FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Maria da Paz Olímpio Lardosa Paz

Analysis of eye tracking data applied to Autism Spectrum Disorder during virtual reality experiments

Thesis submitted to the
University of Coimbra for the degree of
Master in Biomedical Engineering

Supervisors:
Prof. Dr. Miguel Castelo-Branco (Advisor)
Carlos Amaral (Co-Advisor)

Faculty of Medicine, University of Coimbra, Coimbra, Portugal
CIBIT, Coimbra Institute for Biomedical Imaging and Translational Research, ICNAS - Institute of Nuclear Sciences Applied to Health, University of Coimbra, Coimbra, Portugal

Coimbra, 2019

This work was developped in collaboration with:

Faculty of Medicine, University of Coimbra

ICNAS - Institute of Nuclear Sciences Applied to Health, University of Coimbra

Acknowledgments

Em primeiro lugar cabe-me agradecer ao meu orientador, o Professor Doutor Miguel Castelo Branco, pela oportunidade de me deixar trabalhar num tema que me interessa, pelo feedback, que nos faz pensar e chegar mais longe e pelo incentivo, de tentar sempre questionar e fazer melhor.

Ao meu co-orientador Carlos Amaral, aluno de Doutoramento, agradeço o apoio contínuo, a paciência com as minhas dúvidas, mesmo quando os seus afazeres eram outros, a compreensão ao longo de todo este processo e toda a discussão e aconselhamento que me levaram a um enriquecimento pessoal e académico. Desejo-lhe a maior das sortes para este próximo desafio que se avizinha.

Durante estes 5 anos foram muitos os desafios, muitas as noitadas a estudar e a fazer trabalhos ou a preparar eventos, muitos os momentos de stress mas também muitos os momentos de partilha e de companheirismo. Sempre fui apologista de que devemos agarrar as oportunidades que nos aparecem, mesmo que tenhamos que fugir um pouco da nossa zona de conforto. Por isso, agradeço ao Núcleo de Estudantes do Departamento de Física, por se ter tornado a minha segunda casa e pelas portas que abriu, e ao João e à nossa equipa, por juntos termos partilhado e ultrapassado tantos desafios. Agradeço também à Federação Portuguesa da Canoagem por me ter "acolhido" neste último ano. Não consigo descrever o quão gratificante e enriquecedor é trabalhar ao lado de pessoas que o fazem com e por gosto.

Mas não há nada como os amigos. À Joana, à Maggie e ao Oshley, por serem a minha companhia nestes últimos cinco anos, por respeitarem a minha falta de tempo e longas ausências e por me apoiarem em todas as decisões. Pelas longas conversas, pelos bons e maus momentos, pelas gargalhadas e pelo apoio mútuo. À Bia, à Inês, à Maria, à Laura e à Luísa, pela amizade de quase 20 anos, que não se perde, por maior que seja a distância. À Carriço, à Mónica e ao Dani, pela improbabilidade da amizade que se foi construindo entre nós, pelo companheirismo, pelo apoio e pela confiança. À Rute, às gémeas, à Ana, à minha afilhada Inês, e a tantas outras pessoas, que poderia e que teria que agradecer individualmente e que, à sua maneira, marcaram e marcam a pessoa que sou hoje.

Por último, mas não menos importante, à minha família. Ao meu pai e aos meus irmãos, que sempre me apoiaram nas minhas decisões, mesmo quando isso implicava passar menos horas em casa. Em especial ao meu pai, por abdicar de si em prol de nós. Aos meus avós, por sempre olharem por mim, independentemente da distância, pela mensagem ou chamada amiga e por nunca deixarem que nos faltasse nada. À Tetê, à Mima e à Guida, por serem dos maiores marcos na minha educação a todos os níveis, sem a qual não seria a pessoa que hoje sou. A todos, pelos momentos bons e maus que ultrapassámos juntos, por me ajudarem a crescer e por serem o meu suporte na luta pelos meus sonhos.

"Caminante, son tus huellas el camino y nada más; Caminante, no hay camino, se hace camino al andar. Al andar se hace el camino, y al volver la vista atrás se ve la senda que nunca se ha de volver a pisar."

Antonio Machado

Resumo

As Perturbações do Espetro do Autismo (PEA) são uma desordem do neurodesenvolvimento, caracterizadas, entre outros sintomas, por déficits no comportamento social, como a comunicação e a interação. Estes sintomas podem estar associados a falhas na compreensão e uso de linguagem corporal e gestos, e com dificuldades em fazer e manter contacto visual e seguir pistas visuais.

Devido à sua heterogeneidade de sintomas e aos seus diferentes níveis de manifestação, o diagnóstico das PEA torna-se muito desafiante e subjetivo, requerendo sempre a presença de um clínico com treino na área e de uma equipa interdisciplinar, responsável por avaliar tanto o desenvolvimento como o comportamento do indivíduo.

Com os avanços da tecnologia, têm sido usadas novas ferramentas no estudo das PEA. Um destes exemplos são os *eye trackers*, usados para estudar e compreender o comportamento do olhar, do qual está provado existirem diferenças entre indivíduos com Desenvolvimento Típico (DT) e indivíduos com PEA. Entre outras diferenças no olhar, a população autista tem tendência a realizar mais sacadas, fixações mais longas e menos fixações na cara. Para além disso, apresenta também falhas em fenómenos de atenção conjunta.

Neste projeto, estas características foram calculadas a partir dos dados recolhidos durante experimentos de realidade virtual, com dois dispositivos de *eyetracking*, que diferiam no seu nível de imersividade, o *RED* (ecrã com *eye tracker*) e os *Oculus Rift* (óculos de *eyetracking*).

Para analisar e treinar os algoritmos com os dados obtidos, utilizámos técnicas de *Machine Learning (ML)*. Isto permitiu-nos, não só, perceber se estas características são ou não capazes de distinguir as duas classes de indivíduos (PEA e DT), mas também se existem diferenças significativas na aquisição com diferentes dispositivos.

Palavras-chave: Perturbações do Espetro do Autismo (PEA), *eye-tracking*, *Machine Learning (ML)*

Abstract

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized, among others by deficits in social behaviour, like communication and interaction. These symptoms can be associated with impairments in understanding and using body language and gestures, in making eye contact and in following gaze cues.

Due to its heterogeneity of symptoms and different levels of manifestation, the diagnostic of ASD becomes very challenging and subjective, and requires the presence of a trained clinician, and an associated interdisciplinary team, in charge of evaluating both development and behaviour.

With technology advances, different tools have been used in the study of ASD. One of these examples is eye trackers, used to study and understand gaze behaviour, which is proved to show differences between Typically Developing (TD) and ASD individuals. Among other gaze differences, the autistic population show more saccadic movements, longer fixations, lesser fixations in faces and impairments in Joint Attention (JA) phenomena.

In this project we calculated these features from the data obtained from virtual reality experiments with two eyetracking devices, which differ on their levels of immersivity, RED (screen-based eye tracker) and Oculus Rift (eye tracking glasses).

We used Machine Learning (ML) techniques to analyse and train the algorithms with the obtained data. This allowed us, not only to understand if these features were able to distinguish between the two classes of individuals (ASD and TD), but also if there was significant differences in the acquisition with different setups.

Keywords: Autism Spectrum Disorder (ASD), eye-tracking, Machine Learning (ML)

List of Figures

List of Tables

List of Acronyms

ASD - Autism Spectrum Disorder

DSM-V - Diagnostic and Statistical Manual of Mental Disorders

TD - Typically Developing

JA - Joint Attention

ROI - Regions Of Interest

RFE - Recursive Feature Elimination

PCA - Principal Component Analysis

AI - Area of Interest

SVM - Support Vector Machine

TPR - True Positive Rate

FPR - False Positive Rate

ML - Machine Learning

RBF - Radial Basis Function

Contents

1

Introduction

Autism Spectrum Disorder (ASD), its definition, symptoms and causes, have been largely debated since its discovery, in the early 40s [1]. Its current definition can be found, in the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) by the American Psychiatric Association [2]. Many theories are still discussed, concerning the variety of symptoms that are believed to be a part of this heterogeneous syndrome [3]. Those are included in a wide range of manifestations, including communication, social cognition deficits and repetitive behaviour [2]. Despite all the evolution in the understanding of this spectrum there is still a lot of ground to break, in particular from the point of view of eyetracking technologies, which brings us to the work of Klin and his studies with visual fixation patterns in autistic individuals, using eye tracking data [4][5] or the work of Wall, who applies Machine Learning to "shorten observation-based screening and diagnosis of autism" [6]. Both of their works, and several others, make use of new tools and technologies, that can be useful when it comes to diagnose or understand neurodevelopmental disorders.

1.1 Contextualization

ASD is a neurodevelopmental disorder characterized by deficits in social behaviour, which symptoms are usually difficul to detect before 36 months of age, and with no certain cause nor cure, even though some treatments can improve symptoms and ability to function [2]. Diagnosing this pathology can be challenging, since it takes the presence of a trained clinician, and an associated interdisciplinary team, responsible to evaluate both development and behaviour in a subjective way [3]. One of the symptoms that can characterize ASD, is deficits in communication and interaction, which can be associated with difficulties in making eye contact and an impairment in understanding and using body language and gestures [2]. Considering this we can introduce the importance of human eyes as mean to express our motives and interests, and to give information about the attentional focus of a social interaction partner. They can also help us connect with other people and sometimes understand the course of their actions [7], playing an important role in our daily lives. According to Baron Cohen one of the core impairments in ASD is Gaze-Monitoring, which is the ability to look to the same direction as the adult [8] and Joint Attention behavior, which includes not only gaze-monitoring but also pointing or showing, in order to direct someone's attention towards an object or an event [9][10][11].

With this information, it is important to investigate how individuals with ASD show social behaviour differences, when compared to typically development subjects. By making use of virtual reality experiments, that can mimic real life situations, we can acquire eye tracking data, which enables the quantification of eye gaze behaviour, and obtain and analyse features that could possible work as a biomarker for ASD. These tools can potentially also be explored, as biofeedback therapeutic tool in disorders such ASD [4] and to understand impairments in gaze behaviour.

1.2 Motivation

Due to its prevalence, heterogeneity and lack of precise diagnostic tools, and largely unknown underlying neurobiology, ASD is a disorder definitely worth being studied. It is proven that social interaction and JA are impaired in people with ASD. These two phenomena can be studied and evaluated through eye movements analysis, by means of eye tracking methodologies.

1.3 Goals

The main goal of this project is to analyse eyetracking data obtained during virtual reality experiments with people with and without ASD, and identify a potential biomarker for this pathology. Following this, we can divide the work into 3 main questions:

- Which features can we obtain from data containing eye positions and their respective time stamps?
  - To answer this question we searched for metrics obtained from eyetracking data, and which are used to analyse people's behaviour or intentions.

  Also, we investigated eye gaze and joint attention, as an indicator for differences between participants with and without ASD.

- Is any of those features "good" enough to define a potential biomarker for ASD?
  - Here we tested the presence of features that could be extracted from eye tracking data, and that could be used to distinguish both classes.
- Can we predict or confirm the disorder, by analysing eyetracking data with a machine learning approach?
  - We used machine learning algorithms to predict the correspondent class of the eyetracking features, that we could eventually extract from the data.

1.4 Structure

Starting with this chapter, Chapter 1 - Introduction, the document is organised as follows: Chapter 2 focus on the background knowledge that supports this work. In chapter 3 we present the methods used to approach our problem. Then, in chapter 4, we show the obtained results, which are then discussed in chapter 5. Finally, this document ends with chapter 6, where we present the final conclusions of our work, with some possibilities for future work.

2

Background Knowledge

2.1 Autism Spectrum Disorder (ASD)

Being an early neurodevelopmental disorder, ASD is manifest during the first years of life, but is often difficult to diagnose before the 36 months of age [2] and its symptoms can change during development. It is characterized by deficits in social communication and interaction, accompanied by restrictive interests and repetitive behaviours, which can co-occur with other neurodevelopmental disorders [2].

As mentioned before, ASD individuals, among other impairments, have severe difficulties in dealing with the social parts of their lives, which includes understanding and maintaining relationships, understanding and reciprocating nonverbal communication, abnormalities in making eye contact, and impairments in understanding body language or gestures [2]. Despite these symptoms they can not give a viable diagnosis for ASD, when considered alone[2]. This way it is important to consider several sources of information, to search for specific parameters that can possibly give more objective diagnosis criteria. We will add them importance for analysis purposes.

2.1.1 Prevalence

According to Oliveira *et al.* [12] and her study about prevalence of ASD in Portugal, its value is around 0.1% but the current prevalence seems nowadays to be much larger, which can be seen in figure 2.1.

1% is also the average percentage worldwide, both in children and in adults [2][13], which means that, 1 in every 100 people has this disorder, making it an important condition and societal challenge.

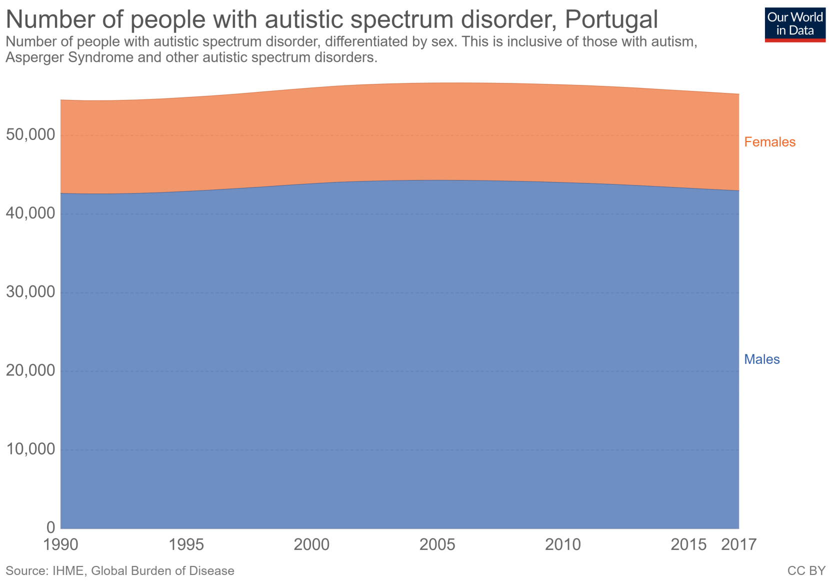

Figure 2.1: Number of people with Autism Spectrum Disorder in Portugal, divided by gender. Retrieved from *Our World in Data*[13]

2.1.1.1 Variability in Worldwide Prevalence

In figure 2.2 we can observe that, the percentage of individuals with ASD, is higher in developed countries, like the United States of America or European countries, which includes Portugal. Nevertheless, this does not necessarily mean that these countries, have a higher percentage of ASD individuals, could only mean that, there is larger public awareness, more diagnostic instruments are available so that the disorder goes unrecognized in underdeveloped countries. It is important to consider that, cultural differences also mean different norms for social interaction and communication and different ways of assessing differences in spite of standardization efforts.

2.1.1.2 Gender-Related Prevalence

When it comes to prevalence related to gender, ASD is diagnosed in more males than females. Epidemiological evidence shows that, ASD is diagnosed four times more often in males [2]. In figure 2.3, we can see the difference between male and female individuals, in the autistic population, which confirms the data from the Diagnostic and Statistical Manual of Mental Disorders.

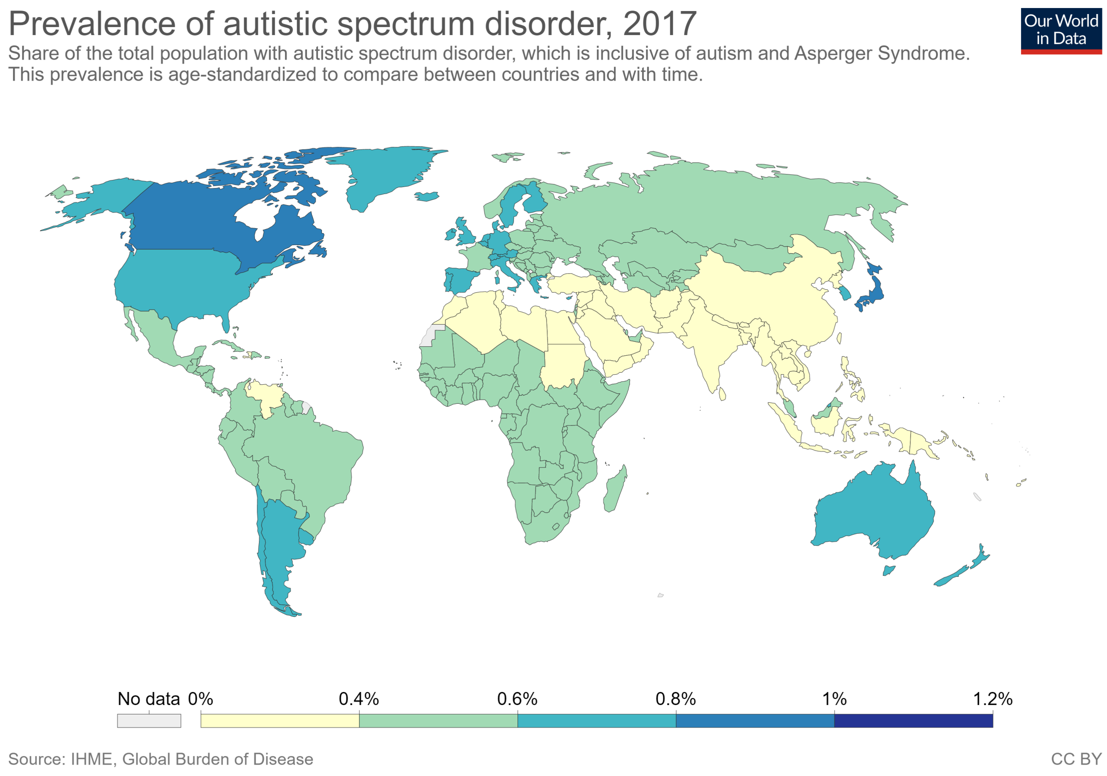

Figure 2.2: Prevalence of Autism Spectrum Disorder in 2017. Retrieved from *Our World in Data*[13]

2.1.2 Causes

Despite not having a known cause, ASD can be associated with some risk factors, like advanced parental age or low birth weight [2]. The literature relates this disorder with high heritability, ranging from 37% to above 90%, and, according to the DSM-V, up to 15% of the cases are associated with a well known mutation in specific genes, related to the disorder in specific families [2].

2.1.3 Diagnosis

ASD is often not diagnosed before the 36 months of age. In some cases, it can be diagnosed earlier, if the symptoms are too severe, or after, if they are more subtle [2]. The first symptoms are associated with lack of social interaction, or with loss of social and language skills. These symptoms are normally based on information given by the parents, or other relatives.

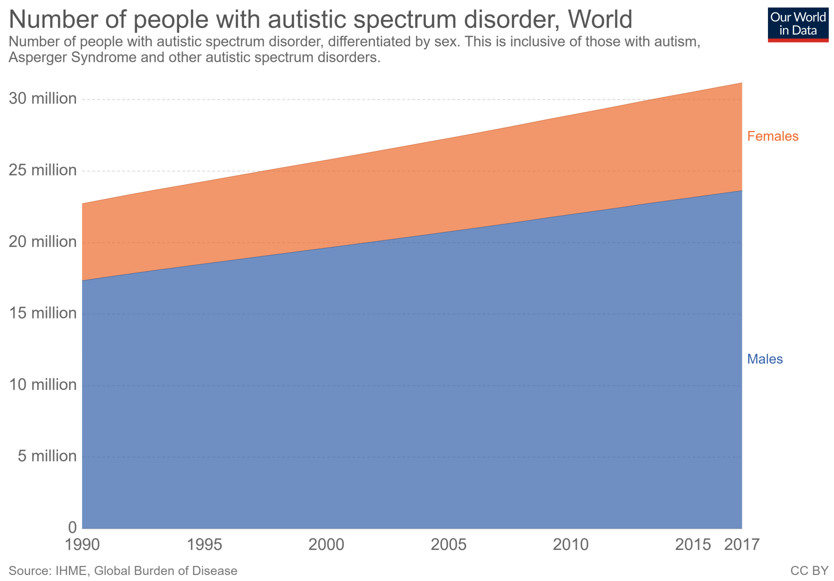

Figure 2.3: Number of people with Autism Spectrum Disorder Worldwide, gender-separated. Retrieved from *Our World in Data*[13]

2.1.3.1 Diagnostic Tools

ASD diagnosis is mostly based on direct or indirect observation, of the individual behaviour [2][14]. The most common diagnostic instruments include the Autism Diagnostic Interview (ADI), the Diagnostic Interview for Social and Communication Disorders (DISCO), the Autism Diagnostic Observation Schedule (ADOS), and their variants [14]. Due to the subjectivity of the diagnosis, a trained specialist is always needed, and a combination of the different tools is recommended [14].

2.2 Visual Attention

The ability to attend to and interpret our surroundings and different stimuli, plays an important role in our daily lives. Without it, we would not be able to detect danger, perceive people's emotions and intentions, or even their actions. But not all we see reaches awareness, and neither can we focus on many objects simultaneously [15], although divided attention is viable for a few objects.

Our brains have the ability to select relevant sensory information, and filter the remaining. This operation refers to "attention" [16]. Attention may be exogenous, resulting from stimuli, not necessarily visual, or endogenous, as a mean to express our interest or curiosity [15] [17]. This concept and its implications represent a large field of study [18]. Many studies include the relation between visual attention, human cognitive neuroscience and infant neurodevelopment [19], important in the study of ASD.

Despite showing development throughout the lifespan, it is during the childhood and adolescence, that attentional processing has its major changes. Around the 2 or 3 years of age, the attention system starts being highly engaged in cognitive, social and emotional tasks [20]. In typical development, basic attentional processes may be a critical foundation for sociocommunicative abilities [21].

Attention can be divided into overt attention and covert attention. The former happens when a person looks directly to what caught his/her attention, which implies eye movements. An example of overt attention is joint attention, a non-verbal coordination of attention of two individuals toward a third object or event [22]. On the other hand, covert attention does not imply eye movement, and therefore not a direct look [16]. A good example of covert attention is what happens when we are driving. Our attention is on the road, but if a pedestrian approaches, we perceive and react without the need to look directly to him/her [16].

2.2.1 Joint Attention

As mentioned in the beginning of this section (section 2.2), Joint Attention (JA) is a non-verbal coordination of attention of two individuals toward a third object or event. It is associated with the development of social communication skills [22], and with the perceiving of people's intentions [23].

It is normal to attend the location in space that is being looked at by another person. This JA phenomena appears during the first months of life. In the work of Bayliss *et al.*, they found that objects that are looked at by other people, have more attention focus than other objects. They also state that this, underlines the importance of other people's relation with objects, in a way that this social interaction, impacts our own impressions, and our own perceiving of the world [24]. JA is also in the basis of learning through imitation [25].

Both communication and the perceiving of people's intentions, are impaired in people with ASD. More specifically, they show deficits in JA abilities, which can be associated with their lack of social interaction skills [10] [11] [25] [26].

2.2.2 The study of attention in ASD individuals

Attention has been a subject largely studied in individuals with ASD, and gaze metrics can be considered markers of this disorder [27][26]. Early attentional dysfunction in ASD, may be related with atypical development of social communication.

In TD children, basic attentional processes may be highly related with the foundation of sociocommunicative abilities. Therefore, the dysfunction in these processes in ASD, may be the cause of the atypical development of social communication [21][28].

Besides this conclusions, it has been shown a preference for objects or other non-social stimuli over faces, other people's actions, and speech, in people with ASD [28][29]. Problems with re-directing attention in these individuals, in both social and nonsocial domains have also been identified [28].

In 1998 Dawson *et al.* [30], implemented a social orienting task, with both social and non-social sounds. They concluded that, children with ASD oriented less frequently to both stimuli, when compared to the control children. Later, in 2004, Dawson *et al.* [31] replicated these findings, revealing impaired social orienting and JA phenomenon in children with ASD. These results were confirmed with several other studies, in the following years, and a lot is yet to be understood concerning the underlying mechanisms [32]. This lack of overt social orienting, is considered one of the most important and relevant features, when it comes to differentiate TD and ASD individuals [33][32].

Features with social context have been studied, and the findings suggest that people living with ASD tend to look less at the eyes, in faces that are expressing emotions, than TD individuals [34][32] or, at least, they look and perceive them differently [35].

Some of these studies were made by simply analysing home videotapes of children's behaviour, where social attention could be studied in real-life situations [32].

2.3 Eye Tracking

Eye tracking allow us to record eye movements, and its ability to function both as a therapy and as a diagnostic tool, has been investigated. Klin has been using this tool to study ASD [4][28], and several articles can be found including this technology and the study of this neurodevelopmental disorder.

By measuring eye movements, eye tracking can be used as a mean to understand and study abilities related with vision, social cognition, interaction and empathy between people and their expression of interests. It can also help us understand the focus of someone's attention and their intentions[7].

An eye tracker records the position of the eyes and their movements. It generally works with near-infrared light, directed to the pupils, which will cause detectable reflections in both the pupil and the cornea. An infrared camera will then track these reflections and return the position of the eyeball.

We can divide these devices in two categories, screen-based eye tracking and eye tracking glasses. The former are the most common among studies with ASD, and require respondents to sit in front of the screen, where the experiment is happening. This allows the recording of data from observations with screen-based stimuli, but its limitations regarding the size of the screen, make these devices seem less like a real life scenario.

On the other hand, the latter, are mobile devices, which are placed near the eyes, just like glasses, allowing complete free movement. This eases the study of eye behaviour in more natural scenarios, without space constraints.

2.3.1 Importance

As mentioned above, human eyes can be a mean to express our interests, and can give information about the main focus of someone's attention. They can also help us connect with other people, and sometimes understand the course of their actions [7]. Eye-tracking allow us to measure eye movements, and it can help us understand social relevant events.

Eye-tracking devices have several advantages, making them an appealing new tool, with a lot of information to explore. They are non-invasive, practical to use and not too expensive.

2.3.2 Eye Movement Analysis

Eye movements are relatively well understood from the neurophysiological point of view [36]. This make them a good candidate to differentiate people, more specifically ASD individuals from TD individuals [26].

When someone intends to study and analyse eye movements, there are some important concepts to consider, and which compose the scanpath over time. Those are, among others, fixations and saccades.

Fixations refer to a relatively stable eye position. The fovea is centered on an object or region of interest. This type of eye movement provides visual acuity [36].

On the other hand, saccades represent the movements between points of fixation. These movements are quite fast, and do not have high visual acuity [36].

The autistic population show unusual fixation trajectories during social scenes, with a high level of complexity [4]. They also show a higher saccade frequency, despite the complexity of the visual task [37] or, at least, abnormalities in saccadic movement [38].

Robert M. Joseph *et al.* found that fixation duration in autistic children were shorter than the ones in TD children. But the number of fixations and their spatial distribution were similar to both groups [39].

When it comes to saccades, the most common saccadic movement metrics include amplitude (size), velocity and duration (time until the target is reached). The relationship between these metrics, is well known in TD individuals. For example, the relationship between amplitude and duration is linear in those subjects [40].

2.3.3 Eye tracking methodologies in individuals with Autism Spectrum Disorder

Eye-tracking can be very useful as a possible new tool to study and understand better ASD.

Recent studies using eye-tracking technologies, examined orienting behaviour during visual attention tasks, which included both social and non-social stimuli, in people with ASD[4][28].

More specifically, Klin *et al.*, used videos representing both social and non-social scenes, to study the looking patterns of children with ASD. They found that, young adults spend more time looking to objects than to faces, when compared to controls. They also related longer fixation times on objects with a higher level of social impairment [4][28].

Other studies looked upon action perception in ASD, founding that people with this disorder, have deficits in understanding the intention of other people's actions [41]. Vivanti [42] found that children with ASD have an impairment in interpreting JA phenomena. These children had difficulties in relating action and intention, when another person turned his head to an object, as demonstration of interest and intention.

In the following year, Falck-Ytter and his colleagues [43], used eye-tracking to record children's eye movements, while they were watching videos with JA actions, which included looking and pointing to a toy. With this study, they found that the ASD group, looked less to the toy than the control group.

These studies with eye-tracking and the study of attention, prove the existence of impairments in JA and in the understanding of intention.

2.4 Data Science Approaches

Data science, as its name indicates, is the science that studies data. It combines several techniques and skills, like feature selection, feature extraction, machine learning, statistics, among others. All these techniques have the purpose to retrieve, and to make us understand the value of and from data, by means of classification.

When approaching a problem using data science there are several general steps we need to consider:

1. Categorize the problem, define its input and its output.
2. Understand the nature of data, and how it can be assessed using visualization techniques and statistical tests, for example.
3. Process and transform the data.
4. Choosing algorithms for the different processing steps and classification.
5. Implement the chosen algorithm or algorithms.
6. Optimize model hyperparameters.

All of the previously mentioned steps, will lead us to the choice of, at least, one algorithm, that might fit better the classification problem. After choosing, the next step is to implement the algorithm or algorithms. Finally, in order to obtain the best possible results we can optimize the hyperparameters of the algorithm.

2.4.1 Feature Engineering

Sometimes, the raw data, are not well suited to the classification problem. When this happens, which is most often the case, it is necessary to transform it into features that could give meaning to the studied subject[44].

In this particular subject, an eye tracking acquisition retrieves the position of the eye, which by itself does not give meaningful information about the problem we want to study. We must then contextualize the position of the eye with the scenario, or the problem the person's eyes were facing. According to information from *iMOTIONS*, a platform that studies human behavior, with eye tracking methodology, there are several metrics that can be calculated from the eye positions [45]:

1. Fixations and gaze points
2. Heatmaps
3. Areas of Interest (AOI)
4. Time to First Fixation
5. Time Spent (looking at a particular AOI)
6. Ratio of times a AOI is looked at
7. Fixation Sequences
8. Revisits
9. First Fixation Duration
10. Average Fixation Duration

These, along with the information about saccades, help to contextualize the data from the eyes and its study.

2.4.2 Visualization

Visualization techniques exist to ease the interpretation of the data. They also allow the comparison between features and the analyses of their distribution.

These techniques have to convert the data values into logical and systematic elements, which will be plotted into the final graphic. Those graphics should contain all the information, without misleading the viewer and his interpretation.

Some common visualization techniques are Boxplots, Swarm plots, Violin plots and Heatmaps.

The former is a well known technique, based on simplicity and without loss of information. Boxplots work well when plotted next to each other, easing the comparison between distributions, in particular when they are non parametric [46].

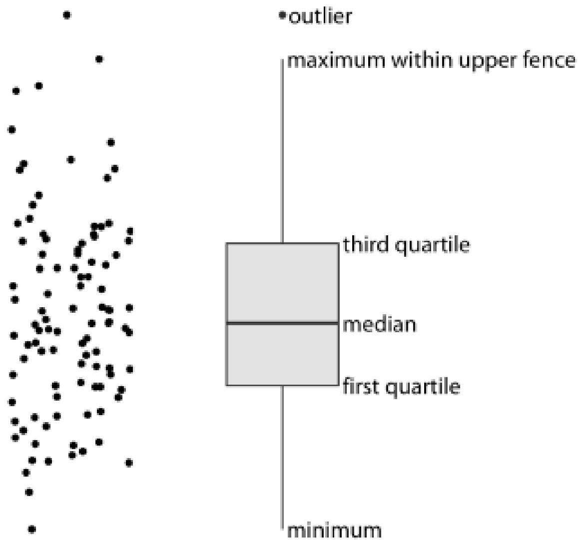

Figure 2.4: Anatomy of a boxplot. Shown are a cloud of points (left) and the corresponding boxplot (right). Retrieved from *Fundamentals of Data Visualization*[46]

The second mentioned plots, violin plots (figure 2.5), are similar to boxplots, because they can also represent data density. These plots can accurately display bimodal data, unlike boxplots.

Other way of looking at data, without the appearance that the data is very dense, when it can be very sparse, or the other way around, is swarm plots. Similar to violin plots but with the representation of single and non-overlapping points.

2.4.3 Data Preprocessing

When building a ML model, one has to consider that, some feature might not give valuable information to the output, and might even influence negatively the model. To solve this problem, one can use feature extraction and feature selection techniques, which will ”clean”, filter and rebuild the dataset, making the overall method less computationally expensive and more accurate.

The key difference between these methods is that feature selection maintains a subset

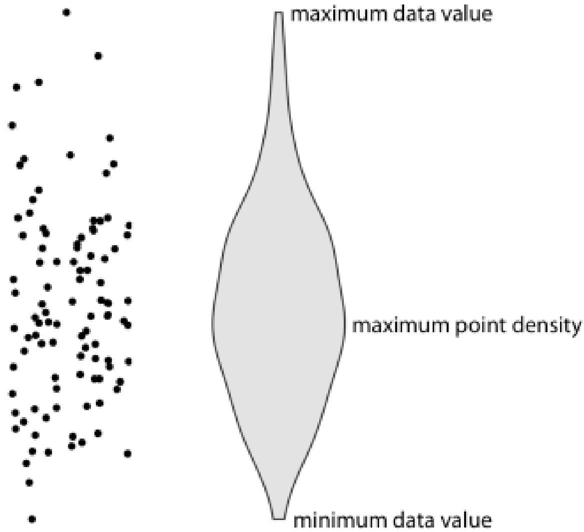

Figure 2.5: Anatomy of a violin plot. Shown are a cloud of points (left) and the corresponding violin plot (right). Retrieved from *Fundamentals of Data Visualization*[46]

of the original features, while feature extraction creates new ones from the original features [47].

2.4.3.1 Feature Selection

Feature selection aims to select the set of information from the data, considered to be more relevant and informative, by removing the redundant and irrelevant information.

Normally, there are two reasons for using feature selection. Or it is because we want to reduce the number of features, to reduce overfitting and obtain a better generalized model. Or to have more insights about the features and their relation with the output [48].

There are several feature selection methods that can be grouped into three categories: filter methods, wrapper methods and embedded methods [49].

- **Filter methods:** Features are ranked according to their relevance. This works by giving them a score and removing the ones below a certain chosen threshold [49].
- **Wrapped methods:** Usually, these methods need a higher computational power, because they also use ML algorithms to estimate the accuracy of

adding, or removing, certain features. Their aim is to find the best optimization for the model by doing it so [50].

- **Embedded methods:** They are similar to wrapped methods. Nevertheless, there is a major difference, because these methods do not separate feature selection from the learning phase. Feature selection is performed during the training of the data, without the splitting into training and testing datasets [51].

The most commonly used methods are the first two, filter methods and wrapped methods. Normally, filters do not resort to ML algorithms, e.g. statistical tests, whereas the latter have on their basis those same algorithms.

- **Correlation Coefficients**

  Correlation coefficients, by definition, measure the relationship between two variables, to see how similar, or not, they are. The highest the positive similarity between features, the more information they have in common. This means that, choosing one of these features does not only affect the information given to the model, but also reduces the dimensionality of the data.

  - **Pearson's Correlation Coefficient or Linear Correlation**

    Pearson correlation coefficient measures the linear correlation between two variables.

    Assuming *var* as the feature variance, *cov* as the covariance and Y as the output (class labels), then the equation is as follows:

    $$R(i) = \frac{cov(X_i,Y)}{\sqrt{var(X_i) \cdot var(Y)}}$$

    The results from this equation lie in the interval [-1;1]. If the result is -1 than the correlation is perfectly negative, if the result is 0, it means that there is no linear correlation and, finally, if the correlation is +1 than the correlation is perfectly positive [48]. Any other value is approximately one of the three above and classifies the relation accordingly.

  - **Spearman's Correlation Coefficient**

    On the other hand, Spearman's rank correlation coefficient is a nonparametric rank statistic. It is normally used when the distribution or the relation between the variables is unknown or likely non-parametric. Spearman's correlation coefficient assesses how well an arbitrary monotonic function describes the relation between two variables, normally a feature and the output, without making assumptions about the real distribution of each variable [52].

The coefficient respects the following equation:

$$\rho = 1 - \frac{6 \sum d_i^2}{n\left(n^2 - 1\right)}$$

Where:

ρ = Spearman rank correlation

d_i = the difference between the ranks of corresponding variables

n = number of observations

- **Linear Regression**

  Linear regression can work as a feature selection method, by finding the relationship between two continuous variables, like one feature and the problem output. It looks for a statistical relationship, obtaining a line that best fits the data and their relation, minimizing the error and the distance between the point and the regression line [53].

- **Regularized Models**

  Regularization works by adding an additional penalty to the model. Its goal is to both prevent overfitting and improve the generalization capacity of the model. So, instead of minimizing a loss function E(X,Y), these methods add a parcel to the equation, becoming: E(X,Y)+α $\|$w$\|$, where w is the vector of model coefficients, either L1 or L2, and α is a tunable free parameter, which specifies the amount of regularization [54].

  The most used regularization methods are L1 and L2 regularization.

  - **L1 regularization or Lasso**

    For this method, a penalty of $\alpha \sum_i^n \mid \mathrm{w}_i \mid$, with i starting at 1, is added to the equation.

    Since each non-zero coefficient adds value to the penalty, weaker features have zero as coefficients. Thus, this method inherently performs feature selection [54]. The final equation is as follows:

$$E(X,Y) + \alpha \sum_i^n |w_i|$$

However, L1 regularization can be unstable, since coefficients can vary significantly when there are correlated features in the data [54].

- **L2 regularization or Ridge Regression**

  On the other hand, L2 regularization or Ridge Regression adds a different penalty to the loss function: $\alpha \sum_i^n \mathrm{w}_i^2$. In this method, the coefficients are now squared, causing a different effect. Contrarily to what happens in Lasso, coefficient values spread more equally [54], meaning that correlated features get similar coefficients. For this method, the final equation is:

$$E(X,Y) + \alpha \sum_i^n w_i^2$$

  All of the above information, makes the model more stable than L1, allowing a better understanding of the features and a not so good feature selection, when compared to L1 regularization.

- **Random Forest**

  The random forest method is considered an embedded method and it consists of several decision trees, each one built randomly. Besides this, random forests also guarantee that not every tree sees the same set of features, reducing the probability of overfitting [55].

  Since it is built by decision trees, each tree works as a sequence of questions, and each node of the tree represents one question. At this node, the dataset is splitted in two according to their resemblances and differences [55]. The importance of each feature is then defined by the purity of each part of the dataset. We can see an example of a decision tree in figure 2.6.

  Normally, since the top of the tree leads with higher amounts of information, features selected in this part of the tree, generally have more importance than the ones selected in the nodes. Besides this, correlated features will have similar importance, but less importance than the ones retrieved from a tree built without correlated features [55].

  Random forests are not very useful at interpretability of the features, but work

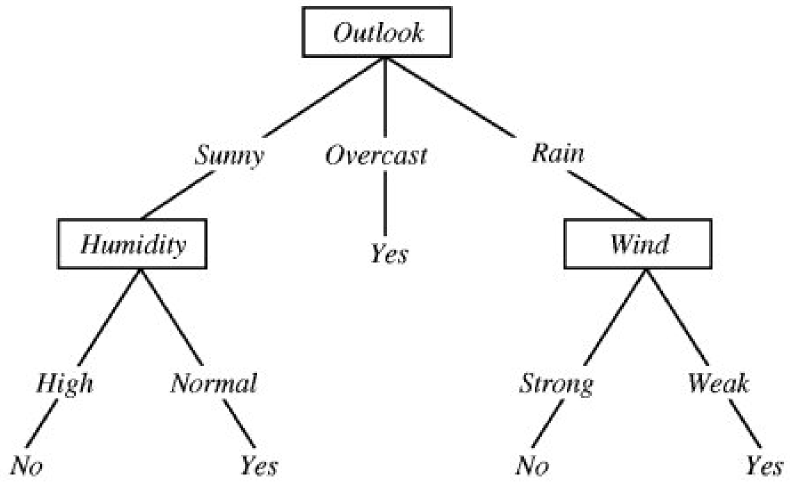

Figure 2.6: Example of a decision tree. Retrieved from [56]

is robust to outliers and to non-linear data [57]

- **Recursive Feature Elimination (RFE)**

  RFE is considered a wrapper method, since it performs a reduction of the dimensions, given an algorithm. This method assigns weights and rankings to each feature, while the algorithm is trained. After each step, the feature with the lowest weight is removed [50]. RFE is applied until the features are exhausted and later features are ranked according to when they were eliminated [58].

Each one of these methods work differently, and can be used depending on the problem's purpose. Either we want to have a better understanding of the data or we want to reduce the number of features to obtain a better classification model with less overfitting. Either way, all of them can be used, one just has to be cautious about the conclusions that can be retrieved.

2.4.3.2 Feature Extraction

- **Principal Component Analysis (PCA)**

  PCA is a feature extraction and reduction technique, which can be used in both visualization and data processing. During this approach, a new set of features is made, based on the old ones, with orthogonalization. This linear transformation, happens without the loss of the majority of the information [55].

Since PCA does not use the classification value, it can be considered an unsupervised method, able to "shrink" data's dimension, focusing on the differences between the data points for each feature.

By definition, PCA is based on the existence of a principal component for each data dimension. The first component is the one with the highest variance, the second component contains the second maximum variance, and so on. PCA intent is to find the lowest number of dimensions (meaning features), that could hold the most important information [55].

2.4.4 Machine Learning Algorithms

Nowadays, with the dissemination of data science approaches, it is easier to deal with a large amount of data. The analysis and treatment of data by humans has been largely replaced by machines and algorithms, responsible to ease this challenges by doing new and complicated tasks.

Other huge advantage of technology advances is Machine Learning (ML). Machine Learning is based on algorithms, an instruction or sequence of instructions to solve a classification problem. But the difference is that these algorithms are learning algorithms, and, by themselves, they apply or sometimes create the rules, based on their specificities. These instructions allow the machine to learn from the given data.

The data given are called training data and the rules are generated based on inferences from the dataset itself. This generates a new algorithm, which we can call machine learning model. The model depends not only on the set of instructions but also on the data. The same learning algorithm can be used in different sets of data, making different ML models.

A high amount of data is also important if it represents different possibilities and situations. This will allow the model to generalize and to be able to solve the problem when new information is given.

There are three general types of ML methods [59]:

- **Supervised learning:** the given data are labeled with the desired output;
- **Unsupervised learning:** the given data are unlabeled and the algorithm needs to search for patterns that could give answer to the problem and define categorization boundaries;

- **Reinforcement learning:** works in a dynamic environment that gives rewards and punishments according to the direction of the output.

Nevertheless, one of the biggest challenges in ML is its "interpretability" or the ability to interpret the output and the logic of the model. These may be the cause of some skepticism regarding this subject and its advantages, specially in the health and medicine fields. To better understand, it would be relevant to know what is happening and why is happening, in every step of the way. ML can be seen as "new eyes" looking at the same data, seeing what hasn't been seen yet, and we might not "grasp" what these "eyes" are really "seeing".

Then again, when training a model to answer a problem, one of the most important aspects is the given data. If a set of data is biased then the output will also be biased, and the confidence in the result will automatically decrease. It is important to filter and understand the data before it can be trained, which justifies the analysis explained in the previous section (section 2.4).

On the other hand, the variability of the data is as double-sword issue, because it will help generalization but also provide challenges.

In biomedicine, the literature mentions different approaches. Both supervised and unsupervised techniques are implemented. For example, the automated interpretation of the ECG signal, or the automated identification of a lung nodule from a x-ray exam, are considered supervised learning, since we know the output boundaries. On the other hand, unsupervised learning, is used, for example, to study the patterns found between data, and not to confirm results. This approach is also useful, since sometimes we can not see the possible patterns in heterogeneous data.

Since we know the type of output of our data, we will only focus on supervised learning. According to a *Towards Data Science* article [44], when choosing a model we should consider:

- Its accuracy;
- Its interpretability;
- Its complexity;
- Its scalability;
- How long does it take to build, test and train it;
- How long does it take to make predictions using it;

- If it meets the goal.

In supervised learning we can divide the methods in regression methods and classification methods.

The former, models and predicts continuous and numeric variables, like student test scores and stock price movements. With these methods the output has to be numeric. On the other hand, the latter represents a set of methods that models and predicts categorical variables, like financial fraud or student letter grades. These algorithms are used to predict a class, and not a real number like regression methods [60].

Since our problem has two categorical classes, we will present some of the most common classification methods, which will later on justify our choice.

- **Logistic Regression**

  Logistic regression is for classification methods, what linear regression is for regression methods.

  This method predicts between 0 and 1, by using the logistic function. Since the model is linear, it works better with linearly separable classes. Nevertheless, logistic regression can use penalizing coefficients to obtain the final classification [60]. Logistic regression then follows the next equation:

$$h(x) = \frac{1}{1 + e^{-b_0 + b_1 x}}$$

  Where $b_0 + b_1 x$ is the linear equation.

  Some of its strengths are its probabilistic interpretations, and the ability to avoid overfitting by regularizing the method. Despite having these advantages, logistic regression tends to underperform when the decision boundaries are non-linear, because its lack of flexibility does not capture more complex relationships [60].

- **Classification Trees**

  Classification trees are the equivalent to regression trees, and they learn in a hierarchical way. The dataset is progressively split into branches that maximize the information gain of each split [60]. An example was already presented in the previous subsection *Data Preprocessing* (subsection 2.4.3).

  These classification models perform very well in practice. They are scalable, robust to outliers and perform well with non-linear decision boundaries. But like everything else, they also have some weaknesses. They are unconstrained and, sometimes, individual trees tend to overfit [60].

- **Support Vector Machine (SVM)**

  A Support Vector Machine is a mechanism to calculate distance between two observations. The main goal of the algorithm is to find a decision boundary that maximizes the distance between the closest members of the two classes [60], called support vectors.

  SVM's are quite robust against overfitting, and are able to model non-linear data. Nevertheless, they require more memory and are tricky to tune [60].

ASD research provides datasets prone to be studied with SVMs.

In clinical medical research SVMs have been used to develop prediction models for both disease diagnosis, and prognosis after given a specific diagnosis [61].

Continuing the above information for Support Vector Machine (SVM)'s, these models are based on the construction of a hyperplane, which separates both cases and controls. In a problem with only two classes, the best fit hyperplane works as a separation line. The optimal hyperplane is built following the equation $w \cdot z - b = 0$. So, if we want to maximize the distance between those members of the two classes, we need first to minimize w, this will give us the next equation:

$$min \frac{1}{2} \|w\|^2 + C \sum_{i=1} \xi_i$$

$$s.t. \qquad y_i(w \cdot z_i - b) \geq 1 - \xi_i, \qquad \xi_i \geq \ 0 \forall i$$

Where:

$y_i -$ class labels, and $y_i \in \{+1, -1\}$

$\xi_i -$ error variable, $i = 1$, ... , nº of training examples

$C -$ margin regularization constant or the penalty parameter of the error term. It is responsible for the trade between a smooth decision boundary and the correct classification of the training points, which is related with the model's complexity.

The following formulation helps to achieve the computational solution for the previous equation:

$$
\begin{aligned}
& max \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j k(x_i x_i) \\
& s.t. \qquad 0 \leq \alpha_i \leq C \ \forall i, \qquad \sum_i y_i \alpha_i = 0
\end{aligned}
$$

Where:

$k(x_i x_j) = \varphi(x_i) \cdot \varphi(x_j) -$ kernel function

$\alpha_i -$ Lagrange multiplier or each training point

SVMs are often used with problems which are not linearly separable, and this is only possible by transforming the data, with the use of Kernel functions [61]. The kernel trick maps the non-linear data into a higher dimensional space, so we can find a hyperplane that separates both classes [62].

So, the effectiveness of this algorithm depends not only on the data, but also on kernel functions and parameters. These parameters include C and σ values, that need to be optimized, in order to obtain the best fitting model, without the danger of overfitting [61].

A lot of work as been done with different kernels [63], but the most common are the linear kernel, the Radial Basis Function (RBF), and the polynomial kernel [62].

- Polynomial kernel: $k(x_i x_j) = (1 + x_i \cdot x_j)^d$,
- Radial Basis Function (RBF) kernel: $k(x_i x_j) = exp\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$.

3

Methods

In this chapter, we will give insights about the approach we used to analyse the eye tracking data. The general scheme of the data analysis can be seen in figure 3.1, and starts with the extraction of features from the raw files, based on cognitive and biological knowledge we addressed in chapter 2. We were then able to generate the dataset where we implemented feature selection and feature extraction methods and classification algorithms, followed by the calculation of metrics for model evaluation.

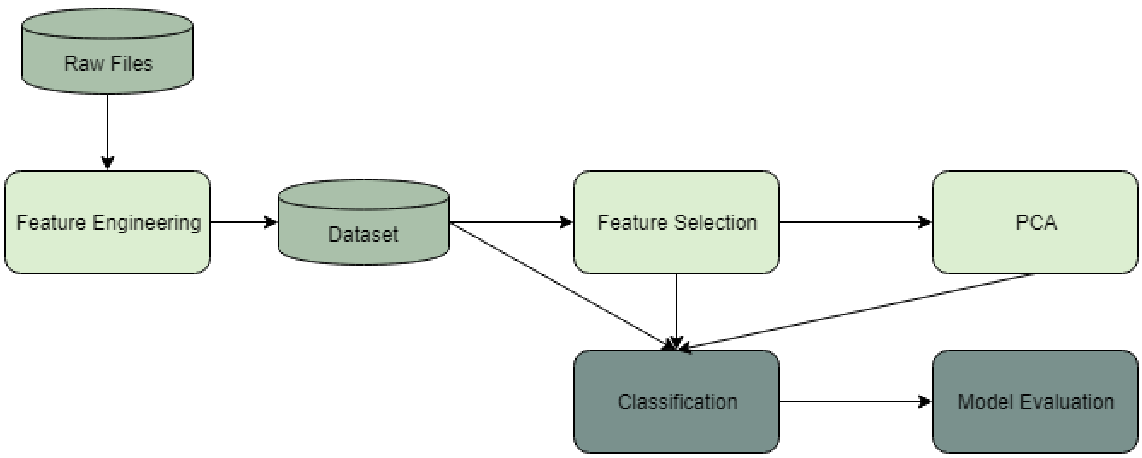

Figure 3.1: General scheme of the approach for the data analysis.

The experimental procedure was implemented using *Python 2.7* with *Spyder* IDE, due to its accessibility, and overall suitability to the different phases of this project. We will then focus on the information eventually retrieved with it, which include data from both ASD and TD classes.

3.1 Dataset

In this subsection, we will describe the different phases we went through in the analysis of the raw eye tracking data. Let us start by the acquisition procedure, previous to this project, but nevertheless important to understand the data structure and the information we could extract from it for both classes.

3.1.1 Acquisition procedure

The data acquisition was done by Carlos Amaral *et al.* in the study *A Feasibility Clinical Trial to Improve Social Attention in Autistic Spectrum Disorder (ASD) Using a Brain Computer Interface* [64].

3.1.1.1 Participants

Carlos Amaral *et al.* collected data from 30 participants, divided in two groups:

ASD group:

- 15 high-functioning ASD patients without intellectual disability;
- mean age: 23 years and 4 months;
- age range: 16 years to 38 years;
- mean full-scale IQ: 103;

TD group:

- 15 typically developing participants;
- mean age: 24 years and 10 months;
- age range: 14 years to 42 years;
- mean full-scale IQ: 117;

3.1.1.2 Apparatus

The task was performed using two setups, which differ in their degree of immersivity:

- **FLAT SCREEN**: participants sat at 70 cm from the screen in a chair in front of a 22-inch flat screen with a resolution of 1680 x 1050 pixels. Eye movements were measured using a SMI RED 500 remote eye tracker (SensoMotoric Instruments GmbH, Germany), with a sampling rate of 500 Hz, an accuracy of 0.4 and a spatial resolution of 0.03. A built in 5-point validation method was used, to ensure the precision of the data collection. In this setup, the viewing perspective is limited to what is shown in the flat screen, since no movement is allowed during the experiment with the Flat Screen.

- **OCULUS RIFT**: participants sat in the same chair and position used in the screen based setup, but with Oculus Rift DK 2 on. Eye movements were recorded with Eye Tracking HMD package from SMI embedded in the Oculus itself, with sampling rate of 60 Hz, and accuracy of 0.5-1. The same 5-point validation method was used. In this setup, the scene had a 360 perspective and a real-time fully immersive experience.

3.1.1.3 Stimuli / Task

The experiment comprised four virtual scenarios. Each scenario was modelled with the help of 3ds Max 2014 (from Autodesk Inc.). The environment texture rendering was done in 3ds Max with a scanline algorithm. The stimulation structure was written using Unity v4.

A depiction of the four scenarios can be found in figure 3.2, and they are as follows:

- **Cafe:** interior of a cafe with a maid (avatar) on the other side of the balcony. The viewer's position is in front of the balcony. Several common objects in a cafe, described in table 3.1, can be found around the avatar. The scenario also has tables and chairs, only visible in the fully immersive experience (Figure 3.2 A);

- **Classroom:** interior of a classroom. The participant finds himself standing in front of a table with a professor (avatar) behind it, with several objects displayed in the table between them. These objects are also described in table 3.1. The scenario has a set of tables and chairs, common to a classroom, visible with the fully immersive experience (Figure 3.2 B);

- **Kiosk:** the participant finds himself standing in front of a street kiosk with an employee (avatar) inside. Around the employee and dispersed on the kiosk, there are several newspapers and magazines, described considering their position, in table 3.1 (Figure 3.2 C);

- **Zebra Crossing:** the participant finds himself standing in one side of a street, waiting to cross the zebra crossing. On the other side of the street, he can find a person surrounded by the objects referred in table 3.1 (Figure 3.2 D).

Table 3.1: List of animations performed by the avatar in each scenario.

Scenario	Joint Attention Animations	Control Animations
Cafe	**Avatar turns the head to:** pack of 'Doritos' pack of 'Lays' pack of 'Lays Bake' bottles, in the back wall shelf with chewing gum left glass in the balcony middle glass in the balcony right glass in the balcony	
Classroom	**Avatar turns the head to:** book notebook pencil and eraser **Avatar points to:** ruler set-square	**Avatar:** Coughs Rolls the head Scratches the head Yawns
Kiosk	**Avatar turns the head to:** magazine above the head magazine in front magazine in left-above magazine in left-middle magazine in left-below magazine in right-above magazine in right-middle magazine in right-below	
Zebra Crossing	**Avatar turns the head to:** dustbin traffic sign traffic light traffic light button city map	

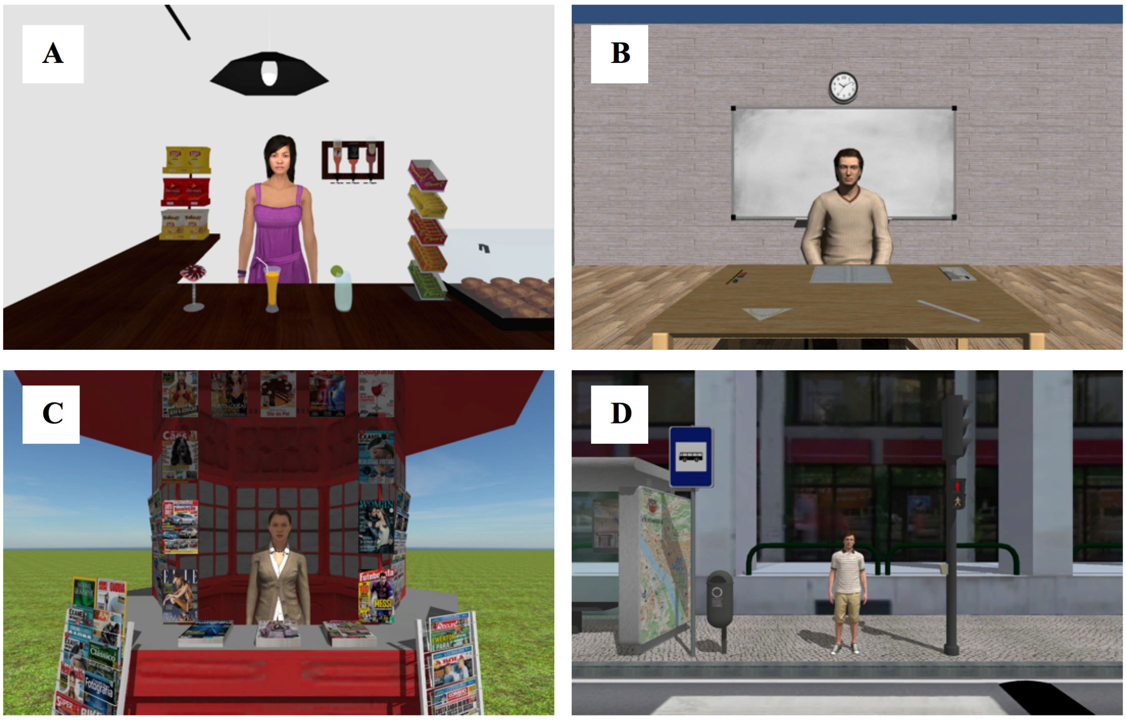

Figure 3.2: Virtual Reality Experiment Scenarios from the work of Carlos Amaral *et al.* [64].

3.1.1.4 Procedure

Each participant performed the task with both apparatuses. The order of the apparatus was alternated from subject to subject, in a balanced manner, and both procedures started with the eye- tracker calibration and validation. Next, the presentation of each scenario was done. The order by which each scenario was presented to every participant was pseudo-randomly defined *a priori* to ensure that every scenario was presented for the first time the same number of times. This order was maintained in both apparatus for each subject. In all the scenarios, the task started with a 30 seconds free-viewing period, followed by a series of avatar animations spaced by between 2 and 2.5 s. The animations are divided in joint attention animations and control animations. The joint attention animations comprise both the avatar's head turning or him pointing to one object of interest. The animations in each scene can be found in table 3.1.

For each experiment, the animations were repeated 2 times in a random order. Since the number of ROIs varies in each scenario, the number of JA animations will also vary. The number of control animations is the same for each scenario, there are 4 different actions, which makes a total of 8 control animations.

Considering each scenario particularly, we have 18 JA animations in the cafe scenario, 10 in the classroom, 16 in the kiosk and 10 in the zebra crossing scenario. The overall animations were 180 with JA and 32 without JA, per setup.

The only instruction given to the participants was to act naturally and they were not aware that their eye movements were being recorded. The goal was to keep the experiment as similar to a real situation as possible.

3.1.2 Data

The raw data was in the form of *txt* files, composed by the starting and ending date and hour (first and last rows), the acquired 3D positions, in the format x, y, z, for the eye projection in the screen, and the record of the trigger for the avatar animations. According to the Cartesian coordinate system, x is the coordinate parallel to the ground, and is seen as the length of the scenario, y is the coordinate perpendicular to the ground, along the height of the scenario, and z gives the 3rd dimension and depth of the scenario(figure 3.3). The default unit in Unity is 1 meter. In this project, a scale of 0.1 was used, but for referring to the distance between the center of the scenario (0,0,0) and a certain position we will use units, since this is only used to check to where the subject is looking at.

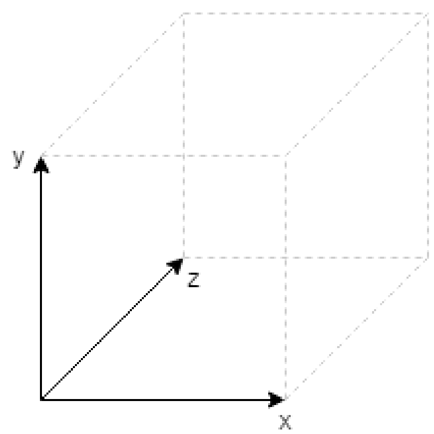

Figure 3.3: Depiction of the used Cartesian coordinate system.

Since this was a temporal acquisition, we could retrieve the procedure duration in milliseconds from the difference between the first and the last lines. An example of the files is shown in table 3.2.

As mentioned before in subsection 3.1.1, the data were acquired from 15 ASD individuals and from 15 TD individuals, for both setups and for the 4 scenarios, which

Table 3.2: Raw Data - *txt* file example from Clinical Trial data with Oculus Rift and Kiosk scenario.

Beginning:2016-2-4-11-45-41-288
(1.2, 3.8, 6.5)
(0.9, 3.9, 6.5)
(1.0, 3.9, 6.5)
(1.0, 3.9, 6.5)
(1.0, 3.9, 6.5)
(1.0, 3.9, 6.5)
(1.1, 4.0, 6.5)
(1.1, 4.0, 6.5)
(1.1, 4.0, 6.5)
...
(0.9, 1.7, 6.5)
A: female@lookleftdown i
(0.8, 1.6, 6.5)
(0.8, 1.6, 6.5)
(0.8, 1.7, 6.5)
(0.9, 1.6, 6.5)
...
(-7.7, 0.1, 38.3)
(-7.7, 0.1, 38.3)
(-6.3, 0.1, 29.5)
(-2.3, 4.2, 6.5)
(1.6, 3.9, 39.5)
(-0.5, 3.9, 6.5)
(1.7, 4.1, 39.5)
(0.3, 3.1, 6.5)
(0.4, 3.0, 6.5)
End:2016-2-4-11-48-4-768

makes a total of 240 files used in this project, 120 for each setup, divided into 60 files per class.

3.1.3 Feature engineering

We investigated features in eye tracking data that could have cognitive and biological meaning for the distinction of ASD and TD groups. We searched for information that might represent social cognitive processes, not focusing only on the ones promoted by the social interactions in the task. These features with potential cognitive and biological meaning were chosen based on the information found in the literature for eye tracking analysis with and without ASD subjects, mentioned in chapter 2.4.1. Tables 3.3 and 3.4 sum up the final set of features. These tables represent the features calculated for different parts of the acquisition, since these parts can depict different levels of interpretation of the situation.

So, table 3.3 contains the set of features obtained during the free viewing part of the acquisition. This part corresponds to the first 30 seconds of the acquisition, where there was no animation (the avatar was just 'breathing'). Here, the level of complexity of the situation can be compared to a visual search [65], where the subject inspects the scenario.

Table 3.3: Features for the acquisition during free viewing.

Code Name	Explanation
fv_gen_gen_n_fixations	Number of fixations regardless of position
fv_gen_gen_avg_time	Average fixation time regardless of position
fv_gen_n_roi	Number of fixations in the pre-established regions of interest
fv_gen_avg_time_roi	Average fixation time in the pre-established regions of interest
fv_gen_n_face	Number of fixations in the avatar's face

Table 3.3 continued from previous page

Code Name	Explanation
fv_gen_avg_time_face	Average fixation time in the avatar's face
fv_gen_max_time_face	Maximum duration of fixation in the avatar's face
fv_total_sac_time	Total time during saccades
fv_n_sac	Number of saccades
fv_avg_sac_time	Average saccade time
fv_sac_velocity	Average saccade velocity
fv_sac_density	Average number of coordinates variation during saccadic movement
fv_avg_sac_size	Average saccadic length
fv_dist_sac_max	Total length of saccadic movement
fv_time_first_fix	Duration of the first fixation

On the other hand, table 3.4 contains the set of features obtained after the occurrence of the first animation. We considered a higher level of complexity in this part, similar to scene perception in reading [65], since the subject had more information to process. The animations started without the subject knowing they were going to happen, and this, added to the expectation of what was going to happen next, increased the complexity.

Table 3.4: Features for the acquisition starting after the 1st animation.

Code Name	Explanation
gen_max_fix_time	Maximum duration of fixation
num_1500_fix	Number of fixations, regardless of position
avg_1500_fix_time	Average fixation time, regardless of position

Table 3.4 continued from previous page

Code Name	Explanation
num_1500_roi_fix	Number of fixations in the pre-established regions of interest
avg_1500_roi_fix_time	Average fixation time in the pre-established regions of interest
num_1500_face_fix	Number of fixations in the avatar's face
avg_1500_face_fix_time	Average fixation time in the avatar's face
num_1500_out_fix	Number of fixations in the positions outside of the scenario limits
avg_1500_out_fix_time	Average fixation time in the positions outside of the scenario limits
num_1500_other_fix	Number of fixations in other positions (that are not considered above)
avg_1500_other_fix_time	Average fixation time in other positions (that are not considered above)
avg_reaction_time	Average reaction time, starting after an animation and ending in the beginning of a fixation
avg_reaction_time_look	Average reaction time, starting after a social animation and ending in the beginning of a fixation
avg_reaction_time_anim	Average reaction time, starting after a nonsocial animation and ending in the beginning of a fixation
total_sac_time	Total time during saccades
n_sac	Number of saccades
avg_sac_time	Average saccade time
sac_velocity	Average saccade velocity

Table 3.4 continued from previous page

Code Name	Explanation
sac_density	Average number of coordinates variation during saccadic movement
avg_sac_size	Average saccadic length
dist_sac_max	Total length of saccadic movement
dens_look	Average number of coordinates variation during saccadic movement within the reaction time after a social animation
dens_anim	Average number of coordinates variation during saccadic movement within the reaction time after a nonsocial animation
vel_look	Average saccade velocity during the reaction time after a social animation
vel_anim	Average saccade velocity during the reaction time after a nonsocial animation
dist_look	Total length of saccadic movement during the reaction time after a social animation
dist_anim	Total length of saccadic movement during the reaction time after a nonsocial animation
match_roi_fix	Number of fixations when a JA phenomenon occurs
match_mean_time_roi_fix	Average fixation time in the region of interest when a JA phenomenon occurs
match_react_time_roi	Average reaction time, starting after a social animation and ending in the beginning of a fixation when a JA phenomenon exists

Table 3.4 continued from previous page

Code Name	Explanation
match_nonsocial_face_fix	Number of fixations in the avatar's face after a nonsocial animation
match_nonsocial_mean_time_face_fix	Average fixation time in the avatar's face after a nonsocial animation
match_react_time_face	Average reaction time, starting after a nonsocial animation and ending in the beginning of a fixation in the avatar's face
match_nonsocial_other_fix	Number of fixations in other positions after a nonsocial animation
match_nonsocial_mean_time_other_fix	Average fixation time in other positions after a nonsocial animation
match_react_time_other	Average reaction time, starting after a nonsocial animation and ending in the beginning of a fixation in other positions
per_out_1500	Percentage of fixations outside the scenario limits - Number of fixations outside the scenario limits divided by the number of fixations, regardless of position
spatial_density	Ratio between the number of regions of interest where a JA phenomenon occured and the total number of regions of interest

To differentiate both parts, we did not only use some different features, to accordingly fit to the situation, but we also varied the minimum time needed to look at a position to be able to consider it a fixation. We call this the fixation time threshold and, after trying several different values, based on the information in table 3.5 from Rayner [65], we chose 180 ms for the fixation time threshold during free viewing and 260 ms for the fixation time threshold during the part after the occurrence of the 1st animation.

Other considerations were used to calculate the features, and we will first mention the ones common to both parts of the acquisition.

Table 3.5: Intervals for fixation duration [65].

	Fixation duration (ms)
Visual Search	180 - 275
Scene Perception	260 - 330

We approached the data using only x and y positions, since we could obtain our features with only these two coordinates, z did not add any meaningful information. To calculate fixations, we established a margin of 0.2 units for each side, meaning that if the acquired coordinates were in the interval [x-0.2; x+0.2] [y-0.2; y+0.2], they would be considered as the same point, reducing some lack of precision in the eye position acquisition. The same principle was followed to calculate saccades, in which we would only consider different points if they were separated by, at least, 0.3 units in any direction.

We then inspected fixations by obtaining their total number and duration. On the other hand, to study saccades we calculated their total number, duration, velocity and density of transition. All of the features are associated with variations, according to the situation or the reaction intended to study.

To check if the subject looked to the areas where the avatar was going to point or look, we needed to delimit them. The x and y values used to delimit the Area of Interest (AI) were taken manually and were based on the pre-established limits in *Unity*, made during the work of Carlos Amaral *et al.* [64] (figure 3.4). Some of the areas were enlarged and others were shortened, to better distinguish the AI. The values used for the new limits are shown in table 3.6.

The next considerations were only used in the second part of the acquisition, where we can assume a social context.

First, we chose a limit of 1500 ms, which started after each animation. We only considered fixations to be relevant for the context when they started before the end of these 1500 ms. The goal here was to give time to the subject to process the animation, search for the object, and fixate it, if he wanted to inspect it. Since the time between animations ranged from 2000 ms and 2500 ms, this threshold allowed us to discard fixations that were not a direct result of the avatar's animation.

Besides those temporal limits, we also delimited the scenarios, like shown in table 3.7, based on the plots from all the points of the files. The examples of the outer limits are shown in figure 3.5, for each one of the scenarios and for one person only. Nevertheless, these limits were calculated for more than one individual, and the

Table 3.6: Values for ROI Limiting

Scenario	ROI	x	y
Cafe	Bottles	[-3.5 ; -1.3]	[5.1 ; 6.7]
	Doritos	[2.1 ; 3.5]	[4.5 ; 5.4]
	Face	[-0.5 ; 0.4]	[5.1 ; 6.3]
	Glass Centre	[0.5 ; 1.4]	[3.2 ; 4.5]
	Glass Left	[-1.4 ; -0.5]	[3.2 ; 4.4]
	Glass Right	[0.7 ; 1.7]	[3.2 ; 4.4]
	Gum	[-2.5 ; -2.1]	[3.4 ; 5.8]
	Lays	[2.1 ; 3.5]	[5.4 ; 6.8]
	Lays Bake	[2.1 ; 3.5]	[3.5 ; 4.4]
	Light	[-0.7 ; 0.9]	[6.3 ; 7.0]
Classroom	Book	[-9.7 ; -7.7]	[2.1 ; 3.1]
	Face	[-9.0 ; -8.4]	[4.1 ; 5.1]
	Pencil & Eraser	[-11.1 ; -9.9]	[2.1 ; 3.2]
	Set-square	[-10.4 ; -9.2]	[2.4 ; 2.7]
	Ruler	[-8.3 ; -6.7]	[2.5 ; 2.7]
	Notebook	[-7.2 ; -6.2]	[2.0 ; 3.2]
Kiosk	Front	[-0.9 ; 0.8]	[3.2 ; 4.0]
	Face	[-0.3 ; 0.3]	[4.9 ; 5.8]
	Left-middle	[-2.7 ; -1.0]	[4.6 ; 5.8]
	Left-below	[-2.7 ; -1.0]	[3.3 ; 4.5]
	Left-above	[-2.7 ; -1.0]	[5.9 ; 7.1]
	Right-middle	[1.0 ; 2.7]	[4.6 ; 5.8]
	Right-below	[1.0 ; 2.7]	[3.3 ; 4.5]
	Right-above	[1.0 ; 2.7]	[5.9 ; 7.1]
	Above	[-0.9 ; 0.8]	[7.0 ; 7.9]
Zebra Crossing	Dustbin	[3.7 ; 8.2]	[0.1 ; 5.7]
	Face	[-0.6 ; 0.7]	[5.1 ; 6.3]
	Hydrant	[-8.7 ; -5.5]	[0.0 ; 4.4]
	City Map	[9.1 ; 10.6]	[0.6 ; 8.6]
	Sign	[4.2 ; 10.4]	[6.4 ; 12.7]
	Traffic Light	[-7.8 ; -2.7]	[6.6 ; 12.8]
	Traffic Light Button	[-6.9 ; -2.7]	[3.1 ; 5.0]

Figure 3.4: Pre-defined ROI limits, based on the work of Carlos Amaral *et al.* [64].

same outer limits were obtained. The black contour, in figure 3.5, represents the limits of each scenario, according to the values in table 3.7.

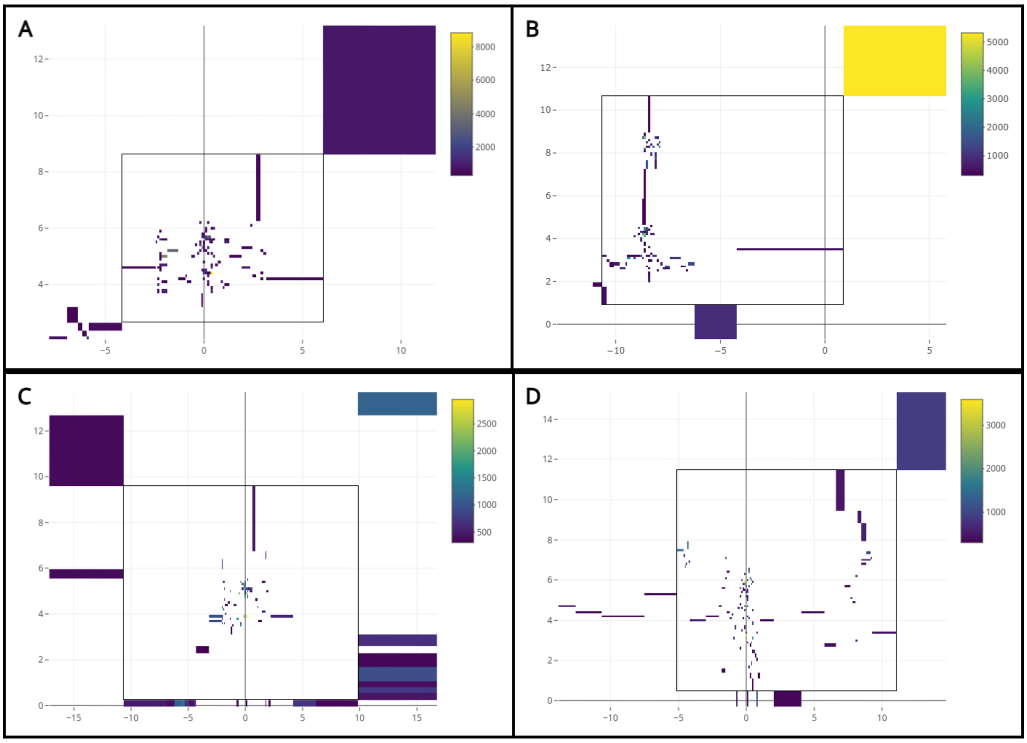

Figure 3.5: Representative heatmaps of each scenario. A - Cafe, B - Classroom, C - Kiosk, D - Zebra Crossing. The black contour limits the scenario seen in figure 3.2.

Table 3.7: Outer limits for each scenario.

Scenario	x	y
Cafe	[-4.2 ; 6.0]	[3.2 ; 8.6]
Classroom	[-10.7 ; 0.9]	[0.9 ; 10.7]
Kiosk	[-10.7 ; 9.9]	[0.2 ; 9.6]
Zebra Crossing	[-5.2 ; 11.1]	[0.5; 11.5]

Since this part considers animations, we first decided to see what kind of reaction the animation produced. So, we focused on where the participant was looking at, regardless of the animation type. We divided the total area into ROIs, avatar's face, positions outside of the scenario and other positions. Then, we divided the animations into social and nonsocial animations. For the former, we considered fixations and the related features, when a JA phenomenon occurred. For the latter, we separated the areas and considered fixations in the avatar's face, outside of the scenario or in the remaining positions, that were neither the previous nor the established ROI.

We also calculated the time it took for the subject to start fixating after each animation, and called it reaction time. Furthermore, the animations were divided again, and the reaction time was calculated for each animation type separately. Ultimately, the saccadic distance, the velocity and the density of transition during this interval were obtained.

3.1.4 Dataset

By the end of the feature generation, we were able to create our final datasets, the matrices containing all the feature information for each participant. Two different datasets were obtained, one for each setup, in order to compare them.

We calculated all 53 features shown in tables 3.3 and 3.4 for each one of the 4 scenarios, making a total of 212 features. The final datasets had 30 rows, one for each individual, and 212 columns, with the information for each one of the 212 features and a column with the classification: 0 for TD and 1 for ASD.

Next, we divided both datasets into a new matrix, containing only the features, and a vector, containing the classification. We will call the matrix and the vector X and y, respectively, for simplification purposes, and this nomenclature will be used in the next sections.

3.2 An approach to feature extraction & selection

To improve our results, we used feature extraction and feature selection algorithms, considering that feature selection selects features without changing them, and feature extraction builds new ones from the ones already existing. By doing this, we could remove unnecessary and repeated information that could be negatively influence the classification algorithms, and create new features able to better distinguish both classes. We divided our general approach in four steps, each one including the implementation of different algorithms.

The first step, represented the first experimental approach, without any feature processing (figure 3.6), used as a control of the experience.

In the second step, to evaluate the correlation between each feature and the target vector, we performed the *Spearman* correlation test and the features with a p-value below 0.05 were chosen, since they presented higher correlation with the target vector. (figure 3.7).

The third step included the implementation of a set of feature selection methods, not only to remove features highly correlated between each other, which would only add unnecessary information, but also to understand which of those features played a larger role in the classification (figure 3.8). So we used the following methods, whose advantages and disadvantages were previously mentioned in subsection 2.4.3:

1. Linear Regression
2. Ridge Regression
3. Recursive Feature Elimination (RFE)
4. Lasso
5. Random Forest
6. Linear Correlation

We used *RandomizedSearchCV*, a function from the *scikit-learn* Machine Learning *Python* library that estimates the best hyperparameters of a model by making random hyperparameters combinations, to find the best solution for the model [66] [67]. With this approach, we could find the best value of alpha for methods 2 and 4, according to the given features.

For each feature selection method, we selected the 3 most relevant features, and crossed the results in order to obtain the final set of features which were thought to perform better.

Finally, in the 4th step, we took the data obtained in the third step, after feature selection, for feature extraction and reduction, we used PCA for both visualization with 2 components, and for the generation of new features to be introduced to our classification model (figure 3.9).

3.3 Machine Learning algorithms

To classify our classes, we used Machine Learning algorithms in all the steps of our approach. This allowed us to verify if our model was being improved according to the feature extraction and selection done. In all of these steps, we used a Support Vector Machine (SVM) with different kernels, addressed in subsection 2.4.4.

The first thing to do, common to all the steps, was to split our dataset into the training data, the information used to train our model, and the testing data, used to evaluate how well it behaved. With this, we were able to test our model with "unseen" data and increase the reliability of our results.

We used the *train_test_split* function from *Python* library *scikit-learn* to split our data randomly. We divided the data in 70% for our training data, and the remaining 30% for our testing data, which means that our model trained with 21 subjects and classified the remaining 9.

3.4 Model evaluation

After testing our model with new data, we obtained a 2x2 matrix with the classification results, the confusion matrix (table 3.8), where we can assess how well our model performed, or not, for both classes.

Table 3.8: General confusion matrix.

		Predicted	
		ASD	**TD**
Real	**ASD**	TP	FP
	TD	FN	TN

In table 3.8, TP (True Positives) gives the number of well classified ASD individuals, and TN (True Negatives) gives the number of well classified TD individuals. On the other hand, FP (False Positives) represents the number of ASD individuals classified as TD individuals and FN (False Negatives) the opposite, meaning the value represents the number of TD individuals classified as having ASD.

We used these values to calculate the metrics mentioned below:

- True Positive Rate (TPR), or Hit Rate or Recall or Sensitivity

$$TPR = \frac{TP}{TP + FN}$$

- False Positive Rate (FPR), or False Alarm Rate

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

- Accuracy

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

- Error Rate

$$ErrorRate = 1 - accuracy = \frac{FP + FN}{TP + TN + FP + FN}$$

- Precision

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

These metrics allowed us to evaluate the performance of our model differentiating the classes TD and ASD

We did it using cross-validation, with a fifty-folds cross validation. The outcomes shown are the mean of each fold obtained for each metric.

Finally, and to summarize our work, figure 3.10 sums the general approach, specifying, some of the methods and algorithms for the different phases, first shown in figure 3.1 and described in this chapter (chapter 3).

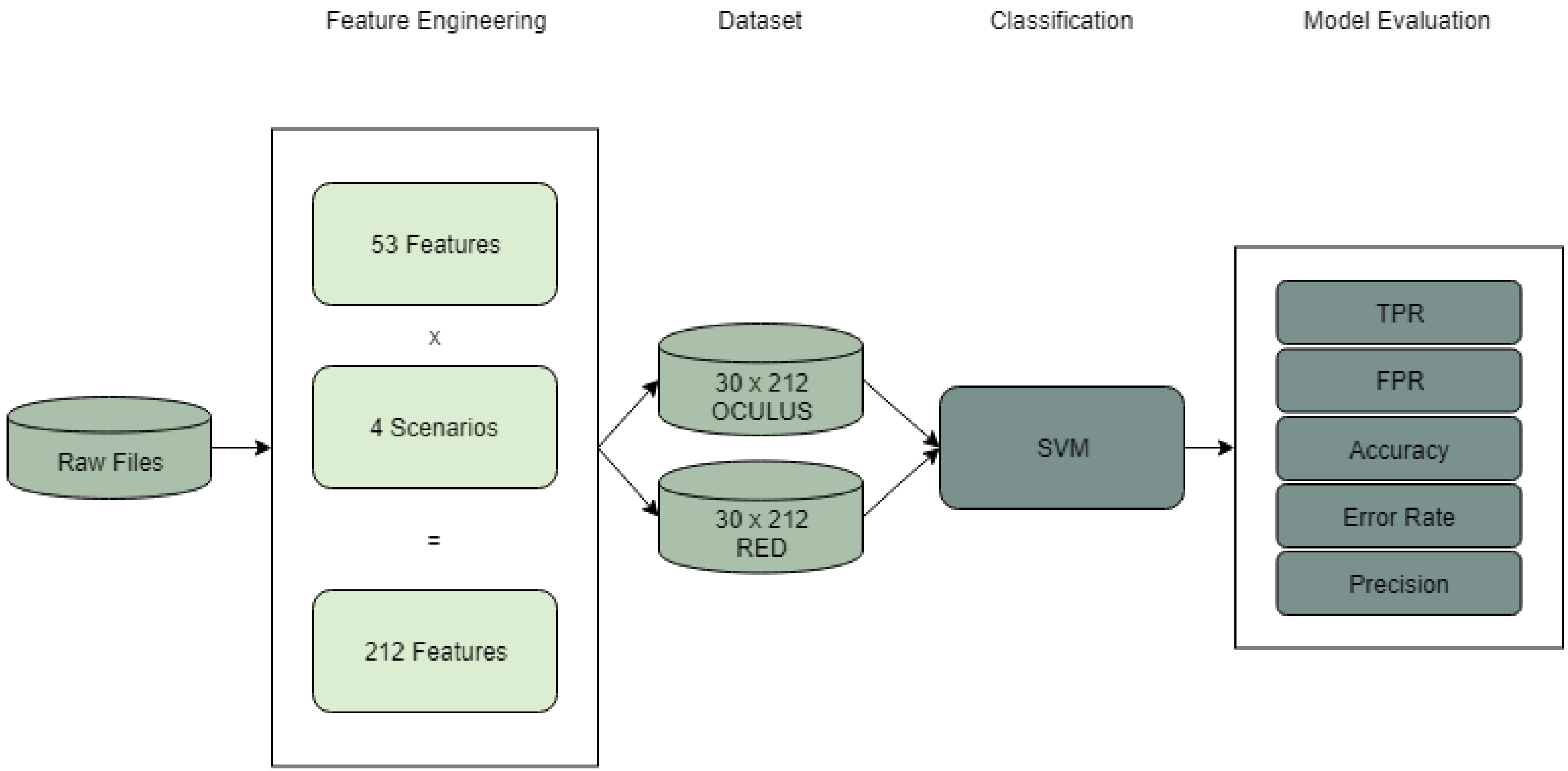

Figure 3.6: Scheme of the first step of the approach for the data analysis.

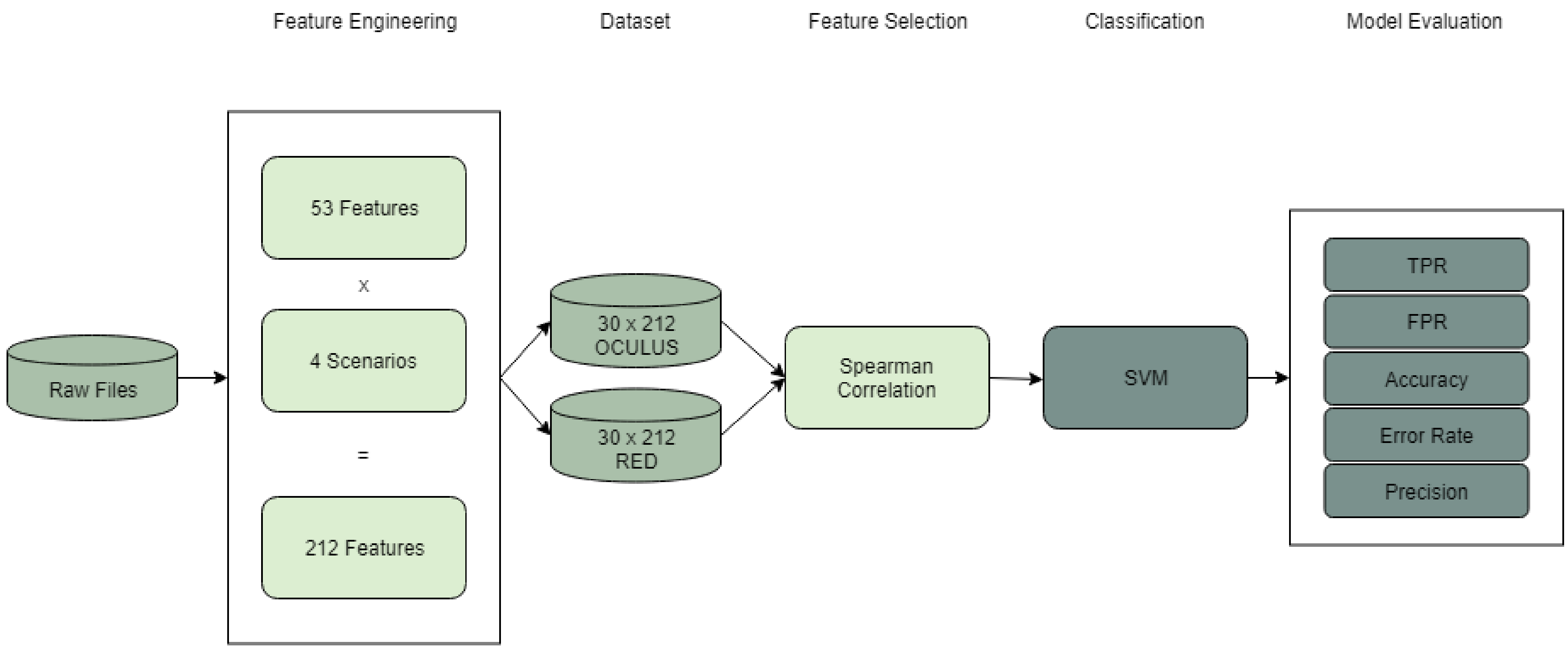

Figure 3.7: Scheme of the second step of the approach for the data analysis.

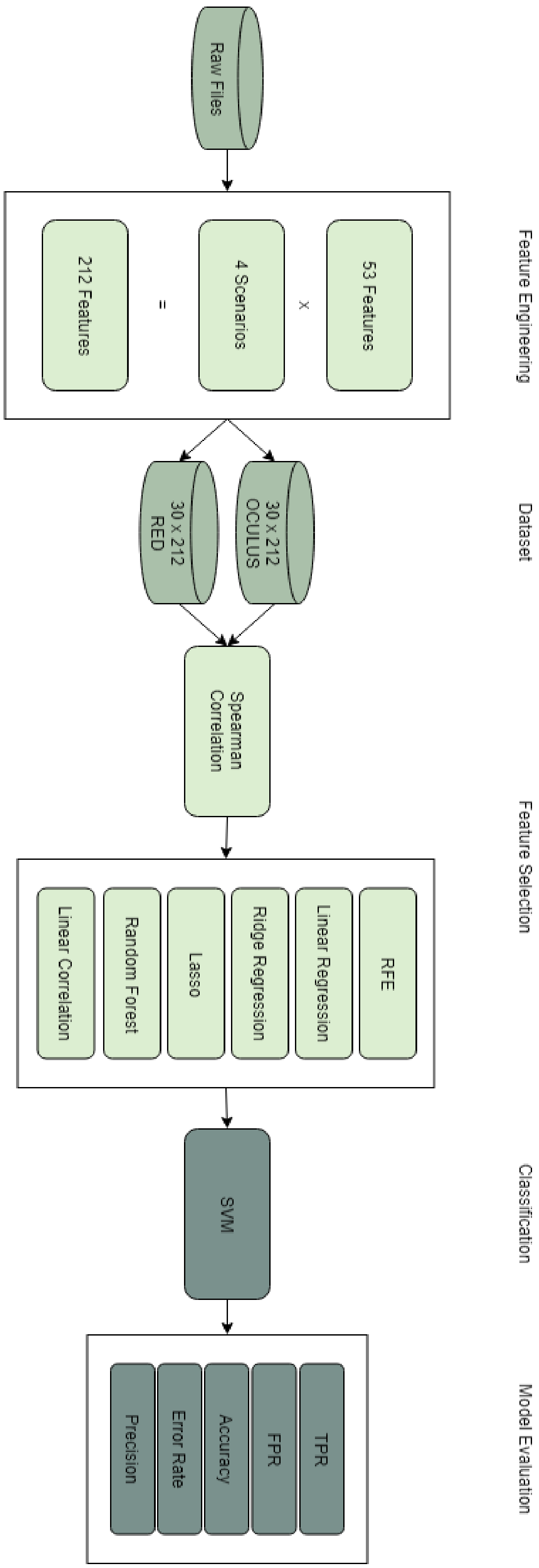

Figure 3.8: Scheme of the third step of the approach for the data analysis.

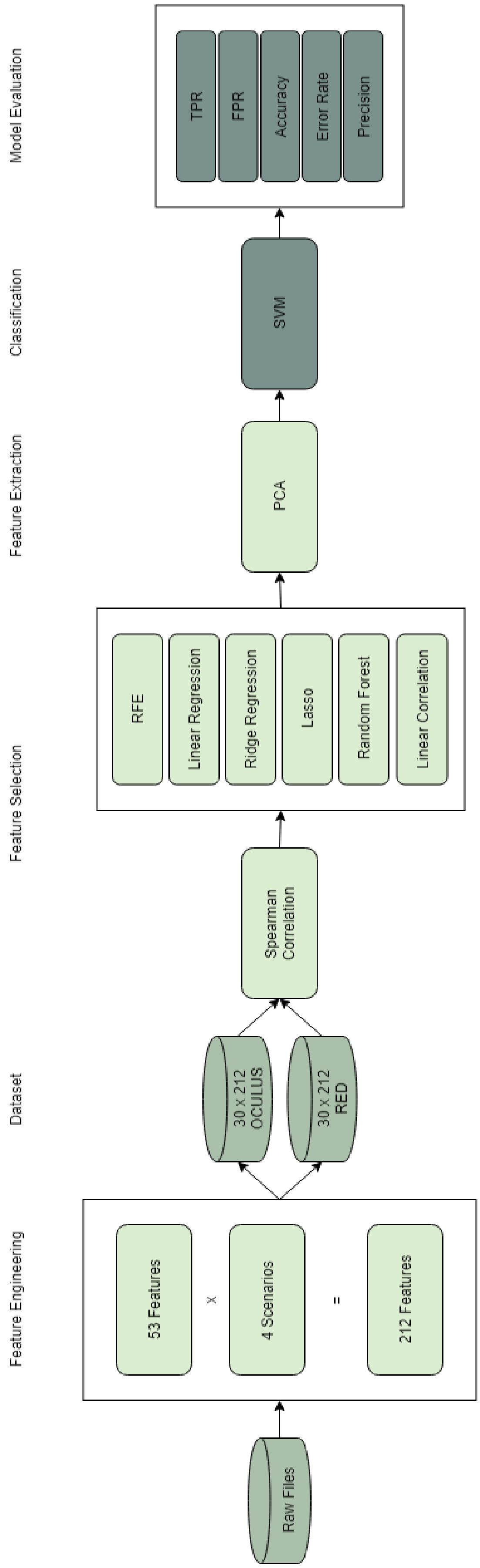

Figure 3.9: Scheme of the fourth step of the approach for the data analysis.

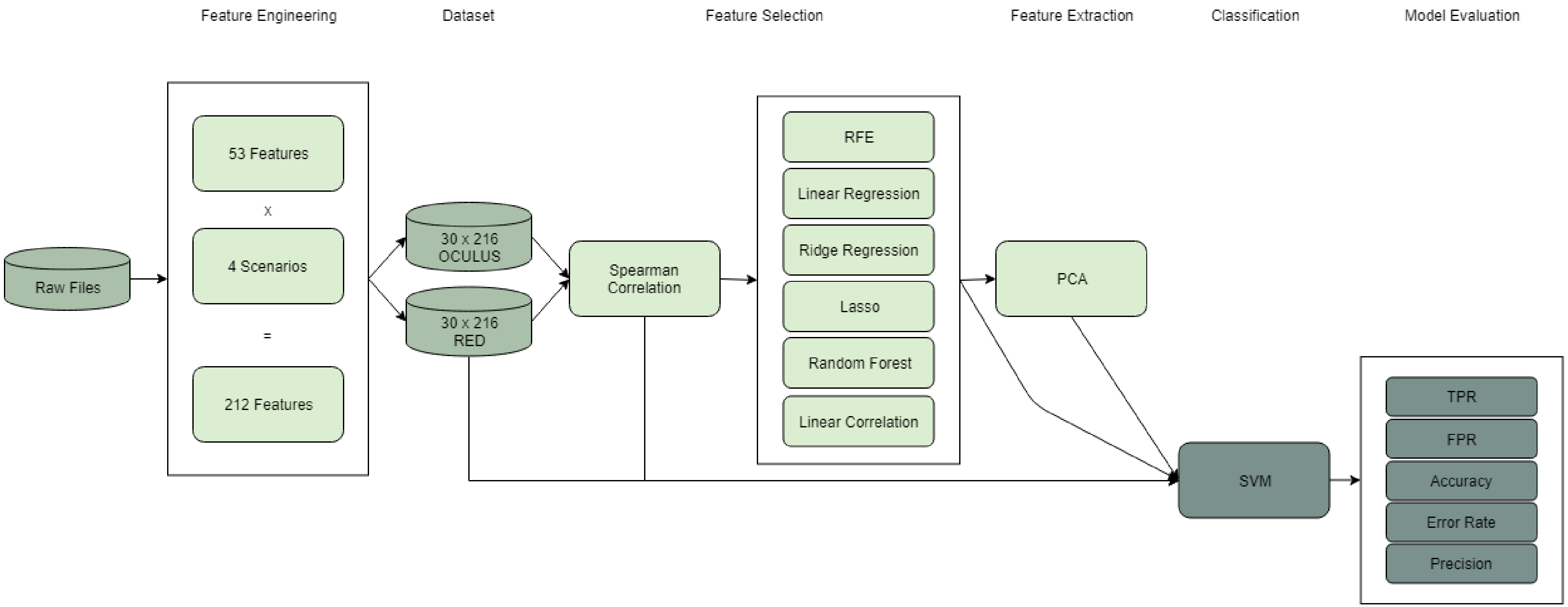

Figure 3.10: Scheme of the approach with specific information.

4

Results

In this chapter, we will report the results obtained with the approach proposed in Chapter 3.

Considering the proposed goals for this project, in Chapter 1, we divided the work in three main questions:

1. Which features can we obtain from data containing eye positions and their respective time stamps?
2. Is any of those features "good" enough to define a potential biomarker for ASD?
3. Can we predict or confirm the disorder, by analysing eyetracking data with a machine learning approach?

The first question was already answered in Chapter 3, where the list of proposed features can be found in table 3.3, for the free viewing part of the acquisition, and in table 3.4 for the part of the acquisition with social interaction. The values obtained for each individual, when those features were calculated, and depending on each scenario, for both the RED setup and the Oculus Rift setup, compose the final datasets for both feature analysis and classification.

4.1 Feature analysis and classification

To answer the third question *Can we predict or confirm the disorder, by analysing eyetracking data with a machine learning approach?* we tested the dataset in four different steps. Each step was incremented to the previous one, according to the obtained results. In each step we tried to overcome the limitations of the data analysis done during it. To evaluate the "quality" of the resulting dataset in each step, we used a Support Vector Machine with two different kernels, Linear kernel and

Radial Basis Function kernel. In this section we will also try to answer the second question *Is any of those features "good" enough to define a potential biomarker for ASD?*, during the second step.

With *GridSearchCV* we obtained the best values for the hyperparameters of both kernels of the SVMs.

In the first experimental step we did not perform any form of feature selection, it worked mainly as control. In all the steps, the whole datasets were split into training and testing data, according to what is mentioned in Chapter 3, 70% for training and 30% for testing, and given to the algorithm.

In the second step, we performed Spearman's correlation test, obtaining datasets with less features. Those datasets were then trained and tested according to the same 1st step principles.

Not satisfied with the obtained results, we tried different feature selection methods, which are described in the 3rd step. With smaller datasets we trained once more a SVM with both kernels.

In the last step we did a PCA to the dataset obtained from the the previous step, followed by training and testing with SVM.

4.1.1 1st step

The results for the first step are shown in table 4.1. Comparing both kernels we see better results with the linear kernel, despite the setup. Where the accuracy is around 50%, 15% higher than with RBF kernel. The percentage of well classified autistic individuals is higher for RED setup, whereas for TD individuals we obtain better results with Oculus rift.

Table 4.1: Results before feature selection, for both linear and RBF kernels, where ASD is the ratio of well classified autistic individuals, among the total number of people within the correspondent class, and TD is the equivalent for the Typically Developing individuals class.

Setup	SVM Kernel	Evaluation Metrics						
		ASD	TD	TPR	FPR	Accuracy	Error Rate	Precision
Oculus	**Linear**	0.487	0.583	0.583	0.513	0.515	0.484	0.500
	RBF	0.460	0.540	0.540	0.540	0.373	0.627	0.193
RED	**Linear**	0.556	0.473	0.473	0.444	0.511	0.489	0.517
	RBF	0.440	0.560	0.560	0.560	0.369	0.631	0.209

4.1.2 2nd step

After this first step, we wanted to understand which were the features with more relevant information for the decision making of the classification algorithm. Our first thought was to train the SVM with the most correlated features with the classes in the data. For this, we used Spearman's correlation test. With it, we analyse the relationship between each feature and the classes of the data, choosing the ones that had a p-value below 0.05, meaning they were more correlated with those classes.

Separating both apparatus, in RED setup, from the 212 features, only 14 features were selected. The chosen features can be seen in table 4.2.

Table 4.2: Features that passed Spearman's correlation test, for RED setup.

	RED
Features	n_sac_kiosk avg_sac_time_kiosk avg_reaction_time_anim_kiosk dens_look_kiosk n_sac_classroom num_1500_fix_classroom num_1500_roi_fix_classroom dist_anim_classroom fv_gen_n_roi_cafe fv_gen_avg_time_roi_cafe fv_gen_gen_n_fixations_zebra fv_gen_n_roi_zebra fv_n_sac_zebra fv_avg_sac_size_zebra

On the other hand, with the values for the Oculus Rift features, we obtained different results. From the same 212 features, 16 were selected as the most correlated with the output. Table 4.3 shows this set of features.

Table 4.3: Features with significant correlation with the classes' values, for the Oculus Rift setup.

	Oculus Rift
Features	total_sac_time_kiosk sac_velocity_kiosk sac_density_kiosk dist_sac_max_kiosk avg_1500_roi_fix_time_kiosk num_1500_face_fix_kiosk dens_look_kiosk avg_sac_size_classroom match_roi_fix_classroom match_mean_time_roi_fix_classroom match_react_time_roi_classroom spatial_density_classroom sac_density_zebra avg_sac_size_zebra num_1500_roi_fix_zebra num_1500_out_fix_zebra

After obtaining this set of features for both apparatus, we trained our SVM models, following the same principles as the 1st step, for data division. We obtained higher values of classification for both classes, than in the 1st step. These results can be found in table 4.4.

Table 4.4: Results after Spearman's correlation test, for both linear and RBF kernels, where ASD is the ratio of well classified autistic individuals, among the total number of people within the correspondent class, and TD is the equivalent for the Typically Developing individuals class.

Setup	**SVM Kernel**	**Evaluation Metrics**						
		ASD	**TD**	**TPR**	**FPR**	**Accuracy**	**Error Rate**	**Precision**
Oculus	**Linear**	0.709	0.743	0.743	0.291	0.702	0.298	0.736
	RBF	0.520	0.480	0.480	0.480	0.389	0.611	0.175
RED	**Linear**	0.716	0.707	0.707	0.284	0.698	0.302	0.727
	RBF	0.568	0.850	0.850	0.432	0.673	0.327	0.631

With a careful analysis, we notice that the results were significantly better with the linear kernel, despite the apparatus. Also, the results for classifying correctly individuals with ASD, were very similar among both apparatus (0.709% with Oculus and 71.6% with RED), whereas for classifying TD individuals, the percentage was slightly higher with the Oculus setup (74.3%) than with RED setup (70.7%). Nevertheless, both accuracy and precision of the model were higher with Oculus.

In figures 4.1 and 4.2 we can have some visual information about the distribution of the features in table 4.2, and give some meaning to the results obtained with the SVM.

The former shows the boxplots of those features. Analysing this figure, we notice slight differences in the distribution of the features and highlight *num_1500_fix_classroom* and *fv_gen_n_roi_zebra*, which are the number of fixations, regardless of position in the classroom scenario, and the number of fixations in ROIs during the free viewing part, in the zebra crossing scenario, respectively.

These differences between distributions in both classes might have contributed to the improvement of the results in this step.

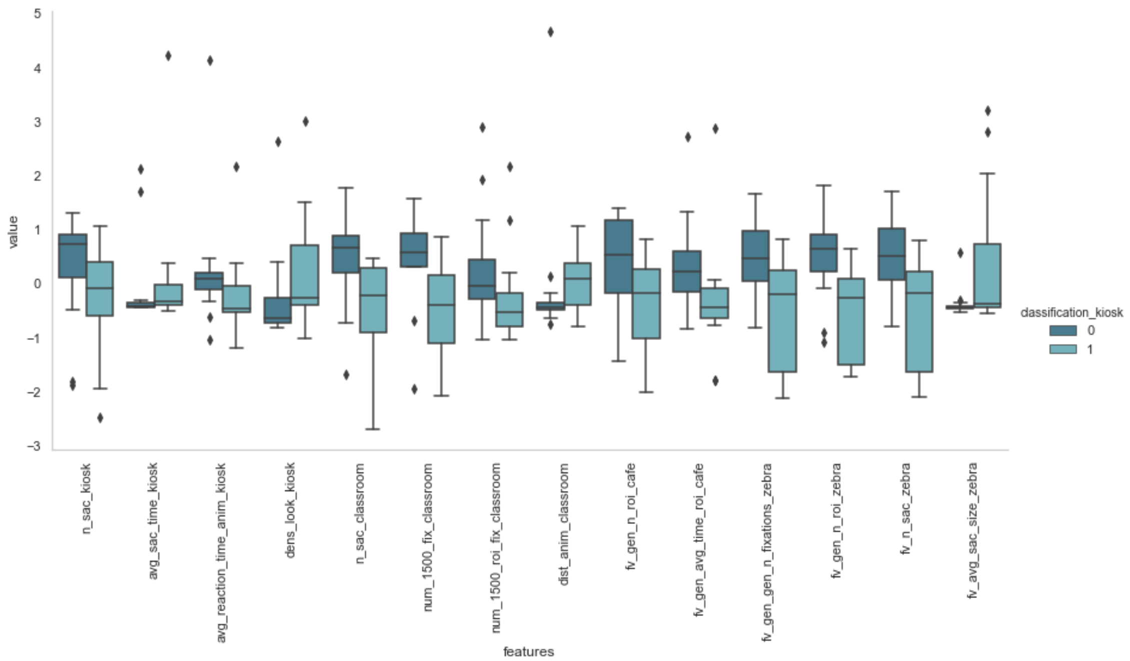

Figure 4.1: Boxplots of the features that passed Spearman's correlation test, with RED setup.

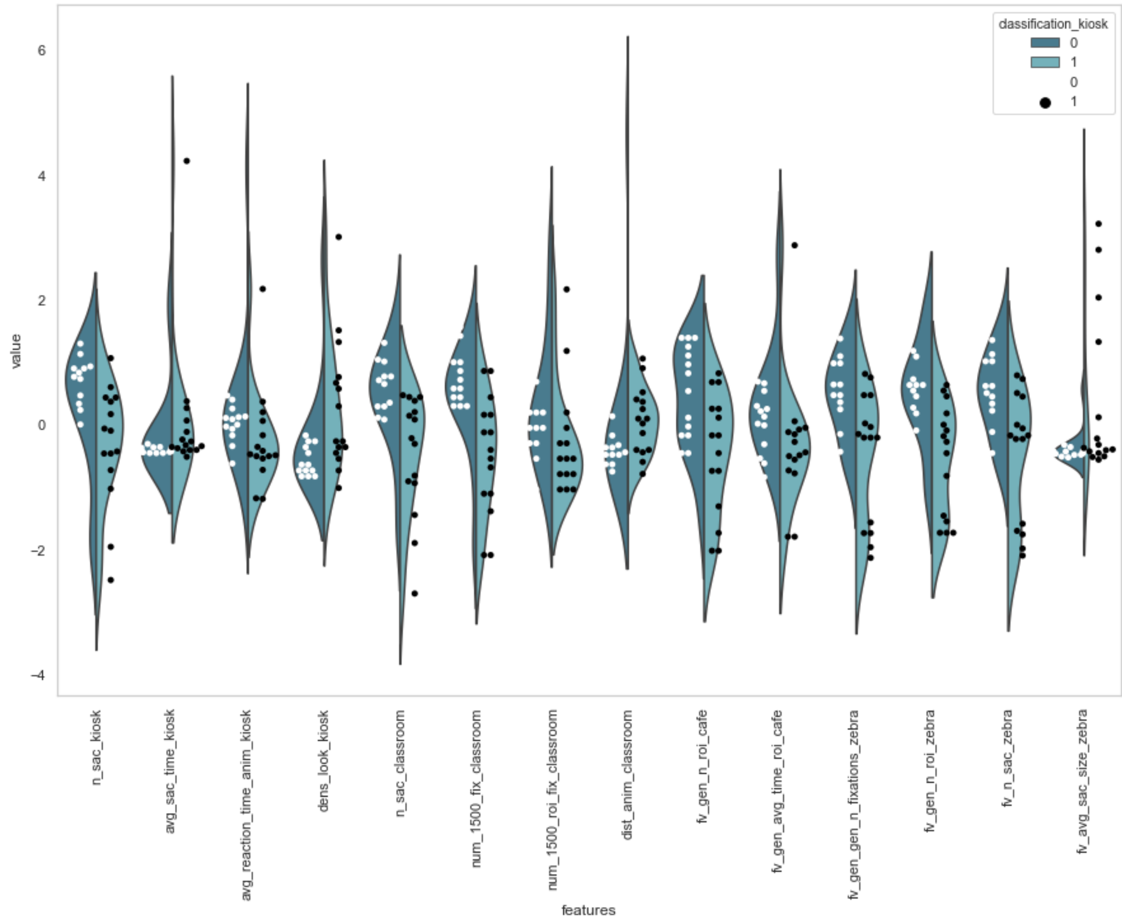

Figure 4.2: Violin plots with swarm plot values distribution, of the features that passed Spearman's correlation test, with RED setup

Similarly to RED setup, with the features obtained with the data from the Oculus Rift acquisition (figure 4.3) we can conclude that these features, from the data obtained with this setup, also show differences in the distribution, that might also be related with the better results in this step, for this setup. Here we highlight the differences in distribution of *sac_density_kiosk*, which represents the average number of coordinates variation during saccadic movement in the kiosk scenario.

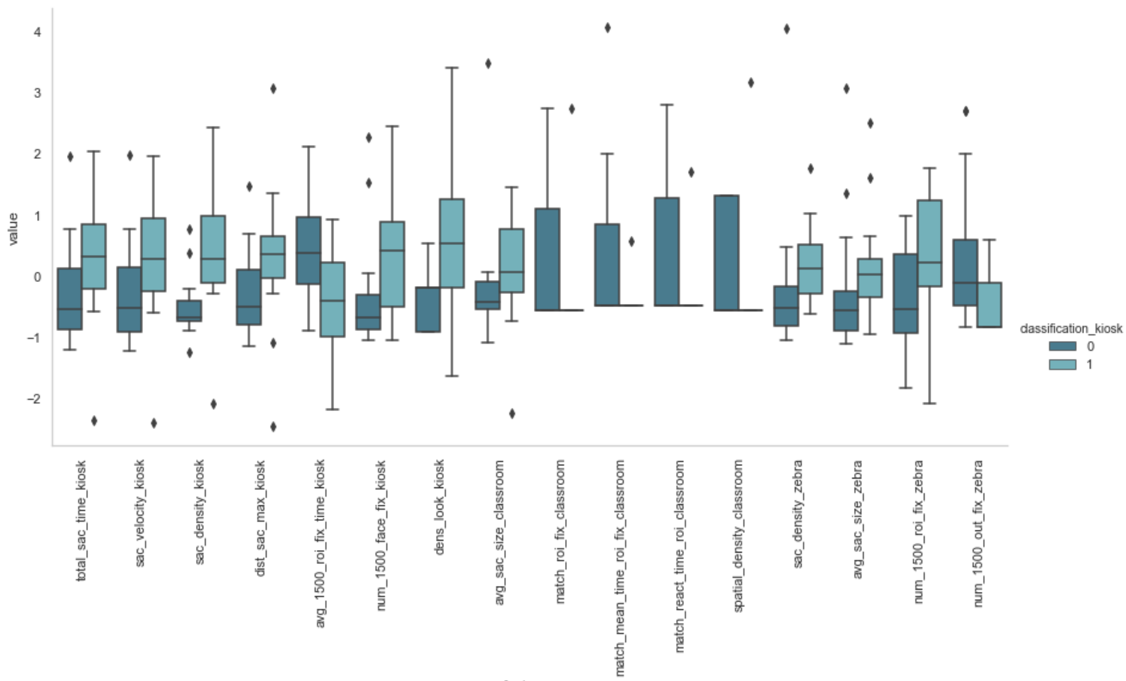

Figure 4.3: Boxplots of the features that passed Spearman's correlation test, with Oculus setup

If we compare both resulting sets of features, we notice that only *dens_look_kiosk*, was correlated with the classes in both systems. Which means that, the average number of coordinates' variation during the saccadic movement within the reaction time, after a social animation, have a high correlation with the classes of the data in both setups.

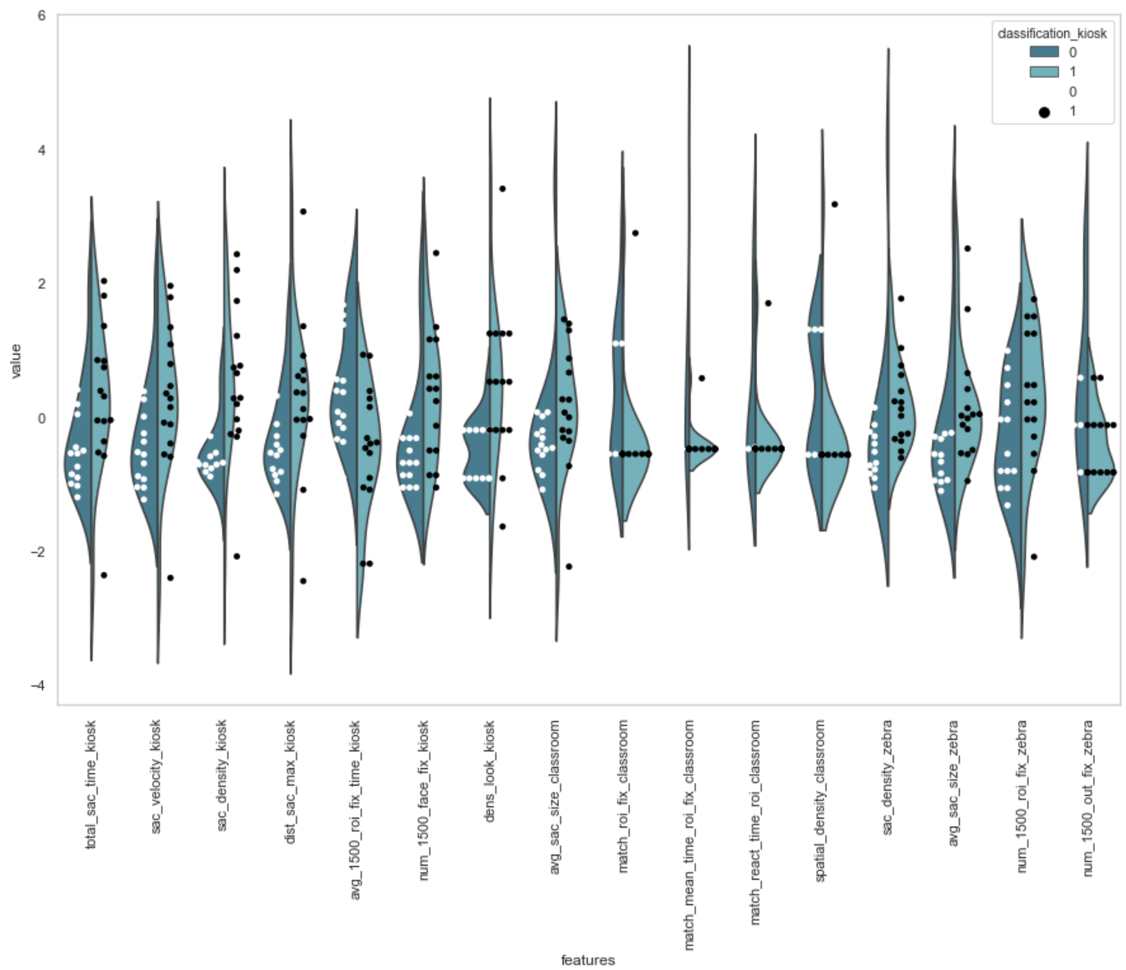

Figure 4.4: Violin plots with swarm plot values distribution, of the features that passed Spearman's correlation test, with Oculus setup

4.1.3 3rd step

The third phase of evaluation is composed by the implementation of different feature selection algorithms, mentioned in chapter 3. Those were linear regression, ridge regression, RFE, lasso, random forest and linear correlation, which we fed with the datasets with the features obtained after Spearman's correlation test.

After trying to combine different numbers of relevant features, we ended up choosing the 3 most relevant features for each method. Their combination resulted in the set of features mentioned in table 4.5 for RED setup, and in table 4.6 for Oculus setup.

In the data from RED setup we see a reduction from 14 to 9, and from 16 to 10 in the data from Oculus setup. We can see that the common feature, *dens_look_kiosk* is also present in both setups.

Table 4.5: Final set of features, after feature selection, for RED setup.

	RED
Features	avg_reaction_time_anim_kiosk dens_look_kiosk n_sac_classroom num_1500_fix_classroom dist_anim_classroom fv_gen_n_roi_cafe fv_gen_avg_time_roi_cafe fv_gen_gen_n_fixations_zebra fv_gen_n_roi_zebra fv_n_sac_zebra

Table 4.6: Final set of features, after feature selection, for Oculus Rift setup.

	Oculus Rift
Features	sac_velocity_kiosk sac_density_kiosk avg_1500_roi_fix_time_kiosk dens_look_kiosk avg_sac_size_classroom match_roi_fix_classroom match_mean_time_roi_fix_classroom spatial_density_classroom sac_density_zebra num_1500_out_fix_zebra

Besides this, the features that we considered having more differences between classes, in the previous step, *num_1500_fix_classroom* and *fv_gen_n_roi_zebra* for RED setup, and *sac_density_kiosk* for Oculus setup, also were also selected with these methods, composing, among others, the features we see in tables 4.5 and 4.6.

We also tested the classes classification with the SVM algorithm with both kernels. The results for this, can be seen in table 4.7.

The percentages for the evaluation metrics in this table (table 4.7) show once more that the classification is generally more accurate with a linear kernel. Nevertheless, here we see that the classification of the data acquired with oculus rift, with a RBF kernel, also show slightly higher evaluation percentages, higher than the values obtained with a linear kernel for the same setup, despite not being significant in terms of value (only 0.7%).

Table 4.7: Results after the remaining feature selection methods, for both linear and RBF kernels, where ASD is the ratio of well classified autistic individuals, among the total number of people within the correspondent class, and TD is the equivalent for the Typically Developing individuals class.

Setup	SVM Kernel	Evaluation Metrics						
		ASD	**TD**	**TPR**	**FPR**	**Accuracy**	**Error Rate**	**Precision**
Oculus	**Linear**	0.731	0.821	0.821	0.269	0.771	0.229	0.773
	RBF	0.580	0.436	0.436	0.420	0.395	0.604	0.247
RED	**Linear**	0.741	0.759	0.759	0.259	0.744	0.255	0.739
	RBF	0.671	0.895	0.895	0.329	0.751	0.249	0.701

Then again, the classification results are similar with both setups (77.1% of accuracy for RED and 74.4% of accuracy for Oculus). However, Oculus setup can give a better classification for TD individuals than the RED apparatus. And regardless of the setup, the models classify more correctly TD individuals than ASD individuals.

4.1.4 4th step

In this step we introduced PCA to the datasets, taking advantage from the variance between the classes. It works by creating new features that encapsulates high variance, from the given set of features.

In this step we used PCA in two different ways.

First, we obtained the number of features that combined, allowed us to have a variance of 95% between them. For both cases, the first component is always the one with higher variance, the second component has the second highest variance, and so on.

Figures 4.5 and 4.6 show the percentage of each original feature, in the making of the new components.

The former corresponds to the data from RED setup, where the given features are the ones in table 4.5. In this case the final number of ideal components is 5, and if we look at the first component, the one with higher variance, we see that *dist_anim_classroom, avg_reaction_time_anim_look* and *dens_look_kiosk* are the features with higher percentage in this new component.

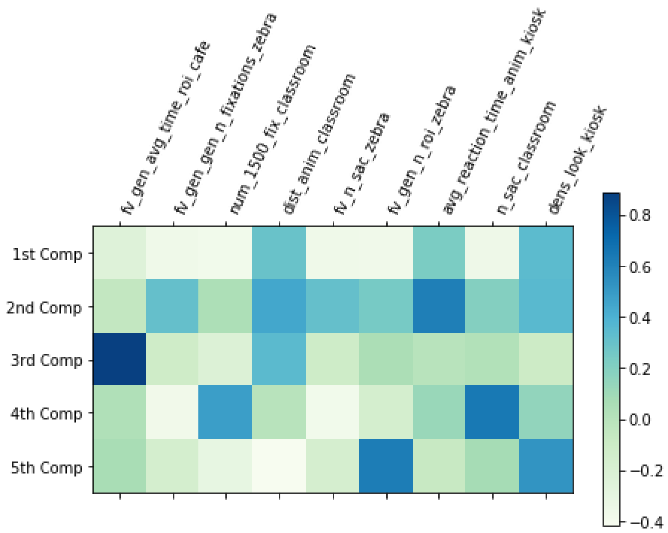

Figure 4.5: Ideal number of PCA components and their feature percentage, for RED setup.

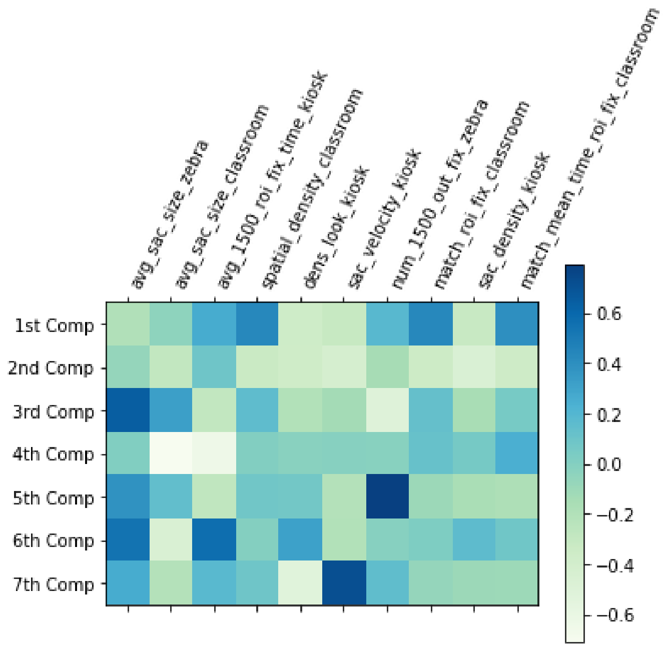

Figure 4.6: Ideal number of PCA components and their feature percentage, for Oculus setup.

In the latter figure (figure 4.6), which corresponds to the data obtained with oculus rift, we have 7 new components, recreated from the features in table 4.6. Here, the features that have higher percentage in the 1st component are *spatial_density_classroom*, *match_roi_fix_classroom* and *match_mean_time_roi_fix_classroom*.

With these new features, we trained once more our SVMs, obtaining the results shown in table 4.8.

Table 4.8: Results after PCA, for both linear and rbf kernels, where ASD is the ratio of well classified autistic individuals, among the total number of people within the correspondent class, and TD is the equivalent for the Typically Developing individuals class.

Setup	SVM Kernel	Evaluation Metrics						
		ASD	TD	TPR	FPR	Accuracy	Error Rate	Precision
Oculus	**Linear**	1.00	0.667	0.667	0	0.800	0.200	1.00
	RBF	1.00	0	0	0	0.400	0.600	0
RED	**Linear**	0.500	0.667	0.667	0.500	0.600	0.400	0.667
	RBF	1.0	0	0	0	0.400	0.600	0

Comparing the results for both setups with linear kernel, we see better results with the data acquired with Oculus rift. Here we obtained 100% in ASD classification and 66.7% in TD classification, which makes a model with 80% of accuracy.

Secondly, to see what was happening behind the PCA, we reduced the dimensions to two, obtaining plots with both classes, according to the transformed new components.

Figure 4.7 shows the plot of the two classes with the dataset obtained with RED setup, whilst figure 4.8 shows the same approach but with the dataset obtained from data acquired with Oculus rift.

Despite none of the plots showing a complete separability of the classes, we can see that figure 4.8, obtained from the data acquired with Oculus rift, shows a higher separability between TD and ASD individuals, with less cases with common information. This results are in agreement with what was obtained after training the SVM with PCA's new components. The components retrieved from the data obtained with Oculus rift show better results with the SVM and seem to be more easily separable in the plot with two components, than the data from RED setup.

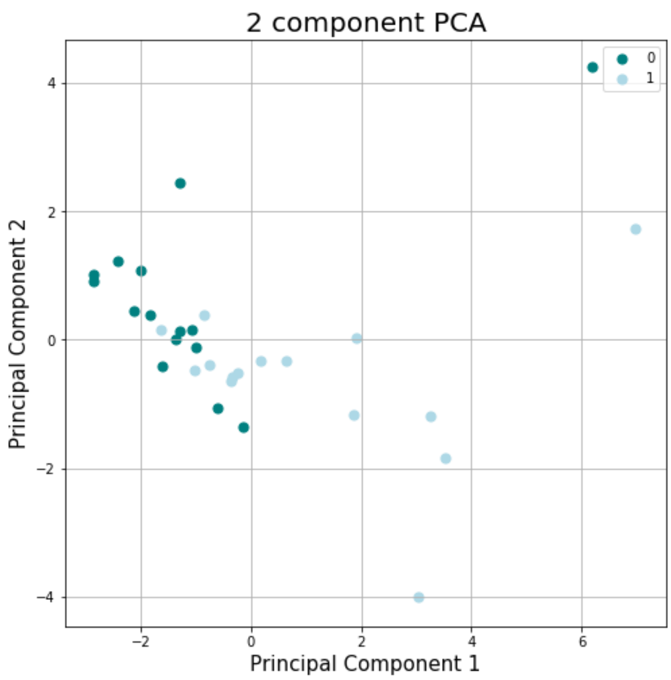

Figure 4.7: 2D PCA for RED setup.

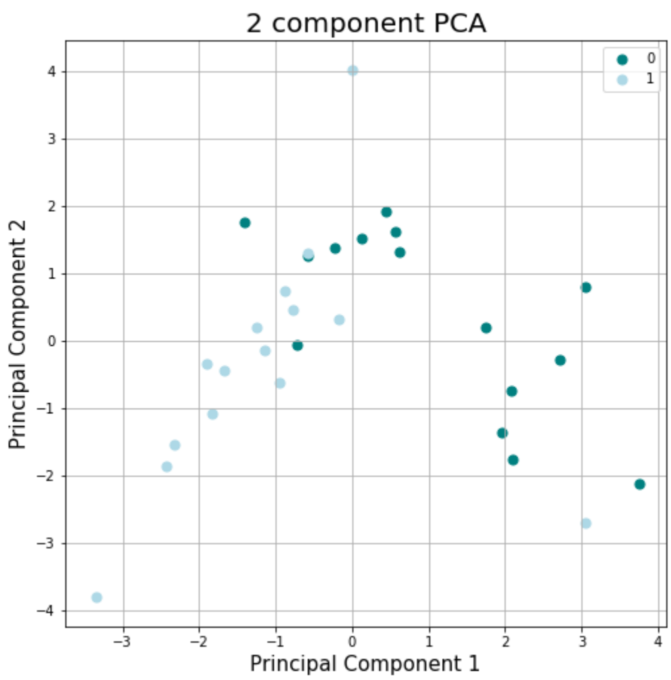

Figure 4.8: 2D PCA for Oculus setup.

5

Discussion

In this project we tried to answer different questions targetting final classification of ASD, which led us to our final approach. The answers to those questions started with the extraction of features from the data obtained with the two setups, RED and Oculus rift.

This extraction was followed by an analysis of the obtained dataset, divided in 4 steps:

- **1st step:** classification of the raw data from both setups, which was trained and tested with a SVM with linear kernel and a SVM with RBF kernel.
- **2nd step:** implementation of Spearman's correlation test in the data from the previous step. This culminated in new datasets with fewer and more relevant features, which were trained and tested with the SVMs with linear and RBF kernels.
- **3rd step:** implementation of different feature selection methods, in the final datasets obtained in the 2nd step. Once again, these new dataset were trained and tested with both SVMs.
- **4th step:** a PCA was used in the dataset from the previous step, this was followed by training and testing with SVM.

From the results obtained in each step, in chapter 4, we see that the SVM with the linear kernel is better suited to distinguish the classes, when compared with the SVM with the RBF kernel.

The summary of the best results for each step can be found in table 5.1.

Nevertheless, the first step and results do not give a viable mean of classification, due to the low percentages for the evaluation metrics.

We see a clear improvement of classification in each phase of the data acquired with

Table 5.1: Compilation of the best results for each step, where ASD is the percentage of well classified autistic individuals, among the total number of people within the correspondent class, and TD is the equivalent for the Typically Developing individuals class. 1st step - before any feature selection; 2nd step - after Spearman's correlation test; 3rd step - after implementing the remaining feature selection methods; 4th step - after PCA implementation.

Setup	Step	Kernel	Evaluation Metrics						
			ASD	TD	TPR	FPR	Accuracy	Error Rate	Precision
Oculus	**1st**	Linear	0.487	0.583	0.583	0.513	0.515	0.484	0.500
	2nd	Linear	0.709	0.743	0.743	0.291	0.702	0.298	0.736
	3rd	Linear	0.731	0.821	0.821	0.269	0.771	0.229	0.773
	4th	Linear	1.00	0.667	0.667	0	0.800	0.200	1.00
RED	**1st**	Linear	0.556	0.473	0.473	0.444	0.511	0.489	0.517
	2nd	Linear	0.716	0.707	0.707	0.284	0.698	0.302	0.727
	3rd	Linear	0.741	0.759	0.759	0.259	0.744	0.255	0.739
	4th	Linear	0.500	0.667	0.667	0.500	0.600	0.400	0.667

a 360^o immersivity. This evolution reveals the importance of feature selection done in each step of our approach. Interestingly, the results in the 4th and last step, show a full ability to classify autistic individuals, and a final model accuracy of 80%.

On the other hand, with RED setup, the results are slightly inferior. Curiously, the results from the 4th step are lower than the ones obtaining in the 2nd step. This suggests that PCA created a loss of information, which is coherent with the idea that the data retrieved from RED setup has less discriminative variance between the classes. This is corroborated by the worst performance of the classifier with the data from RED. We suggest that less immersivity in this setup might be influencing the individual's behaviour, creating less room to translate its true differences in the data.

Taking a closer look to the features and to the scenario they belong to, we notice that there is no coherence among the selected features in each step and in each scenario. The implemented feature selection techniques gave importance to different metrics, and this rejects, once more, the hypothesis of the existence of a single feature, despite the social context, that works better as a biomarker. All together, it seams that the social context influences not only the behaviour of each person, and thus the features with more relevance in that specific context.

Analysing the selected features from both setups (in tables 4.5 and 4.6) we see metrics related to both fixations and saccades, but only with RED setup we have selected features from the free viewing part of the acquisition. We suppose that a higher immersivity, only possible with Oculus rift, is related with higher dispersion of the eye gaze and that the differences in the data from RED are harder to explain from the cognitive neuroscience point of view, due to the reduced immersivity.

According to figures 4.1 and 4.3, of the selected features from the RED and Oculus setups, we see that ASD individuals have a tendency to show more saccades, but with lower velocity and duration, which is according to what was reported [37][38]. They also show a higher reaction time, since the animation occur until they look at the correspondent ROI, nevertheless, they do this lesser than TD individuals, meaning that the JA phenomena occurs less times in these individuals.

Regarding the features related with the fixations within the ASD class, we see a tendency for less fixations in the avatar's face, and a higher number of fixations in the region outside the scenario. Besides this, the tendency is also for longer fixations, when they happen. All of these results were accordingly to what was expected and found in previous studies [28][29][30][31].

Separating both setups, we notice that the selected features from the RED setup come closer to the literature review, and that there was surprisingly not much coherence among those features and the ones resultant from the data acquired with Oculus rift. Since the majority of the studies with eyetracking and ASD have been done with screen-based eye tracking, like RED setup, and the model with the data from Oculus rift performed better.

6

Conclusions

Autism Spectrum Disorder, like previously described, is a neurodevelopmental disorder, characterized by deficits in social behaviour, which include difficulties in making eye contact and an impairment in understanding and using body language and gestures [2].

By making use of virtual reality experiments, that can mimic real life situations, we can put to test the behaviour of ASD individuals, and record and analyse their eye tracking data or, in other words, their behaviour during situations with social context.

The aim of this project was to analyse these eye tracking data, obtained during virtual reality experiments with people with and without ASD, which was acquired during the work of Carlos Amaral *et. al* [64].

This project followed three main questions:

1. Which features can we obtain from data containing eye positions and their respective time stamps?
2. Is any of those features "good" enough to define a potential biomarker for ASD?
3. Can we predict or confirm the disorder, by analysing eyetracking data with a machine learning approach?

The answer to the first question, might have been the most challenging part of this project. Understand what could be meaningful in this data, and how we could retrieve it from the existent files, made up a big and important part of our simple, yet interesting, analysis.

We were able to retrieve a total of 53 features from the data, considering their relevance in previous eye tracking studies and in ASD. They were divided in features acquired during the free viewing part of the experiment, and in features acquired after the first avatar's animation, which represented the starting of the part with a social context. The major difference imposed in these two parts were the time it took to consider a fixation, due to the different complexity of each part.

The retrieved features were then related with fixations and saccades, the two most considered and relevant movements in eye gaze and in its study. The obtained and selected features after all the steps includes the average number of saccades and their variation in space and time, the total number of fixations, the average reaction time it took for the individual to look at the ROI after the avatar's animation, and the total distance traveled during saccadic movement, among others. We analysed each feature individually, without considering the relation that might exist between them. For example, reaction time might not be relevant when considered alone, but it might be related with the animation that preceded it and with the fixation and its time that might succeed it. This increase in complexity in feature extraction can give ground to future work, and approach the complexity of how we behave in social contexts.

From all those features, we were not able to select one that could work as a biomarker for ASD. Despite showing differences between both groups, with none of the features we were able to clearly state that it could separate those groups. Retrospectivly, one might argue that considering individually each feature one will not explain the complexity of ASD, which is proved by our results.

The results showed that, after feature selection and the implementation of PCA, a SVM with a linear kernel and with the data acquired from the Oculus rift setup, could give a reasonable classification of ASD individuals. Despite the inspiring results, this is far from replacing the current diagnostic tools and tests, neither the important role of the trained clinician in all the process.

There were some limitations to our work. This approach does not allow to assess the effects of IQ or severity levels within ASD. Since orienting to social cues and all its underlying neurological processes are considered extremely complex, this approach could suit as a simple starting point to improve the way we analyse and interpret eye tracking data.

Since the data were acquired with two different setups, we notice that the results obtained with a 360^{o} immersity, which give more viewing freedom to the individual, are the best results. These results, and their differences from the obtained with RED, suggest that more studies have to be done with higher immersivity, since they can give us more and different information that screen-based eye tracking could not.

These encouraging results lead us to the future work that can be done in this field.

6.1 Future work

We can divide future work in different parts.

Let us start with the already mentioned limitations. Since our approach was proposed after the data acquisition, we had no control in the way and how the data were obtained. So, the first step in future work might be designing an experiment, and acquiring data that fits better our needs. For example, some findings show that individuals with ASD might exhibit a higher level of impairment in the perception of auditory information, rather than in the perception of non-speech stimuli [68][69].

Since this was a simple approach, other thing we can do is increase the complexity of the features we retrieve from the data. We already noticed differences when considering each parameter as a feature alone, nevertheless, they do not happen isolated from the remaining. So, the relationship between the different metrics might approach what really happens behind our eyes, and might help us understand better human behaviour. Besides this, it might be interesting to complement eye tracking data with other types of data, like EEG signals.

During this project, we tried a more linearly approach, leaving room to the implementation and testing of non-linear methods, which might fit better the data. Since there is no standardized way to analyse data, we are left with several approaches that might be worth trying.

Bibliography

[1] S. Baron-Cohen, "Leo kanner, hans asperger, and the discovery of autism," *The Lancet*, vol. 386, no. 10001, p. 1329–1330, 2015.

[2] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, vol. 189. American Psychiatric Association, may 2013.

[3] "National Institute of Mental Health - Autism Spectrum Disorder." =https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd/index.shtml. Accessed: 2019-06-13.

[4] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism," *Archives of General Psychiatry*, vol. 59, no. 9, p. 809, 2002.

[5] Z. Boraston and S.-J. Blakemore, "The application of eye-tracking technology in the study of autism," *The Journal of Physiology*, vol. 581, no. 3, p. 893–898, 2007.

[6] D. P. Wall, J. Kosmicki, T. F. Deluca, E. Harstad, and V. A. Fusaro, "Use of machine learning to shorten observation-based screening and diagnosis of autism," *Translational Psychiatry*, vol. 2, no. 4, 2012.

[7] "Gaze cueing of attention: Visual attention, social cognition, and individual differences.," *Psychological Bulletin*, vol. 133, pp. 694–724, jul 2007.

[8] S. Baron-Cohen, A. Cox, G. Baird, J. Swettenham, N. Nightingale, K. Morgan, A. Drew, and T. Charman, "Psychological markers in the detection of autism in infancy in a large population," *Early Human Development*, vol. 47, no. 1, p. 98–99, 1997.

[9] S. Baron-Cohen, J. Allen, and C. Gillberg, "Can autism be detected at 18 months?," *British Journal of Psychiatry*, vol. 161, no. 6, p. 839–843, 1992.

[10] S. Baron-Cohen, "The autistic child's theory of mind: a case of specific developmental delay," *Journal of Child Psychology and Psychiatry*, vol. 30, no. 2, pp. 285–297, 1989.

[11] T. Charman, "Why is joint attention a pivotal skill in autism?," no. January, pp. 315–324, 2003.

[12] Guiomar Gonçalves de Oliveira, "EPIDEMIOLOGIA DO AUTISMO EM PORTUGAL Um estudo de prevalência da perturbação do espectro do autismo e de caracterização de uma amostra populacional de idade escolar Gu," 2005.

[13] "Neurodevelopmental disorders - Autism Spectrum Disorder." =https://ourworldindata.org/neurodevelopmental-disordersl. Accessed: 2019-06-15.

[14] D. R. Simmons, A. E. Robertson, L. S. McKay, E. Toal, P. McAleer, and F. E. Pollick, "Vision in autism spectrum disorders," *Vision Research*, vol. 49, no. 22, pp. 2705 – 2739, 2009. Vision Research Special Issue.

[15] A. T. Duchowski, *Eye tracking methodology: theory and practice.* Springer, 2007.

[16] S. A. Mcmains and S. Kastner, "Visual attention," *Encyclopedia of Neuroscience*, p. 4296–4302, 2009.

[17] A. Frischen, A. P. Bayliss, and S. P. Tipper, "Gaze cueing of attention: Visual attention, social cognition, and individual differences.," *Psychological Bulletin*, vol. 133, no. 4, p. 694–724, 2007.

[18] M. Das, D. M. Bennett, and G. N. Dutton, "Visual attention as an important visual function: an outline of manifestations, diagnosis and management of impaired visual attention," *British Journal of Ophthalmology*, vol. 91, no. 11, p. 1556–1560, 2007.

[19] J. Colombo, "The development of visual attention in infancy," *Annual Review of Psychology*, vol. 52, no. 1, p. 337–367, 2001.

[20] J. E. Richards, *The development of visual attention and the brain.* Psychology Press.

[21] B. Keehn, R.-A. Müller, and J. Townsend, "Atypical attentional networks and the emergence of autism," *Neuroscience Biobehavioral Reviews*, vol. 37, no. 2, p. 164–183, 2013.

[22] R. Bakeman and L. Adamson, "Coordinating attention to people and objects in mother-infant and peer-infant interaction," *Child development*, vol. 55, pp. 1278–89, 09 1984.

[23] M. Tomasello, A. C. Kruger, and H. H. Ratner, "Cultural learning," no. September, 1993.

[24] A. P. Bayliss, M. A. Paul, P. R. Cannon, and S. P. Tipper, "Gaze cuing and affective judgments of objects: I like what you look at," *Psychonomic Bulletin Review*, vol. 13, no. 6, p. 1061–1066, 2006.

[25] P. Mundy, M. Sigman, J. Ungerer, and T. Sherman, "Defining the social deficits of autism: The contribution of non-verbal communication measures," *Journal of Child Psychology and Psychiatry*, vol. 27, no. 5, p. 657–669, 1986.

[26] J.-A. Little, "Vision in children with autism spectrum disorder: a critical review," *Clinical and Experimental Optometry*, vol. 101, p. 504–513, Nov 2018.

[27] L. Lonnie Zwaigenbaum and M. Penner, "Autism spectrum disorder: advances in diagnosis and evaluation," *BMJ*, May 2018.

[28] C. Stefanidou, *The neurodevelopmental basis of human action sound processing in typically developing children, children with autism spectrum disorders, and toddlers at risk for autism.* PhD thesis.

[29] G. Chasson and S. R. Jarosiewicz, "Social competence impairments in autism spectrum disorders," *Comprehensive Guide to Autism*, p. 1099–1118, 2014.

[30] G. Dawson, A. N. Meltzoff, J. Osterling, J. Rinaldi, and E. Brown, "Children with autism fail to orient to naturally occurring social stimuli," *Journal of Autism and Developmental Disorders*, 1998.

[31] G. Dawson, K. Toth, R. Abbott, J. Osterling, J. Munson, A. Estes, and J. Liaw, "Early social attention impairments in autism: Social orienting, joint attention, and attention to distress.," *Developmental Psychology*, vol. 40, no. 2, p. 271–283, 2004.

[32] S. Magrelli, P. Jermann, B. Noris, F. Ansermet, F. Hentsch, J. Nadel, and A. Billard, "Social orienting of children with autism to facial expressions and speech: a study with a wearable eye-tracker in naturalistic settings," *Frontiers in Psychology*, vol. 4, 2013.

[33] L. Zwaigenbaum, S. Bryson, C. Lord, S. Rogers, A. Carter, L. Carver, K. Chawarska, J. Constantino, G. Dawson, K. Dobkins, and et al., "Clini-

cal assessment and management of toddlers with suspected autism spectrum disorder: Insights from studies of high-risk infants," *Pediatrics*, vol. 123, no. 5, p. 1383–1391, 2009.

[34] D. Kliemann, I. Dziobek, A. Hatri, R. Steimke, and H. R. Heekeren, "Atypical reflexive gaze patterns on emotional faces in autism spectrum disorders," *Journal of Neuroscience*, vol. 30, no. 37, p. 12281–12287, 2010.

[35] M. B. Harms, A. Martin, and G. L. Wallace, "Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies," *Neuropsychology Review*, vol. 20, no. 3, p. 290–322, 2010.

[36] C. D. Holland and O. V. Komogortsev, "Complex eye movement pattern biometrics: Analyzing fixations and saccades," *2013 International Conference on Biometrics (ICB)*, 2013.

[37] C. Kemner, M. N. Verbaten, J. M. Cuperus, G. Camfferman, and H. V. Engeland, "Abnormal saccadic eye movements in autistic children.," *Journal of Autism and Developmental Disorders*, vol. 28, no. 1, p. 61–67, 1998.

[38] M. Goldberg, A. Lasker, D. Zee, E. Garth, A. Tien, and R. Landa, "Deficits in the initiation of eye movements in the absence of a visual target in adolescents with high functioning autism," *Neuropsychologia*, vol. 40, no. 12, p. 2039–2049, 2002.

[39] R. M. Joseph, B. Keehn, C. Connolly, J. M. Wolfe, and T. S. Horowitz, "Why is visual search superior in autism spectrum disorder?," *Developmental Science*, vol. 12, no. 6, p. 1083–1096, 2009.

[40] L. A. Brenner, K. C. Turner, and R.-A. Muller, "Eye movement and visual search: Are there elementary abnormalities in autism?," *Journal of Autism and Developmental Disorders*, vol. 37, no. 7, p. 1289–1309, 2006.

[41] T. Falck-Ytter, E. Fernell, L. Hedvall, C. V. Hofsten, and C. Gillberg, "Gaze performance in children with autism spectrum disorder when observing communicative actions," *Journal of Autism and Developmental Disorders*, vol. 42, no. 10, p. 2236–2245, 2012.

[42] G. Vivanti, C. Mccormick, G. S. Young, F. Abucayan, N. Hatt, A. Nadig, S. Ozonoff, and S. J. Rogers, "Intact and impaired mechanisms of action understanding in autism.," *Developmental Psychology*, vol. 47, no. 3, p. 841–856, 2011.

[43] E. Thorup, J. L. Kleberg, and T. Falck-Ytter, "Gaze following in children with autism: Do high interest objects boost performance?," *Journal of Autism and Developmental Disorders*, vol. 47, no. 3, p. 626–635, 2016.

[44] Z. A. Almaliki, "Do you know how to choose the right machine learning algorithm among 7 different types?," Jul 2019.

[45] B. Farnsworth, "10 most used eye tracking metrics and terms." = https://imotions.com/blog/7-terms-metrics-eye-tracking, Aug 2018. Accessed: 2019-06-18.

[46] C. Wilke, *Fundamentals of data visualization: a primer on making informative and compelling figures.* OReilly Media, 2019.

[47] "Dimensionality reduction algorithms: Strengths and weaknesses," Jan 2019.

[48] "Feature selection – part i: univariate selection."

[49] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Computing and Applications*, vol. 24, pp. 175–186, Jan 2014.

[50] *Feature Extraction Foundations and Applications.* Springer-Verlag New York Inc, 2016.

[51] J. W. T. Navin Lal, O. Chapelle and A. Elissee, "Embedded methods,"

[52] J. Hauke and T. Kossowski, "Comparison of values of pearsons and spearmans correlation coefficients on the same sets of data," *Quaestiones Geographicae*, vol. 30, p. 87–93, Jan 2011.

[53] S. Swaminathan, "Linear regression - detailed view," Jan 2019.

[54] "Selecting good features – part ii: linear models and regularization."

[55] S. Raschka and V. Mirjalili, *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow.* Packt Publishing, 2017.

[56] J. B. Marshall, "Decision trees."

[57] J. Kho, "Why random forest is my favorite machine learning model," Mar 2019.

[58] "Selecting good features – part iv: stability selection, rfe and everything side by side."

[59] "Artificial intelligence machine learning: Policy paper." Accessed: 2019-08-15.

[60] "Modern machine learning algorithms: Strengths and weaknesses," Jan 2019.

[61] T. J. M. Cleophas and A. H. Zwinderman, *Machine learning in medicine.* Springer, 2015.

[62] Saptashwa, "Support vector machine: Kernel trick; mercer's theorem," Jul 2019.

[63] Y. Ma and G. Guo, *Support vector machines applications.* Springer, 2014.

[64] C. Amaral, S. Mouga, M. Simões, H. C. Pereira, I. Bernardino, H. Quental, R. Playle, R. Mcnamara, G. Oliveira, M. Castelo-Branco, and et al., "A feasibility clinical trial to improve social attention in autistic spectrum disorder (asd) using a brain computer interface," *Frontiers in Neuroscience*, vol. 12, 2018.

[65] K. Rayner, "The 35th sir frederick bartlett lecture: Eye movements and attention in reading, scene perception, and visual search," *Quarterly Journal of Experimental Psychology*, vol. 62, no. 8, p. 1457–1506, 2009.

[66] "Grid search for model tuning," Dec 2018. Accessed: 2019-07-29.

[67] "Scikit-learn." Accessed: 2019-07-29.

[68] A. Klin, "Young autistic childrens listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal," *Journal of Autism and Developmental Disorders*, vol. 21, no. 1, p. 29–42, 1991.

[69] A. J. Whitehouse and D. V. Bishop, "Do children with autism 'switch off' to speech sounds? an investigation using event-related potentials," *Developmental Science*, vol. 11, no. 4, p. 516–524, 2008.