the MR environment. Last row corresponds to the total number of subjects presenting a statistically significant result for each biosignal.

Participant	Respiration		PPG		SpO2		HR		EDA		EEG		Pupil Size	
	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p
C01	11.37 (30.60)	0.14	-2.84 (6.94)	0.12	0.09 (0.42)	0.61	-0.61 (1.48)	0.09	**75.58 (230.87)**	**0.04**	**-2.01 (7.23)**	**0.02**	**-36.35 (7.06)**	**<0.01**
C02	10.48 (29.86)	0.32	-0.73 (3.21)	0.24	0.02 (0.17)	0.98	-1.08 (4.09)	0.11	14.06 (50.96)	0.32	-0.89 (5.63)	0.65	**-36.01 (11.11)**	**<0.01**
C03	**-5.03 (10.06)**	**0.02**	1.69 (4.27)	0.06	-0.01 (0.05)	0.58	**-2.53 (5.28)**	**0.02**	16.62 (61.39)	0.30	-0.40 (2.79)	0.13	**-40.00 (11.59)**	**<0.01**
C04	5.59 (19.94)	0.18	-0.48 (2.26)	0.22	-0.06 (0.21)	0.43	-1.54 (2.01)	0.10	-1.08 (10.69)	0.14	0.14 (0.54)	0.34	**-42.87 (15.18)**	**<0.01**
C05	**9.09 (17.41)**	**0.02**	0.48 (3.34)	0.53	-0.02 (0.35)	0.63	**-1.93 (3.43)**	**0.01**	92.28 (379.99)	0.43	-0.10 (0.63)	0.65	**-48.31 (6.27)**	**<0.01**
C06	**9.56 (16.19)**	**0.01**	-0.38 (2.41)	0.44	-0.12 (0.42)	0.12	-0.31 (2.53)	1.00	0.08 (2.04)	0.98	-0.07 (0.72)	0.68	**-48.86 (5.97)**	**<0.01**
C07	-0.44 (6.93)	0.47	0.11 (4.08)	0.94	0.05 (0.45)	0.67	**-1.39 (3.65)**	**0.03**	0.37 (4.46)	0.79	-0.11 (2.18)	0.77	**-45.90 (7.60)**	**<0.01**
C08	**9.54 (17.06)**	**<0.01**	-0.97 (2.58)	0.08	0.05 (0.15)	0.61	-1.06 (2.70)	0.08	-0.18 (12.91)	0.94	0.27 (2.19)	0.65	**-37.46 (35.88)**	**<0.01**
C09	-2.86 (14.12)	0.26	1.49 (4.09)	0.11	-0.02 (0.43)	0.65	0.20 (2.82)	0.49	0.20 (6.46)	0.19	0.00 (2.34)	0.90	**-44.77 (5.13)**	**<0.01**
C10	**-5.06 (14.41)**	**0.03**	**-0.56 (1.45)**	**0.02**	0.05 (0.28)	0.61	-0.83 (3.72)	0.36	**6.49 (12.66)**	**0.02**	-0.44 (2.46)	0.12	**-48.26 (6.39)**	**<0.01**
C11	0.10 (0.79)	0.55	0.05 (1.34)	0.41	-0.13 (0.61)	0.14	**-1.25 (2.51)**	**0.03**	77.16 (422.99)	0.21	-0.36 (8.13)	0.72	**-21.83 (20.06)**	**<0.01**
C12	3.81 (12.28)	0.14	**1.36 (2.86)**	**0.03**	**-0.05 (0.24)**	**0.03**	-1.13 (3.64)	0.14	10.15 (50.91)	0.35	1.40 (12.73)	0.94	**-44.28 (7.16)**	**<0.01**
C13	**-5.63 (9.09)**	**0.01**	0.48 (1.84)	0.11	-0.07 (0.25)	0.50	-1.23 (3.08)	0.07	1.96 (20.23)	0.37	-0.05 (0.50)	0.36	**-38.44 (9.05)**	**<0.01**
C14	**-12.25 (13.15)**	**<0.01**	0.35 (2.71)	0.61	0.00 (0.41)	0.79			-0.08 (0.92)	0.35	**0.15 (0.31)**	**0.02**	**-48.88 (4.22)**	**<0.01**
C15	-3.43 (33.80)	0.32	0.10 (3.27)	0.72	-0.01 (0.37)	0.96	**-1.85 (3.44)**	**0.01**	-0.45 (16.84)	0.47	**0.29 (0.73)**	**0.04**	**-41.46 (8.85)**	**<0.01**
C16	**-1.07 (14.72)**	**0.01**	-0.18 (6.10)	0.55	0.32 (1.82)	0.91	3.18 (18.73)	0.10	15.46 (33.01)	0.06	**-2.48 (7.31)**	**0.01**	**-39.12 (6.34)**	**<0.01**
Total	8		2		1		5		2		4		16	

Table 5.4 - Average Percent Signal Change values, in percentage, and corresponding standard deviations, in brackets, and p-values for one-sample Wilcoxon Signed Rank Test performed for every biosignal, for each participant in the TD group that performed the task in a common laboratory environment. Last row corresponds to the total number of subjects presenting a statistically significant result for each biosignal.

Participant	Respiration		PPG		EDA		Pupil Size	
	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p
S03	-0.82 (4.06)	0.31	-122.43 (274.15)	0.07	**1.12 (1.59)**	**<0.01**	1.41 (36.38)	0.06
S05	-0.33 (1.64)	0.57	-5.04 (362.90)	0.45	**-12.62 (7.15)**	**<0.01**	**6.37 (25.37)**	**<0.01**
S06	0.83 (3.64)	0.31	-1.66 (102.81)	0.86	**-13.84 (26.09)**	**<0.01**	-2.91 (29.07)	0.09
S07	0.05 (0.78)	0.96	28.28 (321.61)	0.26	**3.08 (24.17)**	**0.04**	-3.60 (15.15)	0.49
S08	0.95 (3.43)	0.21	**-72.35 (196.57)**	**0.02**	11.97 (67.66)	0.28	**12.90 (7.58)**	**<0.01**
S09	-0.01 (1.60)	0.79	-0.23 (104.77)	0.85	**-12.33 (11.56)**	**<0.01**	**7.19 (21.76)**	**0.01**
S11	-0.15 (2.15)	0.53					-7.92 (30.64)	0.61
S12	0.01 (1.04)	0.55	47.31 (158.30)	0.26	**-164.95 (603.00)**	**0.02**	**-13.55 (16.69)**	**<0.01**
S13	-0.20 (9.39)	0.51	1.26 (94.59)	0.94	**-2.75 (19.72)**	**0.01**	1.48 (10.70)	0.34
S14	0.39 (1.78)	0.19	40.18 (483.84)	0.73			-2.44 (25.00)	0.34
S15	0.13 (3.12)	0.88	-47.06 (238.66)	0.31	12.15 (68.26)	0.18	1.44 (8.09)	0.51
S16	0.04 (0.78)	0.61	13.32 (84.58)	0.51	**-4.99 (8.88)**	**<0.01**	1.88 (7.01)	0.26
S17	0.17 (0.46)	0.08	12.96 (51.71)	0.37	0.21 (15.54)	0.47	-4.30 (42.26)	0.37
S18	0.13 (1.49)	0.35	-11.36 (43.64)	0.24	**-3.16 (15.13)**	**<0.01**	5.11 (14.86)	0.17
Total	0		1		9		4	

Table 5.5 - Average Percent Signal Change values, in percentage, and corresponding standard deviations, in brackets, and p-values for one-sample Wilcoxon Signed Rank Test performed for every biosignal, for each of the 3 groups: ASD, TD that performed the task in the MR environment (TD MR) and TD that performed the task in a common laboratory setting (TD Lab).

Group	Respiration		PPG		SpO2		HR		EDA		EEG		Pupil Size	
	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p	PSC (%)	p
ASD	1.06 (18.03)	0.79	**-0.35 (4.86)**	**0.03**	-0.02 (0.38)	0.13	**-2.00 (3.39)**	**<0.01**	26.54 (288.52)	0.20	0.05 (2.09)	0.64	**-21.57 (96.38)**	**<0.01**
TD MR	1.50 (19.30)	0.66	0.00 (3.71)	0.94	0.00 (0.51)	0.25	**-0.96 (5.24)**	**<0.01**	19.29 (156.02)	0.11	**-0.29 (4.97)**	**0.04**	**-41.39 (14.49)**	**<0.01**
TD Lab	0.09 (3.31)	0.47	-9.02 (233.06)	0.18					**-15.84 (183.28)**	**<0.01**			**0.20 (23.80)**	**<0.01**

5.4 Statistical Analysis

For the video runs, differences in the distribution of each feature between High and Low Arousal conditions, and between each valence condition: High Valence vs. Neutral Valence, High Valence vs. Low Valence and Neutral Valence vs. Low Valence, were assessed. Moreover, for each condition, differences between groups were also investigated. For the BART runs, the features were inspected for differences between the events of winning versus losing and winning a social reward versus winning a reward based on personal interest. Differences between groups for each event were also assessed.

The significance levels that resulted from this analysis are present in Tables 5.6 to 5.14. To simplify and facilitate the reader's interpretation, only the features that suggest statistically significant differences for at least one scenario are shown in each table. Nonetheless, the tables with the complete results obtained with the statistical analysis can be consulted in Appendix I.

Table 5.6 - p-values of pairwise comparisons from Wilcoxon Signed Rank Test, performed on the dataset acquired in the MR (High Arousal compared to Low Arousal).

Feature	Database		Self-Assessment	
	Clinical (N = 15)	**Control (N = 16)**	**Clinical (N = 15)**	**Control (N = 16)**
decTimeEDA	**0.05**	0.68	0.56	0.68
meanPPG	0.60	0.18	**0.02**	0.35
maxPPG	0.23	0.41	0.52	**0.04**
minPPG	0.25	**0.04**	0.52	**0.01**
stdPPG	**0.03**	0.33	0.45	**0.01**
VLF	0.19	0.41	**0.03**	0.72
HF	0.17	0.76	**0.05**	0.68
RaLH	0.08	0.57	**0.05**	0.88
RaLVL	0.21	0.11	**0.04**	0.88
meanPeakAmp	**0.02**	0.64	0.89	0.06
stdPeakAmp	0.68	0.44	0.60	**0.02**
alpha	**0.04**	0.88	0.49	0.84
stdRespAmp	0.76	0.21	**0.04**	0.33
stdResp	0.17	0.41	0.21	**0.04**
nBlinks	0.21	**0.03**	**0.02**	0.07
meanHR	**0.02**	1.00	0.10	0.39
maxHR	**0.02**	0.80	0.30	0.64
minHR	**0.01**	0.98	**0.01**	0.89

Table 5.7 - p-values of pairwise comparisons from Wilcoxon Rank Sum Test, for the arousal conditions (Clinical Group compared to Control Group).

Feature	Database		Self-Assessment	
	High Arousal	Low Arousal	High Arousal	Low Arousal
decTimeEDA	0.35	**<0.01**	0.29	**0.02**
pNN50	0.08	**0.02**	**0.04**	0.06
delta	**0.01**	**<0.01**	**0.01**	**0.01**
alpha	0.06	**0.05**	0.07	**0.02**
beta	**0.02**	**<0.01**	**0.01**	**0.01**
stdResp	**0.03**	0.83	0.09	0.89
meanHR	**0.02**	0.91	**0.02**	0.08
maxHR	**0.01**	0.56	**0.03**	0.33
minHR	**0.03**	0.84	**0.02**	0.20

Table 5.8 - p-values of pairwise comparisons between the different valence conditions (High Valence, No Valence, Low Valence) from Wilcoxon Signed Rank Test, for the ASD group.

Feature	Database			Self - Assessment		
	HV vs. NV	HV vs. LV	NV vs. LV	HV vs. NV	HV vs. LV	NV vs. LV
decTimeEDA	0.23	0.52	**0.03**	0.49	0.30	0.15
meanPPG	0.85	0.30	0.36	**<0.01**	**0.01**	0.23
maxPPG	**0.01**	0.17	0.39	0.06	0.09	0.49
minPPG	**0.04**	0.72	0.30	**<0.01**	0.36	0.15
stdPPG	0.11	0.98	0.09	**0.02**	0.80	0.11
VLF	1.00	**0.03**	**0.01**	0.85	0.11	0.07
LF	0.85	**<0.01**	**<0.01**	0.64	0.06	**0.05**
HF	0.80	**<0.01**	**<0.01**	0.76	0.06	**0.04**
RaLH	0.52	**<0.01**	**<0.01**	0.89	0.08	**<0.01**
RaLVL	0.52	0.42	**0.02**	0.93	0.42	0.85
RaHVL	0.98	0.21	**0.04**	0.49	0.19	0.39
meanPeakAmp	**0.04**	0.89	0.06	**0.03**	0.60	0.06
stdPeakAmp	**0.01**	0.25	0.60	0.25	0.12	0.80
delta	0.06	0.45	0.12	**0.02**	0.17	0.19
alpha	**0.03**	0.25	0.28	**0.03**	0.85	**0.04**
beta	**0.02**	0.07	0.39	**<0.01**	0.19	0.30
gamma	0.93	0.52	**0.04**	0.21	0.39	0.36
maxResp	**0.02**	0.28	0.68	0.98	0.52	0.76
minResp	0.21	0.39	0.93	**0.03**	0.52	0.06
stdRespAmp	0.15	**0.03**	0.25	0.15	0.21	0.89
meanRiseDurResp	0.85	0.06	0.25	0.76	**0.02**	**0.01**
SDRiseDurResp	0.64	0.09	**0.04**	0.56	**0.02**	**<0.01**
maxRespInter	0.28	**0.01**	0.19	0.89	0.49	0.39
stdRespInter	0.98	0.06	0.19	0.89	**0.05**	0.12
DRResp	**0.01**	0.21	0.89	0.19	0.60	0.60
nBlinks	0.93	0.24	0.19	**0.02**	0.30	0.68
meanHR	**0.01**	0.36	**<0.01**	0.58	**0.01**	0.33
maxHR	**0.01**	0.81	**0.02**	0.09	**0.03**	0.71
minHR	**<0.01**	0.50	**<0.01**	1.00	**0.01**	0.24
stdHR	0.36	**0.05**	**0.02**	0.36	0.33	0.15

Table 5.9 - p-values of pairwise comparisons between the different valence conditions (High Valence, No Valence, Low Valence) from Wilcoxon Signed Rank Test, for the TD group acquired in the MR.

Feature	Database			Self - Assessment		
	HV vs. NV	HV vs. LV	NV vs. LV	HV vs. NV	HV vs. LV	NV vs. LV
meanAbsFdEDA	0.21	**0.04**	**0.01**	0.88	**0.04**	**0.01**
meanAbsSdEDA	0.21	**0.04**	**0.01**	0.88	**0.04**	**0.01**
meanDerivNegEDA	0.76	0.06	**0.03**	0.68	**0.03**	0.09
rateSCR	0.60	0.20	**0.04**	0.33	0.68	0.16
SDRiseDurSCR	**0.04**	**0.05**	0.90	0.10	0.13	0.63
maxPPG	0.84	**0.02**	0.08	0.84	0.21	0.07
minPPG	0.21	0.26	**0.05**	0.18	0.13	0.11
NNmean	0.72	0.13	0.07	0.76	0.21	**0.05**
NNmin	0.28	**0.01**	0.35	0.57	0.09	**0.03**
SDNN	**0.04**	0.20	0.57	0.07	0.08	0.53
SDSD	0.10	0.06	0.76	**0.05**	0.09	0.21
RMSSD	**0.05**	0.07	0.61	0.06	0.10	0.47
pNN50	**0.05**	0.09	0.35	0.10	**0.03**	0.72
meanPeakAmp	0.76	0.07	0.33	0.47	0.38	**0.04**
meanDeltaNN	0.44	0.61	0.61	0.23	**0.03**	0.76
maxResp	0.38	0.44	0.80	**0.01**	0.18	0.84
meanAbsFdResp	0.20	0.28	0.92	**0.04**	**0.03**	0.64
meanAbsSdResp	0.16	0.33	0.96	**0.04**	**0.03**	0.72
stdRespAmp	**0.04**	0.53	**0.02**	0.15	0.53	0.21
DRResp	0.84	0.53	0.53	**0.01**	0.26	0.33
nBlinks	**0.01**	0.16	0.23	0.50	0.66	0.64
meanAbsFdHR	0.64	**0.02**	**0.05**	1.00	0.07	0.14
minSpO2	0.15	**0.05**	0.15	**0.02**	**0.02**	0.84

Table 5.10 - p-values of pairwise comparisons from Wilcoxon Rank Sum Test, for the valence conditions (Clinical Group compared to Control Group).

Feature	Database			Self - Assessment		
	High Valence	Neutral Valence	Low Valence	High Valence	Neutral Valence	Low Valence
decTimeEDA	0.13	**0.02**	0.86	0.08	**0.02**	0.92
SDRiseDurSCR	0.12	0.77	0.45	**0.04**	0.74	0.35
minPPG	0.80	**0.02**	0.59	0.89	0.06	0.54
stdPPG	0.71	0.17	0.19	0.92	**0.02**	0.23
pNN50	0.92	**<0.01**	0.16	0.95	0.09	**0.03**
meanPeakAmp	0.46	0.16	0.21	0.74	**0.01**	0.26
delta	0.10	**<0.01**	**0.01**	0.09	**0.01**	**0.01**
alpha	**0.05**	**0.03**	0.07	**0.04**	**0.02**	0.09
beta	0.10	**<0.01**	**<0.01**	0.14	**<0.01**	**<0.01**
meanAbsFdResp	**0.01**	0.57	0.62	**0.02**	0.74	0.92
meanAbsSdResp	**0.02**	0.59	0.62	**0.02**	0.74	0.89
stdResp	**0.04**	0.68	0.23	0.08	0.57	0.12
nBlinks	0.58	0.18	0.15	0.89	**0.01**	0.28
meanHR	**0.02**	0.62	**0.03**	0.17	0.23	**0.05**
maxHR	0.12	0.90	**0.03**	0.74	0.39	**0.03**
minHR	**0.02**	0.71	0.10	0.09	0.14	0.08

Table 5.11 - - p-values of pairwise comparisons from Wilcoxon Signed Rank Test, performed on the dataset acquired in a common environment (High Arousal compared to Low Arousal).

Feature	Database	Self-Assessment
decTimeEDA	**0.03**	0.23
meanAmpSCR	**<0.01**	0.42
maxAmpSCR	**<0.01**	0.05
rateSCR	**0.02**	0.06
meanRiseDurSCR	**<0.01**	**0.03**
meanRiseDurResp	**0.04**	0.42
maxPupilSize	0.27	**0.01**

Table 5.12 - p-values of pairwise comparisons between the different valence conditions (High Valence, No Valence, Low Valence) from Wilcoxon Signed Rank Test, for the TD group acquired in a common environment.

Feature	Database			Self - Assessment		
	HV vs. NV	HV vs. LV	NV vs. LV	HV vs. NV	HV vs. LV	NV vs. LV
nFallsEDA	**0.03**	0.34	0.18	0.35	0.83	0.33
maxAmpSCR	**0.04**	0.34	**0.01**	0.27	0.79	0.20
rateSCR	0.05	0.83	**0.02**	0.17	0.24	0.16
meanRiseDurSCR	0.06	0.62	0.05	**0.01**	0.34	**0.01**
NNmean	0.74	0.15	0.19	0.15	**0.03**	0.27
SDNN	**0.04**	0.38	0.89	0.22	0.41	0.89
minResp	0.08	0.27	0.63	**0.04**	0.10	0.86
meanAbsFdResp	0.08	0.07	0.76	0.58	0.14	**0.05**
meanAbsSdResp	0.08	0.07	0.76	0.58	0.14	**0.05**
respRate	0.23	0.18	0.83	0.56	**0.03**	0.90
minRespInter	0.94	1.00	1.00	0.11	0.63	**0.05**
stdResp	**0.02**	**0.04**	0.63	**0.02**	**0.02**	0.33
DRResp	0.15	0.08	0.39	0.17	**0.02**	0.33
meanPupilSize	0.30	**0.05**	0.39	0.08	0.14	0.90
maxPupilSize	0.43	0.06	0.71	**0.05**	**0.02**	0.90

Table 5.13 - p-values of pairwise comparisons from Wilcoxon Signed Rank Test. Second and third columns refer to the comparison of the winning against the losing event, fourth and fifth columns refer to the comparison between the social and interest rewards.

Feature	Winning vs. Losing		Social vs. Interest Reward	
	Clinical (N = 15)	Control (N = 16)	Clinical (N = 15)	Control (N = 16)
meanEDA	**<0.01**	0.09	0.12	0.76
minEDA	**0.02**	0.21	0.36	0.68
nFallsEDA	0.25	**0.04**	**0.01**	0.33
stdPPG	0.21	0.23	**0.05**	0.64
VLF	0.39	**0.04**	0.98	0.88
LF	0.39	**0.04**	1.00	0.96
HF	0.89	**0.04**	0.42	0.96
RaLH	0.56	**0.03**	0.98	0.88
NNmax	**0.02**	0.76	0.30	0.68
SDNN	**0.02**	0.20	0.23	0.50
SDSD	**0.04**	0.28	0.19	0.53
RMSSD	**0.03**	0.26	0.19	0.53
meanPeakAmp	0.15	0.41	**0.03**	0.61
gamma	**0.02**	0.92	0.21	0.26
meanResp	0.49	0.80	**0.05**	0.16
meanRespInter	**0.02**	0.20	0.21	0.64
medianRespInter	**0.03**	0.50	0.25	0.68
stdRespInter	0.64	0.35	0.76	**0.05**
maxPupilSize	0.15	**0.02**	0.68	0.30
meanHR	**0.02**	0.39	0.11	0.45
maxHR	**0.04**	0.46	0.24	0.58
meanAbsFdHR	0.27	0.19	0.09	**0.05**

Table 5.14 - p-values of pairwise comparisons from Wilcoxon Rank Sum Test (Clinical Group compared to Control Group).

Feature	Winning	Losing	Social Reward	Interest Reward
SDRiseDurSCR	0.67	**0.01**	0.78	0.98
minPPG	**<0.01**	0.89	**0.03**	0.29
stdPPG	**0.02**	1.00	**0.03**	0.46
VLF	**0.01**	0.57	**0.03**	**0.04**
LF	**0.01**	0.49	**0.03**	**0.03**
HF	**0.04**	0.49	0.12	**0.03**
RaLH	**0.01**	0.57	**0.03**	**0.03**
RaLVL	0.14	0.51	0.11	**0.04**
RaHVL	**0.05**	0.46	0.07	0.05
NNmax	0.08	0.49	**0.04**	0.35
SDNN	**0.04**	0.44	**0.02**	0.28
SDSD	**0.04**	0.46	0.13	0.33
meanPeakAmp	**0.01**	0.89	**0.03**	0.59
stdPeakAmp	**0.01**	0.86	0.26	0.08
gamma	0.49	**0.04**	0.29	0.77
maxResp	**0.03**	0.19	0.06	0.09
respRate	**0.03**	**0.01**	0.12	**0.03**
meanRespInter	0.74	**0.02**	0.98	0.66
medianRespInter	0.57	**0.01**	0.95	0.75
minRespInter	0.29	**<0.01**	0.95	0.27
stdRespInter	**0.03**	0.59	0.65	**<0.01**
minSpO2	0.49	0.20	**0.04**	0.79

The different statistical analysis performed on the extracted features revealed their overall lack of discriminative power. There are, however, some cases that stand out.

The features exhibiting the most consistent results are the ones extracted from the instantaneous HR signal, particularly meanHR, maxHR and minHR. These features consistently present statistically significant differences between conditions, for the ASD group, namely between HA and LA (Table 5.6), HV and NV, NV and LV (Table 5.8), and between Winning and Losing events (Table 5.13). Also, for the same features and the mentioned conditions, we also found differences between groups. It is important to point out that these differences are present when considering the database mean ratings to label the events, but not when considering the self-assessment answers. The results of the statistical analysis performed between valence conditions, for the ASD group (Table 5.8) also revealed an interesting phenomenon pertaining the HR features. While significant differences are present for these features between HV and NV, and between NV and LV, when considering the database ratings, these differences disappear for those conditions and are instead found between HV and LV conditions, when considering the self-assessment answers to label the events.

When considering the between group analysis for the different conditions (Tables 5.7 and 5.10), the features corresponding to the relative power in the delta and beta bands of the EEG presented statistically significant differences for nearly all cases, with alpha also presenting significant differences for most cases.

For the signals acquired in a common environment, the features derived from the SCRs of the EDA signal all present statistically significant differences between the high and low arousal conditions, with the exception of SDRiseDurSCR, being the only analysis for which this group of features presents consistent differences (Table 5.11).

Another feature group that deserves attention is the frequency domain features of the PRV. These features present statistically significant differences between valence conditions, for the ASD group, particularly for HV vs. LV and NV vs LV, when considering the database ratings to label the events (Table 5.8). Also, for the control group, these features show significantly different values between winning and losing events. Finally, when analysing differences between groups for the BART events of interest, these features also returned significant results for the winning events (Tables 5.13 and 5.14).

According to the PSC results for the Pupil Size measurements obtained with the task conducted outside the MR scanner, we expected to obtain some significant differences between conditions for the features derived from this signal. However, only the maxPupilSize feature presents some significant results, when considering the self-assessment answers, for HA vs. LA, HV vs. NV and HV vs. LV (Tables 5.11 and 5.12).

5.5 Classification

Lastly, the features extracted from the biosignals were used to train 4 different classifiers, a Minimum Distance Classifier, a KNN, and SVM with linear and Radial Basis Function kernels, for the different classification problems mentioned in section 4.10 (see table 4.11).

First, the classifiers were trained and tested with the original features, and later, in an attempt to improve performance, the process was repeated with the extra step of performing PCA on the dataset to reduce its dimensionality.

The performance of the classifiers turned out to be suboptimal for all cases, even when using PCA as a feature reduction algorithm, with no classifier performing better than the other three. For this reason, and to simplify, we chose to present, for illustration purposes, only the outcome of the SVM with a linear kernel. SVM with a linear kernel was chosen for this purpose, for no reason except its popularity. Also, in this section, only the results of analysis A1, A3, A5 and A6 are presented. The remaining results obtained with SVM Linear can be consulted in Appendix II, including the outcomes obtained when performing PCA before classification, and will be referenced here, when relevant.

5.5.1 Video Observations

5.5.1.1 Analysis A1: High Arousal vs. Low Arousal

Results for the HA vs LA classification problem, when considering database ratings as ground truth, are presented next. Results for the dataset acquired in the MR environment are presented in Figures 5.8 and table 5.6. Figure 5.9 and Table 5.7 display the results for the dataset acquired in a common environment.

Considering the database mean ratings as ground truth to label the observations as either low or high arousal resulted in imbalanced classes since 20 of the chosen videos were considered as HA and only 10 as LA (see section 4.2). This helps to explain some of the outcomes, namely for the intersubject modalities with SVM Linear, KNN, and SVM RBF, where the classifier is simply assigning observations in the test set to the most frequent class. This behaviour is observed in both datasets, acquired in the MR and in a common lab, and did not change when performing PCA on the feature space (please refer to Figures 1 and 2 and Tables 10 and 11 of Appendix II).

Changing the ground truth from the database average ratings to the self-assessment answers of each participant did not improve the performance of the classifiers, with all algorithms presenting outcomes coincident with chance level (see Figures 3 and 4 and Tables 12 and 13).

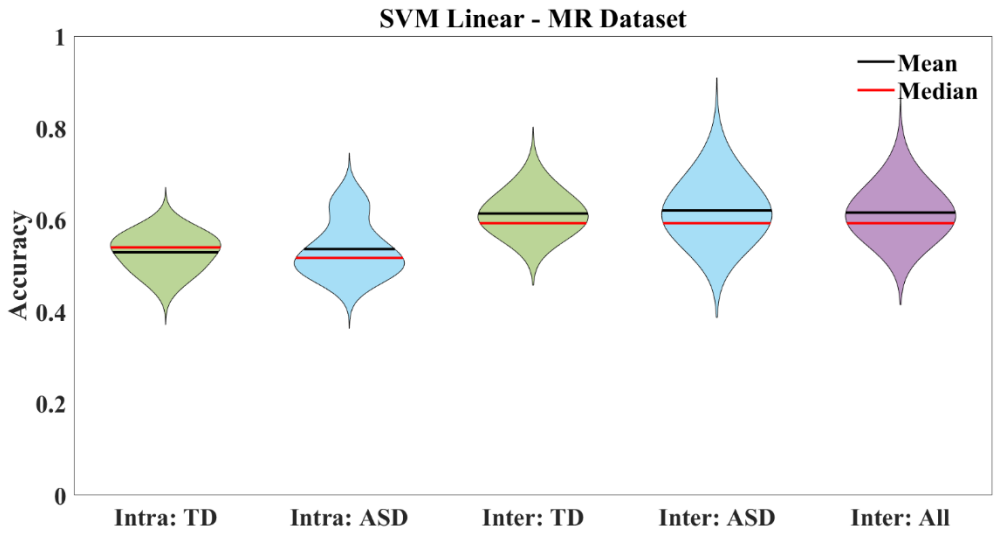

Figure 5.8 - Distribution of accuracies achieved, on the dataset acquired in the MR environment, by SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the CAAV database ratings as ground truth.

Table 5.15 - Performance metrics mean values and respective standard deviations, in brackets, achieved, on the dataset acquired in the MR environment, by SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the CAAV database ratings as ground truth. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject: TD	Intrasubject: ASD	Intersubject: TD	Intersubject: ASD	Intersubject: All
Accuracy	0.53 (0.04)	0.54 (0.06)	0.61 (0.04)	0.62 (0.04)	0.62 (0.04)
Sensitivity	0.78 (0.08)	0.77 (0.11)	1.00 (0.00)	0.99 (0.05)	0.99 (0.03)
Specificity	0.21 (0.13)	0.20 (0.14)	0.00 (0.00)	0.01 (0.05)	0.00 (0.02)
Precision	0.61 (0.05)	0.62 (0.06)	0.61 (0.03)	0.62 (0.04)	0.62 (0.04)
F-Measure	0.65 (0.04)	0.66 (0.06)	0.76 (0.02)	0.76 (0.03)	0.76 (0.03)
p-value	0.10	0.09	0.16	0.16	0.16

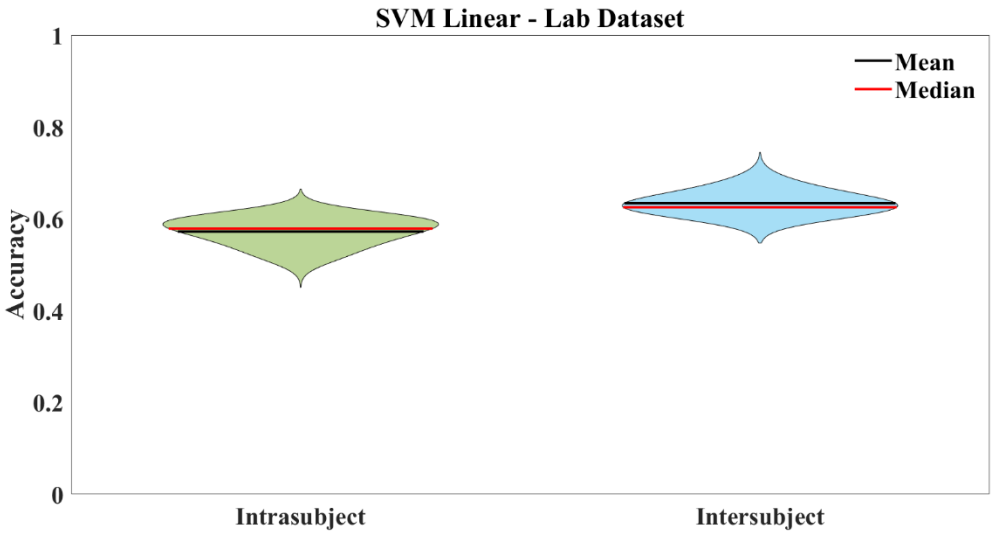

Figure 5.9 - Distribution of accuracies achieved, on the dataset acquired in a common environment, by SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the CAAV database ratings as ground truth.

Table 5.16 - Performance metrics mean values and respective standard deviations, in brackets, achieved, on the dataset acquired in a common environment, by SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the CAAV database ratings as ground truth. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject	Intersubject
Accuracy	0.57 (0.03)	0.63 (0.02)
Sensitivity	0.85 (0.05)	1.00 (0.00)
Specificity	0.17 (0.12)	0.00 (0.00)
Precision	0.64 (0.04)	0.63 (0.02)
F-Measure	0.70 (0.02)	0.78 (0.01)
p-value	0.10	0.14

5.5.1.2 Analysis A3: High Valence, Neutral Valence, and Low Valence

When attempting automatic classification of emotional valence on the video observations, considering the database ratings as ground truth, the outcomes obtained for the intrasubject modality were not significantly higher than random chance, which is, for this classification problem, approximately 33%. And, while accuracy and macro f1-score seem to rise a bit while the p-value lowers for the intersubject modality, possibly explained by the bigger number of observations which improves the generalization capacity of the model, the p-value is still not small enough to affirm that the outcome is not just the result of chance.

Adding PCA to the pipeline results in similar outcomes, and considering the self-assessment answers as ground truth, worsens the results (please see Figures 5 to 8 and Tables 14 to 17). Figure 5.10 and table 5.8 bear the classification outcomes for this analysis, for the dataset acquired simultaneously with fMRI and figure 5.11 and table 5.9 are in respect to the dataset acquired in a common environment.

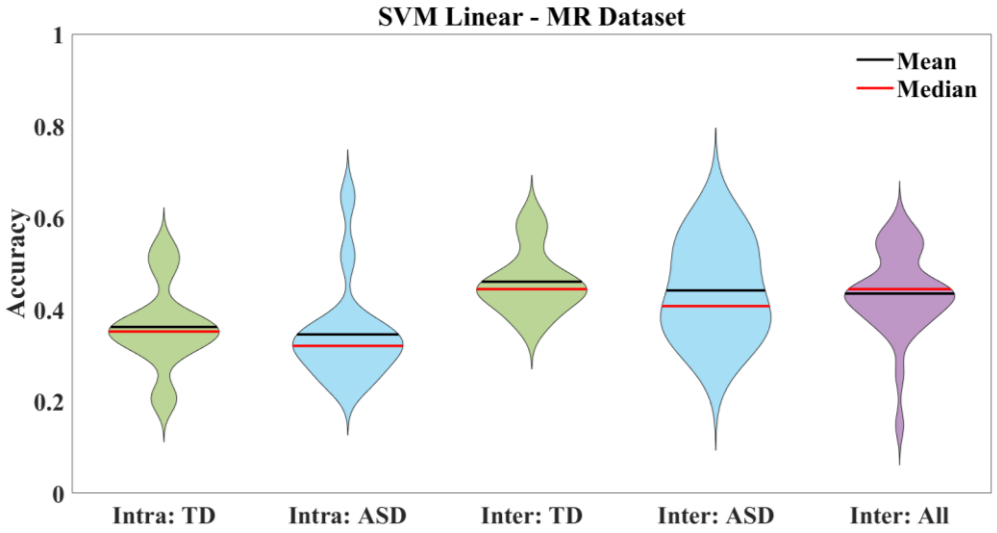

Figure 5.10 - Distribution of accuracies achieved, on the dataset acquired in the MR environment, by SVM with a linear kernel on classifying HV, NV and LV on intra and inter subject modalities using the CAAV database ratings as ground truth.

Table 5.17 - Performance metrics mean values and respective standard deviations, in brackets, achieved, on the dataset acquired in a common environment, by SVM with a linear kernel on classifying High Valence, Neutral Valence and Low Valence, on intra and inter subject modalities using the database mean ratings as ground truth. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject: TD	Intrasubject: ASD	Intersubject: TD	Intersubject: ASD	Intersubject: All
Accuracy	0.36 (0.09)	0.35 (0.11)	0.46 (0.07)	0.44 (0.10)	0.43 (0.09)
Macro F1 - Score	0.35 (0.05)	0.34 (0.09)	0.39 (0.05)	0.39 (0.1)	0.37 (0.09)
p-value	0.47	0.50	0.12	0.20	0.19

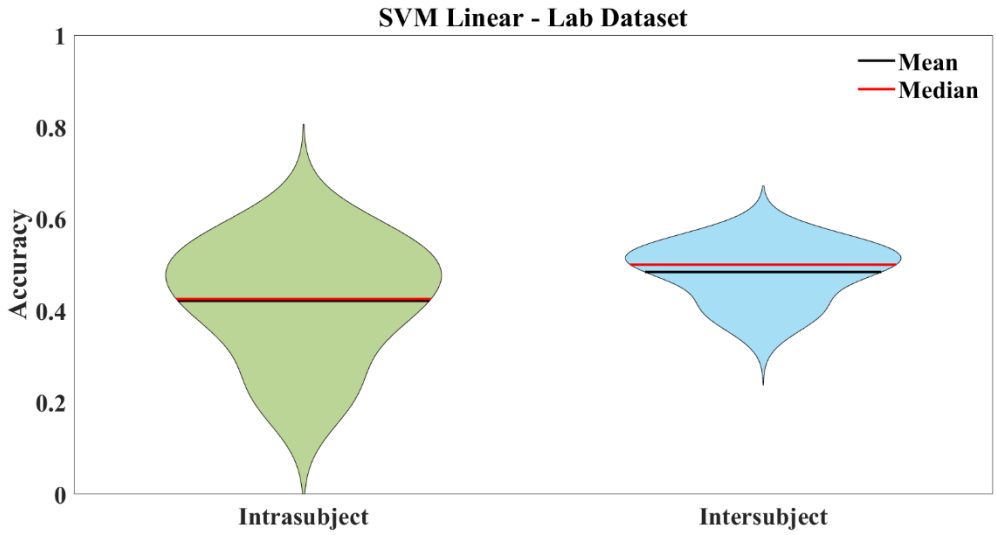

Figure 5.11 - Distribution of accuracies achieved by SVM with a linear kernel on classifying HV, NV and LV on intra and inter subject modalities using the CAAV database ratings as ground truth on the dataset acquired in a common environment.

Table 5.18 - Performance metrics mean values and respective standard deviations, in brackets, achieved by SVM with a linear kernel on classifying High Valence, Neutral Valence and Low Valence, on intra and inter subject modalities using the CAAV database ratings as ground truth on the dataset acquired in a common environment. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject	**Intersubject**
Accuracy	0.42 (0.12)	0.48 (0.06)
Macro F1 - Score	0.37 (0.08)	0.38 (0.05)
p-value	0.42	0.13

5.5.2 BART runs

5.5.2.1 Analysis A5: Winning vs. Losing

Results for the classification problem targeting the BART winning/losing outcomes (Fig. 5.12 and table 5.10) yielded unsatisfactory results, close to chance level, not improving with PCA for dimensionality reduction (see Figure 9 and Table 18).

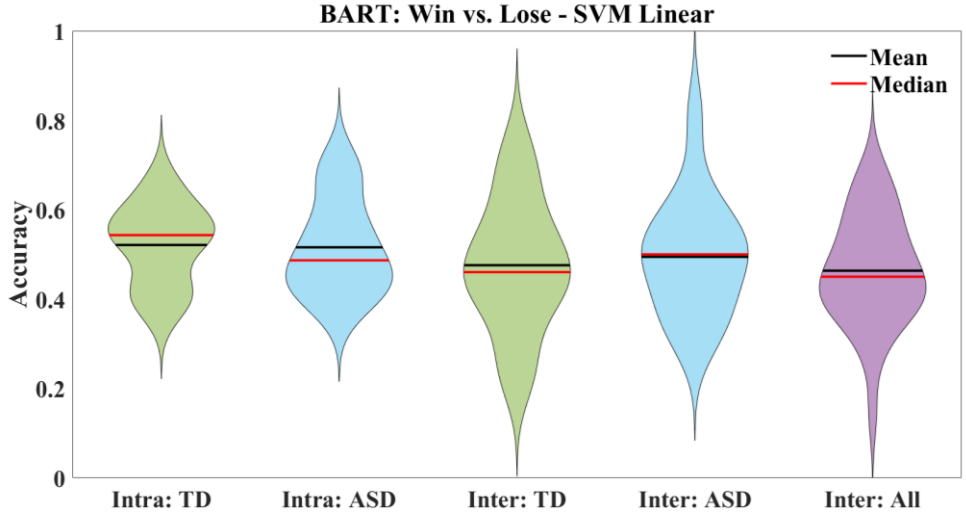

Figure 5.12 - Distribution of accuracies achieved by SVM with a linear kernel on classifying Winning vs Losing events on intra and inter subject modalities.

Table 5.19 - Performance metrics mean values and respective standard deviations, in brackets, achieved by SVM with a linear kernel on classifying Winning vs Losing events on intra and inter subject modalities. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject: TD	Intrasubject: ASD	Intersubject: TD	Intersubject: ASD	Intersubject: All
Accuracy	0.52 (0.09)	0.52 (0.11)	0.48 (0.15)	0.49 (0.13)	0.46 (0.12)
Sensitivity	0.48 (0.29)	0.55 (0.22)	0.38 (0.29)	0.35 (0.36)	0.18 (0.27)
Specificity	0.51 (0.27)	0.47 (0.26)	0.59 (0.27)	0.70 (0.32)	0.78 (0.27)
Precision	0.44 (0.22)	0.48 (0.12)	0.43 (0.25)	0.60 (0.28)	0.40 (0.16)
F-Measure	0.57 (0.14)	0.60 (0.08)	0.49 (0.16)	0.38 (0.21)	0.35 (0.14)
p-value	0.16	0.14	0.55	0.52	0.57

5.5.2.2 Analysis A6: Social Reward vs. Interest Reward

Figure 5.13 and Table 5.11 present the classification outcomes for analysis A6. Similarly to analysis A5, the classification of reward type was unsuccessful with accuracy values around the chance level (50%), moreover, when adding PCA to the process, there were no improvements to the outcome (refer to Figure 10 and Table 19).

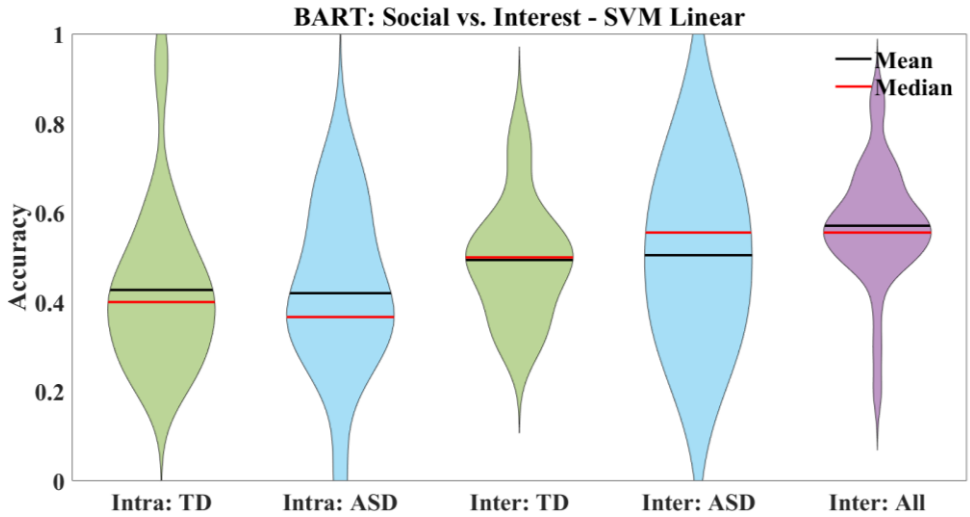

Figure 5.13 - Distribution of accuracies achieved by SVM with a linear kernel on classifying Social vs Interest reward on intra and inter subject modalities.

Table 5.20 - Performance metrics mean values and respective standard deviations, in brackets, achieved by SVM with a linear kernel on classifying Social vs Interest reward on intra and inter subject modalities. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject: TD	Intrasubject: ASD	Intersubject: TD	Intersubject: ASD	Intersubject: All
Accuracy	0.43 (0.18)	0.42 (0.18)	0.49 (0.13)	0.50 (0.18)	0.57 (0.13)
Sensitivity	0.42 (0.25)	0.48 (0.25)	0.28 (0.34)	0.54 (0.31)	0.45 (0.31)
Specificity	0.53 (0.21)	0.46 (0.22)	0.71 (0.27)	0.48 (0.28)	0.69 (0.24)
Precision	0.44 (0.24)	0.48 (0.25)	0.41 (0.27)	0.49 (0.17)	0.59 (0.22)
F-Measure	0.73 (0.14)	0.72 (0.15)	0.52 (0.18)	0.51 (0.19)	0.53 (0.18)
p-value	0.36	0.29	0.47	0.37	0.26

6 Discussion

6.1 Behavioural Analysis

The behavioural analysis revealed that our participants' answers, regardless of their group, correlated strongly with the answers of the CAAV database validation group, validating the use of these stimuli in our groups. Moreover, the answers of our two groups also showed a strong correlation between them. There is, however, a tendency in the ASD group to rate the videos on either extreme of the valence scale. This suggests the existence of some differences in the perception and reporting of emotion, which has been extensively evidenced and documented for ASD (Ben Shalom *et al.*, 2006; Huggins *et al.*, 2021). For arousal, however, there is a notable tendency, in both our groups, to give lower arousal ratings to the videos, which evidences emotion perception differences, when compared to the CAAV database validation group, and indicates that the stimuli may not be having the expected impact on the participants.

6.2 Signal-to-Noise Ratio Quantification

The signal-to-noise ratio results revealed that, in general, the signals acquired in the MR environment present satisfactory SNRs. The EDA signal, however, presents suboptimal SNRs for both settings. A possible explanation is that this quantification of SNR is obtained by averaging over the signal values of every run and dividing it by the standard deviation, for each participant. This process overlooks the variability of the signal between runs, which is particularly high for signals that do not translate an intrinsic rhythmic process, as is the case of EDA, which is the result of sudomotor activity, a function that does not present any particular pattern but is triggered by occasional external and internal stressors, that are different for every run. This translates into biosignals that are very different across runs, which results in a large variance and low SNRs.

The respiration signal displays considerably higher SNR outside of the scanner than in it. The lower value for the MR environment was somewhat expected, taken the added source of noise that is the scanner and its gradient switch, which artifact can be caught in the biosignal recordings. For PPG, however, contrarily to expected, the signals acquired in the MR environment display higher SNR values. This may be explained by considering that first: the sampling rate of the signal outside the MR scanner is approximately 12 times higher than the sampling frequency of the physiological measurement unit of the scanner, which is used to record the PPG during fMRI acquisitions, hence it is prone to capturing an added amount of noise when compared to the lower sampled signal, and second, participants are lying down when they are performing the fMRI task as opposed to sitting when in the common laboratory setting, which can help prevent undesired movement. Moreover, the pulse oximeters used for each environment were different, which results in recordings with differing quality and characteristics.

6.3 Percent Signal Change

From looking at the PSC results we notice a considerable and evident change in pupil size from the baseline to the events of interest, for all cases, but mostly for the signals acquired in the MR environment. This is due to the big difference in luminosity existent between the fixation cross's background of the stimulus used in the MR task, which was a black colour, and the generality of the videos. The muscles of the iris respond to these light changes with greater strength than they do to changes in emotional state or cognition, hence, the changes found in the signals acquired in the MR environment cannot be attributed to an emotional response. After some acquisitions, we became aware that the high contrast between the black screen of the rest period and the videos was causing a big fluctuation in pupil size, unrelated to emotional state and, even though for consistency, the MR task stimulus was not changed, the background of the fixation cross for the stimulus used outside the MR was altered to a light grey colour. This, allied to the fact that pupil size changes are not so pronounced in this environment, even though they seem to be significant for the most part, indicates that this change may be related to an emotional response, which represents a promising result.

The instantaneous heart rate presents significant changes for a large number of subjects, with the most pronounced results for the ASD group. This is in agreement with the statistical analysis performed on the extracted features, which also evidenced significant differences between conditions for the ASD group, for the features derived from this signal. These results match the classical association of cardiac activity with emotional changes (Winton, Putnam and Krauss, 1984; Lang *et al.*, 1993; Legrand *et al.*, 2021), and indicate that, to some extent, the experimental task is indeed inducing physiological variations in the participants. Surprisingly, the PPG signal, also strongly related to cardiac function, did not present similar results. Only 6 out of the 45 individuals that completed the task presented significant changes in this signal between baseline and events of interest, even though the group analysis for the ASD group indicates the overall intragroup changes to be significant. Since PPG is an indirect measure of cardiac function, fluctuations in heart activity related to emotional responses may have a subtler representation in PPG than in the direct instantaneous HR monitorization. Hence, the fact that this group is also the one with the most pronounced HR changes could explain why intragroup PPG changes are significant for ASD but not for TD individuals.

While the intragroup analysis revealed that respiration does not yield any significant variation to baseline, 5 subjects in the ASD group and 8 in the TD group show significant changes for this biosignal, all of them completed the task in the MR scanner. Even though the particularity of each individual's physiological response must always be accounted for when interpreting these results, we should also keep in mind that completing the task inside an MR scanner or in a relatively normal room bears very different conditions for the subjects. Being inside the scanner with a head coil,

several electrodes, and wires connected to the body, with the added constraint of laying as still as possible for approximately 50 minutes is a very different experience to sitting in a chair, in a normal room, surely also with electrodes connected to the body, which were actually wireless. The point is that the MR setting certainly represents a much more uncomfortable experience for the participants than the data acquisitions carried out in a common laboratory. Hence, some physiological responses can be attributed to the uncomfortable setting and not necessarily to the experimental task stimuli. This can help explain the non-existence of changes in respiration signal for the subjects that completed the task outside the scanner.

EEG is considered by many the most important biosignal for emotion recognition. Being a direct measure of the electrical activity of the brain, where emotional processing takes place, a lot of automatic emotion assessment studies have focused on the use of this physiological signal alone, using multi-channel EEG recordings. In this project, we used single-channel EEG, which limits the uptake of brain activity. Even so, we found significant EEG changes for 6 ASD and 4 TD individuals, which indicates that this method may still be capable to detect emotion-related fluctuations in the rhythms of the brain, even in the noisy environment of the MR scanner.

When it comes to EDA, the fact that the signal was high pass filtered with a cut-off frequency of 0.5 Hz as it was acquired, for the MR data acquisitions, resulted in the loss of the tonic component of the signal. This component represents the baseline level of EDA and is an indicator of general arousal. Hence, the fact that it is not present in our MR EDA signals means that the slower emotional state related fluctuations in sweat gland activity were not recorded, which represents a major information loss. For the data acquisitions carried out outside the MR scanner, this issue was corrected, which resulted in signals with very different characteristics and with a clear tonic component. This question, and not the MR-related noise, explains the different results for this signal for the two different settings. The significant results obtained for the group that completed the task outside the MR scanner are in agreement with the extensive literature on the influence of the emotional state on sweat gland activity. Furthermore, this signal is, for this group, the one displaying the most prominent results, with 75% of the subjects which had EDA data available presenting significant changes to the baseline. These results highlight the important role of EDA in emotion and indicate that the experimental task has some influence on this physiological function.

In general, peripheral oxygen saturation does not seem to display any significant changes to baseline, with only 3 out of 31 subjects presenting significant changes. This comes as no surprise and explains the lack of literature relating this measure to the emotional state.

In the end, even though the data acquisitions carried out outside the scanner were meant for comparison purposes, the only two biosignals that are truly comparable are respiration and PPG. Restricting the comparison to the two control groups, we did not find significant changes in the PPG signal for either case. Even though PSC magnitudes are higher for the PPG signals acquired

outside the scanner, their also large variability helps to explain why this does not translate into statistically significant results. As for the case of respiration, for both cases, the group analysis did not return a significant result, nonetheless, half of the individuals in the TD group that completed the MR task present considerable signal changes. However, as mentioned before, the singularity of each individual's physiological response, as well as the uncomfortable setting of the MR scanner, help explain the absence of similar responses in the TD group that completed the task outside the scanner. All in all, despite the limitations of comparability, there is no evidence in our data that the environment of the resonance prevents the successful measurement of physiological signals, which is an encouraging sign, predicting the fruitful combination of fMRI and multi-modal biosignal acquisitions for a variety of applications.

6.4 Statistical Analysis

The statistical analysis revealed that only a very limited number of extracted features present statistically significant differences between conditions, for both the analysis of video and BART runs.

In spite of this, the features extracted from the instantaneous HR, particularly meanHR, maxHR, and minHR seem to be significantly different between conditions for the ASD group, namely between HA and LA, HV and NV, NV and LV, and between the Winning and Losing events, and also between groups for the mentioned arousal and valence conditions. This is in line with the PSC results and the known relation of this signal with emotion (Rainville et al., 2006), shining a light at the role of cardiac function in emotional response and indicating that it may present more reactivity to state alterations than other physiological functions. It is interesting, however, how these differences are present when considering the database ratings to label the data, but not when considering the self-assessment answers. Moreover, for the comparisons between valence conditions, these differences disappear between HV and NV and between NV and LV, when considering the self-assessment answers to label the events, but appear, for the same features, between the HV and LV conditions. This relates to the behavioural analysis findings, revealing more extreme valence responses, given by the subjects of this group, which is in line with reports of altered self-awareness in autism, and evidences the challenge of finding stimuli that induce the desired effects in this population.

For the signals acquired in a common environment, it is worth noting the results obtained for the features derived from the SCRs of the EDA signal, namely 'meanAmpSCR', 'maxAmpSCR', 'rateSCR', 'meanRiseDurSCR' and also 'decTimeEDA', for the HA vs. LA comparison when considering the database mean answers as ground truth. The statistically significant differences for these features between arousal conditions are in line with the literature relating this biosignal with arousing situations, particularly the appearance of phasic SCRs in reaction to such situations. The fact that these differences are present only for the EDA signals acquired outside of the MR scanner,

again shows that the information lost by applying a high-pass filter to this signal during the MR acquisitions, was essential to investigate its relationship with emotion.

While the PSC results led us to expect finding some significant differences in pupil size metrics extracted from the signals acquired in a common environment, which had some luminosity adjustments, this happened only for maxPupilSize, when considering self-assessment answers, which also goes against the norm of finding the most differences when considering the database ratings for event labelling. This suggests that, either the extracted features are not appropriate for emotional state discrimination and the signal should be further explored to find meaningful ones, or the changes found in the signal are unrelated to the emotional content of the task, and are still attributed to light changes, albeit to a lesser magnitude.

6.5 Classification

The statistical analysis and the limited number of features exhibiting statistically significant differences between conditions reveal the low discriminative power of the extracted features for our classification problems and help explain the poor classification outcomes.

A factor that can help explain the poor results of the intraparticipant modality is the low number of observations for each participant, which limits the generalization capacity of the models. However, increasing the number of samples for the interparticipant modality, by joining the observations of every subject, did not increase the performance of the classifiers. This can be justified with the inter-subject variability in ANS response to emotion (Golland, Keissar and Levit-Binnun, 2014). By combining every participant's observations, we are introducing variability in the data which possibly hampers the identification of patterns, compromising the performance of the classifiers. However, we cannot devalue the signs of compromised quality of the features. Even though the positive SNR and PSC calculations imply that the MR environment did not significantly affect the quality of the signals, there are some other limitations, mostly related to the data acquisition step, that potentially did. First, there is the issue with the stimulus contrast between the baseline and the videos, which caused the pupil to strongly react to luminance changes, hiding pupil reactions related to the emotional state. Second, the application of a high-pass filter with a cut-off frequency at 0.5 Hz at the time of acquisition of the EDA caused us to lose an important component of this signal, and its underlying information.

As for the BART runs, according to the statistical analysis performed on the features, the events we hypothesized would result in bigger reactivity, namely winning/losing, and more importantly, the type of reward earned, do not seem to result in major physiological differences between them. This can either be because the stimuli are not causing the desired effect, or because the physiological response that they trigger is similar and unrelated to their quality, but rather to their intensity. For instance, when the participants leveled up and earned a new reward, it meant

they won the game trial, independently of whether they won a social reward or a personalized interest reward, which probably induces a stronger response than the quality of the prize. In the case of winning versus losing, both events can be equally arousing, even though they bear different positive/negative connotations. The arousal dimension of these events is possibly majorly responsible for the physiological response they induce, overshadowing their valence. These points can help explain the unsuccessful automatic discrimination of such events.

The initial analysis and comparison between the quality of the signals acquired in simultaneous with fMRI and outside, and the classification results obtained for the MR dataset, predicted the classification outcomes of the dataset acquired outside the scanner. The SNR and PSC calculations did not indicate any major quality improvement from the MR environment to a normal setting. The only relevant differences encountered were for EDA and Pupil Size, which explanation lies in data acquisition setup and stimulus design, respectively. Like for the MR dataset, the statistical analysis again revealed the lack of discriminative power of the features derived from the signals, explaining the classification outcomes. We must keep in mind, however, the fact that this dataset was comprised of a smaller set of biosignals, limiting the comparisons to the signals that were mutual to both datasets. It would be valuable to compare EEG and instantaneous HR recordings in both environments, especially EEG, which is, out of all the signals acquired, the one subject to more contamination, since the electrodes are literally inside the MR tunnel with the patient.

On a positive note, these results do not point to the rejection of the simultaneous acquisition of physiological signals and fMRI. Instead, they underline the importance that stimulus design and data acquisition steps have on the quality of the acquired biosignals, and mainly, the complexity of automatic emotion assessment, especially when dealing with multi-modal biosignals with different characteristics.

7 Conclusion and Future Work

This project focused primarily on human physiology and its relationship with emotion, with a particular focus on the clinical population of autism spectrum disorder. The understanding of these ties and interdependencies can lead to new rehabilitation targets and the development of valuable tools to complement and increase the efficacy of current therapeutic techniques. The real-time monitoring of an individual's physiological status allows us to identify what stimuli cause a negative reaction, and carefully work around it until it is no longer triggering. This is not only valid for ASD but other clinical populations and even healthy individuals, representing a quality of life improvement (Haji, Mohammadkhani and Hahtami, 2011).

In this sense, the work developed for this dissertation reiterated the challenging nature and complexity behind these physiological interactions. Although the diversity of ANS activity patterns between emotional and cognitive states is evident and strongly evidenced in the literature, their relationship is neither linear nor simple. The biggest challenges are perhaps the singularity of each individual's physiology and subjective emotion interpretation. To get around these obstacles and optimize results, future work must focus on a further search for strong markers of emotional reactivity in the physiological signals that demonstrate consistency between subjects. Considering the different characteristics of each physiological function, in particular, the latency between stimulus onset and response, and response duration, the size of the segments, for each biosignal, can also be tuned and experimented with so that they capture the entirety of the event-related response, and ideally, no more than that. It is also important to consider that the stimuli presented in the experimental task may not have had the desired elicitation effect, hence, future studies must explore stimuli with different characteristics.

On the classification step, feature selection and reduction algorithms must be further explored. The preliminary take on PCA for feature reduction did not improve the results, however, its combination with feature selection or the use of other feature reduction algorithms can help to optimize the extraction of meaningful information from our data.

The PSC results show that some signals display more variations during the events of interest than others when compared to baseline values, which seems to indicate that they do not contribute equally to the discrimination of emotional states. Hence, it is possible that classification would benefit from the selection of features derived from the most relevant biosignals only. This is then one of the paths that should be explored next, and, if successful, it would also reduce the number of signals to acquire for future applications, simplifying data acquisition, and increasing the subject's comfort.

Another important question in this dissertation was the feasibility of acquiring physiological signals simultaneous with fMRI, with all the added obstacles that this entails. The SNR and PSC results do not indicate that there are quality differences between the biosignals that can be attributed

to the environment where they were acquired. This is a positive finding that validates the simultaneous acquisition of physiological signals and fMRI, opening the door to a deeper exploration and understanding of the existing relationships and interactions between brain activity and its translation in the peripheral nervous system. We should keep in mind, however, that only respiration, PPG, EDA, and Pupil Size recordings were compared between settings, and it would be important to understand how other signals may be affected by the MR, mainly the single-channel EEG setup that we use, which probably suffers the most contamination.

Taking the modulation of the brain regions involved in emotion processing, for this experimental task, that fMRI allowed us to assess, we can now relabel the trials according to the activation levels in these regions and analyse the biosignals under this new light. This will improve the precision of the interpretation of variations in the physiological signals, as well as their attribution to different emotional states. Moreover, we can explore the biosignals and the features derived from them to look for the timepoints of greatest autonomic nervous system activity and use it to predict the blood oxygen level dependent fMRI signal, which will allow us to see which regions activate when the emotional stimuli induce a reaction in the autonomic nervous system. This approach has been employed in simultaneous EEG-fMRI data, which is known as EEG-Informed fMRI (Abreu, Leal and Figueiredo, 2018), but similar techniques have been used to correlate HRV and respiration variations with brain activity in different pathological conditions (Macey *et al.*, 2016; Kassinopoulos *et al.*, 2021). Here, we can take advantage of the multimodal biosignals and use them to predict brain responses related to emotional-induced autonomic activity changes.

The work developed in this project resulted in a Full Contributed paper accepted for presentation at the 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (November 2021), which contents can be consulted in Appendix III.

References

Abreu, R., Leal, A. and Figueiredo, P. (2018) 'EEG-informed fMRI: A review of data analysis methods', *Frontiers in Human Neuroscience*, 12(February), pp. 1–23. doi: 10.3389/fnhum.2018.00029.

Allen, J. (2007) 'Photoplethysmography and its application in clinical physiological measurement', *Physiological Measurement*, 28(3). doi: 10.1088/0967-3334/28/3/R01.

Baio, J. *et al.* (2018) 'Prevalence of autism spectrum disorder among children aged 8 Years - Autism and developmental disabilities monitoring network, 11 Sites, United States, 2014', *MMWR Surveillance Summaries*, 67(6). doi: 10.15585/mmwr.ss6706a1.

Baum, S. H., Stevenson, R. A. and Wallace, M. T. (2015) 'Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder', *Progress in Neurobiology*, 134, pp. 140–160. doi: 10.1016/j.pneurobio.2015.09.007.

Bellani, M. *et al.* (2011) 'Virtual reality in autism : state of the art', *Epidemiology and Psychiatric Sciences*, 20, pp. 235–238. doi: 10.1017/S2045796011000448.

Benedek, M. and Kaernbach, C. (2010) 'Decomposition of skin conductance data by means of nonnegative deconvolution', *Psychophysiology*, 47(4), pp. 647–658. doi: 10.1111/j.1469-8986.2009.00972.x.

Blascovich, J. *et al.* (2002) 'Immersive virtual environment technology as a methodological tool for social psychology', *Psychological Inquiry*, 13(2), pp. 103–124. doi: 10.1207/S15327965PLI1302_01.

Bolanos, M., Nazeran, H. and Haltiwanger, E. (2006) 'Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals', *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, (February), pp. 4289–4294. doi: 10.1109/IEMBS.2006.260607.

Bölte, S., Feineis-Matthews, S. and Poustka, F. (2008) 'Brief report: Emotional processing in high-functioning autism - Physiological reactivity and affective report', *Journal of Autism and Developmental Disorders*, 38(4), pp. 776–781. doi: 10.1007/s10803-007-0443-8.

Bradley, M. and Lang, P. J. (1994) 'Self-Assessment Manikin (SAM)', *J.Behav.Ther. & Exp. Psychiat.*, 25(1), pp. 49–59.

Chan, E. D., Chan, Michael M. and Chan, Mallory M. (2013) 'Pulse oximetry: Understanding its basic principles facilitates appreciation of its limitations', *Respiratory Medicine*, 107(6), pp. 789–799. doi: 10.1016/j.rmed.2013.02.004.

Chanel, G. *et al.* (2011) 'Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty', *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 41(6), pp. 1052–1063. doi: 10.1109/TSMCA.2011.2116000.

Chen, L. S. *et al.* (no date) 'Multimodal Human Emotion/Expression Recognition'.

Chennamma, H. R. and Yuan, X. (2013) 'A Survey on Eye-Gaze Tracking Techniques', 4(5), pp. 388–393. Available at: http://arxiv.org/abs/1312.6410.

Di Crosta, A. *et al.* (2020) 'The Chieti Affective Action Videos database, a resource for the study of emotions in psychology', *Scientific Data*, 7(1), pp. 1–6. doi: 10.1038/s41597-020-0366-1.

Damasio, A. and Carvalho, G. B. (2013) 'The nature of feelings: Evolutionary and neurobiological origins', *Nature Reviews Neuroscience*, 14(2), pp. 143–152. doi: 10.1038/nrn3403.

Dawson, M. E., Schell, A. M. and Filion, D. L. (2007) 'The Electrodermal System', *Handbook of Psychophysiology*, pp. 200–223. doi: https://doi.org/10.1017/CBO9780511546396.

Dichter, G. S. *et al.* (2010) 'Affective modulation of the startle eyeblink and postauricular reflexes in autism spectrum disorder', *Journal of Autism and Developmental Disorders*, 40(7), pp. 858–869. doi: 10.1007/s10803-009-0925-y.

Dichter, G. S. *et al.* (2012) 'Reward circuitry function in autism spectrum disorders', *Social Cognitive and Affective Neuroscience*, 7(2), pp. 160–172. doi: 10.1093/scan/nsq095.

Didehbani, N. *et al.* (2016) 'Virtual Reality Social Cognition Training for children with high functioning autism', *Computers in Human Behavior*, 62, pp. 703–711. doi: 10.1016/j.chb.2016.04.033.

Elgendi, M. (2012) 'On the Analysis of Fingertip Photoplethysmogram Signals', *Current Cardiology Reviews*, 8(1), pp. 14–25. doi: 10.2174/157340312801215782.

Essa, I. A. and Pentland, A. P. (1997) 'Coding, analysis, interpretation, and recognition of facial expressions', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pp. 757–763. doi: 10.1109/34.598232.

Freedman, L. W. *et al.* (1994) 'The relationship of sweat gland count to electrodermal activity', *Psychophysiology*, 31(2), pp. 196–200. doi: 10.1111/j.1469-8986.1994.tb01040.x.

Genno, H. *et al.* (1997) 'Using facial skin temperature to objectively evaluate sensations', *International Journal of Industrial Ergonomics*, 19(2), pp. 161–171. doi: 10.1016/S0169-8141(96)00011-X.

Golland, Y., Keissar, K. and Levit-Binnun, N. (2014) 'Studying the dynamics of autonomic activity during emotional experience', *Psychophysiology*, 51(11), pp. 1101–1111. doi: 10.1111/psyp.12261.

Goodwin, M. S. (2008) 'Enhancing and Accelerating the Pace of Autism Research and Treatment: The Promise of Developing Innovative Technology', *Focus on Autism and Other Developmental Disabilities*, 23(2), pp. 125–128. doi: 10.1177/1088357608316678.

Grossard, C. *et al.* (2017) 'Serious games to teach social interactions and emotions to individuals with autism spectrum disorders (ASD)', *Computers and Education*, 113, pp. 195–211. doi: 10.1016/j.compedu.2017.05.002.

Haak, M. *et al.* (2009) 'Detecting Stress Using Eye Blinks And Brain Activity From EEG Signals'. Available at:

http://www.researchgate.net/publication/221024391_Detecting_Stress_using_Eye_Blinks_during_Game_Playing/file/79e4150ac0c38a620b.pdf.

Haji, T. M., Mohammadkhani, S. and Hahtami, M. (2011) 'The effectiveness of life skills training on happiness, quality of life and emotion regulation', *Procedia - Social and Behavioral Sciences*, 30, pp. 407–411. doi: 10.1016/j.sbspro.2011.10.080.

Hall, C. A. and Chilcott, R. P. (2018) 'Eyeing up the future of the pupillary light reflex in neurodiagnostics', *Diagnostics*, 8(1). doi: 10.3390/diagnostics8010019.

He, X. *et al.* (2012) 'The eye activity measurement of mental workload based on basic flight task', *IEEE International Conference on Industrial Informatics (INDIN)*, (1989), pp. 502–507. doi: 10.1109/INDIN.2012.6301203.

Hill, E., Berthoz, S. and Frith, U. (2004) 'Brief report: Cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives', *Journal of Autism and Developmental Disorders*, 34(2), pp. 229–235. doi: 10.1023/B:JADD.0000022613.41399.14.

Huggins, C. F. *et al.* (2021) 'Emotional self-awareness in autism: A meta-analysis of group differences and developmental effects', *Autism*, 25(2), pp. 307–321. doi: 10.1177/1362361320964306.

Johnson, C. P. and Myers, S. M. (2007) 'Identification and Evaluation of Children With Autism Spectrum Disorders', *Pediatrics*, 120(5). doi: 10.1542/peds.2007-2361.

Kandalaft, M. R. *et al.* (2013) 'Virtual reality social cognition training for young adults with high-functioning autism', *Journal of Autism and Developmental Disorders*, 43(1), pp. 34–44. doi: 10.1007/s10803-012-1544-6.

Kassem, K. *et al.* (2017) 'DiVA: Exploring the usage of pupil diameter to elicit valence and arousal', *ACM International Conference Proceeding Series*, pp. 273–278. doi: 10.1145/3152832.3152836.

Kassinopoulos, M. *et al.* (2021) 'Altered Relationship Between Heart Rate Variability and fMRI-Based Functional Connectivity in People With Epilepsy', *Frontiers in Neurology*, 12(June), pp. 1–13. doi: 10.3389/fneur.2021.671890.

Katsis, C. D. *et al.* (2008) 'Toward emotion recognition in car-racing drivers: A biosignal processing approach', *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, 38(3), pp. 502–512. doi: 10.1109/TSMCA.2008.918624.

Kilavik, B. E. *et al.* (2013) 'The ups and downs of beta oscillations in sensorimotor cortex', *Experimental Neurology*, 245, pp. 15–26. doi: 10.1016/j.expneurol.2012.09.014.

Kim, J. and André, E. (2008) 'Emotion recognition based on physiological changes in music listening', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12), pp. 2067–2083. doi: 10.1109/TPAMI.2008.26.

Kim, K. H., Bang, S. W. and Kim, S. R. (2004) 'Emotion recognition system using short-term monitoring of physiological signals', *Medical and Biological Engineering and Computing*, 42(3), pp. 419–427. doi: 10.1007/BF02344719.

Klimesch, W. (1997) ‘EEG-alpha rhythms and memory processes’, *International Journal of Psychophysiology*, 26, pp. 319–340.

Klimesch, W. (1999) ‘EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis’, *Brain Research Reviews*, 29(2–3), pp. 169–195. doi: 10.1016/S0165-0173(98)00056-3.

Kohls, G. *et al.* (2013) ‘Reward system dysfunction in autism spectrum disorders’, *Social Cognitive and Affective Neuroscience*, 8(5), pp. 565–572. doi: 10.1093/scan/nss033.

Kohls, G. *et al.* (2018) ‘Altered reward system reactivity for personalized circumscribed interests in autism’, *Molecular Autism*, 9(1), pp. 1–12. doi: 10.1186/s13229-018-0195-7.

Kragel, P. A. and LaBar, K. S. (2016) ‘Decoding the Nature of Emotion in the Brain’, *Trends in Cognitive Sciences*, 20(6), pp. 444–455. doi: 10.1016/j.tics.2016.03.011.

Kushki, A. *et al.* (2015) ‘A Kalman filtering framework for physiological detection of anxiety-related arousal in children with autism spectrum disorder’, *IEEE Transactions on Biomedical Engineering*, 62(3), pp. 990–1000. doi: 10.1109/TBME.2014.2377555.

Lang, P. J. *et al.* (1993) ‘Looking at pictures: Affective, facial, visceral, and behavioral reactions’, *Psychophysiology*, 30(3), pp. 261–273. doi: 10.1111/j.1469-8986.1993.tb03352.x.

Lang, P. J., Bradley, M. M. and Cuthbert, B. N. (1997) ‘International affective picture system (IAPS): Technical manual and affective ratings’, *NIMH Center for the Study of Emotion and Attention*, pp. 39–58.

Larsen, J. T. *et al.* (2008) ‘The Psychophysiology of Emotion’, in Lewis, M., Haviland-Jones, J. M., and Barrett, L. F. (eds) *Handbook of emotions*. The Guilford Press, pp. 180–195.

Legrand, N. *et al.* (2021) ‘Emotional metacognition: stimulus valence modulates cardiac arousal and metamemory’, *Cognition and Emotion*, 35(4), pp. 705–721. doi: 10.1080/02699931.2020.1859993.

Lejuez, C. W. *et al.* (2002) ‘Evaluation of a behavioral measure of risk taking: The balloon analogue risk task (BART)’, *Journal of Experimental Psychology: Applied*, 8(2), pp. 75–84. doi: 10.1037/1076-898X.8.2.75.

Lisetti, C. L. and Nasoz, F. (2004) ‘Using noninvasive wearable computers to recognize human emotions from physiological signals’, *Eurasip Journal on Applied Signal Processing*, 2004(11), pp. 1672–1687. doi: 10.1155/S1110865704406192.

Liu, C. *et al.* (2008) ‘Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder’, *International Journal of Human Computer Studies*, 66(9), pp. 662–677. doi: 10.1016/j.ijhsc.2008.04.003.

Lohani, M., Payne, B. R. and Strayer, D. L. (2019) ‘A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving’, *Frontiers in Human Neuroscience*, 13(March), pp. 1–27. doi: 10.3389/fnhum.2019.00057.

Lundqvist, M. *et al.* (2018) ‘Gamma and beta bursts during working memory readout suggest roles in

its volitional control', *Nature Communications*, 9(1), pp. 1–12. doi: 10.1038/s41467-017-02791-8.

Macey, P. M. *et al.* (2016) 'Functional imaging of autonomic regulation: Methods and key findings', *Frontiers in Neuroscience*, 9(JAN), pp. 1–23. doi: 10.3389/fnins.2015.00513.

MacIntyre, A. D. (2021) 'breathTimes'. Available at: https://www.mathworks.com/matlabcentral/fileexchange/81066.

Maffei, A. and Angrilli, A. (2019) 'Spontaneous blink rate as an index of attention and emotion during film clips viewing', *Physiology and Behavior*, 204(February), pp. 256–263. doi: 10.1016/j.physbeh.2019.02.037.

Malik, M. *et al.* (1996) 'Heart Rate Variability', *Circulation*, 93(5), pp. 1043–1065. doi: 10.1161/01.CIR.93.5.1043.

McIntire, L. K. *et al.* (2014) 'Detection of vigilance performance using eye blinks', *Applied Ergonomics*, 45(2 PB), pp. 354–362. doi: 10.1016/j.apergo.2013.04.020.

McKee, M. G. (2008) 'Biofeedback: An overview in the context of heart-brain medicine', *Cleveland Clinic Journal of Medicine*, 75(SUPPL.2), pp. 31–34. doi: 10.3949/ccjm.75.Suppl_2.S31.

Niazy, R. K. *et al.* (2005) 'Removal of FMRI environment artifacts from EEG data using optimal basis sets', *NeuroImage*, 28(3), pp. 720–737. doi: 10.1016/j.neuroimage.2005.06.067.

Nishiyama, T. *et al.* (2001) 'Irregular activation of individual sweat glands in human sole observed by a videomicroscopy', *Autonomic Neuroscience: Basic and Clinical*, 88(1–2), pp. 117–126. doi: 10.1016/S1566-0702(01)00229-6.

Nunez, P. L. and Srinivasan, R. (2006) *Electric Fields of the Brain*. Second, *Brain Sciences*. Second. Oxford University Press. doi: 10.1093/acprof:oso/9780195050387.001.0001.

Oliveira, G. *et al.* (2007) 'Epidemiology of autism spectrum disorder in Portugal: Prevalence, clinical characterization, and medical conditions', *Developmental Medicine and Child Neurology*, 49, pp. 726–733.

Oster, J. and Clifford, G. D. (2017) 'Acquisition of electrocardiogram signals during magnetic resonance imaging', *Physiological Measurement*, 38(7), pp. R119–R142. doi: 10.1088/1361-6579/aa6e8c.

Parsons, S. and Cobb, S. (2011) 'State-of-the-art of virtual reality technologies for children on the autism spectrum', *European Journal of Special Needs Education*, 26(3), pp. 355–366. doi: 10.1080/08856257.2011.593831.

Parsons, S. and Mitchell, P. (2002) 'The potential of virtual reality in social skills training for people with autistic spectrum disorders', *Journal of Intellectual Disability Research*, 46(5), pp. 430–443. doi: 10.1046/j.1365-2788.2002.00425.x.

Partala, T. and Surakka, V. (2003) 'Pupil size variation as an indication of affective processing', *International Journal of Human Computer Studies*, 59(1–2), pp. 185–198. doi: 10.1016/S1071-5819(03)00017-X.

Peltola, M. A. (2012) 'Role of editing of R-R intervals in the analysis of heart rate variability', *Frontiers in Physiology*, 3 MAY(May), pp. 1–10. doi: 10.3389/fphys.2012.00148.

Pfurtscheller, G. *et al.* (1997) 'EEG-based discrimination between imagination of right and left hand movement', *Electroencephalography and Clinical Neurophysiology*, 103(6), pp. 642–651. doi: 10.1016/S0013-4694(97)00080-1.

Philip, R. C. M. *et al.* (2012) 'A systematic review and meta-analysis of the fMRI investigation of autism spectrum disorders', *Neuroscience and Biobehavioral Reviews*, 36(2), pp. 901–942. doi: 10.1016/j.neubiorev.2011.10.008.

Philippot, P., Chapelle, G. and Blairy, S. (2002) 'Respiratory feedback in the generation of emotion', *Cognition and Emotion*, 16(5), pp. 605–627. doi: 10.1080/02699930143000392.

Picard, R. W., Vyzas, E. and Healey, J. (2001) 'Toward machine emotional intelligence: analysis of affective physiological state', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10), pp. 1175–1191. doi: 10.1109/34.954607.

Pinheiro, N. *et al.* (2016) 'Can PPG be used for HRV analysis?', *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2016-Octob, pp. 2945–2949. doi: 10.1109/EMBC.2016.7591347.

Rainville, P. *et al.* (2006) 'Basic emotions are associated with distinct patterns of cardiorespiratory activity', *International Journal of Psychophysiology*, 61(1), pp. 5–18. doi: 10.1016/j.ijpsycho.2005.10.024.

Rangayyan, R. M. (2015) *Biomedical Signal Analysis*, *IEEE Press Series in Biomedical Engineering*. Available at: http://dx.doi.org/10.1016/j.cirp.2016.06.001%0Ahttp://dx.doi.org/10.1016/j.powtec.2016.12.055%0Ahttps://doi.org/10.1016/j.ijfatigue.2019.02.006%0Ahttps://doi.org/10.1016/j.matlet.2019.04.024%0Ahttps://doi.org/10.1016/j.matlet.2019.127252%0Ahttp://dx.doi.o.

Ren, P. *et al.* (2013) 'Affective assessment by digital processing of the pupil diameter', *IEEE Transactions on Affective Computing*, 4(1), pp. 2–14. doi: 10.1109/T-AFFC.2012.25.

Samson, A. C. *et al.* (2014) 'Emotion dysregulation and the core features of autism spectrum disorder', *Journal of Autism and Developmental Disorders*, 44(7), pp. 1766–1772. doi: 10.1007/s10803-013-2022-5.

Sapiński, T. *et al.* (2019) 'Emotion recognition from skeletal movements', *Entropy*, 21(7), pp. 1–16. doi: 10.3390/e21070646.

Sarabadani, S. *et al.* (2020) 'Physiological Detection of Affective States in Children with Autism Spectrum Disorder', *IEEE Transactions on Affective Computing*, 11(4), pp. 588–600. doi: 10.1109/TAFFC.2018.2820049.

Sarracanie, M. *et al.* (2015) 'Low-Cost High-Performance MRI', *Scientific Reports*, 5, pp. 1–9. doi: 10.1038/srep15177.

Schmidt, R. *et al.* (2019) 'Beta oscillations in working memory, executive control of movement and

thought, and sensorimotor function', *Journal of Neuroscience*, 39(42), pp. 8231–8238. doi: 10.1523/JNEUROSCI.1163-19.2019.

Schürmann, M., Başar-Eroglu, C. and Başar, E. (1997) 'A possible role of evoked alpha in primary sensory processing: Common properties of cat intracranial recordings and human EEG and MEG', *International Journal of Psychophysiology*, 26(1–3), pp. 149–170. doi: 10.1016/S0167-8760(97)00762-9.

Shaffer, F. and Ginsberg, J. P. (2017) 'An Overview of Heart Rate Variability Metrics and Norms', *Frontiers in Public Health*, 5(September), pp. 1–17. doi: 10.3389/fpubh.2017.00258.

Ben Shalom, D. *et al.* (2006) 'Normal physiological emotions but differences in expression of conscious feelings in children with high-functioning autism', *Journal of Autism and Developmental Disorders*, 36(3), pp. 395–400. doi: 10.1007/s10803-006-0077-2.

Simões, M. *et al.* (2018) 'Virtual Travel Training for Autism Spectrum Disorder : Proof-of-Concept Interventional Study', *JMIR Serious Games*, 6(1). doi: 10.2196/games.8428.

Simões, M. *et al.* (2020) 'Virtual Reality Immersion Rescales Regulation of Interpersonal Distance in Controls but not in Autism Spectrum Disorder', *Journal of Autism and Developmental Disorders*. doi: 10.1007/s10803-020-04484-6.

Soni, R. and Muniyandi, M. (2019) 'Breath rate variability: A novel measure to study the meditation effects', *International Journal of Yoga*, 12(1), p. 45. doi: 10.4103/ijoy.ijoy_27_17.

Strickland, D. (1997) 'Virtual reality for the treatment of autism', *Studies in Health Technology and Informatics*, 44, pp. 81–86. doi: 10.3233/978-1-60750-888-5-81.

Strüber, D. *et al.* (2000) 'Reversal-rate dependent differences in the EEG gamma-band during multistable visual perception', *International Journal of Psychophysiology*, 38(3), pp. 243–252. doi: 10.1016/S0167-8760(00)00168-9.

Tiitinen, H. T. *et al.* (1993) 'Selective attention enhances the auditory 40-Hz transient response in humans', *Nature*, 364(6432), pp. 59–60. doi: 10.1038/364059a0.

De Troyer, A. and Boriek, A. M. (2011) 'Mechanics of the respiratory muscles', *Comprehensive Physiology*, 1(3), pp. 1273–1300. doi: 10.1002/cphy.c100009.

Wang, Y. and Guan, L. (2004) 'An investigation of speech-based human emotion recognition', *2004 IEEE 6th Workshop on Multimedia Signal Processing*, pp. 15–18. doi: 10.1109/mmsp.2004.1436403.

Watson, K. K. *et al.* (2015) 'Increased reward value of non-social stimuli in children and adolescents with autism', *Frontiers in Psychology*, 6(July), pp. 1–7. doi: 10.3389/fpsyg.2015.01026.

Winton, W. M., Putnam, L. E. and Krauss, R. M. (1984) 'Facial and autonomic manifestations of the dimensional structure of emotion', *Journal of Experimental Social Psychology*, 20(3), pp. 195–216. doi: 10.1016/0022-1031(84)90047-7.

Zhang, Y. X. and Cummings, J. R. (2020) 'Supply of Certified Applied Behavior Analysts in the United States: Implications for Service Delivery for Children With Autism', *Psychiatric*

Services, 71(4), pp. 385–388. doi: 10.1176/appi.ps.201900058.

Appendix I – Statistical Analysis of the Extracted Features

Table 1 - p-values of pairwise comparisons from Wilcoxon Signed Rank Test, performed on the dataset acquired in the MR (High Arousal compared to Low Arousal).

Feature	Database		Self-Assessment	
	Clinical (N = 15)	Control (N = 16)	Clinical (N = 15)	Control (N = 16)
meanEDA	0.19	0.28	0.15	0.92
maxEDA	0.23	0.96	0.45	0.68
minEDA	0.36	0.35	0.52	0.76
meanAbsFdEDA	0.85	0.10	0.60	0.07
meanAbsSdEDA	0.85	0.10	0.60	0.07
meanDerivNegEDA	0.85	0.30	0.68	0.06
decTimeEDA	**0.05**	0.68	0.56	0.68
nFallsEDA	0.28	0.98	0.49	0.68
meanAmpSCR	0.28	0.18	0.85	0.80
maxAmpSCR	0.28	0.41	0.49	0.41
rateSCR	0.71	0.17	0.06	0.15
meanRiseDurSCR	0.12	0.44	0.39	0.07
SDRiseDurSCR	0.80	0.81	0.39	0.86
meanPPG	0.60	0.18	**0.02**	0.35
maxPPG	0.23	0.41	0.52	**0.04**
minPPG	0.25	**0.04**	0.52	**0.01**
stdPPG	**0.03**	0.33	0.45	**0.01**
VLF	0.19	0.41	**0.03**	0.72
LF	0.19	0.80	0.06	0.68
HF	0.17	0.76	**0.05**	0.68
RaLH	0.08	0.57	**0.05**	0.88
RaLVL	0.21	0.11	**0.04**	0.88
RaHVL	0.52	0.84	0.21	0.84
NNmean	0.85	0.30	0.93	0.53
NNmin	0.85	0.92	0.85	0.35
NNmax	0.45	0.28	0.60	0.61
SDNN	0.85	0.33	0.80	0.07
SDSD	0.39	0.20	0.45	0.08
RMSSD	0.56	0.23	0.42	0.12
NN50	0.90	0.53	0.22	0.68
pNN50	0.45	0.35	0.56	0.68
meanPeakAmp	**0.02**	0.64	0.89	0.06
stdPeakAmp	0.68	0.44	0.60	**0.02**
meanDeltaNN	0.93	0.53	0.39	0.72
delta	0.21	0.28	0.98	0.15
theta	0.68	0.10	0.85	0.11
alpha	**0.04**	0.88	0.49	0.84
beta	0.08	0.21	0.39	0.61
gamma	0.23	0.61	0.98	0.16
meanResp	0.64	0.16	0.39	0.44
maxResp	0.93	0.13	0.36	0.15
minResp	0.17	0.50	0.98	0.88
meanAbsFdResp	0.19	0.68	0.98	0.76
meanAbsSdResp	0.19	0.76	0.98	0.76
respRate	1.00	0.88	0.04	0.88
meanRespAmp	0.36	0.53	0.76	0.35
medianRespAmp	0.45	0.57	0.89	0.33
stdRespAmp	0.76	0.21	**0.04**	0.33

meanRiseDurResp	0.30	0.88	0.60	0.88
SDRiseDurResp	0.19	0.47	0.15	1.00
meanRespInter	0.89	0.28	0.56	0.23
medianRespInter	0.72	0.23	0.64	0.41
minRespInter	0.52	0.15	0.80	0.20
maxRespInter	0.56	0.53	0.06	0.23
stdRespInter	0.17	0.16	0.17	0.72
respHF	0.30	0.80	0.52	0.50
stdResp	0.17	0.41	0.21	**0.04**
DRResp	0.52	0.23	0.93	0.28
meanPupilSize	0.28	0.06	0.93	0.96
maxPupilSize	0.52	0.60	0.07	0.33
nBlinks	0.21	**0.03**	**0.02**	0.07
meanHR	**0.02**	1.00	0.10	0.39
maxHR	**0.02**	0.80	0.30	0.64
minHR	**0.01**	0.98	**0.01**	0.89
meanAbsFdHR	0.43	0.19	0.39	0.68
meanAbsSdHR	0.76	0.15	0.81	0.52
stdHR	0.09	0.72	0.27	0.93
meanSpO2	0.93	0.41	0.11	0.33
maxSpO2	0.76	0.28	0.80	0.44
minSpO2	1.00	0.72	0.60	0.80

Table 2 - p-values of pairwise comparisons from Wilcoxon Rank Sum Test, for the arousal conditions (Clinical Group compared to Control Group).

Feature	Database		Self-Assessment	
	High Arousal	Low Arousal	High Arousal	Low Arousal
meanEDA	0.54	0.09	0.80	0.19
maxEDA	0.37	0.35	0.44	0.26
minEDA	0.46	0.24	0.29	0.35
meanAbsFdEDA	0.98	0.28	0.86	0.33
meanAbsSdEDA	0.98	0.28	0.86	0.33
meanDerivNegEDA	0.49	0.28	0.89	0.28
decTimeEDA	0.35	**<0.01**	0.29	**0.02**
nFallsEDA	0.23	0.83	0.54	0.92
meanAmpSCR	0.54	0.40	0.40	0.57
maxAmpSCR	0.51	0.37	0.26	0.33
rateSCR	0.91	0.87	0.71	0.95
meanRiseDurSCR	0.23	0.98	0.15	0.98
SDRiseDurSCR	0.29	0.54	0.10	0.59
meanPPG	0.77	0.92	0.28	0.68
maxPPG	0.89	0.24	0.92	0.26
minPPG	0.74	0.12	0.92	0.11
stdPPG	0.51	0.15	0.80	0.49
VLF	0.21	0.11	0.62	0.09
LF	0.20	0.09	0.59	0.09
HF	0.21	0.09	0.54	0.10
RaLH	0.24	0.09	0.68	0.09
RaLVL	0.23	0.28	0.59	0.16
RaHVL	0.31	0.14	0.42	0.12
NNmean	0.46	0.14	0.23	0.13
NNmin	0.68	0.57	0.46	0.42
NNmax	0.46	0.31	0.37	0.29
SDNN	0.92	0.35	0.49	0.09
SDSD	0.17	0.20	0.28	0.19
RMSSD	0.51	0.28	0.42	0.19
NN50	0.12	0.11	0.09	0.15
pNN50	0.08	**0.02**	**0.04**	0.06
meanPeakAmp	0.54	0.14	0.92	0.54
stdPeakAmp	0.51	0.46	0.42	0.31
meanDeltaNN	0.49	0.92	0.28	0.74
delta	**0.01**	**<0.01**	**0.01**	**0.01**
theta	0.65	0.80	0.83	0.74
alpha	0.06	**0.05**	0.07	**0.02**
beta	**0.02**	**<0.01**	**0.01**	**0.01**
gamma	0.29	0.11	0.16	0.16
meanResp	0.80	0.98	0.54	0.42
maxResp	0.42	0.89	0.46	0.46
minResp	0.92	0.74	0.95	0.68
meanAbsFdResp	0.28	0.57	0.13	0.35
meanAbsSdResp	0.31	0.65	0.14	0.42
respRate	0.74	0.72	0.25	0.18
meanRespAmp	0.42	0.92	0.28	0.86
medianRespAmp	0.40	0.89	0.46	0.98
stdRespAmp	0.92	0.98	0.44	0.92
meanRiseDurResp	0.37	0.89	0.42	0.95
SDRiseDurResp	0.71	0.54	0.92	0.42
meanRespInter	0.89	0.54	0.71	0.44
medianRespInter	0.77	0.42	0.77	0.42
minRespInter	0.86	0.46	0.89	0.49

maxRespInter	0.86	0.40	0.35	0.65
stdRespInter	0.80	0.57	0.80	0.68
respHF	0.42	0.16	0.40	0.74
stdResp	**0.03**	0.83	0.09	0.89
DRResp	0.49	0.95	0.65	0.80
meanPupilSize	0.29	0.16	0.46	0.16
maxPupilSize	0.86	0.98	0.17	0.77
nBlinks	0.42	0.08	0.08	0.17
meanHR	**0.02**	0.91	**0.02**	0.08
maxHR	**0.01**	0.56	**0.03**	0.33
minHR	**0.03**	0.84	**0.02**	0.20
meanAbsFdHR	0.53	0.68	1.00	1.00
meanAbsSdHR	0.59	0.59	0.88	0.42
stdHR	0.21	0.56	0.21	0.56
meanSpO2	0.80	0.86	0.86	0.59
maxSpO2	0.64	1.00	0.51	0.89
minSpO2	0.98	0.89	0.86	0.65

Table 3 - p-values of pairwise comparisons between the different valence conditions (High Valence, No Valence, Low Valence) from Wilcoxon Signed Rank Test, for the ASD group.

Feature	Database			Self - Assessment		
	HV vs. NV	HV vs. LV	NV vs. LV	HV vs. NV	HV vs. LV	NV vs. LV
meanEDA	0.64	0.19	0.08	0.60	0.42	0.56
maxEDA	0.36	0.33	0.36	0.56	0.25	0.36
minEDA	0.68	0.98	0.72	0.33	0.93	0.42
meanAbsFdEDA	0.76	0.49	0.85	0.12	0.98	0.76
meanAbsSdEDA	0.76	0.49	0.85	0.12	0.98	0.76
meanDerivNegEDA	0.76	0.49	0.98	0.08	0.98	0.72
decTimeEDA	0.23	0.52	**0.03**	0.49	0.30	0.15
nFallsEDA	0.98	0.12	0.07	0.60	0.21	0.76
meanAmpSCR	0.71	0.68	0.52	0.17	0.64	0.33
maxAmpSCR	0.67	0.68	0.68	0.23	0.49	0.49
rateSCR	1.00	0.33	0.12	0.49	0.21	0.55
meanRiseDurSCR	0.90	0.36	0.42	0.98	0.58	0.52
SDRiseDurSCR	0.13	0.12	0.89	0.54	0.80	0.43
meanPPG	0.85	0.30	0.36	**<0.01**	**0.01**	0.23
maxPPG	**0.01**	0.17	0.39	0.06	0.09	0.49
minPPG	**0.04**	0.72	0.30	**<0.01**	0.36	0.15
stdPPG	0.11	0.98	0.09	**0.02**	0.80	0.11
VLF	1.00	**0.03**	**0.01**	0.85	0.11	0.07
LF	0.85	**<0.01**	**<0.01**	0.64	0.06	**0.05**
HF	0.80	**<0.01**	**<0.01**	0.76	0.06	**0.04**
RaLH	0.52	**<0.01**	**<0.01**	0.89	0.08	**<0.01**
RaLVL	0.52	0.42	**0.02**	0.93	0.42	0.85
RaHVL	0.98	0.21	**0.04**	0.49	0.19	0.39
NNmean	0.49	0.76	0.80	0.36	0.45	0.52
NNmin	0.72	0.60	0.85	0.19	0.23	0.98
NNmax	0.45	0.52	0.76	0.21	0.56	0.33
SDNN	0.72	0.64	0.98	0.21	0.30	0.89
SDSD	0.64	0.17	0.15	0.72	0.30	0.52
RMSSD	0.85	0.23	0.21	1.00	0.19	0.52
NN50	0.72	0.21	1.00	0.97	0.76	0.89
pNN50	0.98	0.52	0.52	0.93	0.42	0.52
meanPeakAmp	**0.04**	0.89	0.06	**0.03**	0.60	0.06
stdPeakAmp	**0.01**	0.25	0.60	0.25	0.12	0.80
meanDeltaNN	0.52	0.42	0.56	0.23	0.60	1.00
delta	0.06	0.45	0.12	**0.02**	0.17	0.19
theta	0.56	0.11	0.76	0.93	0.36	0.28
alpha	**0.03**	0.25	0.28	**0.03**	0.85	**0.04**
beta	**0.02**	0.07	0.39	**<0.01**	0.19	0.30
gamma	0.93	0.52	**0.04**	0.21	0.39	0.36
meanResp	0.60	0.80	0.60	0.93	0.64	0.80
maxResp	**0.02**	0.28	0.68	0.98	0.52	0.76
minResp	0.21	0.39	0.93	**0.03**	0.52	0.06
meanAbsFdResp	0.11	0.08	0.60	0.36	0.98	0.56
meanAbsSdResp	0.11	0.08	0.56	0.36	0.80	0.52
respRate	0.35	0.89	0.49	0.60	0.52	0.81
meanRespAmp	0.80	0.89	0.93	0.28	0.72	0.21
medianRespAmp	0.93	0.72	0.60	0.36	0.98	0.28
stdRespAmp	0.15	**0.03**	0.25	0.15	0.21	0.89
meanRiseDurResp	0.85	0.06	0.25	0.76	**0.02**	**0.01**
SDRiseDurResp	0.64	0.09	**0.04**	0.56	**0.02**	**<0.01**
meanRespInter	0.52	0.23	0.93	0.56	0.52	0.93
medianRespInter	0.64	0.30	0.76	0.49	0.36	0.80
minRespInter	0.72	0.89	0.80	0.49	0.17	0.85

maxRespInter	0.28	**0.01**	0.19	0.89	0.49	0.39
stdRespInter	0.98	0.06	0.19	0.89	**0.05**	0.12
respHF	0.52	0.76	0.72	0.28	0.64	0.49
stdResp	0.07	0.52	0.60	0.36	0.49	0.93
DRResp	**0.01**	0.21	0.89	0.19	0.60	0.60
meanPupilSize	0.72	0.14	0.30	0.28	0.15	0.60
maxPupilSize	0.72	0.76	0.42	0.80	0.25	0.64
nBlinks	0.93	0.24	0.19	**0.02**	0.30	0.68
meanHR	**0.01**	0.36	**<0.01**	0.58	**0.01**	0.33
maxHR	**0.01**	0.81	**0.02**	0.09	**0.03**	0.71
minHR	**<0.01**	0.50	**<0.01**	1.00	**0.01**	0.24
meanAbsFdHR	0.63	0.76	0.22	0.81	0.14	0.15
meanAbsSdHR	0.71	0.63	0.63	0.54	0.39	0.24
stdHR	0.36	**0.05**	**0.02**	0.36	0.33	0.15
meanSpO2	0.68	0.36	0.36	0.56	0.85	1.00
maxSpO2	0.64	0.56	0.72	0.30	0.11	0.72
minSpO2	0.89	0.52	0.76	0.98	0.98	0.89

Table 4 - p-values of pairwise comparisons between the different valence conditions (High Valence, No Valence, Low Valence) from Wilcoxon Signed Rank Test, for the TD group acquired in the MR.

Feature	Database			Self - Assessment		
	HV vs. NV	HV vs. LV	NV vs. LV	HV vs. NV	HV vs. LV	NV vs. LV
meanEDA	1.00	0.30	0.84	0.72	0.30	0.47
maxEDA	0.84	0.06	0.72	0.96	0.57	0.50
minEDA	0.30	0.15	0.26	0.76	0.16	0.06
meanAbsFdEDA	0.21	**0.04**	**0.01**	0.88	**0.04**	**0.01**
meanAbsSdEDA	0.21	**0.04**	**0.01**	0.88	**0.04**	**0.01**
meanDerivNegEDA	0.76	0.06	**0.03**	0.68	**0.03**	0.09
decTimeEDA	0.28	0.12	0.88	0.30	0.30	0.61
nFallsEDA	0.80	0.85	0.89	0.42	0.89	0.80
meanAmpSCR	0.76	0.44	0.12	1.00	0.50	0.53
maxAmpSCR	0.53	0.44	0.21	0.96	0.57	0.44
rateSCR	0.60	0.20	**0.04**	0.33	0.68	0.16
meanRiseDurSCR	0.41	0.33	0.84	0.18	0.92	0.23
SDRiseDurSCR	**0.04**	**0.05**	0.90	0.10	0.13	0.63
meanPPG	0.35	0.30	0.88	0.61	0.88	0.38
maxPPG	0.84	**0.02**	0.08	0.84	0.21	0.07
minPPG	0.21	0.26	**0.05**	0.18	0.13	0.11
stdPPG	0.64	0.13	0.21	0.64	0.35	0.09
VLF	0.68	0.28	0.68	0.92	0.76	0.72
LF	0.38	0.53	0.61	0.61	0.92	0.88
HF	0.35	0.35	0.76	0.68	1.00	0.80
RaLH	0.35	0.33	0.84	0.84	0.76	0.57
RaLVL	0.84	0.13	0.07	0.92	0.30	0.41
RaHVL	0.33	0.26	0.88	0.80	0.61	1.00
NNmean	0.72	0.13	0.07	0.76	0.21	**0.05**
NNmin	0.28	**0.01**	0.35	0.57	0.09	**0.03**
NNmax	0.13	0.38	0.10	0.96	0.47	0.26
SDNN	**0.04**	0.20	0.57	0.07	0.08	0.53
SDSD	0.10	0.06	0.76	**0.05**	0.09	0.21
RMSSD	**0.05**	0.07	0.61	0.06	0.10	0.47
NN50	0.61	0.92	0.29	0.50	0.16	0.30
pNN50	**0.05**	0.09	0.35	0.10	**0.03**	0.72
meanPeakAmp	0.76	0.07	0.33	0.47	0.38	**0.04**
stdPeakAmp	0.50	0.12	0.20	0.96	0.16	0.38
meanDeltaNN	0.44	0.61	0.61	0.23	**0.03**	0.76
delta	0.15	0.11	0.80	1.00	0.06	0.08
theta	0.92	0.50	0.41	0.16	0.88	0.10
alpha	0.80	0.44	0.84	0.64	0.80	0.84
beta	0.64	0.38	0.96	0.72	0.33	0.35
gamma	0.92	0.96	0.13	0.88	0.44	0.64
meanResp	0.18	0.47	0.88	0.44	0.12	0.72
maxResp	0.38	0.44	0.80	**0.01**	0.18	0.84
minResp	0.57	0.16	0.61	0.15	0.57	0.47
meanAbsFdResp	0.20	0.28	0.92	**0.04**	**0.03**	0.64
meanAbsSdResp	0.16	0.33	0.96	**0.04**	**0.03**	0.72
respRate	0.44	0.25	0.47	0.76	0.92	0.98
meanRespAmp	0.68	0.72	0.61	0.80	0.64	0.53
medianRespAmp	0.64	0.88	0.41	0.50	0.21	0.44
stdRespAmp	**0.04**	0.53	**0.02**	0.15	0.53	0.21
meanRiseDurResp	0.72	0.41	0.53	0.92	0.84	0.68
SDRiseDurResp	0.64	0.07	0.13	0.41	0.57	0.68
meanRespInter	0.76	0.30	0.47	0.23	0.41	0.88
medianRespInter	1.00	0.41	0.35	0.23	0.44	0.72
minRespInter	0.76	0.16	0.15	0.16	0.64	0.38

maxRespInter	0.53	0.12	0.50	0.96	0.57	0.88
stdRespInter	0.64	0.96	0.33	0.38	0.96	0.12
respHF	0.88	0.88	0.88	0.72	0.84	0.47
stdResp	0.96	0.12	0.41	0.12	0.96	0.06
DRResp	0.84	0.53	0.53	**0.01**	0.26	0.33
meanPupilSize	0.64	0.28	0.61	0.72	0.28	0.15
maxPupilSize	0.64	0.38	0.44	0.50	0.50	0.76
nBlinks	**0.01**	0.16	0.23	0.50	0.66	0.64
meanHR	0.72	0.60	0.39	0.52	0.52	0.72
maxHR	0.68	0.76	0.89	0.49	0.68	0.89
minHR	0.42	0.19	0.45	0.56	0.36	1.00
meanAbsFdHR	0.64	**0.02**	**0.05**	1.00	0.07	0.14
meanAbsSdHR	0.72	0.42	0.30	0.33	0.80	0.49
stdHR	0.76	0.08	0.56	0.42	0.89	0.64
meanSpO2	0.92	0.44	0.20	0.88	0.80	0.57
maxSpO2	0.16	0.44	0.47	0.53	0.26	0.76
minSpO2	0.15	**0.05**	0.15	**0.02**	**0.02**	0.84

Table 5 - p-values of pairwise comparisons from Wilcoxon Rank Sum Test, for the valence conditions (Clinical Group compared to Control Group).

Feature	Database			Self - Assessment		
	High Valence	Neutral Valence	Low Valence	High Valence	Neutral Valence	Low Valence
meanEDA	0.44	0.31	0.37	0.13	0.28	0.26
maxEDA	0.40	0.29	0.33	0.31	0.42	0.29
minEDA	0.42	0.28	0.44	0.26	0.95	0.57
meanAbsFdEDA	0.57	0.42	0.89	0.80	0.49	0.77
meanAbsSdEDA	0.57	0.42	0.89	0.80	0.49	0.77
meanDerivNegEDA	0.51	0.40	0.95	0.51	0.65	0.98
decTimeEDA	0.13	**0.02**	0.86	0.08	**0.02**	0.92
nFallsEDA	0.64	0.91	0.30	0.54	0.95	0.49
meanAmpSCR	0.33	0.37	0.80	0.59	0.48	0.77
maxAmpSCR	0.59	0.26	0.31	0.40	0.43	0.40
rateSCR	0.86	0.72	0.55	0.51	0.55	0.75
meanRiseDurSCR	0.53	0.83	0.31	0.48	0.32	0.29
SDRiseDurSCR	0.12	0.77	0.45	**0.04**	0.74	0.35
meanPPG	0.28	0.92	0.24	0.65	1.00	0.23
maxPPG	0.98	0.21	0.57	1.00	0.35	0.68
minPPG	0.80	**0.02**	0.59	0.89	0.06	0.54
stdPPG	0.71	0.17	0.19	0.92	**0.02**	0.23
VLF	0.15	0.07	0.40	0.11	0.08	0.33
LF	0.17	0.07	0.44	0.11	0.13	0.35
HF	0.19	0.08	0.44	0.09	0.10	0.35
RaLH	0.14	0.07	0.51	0.09	0.09	0.37
RaLVL	0.49	0.29	0.29	0.24	0.24	0.17
RaHVL	0.21	0.14	0.46	0.11	0.16	0.37
NNmean	0.33	0.12	0.89	0.80	0.15	0.86
NNmin	0.59	0.83	0.33	0.68	0.49	0.19
NNmax	0.74	0.20	0.57	0.35	0.21	0.51
SDNN	0.80	0.31	1.00	0.57	0.57	0.86
SDSD	0.26	0.37	0.54	0.95	0.44	0.89
RMSSD	0.65	0.40	0.65	0.65	0.46	0.95
NN50	0.17	0.08	0.19	0.09	0.20	0.09
pNN50	0.92	**<0.01**	0.16	0.95	0.09	**0.03**
meanPeakAmp	0.46	0.16	0.21	0.74	**0.01**	0.26
stdPeakAmp	0.62	0.35	0.71	0.68	0.65	0.92
meanDeltaNN	0.71	0.59	0.20	0.57	0.09	0.29
delta	0.10	**<0.01**	**0.01**	0.09	**0.01**	**0.01**
theta	0.80	0.92	0.95	0.77	0.59	0.92
alpha	**0.05**	**0.03**	0.07	**0.04**	**0.02**	0.09
beta	0.10	**<0.01**	**<0.01**	0.14	**<0.01**	**<0.01**
gamma	0.31	0.06	0.44	0.21	0.19	0.24
meanResp	0.42	0.62	0.89	0.68	0.92	0.68
maxResp	0.19	1.00	0.74	0.12	0.80	0.71
minResp	0.89	0.98	0.95	0.40	0.24	0.98
meanAbsFdResp	**0.01**	0.57	0.62	**0.02**	0.74	0.92
meanAbsSdResp	**0.02**	0.59	0.62	**0.02**	0.74	0.89
respRate	0.53	0.54	0.83	0.92	0.74	0.71
meanRespAmp	0.95	0.40	0.86	0.62	0.40	0.98
medianRespAmp	0.92	0.40	0.71	0.54	0.35	0.71
stdRespAmp	0.17	0.71	0.98	0.15	0.40	0.68
meanRiseDurResp	0.77	0.65	0.33	0.89	0.68	0.07
SDRiseDurResp	0.24	0.98	0.46	0.54	0.14	0.59
meanRespInter	0.28	0.89	0.42	0.74	0.80	0.89
medianRespInter	0.40	0.80	0.54	0.77	0.62	0.86

minRespInter	0.28	0.71	1.00	0.71	0.44	0.59
maxRespInter	0.35	0.65	0.33	0.86	0.74	0.95
stdRespInter	0.57	0.62	0.65	0.95	0.57	0.92
respHF	0.33	0.16	0.57	0.37	0.31	0.31
stdResp	**0.04**	0.68	0.23	0.08	0.57	0.12
DRResp	0.15	0.74	0.62	0.11	0.98	0.42
meanPupilSize	0.57	0.31	0.09	0.54	0.10	0.11
maxPupilSize	0.21	1.00	0.81	0.14	0.77	0.89
nBlinks	0.58	0.18	0.15	0.89	**0.01**	0.28
meanHR	**0.02**	0.62	**0.03**	0.17	0.23	**0.05**
maxHR	0.12	0.90	**0.03**	0.74	0.39	**0.03**
minHR	**0.02**	0.71	0.10	0.09	0.14	0.08
meanAbsFdHR	0.88	0.68	0.14	1.00	1.00	0.68
meanAbsSdHR	0.98	0.95	0.68	0.74	0.27	0.78
stdHR	0.17	0.53	0.21	0.47	0.71	0.23
meanSpO2	0.98	0.92	0.92	0.86	0.98	0.98
maxSpO2	0.37	0.65	0.57	0.92	0.29	0.92
minSpO2	0.54	0.92	0.38	0.29	0.71	0.51

Table 6 - p-values of pairwise comparisons from Wilcoxon Signed Rank Test, performed on the dataset acquired in a common environment (High Arousal compared to Low Arousal).

Feature	Database	Self-Assessment	Feature	Database	Self-Assessment
meanEDA	0.73	0.15	**NN50**	0.54	0.64
maxEDA	0.20	0.20	**pNN50**	0.59	0.06
minEDA	0.97	0.09	**meanPeakAmp**	0.95	0.68
meanAbsFdEDA	0.20	0.47	**stdPeakAmp**	0.34	0.84
meanAbsSdEDA	0.20	0.47	**meanDeltaNN**	0.07	0.50
meanDerivNegEDA	0.73	0.34	**meanResp**	0.67	0.30
decTimeEDA	**0.03**	0.23	**maxResp**	0.27	0.50
nFallsEDA	0.07	0.27	**minResp**	0.36	0.76
meanAmpSCR	**<0.01**	0.42	**meanAbsFdResp**	0.30	0.58
maxAmpSCR	**<0.01**	0.05	**meanAbsSdResp**	0.30	0.58
rateSCR	**0.02**	0.06	**respRate**	0.13	0.57
meanRiseDurSCR	**<0.01**	**0.03**	**meanRespAmp**	0.11	0.85
SDRiseDurSCR	0.62	0.38	**medianRespAmp**	0.11	0.85
meanPPG	0.50	0.79	**stdRespAmp**	1.00	1.00
maxPPG	1.00	0.79	**meanRiseDurResp**	**0.04**	0.42
minPPG	0.68	0.84	**SDRiseDurResp**	1.00	1.00
stdPPG	1.00	0.84	**meanRespInter**	0.94	0.81
VLF	1.00	1.00	**medianRespInter**	0.58	0.69
LF	0.74	0.89	**minRespInter**	0.47	0.58
HF	0.68	0.74	**maxRespInter**	0.94	0.94
RaLH	0.13	0.34	**stdRespInter**	0.38	0.81
NNmean	0.31	0.79	**stdResp**	0.81	0.58
NNmin	0.74	1.00	**DRResp**	0.63	0.58
NNmax	0.41	0.45	**meanPupilSize**	0.95	0.76
SDNN	0.17	0.89	**maxPupilSize**	0.27	**0.01**
SDSD	0.17	0.54	**nBlinks**	0.29	0.19
RMSSD	0.13	0.74			

Table 7 - p-values of pairwise comparisons between the different valence conditions (High Valence, No Valence, Low Valence) from Wilcoxon Signed Rank Test, for the TD group acquired in a common environment.

Feature	Database			Self - Assessment		
	HV vs. NV	HV vs. LV	NV vs. LV	HV vs. NV	HV vs. LV	NV vs. LV
meanEDA	0.20	0.85	0.91	0.38	0.85	0.68
maxEDA	0.47	0.97	0.27	0.97	0.73	1.00
minEDA	0.30	0.79	0.62	0.85	0.68	0.30
meanAbsFdEDA	0.27	0.57	0.08	0.38	0.97	0.15
meanAbsSdEDA	0.27	0.57	0.08	0.38	0.97	0.15
meanDerivNegEDA	0.20	0.52	0.85	0.09	0.38	0.52
decTimeEDA	0.38	1.00	0.20	0.79	0.85	0.47
nFallsEDA	**0.03**	0.34	0.18	0.35	0.83	0.33
meanAmpSCR	0.23	0.79	0.06	0.23	0.38	0.23
maxAmpSCR	**0.04**	0.34	**0.01**	0.27	0.79	0.20
rateSCR	0.05	0.83	**0.02**	0.17	0.24	0.16
meanRiseDurSCR	0.06	0.62	0.05	**0.01**	0.34	**0.01**
SDRiseDurSCR	0.38	0.91	0.47	0.85	0.41	0.52
meanPPG	0.64	0.34	0.59	0.68	0.74	0.54
maxPPG	0.89	0.64	1.00	0.54	0.38	0.34
minPPG	0.50	0.59	1.00	0.74	0.54	0.74
stdPPG	0.89	1.00	0.95	0.54	0.45	0.27
VLF	1.00	1.00	1.00	1.00	1.00	1.00
LF	0.45	0.45	0.68	0.54	0.17	0.64
HF	0.45	0.89	0.54	0.50	0.17	0.08
RaLH	0.19	0.59	0.74	0.79	0.45	0.38
NNmean	0.74	0.15	0.19	0.15	**0.03**	0.27
NNmin	0.74	0.11	0.38	0.22	0.06	0.45
NNmax	1.00	0.89	0.38	0.68	0.54	0.27
SDNN	**0.04**	0.38	0.89	0.22	0.41	0.89
SDSD	0.11	0.59	0.74	0.22	0.27	0.59
RMSSD	0.08	0.54	0.95	0.17	0.31	0.38
NN50	0.23	0.65	0.64	0.39	0.68	0.59
pNN50	0.79	0.38	0.38	0.79	0.84	1.00
meanPeakAmp	0.84	1.00	0.79	0.68	0.59	0.27
stdPeakAmp	0.45	0.59	0.79	0.84	0.84	0.50
meanDeltaNN	1.00	0.74	0.95	0.45	0.79	0.17
meanResp	0.86	0.76	0.76	0.76	0.22	0.24
maxResp	1.00	0.27	0.17	0.76	0.30	0.10
minResp	0.08	0.27	0.63	**0.04**	0.10	0.86
meanAbsFdResp	0.08	0.07	0.76	0.58	0.14	**0.05**
meanAbsSdResp	0.08	0.07	0.76	0.58	0.14	**0.05**
respRate	0.23	0.18	0.83	0.56	**0.03**	0.90
meanRespAmp	0.08	0.11	0.52	0.83	0.57	0.90
medianRespAmp	0.08	0.11	0.52	0.83	0.57	0.90
stdRespAmp	1.00	1.00	1.00	1.00	1.00	1.00
meanRiseDurResp	0.76	0.68	0.83	0.37	0.38	0.83
SDRiseDurResp	1.00	1.00	1.00	1.00	1.00	1.00
meanRespInter	0.94	0.81	0.69	0.22	0.81	0.11
medianRespInter	0.94	0.81	0.94	0.16	0.63	0.30
minRespInter	0.94	1.00	1.00	0.11	0.63	**0.05**
maxRespInter	1.00	0.69	0.58	0.38	0.81	0.47
stdRespInter	0.58	0.84	0.84	0.69	0.63	0.81
stdResp	**0.02**	**0.04**	0.63	**0.02**	**0.02**	0.33
DRResp	0.15	0.08	0.39	0.17	**0.02**	0.33
meanPupilSize	0.30	**0.05**	0.39	0.08	0.14	0.90
maxPupilSize	0.43	0.06	0.71	**0.05**	**0.02**	0.90
nBlinks	0.95	0.59	0.76	0.22	0.81	0.15

Table 8 - p-values of pairwise comparisons from Wilcoxon Signed Rank Test. Second and third columns refer to the comparison of the winning against the losing event, fourth and fifth columns refer to the comparison between the social and interest rewards.

Feature	Winning vs. Losing		Social vs. Interest Reward	
	Clinical (N = 15)	Control (N = 16)	Clinical (N = 15)	Control (N = 16)
meanEDA	**<0.01**	0.09	0.12	0.76
maxEDA	0.23	0.68	0.11	0.80
minEDA	**0.02**	0.21	0.36	0.68
meanAbsFdEDA	0.28	0.64	0.25	0.38
meanAbsSdEDA	0.28	0.64	0.25	0.38
meanDerivNegEDA	0.17	0.38	0.64	0.41
decTimeEDA	0.52	0.57	0.49	0.38
nFallsEDA	0.25	**0.04**	**0.01**	0.33
meanAmpSCR	0.15	0.92	0.98	0.35
maxAmpSCR	0.09	1.00	0.52	0.64
rateSCR	0.30	0.28	0.36	0.81
meanRiseDurSCR	0.93	0.92	0.08	0.98
SDRiseDurSCR	0.19	0.09	0.16	0.73
meanPPG	0.64	0.50	0.80	0.57
maxPPG	0.23	0.28	0.06	0.92
minPPG	0.14	0.20	0.19	0.84
stdPPG	0.21	0.23	**0.05**	0.64
VLF	0.39	**0.04**	0.98	0.88
LF	0.39	**0.04**	1.00	0.96
HF	0.89	**0.04**	0.42	0.96
RaLH	0.56	**0.03**	0.98	0.88
RaLVL	0.89	0.06	0.42	0.88
RaHVL	0.42	0.07	0.36	0.96
NNmean	0.15	0.41	0.30	0.68
NNmin	0.64	0.12	0.93	0.84
NNmax	**0.02**	0.76	0.30	0.68
SDNN	**0.02**	0.20	0.23	0.50
SDSD	**0.04**	0.28	0.19	0.53
RMSSD	**0.03**	0.26	0.19	0.53
NN50	1.00	0.94	0.11	0.59
pNN50	0.56	0.68	0.08	0.47
meanPeakAmp	0.15	0.41	**0.03**	0.61
stdPeakAmp	0.23	0.15	0.49	0.21
meanDeltaNN	0.60	0.38	0.72	0.96
delta	0.52	0.61	0.89	0.21
theta	0.64	0.88	0.23	0.61
alpha	0.07	0.47	0.85	0.13
beta	0.30	0.61	1.00	0.16
gamma	**0.02**	0.92	0.21	0.26
meanResp	0.49	0.80	**0.05**	0.16
maxResp	0.60	0.64	1.00	0.88
minResp	0.42	0.57	0.25	0.96
meanAbsFdResp	0.36	0.44	0.80	0.21
meanAbsSdResp	0.39	0.47	0.80	0.21
respRate	0.72	0.72	0.37	0.55
meanRespAmp	0.98	0.28	0.15	0.88
medianRespAmp	0.93	0.35	0.28	0.53
stdRespAmp	0.21	0.38	0.19	0.84
meanRiseDurResp	0.98	0.96	0.56	0.96
SDRiseDurResp	0.33	0.16	0.72	0.18
meanRespInter	**0.02**	0.20	0.21	0.64
medianRespInter	**0.03**	0.50	0.25	0.68

minRespInter	0.06	0.41	0.52	0.38
maxRespInter	0.06	0.10	0.60	0.35
stdRespInter	0.64	0.35	0.76	**0.05**
respHF	0.68	0.80	0.80	0.21
stdResp	0.68	0.68	1.00	0.11
DRResp	0.56	0.84	0.68	0.72
meanPupilSize	0.06	0.96	0.64	0.53
maxPupilSize	0.15	**0.02**	0.68	0.30
nBlinks	0.28	0.15	0.35	0.41
meanHR	**0.02**	0.39	0.11	0.45
maxHR	**0.04**	0.46	0.24	0.58
minHR	0.14	0.22	0.34	0.42
meanAbsFdHR	0.27	0.19	0.09	**0.05**
meanAbsSdHR	0.86	0.76	0.95	0.79
stdHR	0.50	0.67	1.00	0.79
meanSpO2	0.17	0.56	0.50	0.86
maxSpO2	0.19	0.85	0.14	0.79
minSpO2	0.19	0.56	0.16	1.00

Table 9 - p-values of pairwise comparisons from Wilcoxon Rank Sum Test (Clinical Group compared to Control Group).

Feature	Winning	Losing	Social Reward	Interest Reward
meanEDA	0.40	0.06	0.71	0.09
maxEDA	0.74	0.33	0.98	0.86
minEDA	0.71	0.09	0.80	0.68
meanAbsFdEDA	0.95	0.49	0.80	0.74
meanAbsSdEDA	0.95	0.49	0.80	0.74
meanDerivNegEDA	0.98	0.86	0.74	0.71
decTimeEDA	0.68	0.89	0.83	0.42
nFallsEDA	0.94	0.94	0.13	0.35
meanAmpSCR	0.74	0.35	0.57	0.83
maxAmpSCR	0.59	0.16	0.95	0.65
rateSCR	0.98	0.41	0.59	0.75
meanRiseDurSCR	0.69	0.46	0.18	0.94
SDRiseDurSCR	0.67	**0.01**	0.78	0.98
meanPPG	0.86	0.16	0.62	0.57
maxPPG	0.16	0.74	0.06	0.77
minPPG	**<0.01**	0.89	**0.03**	0.29
stdPPG	**0.02**	1.00	**0.03**	0.46
VLF	**0.01**	0.57	**0.03**	**0.04**
LF	**0.01**	0.49	**0.03**	**0.03**
HF	**0.04**	0.49	0.12	**0.03**
RaLH	**0.01**	0.57	**0.03**	**0.03**
RaLVL	0.14	0.51	0.11	**0.04**
RaHVL	**0.05**	0.46	0.07	0.05
NNmean	0.16	0.28	0.40	0.29
NNmin	0.98	0.92	0.50	0.95
NNmax	0.08	0.49	**0.04**	0.35
SDNN	**0.04**	0.44	**0.02**	0.28
SDSD	**0.04**	0.46	0.13	0.33
RMSSD	0.05	0.46	0.11	0.35
NN50	0.32	0.68	0.41	0.62
pNN50	0.21	0.71	0.29	0.33
meanPeakAmp	**0.01**	0.89	**0.03**	0.59
stdPeakAmp	**0.01**	0.86	0.26	0.08
meanDeltaNN	0.71	0.68	0.86	0.86
delta	0.15	0.13	0.89	0.07
theta	0.51	0.77	0.98	0.86
alpha	0.74	0.92	0.65	0.80
beta	0.26	0.80	0.95	0.13
gamma	0.49	**0.04**	0.29	0.77
meanResp	0.77	0.57	1.00	0.95
maxResp	**0.03**	0.19	0.06	0.09
minResp	0.71	0.40	0.29	0.89
meanAbsFdResp	0.95	0.77	0.54	0.71
meanAbsSdResp	0.89	0.77	0.49	0.74
respRate	**0.03**	**0.01**	0.12	**0.03**
meanRespAmp	0.10	0.71	0.71	0.06
medianRespAmp	0.16	0.44	0.51	0.05
stdRespAmp	0.92	0.74	0.29	0.72
meanRiseDurResp	0.17	0.17	0.59	0.45
SDRiseDurResp	0.33	0.92	0.74	0.12
meanRespInter	0.74	**0.02**	0.98	0.66
medianRespInter	0.57	**0.01**	0.95	0.75
minRespInter	0.29	**<0.01**	0.95	0.27
maxRespInter	0.62	0.07	0.68	0.18

stdRespInter	**0.03**	0.59	0.65	**<0.01**
respHF	0.65	0.12	0.89	0.44
stdResp	0.16	0.08	0.12	0.62
DRResp	0.40	0.59	0.40	0.33
meanPupilSize	0.62	0.92	0.39	0.76
maxPupilSize	0.21	0.26	0.42	0.25
nBlinks	0.34	0.65	0.51	0.34
meanHR	0.84	0.40	0.40	0.32
maxHR	0.60	0.63	0.34	0.75
minHR	0.63	0.57	0.58	0.27
meanAbsFdHR	0.24	0.40	0.72	0.08
meanAbsSdHR	0.51	0.77	0.98	0.72
stdHR	0.98	0.60	0.90	0.68
meanSpO2	0.56	0.14	0.65	0.98
maxSpO2	0.38	0.05	0.34	0.83
minSpO2	0.49	0.20	**0.04**	0.79

Appendix II – Classification Outcomes

Analysis A1 with PCA

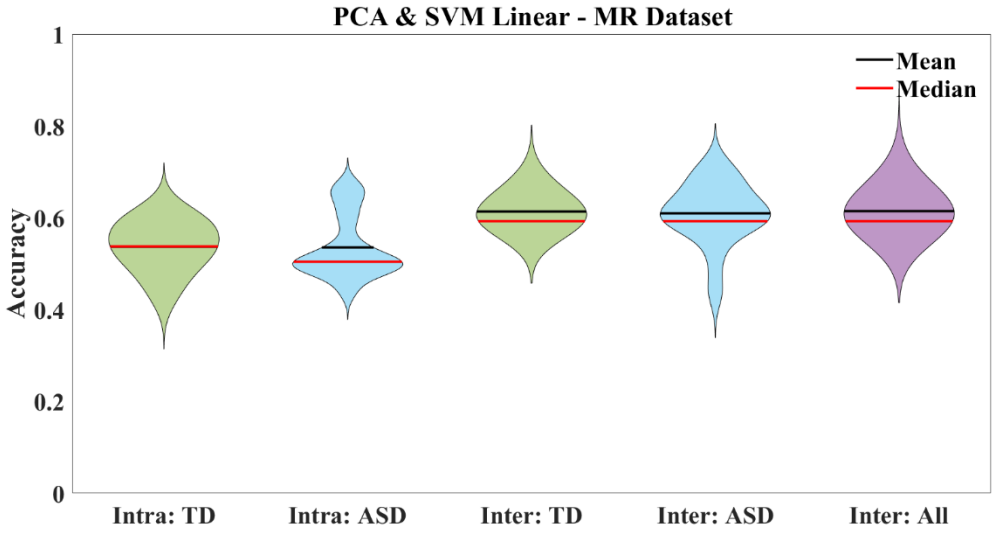

Figure 1 - Distribution of accuracies achieved, on the dataset acquired in the MR environment, using PCA for feature reduction and SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the CAAV database ratings as ground truth.

Table 10 - Performance metrics mean values and respective standard deviations, in brackets, achieved on the dataset acquired in the MR environment, using PCA for feature reduction and SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the CAAV database ratings as ground truth. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject: TD	Intrasubject: ASD	Intersubject: TD	Intersubject: ASD	Intersubject: All
Accuracy	0.54 (0.06)	0.54 (0.07)	0.61 (0.03)	0.61 (0.07)	0.61 (0.03)
Sensitivity	0.78 (0.10)	0.78 (0.09)	1.00 (0.00)	0.91 (0.17)	0.99 (0.05)
Specificity	0.21 (0.12)	0.18 (0.10)	0.00 (0.00)	0.11 (0.18)	0.01 (0.06)
Precision	0.62 (0.04)	0.60 (0.06)	0.61 (0.03)	0.63 (0.05)	0.62 (0.04)
F-Measure	0.66 (0.06)	0.66 (0.06)	0.76 (0.02)	0.74 (0.09)	0.76 (0.03)
p-value	0.12	0.09	0.16	0.20	0.16

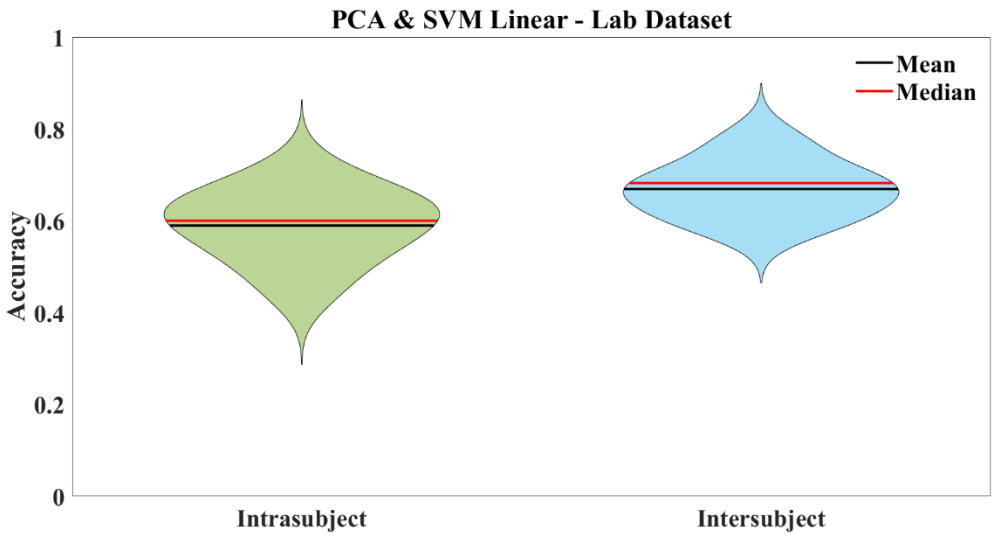

Figure 2 - Distribution of accuracies achieved, on the dataset acquired in a common environment, using PCA for feature reduction and SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the CAAV database ratings as ground truth.

Table 11 - Performance metrics mean values and respective standard deviations, in brackets, achieved on the dataset acquired in a common environment, using PCA for feature reduction and SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the CAAV database ratings as ground truth. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject	**Intersubject**
Accuracy	0.59 (0.08)	0.67 (0.06)
Sensitivity	0.85 (0.08)	0.99 (0.03)
Specificity	0.16 (0.15)	0.07 (0.11)
Precision	0.66 (0.09)	0.67 (0.05)
F-Measure	0.71 (0.06)	0.80 (0.04)
p-value	0.37	0.11

Analysis A2

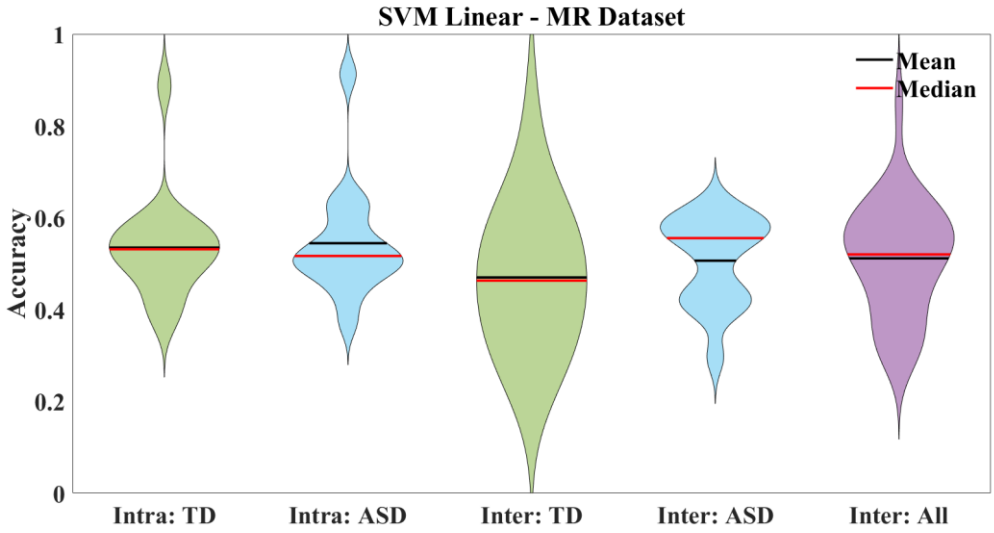

Figure 3 - Distribution of accuracies achieved, on the dataset acquired in the MR environment, by SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the self-assessment answers as ground truth.

Table 12 - Performance metrics mean values and respective standard deviations, in brackets, achieved on the dataset acquired in the MR environment, by SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the self-assessment answers as ground truth. Last row corresponds the p-value estimation obtained with the permutation tests.

	Intrasubject: TD	Intrasubject: ASD	Intersubject: TD	Intersubject: ASD	Intersubject: All
Accuracy	0.53 (0.11)	0.54 (0.12)	0.47 (0.15)	0.51 (0.10)	0.51 (0.12)
Sensitivity	0.50 (0.17)	0.44 (0.11)	0.48 (0.16)	0.42 (0.29)	0.38 (0.24)
Specificity	0.50 (0.18)	0.56 (0.19)	0.47 (0.27)	0.66 (0.25)	0.68 (0.24)
Precision	0.49 (0.18)	0.42 (0.15)	0.46 (0.24)	0.47 (0.26)	0.47 (0.26)
F-Measure	0.53 (0.13)	0.53 (0.15)	0.45 (0.17)	0.33 (0.13)	0.39 (0.15)
p-value	0.14	0.16	0.68	0.62	0.51

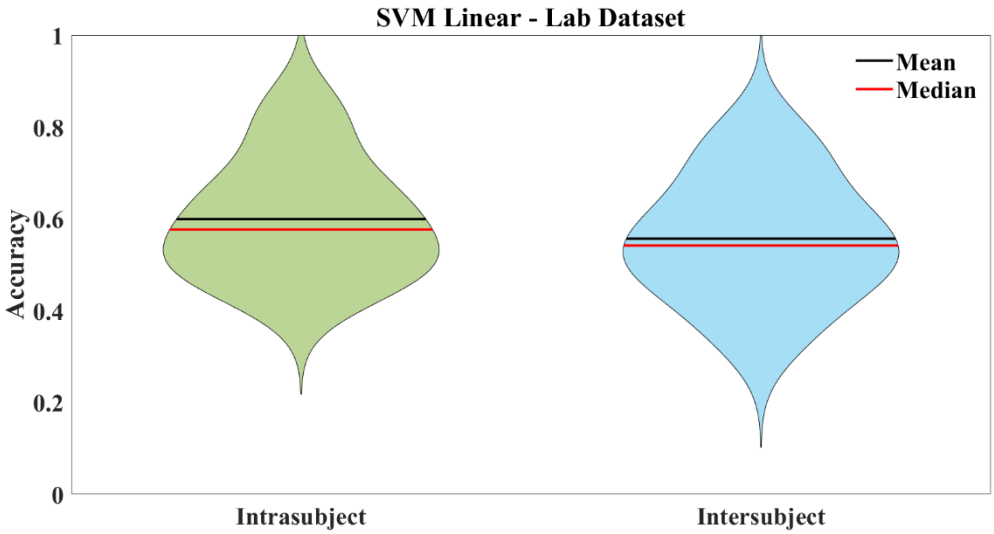

Figure 4 - Distribution of accuracies, achieved on the dataset acquired in a common environment, by SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the self-assessment answers as ground truth.

Table 13 - Performance metrics mean values and respective standard deviations, in brackets, achieved on the dataset acquired in a common environment, by SVM with a linear kernel on classifying HA vs LA on intra and inter subject modalities using the self-assessment answers as ground truth. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject	Intersubject
Accuracy	0.60 (0.13)	0.56 (0.13)
Sensitivity	0.39 (0.30)	0.13 (0.24)
Specificity	0.64 (0.31)	0.89 (0.22)
Precision	0.42 (0.18)	0.45 (0.35)
F-Measure	0.62 (0.18)	0.36 (0.09)
p-value	0.07	0.43

Analysis A3 with PCA

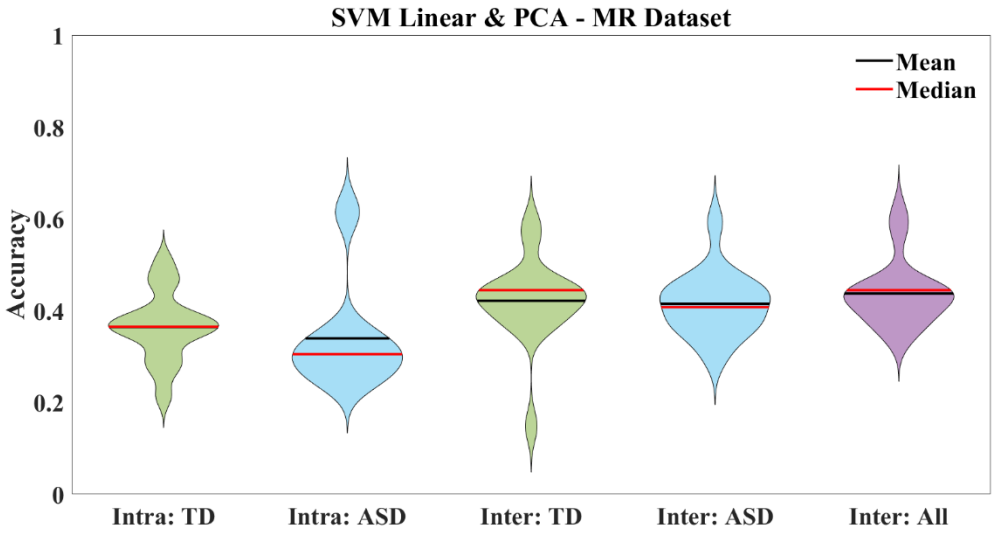

Figure 5 - Distribution of accuracies achieved, on the dataset acquired in the MR environment, using PCA for feature reduction and SVM with a linear kernel on classifying HV, NV and LV on intra and inter subject modalities using the CAAV database ratings as ground truth.

Table 14 - Performance metrics mean values and respective standard deviations, in brackets, achieved, on the dataset acquired in the MR environment, using PCA for feature reduction and SVM with a linear kernel on classifying High Valence, Neutral Valence and Low Valence, on intra and inter subject modalities using the self-assessment answers as ground truth. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject: TD	**Intrasubject: ASD**	**Intersubject: TD**	**Intersubject: ASD**	**Intersubject: All**
Accuracy	0.36 (0.08)	0.34 (0.12)	0.42 (0.09)	0.41 (0.07)	0.44 (0.07)
Macro F1 - Score	0.35 (0.04)	0.34 (0.08)	0.35 (0.07)	0.38 (0.06)	0.36 (0.06)
p-value	0.48	0.51	0.21	0.27	0.18

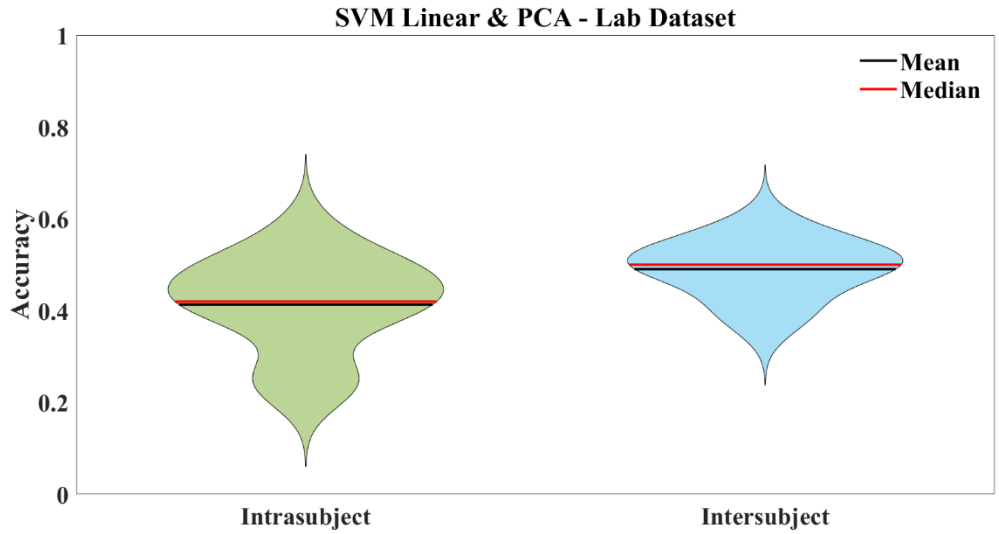

Figure 6 - Distribution of accuracies achieved on the dataset acquired in a common environment, using PCA for feature reduction and SVM with a linear kernel on classifying HV, NV and LV on intra and inter subject modalities using the CAAV database ratings as ground truth.

Table 15 - Performance metrics mean values and respective standard deviations, in brackets, achieved on the dataset acquired in a common environment, using PCA for feature reduction and SVM with a linear kernel on classifying High Valence, Neutral Valence and Low Valence, on intra and inter subject modalities using the CAAV database ratings as ground truth. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject	**Intersubject**
Accuracy	0.41 (0.10)	0.49 (0.06)
Macro F1 - Score	0.38 (0.07)	0.37 (0.05)
p-value	0.42	0.12

Analysis A4

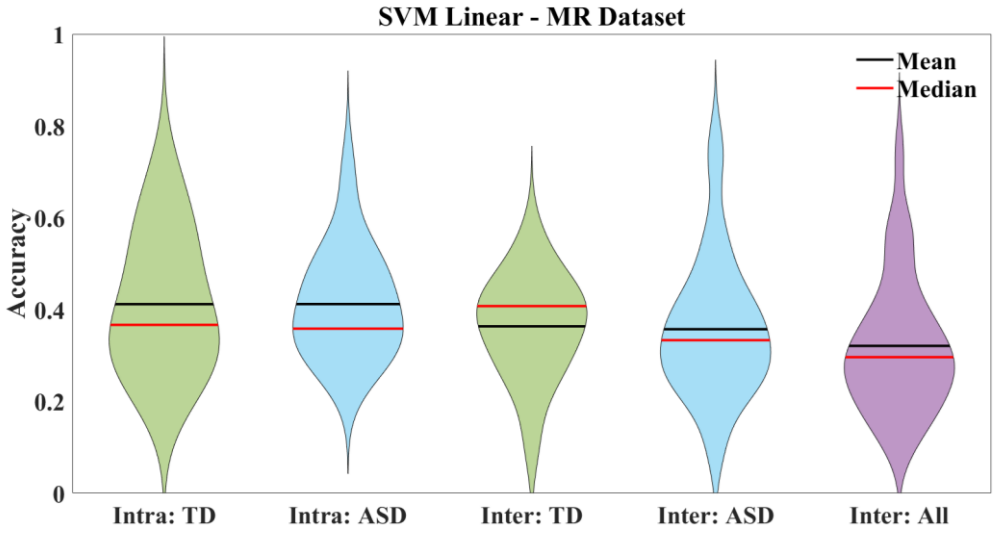

Figure 7 - Distribution of accuracies achieved on the dataset acquired in the MR environment, by SVM with a linear kernel on classifying HV, NV and LV on intra and inter subject modalities using the self-assessment answers as ground truth.

Table 16 - Performance metrics mean values and respective standard deviations, in brackets, achieved on the dataset acquired in the MR environment, by SVM with a linear kernel on classifying High Valence, Neutral Valence and Low Valence, on intra and inter subject modalities using the self-assessment answers as ground truth. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject: TD	Intrasubject: ASD	Intersubject: TD	Intersubject: ASD	Intersubject: All
Accuracy	0.41 (0.15)	0.41 (0.11)	0.36 (0.11)	0.36 (0.15)	0.32 (0.15)
Macro F1 - Score	0.37 (0.08)	0.38 (0.06)	0.35 (0.11)	0.32 (0.10)	0.30 (0.12)
p-value	0.43	0.44	0.50	0.56	0.64

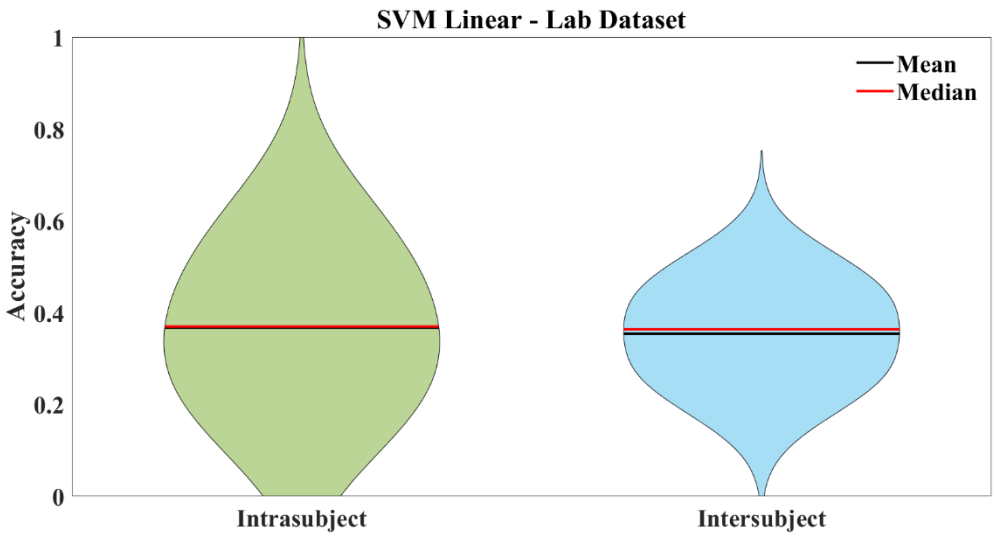

Figure 8 - Distribution of accuracies achieved on the dataset acquired in a common environment, by SVM with a linear kernel on classifying HV, NV and LV on intra and inter subject modalities using the self-assessment answers as ground truth on the dataset acquired in a common environment.

Table 17 - Performance metrics mean values and respective standard deviations, in brackets, achieved on the dataset acquired in a common environment, by SVM with a linear kernel on classifying High Valence, Neutral Valence and Low Valence, on intra and inter subject modalities using the CAAV database ratings as ground truth on the dataset acquired in a common environment. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject	**Intersubject**
Accuracy	0.37 (0.16)	0.35 (0.10)
Macro F1 - Score	0.36 (0.09)	0.28 (0.10)
p-value	0.50	0.53

Analysis A5 with PCA

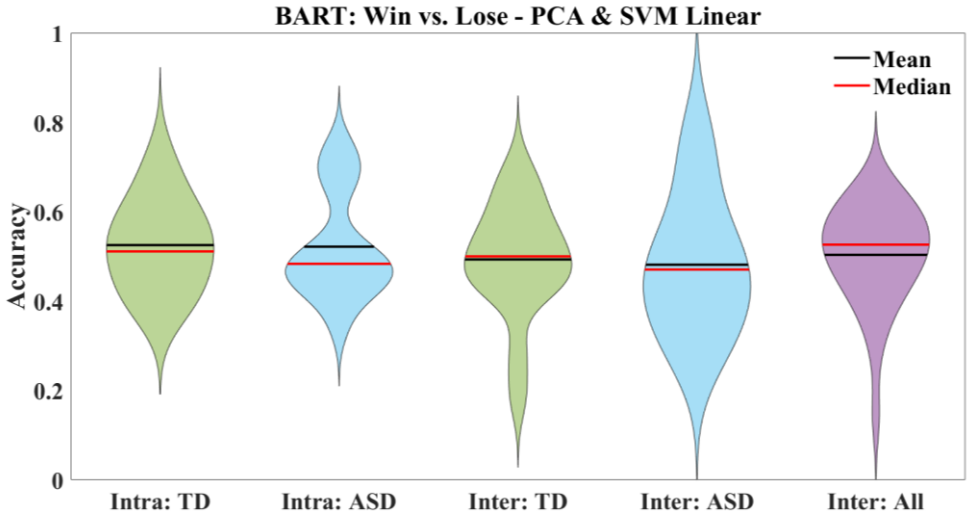

Figure 9 - Distribution of accuracies achieved using PCA for feature reduction and SVM with a linear kernel on classifying Winning vs Losing events on intra and inter subject modalities.

Table 18 - Performance metrics mean values and respective standard deviations, in brackets, achieved using PCA for feature reduction and SVM with a linear kernel on classifying Winning vs Losing events on intra and inter subject modalities. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject: TD	Intrasubject: ASD	Intersubject: TD	Intersubject: ASD	Intersubject: All
Accuracy	0.53 (0.10)	0.52 (0.12)	0.49 (0.13)	0.48 (0.15)	0.50 (0.11)
Sensitivity	0.48 (0.28)	0.55 (0.19)	0.49 (0.27)	0.38 (0.37)	0.37 (0.26)
Specificity	0.51 (0.29)	0.50 (0.27)	0.53 (0.19)	0.65 (0.29)	0.67 (0.26)
Precision	0.47 (0.23)	0.54 (0.13)	0.49 (0.17)	0.50 (0.25)	0.53 (0.20)
F-Measure	0.56 (0.14)	0.62 (0.09)	0.49 (0.13)	0.45 (0.21)	0.44 (0.14)
p-value	0.15	0.16	0.54	0.55	0.46

Analysis A6 with PCA

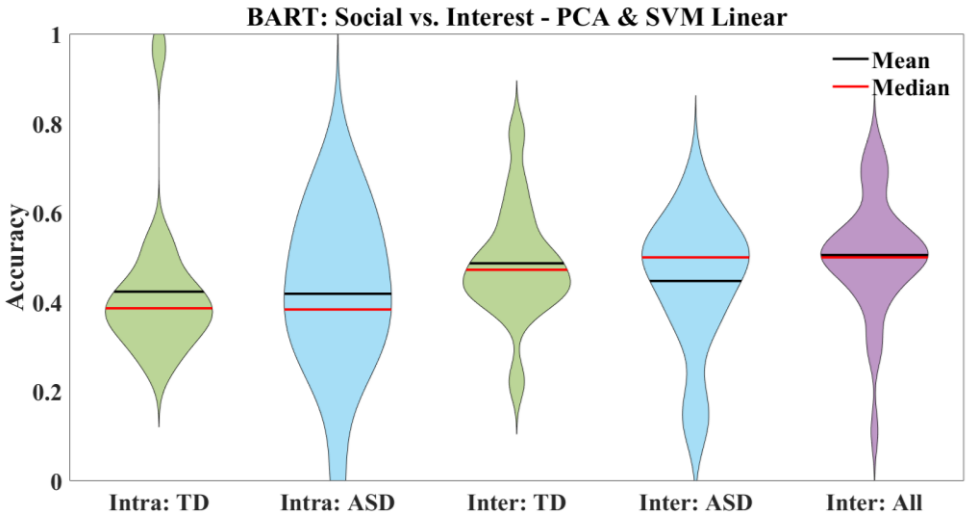

Figure 10 - Distribution of accuracies achieved using PCA for feature reduction and SVM with a linear kernel on classifying Social vs Interest reward on intra and inter subject modalities.

Table 19 - Performance metrics mean values and respective standard deviations, in brackets, achieved using PCA for feature reduction and SVM with a linear kernel on classifying Social vs Interest reward on intra and inter subject modalities. Last row corresponds to the p-value estimation obtained with the permutation tests.

	Intrasubject: TD	Intrasubject: ASD	Intersubject: TD	Intersubject: ASD	Intersubject: All
Accuracy	0.42 (0.16)	0.42 (0.18)	0.49 (0.13)	0.45 (0.15)	0.51 (0.13)
Sensitivity	0.41 (0.23)	0.48 (0.27)	0.32 (0.36)	0.24 (0.32)	0.11 (0.22)
Specificity	0.56 (0.21)	0.40 (0.24)	0.67 (0.26)	0.66 (0.42)	0.88 (0.24)
Precision	0.44 (0.18)	0.46 (0.24)	0.40 (0.22)	0.43 (0.30)	0.50 (0.31)
F-Measure	0.76 (0.16)	0.71 (0.14)	0.47 (0.21)	0.44 (0.17)	0.46 (0.20)
p-value	0.24	0.29	0.46	0.49	0.41

Appendix III - Assessing Arousal Through Multimodal Biosignals: A Preliminary Approach

Rita Correia, Daniel Agostinho, Isabel Catarina Duarte, Daniela Sousa, Ana Pina Rodrigues, Miguel Castelo-Branco and Marco Simões

Abstract

The increase in Autism Spectrum Disorder (ASD) prevalence estimates over the last decades has driven a quest to develop new forms of rehabilitation that can be accessible to a larger part of this population. These rehabilitation approaches often take the form of computer games that are blind to the user's emotional state, which compromises their efficacy. In this study, a set of physiological signals were acquired in simultaneous with functional Magnetic Resonance Imaging (fMRI) with the future prospect of combining both kinds of data to create models capable of assessing the true emotional state of their users based on physiological response as a measure of autonomic nervous system, having as ground truth the activity of targeted brain regions. This paper describes an initial approach, focusing on the information contained on the physiological signals alone. A total of 35 features were extracted from biosignals' segments and subsequently used for automatic classification of arousal state (High Arousal vs. Low Arousal). The suboptimal results, although some extracted features present statistically significant differences, underline the challenging nature of our proposal and the added obstacles of recording physiological signals in the magnetic resonance environment. Further exploration of the measured signals is needed to gather a bigger number of discriminative features that can improve classification outcomes.

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects social and communication skills, as well as normal patterns of behavior, interests and activities [1]. As of 2014, 16.8 per 1000 children aged 8 years were diagnosed with ASD in the United States of America (USA), which represents an increase of 150% when compared with 2000 estimates [2]. Hence, the number of people that can benefit from new and improved rehabilitation approaches is enormous and continues to rise.

Over the last decades, there has been an increasing interest in serious gaming as an alternative or complement approach to the traditional therapeutical interventions. A serious game is a game with an educational purpose, going beyond the sole purpose of entertainment. This rehabilitation tool represents a low-cost option that allows for the repeated practice of different skills that are usually impaired in the ASD population. Autistic individuals, however, generally present an increased sensory

sensitivity, which may compromise the full potential and efficacy of the serious games, the presentation of the wrong type and number of stimuli may lead to the disengagement of the user from the game, or even to a complete rejection of the intervention [3]. Therefore, the next step must be to optimize the serious games based on the emotional state of its users.

In this sense, this project aims to develop biofeedback-based models, particularly designed for the ASD population, that, based on autonomic nervous system (ANS) physiological signals, can infer the state of the user, having as ground truth the neuronal activation evoked by different emotion eliciting stimuli.

While physiology based automatic emotion assessment has been substantially considered for typically developed (TD) individuals [4]–[8], it is underexplored for ASD. To the best of our knowledge, there are only three papers describing automatic emotion classification in autistic subjects. By measuring EDA, PPG, skin temperature, EMG and ECG on children with ASD while they performed computerized cognitive tasks, Liu et al. (2008) successfully attempted to classify emotional states of liking, anxiety and engagement in this population, achieving accuracies of 82.9 % with Support Vector Machines (SVM) [9]. Kushki et al. (2015) classified anxiety-related arousal using metrics derived from the ECG and a modified Kalman filter obtaining an average specificity of 92% and sensitivity of 99% [10]. More recently, Sarabadani et al. (2020) automatically discriminated positive from negative valence during high and low arousal in ASD obtaining accuracies of 78.1% and 84.5% for high arousal and low arousal, respectively, using K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA) and SVM, and combining the outputs of all the classifiers using a Majority Vote to enhance the performance [11].

While these positive outcomes suggest that emotion recognition is a viable approach in ASD, the evidence of emotion dysregulation in this population [12] seems to be overlooked. This evidence means that the use of self-assessment questionnaire responses or labels based on the general population's emotional perception of a stimulus as ground truth is of limited accuracy. For this reason, we believe that, due to its spatial resolution that allows for the precise mapping of brain regions or networks of interest, functional Magnetic Resonance Imaging (fMRI) is the ideal true state indicator. Sessions involving this imaging technique, however, are quite expensive and nonportable, which limits their applicability. With this study, we intend to find ANS physiological patterns that are representative of the targeted brain regions modulation, so that it can be inferred outside the MR scanner.

To this end, respiration, photoplethysmography (PPG), electrodermal activity (EDA), electroencephalography (EEG), pulse oximetry (SpO2) and pupil size were recorded simultaneously with fMRI in ASD patients and a matched TD group, while watching short videos, chosen specifically to induce different kinds of emotional response.

This paper describes an initial approach to the experiment, which includes feature extraction from the physiological signals and subsequent binary classification into high or low arousal states. Given the early nature of this study, the data acquired from the fMRI are not yet considered. Instead,

the ratings of arousal from the database where the videos were taken from were used as classification labels.

2. Methods

A. Participants

Fourteen individuals with ASD (1 female), and 13 typically developed (TD) individuals (2 females) took part in this study. Participants (or their legal representatives) signed an informed consent to participate in the study. Every subject completed the entire task. Table I provides a detailed description of the participants.

Table XXI. Demographic description of the ASD and TD groups, including age, Full-Scale Intelligence Quotient (FSIQ), Empathy Quotient (EQ), Autism Spectrum Quotient (AQ) and the Autism Diagnostic Observation Schedule (ADOS-II) total score. Each score is presented in terms of group average and standard error, in brackets. Group differences were assessed with a two-sample T-test, with p-values on the last column. Groups are matched by age and EQ.

	ASD	**TD**	***P***
N	14	13	
AGE	21.58 (1.36)	23.15 (0.91)	**0.34**
FSIQ	94.50 (2.97)	111.23(4.30)	<0.01
EQ	38.45 (4.54)	45.62 (2.89)	**0.18**
AQ	24.17 (1.49)	15.38 (1.54)	**<0.01**
ADOS-II	11.17 (0.72)	-	-

B. Experimental Task

The task follows a block design. Each block consists of a 15 second video presentation trailed by a self-assessment period and is preceded by a rest period of approximately the same duration. The protocol is composed by 3 task runs, and each run is made up of 10 video trials.

The 30 videos (10 videos x 3 runs) were taken from the Chieti Affective Action Videos (CAAV) database and represent different actions, examples include hugging someone, stealing from another or simply hanging a jacket. For our experiment we chose to use videos recorded in the 1st person perspective and to coincide the gender of the participant to the main actor [7]. Each video in the database is accompanied by the mean rating of valence and arousal given by an evaluation group using a 9-point Self-Assessment Manikin (SAM). Consequently, each video falls into one of the following 3 categories: low valence and high arousal (LVHA); high valence and high arousal (HVHA); no valence and low arousal (NVLA). Thus, 10 videos of each category were selected to integrate the task.

For the self-assessment, the subjects were asked to rate the video they just watched also in the 9-point SAM scale. For this purpose, participants used a joystick.

Before each session, the task was explained and participants were asked to rate some training videos, to guarantee that both concepts of valence and arousal were understood, and that they knew how to operate the joystick.

C. Data Acquisition

EEG, EDA and SpO2 were acquired using the MP150 system and AcqKnowledge 4.2 software (BIOPAC Systems, Inc.). Respiration and PPG were recorded using the Physiological Measurement Unit of the MRI scanner (Siemens Healthcare) and pupil size was registered using the EyeLink 1000 Plus Eye Tracker with the long-range mount (SR Research Ltd.). Due to the hypersensitivity of the ASD population, we tried to simplify and reduce preparation time as much as possible, thus, EEG was acquired using only 3 electrodes, placed on the forehead, and either the right or left earlobe and temporal area. EDA was measured using 2 Ag/AgCl electrodes taped to the proximal phalanges of the index and middle fingers of the participant's nondominant hand. SpO2 and PPG were measured using a pulse oximetry and pulse finger sensors, respectively. Respiration was measured with a respiratory cushion attached to the participant using a respiratory belt. EEG, EDA and SpO2 were recorded with a sampling rate of 5000 Hz, PPG and Respiration were acquired at 400 Hz and pupil size at 500 Hz.

D. Signal Processing

Photoplethysmography

To reduce noise contamination, the PPG signal was bandpass filtered using a 6th order Butterworth filter with a lower cut-off frequency of 0.5Hz and a higher cut-off frequency of 20Hz.

The clean PPG signal was then used to compute the Heart Rate (HR) by identification of the PPG pulse peak. HR is affected by both, the sympathetic and parasympathetic nervous systems, and is one of the most popular measures when it comes to emotion assessment.

Electrodermal Activity

EDA data were high-pass filtered with a 0.5Hz cut-off frequency as it was being collected, and it was later low-pass filtered using a 5th order Butterworth filter with a 1Hz cut-off frequency.

Electroencephalography

As expected, the EEG recordings were considerably contaminated by MR gradient switch artifacts. To correct them, the FMRI Artifact Slice Template Removal (FASTR) algorithm from the FMRIB plug-in for EEGLAB (version 1.21) was used. Feeding the algorithm with the corrupted signal and the events for each slice acquired, it computes an average template for the artifact and subtracts it from the signal, locked to each slice trigger.

Signal Segmentation

Lastly, signals were divided into 30 second segments, time-locked to the beginning of each video, this way each segment includes the 15 seconds of the video, the self-assessment period, and some seconds after.

E. Feature Extraction

For each video trial, a total of 35 features was extracted from the different biosignals. To account for possible carryover effects, the value of the feature for the last 5 seconds of the previous rest period was subtracted after extraction.

For each signal, a brief description of the extracted features is given in Table II.

Table XXXIII. List and description of the features extracted from the physiological signals.

Signal	Feature Name	Extracted Features
EDA	meanEDA	Mean
	maxEDA	Maximum
	minEDA	Minimun
	mean_abs_fd_EDA	Mean Absolute First Difference
	mean_deriv_neg_EDA	Mean of Derivative for Negative Values
PPG	meanPPG	Mean
	maxPPG	Maximum
	minPPG	Minimum
	NNmean	Mean of the Normal-to-Normal (NN) time intervals
	SDNN	Standard Deviation of NN intervals
	SDSD	Standard Deviation of Successive Differences between NN intervals
	RMSSD	Root Mean Square of Successive Differences between NN intervals
	NN50	Number of Successive Differences greater than 50ms
	pNN50	Ratio between NN50 and total number of NN intervals
EEG	delta, theta, alpha, beta, gamma	Relative Power (delta, theta, alpha, beta and gamma bands)
Respiration	meanResp	Mean
	maxResp	Maximum
	minResp	Minimum
	mean_abs_fd_resp	Mean Absolute First Difference
Pupil Size	meanPupilSize	Mean
	maxPupilSize	Maximum
Heart Rate	meanHR	Mean
	maxHR	Maximum
	minHR	Minimum
	VLF	Relative Power (Very Low Frequency, Low Frequency, and High Frequency bands)
	LF	
	HF	
	RaLH	Ratio between Low and High Frequency Powers
SpO2	meanSpO2	Mean
	maxSpO2	Maximum
	minSpO2	Minimum

F. Statistical Analysis

In a first approach, with the intention of inspecting the extracted features for significant differences among conditions and groups, the means of every feature for each subject were computed, for the conditions of High Arousal (HA) and Low Arousal (LA). Given its' subjective nature, for this preliminary study, the valence dimension was not considered. The HA and LA conditions were then obtained by condensing the 3 original ones (LVHA, HVHA, NVLA) and were defined considering both the database ratings and the self-assessed arousal values given by each participant. Using the database ratings, a trial was labeled as LA if the arousal rating for the corresponding video was lower than 4 and labeled as HA otherwise. For the self-assessment, k-means clustering was performed on each participant's answers individually to partition them into 2 clusters. Trials were then classified as HA or LA based on the cluster they fell into. Wilcoxon signed rank tests were then performed to look for statistically significant differences in feature values between HA and LA, for each group, and Wilcoxon rank sum tests were applied to look for differences between groups for the two conditions.

G. Classification

In order to explore the accuracy of automatic emotion assessment in the data acquired with our experimental protocol, 4 classification algorithms were applied. The ratings of the CAAV database for arousal were used to label the data.

The considered classifiers were a Euclidean Minimum Distance Classifier (MDC – Euclidean), a K-Nearest Neighbors (KNN) and SVM using a Radial Basis Function (SVM RBF) kernel and a linear (SVM Linear) kernel. The optimal parameters for the KNN (number of neighbors, K) and SVM (cost, C and Kernel Parameter, γ) were determined by applying a 50/50 partition on the training set 5 times and choosing the parameters that resulted in the smallest classification error. After the hyperparameters were selected, the classifiers were retrained with all training data for the chosen parameters.

The classifiers were then tested for both intraparticipant and interparticipant classification. For the intraparticipant

approach, data from each subject were randomly split using the 70:30 ratio, where 70% of the data were used to train the classifier and the remaining data were used for testing. This process was repeated 30 times to avoid outlier results. As for the interparticipant classification, we employed the Leave One Subject Out (LOSO) method where the data from each participant are used once for testing, while the rest of the participants' data are used to train the classifier.

Finally, to ascertain if the accuracies of the classifiers were significantly greater than chance level (50%), permutation tests were used. For each partition, after testing, the true labels of the test set were iteratively shuffled, and accuracies were calculated using the random labels as the predicted

classes. The number of times that these accuracies were greater than the one obtained with the classes predicted by the classifier, were then counted.

3. Results

The significance levels that resulted from the statistical analysis are present in Tables I and II, for comparisons between conditions and groups, respectively. Only the features that suggest statistically significant differences for at least one scenario are shown in each table.

Apart from the MDC – Euclidean which returned poor accuracies, not significantly higher than chance level, all other 3 classifiers exhibited similar results on classifying HA vs LA. While accuracies for the intraparticipant classification for both groups are highly variable and for the most part, not significantly higher than chance, the interparticipant outcomes have a narrower distribution and exhibit accuracies higher than random chance more than 80% of the times. The median accuracy value for all 3 interparticipant modalities is of approximately 60%. To illustrate these findings, classification results for SVM Linear on intra and inter subject modalities on classifying HA vs LA are displayed in Fig.1.

Table XXIII. p-values of pairwise comparisons from Wilcoxon signed rank test (High Arousal compared to Low Arousal)

Feature	Database		Self-Assessment	
	Clinical (N = 14)	***Control (N = 13)***	***Clinical (N = 14)***	***Control (N = 13)***
meanPPG	0,71	0,31	**0,01**	0,19
minPPG	0,19	0,15	0,76	**0,02**
maxResp	0,67	**0,05**	0,67	0,17
meanPupilSize	0,67	**0,03**	0,76	1,00
meanHR	**0,01**	0,38	0,12	0,84
maxHR	**0,02**	0,45	0,24	0,84
minHR	**< 0,01**	0,59	**0,01**	0,95
RaLH	**0,04**	0,95	**0,01**	0,68

Table XXIV. p-values of pairwise comparisons from Wilcoxon rank sum test (Clinical Group compared to Control Group)

Feature	Database		Self-Assessment	
	High Arousal	***Low Arousal***	***High Arousal***	***Low Arousal***
minPPG	0,87	**0,04**	0,72	**0,03**
delta	**0,05**	**0,01**	**0,04**	**0,03**
beta	**0,03**	**0,01**	**0,03**	**0,03**
gamma	0,08	**0,04**	**0,05**	0,10
meanHR	**0,01**	0,68	**0,01**	0,17
maxHR	**0,01**	0,37	**0,01**	0,72
minHR	**0,01**	0,87	**0,01**	0,15

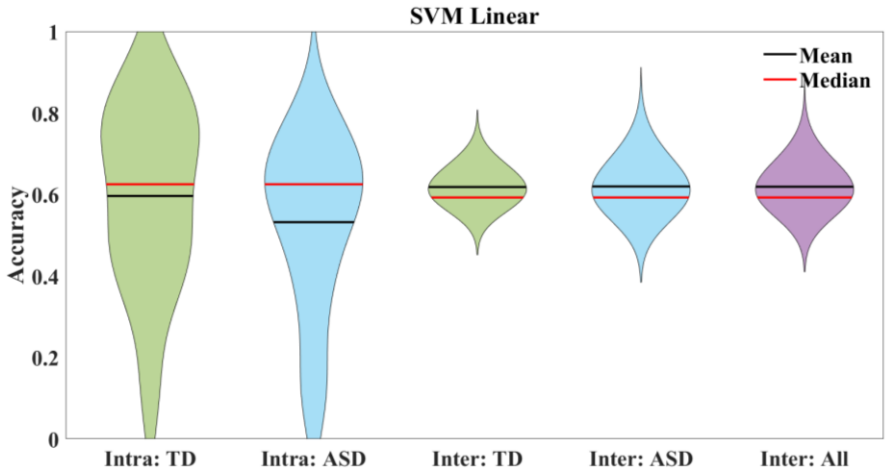

Figure 17. Distribution of accuracies achieved by SVM on classifying HA vs LA on intra and inter subject modalities.

4. Discussion

The statistical analysis revealed that only a limited number of extracted features present statistically significant differences between HA and LA conditions as well as between groups. The time-domain features of the HR (meanHR, maxHR and minHR) seem to be significantly different between conditions for the clinical group, and also between groups for the HA condition, which is in accordance with the known relation of this signal with emotion discrimination [13].

These results help to explain the low classification accuracies and reveal the low discriminative power of the extracted features for distinguishing between HA and LA.

A possible explanation for the poor results of the intraparticipant modality is the low number of observations for each individual participant, which limits the generalization capacity of the models.

The simultaneous acquisition of physiological signals and fMRI represents a great challenge. Besides the common noise sources, there is the added artifact caused by the gradient switch of the MRI scanner that severely contaminates most of the recordings and results in low signal-to-noise ratios (SNR). This interferes with the quality of the features derived from the signals and hinders the appearance of subtle differences between states, which could be valuable for their distinction.

This preliminary approach reiterated the challenging nature of this project and highlighted the need to further explore the biosignals in order to find meaningful features that can optimize the classification results.

References

[1] C. P. Johnson and S. M. Myers, “Identification and Evaluation of Children With Autism Spectrum Disorders,” *Pediatrics*, vol. 120, no. 5, 2007, doi: 10.1542/peds.2007-2361.

[2] J. Baio *et al.*, “Prevalence of autism spectrum disorder among children aged 8 Years - Autism and developmental disabilities monitoring network, 11 Sites, United States, 2014,” *MMWR Surveill. Summ.*, vol. 67, no. 6, 2018, doi: 10.15585/mmwr.ss6706a1.

[3] S. H. Baum, R. A. Stevenson, and M. T. Wallace, “Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder,” *Prog. Neurobiol.*, vol. 134, pp. 140–160, 2015, doi: 10.1016/j.pneurobio.2015.09.007.

[4] J. Kim and E. André, “Emotion recognition based on physiological changes in music listening,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067–2083, 2008, doi: 10.1109/TPAMI.2008.26.

[5] K. H. Kim, S. W. Bang, and S. R. Kim, “Emotion recognition system using short-term monitoring of physiological signals,” *Med. Biol. Eng. Comput.*, vol. 42, no. 3, pp. 419–427, 2004, doi: 10.1007/BF02344719.

[6] R. W. Picard, E. Vyzas, and J. Healey, “Toward machine emotional intelligence: analysis of affective physiological state,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, 2001, doi: 10.1109/34.954607.

[7] C. L. Lisetti and F. Nasoz, “Using noninvasive wearable computers to recognize human emotions from physiological signals,” *EURASIP J. Appl. Signal Processing*, vol. 2004, no. 11, pp. 1672–1687, 2004, doi: 10.1155/S1110865704406192.

[8] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, “Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty,” *IEEE Trans. Syst. Man, Cybern. - Part A Syst. Humans*, vol. 41, no. 6, pp. 1052–1063, Nov. 2011, doi: 10.1109/TSMCA.2011.2116000.

[9] C. Liu, K. Conn, N. Sarkar, and W. Stone, “Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder,” *Int. J. Hum. Comput. Stud.*, vol. 66, no. 9, pp. 662–677, 2008, doi: 10.1016/j.ijhsc.2008.04.003.

[10] A. Kushki, A. Khan, J. Brian, and E. Anagnostou, “A Kalman filtering framework for physiological detection of anxiety-related arousal in children with autism spectrum disorder,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 3, pp. 990–1000, 2015, doi: 10.1109/TBME.2014.2377555.

[11] S. Sarabadani, L. C. Schudlo, A. A. Samadani, and A. Kushski, “Physiological Detection of Affective States in Children with Autism Spectrum Disorder,” *IEEE Trans. Affect. Comput.*, vol. 11, no. 4, pp. 588–600, 2020, doi: 10.1109/TAFFC.2018.2820049.

[12] A. C. Samson, J. M. Phillips, K. J. Parker, S. Shah, J. J. Gross, and A. Y. Hardan, “Emotion

dysregulation and the core features of autism spectrum disorder," *J. Autism Dev. Disord.*, vol. 44, no. 7, pp. 1766–1772, 2014, doi: 10.1007/s10803-013-2022-5.

[13] P. Rainville, A. Bechara, N. Naqvi, and A. R. Damasio, "Basic emotions are associated with distinct patterns of cardiorespiratory activity," *Int. J. Psychophysiol.*, vol. 61, no. 1, pp. 5–18, 2006, doi: 10.1016/j.ijpsycho.2005.10.024.